ISSN 1977-0375

eurostat
Methodologies and Working papers

An assessment of survey errors in EU-SILC

2010 edition

eurostat
EUROPEAN COMMISSION

eurostat
Methodologies and
Working papers

An assessment of survey errors in EU-SILC

2010 edition

eurostat
EUROPEAN COMMISSION

More information on the European Union is available on the Internet (http://europa.eu).

Cataloguing data can be found at the end of this publication

Luxembourg: Publications Office of the European Union, 2010

ISBN 978-92-79-16757-7
ISSN 1977-0375
Doi:10.2785/56062
Cat. No. KS-RA-10-021-EN-N

Theme: Population and social conditions
Collection: Methodologies and working papers

Eurostat is the Statistical Office of the European Union (EU). Its mission is to provide the EU with high-quality statistical information. To that end, it gathers and analyses data from the National Statistical Institutes (NSIs) across Europe and provides comparable and harmonised data for the EU to use in the definition, implementation and analysis of EU policies. Its statistical products and services are also of great value to Europe's business community, professional organisations, academics, librarians, NGOs, the media and citizens. In the social field, the EU Statistics on Income and Living Conditions (EU-SILC) instrument is the main source for statistics on income, poverty, social exclusion and living conditions.

Over the last years, important progress has been made in EU-SILC. This is the result of the coordinated work of Eurostat and the NSIs, *inter alia* in the context of the EU 'Living Conditions' Working Group and various thematic Task-Forces. Despite these significant achievements, EU-SILC data are still insufficiently analysed and used.

It is in this context that Eurostat launched in 2008 a call for applications with the following aims:

(1) develop methodology for advanced analysis of EU-SILC data;
(2) discuss analytical and methodological papers at an international conference;
(3) produce a number of publications presenting methodological and analytical results.

The 'Network for the Analysis of EU-SILC' (Net-SILC), an ambitious 18-partner Network bringing together expertise from both data producers and data users, was set up as in response to this call. The initial Net-SILC findings were presented at the international conference on 'Comparative EU Statistics on Income and Living Conditions' (Warsaw, 25-26 March 2010), which was organised jointly by Eurostat and the Net-SILC network and hosted by the Central Statistical Office of Poland. A major deliverable from Net-SILC is a book to be published by the EU Publications Office at the end of 2010 and edited by Anthony B. Atkinson (Nuffield College and London School of Economics, United Kingdom) and Eric Marlier (CEPS/INSTEAD Research Institute, Luxembourg).

The present methodological paper is also an outcome from Net-SILC. It has been prepared by Vijay Verma, Gianni Betti and Francesca Gagliardi (University of Siena, Italy). Gara Rojas González was responsible at Eurostat for coordinating the publication of the methodological papers produced by Net-SILC members.

It should be stressed that this methodological paper does not in any way represent the views of Eurostat, the European Commission or the European Union. The authors have contributed in a strictly personal capacity and not as representatives of any Government or official body. Thus they have been free to express their own views and to take full responsibility both for the judgments made about past and current policy and for the recommendations for future policy.

This document is part of Eurostat's *Methodologies and working papers* collection which are technical publications for statistical experts working in a particular field. All publications are downloadable free of charge in PDF format from the Eurostat website:
(http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/publications/Methodologies_and_working_papers). Furthermore, Eurostat databases are freely available at this address, as are tables with the most frequently used and requested short- and long-term indicators.

Table of contents

An assessment of survey errors in EU-SILC

Vijay Verma, Gianni Betti and Francesca Gagliardi,
University of Siena, Italy

Abstract: This Working Paper provides a broad assessment of data accuracy and survey errors in EU-SILC, taking comparability of the results across the national surveys as a basic requirement. A typology of errors in survey data is proposed, distinguishing various categories of errors in measurement on the one hand, and errors arising in the process of estimation from the survey sample to the target population on the other. Important sources and types of errors – including conceptual, measurement, and unit and item non-response errors – are explained and illustrated in empirical terms using EU-SILC microdata. Examples are provided of the impact on comparability of national differences in data collection and data treatment procedures. Methodology of, and practical problems in, computing sampling errors and design effects for the complex statistics on income and living conditions which can be constructed from EU-SILC are explained and illustrated in detail. The paper concludes with recommendations on improving the potential for the assessment by researchers and other users of data quality in EU-SILC.

Key words: data quality, survey errors, measurement errors, non-response, variance estimation, design effect, EU-SILC

1. Introduction: a description of errors in survey data

This Working Paper analyses sampling and non-sampling errors in EU-SILC, examining the impact of these errors on comparability across countries and over time. This introductory section provides a typology of survey errors to set the framework for subsequent discussion. Basic aspects of the EU-SILC integrated design are also recapitulated. In the following sections major components and sources of error in EU-SILC are examined in-depth and empirically. Conceptual and measurement errors, unit and item non-response errors, and sampling errors are considered in turn in Sections 2-4.

Quite a number of studies on various aspects of data quality in EU-SILC have been carried out by researchers. These studies tend to focus in-depth on specific problems and issues, and may be specific to particular countries or groups of countries. They also mostly concern ‘non-sampling’ errors (as distinct from sampling variances), especially what we defined below as ‘errors in measurement’. There is little research or information available hitherto on sampling errors for the diverse indicators produced for different subpopulations in EU-SILC, in particular information which is presented and discussed from a comparative prospective. One likely reason for this is the lack of expertise available for the computation and analysis of sampling errors in a survey such as EU-SILC, with its complex sample structure and types of indicators to be estimated.

For these reasons we have chosen in this study to pay additional attention to sampling errors – not simply to the production of large sets of figures covering many variables and countries, but to an exposition of the methodology of variance estimation and the possibilities and problems of its implementation to EU-SILC given the various constraints (Section 5). We also present procedures for computing design effects (Section 6), which measure the relative efficiency of samples, and are also essential for computing ‘effective sample sizes’ as required by EU-SILC regulations.

An important concern of this study is to explore and expose the barriers which researchers, using the restricted information provided in UDB and EU-SILC documentation in the public domain, face in assessing quality of the data. This issue is important for proper use of the data and for the development and improvement of EU-SILC itself, and needs to be brought out prominently (Section 7).

1.1 A typology of errors

Knowledge about data quality is required for their proper use and interpretation. Also, measures of data quality are important for the evaluation and improvement of survey design and procedures. Continued monitoring and improvement of data quality is particularly important in major continuing surveys such as EU-SILC.

There are diverse forms and many different sources of errors in surveys, and various frameworks have been proposed for their classification. Different frameworks emphasise different aspects of the problem. None may be considered as ‘the best’, though some frameworks are more illuminating than others. The following framework is drawn from Verma (1981), further elaborated in Hussmanns *et al* (1990). This framework distinguishes between two groups of errors affecting the survey process:

(a) Errors in measurement

These arise from the fact that what is measured on the units included in the survey can depart from the actual (true) values for those units. These errors concern the accuracy of measurement at the level of individual units enumerated in the survey, and centre on *substantive content of the survey*: definition of the survey objectives and questions; ability and willingness of the respondent to provide the information sought; and the quality of data collection, recording and processing. This group of errors can be studied in relation to various stages of the survey operation.

(b) Errors in estimation

These are errors in the process of extrapolation from the particular units enumerated in the survey to the entire study population for which estimates or inferences are required. These centre on the process of *sample design and implementation*, and include errors of coverage, sample selection and implementation, non-response, and also sampling errors and estimation bias.

The above categorisation, in terms of *errors in measurement* and *errors in estimation*, is more fundamental than the distinction usually made between *sampling* and *non-sampling* errors.

In Figure 1 a third category, namely *item non-response*, has been added as an intermediate category between measurement and estimation errors.

Each group of errors may be further classified in more detail in order to identify specific sources of error, so as to facilitate their assessment and control. However, it is important to note that the various phases of a survey are closely related. While it is useful to classify the total survey error into components, *errors cannot always be attributed to a particular type or source*. The same or similar methods of assessment and control may be suited for measuring more than one type of error, and some of the indicators obtained may provide no more than a general or overall measure of data accuracy without being able to identify specific sources and types of error.

1.2 Errors in measurement

As noted, the broad range of 'errors in measurement' may be classified by source, for example as conceptual, response ('data collection') and processing errors. *Conceptual errors* concern the scope, concepts, definitions and classifications adopted in relation to the survey objectives, and are the most fundamental ones. The distinction between *response errors* concerning the process of data collection, and *processing errors* concerning the subsequent process of transforming the information into a micro database, is a useful one from the point of survey operations and methods of assessing and controlling these errors. Despite this operational distinction, however, the two classes of error are conceptually quite similar.

Various components of measurement error may be distinguished. Further operational classification within each category may be introduced. Each type of error may be decomposed into bias and variance components. These distinctions are useful in so far as the components differ in nature and in methods of assessment and control.

Measurement bias

A part of the error is common to the work of all interviewers (or coder etc.); this gives rise to *response bias*, i.e. more or less systematic errors in obtaining the required information. Bias arises from shortcomings affecting the whole survey operation: basic conceptual errors in defining and implementing the survey content; incorrect instructions affecting all the survey workers; errors in the coding frame or programs for processing the data, etc. Errors also arise from inherent difficulties in collecting certain types of information, more or less independently of the specific technical design and procedures of the survey, given the general social situation and the type of respondents involved.

The first step in identifying bias is through logical and substantive analysis of the internal consistency of the data. Beyond that, the assessment requires comparison with more accurate information: data from external sources and/or data collected with special, improved methods. There are several possibilities in connection with bias assessment. For instance, the study of response bias may involve two interviews on a subsample following the original interview: a *re-interview*, which is an independent replication of the original interview and is aimed at measuring response variance, followed in discrepant cases by a *reconciliation interview* aimed at establishing correct responses and identifying biases and their sources.

Measurement variance

This refers to variable errors in data collection and processing.

In addition to biases common to the whole operation, each interviewer has his/her own particular bias, which affects the interviewer's whole workload. This gives rise to *correlated response variance*, which indicates a lack of uniformity and standardisation in the interviewers' work. By contrast, *simple response variance* is random, not correlated with any particular interviewer. It is an indicator of the inherent instability of particular items in the questionnaire. Its high value indicates the need for better training and supervision of survey work. Its measurement requires comparisons between independent repetitions of the survey under the same general conditions - there is no way, in a single survey, to distinguish between variation among the true values of units (which gives rise to sampling error), and the additional variability arising from random factors affecting individual responses.

1.3 Errors in estimation

Coverage and related errors

Coverage errors arise from discrepancies between the target and the frame populations, and also from errors in selecting the sample from the frame. The condition of ‘probability sampling’ is violated if: (a) the survey population is not fully and correctly represented in the sampling frame; (b) the selection of units from the frame into the sample is not random with known non-zero probabilities for all units; or (c) not all the units selected into the sample are successfully enumerated. Coverage error concerns primarily (a), but also (b); (c) concerns non-response.

Figure 1: Types of errors in surveys

Errors in measurement

1 conceptual errors

- errors in basic concepts, definitions and classifications
- errors in putting them into practice (questionnaire design, preparation of survey manuals, training and supervision of interviewers and other survey workers)

2 response (or 'data collection') errors

- response bias
- simple response variance
- correlated response variance

3 processing errors

- recording, data entry and coding errors
- editing errors
- errors in constructing target variables
- other programming errors

Mixed category

4 item non-response

- only approximate or partial information sought in the survey
- respondents unable to provide the information sought ('don't knows')
- respondents not willing to provide the information ('refusals')
- information suppressed (for confidentiality or whatever reason)

Errors in estimation

5 coverage and related errors

- under-coverage
- over-coverage
- sample selection errors

6 unit non-response

- unit not found or inaccessible
- not-at-home
- unable to respond
- refusal (potentially 'convertible')
- 'hard core' refusal

7 sampling error

- sampling variance
- estimation bias

Non-sampling errors = 1 to 6

+ Comparability, underscoring all aspects of data accuracy

Adapted from Hussmanns *et al* (1990)

Non-response errors

Non-response refers to the failure to obtain a measurement on one or more study variables for one or more sample units. When a whole unit is missed, we have *unit non-response*. When a unit is included but information on some items for it is missed, we have *item non-response*. Non-response causes an increase in variance due to decreased effective sample size and/or due to weighting and imputation introduced to control its impact. More importantly, it causes bias in so far as non-respondents are selective with respect to the characteristic being measured. For instance, one might expect persons with high incomes to be more reluctant to give information on their income; similarly, poorer, unemployed and socially excluded persons are more likely to be missed in surveys such as EU-SILC. Classification of unit non-response according to the reasons or circumstances giving rise to it can be very helpful for identifying and controlling the extent of non-response and assessing its impact. It is most useful when the categories are designed to capture the most important factors in the particular survey, are not too numerous, and are clear and non-overlapping.[1] Examples are units not found or not accessible, not-at-home, unable or refusing to respond. In a repeat survey such as EU-SILC, it can be very useful to distinguish between 'potentially convertible' refusals and 'hard core' refusals which have to be dropped from future rounds.

For composite units (e.g. a multi-adult household), any of the above reasons may result in 'partial unit non-response', in the sense described later.

Sampling error

Sampling error is a measure of the variability between estimates from different samples, disregarding any variable errors and biases resulting from the process of measurement and sample implementation. Of course, sampling error represents only one component of the total survey error. For estimates based on small samples, this component may be the dominant one. In other situations, non-sampling errors, in particular sample selection, non-response and measurement biases, may be much more important. However, even in these cases, sampling error increases progressively as the estimates are produced for smaller and smaller subgroups of the population, such as for social classes or regions of a country. In a small enough subgroup, sampling error may well outweigh non-sampling errors. This consideration is very important in a multi-purpose survey such as EU-SILC, an important objective of which is to study differentials and trends.

[1] 'The non-response rates can be measured well if accurate accounts are kept of all eligible elements that fall into the sample. These are necessary for understanding the sources of non-response, for its control and reduction, for predicting it in future surveys, and for estimating its possible effect on the surveys. … These aims can be better served if the many possible sources of non-response are sorted into a few meaningful classes. A good classification of non-response depends on the survey situation ….'. Kish (1965), Section 13.4A.

1.4 Item non-response

Item non-response can be seen as an intermediate category between errors in measurement and errors in estimation. Like any other error in measurement, item non-response is subject-matter specific. At the same time, it can be viewed simply as an addition to the existing unit non-response in analysis involving the particular item affected, thereby amounting to an error in estimation. Item non-response is particularly important in EU-SILC and similar surveys collecting complex and detailed information on components of household and personal income. Some components such as income from self-employment and capital can be subject to extremely high levels of item non-response.

Information on an item may be incomplete simply because *it is not feasible to seek it exactly or in full detail* in an interview survey; these errors are akin to ‘conceptual errors’. The impact on the results may differ depending on the respondent’s characteristics and circumstances. Often information is missing because the respondent is *unable* to provide it, or the respondent may be *unwilling* to provide information which is considered too sensitive or personal. There can be an added, special reason for item non-response in surveys providing microdata to researchers and other users: this is *deliberate suppression* of some information, presumably based on confidentiality and similar considerations.

For composite items (e.g. an income target variable composed of several individual items), any of the above reasons may result in ‘partial item non-response’, in the sense described later.

As in the case of unit non-response, classification of item non-response according to the reasons or circumstances giving rise to it can be very helpful in identifying and controlling the extent of non-response and assessing its impact.

1.5 Comparability

Comparability is increasingly considered as the central requirement of data quality. This dimension of quality is particularly important in a multi-country undertaking such as EU-SILC, where issues relating to comparability underscore all aspects of data quality, especially data accuracy. It is not possible to assess the extent and impact of sampling and non-sampling errors in EU-SILC without evoking at the same time the extent to which the results can be considered comparable across countries, across time, and also in relation to other data sources. We may also note that indications of accuracy (unit and item non-response rates, sampling variance, etc.) need to be defined and computed following the same procedures.

1.6 EU-SILC structure and content

In order to explore the sources of non-sampling errors in EU-SILC, it is useful to begin by recapitulating some of its main characteristics.

EU-SILC is the major source of comparative statistics on income and living conditions in Europe. It covers data and data sources of various types: cross-sectional and longitudinal; household-level and person-level; economic and social; from registers and interview surveys; from new and existing national sources.

Substantive content

In terms of the substantive content, four types of data are involved: (i) variables measured at the household level; (ii) information on household size and composition and basic characteristics of household members; (iii) income and related variables measured at the personal level, but aggregated to construct household-level variables (which may then be ascribe to each member); and (iv) more complex non-income or 'social' variables collected and analysed at the personal level.

For set (i)-(iii) variables, a sample of households including all household members is required. Among these, sets (i) and (ii) are normally collected from a single, appropriately designated respondent in each sample household. Alternatively, some or all of these data may be compiled from registers or other administrative sources.

Set (iii) variables - concerning mainly, but not exclusively, the detailed collection of household and personal income - must be collected directly at the personal level, *covering all persons in each sample household*. In many countries, these income and related variables are collected through personal interviews with all adults aged 16+ in each sample household. These are the so-called *survey countries*. In other countries by contrast, set (iii) variables are compiled from registers and other administrative sources, thus avoiding the need to interview all members (adults aged 16+) in each sample household. These are the so-called *register countries*.

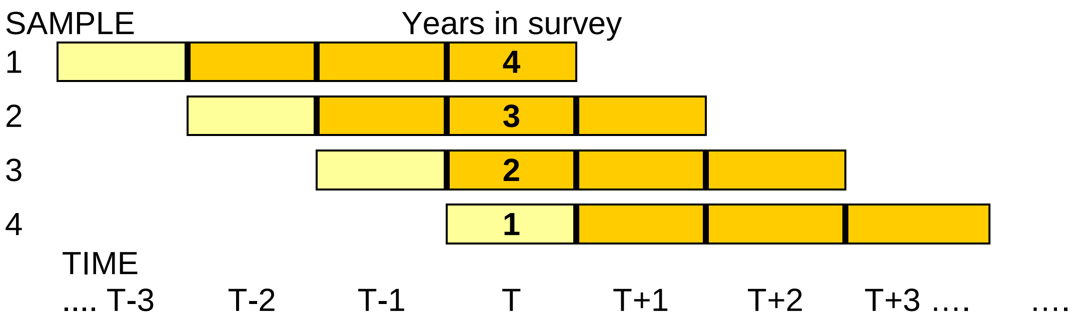
Set (iv) variables are normally collected through direct personal interview in all countries. These are too complex or personal in nature to be collected by proxy, nor are they available from registers or other administrative sources. For the survey countries, this collection is normally combined with that for set (iii) variables, both being based on a sample of complete households, i.e. covering all persons in each sample household. However, from the substantive requirements of EU-SILC, it is not essential — in contrast to set (iii) variables — that set (iv) variables be collected for all adult persons in each *sample household*, since it has been decided in EU-SILC design that these variables need not to be aggregated to the household level. It is therefore sufficient to do this collection on a representative *sample of persons*. This option is normally followed in register countries, since for these countries interviewing all household members for set (iii) is not involved. In countries which choose to do so, the sampling process normally involves the selection of one adult member aged 16+ per household, either directly or through a sample of households.

Cross-sectional and longitudinal data

Another dimension is that both cross-sectional and longitudinal data are required. The *cross-sectional component* covers information pertaining to the current and a recent period such as the preceding calendar year. It aims at providing estimates of cross-sectional levels and of net changes from one period (year) to another. The *longitudinal component* covers information compiled or collected through repeated enumeration of individual units, and then linked over time at the micro-level. It aims at measuring gross (micro-level) change. Both cross-sectional and longitudinal micro-data sets are updated on an annual basis. In most EU-SILC surveys a period of four years is taken as the duration for longitudinal follow-up at the micro level.

The integrated design

Figure 2: Illustration of a simple rotational design



A standard *integrated design* has been adopted by nearly all the participating countries. This integrated design involves a rotational panel in which a new sample of households and persons is introduced each year to replace a part (normally one quarter) of the existing sample. Persons enumerated in each new sample are followed-up in the survey normally for four years. A common rotational sample of this type yields each year a cross-sectional sample as well as longitudinal samples of various durations (Verma, 2001; Verma and Betti, 2006).

At any one time, the sample is made up of 4 subsamples or panels (Figure 2). Each year one new panel is added to stay in the survey for 4 years, and then dropped to be replaced by another new panel. Movers from the original sample are followed-up to their new location for up to the time their panel remains in the survey. This scheme provides both cross-sectional and longitudinal data from the same common set of units. The cross-sectional sample for year T consists of four subsamples, 1-4, one introduced each year from (T-3) to T. A longitudinal sample consists of persons who have remained in the survey since they were first introduced into it. Three overlapping longitudinal samples of different durations are formed: of two year duration from subsamples (2+3+4); of three year duration from subsamples (3+4); and of four year duration from subsample (4).

Sample follow-up (tracing) rules

Another important aspects in the definition of the statistical units in EU-SILC is the tracing of units over time. The initial (Wave 1) sample consists of a sample of households. All usual residents of those households who are over a certain age (stipulated to be no higher than 14 years) are called *sample persons*. The relevant Commission Regulation (European Community, 2003) state that at subsequent waves the eligible population consists of the following.[2]

- The survey is confined to private households. The households covered in any wave consist of those containing at least one sample person still retained in the survey as defined above. This includes newly formed households resulting from the movement of sample members since the last wave, as well as any new households added to the survey. Persons moving out of the country, institutionalised, or moving into a collective household are dropped.

- Sample persons who are still alive and eligible for the survey. Any movers among these persons are followed up to their new address, including children in the original household or born to sample mothers as they reach a certain age (such as 14). In this way the survey population is kept up-to-date for demographic changes except for immigrants into the original population.

[2] The EU-SILC follow-up rules are somewhat more restrictive (and hence imply additional non-response) compared to the ECHP follow-up rules; see Verma and Clemenceau (1996).

- *Co-residents (non-sample persons)*, meaning persons who reside in the same household with one or more sample persons. Such persons are covered using the same procedures; however, the survey does not follow-up non-sample persons who move into households not containing any sample person.

- Register countries normally involve random selection of one sample person per household, designated as a *selected respondent*. The personal interview (mostly on non-income variables) is confined to such selected respondents.

- For practical reasons, a limit is placed on the duration for which a household can remain un-interviewed for it to be retained in the sample for follow-up.

2. Conceptual and measurement errors

There is a great variety of errors arising from conceptual and measurement (collection and processing) sources and the patterns can differ across countries. Below we present a few selected aspects of such errors with reference to the measurement of income, which constitutes the main topic of EU-SILC. Often it is not possible to associate these errors with a single source: usually the observed patterns reflect the combined effect of different sources.

2.1 Reporting of negative and zero values for income components

As illustrations, Table 1 examines variation across countries of the incidence of reporting negative, zero and positive values of some important income components: income from self-employment, capital income, and rental income.

Income from self-employment

Panel (A) shows the variation across countries in the incidence of reporting negative, zero and positive values of income from self-employment. The incidence of reporting a negative amount varies across countries: roughly half the countries in EU-SILC permit the recording of negative values for income from self-employment, and the other half do not. This illustrates the influence of variations in measurement procedures on accuracy and comparability of the data.

Capital income

The impact of conceptual or definitional differences is seen markedly in the case of capital income (from interest, dividends and capital investment). See Panel (B).

No countries record negative values for the component, except for the striking example of Denmark where over half the reported amounts in the 2007 data were negative.[3] This indicates a *conceptual* difference from other countries. This difference of course does not mean that more negative than positive total amount from the source is reported in Denmark: the component accounts for around 20% of total income of households, which is practically identical to the average value of this share over EU countries. The large number of negative values in Denmark are in fact made up of numerous small amounts. The impact of this major conceptual difference on the composition of total income according to source may, therefore, not be too large.

[3] There were a few cases of negative values also in Norway, but amounting to a tiny proportion.

There are also some other conceptual differences in the measurement of capital income. For instance, all but 1-3% households report *zero* income from this source in countries such as Greece and Hungary, while all but 1-3% report a *non-zero* income from this source in register countries like Denmark and Norway. These differences are also reflected in the mean amount per recipient – the values being much higher among the fewer recipients in the former countries compared to the latter. Such differences are likely to arise from, among other factors, differences in the methods of measurement – registers tend to record small values exhaustively, while in personal interviews only larger amounts are likely to be recorded.

Rental income

Similar results for income from rental of property or lands are shown in Panel (C). Any effect of conceptual or methodological differences between countries is less obvious, or at least appears less consistent, in this case. The proportion receiving income from this source is low (1-2%) in some eastern and also in some Scandinavian countries, but much higher in some countries such as Luxembourg (12%) and Greece (18%). The average over countries is 5%. The amount per recipient also varies greatly, from a low of around 5% of national mean household income in Denmark, Sweden and Slovakia, up to a high of 20-25% in a number of countries including Spain, Greece, Italy, Cyprus, Hungary, Austria, Belgium and UK. It is difficult to judge how far these large variations observed are genuine.

Table 1: Percentage of households receiving and mean per recipient: for selected income components

(A) Cash benefits or losses from self-employment (PY050*)

	recipients as % of all households				% negative of recipients	% share of total income	mean per recipient
	% zero*	% negative	% positive	total			
	(1)	(2)	(3)	(4)=(2)+(3)	(5)=(2)/(4)	(6)	(7)=(6)/(4)
BE	89.7	0.1	10.2	10.3	0.7	6.1	59.8
CZ	81.5		18.5	18.5		15.1	81.5
DK	69.2	3.5	27.3	30.8	11.5	5.7	18.4
DE	89.6		10.4	10.4		9.2	88.4
EE	90.6	0.5	9.0	9.4	4.8	2.3	24.4
IE	78.9		21.1	21.1		13.1	61.9
EL	67.9		32.1	32.1		24.3	75.5
ES	85.6	0.8	13.6	14.4	5.7	8.2	56.7
FR	92.3		7.7	7.7		7.0	90.6
IT	72.7	0.2	27.1	27.3	0.8	20.4	74.6
CY	75.3		24.7	24.7		11.2	45.4
LV	91.8	0.2	8.0	8.2	2.5	4.5	54.8
LT	81.3		18.7	18.7		6.4	34.0
LU	93.2	0.1	6.7	6.8	1.5	4.3	62.2
HU	84.4	0.4	15.3	15.6	2.4	8.3	53.2
NL	84.9	2.3	12.8	15.1	15.3	6.1	40.0
AT	83.8	0.4	15.8	16.2	2.2	8.6	53.2
PL	78.3		21.7	21.7		9.9	45.6
PT	79.2		20.8	20.8		12.0	57.4
SI	75.1		24.9	24.9		5.4	21.6
SK	89.2	0.2	10.6	10.8	1.7	7.7	71.6
FI	84.8		15.2	15.2		5.5	36.5
SE	81.3	4.3	14.5	18.8	22.9	2.8	14.8
UK	87.5	0.0	12.5	12.5	0.1	8.8	70.4
IS	82.6		17.4	17.4		3.5	19.9
NO	86.0	2.9	11.1	14.1	20.9	5.8	41.3
Average				17.1		8.5	52.1
cv(%)				40.7		60.1	42.0

Source: Cross-sectional data 2007, weighted by household cross-sectional weight (DB090).

NB: Gross self-employment income (PY050) is aggregated to the household level.
* This column for 'zero' values may contain small numbers of missing values on income.
'Average' refers to simple (unweighted) average over the 26 countries shown.
'cv' is the coefficient of variation of unweighted country values.

Reading note: The table shows that, for example in Belgium (BE), 0.1% of households report a negative amount for income from self-employment and 10.2% a positive amount, giving a total of 10.3% 'recipients'. Negative reports form 0.7% of these recipients. Of the total income received by all households, that from self-employment constitutes 6.1%. However, considering only households with non-zero income from self-employment, the average amount of income from self-employment received by such households amounts to nearly 60% of the total income averaged over all households in Belgium.

Table 1: continued

(B) Interest, dividends, from capital investments (HY090)

	recipients as % of all households: % zero* (1)	% negative (2)	% positive (3)	total (4)=(2)+(3)	% negative of recipients (5)=(2)/(4)	% share of total income (6)	mean per recipient (7)=(6)/(4)
BE	31.5		68.5	68.5		2.5	3.7
CZ	86.0		14.0	14.0		0.8	6.0
DK	1.9	53.2	44.9	98.1	54.2	1.9	2.0
DE	16.6		83.4	83.4		2.7	3.3
EE	58.4		41.6	41.6		0.5	1.2
IE	66.4		33.6	33.6		3.9	11.7
EL	96.6		3.4	3.4		0.3	10.1
ES	69.2		30.8	30.8		1.0	3.1
FR	24.6		75.4	75.4		2.4	3.1
IT	55.5		44.5	44.5		1.6	3.7
CY	88.3		11.7	11.7		2.0	17.4
LV	98.9		1.1	1.1		0.2	22.0
LT	96.3		3.7	3.7		0.4	9.4
LU	46.0		54.0	54.0		1.9	3.5
HU	98.8		1.2	1.2		0.4	31.2
NL	18.9	0.3	80.9	81.1	0.3	4.3	5.3
AT	35.4		64.6	64.6		1.0	1.6
PL	97.5		2.5	2.5		0.5	20.4
PT	87.3		12.7	12.7		0.7	5.8
SI	65.3		34.7	34.7		0.5	1.4
SK	93.2		6.8	6.8		0.1	1.4
FI	25.5		74.5	74.5		3.5	4.6
SE	25.4		74.6	74.6		3.2	4.2
UK	50.3		49.7	49.7		3.2	6.3
IS	33.9		66.1	66.1		8.1	12.3
NO	1.1		98.9	98.9		2.7	2.7
Average				43.5		1.9	7.6
cv(%)				0.8		0.9	1.0

Table 1: continued

(C) Income from rental of property or lands (HY040)

	recipients as % of all households: % zero* (1)	% negative (2)	% positive (3)	total (4)=(2)+(3)	% negative of recipients (5)=(2)/(4)	% share of total income (6)	mean per recipient (7)=(6)/(4)
BE	92.7		7.3	7.3		1.7	23.0
CZ	95.9		4.1	4.1		0.4	10.9
DK	98.4		1.6	1.6		0.1	4.6
DE	91.1		8.9	8.9		1.4	16.3
EE	99.0		1.0	1.0		0.1	9.5
IE	94.8		5.2	5.2		0.9	17.6
EL	82.1		17.9	17.9		3.9	21.9
ES	94.2		5.8	5.8		1.4	24.4
FR	92.6		7.4	7.4		1.3	18.3
IT	93.9		6.1	6.1		1.4	22.2
CY	91.2		8.8	8.8		2.1	23.8
LV	98.7		1.3	1.3		0.1	9.7
LT	94.7		5.3	5.3		0.4	8.1
LU	87.6		12.4	12.4		2.3	18.6
HU	98.2		1.8	1.8		0.4	23.9
NL	96.8		3.2	3.2		0.6	17.5
AT	96.1		3.9	3.9		0.9	23.2
PL	98.7		1.3	1.3		0.2	18.1
PT	96.4		3.6	3.6		0.8	22.7
SI	94.9		5.1	5.1		0.5	9.0
SK	95.9		4.1	4.1		0.2	5.7
FI	93.5		6.5	6.5		1.0	15.3
SE	98.6		1.4	1.4		0.1	5.5
UK	95.6		4.4	4.4		0.9	21.5
IS	95.3		4.7	4.7		0.4	8.8
NO	97.9		2.1	2.1		0.3	12.2
Average				5.2		0.9	15.9
cv(%)				0.7		1.0	0.4

2.2 Total household gross and disposable income (HY010, HY020)

A source of variation affecting comparability is the presence of negative, zero and extreme (very large) values in the distribution of total household income. Often these differences result from different data sources and survey conditions and procedures.

While only a minority (around one-third) of the countries permit negative values for total *gross income,* most (but not all) seem to permit zero values though the proportions of such cases are generally very small. See Table 2(A).

The incidence of negative and zero values is somewhat higher for total household *disposable income*. One of the main uses of this variable is to serve as a measure of economic well-being. However, negative or zero values of disposable income do not provide a useful measure of well-being which can serve as a proxy for living standards. The process of equivalisation of income – which adjusts household income to take into account economies of scale – also makes little sense when applied to negative quantities. In any case, some measures of poverty and inequality cannot be constructed with negative or zero amounts of net disposable income.

The presence of a few large values at the upper end of the income distribution is also problematic in this respect, though not necessarily in the same way as negative or zero incomes. While not affecting measures of poverty, the presence of even a few very large values can markedly affect the computed indicators of inequality such as Gini and (S80/S20). The variance of the estimates may also be greatly inflated. These factors impart instability to the survey estimates, and adversely affect their comparability across time and across countries. For instance, we find (using 2007 data) that, on the average across EU countries, the 99th percentile of total household disposable income is around four times the national median income, with a small coefficient of variation (cv) of 15% across countries. See Table 3. The diversity among countries increases as we move closer to the upper end of the distribution: the cv of the ratio to national median increasing to 20% at the 99.5th percentile, 25% at 99.8th percentile, and to 60% among the largest recorded values in the countries.

2.3 Total household disposable income before social transfers (HY022, HY023)

A major limitation of these variables is the high incidence of zero and negative values. Variable HY022 is constructed from total net income (HY020) by deducting from it social transfers other than pensions, and HY023 is constructed by further deducting pensions as well. A characteristic feature of these variables is the large proportion of zero and negative values encountered: while generally there are only a small proportion of zero and negative values in HY010 or HY020, these proportions become quite significant in the case of HY022, and tend to become very large for HY023. In the 2007 data for instance, averaged over countries 3% of the computed values of total household disposable income before social transfers other than pensions (HY022) are negative or zero. This average figure increases to 17% for total household disposable income before all social transfers (HY023). The last mentioned proportion reaches or exceeds 25% in almost one-fourth of the countries. See Table 2(B). The presence of large proportions of zero and especially of negative values diminishes the usefulness of these variables in providing explanatory or policy-relevant indicators.

Different factors may be involved in different countries in determining the prevalence of negative and zero values in HY022 and HY023.

Table 2: Bottom end of the income distribution
(A) Gross and total disposable income (HY010, HY020)
% of households with income up to x% of median income

	% of households with income up to x% of median household income										
	HY010				HY020						
	Gross household income				Disposable household income						
	median income	% of households x<0	<=0	<=15	median income	% of households x<0		<=0	<=10	<=15	HY020 /HY010
BE	32 345		0.2	1.1	25 016	0.3	0.4	0.6	0.8	1.2	0.77
CZ	10 242		0.0	0.4	8 810		0.0	0.0	0.1	0.4	0.86
DK	42 303	0.4	0.5	1.3	28 824	0.6	0.6	0.7	0.8	1.0	0.68
DE	31 937		0.1	1.9	24 445	0.7	0.8	1.3	1.6	2.1	0.77
EE	7 366	0.0	0.6	1.9	6 404	0.2	0.6	0.9	1.4	1.9	0.87
IE	42 306		0.1	1.1	37 880	0.0	0.1	0.7	0.9	1.0	0.90
EL	20 942		0.4	1.4	16 889	0.3	0.6	0.9	1.2	1.6	0.81
ES	24 000	0.2	0.8	1.9	20 820	0.3	0.9	1.2	1.4	1.9	0.87
FR	30 329			0.4	24 875	0.1	0.1	0.2	0.3	0.6	0.82
IT	28 595	0.1	0.9	2.4	23 051	0.4	1.0	1.3	1.8	2.4	0.81
CY	32 129		0.0	0.8	29 141	0.0	0.0	0.1	0.2	0.5	0.91
LV	6 202		0.8	3.1	5 166	0.4	0.9	1.6	2.1	2.6	0.83
LT	6 047		0.5	2.6	5 213	0.1	0.5	0.9	1.5	2.3	0.86
LU	58 900	0.0	0.0	0.6	47 870	0.0	0.0	0.1	0.3	0.4	0.81
HU	7 510		0.0	0.6	6 540	0.0	0.0	0.2	0.3	0.5	0.87
NL	37 642	0.3	0.3	0.8	27 179	0.5	0.5	0.6	0.7	1.0	0.72
AT	35 583			1.1	27 971	0.0	0.0	0.3	0.7	1.0	0.79
PL	8 134		0.2	1.1	6 285	0.1	0.2	0.3	0.7	1.1	0.77
PT	16 048			1.3	13 800			0.0	0.2	0.9	0.86
SI	21 851			1.7	18 205	0.0	0.0	0.1	0.4	0.8	0.83
SK	7 665		0.2	0.8	6 747	0.2	0.2	0.4	0.4	0.8	0.88
FI	33 105		0.0	0.7	25 702	0.0	0.1	0.2	0.3	0.5	0.78
SE	35 441	0.0	0.3	2.1	25 230	0.2	0.4	0.8	1.3	1.8	0.71
UK	40 004		0.3	1.5	31 891	0.5	0.5	0.8	1.1	1.6	0.80
IS	61 984		0.1	1.2	45 668	0.0	0.1	0.2	0.5	1.0	0.74
NO	48 706	0.2	0.2	2.7	37 024	0.5	0.6	0.9	1.5	2.5	0.76
mean	27 974	0.0	0.3	1.4	22 179	0.2	0.4	0.6	0.9	1.3	0.81
max	61 984	0.4	0.9	3.1	47 870	0.7	1.0	1.6	2.1	2.6	0.91
min	6 047	0.0	0.0	0.4	5 166	0.0	0.0	0.0	0.1	0.4	0.68
cv(%)	58	90	93	52	56	95	86	75	65	55	7

Reading note: The table shows that, for example in BE, 0.3% households report negative disposable income; 0.4% zero or negative income, and 1.2% income at or under 15% of the median disposable income in BE. The mean value of disposable income is 0.77 times the mean gross income per households. The last three rows give simple average over countries, and the maximum and minimum values encountered.

Table 2: continued
(B) Disposable income before social transfers
% of households with income up to x% of median income

	% of households with income up to x% of median household income									
	HY022					HY023				
	Disposable, before transfers (1)					Disposable, before transfers (2)				
	median income	% of households x<0	<=0	<=15	HY022 /HY020	median income	% of households x<0	<=0	<=15	HY023 /HY020
BE	21 662	1.7	5.2	11.4	0.87	17 520	7.0	13.2	34.0	0.70
CZ	7 875	1.0	2.8	4.7	0.89	6 354	14.8	25.3	29.4	0.72
DK	24 449	6.5	6.6	11.7	0.85	21 652	14.1	14.3	29.3	0.75
DE	22 428	2.6	5.1	9.9	0.92	14 619	21.1	24.8	34.3	0.60
EE	5 943	1.3	2.8	4.3	0.93	4 723	13.2	20.6	27.3	0.74
IE	31 765	0.4	7.7	11.0	0.84	26 618	0.5	20.7	26.9	0.70
EL	16 200	0.4	1.1	2.3	0.96	11 500	6.8	18.8	23.9	0.68
ES	19 640	0.4	1.8	3.9	0.94	15 550	4.2	14.5	24.9	0.75
FR	21 796	1.3	2.1	5.3	0.88	14 042	17.8	21.4	31.8	0.56
IT	21 798	0.6	1.5	3.0	0.95	15 867	10.4	16.1	30.1	0.69
CY	27 311	0.2	1.0	1.7	0.94	24 930	5.4	14.5	19.6	0.86
LV	4 716	1.1	2.6	4.7	0.91	3 914	9.0	18.2	24.4	0.76
LT	4 724	0.1	2.4	4.9	0.91	3 881	1.9	21.2	27.9	0.74
LU	43 261	1.5	1.6	3.8	0.90	34 465	7.9	8.0	23.6	0.72
HU	5 467	2.0	3.2	5.8	0.84	3 575	12.1	26.3	31.9	0.55
NL	25 085	1.4	1.4	8.0	0.92	19 544	8.0	8.0	28.7	0.72
AT	24 570	0.2	2.4	5.1	0.88	17 820	1.9	12.0	31.1	0.64
PL	5 682	1.1	2.6	5.2	0.90	3 698	12.7	24.1	31.4	0.59
PT	12 552	0.6	2.0	3.6	0.91	9 190	10.4	24.0	28.7	0.67
SI	15 519	0.5	1.1	5.5	0.85	11 456	2.4	4.3	28.0	0.63
SK	6 048	1.6	2.3	3.3	0.90	4 661	23.4	27.6	31.7	0.69
FI	21 699	1.3	4.3	10.1	0.84	17 625	6.3	11.7	29.9	0.69
SE	21 638	1.3	2.8	8.5	0.86	17 214	11.5	16.9	31.3	0.68
UK	28 495	5.2	6.6	8.6	0.89	21 969	23.4	26.1	31.9	0.69
IS	41 284	1.4	1.5	3.2	0.90	38 558	11.5	11.7	15.5	0.84
NO	31 675	2.4	2.6	11.1	0.86	31 675	2.4	2.6	11.1	0.86
mean	19 742	1.5	3.0	6.2	0.89	15 870	10.0	17.2	27.6	0.70
max	43 261	6.5	7.7	11.7	0.96	38 558	23.4	27.6	34.3	0.86
min	4 716	0.1	1.0	1.7	0.84	3 575	0.5	2.6	11.1	0.55
cv(%)	55	99	62	50	4	62	65	41	20	11

Source: EU-SILC Users' database, cross-sectional data 2007; weighted by DB090

(1) Disposable income, before social transfers other than pensions
(2) Disposable income, before all social transfers

Table 3: Ratio of upper percentiles to the median

Total disposable household income (HY020)

	sample size	median income	minimum reported	Ratio of Pth percentile to median income								
				P = 80	90	95	97	98	99	99.5	99.8	maximum
BE	6 348	25 016	-22 500	1.7	2.2	2.6	3.0	3.3	**3.9**	4.6	5.9	20.3
CZ	9 675	8 810	0	1.6	2.1	2.5	2.9	3.4	**3.8**	4.5	6.4	22.3
DK	5 783	28 824	-618 129	1.8	2.2	2.5	2.8	3.1	**3.7**	4.9	8.3	60.9
DE	14 153	24 445	-186 249	1.7	2.2	2.8	3.2	3.6	**4.6**	6.1	8.9	26.5
EE	5 146	6 404	-1 296	1.9	2.5	3.2	3.8	4.3	**5.1**	6.1	7.2	40.1
IE	5 608	37 880	-736	1.8	2.3	2.8	3.3	3.7	**4.4**	5.9	7.9	39.9
EL	5 643	16 889	-21 759	1.8	2.4	3.0	3.5	4.0	**5.1**	6.3	8.8	21.3
ES	12 329	20 820	-11 171	1.7	2.2	2.7	3.0	3.3	**3.9**	4.7	5.9	10.2
FR	10 498	24 875	-79 373	1.6	2.0	2.5	2.8	3.1	**3.8**	4.4	5.6	41.9
IT	20 982	23 051	-60 090	1.8	2.3	2.9	3.4	3.8	**4.7**	5.8	7.2	25.7
CY	3 505	29 141	-35	1.6	2.1	2.5	2.9	3.5	**4.4**	7.1	11.9	22.8
LV	4 471	5 166	-248	2.0	2.8	3.6	4.0	4.4	**5.6**	7.1	7.8	23.9
LT	4 975	5 213	-579	1.9	2.6	3.2	3.9	4.3	**5.0**	5.8	6.7	14.9
LU	3 885	47 870	-22 977	1.7	2.1	2.6	3.0	3.2	**3.9**	5.0	5.7	75.8
HU	8 737	6 540	-45	1.6	2.0	2.4	2.8	3.1	**3.6**	4.8	6.3	20.8
NL	10 219	27 179	-62 602	1.6	2.1	2.5	2.9	3.3	**4.4**	6.0	10.4	16.5
AT	6 806	27 971	-4 200	1.6	2.1	2.5	2.9	3.3	**4.2**	5.1	6.4	10.2
PL	14 286	6 285	-2 380	1.7	2.3	2.9	3.4	3.9	**4.5**	5.4	6.7	28.5
PT	4 310	13 800	469	1.8	2.5	3.5	4.1	4.6	**5.7**	7.4	9.5	20.0
SI	8 707	18 205	-241 292	1.6	1.9	2.3	2.6	2.8	**3.2**	3.6	4.2	8.9
SK	4 941	6 747	-50	1.6	2.1	2.5	2.9	3.2	**3.8**	4.5	5.2	8.6
FI	10 624	25 702	-3 261	1.7	2.1	2.5	2.9	3.2	**3.8**	4.7	7.1	33.3
SE	7 183	25 230	-8 289	1.7	2.0	2.4	2.6	2.8	**3.3**	4.0	5.2	16.5
UK	9 275	31 891	-12 669	1.8	2.3	2.9	3.4	3.8	**4.8**	5.9	7.6	57.5
IS	2 872	45 668	-228	1.6	2.1	2.6	3.2	3.6	**4.6**	7.4	11.0	23.4
NO	6 013	37 024	-382 721	1.7	2.1	2.4	2.7	3.0	**3.5**	4.3	6.0	23.8
average				1.7	2.2	2.7	3.1	3.5	**4.3**	5.4	7.3	27.5
cv(%)				6.3	9.5	12.5	13.3	14.0	**15.5**	19.5	25.9	60.5

Source: EU-SILC Users' database.

Reading note: The table shows that, for example in Belgium (BE), income of the top 20% households exceeds 1.7 times the national median income and that of the top 1% it exceeds 3.9 times the median. The maximum income recorded in the survey is 20.3 times the national median of euro 25 016. Negative values of income are allowed, and the minimum recorded is minus 22 500 in Belgium. The last two rows show the simple average and the coefficient of variation of the figures across countries.

It is likely that such values appear in large numbers as a result of deducting social transfers from the household's actual disposable income without considering that outgoings (already deducted from income) may be *conditional* on the availability to the household of the social transfer income component which is being removed. An obvious example is a voluntary private transfer paid out by a household, itself dependent on social transfers as the main source of its income. Another important issue concerns the net-gross form of social transfers which are deducted from HY020 in the construction of HY022/HY023. Obviously, this deduction has to be net of taxes and contributions, but in some cases gross amounts seem to have been deducted.

An added disturbing aspect – likely to have an adverse effect on comparability – is the apparently arbitrary choice in recording non-positive values either as zeros or as negative. In some countries negative values while in others zero values predominate. In relation to non-sampling errors in EU-SILC data and their comparability across countries, it is important to investigate how far these markedly differing patterns arise from conceptual and procedural differences among the national surveys.

2.4 The importance of uniform procedures for achieving comparability

Often the presence of negative, zero and very large values of household income is the result of errors in the data introduced at the collection or processing stages. While it cannot be assumed automatically that any such extreme values are erroneous, there is a high chance of that being the case. Empirically we find that country surveys differ greatly in the incidence and patterns of occurrence of extreme values. In part this may result from differences in data sources and national situations, but mostly it seems to be the result of differences in conditions and procedures of data collection, and especially in how the data are recorded and processed. These differences damage the international comparability of the results. EU-SILC data can be made more comparable through greater standardisation across countries of the manner in which negative, zero and very large values are treated.

The use of standardised procedures can of course enhance the data quality of individual EU-SILC surveys. Even more important is the positive effect such standardisation can have on comparability across countries and over time. Improved comparability may be considered as the most important justification for adopting *common procedures* for treating extreme values in the income distribution.

3. Non-response in EU-SILC

Non-response - both unit and item non-response - is a serious problem in EU-SILC surveys. It is clear from the available national and Eurostat reports that *non-response of both types is high in many countries, and very high in some.* Apart from cross-sectional non-response, panel attrition is particularly serious in some cases, affecting also the consistency between cross-sectional and longitudinal results.

3.1 A framework

Though normally a distinction is made merely between unit non-response and item non-response, in the complex data structure and content involved in EU-SILC a more complete classification needs to be employed, such as that in Figure 3.

It would be extremely useful to study these different aspects of non-response in empirical detail. *However, a major practical difficulty is the lack of information on non-response available to researchers with access only to the UDB.* Variables required for the computation and understanding of non-response have been removed from UDB – presumably because of confidentiality concerns.

Figure 3: Components of non-response

Problem	Description	Common solution
(1) Unit non-response	Failure to obtain any information on a sample household, including the household interview and personal interviews in the household	Weighting
(2) Partial unit non-response	Failure to obtain a personal interview with a subset of the eligible adults in a household	Weighting or full-case imputation
(3) Item non-response	Failure to obtain some target variables in an otherwise completed interview. (This generally affects non-income variables in register countries, and all - especially income - variables in survey countries)	Imputation for missing items
(4) Partial item non-response	Refers to the situation when some but not all the information is obtained on a target variable. The most important case is that of detailed income components: a part of the component may be missing, and/or conversion may be required from the collected net to the required gross amount	Micro-simulation (net-gross conversion), in conjunction with imputation for the missing part.

3.2 Unit non-response

Each stage involved in obtaining the interview contributes to unit non-response: successfully contacting the sample address; interviewing the sample household once contacted; and detailed personal interviews with all adults (or, depending on the survey design, with one selected respondent) in the household .

Table 4, cols. (1)-(3) give the response rates at the above three stages for 2007 cross-sectional samples. The figures are confined to the *panel newly introduced* in 2007 in the rotational design. The overall response rate for the personal interview is the product of these rates. Its complement, the overall non-response for the personal interview, is shown in col. (4). A number of points are worth noting.

(1) Non-response rates are very high – exceeding 33% in 8 of the 26 countries, and exceeding 20% in all but 6 countries. Such high rates can be expected to have a major effect on the representativeness of the results.

(2) The potential impact of non-response is further increased because its incidence often varies across different parts of the population with differing characteristics – such as having higher rates of non-response among persons at either end of the income distribution. It is therefore important to analyse non-response rates for subpopulations. Unfortunately this cannot be done for EU-SILC on the basis of microdata available to researchers, since variables concerning the response status of households and individuals have been excluded from those data.[4] The figures reported in the table are merely reproduced from national or Eurostat quality reports.

[4] This applies to response status at the address contact and household interview stages (variables DB120, DB130, DB135) which account for most of the non-response. The only available variable is the response status of personal interview within otherwise completed households (RB245), the contribution of which to overall non-response is minor.

Table 4: Unit non-response (cross-sectional sample 2007)

	New panel Response rate by stage			Overall non-reponse rate for personal interviews	
	address contact	household interview	personal interview	New panel	Total sample
	(1)	(2)	(3)	(4)	(5)
BE	99	48	99	53	36
HU	100	52	100	48	29
DK	86	69	100	41	42
ES	98	63	99	38	24
CZ	96	65	100	38	18
AT	100	65	99	36	23
EE	84	77	99	36	20
PL	99	72	93	34	22
LT	99	68	99	32	17
SI	98	73	100	29	24
IE	100	72	100	28	30
FI	100	75	100	25	17
EL	100	76	100	25	16
NL	95	83	100	22	17
IT	99	81	100	20	15
PT	98	88	100	14	20
DE	91	96	100	13	19
FR	99	88	100	13	15
CY	100	91	100	9	8
SK	100	98	100	2	16
mean	97	75	99	28	22

Source: Compiled from national Intermediate Quality Reports 2007 for EU-SILC.

NB: Countries ordered by col. (4), the overall personal interview response rate for the new panel.

Countries where information for the new panel has not been reported separately are not shown.

Reading note: Overall response rate is the product of the response rates at each stage, cols. (1)-(3). Col. (4) is the complement of the overall response rate: thus 0.53=1-(0.99*0.48*0.99) for Belgium (BE). This column gives overall non-response rates for the newly introduced panel in the rotational design and col. (5) gives those for the whole sample as reported in national quality reports. As explained in the text, these last-mentioned figures have not been correctly computed and generally grossly *under-state* the actual non-response rates.

(3) The table also quotes in col. (5) the reported non-response rates for the cross-sectional sample as a whole. Normally these rates should be higher than the non-response rates in col. (4) for the newly introduced panel, since the older parts of the sample have been subject to additional non-response at previous waves. The reported results are mostly inconsistent with this. It is for the following reason.

In a cross-sectional sample based on a rotational design (Verma and Betti, 2006), proper computation of the rate of non-response must take into account all the losses in the sample which have occurred since the concerned units were first selected into the rotational design. *The reported non-response rates are gross underestimates since their computation has been based entirely on the units present in the current cross-sectional data set*. Units which were selected at an earlier time and remain in-scope of the target population, but were dropped from the survey due to earlier non-response are not taken into account in the computation of the current cross-sectional non-response rates, in so far as such units do not appear in the current cross-sectional data files used as the basis for these computations.

(4) Unit identification numbers in EU-SILC are randomised for confidentiality reasons. This randomisation seems to have been done independently between the cross-sectional and longitudinal data sets, even though in terms of actual units these data sets largely overlap. The problem of correctly computing cross-sectional non-response rates can be resolved only by *retaining the identification of the link between the cross-sectional and longitudinal samples at the micro level, and using the information on longitudinal follow-up rates in the computation of achieved response rates for the cross-sectional sample*. For a sample introduced into the survey at an earlier wave, the actual response rate of its contribution to the current cross-sectional sample is the product of (a) the response rate achieved when it was first introduced into the survey, akin to the complement of col. (4); and (b) the 'wave response rate' at each subsequent wave, similar to the complement of col. (5) per wave.[5]

[5] Wave response rate is the percentage of sample units successfully interviewed in wave t, among in-scope units passed on from wave (t-1) or newly created or added during wave t.

3.3 Within-household ('partial unit') non-response

The overall personal interview response rates discussed above incorporate the effect of within-household non-response, i.e. of failures to obtain personal interview(s) in households otherwise successfully enumerated. In any case, the contribution of such within-household non-response is generally very small at the aggregate level.

However, this is not the case as concerns the *individual households affected*. The income of the household (and hence the equivalised income attributed to each of its members) cannot be measured without including the contribution of all its members.

The reported incidence of within-household non-response is around 1% in most countries, but is higher in a few (for instance in the 2007 survey, around 3% in Latvia and Slovakia, and notably 10% in Poland as a clear outlier). Countries have used quite different methods to deal with the problem, which are as follows (see Table 5).

(1) Full-case imputation of missing personal interviews. It can be a convenient and satisfactory method when the incidence of within-household non-response is small.

(2) Adjustment of total income of the affected household by a factor (UDB variable HY025) determined on the basis of characteristics of the household and of the non-interviewed persons. (Followed by DE, EL, ES, LV, PT, SK).

(3) Taking no action, i.e. making no imputation or weight adjustment for the missing personal interviews. (PL, despite high incidence of within-household non-response).

(4) Deleting from the data all households with one or more missing personal interviews. This inflates the overall household non-response rate. It can be wasteful, and also hides the problem of within-household non-response. Yet, this is a widely used practice. (Followed by CZ, IE, IT, LU, HU, UK).

Table 5: The incidence and treatment of within-household non-response (missing or imputed personal income and related data)

	persons aged 16+	Personal income data completed	not completed number	%	If data collection not completed % of hhs affected	Method used	% hhs with HY025>1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BE	12 322	12 236	86	0.70	1.20	(1)	
EE	11 971	11 875	96	0.80	1.40	(1)	
FR	20 357	20 243	114	0.56	0.89	(1)	
CY	8 470	8 453	17	0.20	0.43	(1)	
LT	10 913	10 885	28	0.26	0.48	(1)	
AT	13 391	13 332	59	0.44	0.82	(1)	
DE	26 399	26 291	108	0.41	0.66	(2)	0.66
EL	12 417	12 346	71	0.57	1.10	(2)	1.10
ES	28 845	28 656	189	0.66	1.11	(2)	1.04
LV	9 442	9 270	172	1.82	3.15	(2)	3.15
PT	9 988	9 947	41	0.41	0.74	(2)	0.65
SK	12 762	12 573	189	1.48	3.26	(2)	5.14
PL	34 888	32 801	2087	5.98	9.85	(3)	none (all=1.0)
CZ	19 384	19 384				(4)	
IE	10 892	10 892				(4)	
IT	44 629	44 629				(4)	
LU	7 913	7 913				(4)	
HU	18 490	18 490				(4)	
UK	17 484	17 484				(4)	
DK	11 610	11 610				(5)	
NL	19 623	19 623				(5)	
SI	24 730	24 730				(5)	
FI	21 773	21 773				(5)	
SE	14 204	14 204				(5)	
IS	6 567	6 567				(5)	
NO	11 706	11 706				(5)	

Source: EU-SILC Users' database. Unweighted figures. 2007 cross-sectional data set.

NB: Countries grouped according to the method used for dealing with within-household non-response

(1) full-case imputation
(2) special weighting of income data (by factor HY025)
(3) nothing done
(4) whole household dropped from data set
(5) Register countries: household dropped if selected respondent not interviewed

Reading note: 1. Col. (4) gives the proportion of adults for which income and related data where not collected in households which were otherwise 'sufficiently complete' to be included in the data base; col. (5) gives the proportion of households with one or more such non-interviewed adults. Where missing information is not imputed, total household income is adjusted by a factor HY025>1 – see col. (7).

(5) All the register countries present a situation similar to (4), but arising from a different mechanism. Here the information on income comes from administrative sources, not subject to non-response. Complex non-income or 'social' variables are collected through personal interview in all countries, including register countries which follow-up one selected respondent per household for the purpose. These interviews are, of course, subject to high rates of non-response.[6] Unfortunately, households where such an interview cannot be conducted are dropped from the survey, hence *losing also the information on income for these households* - even though it would have been possible to compile this information without non-response from registers for all sample households and their members.

Frick *et al* (2010) analyse the problem of within-household non-response using data from more than twenty waves of the German Socio-Economic Panel (SOEP). They evaluate different strategies to deal with this phenomenon, and show how the choice of the technique affects the substantive results and their comparability.

3.4 Item non-response

Unlike unit non-response for which there is a lack of information, EU-SILC is exceptionally good in providing detailed information on item non-response in the microdata files and also in survey documentation. There are few other social surveys which match the EU-SILC standards in this respect.

For every income component, the data provide two 'flags' indicating the form and the degree of completeness of the collected information. Though all income components are recorded gross of taxes and social insurance contributions, the collected amount may be gross, or it may be net of taxes and/or of social insurance contributions. The first flag records the form of collection, which determines whether micro-simulation is required to obtain the target gross amounts. Micro-simulation is similar to imputation in that both involve some form of modelling; micro-simulation tends to be more dependent on external data and relationships, while imputation more on relationships between the variables observed within the dataset.

[6] In fact, the overall personal interview non-response rates in register counties tend to be higher than those in survey countries: the respective average figures being 27% and 21% for the 2007 cross-sectional sample, for instance. As noted, these figures from the national quality reports are themselves under-estimates. It should also be mentioned that within each group there is a wide varation across counties around the above-mentioned averages.

The second flag records the ratio of the amount actually collected to the amount recorded for the component concerned. As explained in notes to Table 6, value ‘0’ means full item non-response – the percentage of cases in which the item has been completely imputed. Value ‘1’ means the amount is recorded exactly as collected, with no imputation or net-to-gross conversion. The remaining cases involve partial item non-response. In this case, the flag gives the combined effect of imputation and net-to-gross conversion. However, in cases where the amounts collected were already in the gross form, no net-to-gross conversion is involved and the flag departs from 1.0 only because of imputation for the part which was missing. In other cases the flag indicates a mixture of imputation and net-to-gross conversion.

For illustration, values of the two flags are shown in Panel (A) for income from self-employment (PY050). Similar information is provided in Panel (B) for capital income and in panel (C) for rental income. These figures underscore the richness of the information available on item non-response in EU-SILC microdata. It has to be admitted, however, that the quality of the available information on flags is not uniformly good and the information is missing in some countries. Having a large proportion of cases with low values of the imputation flag indicates poor quality of the data. Large variability in this index across countries also casts doubt on comparability of the information.

Table 6 Item non-response

(A) Cash benefits or losses from self-employment (PY050), 2007

	% receiving	% distribution of recepients by mode of collection						% distribution acording to imputation factor (= value collected / value recorded)					Total
		-1	1	2	3	4	5	= 0	0-0,5	0,5-1,0	= 1	>1	
BE	6.2		0.1			100		71	0	0	28		100
DK	23.2					100							100
DE	6.1					100		9	4	4	84		100
IE	10.4					100		56	1	6	37		100
ES	7.4					100		29	2	36	33		100
CY	11.3					100		1		0	99		100
LU	5.0					100		46		0.3	53		100
HU	10.2				0.1	100		2			98	0.1	100
NL	9.6					100					100		100
AT	9.7					100		94	4	0.4	2		100
SI	15.8					100		36	5	3	54	1.3	100
SK	4.9					100					100		100
FI	21.3					100							100
SE	13.4				100	0					100		100
UK	7.3					100		22	0.1	0.2	78		100
IS	10.9					100							100
NO	11.2	8				92							100
CZ	7.6	2	18			79			1	16	81	1.3	100
LT	9.4		25			74	0	1	1	14	83	0.8	100
EE	6.7		33			65	3	14	1	18	67		100
PT	10.4		63	19		14	4		2	85	14		100
EL	19.5		100			0					100		100
FR	4.3				100	0			1	94		5.3	100
IT	16.6		100			0		4	19	5	72	0.0	100
LV	4.3		100			0		6	1	9	83		100
PL	10.6		100			0		20	12	15	54	0.0	100

Source: EU-SILC Users' database; cross-sectional data rev. 2; unweighted values.

Code*:* mode of collection: -1 missing; 1 net of tax and social contributions; 2 net only of tax; 3 net only of social contributions; 4 gross; 5 not stated.

Reading note: 'Imputation factor' = '0' means *full item non-response* – the percentage of cases in which the item has been completely imputed. Value = '1' means the amount is recorded as collected, with no imputation or conversion. The remaining cases involve *partial item non-response*; if 'mode of collection' = '4', this partial item non-response is entirely due to a part of the information being missing; in other cases it indicates a mixture of imputation and net-to-gross conversion.

Table 6: continued

(B) Interest, dividends, profit from capital investments (HY090)

	% receiving	% distribution of recepients by mode of collection						% distribution acording to imputation factor (= value collected / value recorded)					Total
		-1	1	2	3	4	5	= 0	0-0,5	0,5-1,0	= 1	>1	
BE	68.0			100		0		69			31		100
DK	99.1					100							100
DE	85.0					100		39	20	18	24		100
IE	32.9		100			0		69	14	16			100
ES	35.0		77			23		39	22	16	23		100
CY	12.6					100					100		100
LU	46.8					100		36	0.2	0.3	64		100
HU	1.1					100					100		100
NL	87.5					100					100		100
AT	66.8					100		21	2	3	74		100
SI	36.4					100		15	0.2	0.3	84	0.4	100
SK	7.0					100					100		100
FI	80.8					100		6	14	14	67		100
SE	78.0		100			0					100		100
UK	52.1					100		21	3	4	72		100
IS	69.5	0				100							100
NO	99.3					100							100
CZ	14.0					100					100		100
LT	4.9					19	81	81	7	6	1	4.9	100
EE	39.0		1			97	3	94	1	3	2		100
PT	13.1	3		97		0				100			100
EL	3.4		100			0					100		100
FR	77.8				100	0		94			6		100
IT	46.2		100			0		15	3	83			100
LV	1.1					100		30			70		100
PL	2.1		100			0		41	0.3	33	26		100

(C) Income from rental of property or lands (HY040)

	% receiving	% distribution of recepients by mode of collection						% distribution acording to imputation factor (= value collected / value recorded)					Total
		-1	1	2	3	4	5	= 0	0-0,5	0,5-1,0	= 1	>1	
BE	7.3					100		8	0.2	0.2	92		100
DK	2.6					100							100
DE	10.1					100		5			95		100
IE	5.7					100		9			91		100
ES	6.4		32			68		5	4	24	68		100
CY	9.6					100					100		100
LU	10.9					100		7			93		100
HU	1.8					100					100		100
NL	3.4					100					100		100
AT	4.1					100		100					100
SI	5.1					100					100		100
SK	4.2					100		5			95		100
FI	10.2					100							100
SE	1.7				100	0					100		100
UK	4.5					100		17	0.2	1	81		100
IS	5.5	1				99							100
NO	2.4					100							100
CZ	4.0					100					100		100
LT	6.0					66	34	34	4	8	53	0.7	100
EE	1.5		91			9		1		90	9		100
PT	4.0	1				99					100		100
EL	17.7		100			0					100		100
FR	8.4				100	0		100			0.3		100
IT	7.0		100			0		10	1	84	5		100
LV	1.4					100		3			97		100
PL	1.2					100		17			83		100

Reading note: Countries presented in exactly the same order as in Table 3(A) for self-employment income.

NB: mode of collection: -1 missing; 1 net of tax and social insurance; 2 net only of tax; 3 net only of social contributions; 4 gross; 5 not stated.

4. Sampling error

4.1 Jackknife Repeated Replication (JRR) for variance estimation

EU-SILC is a set of large-scale household surveys based on complex designs. The surveys are multi-purpose, involving many types of variables, estimates, units of analysis, levels of aggregation of the results, and diverse subpopulations for which estimates of levels, differences and other relationships are required.

Practical procedures for estimating sampling errors for such a survey: (i) must take into account the actual, complex structure of the design; (ii) should be flexible enough to be applicable to diverse designs; (iii) should be suitable and convenient for large-scale application, producing results routinely for diverse statistics and subclasses; (iv) should be robust against departure of the actual sample design from the ideal model assumed in the computation method; (v) should have desirable statistical properties such as small mean-square error of the variance estimator; (vi) should be economical in terms of effort and cost; and (vii) suitable computer software should be available for application of the method (Verma, 1991). Two broad practical approaches to the computation of sampling errors are:

Computation from comparisons among certain aggregates for primary selections or replicates within each stratum, also known as the linearization method.

Computation from comparisons among estimates for replications of the sample, each of which reflects the structure of the full sample.

A major advantage of methods in (2) above is that they do not require an explicit expression for the variance of each particular statistic, and hence can more easily handle complex statistics and designs, including multi-wave and longitudinal situations. The variance estimates take into account the effect on variance of aspects of the estimation process which are repeated for each replication. In principle this can include complex effects such as those of imputation and weighting, though often full repetition of these procedures for each replication is not feasible.

A particular method of class (2), namely the Jackknife Repeated Replication (JRR) method, has been adopted by Eurostat for EU-SILC. Its basic model is as follows. Consider a design in which two or more primary selection units (PSUs) have been selected independently from each stratum. Within each PSU, subsampling of any complexity may be involved, including weighting of the ultimate units. In the standard version, each JRR replication can be formed by eliminating one PSU from a particular stratum at a time, and increasing the weight of the remaining PSUs in that stratum appropriately so as to obtain an alternative but equally valid estimate to that obtained from the full sample. Let z be a full-sample estimate of any complexity, and $z_{(hi)}$ the estimate produced using the same procedure after eliminating primary unit i in stratum h, compensating for that by increasing the sample weight of the remaining $(a_h - 1)$ units in the stratum by an appropriate factor. Let $z_{(h)}$ be the simple average of the $z_{(hi)}$ values over the a_h sample units in h. Variance of z is estimated as[7]:

$$\mathrm{var}(z) = \Sigma_h \left[\left(\frac{a_h - 1}{a_h} \right) . \Sigma_i \left(z_{(hi)} - z_{(h)} \right)^2 \right]. \quad (1)$$

The same relatively simple variance estimation formula holds for z of any complexity. Furthermore, apart from variance estimation of ordinary cross-sectional measures, application of the JRR methodology can be readily extended to more complex indicators based on the EU-SILC rotational panel design. These include longitudinal poverty rates based on union and/or intersection of an individual's poverty statuses at a series of cross-sections, as well as measures of net change and averages over two or more waves (Verma and Betti, 2007).

4.2 Defining sample structure: 'computational' strata and PSUs

Practical variance estimation methods, including the JRR, need to make some basic assumptions about the sample design. These include the following.

(1) The sample selection is independent between strata.
(2) Two or more primary selections are drawn from each stratum.
(3) These primary selections are drawn at random, independently and with replacement.
(4) The number of primary selections is large enough for valid use of the variance estimation procedure described above.

[7] The 'finite population correction', trivial in a survey such as EU-SILC, is negelected in (1).

Though these basic assumptions regarding the structure of the sample for application of the method are met reasonably well in most EU-SILC surveys, often the assumptions are not met exactly. *In many practical situations some aspects of the sample structure need to be redefined to make variance computation possible, efficient and stable.* Of course, any such redefinition is appropriate only if it does not introduce significant bias in the variance estimation.

A very convenient approach in practice is to summarise the most essential information about the sampling design in the form of two variables, coded for each unit in the microdata file: the 'computational stratum' and the 'computational PSU' to which the unit belongs. This can be done in most cases for the type of sample designs used in EU-SILC. Obviously, these two variables must be defined so as to meet the basic requirements (1)-(4) listed above for the application of the variance computation procedure adopted. Normally, we may expect the new variable 'computational stratum' to be related (and sometimes identical) to UDB variable DB050; similarly for 'computational PSU' and DB060. However, very often the UDB variables require some redefinition before they can be used for the purpose of variance estimation. The computation stratum has to incorporate all information about the stratification of the PSUs, including both explicit stratification and, where applicable, implicit stratification resulting from systematic sampling of the PSUs. It has also to ensure that each computational stratum contains at least two computational PSUs (which are then assumed to have been selected at random with replacement). Starting from the actual PSUs, the variable computational PSU should seek to create units 'reasonably' large and uniform in size, and small enough in number so as to avoid excessive computational burden. To do the above in a statistically valid way requires sampling expertise. Apart from codes of the existing sample structure in the microdata files, this requires additional information: (i) detailed description of the sample design, identifying features such as the presence of systematic selection, 'self-representing' PSUs (which are in fact strata), etc; and (ii) information connecting the sample structure codes in the microdata with sufficiently detailed and clear descriptions, on the basis of which sample structure at the level of individual units can be identified.

4.3 Some common procedures for defining computational strata and PSUs

The computational structure can differ from the actual sample structure because of various consideration. Here are some common situations.

Computational strata

1. It may be necessary to regroup ('collapse') strata so as to ensure that each stratum has at least two sample PSUs – the minimum number required for the computation of variance.

2. Units which are included into the sample automatically ('self-representing units') are in fact strata rather than PSUs, and computational PSUs have to be defined at a lower stage within each such unit.

3. In samples selected systematically, the implied implicit stratification is often used to define explicit strata, from each of which an independent sample is supposed to have been selected. Such strata have to be formed by pairing or otherwise grouping of PSUs in the order of their selection from the systematic list, ensuring that each resulting computational stratum has at least two primary selections. The structure of the systematic sample can in fact be more complex, requiring special considerations. For example, the sample may consist of several systematic sample of different sizes, each corresponding to a rotation group.

4. Sometimes the sample may contain too many small PSUs and also too many small strata. Grouping of PSUs to form computational units (see below) may involve joining units across strata, thus resulting also in the grouping of strata.

Computational PSUs

The construction of computational PSUs most commonly involves grouping of existing units. Here are some common motivations for this. Technical procedures for grouping are further elaborated subsequently a procedure like the JRR, the number of replications is equal or at least similar to the number of PSUs in the sample. In a large sample where elements (households, persons) have been selected directly, the number of replications which can be formed will be of the order of the sample size, normally running into thousands. This necessitates forming much fewer computational units, such as creating ‘pseudo-cluster’ from random groupings of sample elements, and then random pairing of these ‘clusters’ to construct computational strata. Such random grouping does not affect the expected value of the variance , but does greatly facilitate the applications of the procedure. The above issue in fact arises in the case of any sample irrespective of its structure when we want to estimate not only variances but also design effects. See next section.

1. Sometimes non-response or attrition can result in the disappearance from the sample of whole PSUs. This can disturb the structure of the sample, such as leaving fewer than two PSUs in some strata. Variance computation requires some redefinition of the computational units to meet the basic requirement of having at least 2 PSUs per stratum. This problem arises more frequently when computing sampling errors for subclasses (subpopulations). The risk can be reduced by aggregating PSUs and strata to create fewer, larger computational units.

2. Generally, steps such as grouping of small PSUs within and across strata, and grouping of strata to form fewer and larger computational units, are desirable to improve stability of the results.

3. Secondly, grouping of units to create fewer, larger computational units is required for efficiency. This is particularly important in the case of re-sampling methods, including the JRR, which tend to be heavy on computational time. The replication approach requires re-computation of the statistic of interest at each replication. For complex statistics such as poverty rates, this may require a considerable amount of computing time, and it can be desirable to reduce the number of times the process has to be repeated.

4. The same also applies to many other forms of complex estimation, especially if it requires iterative procedures. Variance estimation with replications captures the effect on variance of those features of the data treatment and estimation process used in the actual survey which are applied repeatedly to each replication, in the same way that they were applied to the full sample. For instance, in order to fully capture the effect of calibration on variance, it is necessary to recalibrate the sample of each replication using the same procedure as used in the actual sample. The same applies to other aspects of sample weighting, such as adjustment for non-response. Another even more demanding example is imputation for missing data. The need to repeat such heavy procedures at each replication can greatly increase the computational task. Procedures are required to reduce the number of replications involved.

5. Finally, there is an objective which is especially relevant in the context of EU-SILC, which applies both to computational strata and computational PSUs. There are restrictions on the detail with which information identifying individual sampling units, PSUs, strata etc. can be included in the public-release microdata. Grouping of units and strata can help in preserving confidential nature of the data, and make the suppression of the sample structural information unnecessary.

4.4 Some technical procedures for grouping of units and strata

Collapsing of the sample structure for above purposes can involve merging of strata, and of PSUs within strata and also across strata. As noted earlier, any such redefinition of the sample structure is of course appropriate only if it does not increase significantly the mean square error of the variance estimator through introducing bias and/or increasing its variance.

It has been demonstrated that appropriately done collapsing usually does not introduce additional bias or variability in the variance estimates (Rust, 1985). The above-mentioned author notes the following:

> '… beyond about 25 or 30 degree of freedom the quantiles of the t distribution vary little with the number of degrees of freedom, being close to those of the normal distribution. Thus it is common practice to use as 95% confidence intervals of the [normal] form … provided that [the number of replications] r is at least 25 or 30. Hence for the purposes of making inference about a parameter, the precision of variance estimation is not of great importance provided that at least 25 to 30 degrees of freedom are attained. … For stratified designs with many sampled PSUs, 25 to 30 degrees of freedom can often be attained with replicated variance estimators using few more than 30 replicates. However, to attain the required precision with a relatively small number of replicates, *care is often required in the way replicates are formed* …'

Some technical procedures for grouping include the following.

(1) Dropping some of the replications. The 'standard' (or 'full') JRR procedure involves the dropping of one PSU at a time, thus giving the same number of replications as the number of PSUs in the sample. A more general Jack-knife variance estimator requiring fewer replicates can be obtained by omitting more than one PSU from each replicate, and/or omitting only from a subset of the PSUs.

(2) Random grouping PSUs within strata. This method is suitable in designs in which each stratum contains many sample PSUs, and in the sample as a whole there are numerous and small PSUs. Random grouping of units within strata may provide an efficient and unbiased option. An obvious example is creating 'pseudo' PSUs in a simple random or another direct sample of elements. (See above).

(3) Grouping PSUs across strata. This involves combining strata, and simultaneously combining PSUs across the strata in the combination. This is a general form of what Deming (1960) has termed 'thickening of zones'.

(4) Grouping PSUs within and across strata. This is a combination of techniques (2) and (3) above.

The number of computational PSUs to create

The important practical question is: *how many random groups (computational PSUs) should be created?* It is known from theory that such random grouping does not affect the expected value of the variance of the sample. However, it does affect the stability (variance) of the variance estimates. As the number of random groups is reduced, the variance estimates tend to become less stable – we can get different results from repletion of the same procedure, and hence also as the number of random groups is varied. With a larger number of random groups, the computations tend to become stable and insensitive to the exact number of random groups chosen.

From our numerical experience with EU-SILC and similar applications, we have found 200 random groups to be a safe choice in all cases, and even 100 in almost all cases. It is desirable to keep this number small for computational efficiency. But it is desirable to do some numerical testing of the stability of the results with variation in the number of random groups.

The above considerations apply even more strongly to computations for subclasses. Normally, variance estimation for subpopulations does not involve any new procedures: the same formulae apply except that sample elements not members of the subpopulation of interest are simply disregarded. The only complication is the in computations involving subclasses – especially small and not well-distributed subclasses – it can happen that some PSUs and strata contain no elements of interest. This can make the results unstable and biased. The risk can be reduced by aggregating PSUs and strata to create fewer, larger units for the purpose of computation. The above principles also apply to samples (or particular domains of the sample) which, while being multi-stage, involve numerous small PSUs.

As noted, grouping of small PSUs within and across strata, and grouping of strata to form fewer and larger computational units is generally desirable to improve stability of the results. The final choice is always a matter of statistical judgement and numerical experimentation to validate the choice.

The number of computational strata

If the existing sample is an *unstratified* sample of elements, there is no need to create separate computational strata: all the computational PSUs defined by random grouping can constitute a single computational stratum.

If, however, the existing sample is a *stratified* sample of elements, normally the existing stratification can be retained unchanged to constitute the required computational strata, but ensuring that at least *two* random groups are created within each stratum. Larger strata can have more than two random groups each. One scheme can be to assign to each stratum a (rounded) number of random groups proportional to its (weighted) sample size – with the total number of random groups taken as 100-200 – and then adjust this number to ensure that every stratum has at least two such units.

If the existing strata are too small and numerous, merging of strata (on the basis of similarity of stratum characteristics) can also be considered as described earlier.

5. Issues in computing variances in EU-SILC

5.1 Limitations owing to availability of information on sample structure

It is important to re-emphasis a point of great practical relevance in relation to variance estimation by users of EU-SILC data. The major problem is the lack of sufficient information for the purpose: *the UDB does not contain information on sample structure, in particular concerning stratification.* Consequently, from UDB, variances can be computed only for the handful of countries which have employed simple (unstratified) samples of households or persons, or where such a simple structure can be assumed as a reasonable approximation. Generally, however, appropriate coding of the sample structure, in the survey microdata most preferably, is an essential requirement in order to ensure that sampling errors can be computed properly, taking into account the actual sample design. Lack of information on the sample structure in survey data files is a long standing and surprisingly persistent problem in survey work, as for example Kish *et al* (1976) discovered in their attempts to compute sampling errors for achieved survey data sets in the United States.[8]

Consequently, from UDB, variances can be computed only for countries which have employed simple (unstratified) samples of households or persons: Denmark and Iceland. With slight approximation, it is also possible for Sweden, for which the sampling frame of persons is actually ordered by age (providing implicit stratification), though the sample may be considered SRS according to the national quality report; it is also possible for Austria which uses a SRS of dwellings, as we may disregard the small effect of clustering of households within dwellings. In a number of countries, stratified random sample of households or persons are used. For these the effect of stratification *may be* relatively small (at least in comparison with that of stratification in multi-stage designs). These include for instance Cyprus, Estonia and Lithuania. Stratified element samples are also used in Slovakia, Finland and Germany, but with some approximations.

[8] We are fortunate in having received additional information on sample structure (in particular on explicit stratification, variable DB050) from Eurostat for illustrative computation of sampling errors for EU-SILC surveys. But this information still had some major limitations.

Coding of the sample structure in the survey micro-data needs to be complemented with documentation and description of the sampling procedures and the resulting sample in order to ensure that sampling errors can be computed properly, taking into account the actual sample design. Lack of information on the sample structure in survey data files and documentation is a long standing and surprisingly persistent problem in survey work, as for example Kish *et al* (1976) discovered in their attempts to compute sampling errors for achieved survey data sets in the United States.

We are fortunate in having received additional information on sample structure (in particular on explicit stratification, variable DB050) from Eurostat for this and related research. Table 7 provides some essential features of the sample structure of the countries for which the information has been kindly made available by Eurostat for the purpose of this research. This information has some major limitations, however.

1. It is available for only a subset of countries, shown in the table.

2. The sample structure information provided can be linked to only the longitudinal microdata in UDB (through common household identifiers, DB030), but not to the cross-sectional data set because of randomisation of the identifiers.

3. Furthermore, the datasets involved are not truly longitudinal in terms of individuals present continuously in the panel, but are merely the part of the cross-sectional sample comprised of households belonging to rotation groups which have been present in the survey for one or more preceding waves.

4. The information on sample structure, including the specially provided supplementary information, is generally coded only for the newly introduced panels each year, not for the entire data set. Feeding forward this information for panels introduced earlier is not straightforward because of changes in household identifiers, for instance due to household splits. Links have to be established at the personal level, going through personal register (P-file, variable HX030).

5. For some countries, the information on effective stratification is incomplete where implicit stratification resulting from the systematic sampling used. Variable defining the order of selection (DB070) is not coded or not coded in a standard way in some countries.

6. The problem is compounded in some cases because of inconsistencies in the coding of the variable identifying ‘sample rotation groups’ (DB075).

7. The same may be mentioned in cases where some of the PSUs may be 'self-representing', thereby forming effective strata rather than PSUs as coded.

5.2 Creating computational strata and PSUs: illustrations from EU-SILC

(1) Simple random sample of elements

Examples from EU-SILC: DK, IS, SE, AT.

In the application of JRR variance estimation to design involving simple random sampling of elements (dwellings, households or persons), it is necessary in practice to reduce the number of replications we have to deal with. For this purpose, we may randomise the list of sample elements, and create for instance 100 or 200 groups of equal size (or more appropriately groups of equal weighted size).[9] These groups serve as computational PSUs. The whole sample can be taken as a single computational stratum.

Numerically, the results can be affected by exactly how the replications are formed. On the basis of experience, we find that the 'reasonable' number such as 100-300 random groups provides very similar results for EU-SILC samples. We also recommend that the groups formed should be of *uniform weighted size,* that is, as far as possible, constructed by including a constant sum of weights of elementary units in every grouping. Hence the procedure recommended is as follows.

> Ensure that the listing of units in the data base is random (by randomising the order if necessary). Divide the randomised list into nearly equal-sized groups - 'equal size' means the same *sum of unit weights* in the group. Each group as defined above constitute a computational PSU. All the PSUs so defined constitute a single computational stratum.

[9] Numerically, the results can be affected by exactly how the replications are formed. On the basis of experience, we find that the 'reasonable' number such as 100-300 random groups provides very similar results for EU-SILC samples. We also recommend that the groups formed should be of *uniform weighted size,* that is, as far as possible, constructed by including a constant sum of weights of elementary units in every grouping.

Table 7: Basic features of sample structure (selected EU-SILC samples)

	sampling stages	stratifiction (X=yes)	systematic (X=yes)	DB070 available?	register country	USU	Comments
1 DK	1				X	person	simple random sampling (SRS)
2 IS	1				X	person	simple random sampling (SRS)
3 AT	1					dwelling	simple random sampling (SRS)
4 SE	1		X*		X	person	*Sampling frame ordered by age of individual
5 EE	1	X				person*	*Unlike 'register countries', all adults in the household of selected person are interviewed; stratified random sampling
6 LT	1	X				person*	*Unlike 'register countries', all adults in the household of selected person are interviewed; stratified random sampling
7 CY	1	X				household	stratified random sampling
8 DE	1	X				household*	*Lacks strict probability sampling: part quota before 2008; also the 'random' part based on 'recruitment' of households in 'Access Panel' obtained through Microcensus
9 FI	1	X*			X	person	*stratified random sampling; Straification only of Master Sample from which sample drawn in 2 phases, not of the entire frame
10 SK	1	X				household*	*stratified random sampling; households are actually selected by Poisson sampling, rather than simple random sampling
11 CZ	2	X				dwelling	
12 PL	2	X				dwelling	PPS sampling of areas using Hartely-Rao scheme. Many small clusters (3-4 hhs per cluster)
13 EL	2	X	X	X		dwelling	
14 BE	2*	X				household	*PSUs drawn with PPS, with repetions allowed
15 UK	2	X*	X	X		household**	*One PSU per stratum; order of selection available. **Actually postal addresses, which correspond to households
16 LV	2	X	X	X		dwelling*	*Actually addresses, which may correspond to dwellings

Source: EU-SILC Users' database.

In Denmark, Iceland and Sweden, the sample is a SRS of persons aged 16+. Nevertheless, for measuring income, the sample consists of households constructed around each selected person, and all members of that household are included. Income is defined and constructed at the household level, and then the same value ascribed uniformly to each member of the household. Hence for variables related to total household income, such as mean equivalised income or at-risk-of-poverty rate, the randomisation must be at the household level (that is, always keeping all members of a household in the same random group) – even if the unit of analysis is the person, as is normally the case in income distribution and poverty analysis. On the other hand, these are 'register' countries, in which 'social' (non-income) variables are measured and analysed at the person level, with only the selected individuals who are subject to a personal interview entering the analysis. Hence the appropriate random grouping is that of such persons. Of course, in this particular case the grouping corresponds exactly to the grouping by household for income variables since only one respondent is selected from each household in these designs.

In Austria, we have a SRS of dwellings. Dwellings should therefore for the units for random grouping (that is, always keeping all households and all persons in a dwelling in the same random group). This applies to all variables, income and social, since both are measured for all eligible persons in all the households in each sample dwelling. In the absence of information associating household with dwelling, random grouping can only be done at the household level, introducing a (most likely a small) approximation.

(2) Systematic sample of elements

Possible EU-SILC example: SE

The sample of Sweden is actually a little more complex: it is a *systematic sample of persons* from lists ordered by age. The procedure in principle should therefore be as follows, though as noted in the national sample description, in practice the sample may be treated as a simple random sample.

> Divide the list of elements *arranged according to the order of systematic selection* into approximately equal parts (equal sums of unit weights), but each part being twice as large compared to that in case (1) above. Each of these parts forms one computational stratum. Within each part, randomise the order of the elements, and create two randomised groups as in (1) Each randomised groups so created forms a computational PSU.

(3) Stratified random sample of elements

Examples from EU-SILC: CY, EE, LT; also FI, SK ,DE.

In the first subgroup (Cyprus, Estonia, Lithuania), we have strictly stratified random samples of households or persons; there is no ordering and systematic sampling within strata. In the other subgroup, some other complexities are involved.

The Finish sample is actually more complex, involving two phases, with stratification applied only to the master sample obtained after SRS at the first phase.

In Slovakia, the selection within strata is Poisson sampling rather than SRS. This selection method subjects each unit independently to selection with its assigned probability P_i. This can be achieved by assigning to each unit in the frame a random number r_i from uniform distribution [0,1). The unit is in the sample if $r_i <= P_i$, and not in the sample otherwise. Clearly, the method is extremely simple to apply. Its major drawback in this simplest form is the lack of control over the sample size obtained: the achieved sample size is a random variable with large variance. This is because each unit is subject to selection independently, irrespective of how many other units have already been selected in the same way. This is a serious problem when we are dealing with small samples, such as for individual strata. Indeed, when the probability of getting no sample at all is not negligible, the method is hardly worth considering. Fortunately, this tends not to be problem in household surveys such as EU-SILC, which involve relatively large strata and sample sizes. For this design, variance of the Horvitz-Thompson estimator of totals tends to be large, but much more precise is a ratio estimator which takes into account the sample size actually obtained. With the latter type of estimator the procedure can be as efficient as a SRS despite the lack of control over sample size, provided that the stratum sample sizes are not small and the probability of getting a zero sample size is negligible.

The problem is more serious in Germany. Here a large part of the sample is a non-probability quota sample. No information on stratification has been coded for a large part of the sample; presumably this corresponds to the quota part of the sample. Even the so-called 'random' part is based on 'recruitment' of households in 'Access Panel' obtained through Microcensus Hence the survey so far lacks strict probability sampling. Variances can be estimated only by making some serious assumptions about probability nature of the sample.

The following procedure may be used for defining computational strata and PSUs where the design is assumed to be stratified random sample of elements.

Make the groups as defined above in (1), but separately within each explicit stratum of the actual design. The only additional condition is to ensure that at least 2 computational PSUs are created within every original stratum.

(4) Systematic sampling of elements within explicit strata

There are no examples of such a design among the countries for which information on sample structure is available.

In any case, the procedure is exactly the same as (2), except that now it is *applied separately within each explicit stratum*. It is necessary to ensure that at least two computational PSUs are created within every computational stratum so created.

(5) Systematic sample of PSUs in a stratified multi-stage design

Except for any prior grouping on the lines described in case (6) below, we may proceed as follows.

Take the PSUs in the order of their selection.

Group neighbouring PSUs into pairs; taking 3 PSUs into one of the groups if the total number of PSUs involved is odd.

Each group (generally a pair) of PSUs defined above constitutes a computational stratum.

(6) Sample involving many small clusters (PSUs)

The general advice is to group the actual (small and numerous) PSUs into larger, fewer and more uniformly sized computational units using the various techniques described above.

(7) One-PSU-per-stratum design

In such a design it is necessary to combine the actual strata into computational strata such that each computational stratum contains at least 2 (computational) PSUs. *Note that pairing (or other grouping) of strata for this purpose can be subjective.* We can use whatever information we have on the strata to decide on exactly which *strata* to pair or group. It is most appropriate to from each group to contain the strata which are most similar to each other.

But note that the above freedom of choice applies only concerning the grouping of strata – *the grouping must not be based on any known characteristics of the PSUs actually in the sample from the strata concerned.*

(8) Self-representing PSUs

As noted, PSUs which are included into the sample automatically ('self-representing units') are in fact strata, and not real PSUs. *Each such 'PSU' forms a computational stratum*. Within that computational stratum, computational PSUs have to be created as required. The procedure depends on the design within the 'self-representing PSU'. The next stage units within that 'PSU' become the real PSUs to be dealt with. To them, the appropriate procedure from among the various procedures described above can be applied, depending on the details of the sampling design within the self-representing PSU.

5.3 Longitudinal measures of poverty

Longitudinal poverty rates are based on union and/or intersection of an individuals poverty statuses at a series of cross-sections. The JRR methodology can be applied on the following lines.

The basis is provided by the common structure of the sample across waves (cross-sections) of a panel. This structure is defined by where and how the original sample of households and persons was selected when the panel entered the sample. Irrespective of any persons from the original sample moving to new locations over the life of the panel, each person's location in the structure of the sample (stratum, PSU) remains unchanged. This means that a common set of replications can be defined based on the common structure, covering the whole set of cross-sections of interest.

Furthermore, any type of longitudinal poverty status can be defined as an ordinary variable for an individual, based on the person's cross-sectional statuses over the period concerned. The measures (both cross-sectional and longitudinal) are constructed for balanced panels, i.e. for longitudinal samples of individuals present in the survey throughout the duration of interest, weighted so as to represent the corresponding longitudinal population.

Thereafter, the methodology for JRR estimation of standard errors and design effects, as developed for cross-sectional measures, can be applied to longitudinal measures.

5.4 Net change and aggregation

Here we are concerned with sampling errors of measures of net change in the poverty rate from one year (cross-section) to another, and of poverty rates averaged over two or more waves.

With a panel design, the statistical problem is the following. A large proportion of the individuals are common in the different panels. However, a certain proportion of individuals are different from one wave to the other. The cross-sectional samples are not independent, resulting in correlation between measures from different waves. Apart from correlations at the individual level, we have to deal also with additional correlation that arises because of the same structure (stratification and clustering) of the waves of a panel. Such correlation would exist, for instance, in samples coming from the same clusters even if there is no overlap in terms of individual households. For this purpose, the JRR approach can be extended on the following lines (Verma and Betti, 2007).

1. The total sample of interest is formed by the union of all the cross-sectional samples being compared or aggregated. Using as basis the common structure of this total sample, a common set of JRR replications is defined for it in the usual way. Constructing a 'common set of replications' essentially requires that when an element is to be excluded in the construction of a particular replication, it must be excluded simultaneously from every cross-sectional sample (included in the above-mentioned total sample) where it appears).

2. For each replication, the required measure is constructed for each of the cross-sectional samples involved. These replication-specific cross-sectional measures are differenced or aggregated to obtain the required net-change and average measures for the replication.

3. Variance is then estimated from the resulting replicated measures in the usual way.

6. Design effects

6.1 Analysis of design effects in EU-SILC

Design effect (Kish, 1995) is the ratio of the variance (v) under the given sample design, to the variance (v_0) under a simple random sample of the same size:

$$d^2 = v/v_0, \quad d = se/se_0. \quad (2)$$

Computing design effects requires the additional step of estimating the error under simple random sampling (se_0), apart from its estimate under the actual design (se).

Why are design effects needed and useful? EU-SILC regulations require information on *effective sample siz*e, which can be estimated only with information on design effects. Proceeding from standard errors to design effects is also essential for understanding the patterns of variation and determinants of the magnitude of the error, for smoothing and extrapolating the results for diverse statistics and population subclasses, and for evaluating the performance of the sampling design. It is important to note in this context that values of the design effect can differ greatly across variables and subpopulations within the same survey, and it is important to estimate and analyse this variation. (See for instance, Kish *et al* 1976, Verma *et al* 1980, Verma and Lê 1996, as examples from multi-country multi-subject surveys).

Why is analysis of design effects into components needed and useful? The general reasons for analysing design effects into components include the following: to better understand from where inefficiencies of the sample arise; to identify patterns of variation; through that, to improve 'portability' of the results to other statistics, designs, situations, etc. It may also be noted that with JRR (and other replication methods) the total design effect can only be estimated by estimating its components separately. In applications to EU-SILC, there is in addition a most important and special reason for having procedures for appropriate decomposition of the total design effect into its components. Because of the limited information on sample structure included in the microdata available to researchers, direct and complete computation of variances cannot be done in many cases. Decomposition of variances and design effects identifies more 'portable' components, which may be more easily imputed (carried over) from a situation where they can be computed with the given information, to another situation where such direct computations are not possible. On this basis valid estimates of variances can be produced for a wider range of statistics, thus overcoming at least partly the problem due to the lack of information on sample structure in EU-SILC microdata.

6.2 Components of design effect

We may decompose the design effect into components as follows:

$$v = v_0 . d^2 = v_0 . (d_W . d_H . d_D . d_X)^2 . \qquad (3)$$

Here v_0 is the variance (for the statistic concerned) in an equivalent simple random sample of *individual persons*; d_W is the effect of sample weights; if relevant, d_H is the effect of clustering of individual persons into households and d_D the effect of clustering of households into dwellings; and finally, d_X is the effect of other complexities of the design, mainly clustering and stratification.

The effect of weights d_W does not depend on the sample structure, other than the presence of unequal sample weights for the elementary units of analysis. Weighting generally inflates variance (weighting is primarily introduced to reduce bias). With the complex weighting procedures of EU-SILC, variation in weights can become large, inflating the design effect. This effect needs to be evaluated and controlled.

Factor d_H applies if v_0 refers to variance in a simple random sample of individuals, while v refers to a variable measured at the household level. For example, this factor equals square-root of household size for variables relating to household income when the unit of analysis is an individual person and v_0 is defined to refer to a SRS of *individual persons*. For variables constructed at the household level on the basis of separate but correlated observations on individual household members, d_H will be lower than the above depending on the strength of the correlation.

The effect of clustering of households within dwellings or addresses is absent (d_D=1) when we have a direct sample households or persons, or when such units are selected directly within sample areas – as is the case in most of the EU-SILC surveys. This effect is present when the ultimate units are dwellings, some of which may contain multiple households, but it is small in so far as there is generally a one-to-one correspondence between addresses and households.

Table 8: Estimates of standard errors and components of design effects

	n (persons)	estimate	%se*	%se* (rand)	d_X	d_W	d_H	d_D	d	%se* (srs)
	(1)	(2)	(3)	(4)	(5)= (3)/(4)	(6)	(7)	(8)	(9)= (5)*(6)*(7)*(8)	(10)= (3)/(9)
	Mean equivalised disposable income									
PL	32 820	3 686	0.71	0.77	0.94	1.21	1.74	Y	1.98	0.36
UK	15 434	22 686	1.25	0.94	1.33	1.02	1.53	1.00	2.07	0.61
AT	9 516	19 888	0.82	0.82	1.00	1.11	1.58	X	1.75	0.47
BE	8 205	19 274	1.33	1.19	1.11	1.18	1.55	1.00	2.03	0.65
LT	8 036	3 062	1.59	1.61	0.99	1.25	1.64	1.00	2.03	0.79
	At-risk-of-poverty rate									
PL	32 820	18.4	0.45	0.44	1.02	1.08	1.74	Y	1.91	0.24
UK	15 434	18.0	0.95	0.60	1.57	1.07	1.53	1.00	2.56	0.37
AT	9 516	12.0	0.68	0.68	1.00	1.19	1.58	X	1.88	0.36
BE	8 205	14.1	0.68	0.60	1.13	1.05	1.55	1.00	1.85	0.37
LT	8 036	20.0	1.00	0.98	1.02	1.20	1.64	1.00	2.01	0.50
	At-risk-of-poverty rate for children (aged under 16)									
PL	5 798	25.2	0.79	0.80	0.99	1.07	1.27	Y	1.35	0.59
UK	2 995	21.9	1.53	1.42	1.08	1.08	1.31	1.00	1.53	1.00
AT	1 794	14.7	1.47	1.47	1.00	1.12	1.29	X	1.44	1.02
BE	1 617	13.1	1.31	1.12	1.17	1.04	1.31	1.00	1.59	0.83
LT	1 267	23.6	2.18	2.16	1.01	1.21	1.23	1.00	1.49	1.46
	At-risk-of-persistent-poverty rate (two year longitudinal panel)									
PL	32 820	12.7	0.34	0.34	0.99	1.05	1.74	Y	1.82	0.19
UK	15 434	10.4	0.59	0.53	1.12	1.07	1.53	1.00	1.83	0.32
AT	9 516	6.7	0.57	0.57	1.00	1.14	1.58	X	1.80	0.32
BE	8 205	8.9	0.66	0.58	1.15	1.15	1.55	1.00	2.04	0.33
LT	8 036	15.4	0.87	0.89	0.97	1.25	1.64	1.00	1.99	0.44

Source: The computations refer to 2006 data in the 2-year (2005-2006) panel.

Code:

%se*For mean statistics e.g. equivalised disposable income - expressed as percentage of the mean value. For proportions and rates (e.g. poverty rates) - given as absolute percentage points (pp).

d Overall design effect

Components of design effect:

d_X design effect due to clustering and stratification of ultimate sampling units (dwellings or households)

d_W effect of unequal sample weights

d_H effect of clustering of persons within households

d_D effect of clustering of households within dwellings (if applicable)

Y = effect cannot be separately estimated because of lack of information identifying dwellings but is automatically incorporated into the overall design effect d.

X = effect cannot be estimated, and cannot be included in the overall design effect d.

Reading note: In Poland (PL) for example, standard error for mean equivalised disposable income is 0.71% of the mean value (euro 3,686). For at-risk-of-poverty rate of 18.4%, standard error is 0.45 in (absolute) percentage points (implying a 95% confidence interval of 17.5-19.3%, for instance). Col. (4) gives standard error computed by ignoring any clustering and stratification of the ultimate sampling units (dwellings or households). The ratio of the actual to this 'randomised sample' standard error, col. (5), isolates the effect of clustering and stratification of dwellings/households in the sample. Col. (10) is an estimate of standard error which would be obtained in a simple random sample of *persons*, of the same size as shown in col. (1).

The above components of the design effect can be estimated without reference to information on the sample structure, other than weighting and identifiers linking different types of units (e.g. persons with their households). By contrast, computation of d_X, the effect of complexities of the design such as multiple stages and stratification, requires information on the sample structure linking elementary units to their strata and higher stage units. Normally this effect exceeds 1 because the loss in efficiency due to clustering tends to be larger than the gain from stratification. We can expect it to be less than 1 in stratified random samples of elements.

6.3 Illustrative estimates of variances, and of design effect and its components

On the basis of the additional information provided by Eurostat for the purpose of this research, sampling errors have been computed as an illustration for a few countries shown in the Table 8. The results are for the 2006 sample in the *longitudinal data set* for the year 2006. This data set covers the preceding 2 or 3 years depending on the country.[10] The computations illustrated cover three cross-sectional indicators for 2006, and one longitudinal indicator defined over the two years 2005-2006. In the table, col. (3) is the computed standard error based on the actual structure of the sample, and col. (4) is the same statistic computed by treating the sample as a un-clustered and un-stratified sample of households. The ratio of the two, col. (5) gives d_X, the effect due to clustering and stratification of households in the sample.

Two practically important and convenient points may be noted in relation to these results. Firstly, the complexity of the sample design, at stages above the selection of households, is represented by factor d_X only; *all other components of the design effect are independent of this complexity, and hence can be estimated despite any lack of information on sample structure* in EU-SILC data files, except for the identification of individual addresses, households and persons, and their sample weights. Secondly, in many (though not in all) EU-SILC samples with a multi-stage design, only a small number of households or persons have been selected per PSU. In these cases factor d_X tends to be close to 1, thereby not having a major effect on the overall magnitude of the sampling error.

[10] The necessary sample structure information was not available to the authors for the full cross-sectional samples for any year, even 2006.

7. Concluding remarks

7.1 Diverse sources of non-sampling errors in EU-SILC

Following an examination of particular sources of errors in the preceding sections, it is useful to conclude by listing and classifying areas of particular interest and concern on which detailed evaluation studies are needed.

Income variables

- Analysis of the comparability of income distribution by component, especially monetary income from self-employment and capital, and income-in-kind from imputed rent, own production, company car and other sources.

- Assessment of the impact on comparability of the net-to-gross conversion procedures used in different countries, examining how the procedures used fit into the Eurostat/Siena general micro-simulation model SM2 (Betti *et al*, 2010).

- Analysis of outliers and of zero and negative amounts in reported income.

- More detailed study of comparability of self-employment income, considering both the mode of collection and the pattern of resulting data.

Non-monetary deprivation

- Study of comparability of non-income items defining living conditions and deprivation; comparison of indicators used for multidimensional poverty analysis.

Consistency between cross-sectional and longitudinal components

- Examination of national variations in consistency (or lack of it) between longitudinal and cross-sectional components, and its effect on comparability.

Methodological

- Analysis of the impact on comparability of the differences in structure of the EU-SILC instrument between 'register' and 'survey' countries.
- Comparability of basic concepts for data collection and analysis, such as definition of the household, reference person, sample person and tracing rules.
- Comparability of the national questionnaires and modes of data collection.
- Effect of national differences in the cross-sectional non-response and panel attrition rates.
- Study of differences in the weighting procedures used, and an assessment of the effects of such differences on the comparability of the results.
- Comparability of imputation procedures in national surveys.

7.2 Improving the potential for assessment of data quality in EU-SILC

It is obvious from the above discussion of errors in EU-SILC data that the scope and quality of this evaluation would have been improved with better information on sample structure and sample outcome of the surveys. Little information is available in EU-SILC documentation or microdata for an assessment of different types of measurement errors, except perhaps within some individual countries for their own surveys. The microdata available for research do not contain sufficient information on response status for assessing non-response rates, nor do they contain information on sample structure for estimating sampling errors and design effects.

Of course, some limitations on the available information result from genuine concerns about preserving confidentiality of the data on households and persons taking part in the surveys. In this connection, we would like to conclude by pointing out a common misinterpretation which has had a serious negative effect on availability of microdata for research and other legitimate purposes.

It is very important to note a major difference between social data based on sample surveys of small and numerous units such as households and persons, and some other types of data, such as those involving complete enumeration or pertaining to a small number of large units (e.g. enterprises) where there is a danger of exposure at the level of the individual unit (Verma, 1998). 'Problems of confidentiality should not arise in the case of microdatabases concerning surveys where items of the data … have identified numbers which cannot be connected by the user to the corresponding names even if used to relate the information to that from a different source; the [proportionately small] size of the sample … and the fact that named files are considered classified … should [usually] guarantee … sufficient respect for the needs of confidentiality. Problems become more sensitive in the case of microdata based on administrative records that aim to cover … the universe of individuals, families, companies [etc.]. In this case [by contrast], concerns felt about confidentiality would normally be well-founded.' (Frey, 1996).

References

Betti, G., Donatiello, G. and Verma, V. (2010), 'The Siena Micro Simulation Model (SM2) for net-gross conversion of EU-SILC income variables', *International Journal of Microsimulation*, 3(2).

Deming, W.E. (1960), *Sample designs in Business Research.* Wiley.

European Community (2003), Regulation (EC) No 1177/2003 of the European Parliament and of the Council. Official Journal of the European Union, pp. L165/1-9.

Frey, L. (1996), 'The strategy for making the microdata base accessible to users: partnership with researchers', in Future of European Social Statistics. Guidelines and Strategies, Series 0D, Eurostat, pp. 155-159.

Frick, J. R., Grabka, M. M. and Groh-Samberg, O. (2010), 'Dealing with incomplete household panel data in inequality research', monograph for SOEP at DIW Berlin.

Husmanns, R., Mehran, F. and Verma, V. (1990), Surveys of the Economically Active Population, Employment, Unemployment and Underemployment, International Labour Organisation, Geneva.

Kish, L. (1965), Survey Sampling. Wiley.

Kish, L. (1995), 'Methods for design effects', *Journal of Official Statistics*, 11, pp. 55-77.

Kish, L., Groves, R. and Krotki, K. (1976), 'Sampling errors for fertility surveys', WFS Occasional Papers no. 17, International Statistical Institute, The Hague.

Rust, K. (1985), Efficient replicated variance estimation. University of Michigan and Australian Bureau of Statistics.

Verma, V. (1981), 'Assessment of errors in household surveys', Bulletin of the International Statistical Institute, 49(2), pp. 905-919.

Verma, V. (1991), Sampling Methods: Training Handbook, Statistical Institute for Asia and the Pacific (SIAP), Tokyo.

Verma, V. (1998), ‘Data sources and access for comparative analyses’, in R. Walker and M. Taylor (editors), Information Dissemination and Access in Russia and Eastern Europe, IOS Press, Amsterdam, pp. 44-54.

Verma, V. (2001), EU-SILC Sampling Guidelines. EU-SILC doc 51/01. Luxembourg: Eurostat.

Verma, V. and Betti, G. (2006), ‘EU Statistics on Income and Living Conditions (EU-SILC): Choosing the survey structure and sample design’, *Statistics in Transition*, 7(5), pp. 935-970.

Verma, V. and Betti, G. (2007), ‘Cross-sectional and longitudinal measures of poverty and inequality: variance estimation using Jackknife Repeated Replication’, Conference 2007 ‘Statistics under one umbrella’, Bielefeld University.

Verma, V. and Clemenceau, A. (1996), Methodology of the European Community Household Panel. Statistics in Transition, 2(7), pp. 1023-1062.

Verma, V. and Lê, T. (1996), ‘An Analysis of Sampling Errors for the Demographic and Health Surveys’, *International Statistical Review*, 64(3), pp. 265-294.

Verma, V., Scott, C. and O’Muircheartaigh, C. (1980), ‘Sample designs and sampling errors for the World Fertility Survey’, *Journal of the Royal Statistical Society*, A, 143(4), pp. 431-473.

European Commission

An assessment of survey errors in EU-SILC

Luxembourg: Publications Office of the European Union

2010 — 61 pp. — 21 x 29.7 cm

ISBN 978-92-79-16757-7
ISSN 1977-0375
Doi:10.2785/56062
Cat. No KS-RA-10-021-EN-N

Theme: Population and social conditions

Collection: Methodologies and working papers

How to obtain EU publications

Free publications:

- via EU Bookshop (http://bookshop.europa.eu);
- at the European Union's representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu) or by sending a fax to +352 2929-42758.

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Priced subscriptions (e.g. annual series of the *Official Journal of the European Union* and reports of cases before the Court of Justice of the European Union):

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm).

KS-RA-10-021-EN-N

ISBN 978-92-79-16757-7

Publications Office