STATISTICS SWEDEN
Göran Råbäck, BV/SV 2006-03-27

Intermediate quality report for the Swedish EU-SILC, the 2005 cross-sectional component

1. CROSS-SECTIONAL EUROPEAN UNION INDICATORS

The Swedish 2005 cross-sectional EU-SILC survey has been carried out in two parts. The main part is, like 2004 integrated with the Swedish survey of living conditions (ULF). As the sample size in ULF is not sufficiently large we have complemented with data from Panel 2 of the longitudinal EU-SILC. (Panel 2 will be excluded from the survey after 2005. The respondents in this Panel are the only ones who have received the questions from the cross-sectional SILC)

The total micro data registers transmitted to Eurostat contain all 2005 cross-sectional indicators stipulated in the regulation. EU-SILC indicators, which will be included in the 2006 spring report, are covered by these data.

The calculation of unadjusted gender pay gap is based on other sources than EU-SILC (wage statistics).

2. ACCURACY

2.1 Sample design for the ULF-survey.

Every year a systematic sample is drawn from the register of the total population (RTB). For this purpose RTB sorted by age is used. RTB covers the entire population according to the national registration. Such a sample is regarded as a simple random sample. All individuals (selected persons) who have been included in ULF at any time during the preceding seven years are eliminated from the sample.

The final sample also includes a panel of people who participated (incl. non-response) in the survey eight years previously.

For 2005 this means that we have:
a. a new sample, see above
b. a sample initially drawn for the 1981 survey. In 1989, 1997 and 2005, when this sample was re-interviewed, the panel was complemented with immigrants and individuals 16-23 who had ”grown into” the population. Only over-coverage (dead and emigrants) were excluded from the sample. The sample (a+b) is regarded as a simple random sample.

2.1.1 Sample unit

The sample unit was individual. Household members living in the same household as the selected person were mapped according to EU-SILC definitions.

2.1.2 Stratification

No stratification was applied in the sampling procedure.

2.1.3 Sample size (households=selected persons)

	Number	Percent
Gross sample	6008	
Over-coverage	82	
Interview	4 432	74,8
Refusal	957	16,1
Not found	403	6,8
Other reason	132	2,2

2.2 Sample design for the SILC Panel 2-survey.

In 2004 a systematic simple random sample is drawn from the register of the total population (RTB). All individuals (selected persons) who have been included in ULF at any time during the preceding seven years were eliminated from the sample. In 2005 this sample was complemented with a sample among immigrants since 2004 and individuals 16 years old who had ”grown into” the population. Only over-coverage (dead and emigrants) were excluded from the sample.

2.2.1 Sample unit

Like in the ULF-survey the sample unit was individual. Household members living in the same household as the selected person were mapped according to EU-SILC definitions.

2.2.2 Stratification

No stratification was applied in the sampling procedure.

2.2.3 Sample size (households=selected persons)

	Number	Percent
Gross sample	2351	
Over-coverage	19	
Interview	1704	73,1

Refusal	310	13,3
Not found	238	10,2
Other reason	80	3,4

2.3 Weightings - Design factor and non-response adjustment

2.3.1 Combining the samples.

The different samples are added together and divided into 8 strata according to the definitions above

***Table 1**: Stratum definitions.*

	Age in year 2005				
Immigrant status	**16-23**	**24-31**	**32-39**	**40-84**	**85-**
Born in Sweden or immigrated - 1981				1	
Immigrated 1982-1989			2	5	
Immigrated 1990-1997		3	6		
Immigrated 1998-2005	4	7			8

***Table 2**: Sampling year of each stratum and survey.*

Stratum	ULF 2005	ULF panel	SILC panel 2
1	2005	1981	2004/2005
2	2005	1989	2004/2005
3	2005	1997	2004/2005
4	2005	2005	2004/2005
5	2005	1989	2004/2005
6	2005	1997	2004/2005
7	2005	2005	2004/2005
8	2005		2004/2005

2.3.2 Adjusted design-weights.

Within each stratum the design-weights are computed as the inverse of the probability of inclusion. Then the design-weights are adjusted according to the over-coverage.The observed allocation of the combined sample over strata is given in Table 2 below.

***Table 3**: Composition of the combined sample*

Stratum, *h*	**Population 2005,** N_h	**Observed no of obs,** n_h	**Over-cove-rage**	**Adjusted no of obs,** $n_{h,adj}$	**Adjusted weight,** $d_{hi,adj}$
1	4 068 410	4606	28	4578	875,66
2	892 552	993	8	985	892,29
3	812 752	920	13	907	895,33
4	889 460	986	14	972	879,26
5	101 440	118	4	114	898,99
6	164 998	183	7	176	834,08
7	199 317	231	12	219	772,53
8	227 942	322	15	307	694,67
Total	7 356 871	8359	101	8258	

2.3.3 Final weights

To compensate for non-response the data was partitioned into 14 strata, or response homogeneity groups (RHG), according to age and sex as described in Table 4 below.

***Table 4**: Response homogeneity groups, sample and response set sizes*

	Sample, n_g			**Responding,** m_g		
	Sex			**Sex**		
Age	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**
16-23	510	476	986	386	366	752
24-39	1 070	1 034	2 104	765	774	1 539
40-49	690	686	1 376	496	535	1 031
50-64	1 055	991	2 046	767	759	1 526
65-74	427	444	871	313	338	651
75-84	284	370	654	192	252	444
85+	103	219	322	65	128	193
Total	4 139	4 220	8 359	2 984	3 152	6 136

Using the adjusted design weights of Table 3, estimates of the population size of each RHG *g*, is calculated as

$$\hat{N}_g = \sum_{i \in R \cap \text{RHG } g} d_{hi,adj}$$

where R denotes the subset of responding individuals. As we know the true population sizes for each RHG, *a*djustment factors for each RHG g are calculated as the ratio between the true population size and the estimated population size, giving the final weights. Thus, for an individual belonging to stratum h, RHG g and the respons-set R, the weight is calculated as

$$d_{i,\text{final}} = \text{I}(i \in R) \times d_{hi,adj} \times N_g / \hat{N}_g$$

where $\text{I}(i \in R)$ denotes an indicator function equal to 1 when the argument is true.

The calculation of weights does not consider the circumstance that all individuals belonging to a stratum have not initially been drawn with exactly the same sampling probability. The impact of this is however negligible.

So far the text has referred to the sampled person. Weights for all household members (16+) are created analogously. Household weights can be derived from weights for individuals based on information about the number of household members who give the household sampling probability.

2.3.4 Substitutions

Substitution has not been applied.

2.4 Sampling errors

Information concerning effective sample sizes and standard errors for the common cross-sectional EU indicators is available in attached EXCEL files.

2.5 Non-sampling errors

2.4.1 Sampling frame and coverage errors

The sampling frame is RTB, se above. RTB is updated more or less every day. Over-coverage consists of people who have died and people who have left the country but are still registered in Sweden. The sample is drawn several months before the fieldwork start. However a check is made close to the start (the sample is matched to RTB) and people who have died since the sample was drawn are excluded. People who die after that point are registered by the interviewers.

Over-coverage in terms of people who have left Sweden permanently but are still registered in RTB is more difficult to discover. Recent attempts to estimate the size of this over-coverage have given the figure 35 000. Applied on EU-SILC this means 30 individual of which many are discovered by the interviewers. The error is negligible.

If we regard RTB as our population under-coverage by definition does not exist. There are of course people who reside in Sweden illegally or while waiting for residence permit.

2.5.2 Measurement and processing errors

2.5.2.1 *Measurement errors*

Following a basic introductory course in survey methods, new interviewers participate in an additional one-day course that includes approximately six ours of intensive training (ULF including EU-SILC). The various sections of the interview protocol are thoroughly reviewed, and practice in handling certain complicated questions is provided.

The interviewer may miss-understand certain instructions ore responses, which contributes to the survey´s systematic error level. Each interviewer conducts on average roughly 40 interviews per year. Systematic mistakes by an occasional interviewer may not distort the survey data to any great extent, but it is not possible to specify how much error of that sort occurs.

The interviewer´s personality and behaviour may influence the responses, particularly with respect to ”subjective” questions, such as those relating to attitudes.

In some cases interview questions are not presented properly. To the extent that such mistakes cannot subsequently be corrected, there is an increase in partial response.

The respondent may disremember, provide consciously or unconsciously distorted responses or may simply be unable to answer questions.

Most of the ULF/EU-SILC questions refer to the present, for which memory errors can not constitute a major source of error. But there are questions about frequency during a longer reference period that are more complicated. .

The questions in the ULF/EU-SILC protocol are in most cases not very difficult to answer. It is fairly certain that some questions are interpreted differently by different persons. Particular caution should be observed of responses to questions relating to attitudes and frequency in the interpretation.

Data collection methods.

The ULF data are collected primarily through face-to-face interviews. The interview form has been specially designed for this type of survey. Telephone interviews are normally used to follow up non-responses, but are in some cases used as part of the regular interviewing process.

The EU-SILC data are collected primarily through telephone- interviews.

Experiments with split samples have been carried out. The results indicate very little difference between the two interview method.

Indirect interviews can be a source of errors. Applied on appropriate questions experience says that indirect interviews can be an efficient method to collect information.

2.5.2.2 Processing errors

Data processing errors. Data are checked interactively (values, syntax, logics) as an integrated part of the data entry process. (CAPI/CATI is not applied) followed by the Eurostat controll program (after transformation to EU-SILC file format).

All components necessary to derive Gross total income, disposable income etc. are collected from administrative registers. No imputations have been applied for these indictors.

2.5.3 Non-response errors

2.5.3.1 Achieved sample size

See 2.1.3 and 2.2.3 above concerning households and selected persons.

The data file on individuals contains information for 15 526 individuals 16+. Response rate is not possible to calculate as household composition for non-response households is not completely known.

2.3.3.5 Item non-response

Calculations of income variables are based on administrative register data. Imputation procedures are consequently not necessary.

2.4 Mode of data collection

Face-to-face 77,5 percent, 22,5 percent telephone.

2.5 Interview duration

Approximately 65 minutes per household.

3. COMPARABILITY

3.1 Basic concepts and definitions

- The reference population
Short term migration, people who stay in Sweden 3-12 months, is not covered.

- Private household definition
The regulation definition is applied.

- The household membership
The regulation definition is applied

- the income reference period used
Year n-1

- the period for taxes on income and social insurance contributions
Year n-1

- the lag between the income reference period and current variables
The field work is carried out during January-December year N.

- the total duration of the data collection of the sample
12 month, January-December

- basic information on activity status during the income reference period
The twelve calendar months preceding the month of the interview

3.2 Components of income

3.2.1 Differences between national definitions and standard EU-SILC definitions

Only minor deviations with little impact on the results:

- non-cash employee income includes more than company car (housing cost/ interest on loans below market price etc).
- regular inter-household cash transfers paid/received do only consider transactions between parents not living together. Other types of alimonies or cash transfers are not included.

3.2.2 The source or procedure used for collection of income variables

Administrative registers

3.2.3 The form in which income variables at component level have been obtained

Gross but excusive of employers social contributions

3.2.4 The method for obtaining income target variables in the required format

Available from administrative registers.

4. COHERENCE

4.1 Comparison of income target variables and number of persons who receive each ”income component” with external sources

The EU-SILC information is collected from the different administrative sources covering the whole population. The non-response bias has little impact on the estimates.