

# **Final quality report for the Swedish EU-SILC, the 2005 cross-sectional components.**

## **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 2007 spring report, are covered by these data.

In Swedish national statistics the calculation of Equivalised disposable is based on the same principal as in EU-SILC but we use different and more discriminating consumer weights.

In Swedish national statistics 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-part of the survey.**

Every year a systematic sample is drawn from the register of the total population (TPR). For this purpose TPR sorted by age is used. TPR 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.

The total ULF-sample 2005 consists of two independent parts, both of them representing the population aged 16-84 years. To cover the population over 84 years we have made an extra sample:

1. "The ordinary sample". The individuals in this sample have only been interviewed this time and we are not going to include them in

any further sample. The extra sample among the oldest part of the population also belongs to this part.

2. "The panel". The individuals in this sample are drawn 1981 and then we have included them in the survey 1989, 1997 and now 2005 for the 4:th time. In order to cover the total population 16-74 years old 2005 we must complement the original sample with young people and with immigrants.

### 2.1.1 Type of sampling design

The principal of our sampling is a stratified sample with approximately the same sampling fraction within each stratum.

#### 2.1.1 Sample unit

The unit of study of interest is both individuals and households but the sample unit is individuals in TPR aged 16 years and older. Household members living in the same household as the selected person were mapped according to EU-SILC definitions. It is not possible to find all household members using TPR as a sampling frame. We can find persons who are married with the selected persons and who have children under 18 years together with the selected persons and children belonging to these households. Household members in other types of households can not be included in the sampling phase. For this reason it is only possible to detect the correct household consistence for the respondent individuals in the sample.

#### 2.1.2 Stratification

For practical use we can look upon the sample as if no stratification was applied in the sampling procedure. From theoretical point of view we can look at the total sample as a stratified sample from nine strata where the effort has been to make a sampling design that is approximately simple random sampling. We have tried to keep the sample fractions within each stratum as equivalent as possible. In the following schedule we show the contents of the different strata and the sample fractions.

#### 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

### **2.1.5 Sample selection schemes**

From estimation point of view we can look at the total sample as a stratified sample from nine strata. The effort has been to make a sampling design that is approximately simple random sampling. We have tried to keep the sample fractions within each stratum as equivalent as possible. In the following schedule we show the contents of the different strata and the sample fractions.

### **2.1.6 Sample distribution over time**

The original sample was drawn in December 2004 and randomly distributed into four parts. Each part was dedicated to one of the four quarters of 2005 and during three months the data collection was carried out for the whole sample in the actual part.

We made an over-sampling from the oldest part of the population (85- years) at the original time. From each part a sub-sample was made among those still living before the data collection. The reason for this was to avoid over-coverage as much as possible.

### **2.1.7 Renewal of sample: Rotation groups**

The whole sample was divided into the four parts as described above. The respondents in “the original sample” never come back in some form of rotation. To the sample in “the panel” we will return after eight years (2012). In that sample also complemented young people (16-24 years) and immigrants are included.

## **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 (TPR). 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. The unit of study of interest is both individuals and households but the sample unit is individuals in TPR aged 16 years and older. Household members living in the same household as the selected person were mapped according to EU-SILC definitions. It is not possible to find all household members using TPR as a sampling frame. We can find persons who are married with the selected persons and who have children under 18 years together with the selected persons and children belonging to these households. Household members in other types of households can not be included in the sampling phase. For this reason it is only possible to detect the correct household consistence for the respondent individuals in the sample.

### 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

### 2.1.6 Sample distribution over time

The original sample for the SILC-panel was drawn in August 2004 and randomly distributed into four parts, panel1 to panel4. The data collection was carried out for the whole sample in the last quarter of 2006

### 2.1.7 Renewal of sample: Rotation groups

The sample in “panel 2” was included in the survey 2004 and for the last time 2005. Like in the panel-part of the ULF survey the sample is complemented with young people and immigrants included in the population since 2004.

## 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.*

Immigrant status	Age in year 2005				
	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-coverage	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
<b>Total</b>	<b>7 356 871</b>	<b>8359</b>	<b>101</b>	<b>8258</b>	

### 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*

Age	Sample, $n_g$			Responding, $m_g$		
	Sex			Sex		
	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
<b>Total</b>	<b>4 139</b>	<b>4 220</b>	<b>8 359</b>	<b>2 984</b>	<b>3 152</b>	<b>6 136</b>

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, adjustment 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

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

where  $I(i \in R)$  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 the Total Population Register of Sweden (TPR).**

#### **Main outlines for organization of population Statistics**

According to Swedish law, the main rule is that all persons residing in the country shall be registered at the property unit in the parish where they reside.

Since 1 July 1991, local registration functions are performed by the Tax Offices. Between 1686 and 1991, the Parish Offices of the Church of Sweden carried out the local work.

A major means of identifying any person is the personal identity number that is assigned to every individual registered in the Population Registration System. The number follows a person from birth to death and is entered in most personal registers in Sweden, making it possible to identify individuals in different administrative materials and collate data. The personal identity number consists of ten digits. The first six digits show the year, month and day of birth. The next three digits are the birth number which is odd for men and even for women. The last digit is a check digit.

### **Raw data**

As part of the partial computerization of Sweden's continuous population registration in 1966, Statistics Sweden was granted permission to set up and maintain a register of the entire national population, referred to as the Total Population Register (TPR)

The vital statistics are based on notifications of births, deaths, changes in marital status, and changes in citizenship, internal migration, immigration and emigration. The TPR receives these daily from the Tax Authorities.

The notifications relate to the registered population. Thus, vital statistics are based on the National Registration and consequently conform to its concepts and definitions.

### **Quality tests**

Received information is checked mechanically with respect to the validity of the codes and the logical contents of the information. The inspection comprises, among other things, regional codes, connections between age and marital status, etc.

### **Cut-off date**

Beginning in 1998 the cut-off date is 31 January in the year after the event took place. The change in cut-off date in 1998 will have no effect on comparisons between years.

The sampling frame is TPR, see above. TPR 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 TPR) 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 TPR 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 TPR 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 hours 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 or 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 data ULF-part 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.

In the Panel 2-part 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 control 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 indicators.

#### 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**

## **2.5 Interview duration**

Interview duration was approximately 65 minutes per household in the ULF-part and approximately 25 minutes per household in the panel 2-part. The specific cross-sectional questions duration was about 20 minutes.

## **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 exclusive 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.