CENTRAL STATISTICAL OFFICE OF POLAND

FINAL QUALITY REPORT

GRANT AGREEMENT
NO 36401.2007.001-2007.161

ACTION ENTITLED:
EU-SILC 2008

Warsaw, November 2010

"This document has been produced with the financial assistance of the European Community. The views expressed herein are those of the author and can therefore in no way be taken to reflect the official opinion of the European Commission."

CONTENTS Page

PREFACE **3**

1. COMMON LONGITUDINAL EUROPEAN UNION INDICATORS **4**

2. ACCURACY **4**

2.1. Sample design 4

2.1.1. Type of sampling design *4*
2.1.2. Sampling units *4*
2.1.3. Stratification and substratification criteria *5*
2.1.4. Sample size and allocation criteria *5*
2.1.5. Sample selection schemes *5*
2.1.6. Sample distribution over time *5*
2.1.7. Renewal of sample: rotational groups *5*
2.1.8. Weightings *6*
2.1.9. Substitutions *10*

2.2. Sampling errors 11

2.3. Non-sampling errors 15
2.3.1. Sampling frame and coverage errors *15*
2.3.2. Measurement and processing errors *16*
2.3.3. Non-response errors *17*

2.4. Mode of data collection 34

2.5. Imputation procedure 37

2.6. Imputed rent 41

2.7. Company cars 43

3. COMPARABILITY **43**

3.1. Basic concepts and definitions 43

3.2. Components of income 46
3.2.1. Differences between the national definitions and standard EU-SILC definitions *46*
3.2.2. The source or procedure used for the collection of income variables *48*
3.2.3. The form in which income variables at component level have been obtained *48*
3.2.4. The method used for obtaining income target variables in the required form *48*

3.3. Tracing rules 48

4. COHERENCE **49**

4.1. Comparison of EU-SILC and HBS results 49

4.2. Comparison of income data from SNA for the household sector and EU-SILC 52

PREFACE

This quality report is the final quality report on EU-SILC 2008 carried out in Poland according to grant agreement No. 36401.2007.001-2007.161, as provided for in Council Regulation No 1177/2003. It follows the structure outlined in Commission Regulation No. 28/2004. This report provides information on accuracy, comparability and coherence of data with external sources.

The indicator on persistence of poverty, which is presented for the first time in the context of EU-SILC, was calculated using the longitudinal rotation 2005-2008.

1. COMMON LONGITUDINAL EUROPEAN UNION INDICATORS

Persistent-at-risk-of-poverty rate by age and gender (60% median)			
No.	**Age**	**Gender**	**(%)**
1	Total (AGE ≥ 0)	T	10.41
2		M	10.69
3		F	10.15
4	(0 ≤ AGE ≤ 17)	T	15.85
5	(18 ≤ AGE ≤ 64)	T	10.20
6		M	10.80
7		F	9.63
8	(AGE ≥ 65)	T	5.41
9		M	2.66
10		F	7.01

Persistent-at-risk-of-poverty rate by age and gender (50% median)			
No.	**Age**	**Gender**	**(%)**
1	Total (AGE ≥ 0)	T	5.15
2		M	5.56
3		F	4.79
4	(0 ≤ AGE ≤ 17)	T	8.40
5	(18 ≤ AGE ≤ 64)	T	5.22
6		M	5.57
7		F	4.87
8	(AGE ≥ 65)	T	1.40
9		M	0.92
10		F	1.68

2. ACCURACY

2.1. Sample design

2.1.1. Type of sampling design

The two-stage sampling scheme with diversified selection probabilities at the first stage was used. Prior to selection, sampling units were stratified.

2.1.2. Sampling units

The first-stage primary sampling units (PSUs) were census areas, while at the second stage dwellings were selected.

2.1.3. Stratification and substratification criteria

The strata were the voivodships (NUTS2), while within voivodships primary sampling units were classified by class of locality. In urban areas census enumeration areas were grouped by size of town, but in the five largest cities districts were treated as strata. In rural areas strata were represented by rural gminas (NUTS5) of a subregion (NUTS3) or of a few neighbouring poviats (NUTS4). Altogether 211 strata were distinguished.

2.1.4. Sample size and allocation criteria

It was decided that the sample should include some 24 000 dwellings. Proportional allocation of dwellings to particular strata was applied. The number of dwellings selected from a particular stratum was in proportion to the total number of dwellings in the stratum. Furthermore, the number of the first-stage units selected from the strata was obtained by dividing the number of dwellings in the sample by the number of dwellings determined for a given class of locality to be selected from the first-stage unit. In towns with over 100 000 population 3 dwellings per PSU were selected, in towns with 20-100 thousand population – 4 dwellings per PSU, in towns with less than 20 000 population – 5 dwellings per PSU, respectively. In rural areas 6 dwellings from each PSU were selected. Altogether 5912 census areas and 24044 dwellings were selected for the sample. The subsample 5 was selected for the survey in 2006 in order to replace the subsample 1. It consisted of 1476 census areas and 6002 dwellings. For the cross-sectional component of the survey conducted in 2007 a new subsample (the subsample 6) was selected. It replaced the subsample 2, and consisted of 1487 PSUs and 6008 dwellings. For the cross-sectional component of the survey conducted in 2008 the subsample 7 was selected. It replaced the subsample 3, and consisted of 1479 PSUs and 6016 dwellings. The subsample 4 formed the four year longitudinal (panel) component.

2.1.5. Sample selection schemes

Census areas were selected according to the Hartley-Rao scheme. Prior to selection, census areas were put in a random order, for each stratum separately and then the required number of PSUs was selected with probabilities proportionate to the number of dwellings. Next, in each of the census areas belonging to the PSU sample, dwellings were selected using the simple random selection procedure.

2.1.6. Sample distribution over time

The sample is not distributed over time.

2.1.7. Renewal of sample: rotational groups

The selected sample of first-stage units was divided into four subsamples, equal in size. Starting from 2006 one of the subsamples was eliminated and replaced with another one, selected independently as described above. For the 2006 survey the subsample 5 was selected as a replacement of the subsample 1. Then, for the 2007 survey the subsample 6 was selected to replace the subsample 2, and for the 2008 survey the subsample 7 was selected to replace the subsample 3. Rotation comprised first-stage units.

2.1.8. Weightings

Design factor

Design factor – DB080 is equal to the dwelling sampling fraction reciprocal in the h-th stratum i.e.

$$f_h = \frac{n_h * m'_h}{M_h},$$

$$DB080 = \frac{1}{f_h}$$

where:
n_h - number of PSUs selected from the h-th stratum,
m'_h - number of dwellings selected from PSUs in the h-th stratum,
M_h – number of dwellings in the h-th stratum.

Non-response adjustment

DB080 weights were then adjusted with the use of the completeness indicator, estimated for each class of locality separately:

Code of class of locality (p)	**Class of locality**	**Completeness rate (Ra_p*Rh_p)**
	Poland	0.654
1	Warsaw	0.404
2	Towns 500 000 – 1 000 000 inhabitants	0.535
3	Towns 100 000 – 500 000 inhabitants	0.581
4	Towns 20 000 – 100 000 inhabitants	0.637
5	Towns less than 20 000 inhabitants	0.665
6	Rural areas	0.787

The adjusted weights were calculated according to the formula:

$$DB080_p^{corrected} = \frac{DB080_p}{Ra_p * Rh_p},$$

The weights DB080 and $DB080^{corrected}$ were calculated for the subsample 7. The next step consisted in calculating the weights DB090 and RB050 for the households of the subsample 7 with the use of the integrated calibration method as described below in the section "*Adjustments to external data*".

Adjustments to external data

With the use of the integrated calibration method (in its hyperbolic sinus version) weights were calculated for individuals and for households simultaneously. To do this, the following information about households was used: 4 size categories (1-person, 2-person, 3-person and 4- and more person households) and number of persons by age and gender (14 age groups: under 16, 16-19 years, then eleven 5-year groups, 75 years and over). The data at the level of NUTS2, additionally classified by urban/rural areas, were derived from the 2002 Census and current demographic estimates.

Final longitudinal weight

For the subsamples 2, 3 and 4, surveyed for the second time, the base weights were determined by the correction of the base weights from the previous year. The base weight of 2005 is equal to RB050 multiplied by 4. This weight was then adjusted by non-response and households' and individuals' falling out of the population surveyed. The calculations were performed on the subsets of the so called *sample persons* i.e. those who were at the age of 14 and over when participating in the surveyed sample in 2005 and who should be surveyed in 2006. The modifying factor was determined for each subsample (2, 3 and 4) separately according to the class of locality and took the form:

$$\frac{R(1)_p - M}{R(2)_p},$$

where:

$R(t)_p$ – estimated number of respondents belonging to the "*sample person*" group in the p-th class of locality in the subsample surveyed for the t-th time (t = 1, 2),

M – estimated number of "*sample persons*" who belonged to the surveyed population in the first year and in the next year were out of the survey scope.

The base weights of 2005 were used for the calculation of numerator and denominator. The above expression is the reciprocal of the empirical estimate of probability that a given person will be interviewed again in the second year of the survey.

In the second stage of the base weight calculation for the second year of the survey children of "*sample persons*" received the weights of mothers and "co-residents' i.e. additional persons included in the household surveyed were ascribed zero weights. Then the respondents' weights were averaged and all the members of a given household were ascribed such a mean weight. To the base weights thus obtained the trimming of extreme weights was applied.

Adjustment to external data was not made. The panel weight RB062 was calculated by dividing the base weights by 3.

Non-response adjustments – subsequent waves

Third wave

For the subsamples 3 and 4 surveyed for the third time and the subsample 5 surveyed for the second time the base weights were determined by the correction of the base weights from the previous year.
For the subsample 5 the following method was applied:
The base weight of 2006 is equal to RB050 multiplied by 4. This weight was then adjusted by non-response and households' and individuals' falling out of the population surveyed. The calculations were made on the subsamples of the so called sample persons i.e. those who were at the age of 14 and over when participating in the surveyed sample in 2006 and who should be surveyed in 2007. The modifying factor was determined according to the class of locality and took the form:

$$\frac{R(1)_p - M}{R(2)_p}$$

where:

$R(t)_p$ – estimated number of respondents belonging to the sample person group in the p-th class of locality in the subsample surveyed for the t-th time,
M – estimated number of sample persons who belonged to the surveyed population in the first year and in the next year were out of the survey scope.

The base weights of 2006 were used for the calculation of numerator and denominator. The above expression is the reciprocal of the empirical estimate of probability that a given person will be interviewed again in the second year of the survey. In the second stage of the base weight calculation for the second year of the survey children of "sample persons" received the weights of mothers and "co-residents' i.e. additional persons included in the household surveyed were ascribed zero weights.

For the subsamples 3 and 4 (surveyed for the third time) the algorithm based on the method described for the subsample 5 was used. Additionally, re-entries, i.e. persons who were surveyed in 2005, not surveyed in 2006, and surveyed in 2007 again, were taken into account. The base weights for such persons were computed by the correction of base weights from 2005 on the data for 2005 and 2007 (wit no data for 2006). The inclusion of re-entries in the subsamples surveyed in 2007 made it necessary to additionally correct the base weights for the persons surveyed in the three successive years. Coefficients of these corrections were computed separately according to the class of locality as ratios between the weighted number of respondents surveyed in all the three years and the weighted number of respondents in the last year of survey (i.e. with re-entries); the weight used in these calculations was the weight RB050 for 2005.
The last stage of the base weight calculation for the third year of the survey consisted in receiving weights of mothers by children of "sample persons" and zero weights by "coresidents' i.e. additional persons included in the households.

Fourth wave

For the subsample 4 surveyed for the fourth time, the subsample 5 surveyed for the third time, and the subsample 6 surveyed for the second time, the base weights were determined by the correction of the base weights from the previous year.
For the subsample 6 the following method was used:
The base weight of 2007 is equal to RB050 multiplied by 4. This weight was then adjusted by non-response and households' and individuals' falling out of the population surveyed. The calculations were made on the subsamples of the so called sample persons i.e. those who were at the age of 14 and over when participating in the surveyed sample in 2007 and who should be surveyed in 2008. The modifying factor was determined according to the class of locality and took the form:

$$\frac{R(1)_p - M}{R(2)_p}$$

where:

$R(t)_p$ – estimated number of respondents belonging to the sample person group in the p-th class of locality in the subsample surveyed for the t-th time,
M – estimated number of sample persons who belonged to the surveyed population in the first year and in the next year were out of the survey scope.

The base weights of 2007 were used for the calculation of numerator and denominator. The above expression is the reciprocal of the empirical estimate of probability that a given person will be interviewed again in the second year of the survey. In the second stage of the base weight calculation for the second year of the survey children of "sample persons" received the weights of mothers and "co-residents' i.e. additional persons included in the household surveyed were ascribed zero weights.
For the subsamples 4 and 5 (surveyed for the fourth and third time respectively) the algorithm based on the method described for the subsample 6 was used. Additionally, re-entries, i.e. persons who were surveyed in 2006, not surveyed in 2007, and surveyed in 2008 again, were taken into account. The base weights for such persons were computed by the correction of the base weights from 2006 on the data for 2006 and 2008 (with no data for 2007). The inclusion of re-entries in the subsamples surveyed in 2008 brought about the necessity to additionally correct the base weights for persons surveyed in the three successive years. Coefficients of these corrections were computed separately according to the class of locality as ratios between the weighted number of respondents surveyed in all the three years and the weighted number of respondents in the last year of the survey (i.e. with re-entries); the weight used in these calculations was the weight RB050 for 2006.

Adjustments to external data

Adjustment to external data was not applied.

Final longitudinal weights

The panel weight RB062 was calculated:

1) by taking the base weights for subsamples 4, 5 and 6,
2) by giving a zero value to people not present in the two waves (like for example the newly born),
3) by dividing the obtained weights by 3.

The panel weight RB063 was calculated with a similar procedure, that is:

1) by taking the base weights for subsamples 4 and 5,
2) by giving a zero value to people not present in the three waves,
3) by dividing the obtained weights by 2.

The panel weight RB064 was also calculated with a similar procedure, that is:

1) by taking the base weights for the subsample 4,
2) by giving a zero value to people not present in the four waves.

Final household cross-sectional weight

The last stage of calculations consisted in combining the four independent subsamples, applying the integrated calibration and trimming of extreme weights. As a result the following cross-sectional weights were calculated for households and individuals from the samples 2, 3, 4 and 5 in EU-SILC 2006:

DB090 – weight for households,
RB050 – weight for all household members but
$RB050_{ij} = DB090_i$

where:
i – household number,
j – person number in the i-th household.

PB040 – weight for respondents at the age of 16 and over who had individual interviews. This weight is obtained by the adjustment of RB050 separately in the groups according to gender and age in each voivodship by urban and rural area,

RL070 – weight for children at the age of 0–12 years. It is obtained by the adjustment of RB050 weight in 26 groups, i.e. 13 years of birth and gender.

Final cross-sectional weights for EU-SILC 2007 were calculated in a similar way for households and individuals from the samples 3, 4, 5 and 6. This is documented in the EU-SILC 2007 Intermediate Quality Report.
Final cross-sectional weights for EU-SILC 2008 were calculated in a similar way for households and individuals from the samples 4, 5, 6 and 7. This is documented in the EU-SILC 2008 Intermediate Quality Report.

2.1.9. Substitutions

No substitution was applied if the household did not enter the survey.

2.2. Sampling errors

Standard error and effective sample size

Estimation of standard errors was based on the resampling approach. We used a bootstrap method which resamples 200 times from each stratum $n_h - 1$ PSUs (primary sampling units) with replacement (McCarthy and Snowden method (1985)), where n_h the number of PSUs selected for the sample size selected from each PSUs in the *h*th stratum. After resampling the original weights were properly rescaled and bootstrap variance estimate of the corresponding indicator was obtained by the usual Monte Carlo approximation based on the independent bootstrap replicates. Computations were carried out using SAS software. Additionally, we implemented the linearization method of variance estimation for the main poverty indicators, and the results were similar to those obtained with the bootstrap method.

Cross-sectional component

The mean, the total number of observations (before and after imputation) and the standard errors for the following income components

Income components	Mean[1]	Standard error	Mean[2]	Standard error	Number of observations	
					Before imputation	After imputation
Total household gross income (HY010)	45575	440	45447	439	4862	13949
Total disposable household income (HY020)	34578	320	34543	319	9751	13970
Total disposable household income before social transfers other than old-age and survivor's benefits (HY022)	32831	324	32396	321	9759	13808
Total disposable household income including old-age and survivor's benefits (HY023)	27490	366	24445	340	8653	12455
Net income components at household level						
HY040N	6968	692	98	13	119	188
HY050N	3165	72	568	18	2710	2830
HY060N	1643	73	70	5	601	628
HY070N	1455	43	58	3	538	554
HY080N	5405	215	325	19	732	823
HY090N	7218	1071	165	27	150	277
HY110N	2340	147	68	6	500	514
HY120N	216	4	119	2	7823	7823
HY130N	4740	236	234	16	630	678
HY140N	10883	124	10688	123	13750	13750
HY145N	-364	46	-137	17	5130	5130

[1] Taking into account only households/persons receiving such income.
[2] Taking into account whole population (households/persons) surveyed.

Income components	Mean[1]	Standard error	Mean[2]	Standard error	Number of observations	
					Before imputation	After imputation
Gross income components at household level						
HY040G	7952	824	111	15	158	188
HY050G	3262	78	585	19	2580	2830
HY060G	1643	73	70	5	601	628
HY070G	1455	43	58	3	538	554
HY080G	5405	215	325	19	732	823
HY090G	8821	1308	201	33	59	277
HY110G	2482	166	72	6	450	514
HY120G	216	4	119	2	7823	7823
HY130G	4740	236	234	16	630	678
HY140G	10725	130	10550	129	13772	13772
Net income components at personal level						
PY010N	19818	229	8859	127	10485	13795
PY020N	1052	43	150	7	2757	4338
PY035N	1898	58	69	4	1051	1051
PY050N	14177	410	1296	46	2109	3228
PY080N	19873	6323	4	2	5	5
PY090N	5429	303	141	9	742	907
PY100N	13397	79	3276	48	8182	9020
PY110N	8217	330	94	7	379	441
PY120N	3080	672	11	3	102	128
PY130N	7573	93	440	13	1861	2065
PY140N	2183	142	23	2	316	358
Gross income components at personal level						
PY010G	27220	306	12167	173	4962	13795
PY020G	1052	43	150	7	2757	4338
PY035G	1898	58	69	4	1051	1051
PY050G	18690	522	1864	64	1817	3561
PY080G	23045	7332	5	3	1	5
PY090G	6360	405	166	12	498	907
PY100G	15532	94	3798	56	4905	9020
PY110G	9455	386	109	8	168	441
PY120G	3590	799	13	3	63	128
PY130G	8627	108	501	15	1013	2065
PY140G	2183	142	23	2	316	358
PY200G	2269	20	842	11	27770	30031

The mean, the number of observations (before and after imputation) and the standard errors for the equivalised disposable income breakdown by sex, age groups and household size

Equivalised disposable income	Mean[3]	Standard error	Mean[4]	Standard error	Number of observations	
					Before imputation	After imputation
Subclasses by household size						
1 household member	16911	330	16883	329	2283	2634
2 household members	21239	406	21201	405	5852	7648
3 household members	21007	332	21001	332	5478	8553
4 and more	17312	233	17307	232	12865	22336
Population by age group						
<25	17325	211	17303	211	8155	13232
25 to 34	20773	295	20771	295	3033	5244
35 to 44	19440	356	19432	356	3245	5101
45 to 54	19432	269	19409	268	3953	6423
55 to 64	19602	391	19597	391	3469	5176
65+	17049	159	17049	159	4623	5995
Population by sex						
Male	18933	181	18916	180	12479	19756
Female	18477	160	18469	159	13999	21415

Longitudinal component

The mean, the total number of observations (before and after imputation) and the standard errors for the following income components , subsample 4.

Income components	Mean[5]	Standard error	Mean[6]	Standard error	Number of observations	
					Before imputation	After imputation
Total household gross income (HY010)	43486	730	43392	729	1228	3204
Total disposable household income (HY020)	33262	524	33216	523	2286	3206
Total disposable household income before social transfers other than old-age and survivor's benefits (HY022)	31380	534	31030	535	2285	3178
Total disposable household income including old-age and survivor's benefits (HY023)	25409	611	22153	544	1973	2831

[3] Taking into account only households/persons receiving such income.
[4] Taking into account whole population (households/persons) surveyed.
[5] Taking into account only households/persons receiving such income.
[6] Taking into account whole population (households/persons) surveyed.

Income components	Mean[5]	Standard error	Mean[6]	Standard error	Number of observations	
					Before imputation	After imputation
Net income components at household level						
HY040N	7537	1622	93	25	25	39
HY050N	3185	129	564	36	627	647
HY060N	1689	143	71	10	139	143
HY070N	1441	81	63	7	139	141
HY080N	4614	336	286	29	188	204
HY090N	7466	2587	183	66	25	69
HY110N	2276	222	60	8	99	102
HY120N	216	7	118	4	1755	1755
HY130N	4144	306	212	22	153	164
HY140N	10135	216	9981	214	3162	3162
HY145N	-395	52	-136	18	1103	1103
Gross income components at household level						
HY040G	8664	1887	107	29	35	39
HY050G	3249	135	576	38	602	647
HY060G	1689	143	71	10	139	143
HY070G	1441	81	63	7	139	141
HY080G	4614	336	286	29	188	204
HY090G	9137	3149	224	80	10	69
HY110G	2392	245	63	8	87	102
HY120G	216	7	118	4	1755	1755
HY130G	4144	306	212	22	153	164
HY140G	9991	220	9845	218	3165	3165
Net income components at personal level						
PY010N	18993	354	8063	197	2344	3063
PY020N	1021	75	138	12	592	955
PY035N	1877	109	64	7	219	219
PY050N	13818	783	1250	93	483	729
PY080N	8400	0	1	1	1	1
PY090N	6156	1034	159	29	164	199
PY100N	13496	154	3696	103	1949	2096
PY110N	8382	571	115	14	110	117
PY120N	3056	969	7	3	19	19
PY130N	7477	165	455	24	466	493
PY140N	1869	264	19	4	73	79
Gross income components at personal level						
PY010G	26138	494	11096	274	1184	3063
PY020G	1021	75	138	12	592	955
PY035G	1877	109	64	7	219	219
PY050G	18062	1008	1776	129	423	801
PY080G	9741	0	1	1	0	1
PY090G	7283	1391	188	38	122	199
PY100G	15641	181	4283	120	1265	2096
PY110G	9636	659	133	16	51	117
PY120G	3574	1159	8	4	9	19
PY130G	8488	191	516	27	267	493
PY140G	1869	264	19	4	73	79
PY200G	2214	41	784	20	6311	6816

The mean, the number of observations (before and after imputation) and the standard errors for the equivalised disposable income breakdown by sex, age groups and household size (mean and standard errors based on weighted data, the number of observations based on unweighted results), subsample 4.

Equivalised disposable income	Mean[7]	Standard error	Mean[8]	Standard error	Number of observations	
					Before imputation	After imputation
Subclasses by household size						
1 household member	15544	450	15484	447	620	716
2 household members	20212	489	20212	489	1376	1736
3 household members	21242	798	21242	798	1158	1857
4 and more	16753	352	16735	351	2804	4797
Population by age group						
<25	16809	350	16793	349	1771	2839
25 to 34	19615	535	19604	535	587	1028
35 to 44	19070	600	19049	601	770	1180
45 to 54	19187	435	19136	433	915	1483
55 to 64	19190	578	19190	578	773	1161
65+	17234	288	17234	288	1142	1415
Population by sex						
Male	18371	303	18344	303	2793	4348
Female	18066	291	18058	291	3165	4758

2.3. Non-sampling errors

2.3.1. Sampling frame and coverage errors

The samples for EU-SILC 2005 and EU-SILC 2006 were selected from the sampling frame based on the TERYT system, i.e. the *Domestic Territorial Division Register*. Two kinds of primary sampling units (PSUs) were distinguished in the sampling frame:

- about 178 000 *CEA – census enumeration areas* including about 68 dwellings each,
- about 33 000 *ESD – enumeration statistical districts* including about 377 dwellings each.

[7] Taking into account only households/persons receiving such income.
[8] Taking into account whole population (households/persons) surveyed.

The whole territory of Poland is divided into enumeration statistical districts and census enumeration areas. In EU-SILC census enumeration areas are used as primary sampling units. The secondary sampling units are dwellings. For each census enumeration area a list of dwellings was made up to form the secondary sampling frame. All the households from the selected dwellings are supposed to enter the survey.
The TERYT system is updated annually with respect to the territorial division into statistical districts and census enumeration areas. The lists of dwellings, names of towns, villages and streets are updated. Other changes due to new construction, dismantle of buildings and administrative division modifications are also introduced.
In the longitudinal (panel) component consisting of the subsample 4, some 6.6% of dwellings were found to be non-existing (cancelled, changed for non-residential units), uninhabited or temporarily inhabited.

2.3.2. Measurement and processing errors

Very much like any other statistical survey, EU-SILC may be burdened with non-sampling errors which occur at various stages of the survey and which cannot be eliminated completely. This mainly applies to interviewers' errors at the stage of collecting the information, errors due to the respondents' misunderstanding of questions and inaccurate or sometimes even false answers as well as the errors taking place at the stage of data recording.
According to the interviewers, who after filling in the household and individual interview were obliged to answer a few questions concerning interview performance for EU-SILC 2005, over 70% of the respondents showed a favourable attitude towards the survey, while about 5% were unwilling towards it. In the interviewers' opinion, in about 86% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
For EU-SILC 2006 it is possible to state that about three quarters of respondents (78% of those filling in the household questionnaire and 75% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 3% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 88% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
For EU-SILC 2007 and EU-SILC 2008 the figures were almost the same, about three quarters of respondents (80% of those filling in the household questionnaire and 78% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 3% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 89% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
The quality of income data in 2005, 2006, 2007 and 2008 was evaluated as slightly worse, mainly because of item non-response. It should also be pointed out that, in our opinion, the quality of data concerning net income categories is much higher than that of gross income. This is due to the fact that non-response was much more frequent for the information on taxes and social and health insurance contributions.

In Poland EU-SILC was carried out in May/June 2005, 2006, 2007 and 2008.
During the years 2005, 2006, 2007 and 2008 the data collection was performed by a face-to-face interview technique with the use of paper form questionnaires (the so called PAPI method). Two types of questionnaire: individual and household questionnaire were applicable.
The organisation and performance of the survey in the field was within the responsibility of regional statistical offices. Many interviewers were regular employees of the statistical offices and had experience in other social surveys. Survey performance in the field was preceded by a series of trainings organised in 2005, 2006, 2007 and 2008. Regional survey coordinators were instructed by CSO Social Statistics Division staff members and then the regional survey coordinators trained the interviewers at the regional statistical offices.
Interviewers' visits to households were preceded by the introductory letter of the CSO President.
The interviewers received written instructions concerning the survey performance.
Small gifts were given to the families participating in the survey. Each statistical office chose the type of gift for its respondents.

Data recording and check-up took place in regional statistical offices and was done with the use of Microsoft Visual FoxPro. After all the questionnaires for a given household had been recorded (the identifiers being voivodship number, dwelling number and household number), it was possible to make the household screening which consisted of logical and calculation check-up at the section, inter-section and inter-questionnaire levels. The regional files were then transferred to the CSO Computing Centre and combined to make up the general files at the national level. The national file completeness was also checked with the use of Microsoft Visual FoxPro. Additional check-up was made with SAS checking programmes.
On the basis of overall data files it was possible to create files for Eurostat. Some of the primary target variables could be found directly in the questionnaires, others had to be calculated with the algorithms especially prepared for this purpose.
The tables of EU-SILC results were compiled with the use of: SAS, SPSS, Microsoft Visual FoxPro.

2.3.3. Non-response errors

Achieved sample size

subsample 4:

Sample size	**wave**			
	1	2	3	4
A	4068	3614	3368	3210
B	9308	8525	7829	7445

A - number of households for which an interview is accepted for the database

B - number of persons of 16 years or older who are members of the households for which the interview is accepted for the database, and who completed a personal interview

Unit non-response

wave 1 (subsample 4):

- Household non-response rate NRh = [1 – (Ra*Rh)]*100,

Ra = 0.993
Rh = 0.703
NRh = 30.16

- Individual non-response rates NRp = (1 – Rp)*100,

Rp = 0.947
NRp = 5.20

- Overall individual non-response rates *NRp = [1 – (Ra*Rh*Rp)]*100,

*NRp = 33.89.

Response rate for household:

Comparison between wave 2 and wave 1 (subsample 4).

- Wave response rate = 0.889
(percentage of households successfully interviewed (DB135 = 1) which were passed on to wave t (from wave *t-1*) or newly created or added during wave *t*, excluding those out of scope (under the tracing rules) or non-existent)

- Longitudinal follow-up rate = 0.909
(percentage of households which are passed on to wave *t+1* for follow-up within the households received into wave *t* from wave *t-1*, excluding those out of scope (under the tracing rules) or non-existent)

- Follow-up rate = 0.917
(Number of households passed on from wave *t* to wave *t+1* in comparison to the number of households received for follow-up at wave *t* from wave *t-1*)

- Achieved samples size ratio = 0.897
(ratio of the number of households accepted for the database (DB135 = 1) in wave *t* to the number of households accepted for the database (DB135 = 1) in wave *t-1*)

Comparison between wave 3 and wave 2 (subsample 4).

- Wave response rate = 0.909
- Longitudinal follow-up rate = 0.925
- Follow-up rate = 0.936
- Achieved samples size ratio = 0.932

Comparison between wave 4 and wave 3 (subsample 4).

- Wave response rate = 0.934
- Longitudinal follow-up rate = 0.947
- Follow-up rate = 0.956
- Achieved samples size ratio = 0.958

Response rate for persons:

Personal interview response rates in wave 2 (subsample 4).

- Wave response rate = 0.957

(percentage of sample persons successfully interviewed (RB250 = 11, 12, 13) among those passed on to wave *t* (from wave *t-1*) or newly created or added during wave *t*, excluding those out scope (under the tracing rules).

- Wave response rate of co-residents = 0.000

(percentage of co-residents selected in wave 1, successfully interviewed (RB250 = 11, 12, 13) among those passed on to wave *t* (from wave *t-1*))

- Longitudinal follow-up rate = 0.858

(percentage of sample persons successfully interviewed (RB250 = 11, 12, 13) in wave *t* out of all sample persons selected, excluding those who have died or been found ineligible (out of scope), breakdown by causes of non-response).

- Achieved sample size ratio = 0.904

(ratio of the number of completed personal interviews (RB250 = 11, 12, 13) in wave *t* to the number of completed personal interviews in wave *t-1*. This ratio will be defined for sample persons and for all persons, including non-sample persons aged 16+ and for co-residents aged 16+ selected in the first wave).

- Response rate for non-sample persons = 0.908

(ratio of the number of completed personal interviews (RB250 = 11, 12, 13) of non-sample persons aged 16+ in wave *t* to all non-sample persons aged 16+ listed in the households accepted for the database (DB135 = 1) in wave *t* or listed in the most recently conducted household interviews for households, which were forwarded from wave *t-1* to wave *t* for follow-up, but could not be successfully interviewed in wave *t*).

Personal interview response rates in wave 3 (subsample 4).

- Wave response rate = 0.947
- Wave response rate of co-residents = 0.000
- Longitudinal follow-up rate = 0.874
- Achieved sample size ratio = 0.901
- Response rate for non-sample persons = 0.843

Personal interview response rates in wave 4 (subsample 4).

- Wave response rate = 0.944
- Wave response rate of co-residents = 0.000
- Longitudinal follow-up rate = 0.889
- Achieved samples size ratio = 0.929
- Response rate for non-sample persons = 0.830

Distribution of households by household status (DB110)*, by record of contact at address* (DB120)*, by household questionnaire result* (DB130) *and by household interview acceptance* (DB135)

Wave 1 (subsample 4*).*

Household questionnaire result

DB130	Total	%
Total	5787	100.0
11 – household questionnaire completed	4100	70.8
21 – refusal to co-operate	1177	20.3
22 – entire household temporarily away for duration of fieldwork	279	4.8
23 – household unable to respond (illness, incapacity,…)	174	3.0
24 – other reasons	57	1.0
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	4100	100.0
1 – interview accepted for database	4068	99.2
2 – interview rejected	32	0.8

Wave 2 (subsample 4*).*

Household status

DB110	Total	%
Total	4108	100.0
1 – at the same address as last interview	3882	94.5
2 – entire household moved to a private household within the country	68	1.7
3 – entire household moved to a collective household or institution within the country	1	0.0
4 – household moved outside the country	16	0.4
5 – entire household died	21	0.5
6 – household does not contain sample persons	4	0.1
7 – address non-contacted (unable to access, lost - no information on record on what happened to the household)	75	1.8
8 – split –off household	40	1.0
10 – fusion	1	0.0

Record of contact at address

DB120	Total	%
Total	108	100.0
11 – address contacted	67	62.0
21 – address cannot be located	0	0.4
22 – address unable to access	0	0.0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	41	38.0
Missing	0	0.0

Household questionnaire result

DB130	Total	%
Total	3949	100.0
11 – household questionnaire completed	3614	91.5
21 – refusal to co-operate	254	6.4
22 – entire household temporarily away for duration of fieldwork	52	1.3
23 – household unable to respond (illness, incapacity,…)	21	0.5
24 – other reasons	8	0.2
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3614	100.0
1 – interview accepted for database	3614	100.0
2 – interview rejected	0	0.0

Wave 3 (subsample 4*)*.

Household status

DB110	Total	%
Total	3740	100.0
1 – at the same address as last interview	3554	95.0
2 – entire household moved to a private household within the country	55	1.5
3 – entire household moved to a collective household or institution within the country	5	0.1
4 – household moved outside the country	8	0.2
5 – entire household died	25	0.7
6 – household does not contain sample persons	2	0.1
7 – address non-contacted (unable to access, lost - no information on record on what happened to the household)	46	1.2
8 – split –off household	45	1.2
10 – fusion	0	0.0

Record of contact at address

DB120	Total	%
Total	100	100.0
11 – address contacted	75	75.0
21 – address cannot be located	1	1.0
22 – address unable to access	0	0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	24	24.0
Missing	0	0

Household questionnaire result

DB130	Total	%
Total	3629	100.0
11 – household questionnaire completed	3368	92.8
21 – refusal to co-operate	150	4.1
22 – entire household temporarily away for duration of fieldwork	67	1.9
23 – household unable to respond (illness, incapacity,…)	33	0.9
24 – other reasons	11	0.3
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3368	100.0
1 – interview accepted for database	3368	100.0
2 – interview rejected	0	0.0

Wave 4 (subsample 4*)*.

Household status

DB110	Total	%
Total	3479	100.0
1 – at the same address as last interview	3334	95.8
2 – entire household moved to a private household within the country	52	1.5
3 – entire household moved to a collective household or institution within the country	4	0.1
4 – household moved outside the country	10	0.3
5 – entire household died	17	0.5
6 – household does not contain sample persons	0	0.0
7 – Household unable to access (due to for example climatic conditions)	0	0.0
8 – split –off household	32	0.9
10 – fusion	2	0.1
11 – lost household (no information on record on what happened to the household)	28	0.8

Record of contact at address

DB120	Total	%
Total	84	100.0
11 – address contacted	64	76.2
21 – address cannot be located	2	2.4
22 – address unable to access	0	0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	18	21.4
Missing	0	0

Household questionnaire result

DB130	Total	%
Total	3404	100.0
11 – household questionnaire completed	3210	94.5
21 – refusal to co-operate	108	3.2
22 – entire household temporarily away for duration of fieldwork	63	1.9
23 – household unable to respond (illness, incapacity,…)	13	0.4
24 – other reasons	4	0.1
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3210	100.0
1 – interview accepted for database	3210	100.0
2 – interview rejected	0	0.0

Distribution of persons for membership status (RB110)

Wave 2 (subsample 4*).*

Distribution of persons for membership status (RB110)

	Total	Current household members				No current household members		
		RB110=1	RB110=2	RB110=3	RB110=4	RB120 = 2 to 4	RB110=6	RB110=7
Total	11030	10636	57	151	103	83	0	0
%	100.0	96.4	0.5	1.4	0.9	0.8	0.0	0.0

Distribution of persons moving out by variable RB120.

	Total	**RB110 = 5**				
		RB120 = 1		**RB120 = 2**	**RB120 = 3**	**RB120 = 4**
		A	**B**			
Total	204	57	64	34	41	8
%	100	27.9	31.4	16.7	20.1	3.9

A – this person is a current household member in this wave

B - this person is not a current household member

Wave 3 (subsample 4*).*

Distribution of persons for membership status (RB110)

	Total	Current household members				No current household members		
		RB110=1	RB110=2	RB110=3	RB110=4	RB120 = 2 to 4	RB110=6	RB110=7
Total	10230	9813	74	127	90	126	0	0
%	100.0	95.2	0.8	1.2	0.9	1.2	0.7	0.0

Distribution of persons moving out by variable RB120.

	Total	**RB110 = 5**				
		RB120 = 1		**RB120 = 2**	**RB120 = 3**	**RB120 = 4**
		A	**B**			
Total	264	74	64	28	81	17
%	100	28.0	24.3	10.6	30.7	6.4

A – this person is a current household member in this wave

B - this person is not a current household member

Wave 4 (subsample 4*)*.

Distribution of persons for membership status (RB110)

	Total	**Current household members**				**No current household members**		
		RB110=1	RB110=2	RB110=3	RB110=4	RB120 = 2 to 4	RB110=6	RB110=7
Total	9661	9283	59	106	68	145	0	0
%	100.0	96.1	0.6	1.1	0.7	1.5	0.0	0.0

Distribution of persons moving out by variable RB120.

	Total	**RB110 = 5**				
		RB120 = 1		**RB120 = 2**	**RB120 = 3**	**RB120 = 4**
		A	**B**			
Total	258	59	54	26	99	20
%	100	22.9	20.9	10.0	38.4	7.8

A – this person is a current household member in this wave

B - this person is not a current household member

Item non-response (income variables)

2005

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	41.2	6.1	52.0
Total disposable household income	67.7	5.2	26.6
Total disposable household income before social transfers other than old-age and survivor's benefits	67.5	8.0	22.7
Total disposable household income before social transfers, including old-age and survivor's benefits	60.8	11.0	16.4
Net income components at household level			
HY040N	1.4	0.1	0.0
HY050N	23.2	0.3	0.4
HY060N	4.3	0.2	0.0
HY070N	6.1	0.2	0.0
HY080N	5.1	0.6	0.0
HY090N	1.0	0.6	0.0
HY110N	2.3	0.1	0.0
HY120N	42.1	4.6	0.0
HY130N	5.0	0.2	0.0
HY140N	40.9	32.7	22.8
HY145N	46.0	3.3	0.0
Gross income components at household level			
HY040G	1.4	0.1	0.0
HY050G	22.6	0.3	0.9
HY060G	4.3	0.2	0.0
HY070G	6.1	0.2	0.0
HY080G	5.1	0.6	0.0
HY090G	0.5	0.6	0.5
HY110G	1.9	0.1	0.4
HY120G	42.1	4.6	0.0
HY130G	5.0	0.2	0.0
HY140G	41.2	32.8	23.2

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	29.9	6.4	0.1
PY021N	0.1	0.5	0.0
PY035N	3.0	0.7	0.0
PY050N	4.4	3.1	0.3
PY080N	0.0	0.0	0.0
PY090N	3.7	0.3	0.0
PY100N	22.6	1.3	0.4
PY110N	1.5	0.1	0.0
PY120N	0.3	0.1	0.0
PY130N	6.6	0.2	0.0
PY140N	0.9	0.0	0.0
Gross income components at personal level			
PY010G	16.7	6.4	13.3
PY021G	0.1	0.5	0.0
PY035G	3.0	0.7	0.0
PY050G	4.8	1.5	3.1
PY080G	0.0	0.0	0.0
PY090G	2.2	0.3	1.5
PY100G	15.6	1.2	7.5
PY110G	0.9	0.1	0.6
PY120G	0.1	0.1	0.2
PY130G	4.5	0.2	2.1
PY140G	0.9	0.0	0.0

2006

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	41.7	5.4	52.8
Total disposable household income	72.9	4.5	22.5
Total disposable household income before social transfers other than old-age and survivor's benefits	72.6	6.2	19.7
Total disposable household income before social transfers, including old-age and survivor's benefits	66.3	8.2	15.5
Net income components at household level			
HY040N	1.0	0.2	0.2
HY050N	22.8	0.3	0.3
HY060N	5.5	0.1	0.0
HY070N	5.7	0.2	0.0
HY080N	6.1	0.5	0.0
HY090N	0.8	0.5	0.0
HY110N	3.7	0.1	0.0
HY120N	46.0	4.6	0.0
HY130N	5.2	0.3	0.0
HY140N	40.8	31.6	25.0
HY145N	45.8	2.5	0.0
Gross income components at household level			
HY040G	1.2	0.2	0.0
HY050G	21.8	0.3	1.2
HY060G	5.5	0.1	0.0
HY070G	5.7	0.2	0.0
HY080G	6.1	0.5	0.0
HY090G	0.4	0.5	0.4
HY110G	3.5	0.1	0.3
HY120G	46.0	4.6	0.0
HY130G	5.2	0.3	0.0
HY140G	40.7	31.4	25.5

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	31.1	7.2	0.0
PY021N	0.1	0.2	0.0
PY035N	2.8	0.7	0.0
PY050N	5.8	2.7	0.4
PY080N	0.0	0.0	0.0
PY090N	3.3	0.3	0.0
PY100N	22.2	1.7	0.3
PY110N	1.6	0.1	0.0
PY120N	0.4	0.0	0.0
PY130N	6.2	0.5	0.0
PY140N	1.3	0.1	0.0
Gross income components at personal level			
PY010G	16.1	7.2	15.1
PY021G	0.1	0.2	0.0
PY035G	2.8	0.7	0.0
PY050G	5.1	1.9	3.2
PY080G	0.0	0.0	0.0
PY090G	2.1	0.3	1.2
PY100G	15.8	1.7	6.7
PY110G	1.0	0.1	0.6
PY120G	0.2	0.0	0.1
PY130G	4.4	0.5	1.8
PY140G	1.3	0.1	0.0

2007

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	36.0	5.3	58.6
Total disposable household income	73.4	5.2	21.4
Total disposable household income before social transfers other than old-age and survivor's benefits	73.2	7.2	18.3
Total disposable household income before social transfers, including old-age and survivor's benefits	65.1	9.7	14.2
Net income components at household level			
HY040N	0.9	0.2	0.2
HY050N	21.7	0.3	0.4
HY060N	5.1	0.1	0.1
HY070N	5.0	0.1	0.0
HY080N	5.7	0.6	0.0
HY090N	1.2	0.9	0.0
HY100N	1.1	1.9	0.0
HY110N	4.0	0.1	0.0
HY120N	49.0	5.1	0.0
HY130N	4.3	0.4	0.0
HY140N	34.7	37.8	25.4
HY145N	36.7	2.5	0.0
Gross income components at household level			
HY040G	1.1	0.2	0.0
HY050G	20.6	0.3	1.5
HY060G	5.1	0.1	0.1
HY070G	5.0	0.1	0.0
HY080G	5.7	0.6	0.0
HY090G	0.5	0.9	0.7
HY100G	1.1	1.9	0.0
HY110G	3.6	0.1	0.4
HY120G	49.0	5.1	0.0
HY130G	4.3	0.4	0.0
HY140G	34.5	37.0	26.5

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	31.6	7.8	0.1
PY020N	7.5	3.0	1.0
PY021N	0.2	0.2	0.0
PY035N	2.4	0.7	0.0
PY050N	5.8	2.9	0.3
PY070N	6.0	1.4	0.0
PY080N	0.0	0.0	0.0
PY090N	2.5	0.4	0.0
PY100N	23.1	1.8	0.2
PY110N	1.3	0.2	0.0
PY120N	0.4	0.0	0.0
PY130N	6.0	0.5	0.0
PY140N	1.4	0.1	0.0
Gross income components at personal level			
PY010G	15.7	7.8	16.0
PY020G	7.5	3.0	1.0
PY021G	0.2	0.2	0.0
PY035G	0.0	20.0	2.7
PY050G	2.4	0.7	0.0
PY070G	5.7	2.2	2.8
PY080G	6.0	1.4	0.0
PY090G	0.0	0.0	0.0
PY100G	1.2	0.4	1.3
PY110G	13.8	1.8	9.5
PY120G	0.5	0.2	0.7
PY130G	0.2	0.0	0.2
PY140G	3.0	0.5	3.0

2008

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	36.2	5.9	57.6
Total disposable household income	71.0	5.8	23.1
Total disposable household income before social transfers other than old-age and survivor's benefits	71.0	8.0	19.7
Total disposable household income before social transfers, including old-age and survivor's benefits	62.7	11.4	14.9
Net income components at household level			
HY040N	0.9	0.2	0.3
HY050N	20.4	0.4	0.4
HY060N	4.4	0.2	0.0
HY070N	4.1	0.1	0.0
HY080N	5.5	0.6	0.0
HY090N	2.3	0.2	0.0
HY100N	0.9	0.9	0.0
HY110N	1.0	2.1	0.0
HY120N	3.8	0.1	0.0
HY130N	50.2	5.5	0.0
HY140N	4.5	0.3	0.0
HY145N	0.9	0.1	0.0
Gross income components at household level	35.5	38.8	24.1
HY040G	32.1	3.3	0.0
HY050G			
HY060G	1.1	0.2	0.0
HY070G	19.5	0.4	1.4
HY080G	4.4	0.2	0.0
HY090G	4.1	0.1	0.0
HY100G	5.5	0.6	0.0
HY110G	2.3	0.2	0.0
HY120G	0.4	0.9	0.6
HY130G	1.0	2.1	0.0
HY140G	3.5	0.1	0.4

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	31.7	9.5	0.1
PY020N	8.0	3.3	1.2
PY021N	0.2	0.2	0.0
PY035N	2.4	0.6	0.0
PY050N	6.2	2.8	0.3
PY070N	6.2	1.5	0.0
PY080N	0.0	0.0	0.0
PY090N	2.2	0.5	0.0
PY100N	24.0	2.0	0.2
PY110N	1.3	0.2	0.0
PY120N	0.3	0.1	0.0
PY130N	5.6	0.5	0.0
PY140N	1.0	0.1	0.0
Gross income components at personal level			
PY010G	15.2	9.5	16.6
PY020G	8.0	3.3	1.2
PY021G	0.2	0.2	0.0
PY035G	2.4	25.2	0.3
PY050G	2.4	24.9	0.0
PY070G	2.4	0.6	0.0
PY080G	5.4	1.8	3.1
PY090G	6.2	1.5	0.0
PY100G	0.0	0.0	0.0
PY110G	1.6	0.5	0.7
PY120G	14.9	2.0	9.3
PY130G	0.6	0.2	0.7
PY140G	0.2	0.1	0.1

2.4. Mode of data collection

EU-SILC is a non-obligatory, representative survey of individual households, performed by a face-to-face interview technique with the use of paper form questionnaires (the so called PAPI method). Two types of questionnaire: individual and household questionnaire were applicable.

Wave 1 (subsample 4*)*.

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	**RB250=11**	**RB250=14**
Total	9824	9308	516
%	100.0	94.7	5.3

Distribution of household members by RB260

Household members 16+ (RB245 = 1 to 3 and RB250 = 11 or 13)

	Total	**RB260 = 1**	**RB260 = 2**	**RB260 = 3**	**RB260 = 4**	**RB260 = 5**
Total	9308	7501	0	0	0	1807
%	100.0	80.6	0.0	0.0	0.0	19.4

Wave 2 (subsample 4*)*.

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	**RB250=11**	**RB250=14**
Total	8916	8525	391
%	100.0	95.6	4.4

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1)

	Total	**RB250=11**	**RB250=14**
Total	8796	8416	380
%	100.0	95.7	4.3

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2)

	Total	**RB250=11**	**RB250=14**
Total	120	109	11
%	100.0	90.8	9.2

Distribution of household members by RB260

Household members 16+ (RB245 = 1 to 3) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	8525	6940	0	0	0	1585
%	100.0	81.4	0	0	0	18.6

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	8416	6862	0	0	0	1554
%	100.0	81.5	0.0	0.0	0.0	18.5

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	109	78	0	0	0	31
%	100.0	71.6	0.0	0.0	0.0	28.4

Wave 3 (subsample 4*).*

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	8282	7829	453
%	100.0	94.5	5.5

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1)

	Total	RB250=11	RB250=14
Total	8083	7654	429
%	100.0	94.7	5.3

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2)

	Total	RB250=11	RB250=14
Total	199	175	24
%	100.0	87.9	12.1

Distribution of household members by RB260

Household members 16+ (RB245 = 1 to 3) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	7829	6467	0	0	0	1362
%	100.0	82.6	0	0	0	17.4

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	7654	6337	0	0	0	1317
%	100.0	82.8	0.0	0.0	0.0	17.2

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	175	130	0	0	0	45
%	100.0	74.3	0.0	0.0	0.0	25.7

Wave 4 (subsample 4*).*

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	7902	7445	457
%	100.0	94.2	5.8

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1)

	Total	RB250=11	RB250=14
Total	7466	7047	419
%	100.0	94.4	5.6

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2)

	Total	RB250=11	RB250=14
Total	436	398	38
%	100.0	93.3	8.7

Distribution of household members by RB260

Household members 16+ (RB245 = 1 to 3) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	7445	6032	0	0	0	1413
%	100.0	81.0	0	0	0	19.0

Sample persons 16+ (RB245 = 1 to 3 and RB100 = 1) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	7047	5765	0	0	0	1282
%	100.0	81.8	0.0	0.0	0.0	18.2

Co-residents 16+ (RB245 = 1 to 3 and RB100 = 2) and RB250 = 11 or 13

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	398	267	0	0	0	131
%	100.0	67.1	0.0	0.0	0.0	32.9

As for individual interviews, in 2005, 2006, 2007 and in 2008 a relatively high share (19.2%, 18.4%, 17.0% and 17.0%) of proxy interviews was noted. This was thoroughly discussed with the survey coordinators in the field.

The interviewers decided on proxy interviews only if the substitute respondents were well informed about the situation in the household and there was no other possibility to get the information. Proxy interviews were performed in the following situations:
- no contact with the respondent because of long-term absence (e.g. work in another town or abroad);
- respondent's disability, illness or pathology (such as alcoholism);
- according to other members of the household, the respondent was only available late at night and was not willing to participate in such a long interview, while at the same time the proxy could provide detailed information, sometimes even based on the documents, such as tax statements.

2.5. Imputation procedures

Imputation is aimed at obtaining complete records at the level of target variables. Target variables do not simply reflect questionnaire variables and their calculation algorithm is often complicated, although it principally consists in aggregation. So it is necessary to decide what aggregation level the imputation should take place at. There are three possible options:
- the level of questionnaire variables,
- the level of partly aggregated components,
- the level of ready-calculated target variables.

Since the only formal requirement is to obtain imputed target variables, all the above options are permissible and practicable, depending on the specific character of variables. However, the most frequent practice is the imputation at the level of questionnaire variables. There are certain arguments for this approach, on condition that the quantity of data and calculation algorithm details allow for it without much complication.

First of all, imputation at the lowest aggregation level can be desirable for the principal reasons related to the quality of imputation when:

- a target variable implies components of different character (i.e. taking different but rather predictable values, e.g. various social benefits, or dependent on a number of explanatory variables and thus easier to be modelled separately);
- target variables include many components and it is often the case that some of them have the missing items, while others have the correct ones which would be lost during the imputation of an aggregated variable.

Secondly, there are practical arguments for the imputation of disaggregated variables, as the same data serve as a basis for calculating national variables differing from the Eurostat's target variables. Thus the imputation of disaggregated components may be required so as to ensure the imputed data needed for other calculations.
The imputation at the target variable level is carried out only when the above circumstances do not occur or when overcoming the practical difficulties is easier than the imputation of disaggregated data.

There are several methods of component imputation. They can be classified as deterministic and stochastic methods. In case of deterministic methods the selected method and the set of explanatory variables (algorithm) clearly determine the imputation values for each record. In stochastic methods the imputation value is determined with the use of a random component. That is why it may happen that with the same algorithm and the same data file each algorithm realisation will give slightly different imputation results. Although the stochastic methods slightly increase estimator variance (introducing an additional random error component), they do not distort variance or original data distribution characteristics and allow for the correct estimation of random error. Deterministic imputation brings about variable variance reduction in the file and random error underestimation; it also distorts to a greater extent the correlation structure (increasing correlations with explanatory variables). According to item 2.7 of Regulation 1981/2003, it is recommended that for EU-SILC imputation the methods retaining distribution characteristics should be applied, which means the preference for the stochastic methods.

From among the stochastic methods the following were used in the task presented here:

- Hot-deck method

Random selection of a representative (donor) out of the correct records.
If auxiliary categorizing variables are used in the hot-deck method, a random representative is selected out of the records showing adequate values of auxiliary variables. If it is not possible to find a donor with the equivalent values for all the auxiliary variables, the so called sequence approach is applied. The categorising variables were ranked from the most to the least significant ones. If no donors are available, categorization is carried out with the subsequent explanatory variables being left out, starting from the least significant ones so as to obtain a subset containing donors.

- Stochastic regression imputation

Auxiliary variables are the explanatory variables of the regression model. The model takes the linear form or the logarithmic transformation is used. It is fitted on the basis of the correct records. The imputed value (or its logarithm in the case of transformed models) is a sum of the theoretical value derived from the model and a randomly selected model residual. The set of records of which the residual is selected is restricted to those which are nearest to the record imputed for the theoretical value derived from the model.

Out of the deterministic methods the following are applied:

- Regression deterministic imputation

The theoretical value from the model is adopted as the imputation value.

- Deduction imputation

The imputation value is directly determined on the basis of the relationships between variables.

In the case of imputation at the target variable level or imputation of the most significant components of target variables, stochastic imputation is applied in order to retain the variable property distribution as required by Regulation 1981/2003.

The application of stochastic regression imputation requires a model which describes well the formation of a variable with relatively small variance of an error term and good statistical qualities. With high variance of an error term, there is a danger of getting accidental values which are not typical of the correct part of the dataset. That is why in the cases where, in accordance with the assumption referred to above, stochastic imputation is required, the hot-deck method is applied in preference to regression imputation. This is particularly justified when the number of records for imputation is rather low, or when the number of correct records is too small for a suitable model fitting.

Stochastic regression imputation is most commonly used for income from hired employment, as:

- it is an important income category, declared by a significant rate of respondents which, if present, has a significant share in the total household income;
- this category can be successfully modelled with the use of the variables included in the questionnaire;
- there is a large (absolute) number of missing data, the percentage, however, being rather small; a large number of correct records make it possible to design a well-fitted model.

In case of incomes from hired employment stochastic regression imputation is applied to the majority of records with missing items, both those for which observations from the previous year are available (panel sample) and the new ones in the sample. In case of other income categories stochastic regression imputation is used as the basic imputation method when incomes of the same type for a given person/household are known from the previous year. If such income data from the previous year are not available, the hot-deck method is applied. The hot-deck method is also applied when the income data are known from the previous year but a suitable model fitting is difficult. In such a case the income from the previous year is used as a grouping variable. If the quantitative categorizing variable is applied in the hot-deck method, the categorization criterion is a breakdown into deciles.

Considering a relatively wide application of the stochastic regression imputation, supplementary protection against the effects of potential insufficient model adequacy was introduced. The residuals are not generated from the distribution of residuals for the whole sample, but they are selected from a restricted subset. Although in an ideal model residuals should be in the form of white noise, showing no trend whatsoever, in reality some trends can be observed in the distribution of residuals which are not detected by the model (like those related to non-linearity of relationships which cannot be removed by known transformations). In the case of 2005 data, where the use of stochastic regression is not so wide, imputed residuals are generated as pseudo-random numbers of normal distribution with variance corresponding to the estimated variance of an error term in the model.

In such a case, if we used residuals from the whole range, we could combine a particular theoretical value obtained from the model with the residual which occurs in the whole distribution but is quite improbable in combination with this particular theoretical value. So we could generate values significantly diverging from the real variable distribution. The use of residuals from the restricted range just reduces that risk.

Deterministic imputation is applied where missing data occur in less significant components of target variables (taxes, burdens to the main component, additions, etc.) in the situation when the main component is known. In such cases deterministic regression imputation is usually applied. Gross/net conversion is carried out with the use of the deterministic regression method. Deduction imputation is employed in rare cases of obvious relationships and can be treated as a supplementary stage of data editing.

The explanatory variables in the models and the grouping ones in the case of hot-deck method were selected to represent the relationships which, according to logics and knowledge about the phenomena studied, should occur in the data set, taking into account accessibility of the potential variables in the questionnaire. The relationships were tested on the file of correct data and in the majority of cases they proved to be significant. Some of the explanatory variables were retained, even if their impact on the imputed variable has not been statistically confirmed, if they expressed an economically important relationship or provided a grouping condition (interpretation criterion) in the calculation algorithm.

For the persons and households not surveyed in the previous year (a new sample, new household members, persons who could not be interviewed and all 2005 data) or for those who did not gain a particular type of income in the previous year, explanatory variables derived from the current data file are applied.
Wherever the same type of income is found in the data for the previous year, its value is treated as the main explanatory (categorizing) variable, both in the case of variables subjected to regression imputation and the hot-deck method. The current variables can be treated as additional explanatory variables.

The imputation of the missing individual questionnaires is carried out with the use of the hot-deck method. A wide set of variables providing household's characteristics (main source of maintenance) and variables from R set determining the person's position in the household and on the labour market is used as the categorization criterion. All the primary target variables related to the donor are transferred to the taker's record and then they are used for the calculation of household's total income.

2.6. Imputed rents

Definitions

Imputed housing costs should be understood as the estimated amount consisting of the actual payments effected by the owners (i.e. charges for water, electricity, gas, other fuels, maintenance and repair of the dwelling and other services relating to the dwelling) as well as imputed rentals that should be ascribed to the owners of flats or houses for their unpaid use of accommodation resources.

Imputed rentals should be understood as the estimated amount of profit gained in the form of a surplus of the rent over the dwelling maintenance costs, being the landlord's net profit, equal to the amount which could be gained by owners if they wanted to hire their dwellings in the same conditions at market prices.

Actual housing costs should be understood as rentals (charges for water, electricity, gas, other fuels, maintenance and repair of the dwelling and rent) paid by the tenants renting dwellings at market prices.

Actual rentals should be understood as the profit being a surplus of the rent over the dwelling maintenance costs, which is the landlord's net profit gained by the landlords hiring their dwellings at market prices.

Data

Data carried out in May-June 2008. The sample consisted of 13984 households. The analysis was made with the use of data from Section 2 of EU-SILC-1G Questionnaire "Dwelling Conditions", supplemented by the information about household size, number of households in a dwelling and locality.

Methodology

For the purposes of imputed rent estimation, regression analysis was used. It was decided to apply econometric methods, and particularly regression analysis. Taking into account the type of the two surveys, based on the representative method, the weighted least square method (WLSM) was employed. In EU-SILC the weights were also calculated on the basis of the number of persons in a household, locality (urban/rural areas) and number of households (based on the National Census 2002), but the number of persons was adjusted with demographic data. It is a commonly adopted approach (WLSM) which gives satisfactory results. The first step consists in the estimation of a hedonic price function according to which rents paid by tenants depend on the main characteristics of dwellings. In the second step all tenants, absolute owners, households with full and restricted rights to cooperative dwellings and do not pay any rent are ascribed forecasted rents and tenants paying rents below market prices are ascribed the difference between the forecasted and reduced rents, based on their dwelling characteristics. If rentals at a reduced rate are higher than the forecasted rentals, the rent actually paid is taken into account. In this way imputed rentals are estimated.

Subsamples of tenants

In the survey the function representing the relationship between the rentals and dwelling characteristics is determined based on the observations of households being tenants who pay rents at market prices.

The sample of 2008 covered 326 such households, of which 282 (2,02% of all the households participating in the survey) gave the amount of rentals, while 44 households did not, although they declared such payments.

Dependent variable
Following Eurostat's recommendations the variable explained is equal to the monthly rent for a dwelling. For the purposes of this model the value was calculated per 1 m sq. of the usable dwelling area occupied by the household or introduced as a total. If in the time of the survey a household did not pay any rent, it could declare the monthly rent paid last.

The dependent variable is introduced in the form of natural logarithm.

Explanatory variables
The set of explanatory variables in the rent function consisted of regressors describing flat/house location, building and environment standard as well as dwelling standard (arrangement and equipment).

In this survey only three explanatory variables (dwelling area occupied by the household, number of rooms, the year the household moved in) are quantitative, while all the others are dummies. Variable coding was much simpler, because the majority of questions were yes/no questions and the answers could be directly represented in the form of binary variables. Most of the variables for this survey were imposed by Eurostat as obligatory and the remaining ones were included in EU-SILC-G1 questionnaire by the team of experts preparing the survey on the basis of their experience and from HBS.

Form of the function
It was assumed that the estimated function of rentals is an exponential function which means that in the estimation form the dependent variable is a logarithm of variable. This is a convenient solution, ensuring that the theoretical values (calculated and forecasted) will be positive, which could not be guaranteed by the linear function. Trial estimates of rentals based on the data showed that for some of the households the forecasted values, estimated with the linear function, were negative.

Estimation technique
All the calculations were made with the use of SPSS package and the ENTER or STEPWISE estimation method. The final results are given below in detail.
To start with, the functions including all the variables were applied so as to check: what is the maximum R^2 that can be obtained, which of the variables describe the rentals level best, which seem irrelevant and which provide estimates difficult to interpret. Then the explanatory variable file was reduced by rejection of the variables which were not fitted. Estimation was applied by introducing into the function regressors according to their statistical significance (STEPWISE) and thus the correct versions, both as regards statistics and substance, were obtained.
Using the acceptable versions (upon introduction of the data from the auxiliary model and based on the expert model– see the paragraph "Results"), imputed rentals were determined for all the households except those paying the actual rentals.

Main characteristics

Sample size	13984
Number of observations on tenants at market prices	327
Number of explanatory variables	17
R2	0.730
R2 corrected	0.712
R	0.854
Imputed rentals (in PLN per household, per year)	
Averages for the total number of households in Poland	9303
of which:	
dwelling owners	8439
other households	864
Averages for owners' subsample	10059
Averages for subsample of „others"	6133
Actual rentals (in PLN per household, per year)	
Averages for the total number of households in Poland	132
Averages for subsample of tenants paying rentals at market prices	5022

2.7. Company cars

The information on the private use of the company car is collected in the individual questionnaire. The data covers the estimated amount the respondent gained by using the company car for private purposes. In case of the missing value (the respondent was using the company car but did not estimate the amount gained) imputation is applied with the use of the hot-deck and regression imputation with simulated residuals methods.

3. COMPARABILITY

3.1. Basic concepts and definitions

The reference population

There were no essential differences between the national concepts and standard EU-SILC concepts.
The survey unit was a household and all the household members who had completed 16 years of age by:

- December 31, 2004 for EU-SILC 2005;
- December 31, 2005 for EU-SILC 2006;
- December 31, 2006 for EU-SILC 2007;
- December 31, 2007 for EU-SILC 2008.

The survey did not cover collective accommodation households (such as boarding house, workers' hostel, pensioners' house or monastery), except for the households of the staff members of these institutions living in these buildings in order to do their job (e.g. hotel manager, tender etc.).
The households of foreign citizens should participate in the survey.

The private household definition

No difference to the common definition in either wave (EU-SILC 2005, EU-SILC 2006, EU-SILC 2007 and EU-SILC 2008).

Household is a group of persons related to each other by kinship or not, living together and sharing their income and expenditure (multi-person household) or a single person, not sharing his/her income or expenditure with any other person, whether living alone or with other persons (one-person household).
Family members living together but not sharing their income and expenditure with other family members make up separate households.
The household size is determined by the number of persons comprised by the household.

The household membership

No difference to the common definition in either wave (EU-SILC 2005, EU-SILC 2006, EU-SILC 2007 and EU-SILC 2008).

The household composition accounted for:
- persons living together and sharing their income and expenditure who have been in the household for at least 6 months (either the real or the intended time of staying in the household should be considered),
- persons absent from the household because of their occupation, if their earnings are allocated to the household's expenditure,
- persons at the age of up to 15 years (inclusive), absent from the household for education purposes, living in boarding houses or private dwellings,
- persons absent from the household at the time of the survey, staying at education centres, welfare houses or hospitals, if their real or intended stay outside the household is less than 6 months.

The household composition did not account for:
- persons at the age of over 15 years, absent from the household for education purposes, living in boarding houses, students' hostels or private dwellings,
- men in military service (those performing substitute military service by working in companies and living at home are included in the household),
- persons in prison,
- persons absent from the household at the time of the survey, staying at education centres, welfare houses or hospitals, if their real or intended stay outside the household is more than 6 months,
- persons (household's guests) staying in the household at the time of the survey who have been or intended to be there for less than 6 months,
- persons renting a room, including students (unless they are treated as household members),
- persons renting a room or bed for the time of work in a given place (including such works as land melioration, geodetic measurements, forest cut-down or building constructions),

- persons living in the household and employed as au pairs, helping personnel on the farm, craft apprentices or trainees.

The income reference period(s) used

No difference to the common definition in either wave. The income reference year for:
EU-SILC 2005 was 2004,
EU-SILC 2006 was 2005,
EU-SILC 2007 was 2006,
EU-SILC 2008 was 2007 year.

Reference period for taxes on income and social insurance contributions

No difference to the common definition in either wave (EU-SILC 2005, EU-SILC 2006, EU-SILC 2007 and EU-SILC 2008).
The reference periods for income tax prepayment and compulsory social insurance contributions were again the years 2004, 2005, 2006, 2007. The account clearance with the Treasury Office (including payments and returns) effected in 2004 refers to the income for 2003, in 2005 for 2004 in 2006 for 2005 and in 2007 for 2006.

The reference period for taxes on wealth

No difference to the common definition in either wave (EU-SILC 2005, EU-SILC 2006, EU-SILC 2007 and EU-SILC 2008).
Taxes on wealth paid during the income reference period were recorded correspondingly - 2004, 2005, 2006 or 2007.

The lag between the income reference period and current variables

The lag between the income reference period and current variables is about 5 months for each wave.

The total duration of the data collection

EU-SILC was performed on the territory of the whole country in:
2005 - between May 2 and June 17,
2006 - between May 2 and June 19,
2007 - between May 2 and June 19,
2008 - between May 2 and June 26.

Basic information on activity status during the income reference period

<u>Differences concerning EU-SILC 2005:</u>
The following variables were not recorded:
- Change of job since last year (PL160);
- Reason for change (PL170);
- Most recent change in the individual's activity status (PL180).

Starting from EU-SILC 2006, they have been taken into account.

Differences concerning EU-SILC 2006:
In EU-SILC 2006 the definition of retired person was changed (now this definition is in accordance with the international recommendations). In EU-SILC 2005 people obtaining disability pensions were included in PL085 and in variable PL210 in category 6, while in EU-SILC 2006 they were included in variable PL090 and in variable PL210 in category 8.

3.2. Components of income

3.2.1. Differences between the national definitions and standard EU-SILC definitions

There were no essential differences between the national concepts and standard EU-SILC concepts.

The income components where no difference between national and standard definitions can be found are not mentioned. The differences between the national and the EUROSTAT definitions refer to four waves likewise unless it is indicated otherwise.

HY010 – Total household gross income
HY020 – Total disposable household income
HY022 – Total disposable household income before social transfers other than old-age and survivor's benefits
HY023 - Total disposable household income before social transfers including old-age and survivor's benefits

In accordance with EU-SILC 065 (2008 operation) the new income components, mandatory from 2007 operation onwards, namely:

- PY020G – NON-CASH EMPLOYEE INCOME;
- PY030G – EMPLOYER'S SOCIAL INSURANCE CONTRIBUTION;
- PY070G – VALUE OF GOODS PRODUCED FOR OWN CONSUMPTION;
- PY080G – PENSION FROM INDIVIDUAL PRIVATE PLANS;
- HY030G – IMPUTED RENT;
- HY100G – INTEREST REPAYMENTS ON MORTGAGE

have been recorded at component level only and they are not included in the household`s total income (variables: HY010G; HY020G; HY22G; HY023G).

Since 2007, revenue from the use of a company car for private purposes (PY021) has been added as a component of the aggregate income.

PY010 - Cash or near-cash employee's income
This variable does not account for:
- assistance for foster families; since granting the benefit is not connected with quitting the job, this benefit has been qualified to the category of „Family related allowances' (HY050),
- benefit granted to the families when the only person providing income for the family is called up to the active military service; since this benefit is only granted when the only family supporter has been called to the military service, it has been included in the category of „Family related allowances' (HY050).

PY020 - Non-cash employee income
In EU-SILC 2005 and EU-SILC 2006 this information only refers to the income gained from the use of the company car for private purposes. Since 2007, the information about other non-cash employee income has been collected, but at the component level only and it is not included in the income.

PY030 – Employer's social insurance contributions
The variables have been collected since EU-SILC 2006, but only at the component level.

PY031 – Optional employer's social insurance contributions
The variables have been collected since EU-SILC 2008, but only at the component level.

PY050 - Cash profits or losses from self-employment (including royalties)
The data on income from self-employment were collected in two different ways: the respondents were asked about the company's costs and profits and also about the amount of money gained from self-employment which was allocated to the household's expenditure. After a detailed analysis of data it was decided that the income from self-employment would be equal to the amount allocated to the household's needs.

HY080 and HY130 - Regular inter-household cash transfer received
In EU-SILC2008 from both variables on regular cash transfers (HY080 and HY130) two additional variables were distinguished: Alimonies received - compulsory + voluntary (HY081), and Alimonies paid – compulsory + voluntary (HY131).
HY081 variable is contained in the variable HY080 and similarly, HY131 is contained in HY130.

PY110 – Survivor's benefits
Death grants are not included in the income because the whole sum is used to cover the cost of the funeral.

PY120 - Sickness benefits
Sickness and childcare benefits are not included (a childcare benefit is granted to the working parent of a sick child), because they are paid by the employer and cannot be detached from the income from hired employment. Therefore, they are accounted for in the income from hired employment.

PY200 – Gross monthly earnings for employees
The variables have been collected since EU-SILC 2006.

All the other variables not listed above

Dwelling conditions and material deprivation items
The analysis of questions and explanatory notes from the guideline for interviewers concerning dwelling conditions and material deprivation items showed that some records differed from those included in document 065/04:

Arrears on mortgage payment – it was not clarified that only arrears on mortgage should be taken into account, so other dwelling related credits might have been included.
Arrears on hire purchase instalments other than loan payments – this question included arrears on hire purchase and credits other than dwelling-related ones.

Capacity to afford paying for one week annual holiday away from home – first of all, the question included the expression "if the household wants"; secondly, a family as such was concerned and it was not pointed out that the question referred to the household as a whole.

Leaking roof, damp walls/floors/foundation, or rot in window frames or floor – the question was formulated in a different way, namely: "Do you think your dwelling requires renovation because of…?"

Indoor flushing toilet for sole use of the household – the toilet could have been shared with other households.

Additionally, for the variables from HS010 to HS050 no information was given that paying through borrowing meant that household was not in arrears.

There were no other major divergences from common definitions.

In 2008 changes were introduced to bridge these differences according with document 065/04.

3.2.2. The source or procedure used for the collection of income variables

The income data were collected during the interviews with respondents. The target income variables were split into components corresponding to particular benefits applicable in the Polish conditions.

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

The respondents were asked to give the net incomes and contributions (income tax prepayments and compulsory social insurance). Only in the case of income from rental of a property (HY040) the respondents were asked to give the gross income and the amount of tax paid.

3.2.4. The method used for obtaining income target variables in the required form

The gross income was obtained by summing up net value, income tax prepayments and compulsory social insurance contributions. If the information on tax and insurance contributions was missing, the amounts were imputed on the basis of the results obtained. Only in the case of income from rental of property, the tax paid was subtracted from the gross income.

3.3. Tracing rules

Standard EU-SILC tracing rules are applied.

4. COHERENCE

The calculation in point 4.1 and 4.2 was made taking into account the cross-sectional data of EU-SILC 2007.

4.1. Comparison of EU-SILC and HBS results

The objective of this section is to compare HBS (Household Budget Survey) and EU-SILC results.
Up to 2004 the HBS provided the main source of data on the living conditions of the Polish population, among others on incomes, dwelling conditions and households' equipment.
The HBS has been regularly conducted every year since 1993 up to now with the use of the rotational method. The households are surveyed in the two year panel.
In HBS the main source of data on income and expenditure is provided by the diaries, while that concerning dwelling-related expenditure and utilities – by BR-01a questionnaire. In addition, three other questionnaires are filled in.

When comparing these two sources we must take into account some discrepancies. The differences are to a great extent brought about by the methodological diversity. Here are the main diverging points:

- Different reference periods for income variables – in HBS the reference period is 1 month and, following Eurostat's recommendation, the annual income is the monthly income multiplied by 12, which in the case of irregular income, like that from farming, can bring about significant distortions. In EU-SILC the reference period is a calendar year preceding the survey;
- EU-SILC does not take into account the so called lump sums and irregular donations, which is the case in HBS. Moreover, in EU-SILC disposable income covers non-monetary profit related to the use of the company car;
- Different ways of data collection – in HBS the respondents make records in the so called diary. They have to determine the data sources themselves and do not have them listed in the diary. This may cause omissions. In EU-SILC each respondent is asked detailed questions. In EU-SILC all the income missing data are imputed, while there is no imputation in HBS;
- Different ways of sample selection – in HBS dwellings in which all the households refused to participate in the survey are replaced with new ones from the so called reserve list;
- Slightly different weighting of results.

In some tables given below the data are presented in the breakdown by socio-economic group and household size. The household survey results are usually released by CSO in the breakdown by socioeconomic group and household size.
The main criterion for socioeconomic group classification is the prevailing source of income.
In the tables below only weighted data are presented.

Tab. 1. Structure of population by age

Specification	EU-SILC 2008	HBS 2008
	In %	
Total	100.0	100.0
0-14	15.7	18.1
15-24	15.1	15.8
25-54	44.0	41.6
55-64	11.8	12.1
65+	13.5	12.4

Tab. 2. Structure of population by level of education

Specification	EU-SILC 2008	HBS 2008
	In %	
Total	100.0	1000
No school education	1.9	0.7
Completed primary	16.4	17.0
Lower secondary	5.1	6.8
Elementary vocational	26.8	27.0
Secondary	34.4	34.5
Higher	15.4	14.0

Tab. 3. Structure of households and persons in households by socioeconomic group

Households	Households		Persons in households	
	EU-SILC 2008	HBS 2007	EU-SILC 2008	HBS 2007
Total	13253090	13332283	37637658	37707821
Total = 100				
Employees	52.3	47.5	62.7	56.3
Farmers	2.6	4.3	3.5	6.6
Self-employed	4.6	6.4	5.2	7.6
Retirees	27.4	27.6	18.6	18.9
Pensioners	8.4	9.2	5.2	6.4
Maintained from non-earned sources	4.8	5.0	4.7	4.2

Tab. 4. Average yearly equivalent income in PLN by socio-economic group

Households	Disposable income		Income from hired work	
	EU-SILC 2008	HBS 2007	EU-SILC 2008	HBS 2007
Total	18684	16549	11719	8857
Employees	20899	17120	17609	14287
Farmers	13550	16879	1356	1669
Self-employed	21348	23317	3201	3412
Retirees	15618	15089	1545	1694
Pensioners	12061	11472	1178	1234
Maintained from non-earned sources	9665	10327	2401	936

Tab. 5. Average yearly equivalent income in PLN by number of persons

Households	Disposable income		Income from hired work	
	EU-SILC 2008	HBS 2007	EU-SILC 2008	HBS 2007
Total	18684	16549	11719	8857
1-person	16883	15540	5675	4627
2-persons	21201	18849	9830	7426
3-persons	21001	18690	14861	11635
4-persons	19273	16936	14363	11036
5-persons	16238	14498	11012	8179
6-persons and more	14947	12605	9019	5679

Tab. 6. Households provided with selected durables

Specification	EU-SILC 2008	HBS 2008
	In %	
Fixed telephone	68.5	64.2
Mobile telephone	79.9	83.5
Television set	97.5	98.5
Computer	54.5	56.4
Printer	40.0	37.1
Internet connection	43.0	45.6
Microwave oven	41.9	46.1
Dishwasher	11.5	9.6
Refrigerator	97.7	98.4
Washing machine	96.8	97.3
Passenger car	56.2	54.7

4.2. Comparison of income data from SNA for the household sector and EU-SILC

The comparison covered disposable income and its main components: income from hired employment, self-employment (in and outside farming), as well as social benefits.
When comparing these two sources we must take into account some discrepancies. The differences are to a great extent brought about by the methodological diversity. Here are the main diverging points:

1. In SNA the household sector includes collective households which do not enter EU-SILC.
2. Both systems employ different methods of measuring income from self-employment.
3. Accounts of primary and secondary income distribution in SNA used for the determination of disposable income include some items not covered by EU-SILC 2008 or not taken into account in the calculation of its results. The most important of them are imputed rents.

In SNA income from self-employment is determined as the so called operation surplus which is the balance between the global production and current production inputs (i.e. intermediate consumption) and hired employees' wages. This difference is reduced by taxes and increased by subsidies. The operation surplus thus calculated is allocated to the household's consumer needs, housing-related investment as well as production-related investment. In the Polish EU-SILC the question about income from self-employment concerns just the amount allocated to the household's consumer needs and its housing-related investment. In addition, SNA takes into account consumption from own production which is not taken into consideration by EU-SILC for farmers' households.
Due to these differences incomes from self-employment according to EU-SILC made 24% - 26% of the operation surplus only (after deduction of section K).

Incomes from hired employment in EU-SILC 2008 are equal to 101% of the corresponding figure in SNA, while social benefits – 92% respectively, which seems to be a good result.

In EU-SILC 2008, as compared with EU-SILC 2007, the data coherence for disposable income with SNA increased by 4 percentage points. This is due to the coherence increase for income from hired employment by 2 percentage points, i.e. from 99% to 101%. In terms of value, wages and other incomes related to hired employment provide the most important component of disposable income in SNA. This category made up over 50% of the disposable income for 2007.

Taking into consideration the methodological changes in both surveys, the real increase of the coherence of income from hired employment was as high as 7 percentage points, rising from 95% to 102%. In EU-SILC 2008 RB050 weight was used for the calculation of individual incomes, while in EU-SILC 2007 – PB040, respectively. In SNA incomes of the employees working abroad were calculated in a different way. However, these methodological changes do not explain the increased coherence of incomes from hired employment. The change of weight in EU-SILC could justify an increase by 1 percentage point only. The methodological

changes of SNA bring about reduced coherence between SNA data and EU-SILC data, since they lead to an increase in wages and other incomes from hired employment in SNA (for 2006 by over PLN 15 million). Considering the fact that SNA data are based on the results of the enterprise surveys, it can be judged that the increased coherence of incomes from hired employment might be due to some deterioration of the quality of enterprise survey results in the scope of wages. Unlike for EU-SILC 2006 and 2007, it is less probable that the increased coherence of SNA results in the area of hired employment could be brought about by a higher quality of EU-SILC results, as the coherence for all the other significant economic categories remained more or less at the same level.

The data coherence for social benefits between SNA and EU-SILC was 92%, and taking into account the change of the weighting method – 93%, which is equivalent to EU-SILC 2007. The coherence of income from self-employment between EU-SILC 2008 and SNA was 26% and taking into account the change of the weighting method – 27%, which is by 1 percentage point more than in EU-SILC 2007.

Comparison between SNA results for the household sector and EU-SILC

Category in SNA	Category description in EU-SILC	Incomes: 2004 r.			Incomes: 2005 r.			Incomes: 2006 r.			Incomes: 2007 r.		
		SNA in mln PLN	EU-SILC in mln PLN	**SNA = 100%**	SNA in mln PLN	EU-SILC in mln PLN	**SNA = 100%**	SNA in mln PLN	EU-SILC in mln PLN	**SNA = 100%**	SNA in mln PLN	EU-SILC in mln PLN	**SNA = 100%**
Gross disposable income (net)	Total disposable household income (net)	627 766	340 299	**54**	651 512	369 046	**57**	683 483	398 939	**58**	742 374	457 807	**62**
Wages, salaries and other income connected with hired work (gross)	Employee cash or near cash income (gross)	288 540	258 818	**90**	303 358	290 140	**96**	325 030	323 956	**99**	375 358	380 422	**99**
Gross operating surplus (gross) with the exception of section K	Self-employment income (gross) - value allocated to household`s consumption and dwelling-related investment	182 152	44 141	**24**	189 378	50 167	**27**	201 601	53 219	**26**	220 168	58 291	**26**
Social security benefits and social assistance benefits (gross)	Social benefits (gross)	148 289	132 093	**89**	153 946	141 334	**92**	161 336	149 258	**93**	166 880	153 565	**93**

Notes:

1. Remarks in brackets: “net” or “gross” refer to including or not including income tax and social security contributions, while the word “gross” in SNA names of categories refers to including fixed assets deppreciation.

2. Data for gross operating surplus in SNA has been taken into consideration with the exception of section K, which allows for better comparability with EU-SILC data on self-employment income (PY050G). The data for section K mainly cover imputed rents, not included in the results of EU-SILC, and market income from renting of real estate included in EU-SILC as variable HY040G.