

CENTRAL STATISTICAL OFFICE OF POLAND

FINAL QUALITY REPORT

**ACTION ENTITLED:
*EU-SILC 2010***

Warsaw, January 2013

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. <i>Sampling units</i>	4
2.1.3. <i>Stratification and substratification criteria</i>	5
2.1.4. <i>Sample size and allocation criteria</i>	5
2.1.5. <i>Sample selection schemes</i>	5
2.1.6. <i>Sample distribution over time</i>	5
2.1.7. Renewal of sample: rotational groups	6
2.1.8. <i>Weightings</i>	6
2.1.9. <i>Substitutions</i>	11
2.2. Sampling errors	12
2.3. Non-sampling errors.....	16
2.3.1. <i>Sampling frame and coverage errors</i>	16
2.3.2. Measurement and processing errors	17
2.3.3. <i>Non-response errors</i>	18
2.3.3.5. <i>Item non-response</i>	27
2.4. Mode of data collection.....	34
2.5. Imputation procedure	39
2.6. Imputed rents.....	42
2.7. Company cars	44
3. COMPARABILITY	44
3.1. Basic concepts and definitions	44
3.2. Components of income.....	46
3.2.1. <i>Differences between the national definitions and standard EU-SILC definitions</i>	46
3.2.2. <i>The source or procedure used for the collection of income variables</i>	49
3.2.3. <i>The form in which income variables at component level have been obtained</i>	49
3.2.4. <i>The method used for obtaining income target variables in the required form</i>	49
3.3. Tracing rules.....	49
4. COHERENCE	50
4.1. Comparison of EU-SILC and HBS results.....	50
4.2. Comparison of income data from SNA for the household sector and EU-SILC	55

PREFACE

This quality report is the final quality report on EU-SILC 2010 carried out in Poland 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 in the context of EU-SILC, was calculated using the longitudinal rotation 2007-2010.

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	10.47
2		M	10.24
3		F	10.68
4	(0 ≤ AGE ≤ 17)	T	14.79
5	(18 ≤ AGE ≤ 64)	T	10.17
6		M	10.10
7		F	10.24
8	(AGE ≥ 65)	T	6.79
9		M	4.62
10		F	8.10

Persistent-at-risk-of-poverty rate by age and gender (50% median)			
No.	Age	Gender	(%)
1	Total (AGE ≥ 0)	T	6.10
2		M	6.41
3		F	5.83
4	(0 ≤ AGE ≤ 17)	T	9.00
5	(18 ≤ AGE ≤ 64)	T	6.08
6		M	6.26
7		F	5.91
8	(AGE ≥ 65)	T	2.78
9		M	2.84
10		F	2.75

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. 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.

For the cross-sectional component of the survey conducted in 2009 the subsample 8 was selected. It replaced the subsample 4, and consisted of 1479 PSUs and 6017 dwellings.

For the cross-sectional component of the survey conducted in 2010 the subsample 5 was replaced by the subsample 9 (coded as 1) which consisted of 1477 census areas and 6003 dwellings.

The subsample 6 formed the four year longitudinal (panel) component.

In official cross-sectional and longitudinal data for EU-SILC 2010 operation the following coding was used: variable DB075 (rotation group) equals 1 for subsample 1, 2 for subsample 6, 3 for subsample 7, and 4 for subsample 8.

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, for the 2008 survey the subsample 7 was selected to replace the subsample 3, for the 2009 survey the subsample 8 was selected to replace the subsample 4, and for the year 2010 the new subsample 1 replaced the subsample 5. 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	0.642
1	Warsaw	0.386
2	Towns 500 000 – 1 000 000 inhabitants	0.519
3	Towns 100 000 – 500 000 inhabitants	0.589
4	Towns 20 000 – 100 000 inhabitants	0.655
5	Towns less than 20 000 inhabitants	0.677
6	Rural areas	0.742

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 new subsample 1. The next step consisted of calculating the weights DB090 and RB050 for the households of the subsample 1 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

Second wave

For 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 subsample 7 (new in this wave) the base weight is equal to RB050 (cross-sectional weight) multiplied by 4.

Non-response adjustments – subsequent waves

Third wave

For the subsample 6 surveyed for the third time and the subsample 7 surveyed for the second time the base weights were determined by the correction of the base weights from the previous year.

For the subsample 7 the following method was applied:

The base weight of 2008 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 2008 and who should be surveyed in 2009. 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 2008 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 subsample 6 (surveyed for the third time) the algorithm based on the method described for the subsample 7 was used. Additionally, re-entries, i.e. persons who were

surveyed in 2007, not surveyed in 2008, and surveyed in 2009 again, were taken into account. The base weights for such persons were computed by the correction of base weights from 2007 on the data for 2007 and 2009 (with no data for 2008). The inclusion of re-entries in the subsamples surveyed in 2009 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 2007.

The last stage of the base weight calculation for the third year of the survey consisted of receiving weights of mothers by children of “sample persons” and zero weights by “co-residents” i.e. additional persons included in the households.

For the subsample 8 (new in this wave) the base weight is equal to RB050 (cross-sectional weight) multiplied by 4.

Fourth wave

For the subsample 6 surveyed for the fourth time, the subsample 7 surveyed for the third time, and the subsample 8 surveyed for the second time, the base weights were determined by the correction of the base weights from the previous year.

For the subsample 8 the following method was used:

The base weight of 2009 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 2009 and who should be surveyed in 2010. 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 2009 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 6 and 7 (surveyed for the fourth and third time respectively) the algorithm based on the method described for the subsample 8 was used. Additionally, re-entries, i.e. persons who were surveyed in 2008, not surveyed in 2009, and surveyed in 2010 again, were taken into account. The base weights for such persons were computed by the correction of the base weights from 2008 on the data for 2008 and 2010 (with no data for 2009). The inclusion of re-entries in the subsamples surveyed in 2010 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 2008.

Adjustments to external data

Adjustment to external data was not applied.

Final longitudinal weight

The panel weight RB062 was calculated:

- 1) by taking the base weights for subsamples 6, 7 and 8,
- 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 6 and 7,
- 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 6,
- 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 of 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 6, 7, 8 and 1 in EU-SILC 2010:

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 equals the weight RB050,

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 previous EU-SILC operations were documented in the previous EU-SILC Intermediate Quality Reports.

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 500 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)	52978	458	52918	457	3994	12920
Total disposable household income (HY020)	41020	337	40996	337	8519	12925
Total disposable household income before social transfers other than old-age and survivor's benefits (HY022)	39158	340	38806	340	8539	12821
Total disposable household income including old-age and survivor's benefits (HY023)	32340	370	29348	357	7647	11713
Net income components at household level						
HY040N	8187	713	130	16	125	204
HY050N	3704	94	557	20	2047	2166
HY060N	2151	114	80	6	494	532
HY070N	1574	61	45	4	331	356
HY080N	6027	250	377	23	668	772
HY090N	7028	1254	208	38	200	355
HY110N	2274	206	57	6	377	401
HY120N	225	4	136	3	0	8009
HY130N	5228	296	248	19	605	642
HY140N	12169	125	11969	126	12738	12738
HY145N	-914	30	-432	15	0	6017

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	9239	826	146	18	169	204
HY050G	3854	102	580	21	1896	2166
HY060G	2151	114	80	6	494	532
HY070G	1574	61	45	4	331	356
HY080G	6027	250	377	23	668	772
HY090G	8442	1451	250	44	76	355
HY110G	2423	237	60	7	341	401
HY120G	225	4	136	3	6730	8009
HY130G	5228	296	248	19	605	642
HY140G	11721	126	11537	127	12752	12752
Net income components at personal level						
PY010N	22953	212	10291	126	9317	12568
PY020N	1072	40	161	7	2529	4164
PY035N	2003	92	74	5	996	996
PY050N	17212	528	1730	63	1919	3122
PY080N	0	0	0	0	0	0
PY090N	4699	195	123	7	677	815
PY100N	15592	99	3886	59	7588	8589
PY110N	9422	419	106	8	322	378
PY120N	3343	508	20	4	154	188
PY130N	8230	109	467	15	1556	1832
PY140N	2810	471	26	5	254	293
Gross income components at personal level						
PY010G	30791	290	13806	170	4309	12568
PY020G	1072	40	161	7	2529	4164
PY035G	2003	92	74	5	996	996
PY050G	23297	688	2518	89	1539	3387
PY080G	0	0	0	0	0	0
PY090G	5426	235	142	9	235	815
PY100G	18053	117	4500	69	4018	8589
PY110G	10847	485	122	9	132	378
PY120G	3691	539	22	4	118	188
PY130G	9354	129	531	17	806	1832
PY140G	2810	471	26	5	254	293
PY200G	2613	23	1042	13	23967	27091

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	20484	481	20436	480	2025	2453
2 household members	24812	370	24812	370	5496	7540
3 household members	25698	385	25698	385	4554	7686
4 and more	20216	224	20216	224	10435	19695
Population by age group						
<25	20495	217	20495	217	6479	11403
25 to 34	24305	361	24301	360	2698	4760
35 to 44	23566	389	23550	388	2619	4605
45 to 54	23088	309	23083	309	3183	5619
55 to 64	22869	305	22861	305	3467	5341
65+	20063	189	20063	189	4064	5646
Population by sex						
Male	22400	198	22392	198	10525	17852
Female	21911	169	21909	169	11985	19522

Longitudinal component

The mean, the total number of observations (before and after imputation) and the standard errors for the following income components , subsample 6.

Income components	Mean ⁵	Standard error	Mean ⁶	Standard error	Number of observations	
					Before imputation	After imputation
Total household gross income (HY010)	53242	909	53213	910	983	3036
Total disposable household income (HY020)	41282	662	41260	663	2037	3036
Total disposable household income before social transfers other than old-age and survivor's benefits (HY022)	39449	659	39044	668	2032	3004
Total disposable household income including old-age and survivor's benefits (HY023)	33012	735	29601	703	1808	2737

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	6727	819	116	24	31	49
HY050N	3525	191	564	41	525	556
HY060N	2025	178	86	12	143	154
HY070N	1753	145	53	9	84	92
HY080N	5503	441	344	43	155	184
HY090N	12006	4735	260	106	34	65
HY110N	2016	357	50	11	94	98
HY120N	229	8	134	6	0	1862
HY130N	4981	463	238	31	139	144
HY140N	12215	265	12019	261	2993	2993
HY145N	-947	78	-437	37	0	1415
Gross income components at household level						
HY040G	7735	947	133	27	43	49
HY050G	3641	203	582	43	497	556
HY060G	2025	178	86	12	143	154
HY070G	1753	145	53	9	84	92
HY080G	5503	441	344	43	155	184
HY090G	13726	4900	297	111	11	65
HY110G	2139	411	53	12	87	98
HY120G	229	8	134	6	1610	1862
HY130G	4981	463	238	31	139	144
HY140G	11757	265	11582	263	2995	2995
Net income components at personal level						
PY010N	23444	423	10487	263	2143	2891
PY020N	876	55	125	9	554	904
PY035N	2009	214	65	10	187	187
PY050N	16362	991	1664	125	468	746
PY080N	0	0	0	0	0	0
PY090N	4685	336	124	13	151	187
PY100N	15580	188	3879	115	1760	1953
PY110N	9939	1004	113	19	83	98
PY120N	5532	1890	28	11	27	30
PY130N	8476	209	469	29	395	449
PY140N	2295	249	18	3	59	62
Gross income components at personal level						
PY010G	31403	579	14048	356	1015	2891
PY020G	876	55	125	9	554	904
PY035G	2009	214	65	10	187	187
PY050G	22344	1315	2438	179	393	814
PY080G	0	0	0	0	0	0
PY090G	5417	401	143	16	52	187
PY100G	18034	222	4490	134	985	1953
PY110G	11408	1163	129	22	37	98
PY120G	5947	1942	30	12	22	30
PY130G	9641	247	534	33	213	449
PY140G	2295	249	18	3	59	62
PY200G	2640	47	1056	28	5545	6288

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 6.

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	19030	717	18987	714	528	646
2 household members	23923	581	23923	581	1232	1658
3 household members	26317	792	26317	792	1086	1776
4 and more	20900	505	20900	505	2444	4584
Population by age group						
<25	21181	500	21181	500	1495	2629
25 to 34	24834	672	24834	672	561	1006
35 to 44	22877	663	22845	664	636	1108
45 to 54	23865	605	23865	605	811	1380
55 to 64	22781	598	22781	598	855	1271
65+	19873	347	19873	347	932	1270
Population by sex						
Male	22835	379	22826	379	2441	4094
Female	22006	337	22006	337	2849	4570

2.3. Non-sampling errors

2.3.1. Sampling frame and coverage errors

Since 2005 EU-SILC survey the samples for each year 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.

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.

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.

For EU-SILC 2009 about three quarters of respondents (83% of those filling in the household questionnaire and 81% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 2% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 74% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 2% - as doubtful.

For EU-SILC 2010 about three quarters of respondents (85% of those filling in the household questionnaire and 82% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 2% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 90% 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 2007, 2008, 2009 and 2010 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 2007, 2008, 2009 and 2010.

During the years 2007, 2008, 2009 and 2010 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 2007, 2008, 2009 and 2010. 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 6:

Sample size	wave			
	1	2	3	4
A	3830	3460	3169	3037
B	8606	7824	7105	6872

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 6):

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

$$Ra = 0.986$$

$$Rh = 0.720$$

$$NRh = 28.9$$

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

$$Rp = 0.931$$

$$NRp = 6.9$$

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

$$*NRp = 33.9$$

Response rate for household:

Comparison between wave 2 and wave 1 (subsample 6).

- Wave response rate = 0.907

(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 = 0.931

(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 = 0.941

(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 = 0.910

(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 6).

- Wave response rate = 0.914

- Longitudinal follow-up rate = 0.936

- Follow-up rate = 0.947

- Achieved samples size ratio = 0.922

Comparison between wave 4 and wave 3 (subsample 6).

- Wave response rate = 0.936

- Longitudinal follow-up rate = 0.950

- Follow-up rate = 0.956

- Achieved samples size ratio = 0.962

Response rate for persons:

Personal interview response rates in wave 2 (subsample 6).

- Wave response rate = 0.930
(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 = 0.842
(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 = 0.847
(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 6).

- Wave response rate = 0.924
- Wave response rate of co-residents = 0.000
- Longitudinal follow-up rate = 0.839
- Achieved sample size ratio = 0.841
- Response rate for non-sample persons = 0.821

Personal interview response rates in wave 4 (subsample 6).

- Wave response rate = 1.000
- Wave response rate of co-residents = 0.000
- Longitudinal follow-up rate = 0.945
- Achieved samples size ratio = 0.855
- Response rate for non-sample persons = 1.000

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 6).

Household questionnaire result

DB130	Total	%
Total	5318	100.0
11 – household questionnaire completed	3834	72.1
21 – refusal to co-operate	1171	22.0
22 – entire household temporarily away for duration of fieldwork	166	3.1
23 – household unable to respond (illness, incapacity,...)	114	2.1
24 – other reasons	33	0.6
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3834	100.0
1 – interview accepted for database	3830	99.9
2 – interview rejected	4	0.1

Wave 2 (subsample 6).

Household status

DB110	Total	%
Total	3874	100.0
1 – at the same address as last interview	3693	95.3
2 – entire household moved to a private household within the country	62	1.6
3 – entire household moved to a collective household or institution within the country	2	0.1
4 – household moved outside the country	8	0.2
5 – entire household died	33	0.9
6 – household does not contain sample persons	1	0.0
7 – household unable to access	1	0.0
8 – split –off household	44	1.1
10 – fusion	1	0.0
11 – lost household	29	0.8

Record of contact at address

DB120	Total	%
Total*	3799	100.0
11 – address contacted*	3767	99.2
21 – address cannot be located	1	0.0
22 – address unable to access	0	0.0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	31	0.8

*There were added households from previous wave which were at the same address as last interview (DB110=1).

Household questionnaire result

DB130	Total	%
Total	3767	100.0
11 – household questionnaire completed	3460	91.2
21 – refusal to co-operate	217	5.8
22 – entire household temporarily away for duration of fieldwork	68	1.8
23 – household unable to respond (illness, incapacity,...)	12	0.3
24 – other reasons	10	0.3
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3460	100.0
1 – interview accepted for database	3460	100.0
2 – interview rejected	0	0.0

Wave 3 (subsample 6).

Household status

DB110	Total	%
Total	3589	100.0
1 – at the same address as last interview	3406	94.9
2 – entire household moved to a private household within the country	43	1.2
3 – entire household moved to a collective household or institution within the country	3	0.1
4 – household moved outside the country	10	0.3
5 – entire household died	27	0.8
6 – household does not contain sample persons	3	0.1
7 – household unable to access	3	0.1
8 – split –off household	43	1.2
10 – fusion	1	0.0
11 – lost household	50	1.4

Record of contact at address

DB120	Total	%
Total*	3492	100.0
11 – address contacted*	3475	99.5
21 – address cannot be located	1	0.0
22 – address unable to access	0	0.0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	16	0.5
Missing	0	

*There were added households from previous wave which were at the same address as last interview (DB110=1).

Household questionnaire result

DB130	Total	%
Total	3475	100.0
11 – household questionnaire completed	3169	91.2
21 – refusal to co-operate	183	5.3
22 – entire household temporarily away for duration of fieldwork	92	2.7
23 – household unable to respond (illness, incapacity,...)	28	0.8
24 – other reasons	3	0.1
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3169	100.0
1 – interview accepted for database	3169	100.0
2 – interview rejected	0	0

Wave 4 (subsample 6).

Household status

DB110	Total	%
Total	3282	100.0
1 – at the same address as last interview	3150	96.0
2 – entire household moved to a private household within the country	64	2.0
3 – entire household moved to a collective household or institution within the country	1	0.0
4 – household moved outside the country	6	0.2
5 – entire household died	14	0.4
6 – household does not contain sample persons	2	0.1
7 – Household unable to access (due to for example climatic conditions)	2	0.1
8 – split –off household	25	0.8
10 – fusion	1	0.0
11 – lost household (no information on record on what happened to the household)	17	0.5

Record of contact at address

DB120	Total	%
Total*	3239	100.0
11 – address contacted*	3216	99.3
21 – address cannot be located	1	0.0
22 – address unable to access	1	0.0
23 – address does not exist or is non-residential or unoccupied or is not principal residence	21	0.7

*There were added households from previous wave which were at the same address as last interview (DB110=1).

Household questionnaire result

DB130	Total	%
Total	3216	100.0
11 – household questionnaire completed	3037	94.4
21 – refusal to co-operate	89	2.8
22 – entire household temporarily away for duration of fieldwork	62	1.9
23 – household unable to respond (illness, incapacity,...)	26	0.8
24 – other reasons	2	0.1
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	3037	100.0
1 – interview accepted for database	3037	100.0
2 – interview rejected	0	0.0

Distribution of persons for membership status (RB110)

Wave 2 (subsample 6).

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	10482	10057	62	106	88	96	73	0
%	100.0	95.9	0.6	1.0	0.8	0.9	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	226	62	68	22	60	14
%	100.0	27.4	30.1	9.7	26.7	6.2

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

B - this person is not a current household member

Wave 3 (subsample 6).

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	9570	9105	64	132	89	101	79	0
%	100.0	95.1	0.7	1.4	0.9	1.1	0.8	0.0

Distribution of persons moving out by variable RB120.

	Total	RB110 = 5				
		RB120 = 1		RB120 = 2	RB120 = 3	RB120 = 4
		A	B			
Total	253	64	88	31	54	16
%	100.0	25.3	34.8	12.3	21.3	6.3

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

B - this person is not a current household member

Wave 4 (subsample 6).

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	9115	8781	42	96	72	62	62	0
%	100.0	96.3	0.5	1.1	0.8	0.7	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	154	42	50	13	41	8
%	100.0	27.3	32.5	8.4	26.6	5.2

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

B - this person is not a current household member

2.3.3.5. Item non-response

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	34.5	5.6	59.7
Total disposable household income	72.0	5.3	22.6
Total disposable household income before social transfers other than old-age and survivor's benefits	71.9	7.3	19.3
Total disposable household income before social transfers, including old-age and survivor's benefits	64.1	9.8	15.1
Net income components at household level			
hy040n	1.0	0.2	0.2
hy050n	20.8	0.3	0.4
hy060n	5.0	0.1	0.1
hy070n	4.6	0.2	0.0
hy080n	5.3	0.5	0.0
hy090n	1.2	0.9	0.0
hy100n	1.4	2.4	0.0
hy110n	4.0	0.1	0.0
hy120n	49.4	5.2	0.0
hy130n	4.0	0.3	0.0
hy140n	33.3	40.6	23.8
hy145n	38.4	2.6	0.0
hy170n	14.1	2.8	0.0
Gross income components at household level			
hy040g	1.2	0.2	0.0
hy050g	19.6	0.3	1.6
hy060g	5.0	0.1	0.1
hy070g	4.6	0.2	0.0
hy080g	5.3	0.5	0.0
hy090g	0.6	0.9	0.7
hy100g	1.4	2.4	0.0
hy110g	3.7	0.1	0.4
hy120g	49.4	5.2	0.0
hy130g	4.0	0.3	0.0
hy140g	33.3	39.8	24.8
hy170g	14.1	2.8	0.0

	% 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	7.9	0.1
py020n	8.1	3.1	1.1
py021n	0.2	0.2	0.0
py035n	2.7	0.7	0.0
py050n	5.6	3.1	0.2
py080n	0.0	0.0	0.0
py090n	2.4	0.4	0.0
py100n	22.5	1.9	0.3
py110n	1.1	0.2	0.0
py120n	0.4	0.1	0.0
py130n	5.6	0.7	0.0
py140n	1.5	0.1	0.0
Gross income components at personal level			
py010g	15.1	7.9	16.6
py020g	8.1	3.1	1.1
py021g	0.2	0.2	0.0
py030g	0.0	20.5	2.5
py035g	2.7	0.7	0.0
py050g	5.7	2.5	2.7
py080g	0.0	0.0	0.0
py090g	1.1	0.4	1.3
py100g	12.6	1.9	10.1
py110g	0.5	0.2	0.7
py120g	0.2	0.1	0.2
py130g	2.8	0.7	2.8
py140g	1.5	0.1	0.0

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	32.5	7.5	59.8
Total disposable household income	68.4	7.1	24.4
Total disposable household income before social transfers other than old-age and survivor's benefits	68.5	9.3	20.8
Total disposable household income before social transfers, including old-age and survivor's benefits	61.1	12.8	15.3
Net income components at household level			
hy040n	0.9	0.2	0.2
hy050n	18.5	0.4	0.5
hy060n	4.3	0.2	0.0
hy070n	3.5	0.1	0.0
hy080n	5.0	0.6	0.0
hy081n	2.2	0.2	0.0
hy090n	1.2	0.8	0.0
hy100n	1.2	2.7	0.0
hy110n	3.4	0.1	0.0
hy120n	50.2	5.7	0.0
hy130n	4.4	0.4	0.0
hy131n	1.1	0.1	0.0
hy140n	32.1	42.7	23.4
hy145n	33.7	4.0	0.1
hy170n	15.2	3.4	0.0
Gross income components at household level			
hy040g	1.2	0.2	0.0
hy050g	17.5	0.4	1.5
hy060g	4.3	0.2	0.0
hy070g	3.5	0.1	0.0
hy080g	5.0	0.6	0.0
hy081g	2.2	0.2	0.0
hy090g	0.5	0.8	0.7
hy100g	1.2	2.7	0.0
hy110g	3.1	0.1	0.4
hy120g	50.2	5.7	0.0
hy130g	4.4	0.4	0.0
hy131g	1.1	0.1	0.0
hy140g	31.8	42.2	24.3
hy170g	15.2	3.4	0.0

	% 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	30.4	10.0	0.1
py020n	8.5	3.5	1.4
py021n	0.2	0.3	0.0
py035n	2.5	0.8	0.0
py050n	6.2	3.2	0.3
py080n	0.0	0.0	0.0
py090n	2.1	0.5	0.0
py100n	24.0	2.5	0.3
py110n	0.9	0.2	0.0
py120n	0.3	0.1	0.0
py130n	5.1	0.8	0.0
py140n	0.9	0.1	0.0
Gross income components at personal level			
py010g	14.0	10.0	16.5
py020g	8.5	3.5	1.4
py021g	0.2	0.3	0.0
py030g	2.5	24.8	0.3
py031g	2.5	24.6	0.0
py035g	2.5	0.8	0.0
py050g	5.3	2.1	3.4
py080g	0.0	0.0	0.0
py090g	1.4	0.5	0.8
py100g	13.6	2.5	10.7
py110g	0.4	0.2	0.5
py120g	0.2	0.1	0.1
py130g	2.7	0.8	2.4
py140g	0.9	0.1	0.0

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	31.0	7.3	61.7
Total disposable household income	66.3	6.3	27.4
Total disposable household income before social transfers other than old-age and survivor's benefits	66.4	8.5	24.1
Total disposable household income before social transfers, including old-age and survivor's benefits	58.9	12.1	18.2
Net income components at household level			
hy040n	0.9	0.2	0.3
hy050n	17.4	0.4	0.7
hy060n	3.8	0.2	0.0
hy070n	2.9	0.1	0.0
hy080n	5.2	0.6	0.0
hy081n	2.3	0.2	0.0
hy090n	1.2	1.1	0.0
hy100n	1.7	3.1	0.0
hy110n	3.1	0.2	0.0
hy120n	49.6	7.8	0.0
hy130n	4.5	0.3	0.0
hy131n	0.9	0.1	0.0
hy140n	30.5	44.8	23.5
hy145n	39.2	5.0	0.0
hy170n	16.8	3.3	0.0
Gross income components at household level			
hy040g	1.2	0.2	0.0
hy050g	16.4	0.4	1.8
hy060g	3.8	0.2	0.0
hy070g	2.9	0.1	0.0
hy080g	5.2	0.6	0.0
hy081g	2.3	0.2	0.0
hy090g	0.5	1.1	0.7
hy100g	1.7	3.1	0.0
hy110g	2.8	0.2	0.3
hy120g	49.6	7.8	0.0
hy130g	4.5	0.3	0.0
hy131g	0.9	0.1	0.0
hy140g	30.2	44.5	24.1
hy170g	16.8	3.3	0.0

	% 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	30.6	10.6	0.1
py020n	8.6	3.9	1.6
py021n	0.2	0.3	0.0
py035n	2.4	0.8	0.0
py050n	6.4	3.3	0.6
py080n	0.0	0.0	0.0
py090n	2.0	0.6	0.0
py100n	24.3	2.7	0.4
py110n	0.9	0.2	0.0
py120n	0.5	0.1	0.0
py130n	5.1	0.8	0.0
py140n	0.9	0.1	0.0
Gross income components at personal level			
py010g	13.8	10.6	16.9
py020g	8.6	3.9	1.6
py021g	0.2	0.3	0.0
py030g	2.0	26.8	0.4
py031g	0.5	4.4	0.0
py035g	2.4	0.8	0.0
py050g	5.3	2.1	3.6
py080g	0.0	0.0	0.0
py090g	0.8	0.6	1.2
py100g	13.0	2.7	11.7
py110g	0.4	0.2	0.6
py120g	0.3	0.1	0.2
py130g	2.7	0.8	2.5
py140g	0.9	0.1	0.0

2010

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	32.0	7.5	60.5
Total disposable household income	66.2	6.6	27.2
Total disposable household income before social transfers other than old-age and survivor's benefits	66.3	8.8	24.1
Total disposable household income before social transfers, including old-age and survivor's benefits	59.1	13.2	18.0
Net income components at household level			
hy040n	0.9	0.3	0.3
hy050n	16.3	0.4	0.5
hy060n	3.9	0.2	0.0
hy070n	2.8	0.2	0.0
hy080n	5.1	0.8	0.0
hy081n	2.5	0.2	0.0
hy090n	1.4	1.2	0.0
hy100n	1.8	3.2	0.0
hy110n	2.9	0.2	0.0
hy120n	52.1	9.5	0.0
hy130n	4.6	0.3	0.0
hy131n	0.9	0.1	0.0
hy140n	31.4	43.4	23.8
hy145n	40.6	5.3	0.0
hy170n	17.5	3.6	0.0
Gross income components at household level			
hy040g	1.2	0.3	0.0
hy050g	15.3	0.4	1.5
hy060g	3.9	0.2	0.0
hy070g	2.8	0.2	0.0
hy080g	5.1	0.8	0.0
hy081g	2.5	0.2	0.0
hy090g	0.5	1.2	1.0
hy100g	1.8	3.2	0.0
hy110g	2.6	0.1	0.4
hy120g	52.1	9.5	0.0
hy130g	4.6	0.3	0.0
hy131g	0.9	0.1	0.0
hy140g	31.1	43.3	24.3
hy170g	17.5	3.6	0.0

	% 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	29.8	10.4	0.1
py020n	8.1	3.7	1.3
py021n	0.2	0.3	0.0
py035n	2.1	0.8	0.0
py050n	6.5	3.5	0.5
py080n	0.0	0.0	0.0
py090n	2.2	0.4	0.0
py100n	24.8	2.7	0.5
py110n	1.1	0.2	0.0
py120n	0.5	0.1	0.0
py130n	5.2	0.8	0.1
py140n	0.8	0.1	0.0
Gross income components at personal level			
py010g	14.2	10.4	15.6
py020g	8.1	3.7	1.3
py021g	0.2	0.3	0.0
py030g	2.1	28.0	0.3
py031g	0.5	4.2	0.0
py035g	2.1	0.8	0.0
py050g	5.3	2.1	3.9
py080g	0.0	0.0	0.0
py090g	0.8	0.4	1.5
py100g	13.5	2.7	11.7
py110g	0.5	0.2	0.6
py120g	0.4	0.1	0.1
py130g	2.8	0.8	2.5
py140g	0.8	0.1	0.0

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 6).

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	9243	8606	637
%	100.0	93.1	6.9

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	8606	7134	0	0	0	1472
%	100.0	82.9	0.0	0.0	0.0	17.1

Wave 2 (subsample 6).

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	8414	7824	590
%	100.0	92.0	3.0

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

	Total	RB250=11	RB250=14
Total	8316	7741	575
%	100.0	93.1	6.9

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

	Total	RB250=11	RB250=14
Total	98	83	15
%	100.0	84.7	15.3

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	7824	6387	0	0	0	1437
%	100.0	81.6	0.0	0.0	0.0	18.4

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	7741	6328	0	0	0	1413
%	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	83	59	0	0	0	24
%	100.0	71.1	0.0	0.0	0.0	28.9

Wave 3 (subsample 6).

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	7701	7105	596
%	100.0	92.3	7.7

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

	Total	RB250=11	RB250=14
Total	7514	6946	568
%	100.0	92.4	7.6

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

	Total	RB250=11	RB250=14
Total	187	159	28
%	100.0	85.0	15.0

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	7105	5739	0	0	0	1366
%	100.0	80.8	0.0	0.0	0.0	19.2

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	6946	5626	0	0	0	1320
%	100.0	81.0	0.0	0.0	0.0	19.0

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	159	113	0	0	0	46
%	100.0	71.1	0.0	0.0	0.0	28.9

Wave 4 (subsample 6).

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=14
Total	7438	6872	566
%	100.0	92.4	7.6

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

	Total	RB250=11	RB250=14
Total	7056	6546	510
%	100.0	92.8	7.2

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

	Total	RB250=11	RB250=14
Total	382	326	56
%	100.0	85.3	14.7

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	6872	5510	0	0	0	1362
%	100.0	80.2	0.0	0.0	0.0	19.8

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	6545	5290	0	0	0	1256
%	100.0	80.8	0.0	0.0	0.0	19.2

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	326	220	0	0	0	106
%	100.0	67.5	0.0	0.0	0.0	32.5

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 – the correct ones which would be missed 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 values. 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.

Out of 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 there are no donors 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 properties 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 widely used for incomes from hired employment, as:

- it is an important category of income, declared by a significant rate of respondents which, if present, has a significant share in the total household's 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 break-down 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 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 only reduces that risk.

Deterministic imputation is applied where missing data concern 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 have been selected so as 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 have been tested on the file of correct data and in the majority of cases they proved to be significant. Some of the explanatory variables have been 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) 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

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.

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 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.

Methodology

For the purposes of imputed rent estimation, regression analysis has been used. It was decided to use econometric methods, and especially regression analysis.

The first step consists in the estimation of a hedonic price function according to which actual rents paid by tenants depend on the main characteristics of dwellings. In the second step an imputed rent is ascribed to all households, which do not pay rent at market price. In case of households, which do not pay any rent (e.g. owners), the imputed rent is equal to the forecast from the hedonic model (based on their dwelling characteristics). In case of tenants, who pay rent below market prices (reduced rent), the imputed rent is equal to the difference between the forecasted rent and the actual rent (if the forecasted rent is lower than the actual rent, the imputed rent equal to zero is ascribed).

Data in the panel dataset refer to the years 2007 – 2010. The analysis has been made separately for each year. There are small differences in description of the flat/house on questionnaire in particular years (what implies small differences in the model form, regressors set) but the general methodology is the same. Detailed description of the sample and modeling and statistics presented below refers to the 2010 subset.

Subsets of tenants

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

The sample of 2010 covered 305 such households, of which 270 (2,1% of all households participating in the survey) gave the amount of rentals, while 35 households did not, although they declared such payments.

Form of the hedonic function

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 a total rent was introduced. If in the time of the survey a household did not pay any rent, it could declare the monthly rent paid last.

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.

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).

There was relatively wide set of potential explanatory variables, which described housing condition considering the aspects mentioned above. The set contained mandatory target variables for EU-SILC survey and other variables included in EU-SILC-G1 questionnaire, collected for domestic use. The set of potential regressors was made according to Eurostat recommendations included in the handbook.

Final set of explanatory variables was obtained by statistical and logical verification of the model. It was selected taking into account parameters values (and signs) and statistical significance of potential regressors. The final set of explanatory variables contains: class of locality, region, dwelling area, dwelling standard, housing infrastructure (water supply, bathroom, heating), crime threat, form of property and – finally – the indicator, if it is self-contained dwelling or not.

Estimation technique

Taking into account the fact, that the survey is based on the representative method, the weighted least square method (WLS) was applied. As weights for the regression procedure the survey weights are used. They are contained in the variable DB090.

Using the final version of regression model imputed rentals were determined for all the households except those paying the actual rentals.

Main characteristics

Sample size	12 930
Number of observations on tenants at market prices	270
R2 (adjusted)	0.68
Imputed rentals (in PLN per household, per year)	
Averages for all households (paying and do not paying actual rent)	7 415
Averages for the households, which do not pay any actual rent	7 514
Actual rentals (in PLN per household, per year)	
Averages for tenants paying rentals at market prices	7 143

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, 2006 for EU-SILC 2007;
- December 31, 2007 for EU-SILC 2008;
- December 31, 2008 for EU-SILC 2009;
- December 31, 2009 for EU-SILC 2010.

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 2007, EU-SILC 2008, EU-SILC 2009 and EU-SILC 2010).

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 2007, EU-SILC 2008, EU-SILC 2009 and EU-SILC 2010).

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 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 2007 was 2006,

EU-SILC 2008 was 2007,

EU-SILC 2009 was 2008,

EU-SILC 2010 was year 2009.

Reference period for taxes on income and social insurance contributions

No difference to the common definition in either wave (EU-SILC 2007, EU-SILC 2008, EU-SILC 2009 and EU-SILC 2010).

The reference period for income tax prepayment and compulsory social insurance contributions were again years 2006, 2007, 2008, 2009. The account clearance with the Treasury Office (including payments and returns) effected in 2007 refers to the income for 2006, in 2008 for 2007 in 2009 for 2008 and in 2010 for 2009.

The reference period for taxes on wealth

No difference to the common definition in either wave (EU-SILC 2007, EU-SILC 2008, EU-SILC 2009 and EU-SILC 2010).

Taxes on wealth paid during the income reference period were recorded, properly for year 2006, 2007, 2008 or 2009.

The lag between the income reference period and current variables

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

The total duration of the data collection of the sample

EU-SILC was performed on the territory of the whole country in:
2007 year between May 2 and June 19,
2008 year between May 2 and June 26,
2009 year between May 4 and June 26,
2010 year between May 4 and June 28.

3.2. Components of income

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

Differences between the national definitions and standards EU-SILC definitions, and an assessment:

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

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

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 three waves likewise unless it is indicated.

PY010 - Cash or near-cash employee 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 2006 the information collected only referred to the income gained from the use of the company car for private purposes. Since 2007 the collected information also refers to other non-cash employee income, but only at component level (PY021) and is not included in the income.

PY080 and PY130 - Regular inter-household cash transfer received

Since 2008 these variables include alimonies (compulsory and voluntary). It will be created to new separate alternate variables (PY081G/PY081N – Alimonies received: compulsory + voluntary and HY131G/HY131N – Alimonies paid: compulsory + voluntary).

PY110 - Survivors` 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), as 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.

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 065/10.

Arrears on utility bills – lack of information that the question applies to the main dwelling.

Ability to keep home adequately warm – the question included the expression: “if the households wants”.

Capacity to afford paying for one week annual holiday away from home – the question included the expression “if the households wants”

Capacity to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day – the question included the expression “if the households wants”

Ability to make ends meet – the question does not specify that "net monthly income" is meant. However, a detailed explanation is given in the guidelines for the interviewer in accordance with the recommendations of Eurostat. In the Polish conditions it is not necessary to use the

term "monthly net income". If we ask about the income, it is clear for the respondent that we mean monthly net income.

The lowest monthly income to make ends meet – the question does not specify that "net monthly income" is meant. Also in this case a detailed explanation is given in the guidelines for the interviewer in accordance with the recommendations of Eurostat.

Pollution, grime or other environmental problems – the question was formulated in a different way, namely: "Is your dwelling situated in a particularly inconvenient environment due to environment pollution (i.e. dust, smoke, unpleasant smell, polluted water)?".

Tenure status of dwelling – change the definition in compared to previous years.

In the case of housing cooperative dwellings users can have two different titles to the flat which they live in, i.e.:

- cooperative member's tenancy right
- cooperative member's ownership right

Cooperative member's tenancy right - it is an inalienable right: you cannot sell the dwelling, donate it, inherit it, nor lose it through enforcement by execution.

Cooperative member's ownership right – this is a tradable right: dwelling can be sold, donated rented/lent for use, inherited, but also lost through enforcement by execution.

In Polish law the cooperative member's ownership right means that one has the right to the premises, but the building and the land beneath it are owned by a housing cooperative. This so-called "limited property right" to the property, which is not separate property. For a cooperative member's ownership right to the property a land and mortgage register can be kept - which allows taking out a mortgage.

Given the above, since 2010:

1. Dwellings occupied under cooperative member's ownership right are included, accordingly in the following categories:

- outright owner HH021 = 1 (not encumbered with a mortgage)
- owners paying mortgage HH021 = 2

1. Dwellings occupied under cooperative member's tenancy right are included in the category: accommodation is provided free (HH021 = 5).

In 2005-2009 all cooperative dwellings were included in the category: accommodation is provided free (HH020 = 4)

In 2008 changes were introduced to these differences according with document 065/04. There were no other major divergences from common definitions.

VARIABLES ON THE ECONOMIC STATUS

In 2009 the recommendations from EU-SILC 065 (2009 operation) concerning economic status variables were applied. The variables introduced:

PL031: SELF-DEFINED CURRENT ECONOMIC STATUS (replaced PL030)

PL073: NUMBER OF MONTHS SPENT AT FULL-TIME WORK AS EMPLOYEE

PL074: NUMBER OF MONTHS SPENT AT PART-TIME WORK AS EMPLOYEE

**PL075: NUMBER OF MONTHS SPENT AT FULL-TIME WORK AS SELF-EMPLOYED
(INCLUDING FAMILY WORKER)**

**PL076: NUMBER OF MONTHS SPENT AT PART-TIME WORK AS SELF-EMPLOYED
(INCLUDING FAMILY WORKER)**

PL086: NUMBER OF MONTHS SPENT AS DISABLED OR/AND UNFIT TO WORK

PL088: NUMBER OF MONTHS SPENT IN COMPULSORY MILITARY SERVICE

**PL089: NUMBER OF MONTHS SPENT FULFILLING DOMESTIC TASKS AND CARE
RESPONSIBILITIES**

PL111: NACE REV.2 (since 2009 without PL110)

PL211: MAIN ACTIVITY (JANUARY – DECEMBER) –replaced PL210

Additionally, the following variables were removed:

PL070: NUMBER OF MONTHS SPENT AT FULL-TIME WORK

PL072: NUMBER OF MONTHS SPENT AT PART-TIME WORK

No difference to the common definition.

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 - 2010.

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, as well as data 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 is 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.

The differences in distributions may be due to different sample sizes of both surveys (HBS has approximately three times larger sample than the EU-SILC).

Tab. 1. Structure of population by age

Specification	EU-SILC 2007	HBS 2007	EU-SILC 2008	HBS 2008	EU-SILC 2009	HBS 2009	EU-SILC 2010	HBS 2010
	<i>In %</i>							
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0-14	16.0	18.1	15.7	18.1	15.5	18.0	15.3	18.0
15-24	15.5	16.1	15.1	15.8	14.6	15.7	14.0	15.1
25-54	44.0	41.9	44.0	41.6	44.0	41.1	44.0	41.2
55-64	11.1	11.6	11.8	12.1	12.4	12.7	13.1	13.2
65+	13.4	12.3	13.5	12.4	13.6	12.5	13.7	12.4

There are no significant differences when comparing the results from both surveys conducted between 2007 and 2010. When analyzing the structure of the age groups, the largest differences can be observed in the range 25 - 54 years (only 2.9 percentage points in 2009 and 2.8 in 2010). At the same time, in the EU-SILC survey, this age group remained at the same level throughout this period.

Tab. 2. Structure of population by level of education

Specification	EU-SILC 2007	HBS 2007	EU-SILC 2008	HBS 2008	EU-SILC 2009	HBS 2009	EU-SILC 2010	HBS 2010
	<i>In %</i>							
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.00	100.00
No school education	2.1	0.9	1.9	0.7	1.6	0.7	1.5	0.6
Completed primary	18.5	19.1	16.4	17.0	16.6	16.2	15.8	15.4
Lower secondary	5.1	6.6	5.1	6.8	4.9	6.7	4.8	6.6
Elementary vocational	26.8	26.6	26.8	27.0	26.7	27.0	26.8	27.5
Secondary	33.6	33.9	34.4	34.5	34.2	34.5	34.0	33.2
Higher	13.8	12.9	15.4	14.0	16.0	14.9	17.1	16.7

The number of people without education decreases from year to year in both surveys. In the EU-SILC the phenomenon is more evident due to a larger share of this category (HBS: 0.9% in 2007 – 0.6% in 2010; EU-SILC: 2.1% in 2007 – 1.5% in 2010).

There were no significant differences either between the years or between the results from both surveys.

Tab. 3. Structure of households by socioeconomic group

Households of	Households							
	EU-SILC 2007	HBS 2007	EU-SILC 2008	HBS 2008	EU-SILC 2009	HBS 2009	EU-SILC 2010	HBS 2010
<i>Total = 100</i>								
Employees	49.5	45.1	52.3	47.5	53.3	49.4	52.5	48.9
Farmers	2.6	4.6	2.6	4.3	2.5	4.1	2.7	4.1
Self-employed	4.8	6.1	4.6	6.4	4.8	6.6	5.5	6.9
Retirees	27.6	27.9	27.4	27.6	27.3	28.0	28.1	28.3
Pensioners	9.3	10.3	8.4	9.2	7.6	7.8	7.2	7.6
Maintained from non-earned sources	6.1	6.0	4.8	5.0	4.4	4.1	3.9	4.2

Tab. 4. Structure of persons in households by socioeconomic group

Households of	Persons in households							
	EU-SILC 2007	HBS 2007	EU-SILC 2008	HBS 2008	EU-SILC 2009	HBS 2009	EU-SILC 2010	HBS 2010
<i>Total = 100</i>								
Employees	59.3	53.6	62.7	56.3	64.0	58.7	62.9	58.0
Farmers	3.6	7.0	3.5	6.6	3.4	6.2	3.7	6.3
Self-employed	5.6	7.1	5.2	7.6	5.4	7.7	6.2	8.2
Retirees	19.4	19.5	18.6	18.9	18.2	18.7	18.9	19.0
Pensioners	6.4	7.4	5.2	6.4	4.6	5.3	4.5	5.0
Living on uneamed sources	5.9	5.4	4.7	4.2	4.4	3.4	3.8	3.5

Analyzing the structure of households in terms of socio-economic groups in both surveys, we observed significant differences in two groups: persons employed (higher percent in the EU-SILC) and farmers (higher percent in HBS).

Modifications to sampling methodology in the HBS survey (before 2006) resulted in more precise data in a study of farmers' income. These changes consisted of increasing the number of rural areas in the sample. In EU-SILC the sampling methodology has also been changed, introducing in the 2012 survey a new stratification method. It is expected that this will improve the quality of data on the income of rural areas.

Tab. 5. Average yearly equivalent income in PLN by socio-economic group

Households of	EU-SILC 2007	HBS 2006	EU-SILC 2008	HBS 2007	EU-SILC 2009	HBS 2008	EU-SILC 2010	HBS 2009
Disposable income								
Total	16166	14767	18684	16549	21018	18645	22142	19840
Employees	18140	15455	20899	17120	23282	19621	24432	20966
Farmers	10550	13891	13550	16879	13596	17556	14904	17747
Self-employed	17828	20416	21348	23317	26554	24945	26142	26132
Retirees	14670	14278	15618	15089	17209	16445	18710	17700
Pensioners	10888	10478	12061	11472	12930	12104	13598	12775
Living on unearned sources	8723	9272	9665	10327	11063	11636	11951	11684
Income from hired work								
Total	9649	7482	11719	8857	13264	10289	13617	10921
Employees	15195	12677	17609	14287	19468	16053	20341	17204
Farmers	1042	1367	1356	1669	1662	1862	1504	2096
Self-employed	2700	2805	3201	3412	4711	3934	4158	4332
Retirees	1466	1423	1545	1694	1746	1823	2025	1849
Pensioners	1175	1022	1178	1234	1330	1314	1360	1275
Living on unearned sources	1665	736	2401	936	2576	1134	1876	965

The most similar results in both surveys were obtained in households of pensioners and retirees. Households dependent on agriculture show higher incomes in the HBS survey than in the EU-SILC. In HBS these incomes include also the value of goods produced for own consumption.

Differences in income from employment between surveys arise from the method of data collection. In EU-SILC the annual income data is collected soon after the settlement with the Tax Office and often respondents provide these data on the basis of relevant documents. By contrast, the HBS survey enquires about expenditure and consumption, while income is only additional information. HBS data is collected each month which results in lower availability of evidence of the income. However, in the case of income from agriculture, this method of data collection gives better results.

Tab. 6. Average yearly equivalent income in PLN by number of persons

Households with specified number of persons	EU-SILC 2007	HBS 2006	EU-SILC 2008	HBS 2007	EU-SILC 2009	HBS 2008	EU-SILC 2010	HBS 2009
Disposable income								
Total	16166	14769	18684	16549	21018	18645	22142	19840
1	15412	14686	16883	15540	19096	17121	20436	18377
2	18725	17225	21201	18849	23652	21145	24812	22742
3	18561	16511	21001	18690	23830	21056	25698	22516
4	16235	14862	19273	16936	21879	19168	22414	20369
5	13733	12716	16238	14498	18455	16552	19880	17269
6 and more	12456	11187	14947	12605	16184	14183	16750	14936
Income from hired work								
Total	9649	7482	11719	8857	13264	10289	13617	10921
1	4713	4103	5675	4627	6549	5551	7154	5931
2	7980	6145	9830	7426	11549	8726	11793	9513
3	126892	9955	14861	11635	16760	13546	17738	14146
4	11714	9393	14363	11036	16306	12611	16247	13554
5	88895	6735	11012	8179	12562	9542	13174	9826
6 and more	7166	4684	9019	5679	9592	6678	9480	7119

In all the presented years and groups of households, incomes shown in the EU-SILC are higher than in HBS. This results from the previously mentioned methodological differences and method of calculating annual income of the HBS survey (average monthly income multiplied by 12).

In both surveys there is however a comparable rate of income's growth over the reported years

Tab. 7. Households provided with selected durables

Specification	EU-SILC 2007	HBS 2006	EU-SILC 2008	HBS 2007	EU-SILC 2009	HBS 2008	EU-SILC 2010	HBS 2009
<i>In %</i>								
Fixed telephone	71.6	67.9	68.5	64.2	64.9	60.7	61.7	55.2
Mobile telephone	75.5	79.3	79.9	83.5	84.1	86.5	85.9	88.8
Television set	97.1	98.5	97.5	98.5	97.8	98.5	97.6	98.4
Computer	48.7	50.1	54.5	56.4	59.3	60.8	62.8	64.9
Printer	35.7	33.6	40.0	37.1	43.9	40.1	46.7	41.4
Internet connection	34.8	36.6	43.0	45.6	51.1	53.4	55.8	59.6
Microwave oven	37.9	42.4	41.9	46.1	46.6	50.1	49.7	52.6
Dishwasher	9.3	7.4	11.5	9.6	13.4	12.3	15.7	15.7
Refrigerator	97.6	98.9	97.7	98.4	98.5	97.9	98.6	98.5
Washing machine	96.6	97.1	96.8	97.3	97.6	97.9	97.9	98.0
Passenger car	53.6	52.5	56.2	54.7	58.9	57.4	61.3	59.5

Results from both studies are very similar. The greatest differences (above 5 percentage points) are observed in the Fixed Telephone and Printer categories.

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 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 26% - 33% of the operation surplus only (after deduction of section K).

Incomes from hired employment in EU-SILC 2009 are equal to 91% of the corresponding figure in SNA, while social benefits – 98% respectively, which seems to be a good result.

In EU-SILC 2009, as compared with EU-SILC 2007, the data coherence for disposable income with SNA increased by 17 percentage points. This is due to the coherence increase for self-employment income by 7 percentage points, i.e. from 26% to 33% and for social benefits by 5 percentage points. 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 2009.

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 is 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 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.

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

Category in SNA	Category description in EU-SILC	Incomes: 2006 r.			Incomes: 2007 r.			Incomes: 2008 r.			Incomes: 2009 r.		
		SNA	EU-SILC	SNA = 100%	SNA	EU-SILC	SNA = 100%	SNA	EU-SILC	SNA = 100%	SNA	EU-SILC	SNA = 100%
		in mln PLN	in mln PLN		in mln PLN	in mln PLN		in mln PLN	in mln PLN		in mln PLN	in mln PLN	
Gross disposable income (net)	Total disposable household income (net)	683 483	398 939	58	742 374	457 807	62	799 085	513 594	64	860 625	642 631	75
Wages. salaries and other income connected with hired work (gross)	Employee cash or near cash income (gross)	325 030	323 956	99	375 358	380 422	99	426 735	423 255	99	505 401	460 369	91
Gross operating surplus (gross) with the exception of section K	Self-employment income (gross) - value allocated to household`s consumption and dwelling-related investment	201 601	53 219	26	220 168	58 291	26	226 352	67 485	30	238 613	78 736	33
Social security benefits and social assistance benefits (gross)	Social benefits (gross)	161 336	149 258	93	166 880	153 565	93	179 112	163 289	91	197 755	193 177	98

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 depreciation.
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.