CENTRAL STATISTICAL OFFICE OF POLAND

FINAL QUALITY REPORT

GRANT AGREEMENT
NO 36401.2006.001-2006.199

ACTION ENTITLED:
EU-SILC 2007

Warsaw, November 2009

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

CONTENTS Page

PREFACE

The present quality report is the final quality report on EU-SILC 2007 carried out in Poland according to grant agreement No. 36401.2006.001-2006.199, 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 EU-SILC operations in Poland started in 2005, so by the year 2007 the panel had not been accomplished yet. Two of the 4 sub-samples (which form the total sample of the EU-SILC 2007) make up a panel lasting for the three consecutive years: 2005, 2006 and 2007 and one sub-sample makes up a panel lasting for two years: 2006 and 2007.

1. COMMON LONGITUDINAL EUROPEAN UNION INDICATORS

Longitudinal indicators are not available, as no rotational group has yet been in the survey for four years.

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 (PSU) 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 population number of dwellings in the stratum. Furthermore, the number of the first-stage units selected from the strata was obtained by dividing the number of dwellings in the sample by the number of dwellings determined for a given class of locality to be selected from the first-stage unit. In towns with over 100 000 population 3 dwellings per PSU were selected, in towns with 20-100 thousand population – 4 dwellings per PSU, in towns with less than 20 000 population – 5 dwellings per PSU, respectively. In rural areas 6 dwellings from each PSU were selected. Altogether 5912 census areas and 24044 dwellings were selected for the sample. The subsample 5 was selected for the survey in 2006 in order to replace the sub-sample 1. It consisted of 1476 census areas and 6002 dwellings. Sub-samples 2, 3, 4 formed the longitudinal (panel) component. For cross-sectional component of the survey conducted in 2007 year the new subsample number 6 was selected. This subsample replaced subsample number 2, and consisted of 1487 PSU and 6008 dwellings.

2.1.5. Sample selection schemes

Census areas were selected according to the Hartley-Rao scheme. Prior to selection census areas were put in random order separately for each stratum and then the determined number of PSU was selected with probabilities proportionate to the number of dwellings. Then 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 is 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 which replaced the subsample 2. 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 PSU selected from the h-th stratum,
m'_h - number of dwellings selected from PSU in the h-th stratum,
M_h – number of dwellings in the h-th stratum.

Non-response adjustments

DB080 weights were then adjusted with the use of 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	0.699
1	Warsaw	0.399
2	Towns 500 000 – 1 000 000 inhabitants	0.567
3	Towns 100 000 – 500 000 inhabitants	0.636
4	Towns 20 000 – 100 000 inhabitants	0.695
5	Towns less than 20 000 inhabitants	0.740
6	Rural areas	0.823

The adjusted weights were calculated according to the formula:

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

The weights DB080 and DB080^{corrected} were calculated for the subsample 5. The next step consisted in calculating the weights DB090 and RB050 for the households of the subsample 5 with the use of the integrated calibration method as described below in section *Adjustments to external data*.

Adjustments to external data

Using the integrated calibration method (in hyperbolic sinus version) weights were calculated for individuals and for households simultaneously. To do this the 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). This information at the level of NUTS2, additionally classified by urban/rural areas were derived from the 2002 Census and current demographic estimates.

Final longitudinal weight

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

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

where:

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

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

The base weights of 2005 were used for the calculation of numerator and denominator. The above expression is the reciprocal of the empirical estimate of probability that a given person will be interviewed again in the second year of the survey.
In the second stage of the base weight calculation for the second year of the survey children of "*sample persons*" received the weights of mothers and "co-residents' i.e. additional persons included in the household surveyed were ascribed zero weights. Then the respondents' weights were averaged and all the members of a given household were ascribed such a mean weight. Then for the base weights thus obtained the trimming of extreme weights was applied. Adjustment to external data was not applied. The panel weight RB062 was calculated by dividing the base weights by 3.

Non-response adjustments – subsequent waves

done with the use of the integrated calibration method (by algorithm described in ch.2.1.8.2 for subsample 5).

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

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

where:

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

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

For the subsamples 3 and 4 (surveyed for the third time) algorithm based on method described for the subsample 5 was used. Additionally, re-entries occurrence was taken into account i.e. persons who were surveyed in 2005, not surveyed in 2006, and again surveyed in 2007 year. The base weights for such persons were computed by correction of base weights from year 2005 on data for years 2005 and 2007 (without information from 2006 year). Inclusion of re-entries to the subsamples surveyed in 2007 year caused the necessity of additional correction of the base weights for persons surveyed in the three successive years. Coefficients of these

corrections were computed separately according to classes of locality as ratios: weighted number of respondents surveyed in all three years to the weighted number of respondents in the last survey year (i.e. with re-entries); weight used in these calculations was the weight RB050 for year 2005.
The last stage of the base weight calculation for the third year of the survey consisted in receiving weights of mothers by children of "sample persons" and zero weights by "coresidents' i.e. additional persons included in the households.

Adjustments to external data

Adjustment to external data was not applied.

Final longitudinal weight

The panel weight RB062 was calculated by dividing the base weights for subsamples 3, 4, and 5 by 3.
The panel weight RB063 was calculated by dividing the base weights for subsamples 3 and 4 by 2.

Final household cross-sectional weight

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

DB090 – weight for households,

RB050 – weight for all household members but

$RB050_{ij} = DB090_i$

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

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

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

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

2.1.9. Substitutions

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

2.2. Sampling errors

Standard error and effective sample size

Estimation of standard errors was based on resampling approach. We used a bootstrap method which resamples 200 times from each stratum $n_h - 1$ PSU's (primary sampling units) with replacement (McCarthy and Snowden method (1985)), where n_h denotes the sample size of PSU's 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 also implemented linearization method of variance estimation for main poverty indicators, and the results of comparisons with bootstrap method were very similar.

Cross-sectional component

The mean, the total number of observations (before and after imputation) and the standard errors for the following income components (mean and standard errors based on weighted data while the number of observations based on unweighted results).

Income components	**Mean[1]**	**Standard error**	**Mean[2]**	**Standard error**	**Number of observations**	
					Before imputation	**After imputation**
Total household gross income (HY010)	40207	366	40141	367	5174	14272
Total disposable household income (HY020)	30079	256	30036	256	10464	14275
Total disposable household income before social transfers other than old-age and survivors benefits (HY022)	28216	261	27811	263	10438	14091
Total disposable household income including old-age and survivors benefits (HY023)	22955	294	20338	276	9308	12717
Net income components at household level						
HY040N	6306	495	83	10	124	178
HY050N	2693	62	519	16	3125	3225
HY060N	1625	72	78	5	717	741
HY070N	1422	41	70	4	692	709
HY080N	5252	232	368	22	821	907
HY090N	6627	1775	164	46	173	293
HY110N	2217	120	75	5	559	581
HY120N	215	5	114	3	7014	7714
HY130N	4255	220	199	14	620	662
HY140N	10127	117	9893	115	5016	13990
HY145N	-259	38	-100	15	5104	5469

[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	7255	595	96	11	148	178
HY050G	2775	67	535	17	2977	3225
HY060G	1625	72	78	5	717	741
HY070G	1422	41	70	4	692	709
HY080G	5252	232	368	22	821	907
HY090G	8197	2234	203	57	77	293
HY110G	2400	141	81	5	503	581
HY120G	215	5	114	3	7014	7714
HY130G	4255	220	199	14	620	662
HY140G	10016	120	9792	119	4996	14003
Net income components at personal level						
PY010N	17213	183	7645	114	10897	13709
PY020N	1048	52	139	8	2571	3948
PY035N	1882	65	71	4	833	1068
PY050N	12706	337	1150	39	2021	3109
PY080N	9049	4650	4	3	6	7
PY090N	5553	188	175	9	900	1052
PY100N	12863	74	3202	51	8122	8845
PY110N	7912	295	100	7	428	488
PY120N	1870	146	8	1	129	147
PY130N	7427	89	482	14	2113	2312
PY140N	1813	104	25	2	464	508
Gross income components at personal level						
PY010G	24181	254	10755	160	5379	13709
PY020G	1048	52	139	8	2571	3948
PY035G	1882	65	71	4	833	1068
PY050G	16703	436	1767	58	1981	3689
PY080G	11952	6176	5	4	3	7
PY090G	6422	222	202	10	433	1052
PY100G	14996	89	3734	59	4888	8845
PY110G	9194	351	116	8	181	488
PY120G	2253	187	9	1	68	147
PY130G	8549	106	556	16	1083	2312
PY140G	1813	104	25	2	464	508
PY200G	1989	18	808	11	9943	12626

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 while the number of observations based on unweighted results).

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	15481	277	15412	275	2346	2640
2 household members	18734	303	18725	302	5954	7498
3 household members	18575	353	18561	353	5850	8748
4 and more	14579	161	14577	161	14466	23966
Population by age group						
<25	14646	157	14636	157	9027	14176
25 to 34	17692	264	17677	264	3482	5628
35 to 44	16724	301	16719	302	3427	5216
45 to 54	16673	234	16651	234	4384	6851
55 to 64	17734	286	17717	287	3548	5020
65+	15630	131	15630	131	4748	5961
Population by sex						
Male	16297	138	16281	139	13521	20553
Female	16066	135	16059	135	15095	22299

Longitudinal component

The mean, the total number of observations (before and after imputation) and the standard errors for the following income components (mean and standard errors based on weighted data while the number of observations based on unweighted results), subsamples 3 and 4.

Income components	Mean[5]	Standard error	Mean[6]	Standard error	Number of observations	
					Before imputation	After imputation
Total household gross income (HY010)	39094	440	39066	440	2544	6820
Total disposable household income (HY020)	29508	310	29481	310	5030	6820
Total disposable household income before social transfers other than old-age and survivors benefits (HY022)	27529	309	27160	311	5024	6739
Total disposable household income including old-age and survivors benefits (HY023)	21553	348	18980	330	4476	6069

[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	6523	734	79	13	57	79
HY050N	2738	85	546	21	1525	1570
HY060N	1660	100	81	7	339	349
HY070N	1461	52	71	5	333	339
HY080N	4721	249	320	25	401	447
HY090N	4507	1319	110	34	85	144
HY110N	2179	158	72	7	247	259
HY120N	208	6	109	4	3303	3614
HY130N	3739	205	187	14	310	327
HY140N	9605	135	9397	137	2446	6686
HY145N	-293	29	-108	11	2337	2511
Gross income components at household level						
HY040G	7355	871	89	15	66	79
HY050G	2813	90	561	23	1464	1570
HY060G	1660	100	81	7	339	349
HY070G	1461	52	71	5	333	339
HY080G	4721	249	320	25	401	447
HY090G	5517	1624	134	41	38	144
HY110G	2373	186	79	8	217	259
HY120G	208	6	109	4	3303	3614
HY130G	3739	205	187	14	310	327
HY140G	9488	137	9289	139	2432	6691
Net income components at personal level						
PY010N	16885	227	6871	124	5144	6510
PY020N	920	64	107	8	1167	1830
PY035N	1856	90	60	5	358	468
PY050N	13149	476	1114	54	974	1479
PY080N	1466	450	1	0	4	4
PY090N	5726	210	180	12	448	514
PY100N	12995	103	3415	66	3990	4320
PY110N	8309	451	106	10	207	237
PY120N	2119	239	9	2	64	74
PY130N	7382	109	491	18	1061	1150
PY140N	1596	139	18	2	202	224
Gross income components at personal level						
PY010G	23687	312	9639	172	2563	6510
PY020G	920	64	107	8	1167	1830
PY035G	1856	90	60	5	358	468
PY050G	17110	623	1678	79	930	1724
PY080G	1854	589	1	1	2	4
PY090G	6621	248	208	14	217	514
PY100G	15152	123	3981	77	2487	4320
PY110G	9662	542	123	12	80	237
PY120G	2569	314	11	2	35	74
PY130G	8499	129	566	21	561	1150
PY140G	1596	139	18	2	202	224
PY200G	1951	22	679	12	4723	6020

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 while the number of observations based on unweighted results), subsamples 3 and 4.

Equivalised disposable income	Mean[7]	Standard error	Mean[8]	Standard error	Number of observations	
					Before imputation	After imputation
Subclasses by household size						
1 household member	14925	387	14896	387	1164	1298
2 household members	18485	349	18485	349	2820	3540
3 household members	18271	368	18245	369	2778	4110
4 and more	14440	200	14436	200	6984	11525
Population by age group						
<25	14482	188	14475	188	4334	6751
25 to 34	17082	371	17053	372	1594	2643
35 to 44	16509	385	16509	385	1616	2437
45 to 54	16450	261	16442	261	2112	3271
55 to 64	17583	316	17558	318	1745	2454
65+	15761	180	15761	180	2345	2917
Population by sex						
Male	16075	170	16063	170	6503	9813
Female	15888	162	15879	162	7243	10660

2.3. Non-sampling errors

2.3.1. Sampling frame and coverage errors

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

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

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

The whole territory of Poland is divided into enumeration statistical districts and census enumeration areas. In EU-SILC census enumeration areas are used as primary sampling units. The secondary sampling units are dwellings. For each census enumeration area a list of dwellings was made up to form the secondary sampling frame. All the households from the selected dwellings are supposed to enter the survey.
The TERYT system is updated annually with respect to the territorial division into statistical districts and census enumeration areas. The lists of dwellings, names of towns, villages and streets are updated. Other changes due to new construction, dismantle of buildings and administrative division modifications are also introduced.
In the longitudinal (panel) component consisting of the subsamples 3 and 4 some 6.7% of dwellings were found to be non-existing (cancelled, changed for non-residential units) as well as uninhabited or temporarily inhabited.

2.3.2. Measurement and processing errors

As with any other statistical survey, EU-SILC may be burdened with non-sampling errors which occur at various stages of the survey and which cannot be eliminated completely. This mainly applies to interviewers' errors at the stage of collecting the information, errors due to the respondents' misunderstanding of questions and inaccurate or sometimes even false answers as well as the errors taking place at the stage of data recording.
According to the interviewers, who after the household and individual interview completion were obliged to answer a few questions concerning interview performance for EU-SILC 2005, over 70% of the respondents showed a favourable attitude towards the survey, while about 5% were unwilling towards it. In the interviewers' opinion, in about 86% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
For EU-SILC 2006 on the basis of material it is possible to state that about three quarters of respondents (78% of those filling in the household questionnaire and 75% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 3% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 88% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
For EU-SILC 2007 about three quarters of respondents (80% of those filling in the household questionnaire and 78% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 3% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 89% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful.
The quality of income data in 2005, 2006 and 2007 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 in the case of gross income. The reason is that non-response to the highest degree affected the information on taxes and social and health insurance contributions.

In Poland EU-SILC was carried out in May/June 2005, 2006 and 2007.
During the years 2005, 2006 and 2007 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 having experience in other social surveys. Survey performance in the field was preceded by a series of trainings in 2005, 2006 and in 2007. Regional survey coordinators were instructed by CSO Social Statistics Division staff members and then the regional survey coordinators trained 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 together 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.
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

subsamples 3 and 4:

Sample size	**wave**		
	1	2	3
A	8121	7208	6733
B	18705	16978	15611

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

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

Unit non-response

wave 1 (subsamples 3 and 4):

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

Ra = 0.993
Rh = 0.705

NRh = 30.01

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

Rp = 0.948
NRp = 5.20

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

*NRp = 33.65.

Response rate for household:

Comparison of results codes between wave 2 and wave 1 (subsamples 3 and 4).

- Wave response rate = 0.887
(percentage of households successfully interviewed (DB135 = 1) which were passed on to wave (from wave *t-1*) or newly created or added during wave *t*, excluding those out of scope (under the tracing rules) or non-existent)

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

- Follow-up rate = 0.918
(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 = 0.888
(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 of results codes between wave 3 and wave 2 (subsamples 3 and 4).

- Wave response rate = 0.924
- Longitudinal follow-up rate = 0.940
- Follow-up rate = 0.952
- Achieved samples size ratio = 0.954

Response rate for persons:

Personal interview response rates in wave 2 (subsamples 3 and 4).

- Wave response rate = 0.955

(percentage of sample person 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 = 0.851

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

- Achieved samples size ratio = 0.896

(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 first wave).

- Response rate for non-sample persons = 0.901

(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 (subsamples 3 and 4).

- Wave response rate = 0.946
- Wave response rate of co-residents = 0.000
- Longitudinal follow-up rate = 0.852
- Achieved samples size ratio = 0.901
- Response rate for non-sample persons = 0.870

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 (subsamples 3 and 4*).*

Household questionnaire result

DB130	Total	%
Total	11525	100.0
11 – household questionnaire completed	8188	71.0
21 – refusal to co-operate	2312	20.1
22 – entire household temporarily away for duration of fieldwork	600	5.2
23 – household unable to respond (illness, incapacity,…)	319	2.8
24 – other reasons	106	0.9
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	8188	100.0
1 – interview accepted for database	8121	99.2
2 – interview rejected	67	0.8

Wave 2 (subsamples 3 and 4*).*

Household status

DB110	Total	%
Total	8206	100.0
1 – at the same address as last interview	7747	94.4
2 – entire household moved to a private household within the country	145	1.8
3 – entire household moved to a collective household or institution within the country	3	0.0
4 – household moved outside the country	29	0.4
5 – entire household died	50	0.6
6 – household does not containg sample person	9	0.1
7 –address non-contacted (unable to access, los-no information on record on what happened to the household-)	137	1.7
8 – split –off household	85	1.0
10 – fusion	1	0.0

Record of contact at address

DB120	Total	%
Total	230	100.0
11 – address contacted	152	66.1
21 – address cannot be located	1	0.4
22 – address unable to access	0	0.0
23 – address does not exit or is non-residential address or is unoccupied or not principal residence	77	33.5
Missing	0	0.0

Household questionnaire result

DB130	Total	%
Total	7899	100.0
11 – household questionnaire completed	7209	91.3
21 – refusal to co-operate	512	6.5
22 – entire household temporarily away for duration of fieldwork	107	1.4
23 – household unable to respond (illness, incapacity,…)	52	0.6
24 – other reasons	19	0.2
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	7209	100.0
1 – interview accepted for database ,	7208	100.0
2 – interview rejected	1	0.0

Wave 3 (subsamples 3 and 4*)*.

Household status

DB110	Total	%
Total	7478	100.0
1 – at the same address as last interview	7116	95.2
2 – entire household moved to a private household within the country	98	1.3
3 – entire household moved to a collective household or institution within the country	8	0.1
4 – household moved outside the country	28	0.4
5 – entire household died	37	0.5
6 – household does not containg sample person	3	0.0
7 –address non-contacted (unable to access, los-no information on record on what happened to the household-)	97	1.3
8 – split –off household	91	1.2
10 – fusion	0	0.0

Record of contact at address

DB120	Total	%
Total	189	100.0
11 – address contacted	144	76.2
21 – address cannot be located	1	0.5
22 – address unable to access	0	0
23 – address does not exit or is non-residential address or is unoccupied or not principal residence	44	23.3
Missing	0	0

Household questionnaire result

DB130	Total	%
Total	7260	100.0
11 – household questionnaire completed	6733	92.7
21 – refusal to co-operate	309	4.3
22 – entire household temporarily away for duration of fieldwork	139	1.9
23 – household unable to respond (illness, incapacity,…)	57	0.8
24 – other reasons	22	0.3
Missing	0	0.0

Household interview acceptance

DB135	Total	%
Total	6733	100.0
1 – interview accepted for database ,	6733	100.0
2 – interview rejected	0	0.0

Distribution of persons for membership status (RB110)

Wave 2 (subsamples 3 and 4*).*

Distribution of persons for membership status (RB110)

	Total	Current household members				No current household members		
		RB110=1	RB110=2	RB110=3	RB110=4	RB120 = 2 to 4	RB110=6	RB110=7
Total	22277	21276	124	310	192	233	140	2
%	100.0	95.5	0.6	1.4	0.9	1.0	0.6	0.0

Distribution of persons moving out by variable RB120.

	Total	**RB110 = 5**				
		RB120 = 1		**RB120 = 2**	**RB120 = 3**	**RB120 = 4**
		A	**B**			
Total	474	124	117	63	143	27
%	100	26.1	24.7	13.3	30.2	5.7

A – this person is a current household number of a household this wave

B - this person is not a current household member

Wave 3 (subsamples 3 and 4*).*

Distribution of persons for membership status (RB110)

	Total	Current household members				No current household members		
		RB110=1	RB110=2	RB110=3	RB110=4	RB120 = 2 to 4	RB110=6	RB110=7
Total	20642	19642	160	248	180	257	155	0
%	100.0	95.2	0.8	1.2	0.9	1.2	0.7	0.0

Distribution of persons moving out by variable RB120.

	Total	**RB110 = 5**				
		RB120 = 1		**RB120 = 2**	**RB120 = 3**	**RB120 = 4**
		A	**B**			
Total	554	160	137	63	155	39
%	100	28.9	24.7	11.4	08.0	7.0

A – this person is a current household number of a household this wave

B - this person is not a current household member

Item non-response (income variables)

2005

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	39.4	6.4	53.7
Total disposable household income	67.4	5.7	26.6
Total disposable household income before social transfers other than old-age and survivor's benefits	67.0	8.6	22.4
Total disposable household income before social transfers, including old-age and survivor's benefits	60.3	11.5	16.5
Net income components at household level			
HY040N	1.3	0.1	0.0
HY050N	23.4	0.3	0.6
HY060N	4.2	0.1	0.0
HY070N	6.0	0.2	0.0
HY080N	5.2	0.7	0.0
HY090N	1.0	0.7	0.0
HY110N	2.3	0.1	0.0
HY120N	42.2	4.8	0.0
HY130N	4.9	0.3	0.0
HY140N	39.2	34.5	22.8
HY145N	45.0	3.4	0.0
Gross income components at household level			
HY040G	1.3	0.1	0.0
HY050G	23.0	0.3	1.0
HY060G	4.2	0.1	0.0
HY070G	6.0	0.2	0.0
HY080G	5.2	0.7	0.0
HY090G	0.5	0.7	0.5
HY110G	1.9	0.1	0.4
HY120G	42.2	4.8	0.0
HY130G	4.9	0.3	0.0
HY140G	39.4	34.5	23.3

	% 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	29.2	6.9	0.1
PY021N	0.1	0.4	0.0
PY035N	2.9	0.6	0.0
PY050N	4.5	3.1	0.2
PY080N	0.0	0.0	0.0
PY090N	3.7	0.2	0.0
PY100N	22.6	1.3	0.4
PY110N	1.4	0.1	0.0
PY120N	0.3	0.1	0.0
PY130N	6.5	0.3	0.0
PY140N	0.9	0.0	0.0
Gross income components at personal level			
PY010G	15.6	6.9	13.7
PY021G	0.1	0.4	0.0
PY035G	2.9	0.6	0.0
PY050G	4.7	1.6	3.1
PY080G	0.0	0.0	0.0
PY090G	2.1	0.2	1.6
PY100G	15.2	1.3	7.8
PY110G	0.8	0.1	0.6
PY120G	0.1	0.1	0.3
PY130G	4.5	0.3	2.1
PY140G	0.9	0.0	0.0

2006

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	41.4	5.3	53.2
Total disposable household income	72.8	4.5	22.6
Total disposable household income before social transfers other than old-age and survivor's benefits	72.5	6.3	19.7
Total disposable household income before social transfers, including old-age and survivor's benefits	66.1	8.4	15.3
Net income components at household level			
HY040N	1.0	0.2	0.1
HY050N	23.2	0.3	0.3
HY060N	5.4	0.1	0.0
HY070N	5.6	0.2	0.0
HY080N	6.2	0.5	0.0
HY090N	0.8	0.5	0.0
HY110N	3.7	0.1	0.0
HY120N	46.2	4.3	0.0
HY130N	4.9	0.3	0.0
HY140N	40.7	31.9	24.9
HY145N	44.7	2.7	0.0
Gross income components at household level			
HY040G	1.1	0.2	0.0
HY050G	22.3	0.3	1.2
HY060G	5.4	0.1	0.0
HY070G	5.6	0.2	0.0
HY080G	6.2	0.5	0.0
HY090G	0.4	0.5	0.4
HY110G	3.4	0.1	0.3
HY120G	46.2	4.3	0.0
HY130G	4.9	0.3	0.0
HY140G	40.6	31.6	25.4

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	30.7	7.2	0.1
PY021N	0.1	0.2	0.0
PY035N	2.7	0.7	0.0
PY050N	5.7	2.8	0.3
PY080N	0.0	0.0	0.0
PY090N	3.4	0.3	0.0
PY100N	22.3	1.7	0.3
PY110N	1.5	0.1	0.0
PY120N	0.4	0.0	0.0
PY130N	6.3	0.5	0.0
PY140N	1.3	0.1	0.0
Gross income components at personal level			
PY010G	15.6	7.2	15.2
PY021G	0.1	0.2	0.0
PY035G	2.7	0.7	0.0
PY050G	5.1	1.8	3.1
PY080G	0.0	0.0	0.0
PY090G	2.0	0.3	1.3
PY100G	15.8	1.7	6.8
PY110G	0.9	0.1	0.6
PY120G	0.2	0.0	0.2
PY130G	4.5	0.5	1.8
PY140G	1.3	0.1	0.0

2007

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	36.9	5.2	57.8
Total disposable household income	73.9	5.2	20.9
Total disposable household income before social transfers other than old-age and survivor's benefits	73.6	7.3	17.8
Total disposable household income before social transfers, including old-age and survivor's benefits	65.6	10.0	13.4
Net income components at household level			
HY040N	0.8	0.2	0.2
HY050N	22.4	0.3	0.4
HY060N	5.0	0.1	0.0
HY070N	4.9	0.1	0.0
HY080N	5.9	0.6	0.0
HY090N	1.2	0.8	0.0
HY100N	0.7	1.6	0.0
HY110N	3.9	0.2	0.0
HY120N	49.1	4.8	0.0
HY130N	4.5	0.3	0.0
HY140N	35.5	36.4	26.1
HY145N	34.8	2.5	0.0
Gross income components at household level			
HY040G	1.0	0.2	0.0
HY050G	21.4	0.3	1.4
HY060G	5.0	0.1	0.0
HY070G	4.9	0.1	0.0
HY080G	5.9	0.6	0.0
HY090G	0.5	0.8	0.7
HY100G	0.7	1.6	0.0
HY110G	3.5	0.2	0.4
HY120G	49.1	4.8	0.0
HY130G	4.5	0.3	0.0
HY140G	35.3	35.8	27.0

	% 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	31.1	7.9	0.1
PY020N	7.1	3.0	0.9
PY021N	0.2	0.2	0.0
PY035N	2.3	0.7	0.0
PY050N	5.9	2.7	0.3
PY070N	6.0	1.6	0.0
PY080N	0.0	0.0	0.0
PY090N	2.6	0.4	0.0
PY100N	23.6	1.8	0.2
PY110N	1.3	0.2	0.0
PY120N	0.4	0.0	0.0
PY130N	6.2	0.5	0.0
PY140N	1.3	0.1	0.0
Gross income components at personal level			
PY010G	15.5	7.9	15.7
PY020G	7.1	3.0	0.9
PY021G	0.2	0.2	0.0
PY035G	2.3	0.7	0.0
PY050G	5.7	1.9	2.9
PY070G	6.0	1.6	0.0
PY080G	0.0	0.0	0.0
PY090G	1.3	0.4	1.4
PY100G	14.5	1.8	9.3
PY110G	0.5	0.2	0.7
PY120G	0.2	0.0	0.2
PY130G	0.0	0.5	6.2
PY140G	0.0	0.1	1.3

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 (subsamples 3 and 4*).*

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	**RB250=11**	**RB250=21**	**RB250=23**	**RB250=31**	**RB250=32**	**RB250=33**
Total	19732	18705	82	469	418	58	0
%	100.0	94.8	0.4	2.4	2.1	0.3	0.0

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	18705	15120	0	0	0	3585
%	100.0	80.8	0.0	0.0	0.0	19.2

Wave 2 (subsamples 3 and 4*).*

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	**RB250=11**	**RB250=21**	**RB250=23**	**RB250=31**	**RB250=32**	**RB250=33**
Total	17788	16978	56	403	314	35	2
%	100.0	95.4	0.3	2.3	1.8	0.2	0.0

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

	Total	**RB250=11**	**RB250=21**	**RB250=23**	**RB250=31**	**RB250=32**	**RB250=33**
Total	17545	16760	55	389	307	33	2
%	100.0	95.5	0.3	2.2	1.7	0.2	0.0

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

	Total	**RB250=11**	**RB250=21**	**RB250=23**	**RB250=31**	**RB250=32**	**RB250=33**
Total	242	218	1	14	7	2	0
%	100.0	90.1	0.4	5.8	2.9	0.8	0.0

Distribution of household members by RB260

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

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	16978	13848	0	0	0	3130
%	100.0	81.6	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	16760	13689	0	0	0	3071
%	100.0	81.7	0.0	0.0	0.0	18.3

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	218	159	0	0	0	59
%	100.0	72.9	0.0	0.0	0.0	27.1

Wave 3 (subsamples 3 and 4*).*

Distribution of household members by RB250

Household members 16+ (RB245 = 1 to 3)

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	16529	15611	57	431	358	72	0
%	100.0	94.4	0.3	2.6	2.2	0.4	0.0

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

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	16135	15260	55	411	339	70	0
%	100.0	94.6	0.3	2.5	2.1	0.4	0.0

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

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	394	351	2	20	19	2	0
%	100.0	89.1	0.5	5.1	4.8	0.5	0.0

Distribution of household members by RB260

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

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	15611	12960	0	0	0	2651
%	100.0	83.0	0	0	0	17.0

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

	Total	RB260 = 1	RB260 = 2	RB260 = 3	RB260 = 4	RB260 = 5
Total	15260	12702	0	0	0	2558
%	100.0	83.2	0.0	0.0	0.0	16.8

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	351	258	0	0	0	93
%	100.0	73.5	0.0	0.0	0.0	26.5

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

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

2.5. Imputation procedures

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

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

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

- a target variable implies components of different character (i.e. taking different but rather predictable values, e.g. various social benefits, or dependent on a number of explanatory variables and thus easier to be modelled separately);
- target variables include many components and it is often the case that some of them have the missing items, while others – 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 case of 2005 data, where the use of stochastic regression is not so wide, imputed residuals are generated as pseudo-random numbers of the normal distribution with variance corresponding to the estimated variance of an error term in the model.

In such a case, if we used residuals from the whole range, we could combine a particular theoretical value obtained from the model with the residual which occurs in the whole distribution but is quite improbable in combination with this particular theoretical value. So we could generate values significantly diverging from the real variable distribution. The use of residuals from the restricted range 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 and all 2005 data) or for those who did not gain a particular type of income in the previous year, explanatory variables derived from the current data file are applied.
Wherever the same type of income is found in the data for the previous year, its value is treated as the main explanatory (categorizing) variable, both in the case of variables subjected to regression imputation and the hot-deck method. The current variables can be treated as additional explanatory variables.

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

2.6. Imputed rent

Definitions
Imputed housing costs should be understood as the estimated amount consisting of the actual payments effected by the owners (i.e. charges for water, electricity, gas, other fuels, maintenance and repair of the dwelling and other services relating to the dwelling) as well as imputed rentals that should be ascribed to the owners of flats or houses for their unpaid accommodation resources.
Imputed rentals should be understood as the estimated amount of profit gained in the form of a surplus of the rent over the dwelling maintenance costs, being the landlord's net profit, equal to the amount which could be gained by owners if they wanted to hire their dwellings in the same conditions at market prices.
Actual housing costs should be understood as rentals (charges for water, electricity, gas, other fuels, maintenance and repair of the dwelling and rent) paid by the tenants renting dwellings at market prices.
Actual rentals should be understood as the profit being a surplus of the rent over the dwelling maintenance costs, which is the landlord's net profit gained by the landlords hiring their dwellings at market prices.

Data
The data collected in May-June 2007. The sample consisted of 14286 households. The analysis was made with the use of data from Section 2 of EU-SILC-1G Questionnaire "Dwelling Conditions", supplemented by the information about households' disposable income, household size, number of households in a dwelling and locality.

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

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

The sample of 2007 covered 359 such households, of which 313 (2.19% of all the households participating in the survey) gave the amount of rentals, while 46 households did not, although they declared such payments.

Dependent variable

Following Eurostat's recommendations the variable explained is equal to the monthly rent for a dwelling. For the purposes of this model the value was calculated per 1 m sq. of the usable dwelling area occupied by the household or 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.

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

Explanatory variables

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

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

Form of the function

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

Estimation technique

All the calculations have been made with the use of SPSS package and ENTER or STEPWISE estimation method. The final detailed results are given below.

To start with, functions including all the variables were applied so as to check what maximum R^2 could be obtained, which of the variables described best the rentals level, which seemed irrelevant and which provided estimates difficult to interpret. Then the explanatory variable file was reduced by rejection of non-fitting variables. Estimation was applied by introducing into the function regressors according to their statistical significance (STEPWISE) and thus the correct versions, both as regards statistics and substance, were obtained.

Using the acceptable versions (upon introduction of the data from the auxiliary model and based on the expert model– see the paragraph "Results") imputed rentals were determined for all the households except those paying the actual rentals.

Main characteristics

Sample size	14286
Number of observations on tenants at market prices	359
Number of explanatory variables	21
R2	0.652
R2 corrected	0.627
R	0.808
Imputed rentals (in PLN per household, per year)	
Averages for the total number of households in Poland	4950
of which:	
dwelling owners	3162
other households	1788
Averages for owners' subsample	5085
Averages for subsample of „others",	4728
Actual rentals (in PLN per household, per year)	
Averages for the total number of households in Poland	123
Averages for subsample of tenants paying rentals at market prices,	4488

2.7. Company cars

The information on the private use of the company car is collected in the individual questionnaire. Here belongs the respondent's estimated amount he/she has 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 estimated the amount gained) imputation is applied with the use of hot-deck and regression imputation with simulated residuals methods.

3. COMPARABILITY

3.1. Basic concepts and definitions

The reference population

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

The survey unit was a household and all the household members who had completed 16 years of age by:

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

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

The private household definition

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

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

The household membership

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

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 2005 was 2004,
EU-SILC 2006 was 2005,
EU-SILC 2007 was 2006 year.

Reference period for taxes on income and social insurance contributions

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

The reference period for taxes on wealth

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

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:
2005 year between May 2 and June 17,
2006 year between May 2 and June 19,
2007 year between May 2 and June 19.

Basic information on activity status during the income reference period

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

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

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

3.2. Components of income

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

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.

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

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

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

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

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

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 2005 and EU-SILC 2006 this information collected only refers to the income gained from the use of the company car for private purposes. Since 2007, also collected information about other non-cash employee income, but only at component level (PY021) and not included in the income.

PY030 – Employer's social insurance contributions

Variables was collected since EU-SILC 2006 only at component level.

PY050 - Cash profits or losses from self-employment (including royalties)

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

PY080 and PY130 - Regular inter-household cash transfer received

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) in EU-SILC2008.

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), because they are paid by the employer and cannot be detached from the income from hired employment. Therefore, they are accounted for in the income from hired employment.

PY200 – Gross monthly earnings for employees

Variables were collected since EU-SILC 2006.

All the other variables not listed above

Dwelling conditions and material deprivation items

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

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

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

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

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

Additionally, for the variables from HS010 to HS050 no information was given that paying through borrowing meant that household was not in arrears.
In 2008 changes were introduced to bridge these differences.

There were no other major divergences from common definitions.

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

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

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

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

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

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

3.3. Tracing rules

Standard EU-SILC tracing rules are applied.

4. COHERENCE

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

4.1. Comparison of EU-SILC and HBS results

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

When comparing these two sources we must take into account the discrepancies. The differences are to 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 considerable distortions. In EU-SILC the reference period is a calendar year preceding the survey;
- Different types of income are taken into account i.e. in HBS the information is collected both about the income in cash and in kind, while in EU-SILC – only about the income in cash (with a few exceptions), which may be important for the income from farming and social benefits other than retirement pay and pension. Moreover, EU-SILC does not take into account the so called lump sums which is the case in HBS;
- Different way of data collection – in HBS the respondents make records in the so called diary. They have to determine the data sources themselves and do not have them listed in the diary. This may cause omissions. In EU-SILC each respondent is asked detailed questions. In EU-SILC all the income missing data are imputed, while there is no imputation in HBS;
- Different way 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 socioeconomic group and household size. The household survey results are usually released by the CSO in the breakdown by socioeconomic group and household size.
The main criterion for socio-economic group classification is the prevailing source of income.
In tables below only weighted data are presented.

Tab. 1. Structure of population by age

Specification	EU-SILC 2007	HBS 2007
	in %	
Total	100.0	100.0
0-14	16.0	18.1
15-24	15.5	16.1
25-54	44.0	41.9
55-64	11.1	11.6
65+	13.4	12.3

Tab. 2. Structure of population by level of education

Specification	EU-SILC 2007	HBS 2007
	in %	
Total	100.0	1000
No school education	2.1	0.9
Completed primary	18.5	19.1
Lower secondary	5.1	6.6
Elementary vocational	26.8	26.6
Secondary	33.6	33.9
Higher	13.8	12.9

Tab. 3. Structure of households and persons in households by socio-economic group

Households	Households		Persons in households	
	EU-SILC 2007	HBS 2006	EU-SILC 2007	HBS 2006
Total	13281985	13332332	37719639	37703168
Total = 100				
Employees	49.5	45.1	59.3	53.6
Farmers	2.6	4.6	3.6	7.0
Self-employed	4.8	6.1	5.6	7.1
Retirees	27.6	27.9	19.4	19.5
Pensioners	9.3	10.3	6.4	7.4
Maintained from non-earned sources	6.1	6.0	5.9	5.4

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

Households	Disposable income		Income from hired work	
	EU-SILC 2007	HBS 2006	EU-SILC 2007	HBS 2006
Total	16166	14767	9649	7482
Employees	18140	15455	15195	12677
Farmers	10550	13891	1042	1367
Self-employed	17828	20416	2700	2805
Retirees	14670	14278	1466	1423
Pensioners	10888	10478	1175	1022
Maintained from non-earned sources	8723	9272	1665	736

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

Households	Disposable income		Income from hired work	
	EU-SILC 2007	HBS 2006	EU-SILC 2007	HBS 2006
Total	16166	14769	9649	7484
1-person	15412	14686	4713	4103
2-persons	18725	17225	7980	6145
3-persons	18561	16511	12689	9955
4-persons	16235	14862	11714	9393
5-persons	13733	12716	8889	6735
6-persons and more	12456	11187	7166	4684

Tab. 6. Households provided with selected durables

Specification	EU-SILC 2007	HBS 2007
	in %	
Fixed telephone	71.6	67.9
Mobile telephone	75.5	79.3
Television set	97.1	98.5
Computer	48.7	50.1
Printer	35.7	33.6
Internet connection	34.8	36.6
Microwave oven	37.9	42.4
Dishwasher	9.3	7.4
Refrigerator	97.6	98.9
Washing machine	96.6	97.1
Passenger car	53.6	52.5

4.2. Comparison of 2006 results of SNA and EU-SILC 2007 (data for 2006) for Poland

The comparison covered the disposable income and its main components: income from hired employment, self-employment (in and outside farming) and social benefits.
It was confirmed that in EU-SILC 2006 the disposable income was 58% of the respective category in SNA. This has been brought about by the following reasons:

1. The household sector in SNA includes collective households which are not covered by EU-SILC.
2. Each of the systems applies a different method of measuring income from self-employment.
3. The estimates of primary and secondary distribution of income in SNA, used as a basis for the calculation of disposable income refer to some items not covered by EU-SILC 2007 or not taken into account when calculating EU-SILC results. The most important of these items is imputed rents.

In SNA income from self-employment is calculated as the so called operation surplus which is a balance between global production and current production inputs, i.e. intermediate consumption and hired employees' remunerations. This amount is reduced by taxes and increased by subsidies. The operation surplus calculated in this way is allocated to households' consumption needs as well as dwelling- and business-related investment. In the Polish EU-SILC the question about income from self-employment refers only to the amount spent on household's consumption and its dwelling-related investment. Besides, SNA takes into account consumption from own production, which was not covered by EU-SILC 2007 for farming. These differences are responsible for the fact that income from self-employment in EU-SILC 2007 amounted only to 26% of the operation surplus in SNA (after section K deduction).
The income from self-employment in EU-SILC 2006 is equal to 99% of the respective category in SNA, while social benefits – 93%, respectively, which seems to be a satisfactory outcome.
As compared with EU-SILC 2006, there was a higher convergence between EU-SILC 2007 data and SNA: for disposable income by 1 percentage point, for incomes from hired employment – by 3 p.p. and for social benefits – by 1 p.p. This marks further improvement of the quality of data. Some improvement of the data convergence with SNA was already noticed for 2005. The only decrease in convergence of data between SNA and EU-SILC was noted for the income from self-employment which dropped in EU-SILC 2007 by 1 p.p. as compared with EU-SILC 2006.

Comparison of 2006 results of SNA and EU-SILC 2007 for Poland

Category in SNA	*Variables in EU-SILC 2007*	*Category description in EU-SILC 2007*	*SNA in mln PLN*	*EU-SILC in mln PLN*	*SNA = 100%*	*SNA = 100% EU-SILC 2006*
Gross disposable income (net)	HY020	Total disposable household income (net)	683 483	398 939	58	57
Wages, salaries and other income connected with hired work (gross)	PY010G	Employee cash or near cash income (gross)	325 030	323 956	99	96
Gross operating surplus (gross) with the exception of section K	PY050G	Self-employment income (gross) - value allocated to household's consumption and dwelling-related investment	201 601	53 219	26	27
Social security benefits and social assistance benefits (gross)	PY90G + PY100G + PY110G + PY120G + PY130G + PY140G + HY050G + HY060G + HY070G	Social benefits (gross)	161 336	149 258	93	92

Remarks:

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 refer to including of depreciation of fixed assets.

2. Data for gross operating surplus in SNA has been taken into consideration with the exception of section K what allows for better comparability with EU-SILC data on self-employment income (PY050G). The data for section K includes mainly imputed rents, not included in the results of EU-SILC 2007 (data for 2006), and market income from renting of real estate included in EU-SILC as the variable HY040G.