Final Quality Report SILC2010- BELGIUM

Longitudinal report (2007-2010)

0. Introduction

This report contains a description of the accuracy, precision and comparability of the Belgian SILC2007 to SILC2010-surveydata. It is structured following the guidelines in the commission regulation (EC) no. 28/2004. This results in three chapters:

1. Indicators
2. Accuracy
3. Comparability
4. Coherence

1. Indicators

For the common longitudinal EU indicators based on the longitudinal sample of EU-SILC we refer the readers to the EUROSTAT website where these indicators are available in a dynamic way.

2. Accuracy

For second and following waves of the longitudinal component the following information has to be provided:

2.1 Sampling Design

2.1.1 Type of sampling (stratified, multi-stage, clustered)

The Belgian EU-SILC 2010 survey follows a stratified 2-stage sampling.

2.1.2 Sampling units (one stage, two stages)

Primary units:

The *Primary Sampling Units* are the municipalities (or part thereof in the larger ones); in each of the 11 strata, they were drawn PPS, i.e. with repetitions allowed (for instance, Schaerbeek was drawn 6 times). In total, 275 draws were made in 2004, once forever (for the whole duration of EU-SILC).

Secondary units:

The *Final Sampling Units* are the (private) households.

Recall that, in 2004, 40 households had been selected in each PSU, numbered 1 to 40.

The first 10 (whether or not they responded irrelevant) vanished from the panel in 2005, the other 30 (including possible split-offs) were followed according to the tracing rules.

Hence, the (cross-sectional) sample of SILC 2010 consists of

- "old" households (drawn between 2007 and 2009)

and

- "new" households (drawn in 2010, staying until 2012).

In fact, it is only the selection of the new households that gave us some degree of freedom (see in particular 2.1.4).

In the D-file, three variables have been added:

- ✓ DB061 is the identification of the primary units (concatenation of 5 digits for the municipalities and one letter).
- ✓ DB063 is the 'multiplicity order', the number of times each PSU was drawn in the sample.
- ✓ DB071 is the order of selection of the new households within each letter.

2.1.3 Stratification and sub-stratification criteria

The stratification criterion is the region (NUTS2 level). The 11 strata are the 10 provinces of Belgium and the Brussels Capital Region.

2.1.4 Sample size and allocation criteria

In 2010 we managed to keep the number of responding households above 6000, drawing 17 new hh in each PSU**.**

Table 1: sample size and achieved response by NUTS2-units

NUTS2	Name	Old (or strange) hh	New hh	Total hh	Accepted hh (DB135=1)
		53	17	70	0
BE10	Brussels	720	836	1556	816
BE21	Antwerpen	712	731	1443	791
BE22	Limburg	360	281	641	447
BE23	Oost-Vlaanderen	631	550	1181	780
BE24	Vlaams-Brabant	474	442	916	554
BE25	West-Vlaanderen	533	360	893	702
BE31	Brabant Wallon	136	136	272	176
BE32	Hainaut	751	646	1397	889
BE33	Liège	438	413	851	558
BE34	Luxembourg	164	102	266	196
BE35	Namur	182	161	343	223
Total	**Belgium**	5154	4675	9829	6132

2.1.5 Sample selection schemes

Systematic sampling of secondary units (new households) in each primary unit selected, the households have been ordered according to the age of the reference person.

2.1.6 Sample distribution over time

2.1.7 Renewal of sample: Rotational groups

See above.

2.1.8 Weightings

Recall that, for the first year of the panel (=SILC 2004 in Belgium), the computation of weights involved three stages (described in 134-04)

(a) initial weights
(b) weights corrected for nonresponse
(c) final (calibrated) weights.

For 2010, a distinction has to be made between :

-"old" households i.e. households that contain at least one sample person who took part in 2009, and had to be surveyed again in 2010 according to the rotation and tracing rules (excluding the outgoing fourth) (household composition may have changed, whence quotations marks);

-"new" households i.e. households that were drawn for the first time in 2010, among those households not containing any sample person already drawn before (quotations marks superfluous).

This distinction pertains to initial weights and nonresponse correction:

Since the "old" households are selected indirectly from the 2007, 2008 or 2009 samples, and household composition may have changed, some kind of "weight sharing" must be applied to determine the (2010) initial weights, or rather base weights. On the other hand, "new" households have their own inclusion probability, whose inverse gives the initial weights;

For the "old" households, (2010) nonresponse=attrition can be linked with (2009) SILC information. For the "new" households, all we can rely upon to explain initial nonresponse is auxiliary information from the Population Register (household size, urban/rural character) and the Financial Statistics (median fiscal income by municipality).

On the other hand,

Calibration can be done together for "old" and "new" households. With respect to our 2004 model, we decided in 2005 to relax the constraints (basically, calibrating at NUTS1-level instead of NUTS2), in order to decrease the standard deviation of weights.

This introduces the following sections:

2.1.8.1 Initial weights for the new households
2.1.8.2 Nonresponse correction for the new households
2.1.8.3 Attrition for the old households
2.1.8.4 Weight sharing
2.1.8.5 Calibration
2.1.8.6 Final longitudinal weights

2.1.8.7 Final cross-sectional weights

2.1.8.1 Initial weights for the new households

Belgium chose to draw the Primary Sampling Units (= municipalities or parts thereof) "forever", and to rotate the Secondary Sampling Units (=households) within the selected PSU's.

The 2004 PPS two-stage sampling design was self-weighting within each stratum (*h*: *x* denoting any households in municipality *X*), we had (in 2004)

P (*x* drawn) = P(*x* drawn|*X* drawn) . P(*X* drawn) = n_h/N_X . N_X/N_h . g_h = n_h/N_H . g_h, where

n_h denotes the number of households to be drawn in the (selected) PSU (viz. 40)

N_X the number of households in the PSU (in 2004)

N_h the number of households in the stratum (in 2004)

g_h the number of PSU's drawn in the stratum.

(This is an oversimplification, since PSU are drawn with repetition; the selection probability for a PSU should be replaced by the expectation of selection multiplicity, and the term 40 by a multiple depending on the selection multiplicity…but the idea is the same).

In 2010, the picture has become

P (*x* drawn) = P(*x* drawn|*X* drawn) . P(*X* drawn) = m_h/M_X . N_X/N_h . g_h, where

- m_h is the number of households to be drawn in the (selected) PSU (depending on *h*)
- M_X is the number of households in the PSU (in 2010)

The factor N_X/M_X indicates the increase-decrease in inclusion probabilities in PSU *X* (still assuming *X* has been drawn) between 2010 and 2004.

Now it would seem logical to replace N_X by a smaller number, to account for the households[1] already drawn in 2004, 2005, 2006, 2007, 2008 or 2009 whence immunized from being drawn again in 2010.

However, the following argument shows that (assuming momentarily that *X* has been drawn and that the population figures N_X and M_X remain stable) matters are not so easy:

P(*x* drawn in 2010) =

(P(*x* drawn in 2010|*x* drawn before) . P(*x* drawn before)) +

(P(drawn in 2010|*x* not drawn before) . P(*x* not drawn before),

the first term vanishes and the second equals $n_h/(M_X-b)$. $(N_X-b)/N_h$, where *b* denotes the number of hh already drawn; since both fraction terms are much larger than *b* (at least 900 in all selected PSU's), the ratio $(N_X-b)/(M_X-b)$ is (close to 1, and) very close to N_X/M_X. Since the term *b* is an approximation anyway, we chose to stick to m_h/M_X . N_X/N_h. g_h as inclusion probabilities, and its inverse for initial weights **INIwei=DB080**. Note that, with this concept of DB080, the "new" hh correspond to the total Belgian population (some 4,5 millions private hh); before calibrating, theses weights will be scaled down "to make room" for the old hh; recovering the strange hh means that the sum of

[1] Perhaps a bit less (households that vanished already subtracted) or a bit more (split households, both components of which stayed in PSU, should be subtracted twice)

the pre-calibration weights will be slightly larger than 4,5 millions (average of g-weights slightly less than 1).

2.1.8.2 Nonresponse correction for the new households

Following Eurostat's suggestion (see Document 065, WEIGHTING II. WEIGHTING FOR THE FIRST YEAR OF EACH SUB-SAMPLE), we replaced the homogeneous response groups (based on household size crossed with urbanity) ratio by a multiple regression model (based on the same dummy variables). By "responding", we mean only those households whose results were accepted (DB135=1). For technical reasons, we used linear regression instead of logistic; since the (predicted) response turned out to be close to 50% for all categories, this is harmless.

The file was split by NUTS1 and the following variables were used

- Everywhere: Household size, recoded into the four values "one", "two", "three" and "four or more" (so three dummies);
- Out of Brussels: DB100 = urbanity ;
- In Brussels = BE10: median fiscal income of municipality.

The regression produced a new variable "expresp", allowing us to define

NRwei = INIwei/expresp

2.1.8.3 Attrition for the old households

Before "sharing" the 2009 weights, a correction for attrition should be introduced. This year, we elected to perform this correction at the level of individuals, since a 2009 sample person either stays in the panel or leaves it (rotated out, left population, noncontact, refusal or inability to respond, while the structure of a household can change. Note that all household characteristics (e.g. HH020) can be distributed to the members.

We separated the "Children" (for which only basic personal information from the R-file and the distributed H-file is available) from the "Adults" (present in the 2009 P-file as well), i.e. those persons born in 1990 or before.

In the children's model, the following predictors (all, except the last, from the 2005 file – although this does not matter much for group A) were used, grouped by type

A. <u>individual demographic information</u>: age[2] from RB080, sex = RB090, country of birth (= pb210 for adults, but available for children too in our Belgian files);

[2] Let us start with a picture (Z in function of age class, "1" denoting the range 0-4, …, "17" the range "80-84", "18' corresponds to '85 or older", age computed here as 2007-rb080)

The highest 2 scores are depicted in white, the lowest 2 in dark blue. We distinguish two local maxima (one among children 5-9, the other one in the area of "old but not too old") and two local minima (one among "young adults" and one for "very old".

B. housing information: dwelling type = HH010 and tenure = HH020
C. household type: a limited number of dummies, as there is at least one dependent child;
D. monetary indicators: we refrained from taking the equivalised income (outliers), but took a transform of it, as well as the dummy “poor or not” and the subjective ability to make ends meet = HS120
E. sampling and rotation: number of years in panel (from DB075) and urbanisation (=DB100)
F. one variable (paradata) related to fieldwork (computed from HB040 and HB050)

For the adults, the same predictors were used, and moreover

G. variables from the P-file (related to education level and health);

were integrated.

We used linear regression; (with some truncation, when the estimated response propensity turned out to be larger than one).

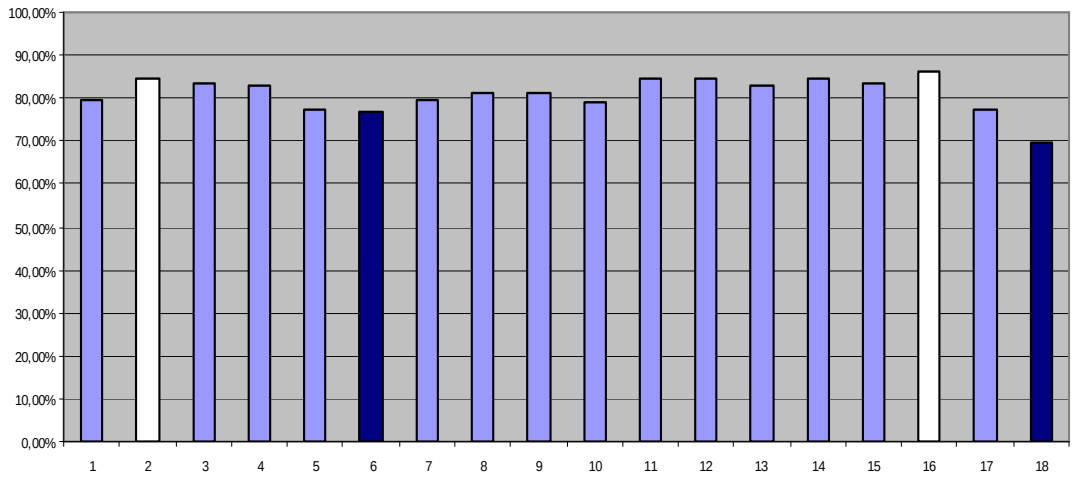
2.1.8.4 Weight sharing

We followed Eurostat’s recommendation "EU-SILC weighting procedures: an outline" and shared the calibrated 2009 weights, after correcting for attrition (instead of the initial weights, see Lavallée).

This can be illustrated by an imaginary example, dealing simultaneously with fusions (persons *A&B* in same 2009 hh, *C* in another 2009 hh, so “fusion” in the sense of DB110 occurs), new members (a baby like *E* or already in population like *D*); we focus on the 2010 hh, what happened to those who co-resided with *A* and *B* or with *C* in 2009 (left or split) is irrelevant!

Note that

- RB050 = weight 2009: same for *A* & *B*, vacuous for *D* and *E*
- Newi: in general a bit larger than RB050; *A*’s differs from *B*’s (attrition correction at individual level)
- Somwe = 950+1000+850 involves only *A*, *B* and *C*



- *Weiind:* = ¼ * **somwe** (*A B C D* : four contribute to the denominator)[3]

Person in 2010 hh	*A*	*B*	*C*	*D*	*E*
RB110 (2010)	1	1	2	3	4
RB050 (weight 2009)	800	800	600	---	---
Newi = Weight 2009 (after attrition correction)	950	1000	850	---	---
Somwe (sum Newi over 2010 hh)	2800	2800	2800	2800	2800
Weiind	700	700	700	700	700

Weiind will be injected as "initial" weight in the final calibration job.

2.1.8.5 Calibration

We first put the pieces together: weiind is defined as

- (new = started in 2010) : initial weight, corrected for initial nonresponse, scaled, see 2.1.8.1)
- (old = took part in 2009) 2007 weight, corrected for attrition and weight sharing if necessary, see 2.1.8.4)
- (strange = did not take part in 2009 but before) initial weight, no correction)

In terms of persons, the weiind statistics were

Type	# ind	Mean of weiind
NEW	4702	596,78
OLD	9385	804,25
BACK	667	561,72
Total	14754	727,17

Recall that 11 *sampling* strata were used (provinces= NUTS2); we use 3 *extrapolation* strata (the 3 NUTS1 regions BRUssels=BE1, VLAanderen=BE2 and WALlonia=BE3)

Calibration model

- ***VLA, WAL:***
  - SIZE4+(AGE8XSEX2)+PROV5 ;
  - →20 individual[4] + 4 household constraints.

[3] Do we abide by the Eurostat rules (starting from base weights, it is unclear whether "their" attrition correction precedes or follows weight sharing) ?

There remain some additional categories of persons to be considered:

-Children born to sample women. They receive the weight of the mother (this assumes that the baby belongs toto his/her mother's hh)

-Persons moving into sample households from outside the survey population. They receive the average of base weights of existing household members (vacuous here, as RB110 enables us to identify the newborns, but not the immigrants or the –few- persons moving from a collective to a private hh)

-Persons moving into sample households from other non-sample households in the population – these are "co-residents" and are given zero base weight.

[4] Five provinces and 16 age*sex categories, but sum over provinces = sum over age*sex

- ***BRU:***
  - SIZE4+(AGE8XSEX2) ;
  - →16 individual + 4 household constraints;
  - Prov = province where interviewed (differs from DB040 in some cases);
  - Individual constraints : 27=16+11 (age*sex + prov; note that each province belongs to one single region (extrapolation stratum), for the other two regions, the total is set to 0 and the condition is vacuous);
  - Household constraints : 4 (size: "1", "2", "3 or "4 & more");
  - Calibration type (after some trials and errors…): truncated.

2.1.8.6 Final longitudinal weights

Combination of steps above…

2.1.8.7 Final cross-sectional weights

Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Final weights	6132	97.22	4523.09	763.24	359.80

2.1.9 Substitutions

No substitution was applied in our survey.

2.2 Sampling errors

Income components	Mean	Number of observations before imputation	Number of observation after imputation	Standard error
HY010	43974.71	2010	6129	577.19
HY020	33118.89	1041	6129	386.58
HY022	29535.72	958	5998	403.04
HY023	23940.96	874	5855	427.71
Net income components at household level				
HY030N				
HY040N				
HY090N	1327.49	1271	4133	102.25
HY050N	3738.67	2080	2137	92.29
HY060N				
HY070N				

HY080N				
HY100N				
HY110N				
HY120N				
HY130N				
HY140N				
HY145N				
Gross income components at household level				
HY030G				
HY040G	8410.40	518	549	451.06
HY090G	1327.49	1271	4133	102.25
HY050G	3776.77	2032	2137	93.12
HY060G	3623,437	128	129	267.44
HY070G	1815.40	42	46	574.70
HY080G	3652.04	450	466	257.86
HY100G	3221.78	1905	1974	76.32
HY110G	2840.77	12	14	890.22
HY120G				
HY130G	3345.48	578	597	195.89
HY140G	11406.45	1668	5671	218.93
net income components at personal level				
PY010N	20624.37	5449	5635	215.35
PY020N	1439.96	2196	2601	28.12
PY021N	1921.87			
PY035N				
PY050N	19978.20	552	714	837.36
PY070N				
PY080N	87976.47	28	28	17227.61
PY090N	7680.48	1330	1441	193.66
PY100N	15112.38	2092	2423	354.53
PY110N	13193.31	96	103	483.76
PY120N	5912.94	161	185	459.54
PY130N	9977.70	454	471	345.39
PY140N	911.625	608	658	108.61
gross income components at personal level				
PY010G	31062.88	5233	5635	377.57
PY020G	1564.71	1738	2601	34.23
PY030G				
PY035G				
PY050G	25958.48	270	714	1180.65
PY070G				

PY080G	87976.47	28	28	17227.61
PY090G	8162.80	793	1441	214.99
PY100G	17251.49	1258	2423	384.26
PY110G	14191.18	66	103	591.27
PY120G	6231.42	86	185	481.45
PY130G	10335.71	269	471	379.74
PY140G	911.625	608	658	108.61
PY200G				

Equivalised disposal income	Mean	Number of observations before imputation	Number of observation after imputation	Standard error
Subclasses by household size				
1 household member	18619.03	1395	1761	315.35
2 household members	22367.54	3526	4158	397.66
3 household members	22720.52	2532	2925	408.56
4 and more	20986.92	4870	5906	364.42
Population by age group				
<25	20130.18	3668	4502	306.83
25 to 34	22831.17	1553	1847	373.98
35 to 44	22487.12	1705	2107	311.63
45 to 54	23344.45	1763	2077	365.95
55 to 64	23037.23	1619	1887	525.47
65+	18323.59	2015	2330	432.23
Population by sex				
Male	21802.68	6077	7259	242.89
Female	20917.62	6246	7491	219.33

2.3 Non-sampling errors

2.3.2 Measurement and processing errors

- ***Mismatch in time between household composition and household income (see also §3.1)***

A number of inconsistencies result from a mismatch between the composition of the household at the moment of the interview (between September and December of year x) and the income of the previous year (year x-1).

This mismatch can bias the measurement of poverty status in several ways. For example:

- ✓ Persons who were full-time students in year x-1 (and depending on their parents), but were employed at the time of the interview (and living independently in a one person household for example)

will report an income equal to 0 in year x-1 and will be wrongly classified as a poor household.

Other examples can also occur for persons where the household composition changed:

- ✓ For a housewife who was married in year x-1, but divorced and is working at the time of the survey there will also be a mismatch
- ✓ For a household which received family allowances for a student in year x-1, but where the student is no longer part of the household in year x there will also be a mismatch
- ✓ For a household with a person working in year x-1, but retired at the moment of the survey (in year x) a mismatch will also occur. Take notice of the fact that, as the examples show the bias can go in both directions: under and over reporting of income. In each one of the examples, the choice to situate the income reference period in the past is the cause, however.

- ***Error in the routing wave 2007***

There was one error in the routing. In the household questionnaire, in the part concerning childcare, the selection was made on the base of actual age instead of age in the income reference period. So we missed information for some children born in 1994.

- ***Error in the routing wave 2008***

See wave 2007.

- ***Processing errors***

Belgium used the CAPI–method to interview the persons. The questionnaire was programmed in Blaise. So processing errors due to data entry (from a written to an electronic format) were reduced to a minimum.

Statistics Belgium programmes several data entry and coding controls in the Blaise program. Those were identical for both waves.

Next to these controls, some warnings were implemented **in 2005** in order to ask the interviewer to verify the introduced data in the case of abnormally high or low amounts. A warning is a simple text box with a message such as ‘This amount is very low, are you sure the amount is right?’ or ‘This amount is very high, are you sure the amount is right?’. The interviewer has then to confirm the value or to change it in case of error.

Household questionnaire	
H16	If lower than 500 or higher than 1000000
H22 (monthly)	If lower than 20 or higher than 2000
H22 (half-yearly)	If lower than 100 or higher than 10000
H22 (yearly)	If lower than 200 or higher than 20000
H23 (monthly)	If lower than 20 or higher than 2000

H23 (half-yearly)	If lower than 100 or higher than 10000
H23 (yearly)	If lower than 200 or higher than 20000
H26	If lower than 25 or higher than 5000
H33	If lower than 50 or higher than 10000
H34, H37, H41	If lower than 100 or higher than 5000
H43, H77, H84	If lower than 25 or higher than 1000
H66	If lower than 100 or higher than 25000
H71B	If lower than 25 or higher than 750
H79, H86	If lower than 300 or higher than 12000
H93	If lower than 100 or higher than 1500
Individual questionnaire	
I25, I27, I47, I50, I90, I91	If lower than 500 or higher than 5500
I53, I86, I93, I94	If lower than 6000 or higher than 66000
I58	If higher than 1200
I98B, I98C, I115B, I115C	If higher than 1350
I99, I102B, I102C	If higher than 5400

Some warnings concern other values than amounts. It's the case for H17 when the value is higher than 30 years ('A period of 30 years is really exceptional, are you sure it is right?') and for H18 when the interest equals 0 or is higher than 15.

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

- number of households for which an interview is accepted in the **longitudinal database 2007-2010**:

	2007	2008	2009	2010
Households	2026	3188	4621	4064

- number of persons 16 years or older, number of sample persons and number of co-residents, members of households for which an interview is accepted in the **longitudinal database 2007-2010** and who completed a personal interview:

	2007	2008	2009	2010
Persons 16 y and more	3855	6037	8757	7738
Sample persons	3855	5962	8547	7315
Co-residents with interview	0	75	210	423

2.3.3.2. Unit non-response

A. Response rate for households

		SAMPLE OUTCOME IN WAVE 7											
		DB130=11											Total
		DB135=1 (A)	DB135=2 (B)	DB120=22 (C)	DB130=22 (D)	DB130=23 (E)	DB130 =24 (F)	DB130= 21 (G)	DB120= 21 (H)	NC (I)	DB110 =10 (J)	DB120= 23 (K)	(T)
		6016	24	18	124	61	1185	1955	86	22		173	9664
SAMPLE OUTCOME IN WAVE 6													
DB130=11	DB135=1	3892	10	1	49	13	273	354	13	18	0	6	4629
	DB135=2	0	0	0	0	0	1	0	0	1	0	0	2
DB120=21 to 23 or DB130=21 to 24		104	0	0	9	5	65	55	1	3	0	0	242
TOTAL		3996	10	1	58	18	339	409	14	22	0	6	4873
NEW HOUSEHOLDS IN WAVE 7													
DB110=8		68	1	0	1	0	27	17	1	0	.	1	116
DB110=9		1952	13	17	65	43	819	1529	71	0	.	166	4675

- Wave response rate:
  - Wave response rate: $\frac{6016}{9664-173}=63\%$
  - Refusal rate : $\frac{1955}{9664-173}=21\%$
  - No-contacted and others: $\frac{1459}{9664-173}=15\%$
- Longitudinal follow-up rate:
  - Longitudinal follow-up rate: $\frac{4238}{4238+387}=92\%$
- Follow-up ratio:
  - Follow-up ratio: $\frac{4238+2062}{4238+387}=1.36$
- Achieved sample size ratio:
  - Achieved sample size ratio: $\frac{6016}{4629}=1.3$

B. Personal interview response rates : table and rates under construction

		PERSONAL INTERVIEW OUTCOME IN WAVE t										
		rb250 = 11	Not completed because of									
			rb250 = 21	rb250 = 22	rb250 = 23	rb250 = 31	rb250 = 32	rb250 = 33	HHnc	Pn	Pl	Total
RESIDENCE AS OF THIS WAVE t SAMPLE PERSONS (RB100=1 and RB245 = 1-3) FROM THE SAMPLE FORWARDED FROM LAST WAVE (t-1)												
[01] rb110 in (1,2)												7285
[02] rb110 = 6												
[03] rb110_f = -1												
[04] rb120 = 2												
[05] rb120 = 3												
[06] rb120 = 4												
[07] db135 in (2, -1) or db110=7 or db120 in (21-23, -1) or db130 in (21-24, -1)												
[08] db110 in (3-6)												
NEW SAMPLE PERSONS												
[09] reached age 16												
[10] sample additions												
NON SAMPLE PERSONS 16+												
[11] this wave	From W1											
	No in wave 1											
[12] earlier wave	From W1											
	No in wave 1											
SAMPLE PERSONS FROM SAMPLE NOT FORWARDED FROM LAST WAVE t-1												
[13]	No in wave 1											
SUM OF ROWS												
[1], [3], [6], [7], [9], [10]												
[1], [3], [6], [7], [9], [10], [13]												
[1], [3], [6], [7], [9], [10], [11]												

- Wave response rate
    - Wave response rate of sample persons :

    - Wave response rate of non sample persons:

- Longitudinal follow-up rate:

Rate (RB250=21)
Rate (RB250=23)
Rate (RB250=31)
Rate (RB250=32)
Rate (RB250=33)

- Achieved sample size ratio for sample persons

- Achieved sample size ratio for sample and co-residents

- Response rate for non-sample persons

2.3.3.3 Distribution of households by household status, by record of contact at address, by household questionnaire result, by household acceptance

Household status

DB110=

	Total	1	2	3	4	5	6	7	8	9	10
Total	4657	4657	194	10	55	5	2	0	116		4
%	100	93.3%	4%	0.2%	0.1%	0.1%	0.0%	0.0%	2.3%	0.0%	0.0%

Record of contact at address

DB120=

	Total	11	21	22	23
Total (DB110=2,8,10)	310	287	15	1	7
%	100	92.6%	4.8%	0.3%	2.3%

Household questionnaire result

DB130=

	Total	11	21	22	23	24
Total (DB120=11 or DB110=1)	4944	4075	426	59	18	366
%	100	82.4%	8.6%	1.2%	0.4%	7.4%

Household interview acceptance

DB135=

	Total	1	2
Total(DB130=11)	4075	4064	11
%	100	99.7%	0.3%

2.3.3.4 Distribution of persons for membership status (RB110)

	Total	Current HH member				No current HH member		
		RB110 =1	RB110 =2	RB110 =3	RB110 =4	RB110 =5	RB110 =6	RB11 0=7
Total	10013	9424	93	161	88	236	11	0
%	100	94.12%	1.0%	1.6%	1.%	2.4%	0.11	0.0%

Distribution of persons moving out by variable RB120

	Total	RB110=5				
		RB120=1		RB120=2	RB120=3	RB120=4
		This person is a current HH member	This person is not a current HH member			
Total	236	0	208	9	17	2
%	100	0.0%	88.1%	3.8%	7.2%	0.8%

2.3.3.5 Item non-response

In the following table an overview of the item non-response for all income variables is presented. The percentage households having received an amount, the percentage of households with missing values and the percentage of households with partial information is calculated.

These percentages are calculated as follows:

- % of households having received an amount : number of households (or persons) who have received something (yes to a filter) / total
- % of households with missing values : number of households (or persons) who said that they have received something but did not give any amount (no partial information) / number of households (or persons) who have received something (yes to a filter)
- % of households with partial information: number of households (or persons) who said that they have received something but gave partial information (amounts were not given for all components) / number of households (or persons) who have received something (yes to a filter)

Overview of the non-response for the income variables - % households having received an amount, % of households with missing values and % of households with partial information.

Item non-response	% of households having received an amount	% of households with missing values	% of households with partial information
Total gross household income (HY010)	99.98	12.31	55.33
Total disposable household income (HY020)	99.98	9.87	73.25
Total disposable household income before social transfers except old-age and survivor's benefits (HY022)	97.83	7.82	76.21
Total disposable household income before social transfers including old-age and survivor's benefit (HY023)	95.40	2.40	82.67
Net income components at household level			
Family related allowances (HY050N)	35.51	1.32	1.18
Interests, dividends, etc. (HY090N)	68.13	69.19	.
Gross income components at household level			

Income from rental of a property or land (HY040G)	9.25	3.99	.
Family related allowances (HY050G)	35.51	1.32	3.26
Social exclusion not elsewhere classified (HY060G)	1.92	1.28	.
Housing allowance (HY070G)	0.57	4.35	.
Regular inter-household cash transfer received (HY080G)	7.46	2.97	.
Interest repayments on mortgage (HY100G)	32.53	3.25	0.15
Income received by people aged < 16 (HY110G)	0.25	20.00	.
Regular inter-household cash transfer paid (HY130G)	9.50	2.59	0.26
Tax on income and social contributions (HY140G)	92.89	19.05	50.70
Net income components at personal level			
Employee cash or near cash income (PY010N)	48.45	1.17	1.39
Cash benefits or losses from self-employment (PY050N)	5.82	20.22	.
Pension from individual private plans (PY080N)	0.19	.	.
Unemployment benefits (PY090N)	12.72	7.72	.
Old age benefits (PY100N)	21.06	12.02	0.43
Survivor' benefits (PY110N)	1.01	6.41	.
Sickness benefits (PY120N)	1.52	11.86	.
Disability benefits (PY130N)	3.93	2.96	.
Gross income components at personal level			
Employee cash or near cash income (PY010G)	48.45	1.31	5.39

Non cash employee income (PY020G)	22.37	15.02	17.45
Non cash employee income: company car (PY021G)	4.68	29.28	.
Cash benefits or losses from self-employment (PY050G)	5.82	56.67	.
Pension from individual private plans (PY080G)	0.19	.	.
Unemployment benefits (PY090G)	12.72	44.92	.
Old age benefits (PY100G)	21.06	46.26	1.47
Survivor' benefits (PY110G)	1.01	30.77	.
Sickness benefits (PY120G)	1.52	48.31	.
Disability benefits (PY130G)	3.93	42.43	.
Education-related allowances (PY140G)	5.18	4.24	.

2.4 Mode of data collection

Distribution of household members aged 16 and over by RB250

(Household members RB245=1)

	total	**RB250=11**	**RB250=14**	**RB250=21**	**RB250=23**	**RB250=31**	**RB250=32**	**RB250=33**
Total	**11605**	11377		25	22	72		109
	100%	98.04%	0%	0.22%	0.19%	0.62%	0%	0.94%
db075 = 1 (wave 2009)	2851	2820	0	5	5	16	0	5
db075 = 2 (wave 2010)	3773	3639	0	8	7	27	0	92
db075 = 3 (wave 2007)	2668	2632	0	8	5	13	0	10
db075 = 4 (wave	2313	2286	0	4	5	16	0	2

2008)								

Distribution of household members aged 16 and over by RB260

(Household members RB250=11)

	total	**2**	**5**	**Missing**
Total	**11377**	10399	978	
	100%			

db075 = 1 (wave 2009)	2820	2576	244	
db075 = 2 (wave 2010)	3639	3346	293	
db075 = 3 (wave 2007)	2632	2392	240	
db075 = 4 (wave 2008)	2286	2085	201	

2.5 Imputation procedure

2.5.0 Preceding important remark

The calendar question (i40 in the questionnaire) was presented to every respondent rather then only to those who indicated that had been a change in their social-economic position. It enabled us to assess and check much thoroughly the link between the social-economic position and the income variables. Notably for the self-employed this resulted in a substantive number of cases (being identified as being self-employed) who would be otherwise not identified as being self-employed. These cases mainly concern people in jobs ‘somewhere on the bridge’ between being self-employed and employee but who nevertheless indicated in the calendar that they were self-employed.

2.5.1 Overall strategy: Emphasis on internal information and integration of outlier detection- , imputation- and control-phases.

Overall strategy has not changed between 2007 and 2010. We refer the readers to the 2007 Quality rapport for details.

2.5.2 Description on imputation per target variable

In the following table is shown which imputation method we used for each target variable (and also for each component within the Belgian questionnaire). The percentage of imputed cases and the total number of observations is added.

Percentage of imputation over the total number of observations per (target) variable

% Imputation method over the total number of observations per (target) variable – gross variables on household level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
HY040G	96.01	0.00	3.99	0.00
HY050G	95.43	4.57	0.00	0.00
HY060G	100.00	0.00	0.00	0.00
HY070G	95.65	0.00	4.35	0.00
HY080G	97.03	0.00	2.97	0.00
HY081G	99.52	0.00	0.48	0.00
HY090G	30.81	0.00	69.19	0.00
HY100G	96.60	0.00	3.40	0.00
HY110G	80.00	0.00	20.00	0.00
HY120G	.	.	.	.
HY130G	97.15	0.00	2.85	0.00
HY131G	99.44	0.00	0.56	0.00

% Imputation method over the total number of observations per (target) variable – NET variables on household level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
HY040N	.	.	.	.
HY050N	97.51	2.49	0.00	0.00
HY060N	.	.	.	.
HY070N	.	.	.	.
HY080N	.	.	.	.
HY081N	.	.	.	.
HY090N	30.81	0.00	69.19	0.00
HY100N	.	.	.	.
HY110N	.	.	.	.
HY120N	.	.	.	.
HY130N	.	.	.	.

HY131N	.	.	.	.

% Imputation method over the total number of observations per (target) variable – gross variables on Personal level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
PY010G	93.09	0.24	1.20	5.47
PY020G	67.53	0.00	0.00	32.47
PY021G	70.72	0.00	29.28	0.00
PY030G	100.00	0.00	0.00	0.00
PY035G	.	.	.	.
PY050G	43.33	0.00	0.67	56.00
PY070G	.	.	.	.
PY080G	0.00	0.00	100.00	0.00
PY090G	99.80	0.00	0.20	0.00
PY100G	52.27	0.00	13.25	34.48
PY110G	69.23	0.00	0.00	30.77
PY120G	51.69	22.03	0.85	25.42
PY130G	57.57	10.20	0.33	31.91

% Imputation method over the total number of observations per (target) variable – Net variables on Personal level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
PY010N	97.44	0.27	1.92	0.37
PY020N	84.63	15.37	0.00	0.00
PY021N	70.72	0.00	0.00	29.28
PY030N	.	.	.	.
PY035N	.	.	.	.
PY050N	79.78	0.89	15.33	4.00
PY070N	.	.	.	.
PY080N	0.00	0.00	100.00	0.00
PY090N	99.80	0.00	0.20	0.00
PY100N	87.55	0.00	12.21	0.25
PY110N	93.59	0.00	6.41	0.00
PY120N	88.14	5.93	1.69	4.24
PY130N	97.04	0.00	1.32	1.64

2.6 Imputed rent

2.8 Imputed rent

From 2007 onwards a measure for 'imputed rent' needs to add to the data.
Below we briefly explain the implementation of imputed rent (IR – hereafter) in the Belgian EU-SILC 2007 data. The text gives insight in the variables and methods used and in the results but is, overall, non-technical. For more in-depth technical background on the subject please turn to the appropriate documentation available via Eurostat (Doc. EU-SILC/162/06/EN).

In order to asses IR it was agreed on with Eurostat to use a (two-step) Heckman regression. The Heckman method involves in essence (A) the resolution of a probit regression model with tenure status of the household dwelling (dichotomy tenant/non-tenant) as dependent variable and conventional explanatory variables (Doc. EU-SILC/162/06/EN). (B) The coefficients found for the inverse of Mills ratio are then introduced in a regression model to counter selection bias in the estimated IR outcomes.

One difficulty in the first step is choosing the right variables. The Eurostat guidelines were closely followed for that purpose and also previous work on the subject of IR for the household budget survey was helpful. The following variables - or rather sets of variables - were selected:

- Characteristics and 'state' of the dwelling: type, number of rooms, presence of problems with the dwelling
- A number of neighborhood characteristics (with some emphasis on the presence of problems).
- Characteristics of the household: ages of the members of the household, their activity status, educational attainment, household type, number of children, number of persons in the household

One difficulty was that individual characteristics (age, activity status, educational attainment) needed to be aggregated on the household level. That was done by the creation of dummy variables for each category of the individual characteristics measuring the presence or the absence of that category on the level of the household. The table below gives an overview.

Not all variables originated from the SILC-database. Calculated for each municipality from the Belgian census 2001 — the distribution renters/owners was added to the equation.

Table: Overview of the variables in the analysis.

Label in output-files	Variable	Operationalisation/ measurement level
HH_INC_Q	Household income – HY020	quintiles
HT	householdtype	Categorical – see EUR.doc….
N_HH	Number of persons in the household	Metric
HH010	Dwelling type	Categorical – see EUR.doc.065
HH030	Number of rooms	Metric
HH050	Ability to keep dwelling warm	Categorical
HH080	Bath or shower	Categorical
HH090	Indoor flushing toilet	Categorical
HS160	Problems with dwelling	Categorical
HS170	Noise from neighbours	Categorical
HS180	Pollution	Categorical
HS190	Crime, violence or vandalism	Categorical
PERC_RENT	% HH renting in community of residence	Source census 2001
AGE_1	<18 yrs.	Dummy
AGE_2	>= 18 yrs. - < 25 yrs.	Dummy
AGE_3	>= 25 yrs. - < 45 yrs.	Dummy
AGE_4	>= 45 yrs. - < 65 yrs.	Dummy
AGE_5	>= 65 yrs.	Dummy
ACTSTA_1	Activity status – working	Dummy
ACTSTA_2	Activity status – unemployed	Dummy
ACTSTA_3	Activity status – retired	Dummy
ACTSTA_4	Activity status – non active	Dummy
EDUC_1	ISCED – 0 – 1	Dummy
EDUC_2	ISCED – 2	Dummy
EDUC_3	ISCED – 3 – 4	Dummy
EDUC_4	ISCED – 5 – 6	Dummy

EXPLORATORY ANALYSIS.

To get a first insight in the impact of each of the variables on the dependent variable tenure status (tenant/owner) a number of (mainly) bivariate logistic regressions were done.

Overall, the results show that the majority of the variables are associated with tenure status. All variables were therefore further kept in the analysis.

The explanatory analysis also resulted in the identification of a small number of missing values on some of the variables. Imputations were necessary to avoid distortion of further analysis.

PROBIT-REGRESSION.

The probit-regression part of the analysis was done in SAS. The output of this analysis is available on demand.

LINEAIR-REGRESSION.

The final estimation of IR is based on a linear regression model in which the observed rent for the renters is the dependent quantity and a number of dwelling-related characteristics are the independent variables.

An important note here is that, that dummy variables for the arrondissement of residence – variables ARR in the output – were introduced in the model. Arrondissements are (in fact) a (juridical – not political) administrative level between municipalities and provinces. We believe they are excellent indicators of regional differences and tendencies on scale smaller than provinces but bigger than municipalities.

The inverse-mills coefficient was significant at <0.001 level.

The output of the final regression is available on demand.

2.7 Collection variable company Car

Since 2005, we decided to work with **the national rules of the tax authorities**. The benefit for individuals of using a company car for private goals was not directly assessed at the interview but afterwards calculated by applying the applicable taxation rules.

The fiscal benefit of all nature that a person has - due to disposition of a company car for private goals - is calculated by multiplying a fixed amount of kilometres driven for private use by a coefficient. To calculate the latest we need the fiscal cylinder capacity of the car. This fixed amount of kilometres driven for private use is for the tax authorities 5000 km if the distance home-work is less than 25 km, and 7500 if it's more than 25 km.

Since 2005, we asked directly the fiscal cylinder capacity and the distance between work and home. In case of non response of the cylinder capacity, we asked the mark, type and registration year of the car. Then we had to use an imputation method.

Imputation: To calculate the cylinder capacity, we did the following. We assumed that a company car is mostly diesel driven. We looked up for each mark, type and diesel engine what the corresponding cylinder capacity is. If we had several cylinder capacities for the type of the mark, we calculated the weighted mean of the cylinder capacity. If there is not diesel version for a type of car, we did the same logic but than for petrol.

Once we had that we could easily find the corresponding fiscal coefficient. Than we only had to multiply it by the fixed amount of kilometres driven for private use to obtain the fiscal benefit of all nature

Example:

Type of car	Fiscal cylinder capacity	Forfait	Distance home work	Fixed amount	Fiscal benefit of all nature
Smart fortwo	5	0,1864	< 25 km	5000	931 €
Smart fortwo	5	0,1864	> 25 km	7500	1396 €

After we calculated the fiscal benefit of all nature for a whole year, we weighted it for respondents who didn't dispose for a whole year of the company car. **The fiscal benefit of all nature is a gross non-cash employee income**.

3.Comparability

All household members of 16 year and older **at the time of the interview**, are selected for a personal interview. From 2006 on the age of 16 will be calculated at the end of the income reference period.

3.1 Basic concepts and definition

Only changes from first wave are reported.

Basic information on activity status during the income reference period

Basic information on activity status during the income reference period was mainly obtained via the calendar question (I40) in contrast to 2004 where it was obtained by combining the answer for question I8 (PL030) with the answer(s) for question(s) I38 (PL200) and for those with a change I40 (calendar question)). ALSO SEE REMARK 2.5.0.

3.2 Components of income

3.2.1 Differences between the national definitions and standard EU-SILC definitions, and an assessment, if available, of the consequences of the differences mentioned will be reported for the following target variables.

Total household gross income

HY010 = PY010 + PY020G + PY050G + PY090G + PY100G + PY110G + PY120G + PY130G + PY140G + HY040G + HY050G + HY060G + HY070G + HY080G + HY090G + HY110 G.

PY020G was not part of HY010 for 2004.

For 2005 and 2006 PY020G only contains the value of company cars.

Family/children related allowances

For the SILC 2004 Belgium asked allowances received from the federal government. From 2005 on it also includes birth grants given by some local authorities and medical organizations.

Income received by people aged under 16: in 2004 we asked the amount for last month (current) but the reference period for the variable is income reference period (year 2003). This was corrected for 2005 and the question aimed at the total income received last year by people aged fewer than 16.

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

No change from the previous wave.

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

No change from the previous wave.

3.2.4 The method used for obtaining income target variables in the required form (i.e. gross values)

See above for information on control, correction, imputation and creation of the gross target variables.

Tracing rules

4. Coherence

The results of the Belgian EU-SILC2010 are in an acceptable way coherent with the results of previous waves. In depth studies to demonstrate this are currently in process at Statistics Belgian. Nevertheless however, these analysis (not the general analysis that is done during the validation phase at Eurostat) always has to be weighted against other priorities such as timelines f. ex.