

Annex 3:

Drawing of the SILC 2016 sample: stratification for Brussels with tax quantiles

1. Introduction

In the framework of Eurostat's evolving requirements and of the survey reform scheduled in 2018, Statistics Belgium decided to modify the sampling design for the Brussels region as from SILC 2016. Starting with a primary and then secondary unit drawing, as for the whole country, we have chosen to proceed with a stratified sampling based on the new administrative data available. The aim of this report is to summarise the works carried out to justify and implement a new stratification for the SILC survey.

Over the last few years, the standard deviation of the AROPE is between 2.28% and 3.26% (see Table 1). In its last recommendations¹, the maximum standard deviation accepted by Eurostat for Brussels is 2.36%. This low precision is mainly due to a very high AROPE in Brussels (around 40%). Therefore, in order to improve this precision, we propose to modify the sampling design for Brussels, by stratifying now according to the tax data of households in Brussels.

As the SILC survey works with panels, these new changes will only have a full effect on the precision of our estimators after four whole years, i.e. for SILC 2019. However, given the relatively high attrition rate, we can anticipate that we will be able to estimate the impact of the new stratification from two or three years of the panel, i.e. 2017 or 2018. This way, the work will have significant effects on the SILC reform scheduled in 2019 (the questionnaire will be replaced by administrative data).

Table 1: Estimator of the standard deviation of the AROPE for the Brussels region

Year	Standard deviation of the AROPE
2010	2.81 %
2011	3.17 %
2012	3.26 %
2013	3.13 %
2014	2.28 %

All the results presented here concern the years 2010 to 2014 of the SILC survey. For these five years, we have the SILC samples of individuals and households and the universes of individuals and households. Moreover, we have complete fiscal data for these four databases. Here, we confine ourselves to the households and individuals situated in the region of Brussels.

¹ « Working Group meeting " Statistics on Living Conditions " », Doc LC/132/15/EN, Eurostat, 10-12 June 2015

2. Calculation of tax quantiles

Regarding the fiscal data, we first use the equivalised household taxable income². This income is currently the closest to the equivalised household income calculated by the SILC survey. It should therefore be a good indicator of the monetary poverty risk (AROP).

This equivalised household taxable income is calculated per household (within the meaning of the National Register) for a year N, and is then assigned to each member of this household. Based on the equivalised taxable income per individual, we calculate for the year N the tax quantiles (we always calculate 5 quantiles; we demonstrate hereafter that the results are not very sensitive to the increase of the number of quantiles).

We know that, at the time of drawing the sample and so of the stratification per tax quantiles for a year N, only the taxable income for the year N-3 will be available. Therefore, in order to conduct our study, we take this discrepancy into account in the data accessibility. So we couple, at individual level, the SILC N base with the N-3 base including the tax data. As a result, we obtain for each SILC individual of the year N his N-3 tax quantile. Finally, we group the individuals per household (according to the definition in the National Register), by choosing the highest quantile among the household members. This gives us, for the SILC years 2011 to 2014, the tax quantiles of the households from the tax bases 2008 to 2011.

Table 2 shows the share of households according to the tax quantiles, for the universe and for the SILC sample. We note that the distribution of households in the SILC sample generally corresponds to that in the universe, with two exceptions:

- The share of households whose tax situation is unknown is largely underestimated in our sample for the four years.
- The share of households in the first (last) quantile seems structurally underestimated (overestimated).

We will correct these trends when choosing the size of the tax strata. The analysis remains the same if we use as basis the tax data for N-2 or N-1 instead of N-3.

² Data calculated by Statistics Belgium based on the fiscal data.

Table 2: Share (%) of households in the tax quantiles, according to the universe or the SILC sample

SILC year	Fiscal year	Database	Quantiles					Unknown
			1	2	3	4	5	
2011	2008	Universe	16.22	16.36	17.08	17.45	17.91	14.98
		Sample	15.20	17.73	18.34	18.46	22.68	7.60
2012	2009	Universe	16.13	16.00	16.96	17.43	17.86	15.62
		Sample	15.03	18.11	17.20	21.41	21.75	6.49
2013	2010	Universe	16.85	15.93	17.68	17.93	18.48	13.13
		Sample	15.05	18.08	19.02	21.00	20.17	6.69
2014	2011	Universe	16.72	15.90	17.83	18.16	18.78	12.62
		Sample	16.67	16.67	19.49	19.49	20.81	6.87

3. Fiscal LWI

Based on the fiscal data, we also calculate all the non-labour income collected by the household. By doing so, we hope to get a proxy for the indicator of low work intensity of the household (LWI).

The calculation is done on the same basis as for the tax quantiles (see 2.2), by adding up all the non-labour income (unemployment, sickness, retirement pensions, disability, etc.). In order to respect the availability of tax data, we also use the N-3 data to conduct the analysis.

We have chosen to build, from the sum of the non-labour income of a household, a binary indicator approximating the LWI indicator with the fiscal data. We have set the non-labour income threshold at 6,000€ per year, based on a study on the closeness between this non-labour income and the LWI indicator in the SILC samples (see annex 2 on small area estimation for more information on this choice). Our objective is to couple this binary indicator with the previously calculated tax quantiles.

Just like Table 2, Table 3 shows the share of households in fiscal LWI, according to the universe or the SILC sample. Again, we note that the shares are fairly stable and fairly close between the universe and the sample.

Table 3: Share (%) of households considered or not in fiscal LWI, according to the universe or the SILC sample

SILC year	Fiscal year	Database	Fiscal LWI	
			NO	YES
2011	2008	Universe	77.01	22.99
		Sample	74.07	25.93
2012	2009	Universe	75.03	24.97
		Sample	70.62	29.38
2013	2010	Universe	72.96	27.04
		Sample	69.59	30.41
2014	2011	Universe	71.29	28.71
		Sample	70.00	30.00

4. Links with poverty indicators

We will analyse here three poverty indicators: AROP, LWI and AROPE. The AROP indicator should be highly correlated with the tax quantiles, because the measure is very similar. The LWI indicator should be highly correlated with the fiscal LWI. Finally, the AROPE indicator is our objective of precision for Eurostat.

To be useful, the stratification variables should distinguish as well as possible between the households' profiles according to their poverty indicators. A strata where the share of an indicator is very different from the average share for the whole sample is useful.

Tables 4 and 5 show the share of households in the three poverty indicators, according to the tax quantile (Table 4) or the fiscal LWI (Table 5). We note that the tax quantiles make it possible to "distinguish" between the households' profiles according to the three poverty indicators. For example, for the SILC year 2014, being in both quantiles 1 and 2 induces a share in AROP, LWI and AROPE which is much higher than the average of the sample (38%). On the contrary, being in the quantiles 4 and 5 leads to a share in AROP, LWI and AROPE much lower than the average of the sample. These results are stable over the four years. Therefore, the stratification by tax quantile should, in all likelihood, improve the precision of the three indicators.

Table 4: Share of households over the three poverty indicators according to the tax quantile

SILC year	Fiscal year	Database données	Fiscal quantiles					Total	
			1	2	3	4	5		Unknown
2011	2008	AROP	58%	59%	24%	8%	4%	56%	31%
		LWI	47%	55%	24%	12%	3%	38%	27%
		AROPE	69%	71%	37%	14%	7%	57%	38%
2012	2009	AROP	66%	62%	31%	10%	5%	40%	32%
		LWI	55%	50%	27%	11%	5%	27%	27%
		AROPE	71%	79%	44%	17%	8%	44%	41%
2013	2010	AROP	69%	63%	25%	11%	3%	33%	32%
		LWI	60%	51%	29%	9%	4%	13%	28%
		AROPE	76%	73%	38%	19%	6%	34%	39%
2014	2011	AROP	64%	56%	26%	10%	4%	24%	29%
		LWI	58%	51%	28%	11%	6%	16%	29%
		AROPE	73%	70%	38%	18%	9%	29%	38%

Reading: 58% of the households in Brussels in the first fiscal quantiles in 2011 were in AROP.

However, the results for the fiscal LWI are more mixed. As expected, the fiscal LWI can only separate the households according to the LWI indicator. And even so, this separation is only valid where the fiscal LWI equals 1. In the opposite case, the share of LWI is still very close to the average of the sample. Consequently, it is certain that the fiscal LWI should be used, *a minima*, with the tax quantiles. And it is possible that the explanatory power of the fiscal LWI is captured by that of the tax quantiles when both are used in combination. For example, the separation "fiscal LWI Unknown" leads to shares in AROP, LWI and AROPE very different from the average of the sample, but this information is strictly identical to that obtained with the tax quantiles.

Table 5: Share of households for the three poverty indicators according to the fiscal LWI

SILC year	Fiscal year	Database	Fiscal LWI			Total
			NO	YES	Unknown	
2011	2008	AROP	28%	30%	56%	31%
		LWI	20%	54%	38%	27%
		AROPE	36%	40%	57%	38%
2012	2009	AROP	33%	28%	40%	32%
		LWI	22%	52%	27%	27%
		AROPE	41%	40%	44%	41%
2013	2010	AROP	33%	29%	33%	32%
		LWI	21%	61%	13%	28%
		AROPE	40%	40%	34%	39%
2014	2011	AROP	32%	26%	24%	29%
		LWI	23%	62%	16%	29%
		AROPE	39%	39%	29%	38%

Reading: 31% of the households in Brussels situated in the fiscal LWI in 2011 were in AROP.

5. Variance analysis of the poverty indicators

We continue the analysis with a test of the variance analysis obtained when using the Poulpe macro in SAS, in order to test our strata in near real situation. Still using only the data for Brussels, we calculate the variance by stratifying the households in two different ways:

- Using only the tax quantiles
- With the tax quantiles and the fiscal LWI: in that case, with the aim of not increasing too much the number of strata, we carried out aggregations (Table 6) on the basis of the closeness of these strata.

We compare these two situations with the hypothetical situation in which the variance of the indicators is calculated by considering the groups as strata (in order to test a situation without strata). We also tested the calibration of these tax data (this time with a shorter timescale given that the tax data are mostly N-1).

Table 6: Proposed strata according to the tax quantiles and the fiscal LWI

		Tax quantiles					
		1	2	3	4	5	Unknown
Fiscal LWI	YES	1	1	3	5	7	9
	NO	2	2	4	6	8	9

Table 7 shows the estimated standard deviations for the various scenarios and the various poverty indicators. We note that stratification slightly reduces the standard deviation of the indicators, but that this effect is reinforced when calibration of the tax data is applied additionally. Double stratification of the tax quantiles and the fiscal LWI doesn't provide a much higher precision than the stratification of the tax quantiles only.

Table 7: Standard deviation of the poverty indicators according to the stratification scenario

Year stratum	Year calibration	Database	No stratification	Stratification quantiles		Stratification quantiles + fiscal LWI	
				Without calibration	With calibration	Without calibration	With calibration
2008	2010	AROP	1.990%	1.874%	1.387%	1.871%	1.378%
		LWI	1.971%	1.713%	1.579%	1.711%	1.583%
		AROPE	2.238%	1.908%	1.401%	1.905%	1.374%
2009	2011	AROP	2.121%	1.837%	1.390%	1.833%	1.383%
		LWI	1.715%	1.679%	1.521%	1.677%	1.501%
		AROPE	2.327%	1.869%	1.380%	1.866%	1.367%
2010	2012	AROP	1.860%	1.765%	1.356%	1.762%	1.335%
		LWI	1.572%	1.582%	1.489%	1.580%	1.479%
		AROPE	1.922%	1.794%	1.365%	1.791%	1.327%
2011	2013	AROP	1.831%	1.749%	1.283%	1.746%	1.280%
		LWI	1.634%	1.555%	1.417%	1.553%	1.419%
		AROPE	1.871%	1.775%	1.252%	1.772%	1.243%

6. Choice of the stratification

On the basis of the results of the previous subsection we opt for the stratification of the tax quantiles only. Fiscal LWI didn't provide any additional benefit for the stratification of Brussels. We therefore decide to break down the households of Brussels into 5 tax quantiles³, or 6 tax strata.

In so far as the SILC survey is a 4 year panel survey, the precision improvements mentioned here will not be visible immediately. We can expect a significant decrease of the standard deviation of the various indicators only after 3 or 4 years. We can already see a decrease of the variance for AROP and AROPE in 2016 (see figure 1 below) but it is still not clear whether this decrease results from the new stratification.

³ Other tests were carried out under the same conditions by increasing the number of tax quantiles. It didn't result in a significant improvement in the precision.

Figure 1: Standard deviation of the poverty indicators per year

