EUROMOD WORKING PAPER SERIES

EM16/14

Gross incomes in the Belgian SILC dataset: An analysis by means of EUROMOD

André Decoster, Dieter Vandelannoote, Toon Vanheukelom, and Gerlinde Verbist

August 2014

Gross incomes in the Belgian SILC dataset: An analysis by means of EUROMOD[1]

André Decoster[a]
Dieter Vandelannoote[b]
Toon Vanheukelom[a]
Gerlinde Verbist[b]

[a]Centre for Economic Studies, Katholieke Universiteit Leuven
[b]Herman Deleeck Centre for Social Policy, University of Antwerp

Abstract

Income aggregates and poverty and inequality measures tend to show important differences when calculated either with disposable income reported in SILC data, or with the same income concept calculated on the basis of the microsimulation model EUROMOD, which starts from gross incomes in SILC. This is one of the reasons why gross income data in SILC are often regarded with some suspicion. In this paper we try to shed light on this question by 1) comparing gross incomes in SILC with gross incomes reported on the fiscal form for the same individuals; 2) testing a re-calibration of gross incomes proposed by Immervoll and O'Donoghue (2001), to make them consistent with reported net incomes. We find that on average fiscally reported gross incomes exceed gross incomes in the survey. It is not clear however, whether the method of constructing updated gross incomes through an iterative method using the observed net incomes and withholding tax rules, is a genuine improvement upon the reported gross income distribution

JEL classification: C81, R20

Keywords: household income, administrative data, survey data, microsimulation model

Corresponding author:
Dieter Vandelannoote
E-mail: dieter.vandelannoote@uantwerpen.be

[1] The paper uses EUROMOD F6.16. EUROMOD is continually being improved and updated and the results presented here represent the best available at the time of writing. Any remaining errors, results produced, interpretations or views presented are the author's responsibility. The process of extending and updating EUROMOD is financially supported by the Directorate General for Employment, Social Affairs and Inclusion of the European Commission [Progress grant no. VS/2011/0445]. We make use of micro-data from the Belgian Survey on Income and Living Conditions (SILC) dataset of 2010 provided by the Directorate-General of Statistics within the FPS Economy, SMEs, Self-Employed and Energy. The data is matched with IPCAL data containing the fiscal forms of the households in the survey. The latter dataset was obtained as part of the SBO-project IWT090044 "FLEMOSI: A tool for ex ante evaluation of socio-economic policies in Flanders", funded by IWT Flanders (see www.flemosi.be).

1. INTRODUCTION

A comparison of yearly disposable income according to the Belgian SILC data of 2010 (BE-SILC 2010) with the same income concept calculated on the basis of EUROMOD shows some important differences: yearly incomes simulated with EUROMOD, for instance, are on average € 1,500 below the ones reported in BE-SILC 2010. Also poverty rates and inequality indicators are much lower when compared to the external statistics based on SILC. For 2009, EUROMOD calculates that 10.25% of the individuals are at risk of poverty (measured as 60% of median equivalent household income). This is more than 4 percentage points lower than the official statistics, based on reported incomes in SILC (14.56%). Also the GINI coefficient is lower compared to the official statistics (0.22 versus 0.26). These differences may come as a surprise, given that the SILC data are used as the input data source for the calculation of disposable incomes in EUROMOD (see Hufkens et al. 2013 for more details).

In principle, these differences can be due to the following reasons:

- Gross income data in SILC, which are used as starting variables for EUROMOD, are unreliable;
- Net income data as reported in SILC, and which are used to compare EUROMOD output, are unreliable;
- Tax-benefit rules are inadequately implemented in EUROMOD, e.g. due to missing information which is needed for accurate tax-benefit calculations[2];
- A combination of the above reasons.

In this paper we try to gain a better understanding of the possible sources of this discrepancy. In order to gauge the quality of the SILC income data, we compare them with actual administrative income data on which taxes have actually been calculated and that form the basis of the official aggregate government tax revenue numbers. Indeed, as part of the Flemosi project we dispose of a unique dataset which combines tax return data with survey data for the same individuals. [3] More specifically for the SILC individuals of the BE-SILC 2010, we obtained the information of the tax forms from the administrative dataset (known as the IPCAL dataset). We describe this 'enriched' SILC-dataset in section 2. This additional information on individuals in SILC allows us to compare gross incomes from the administrative source (in this case personal income tax forms) with gross incomes reported in SILC. This comparison is described in

2 An example are the many possibilities for tax deductions (e.g. for energy saving investments; real professional costs, etc.) which are only partially covered in EUROMOD due to lack of necessary information in the SILC-data.

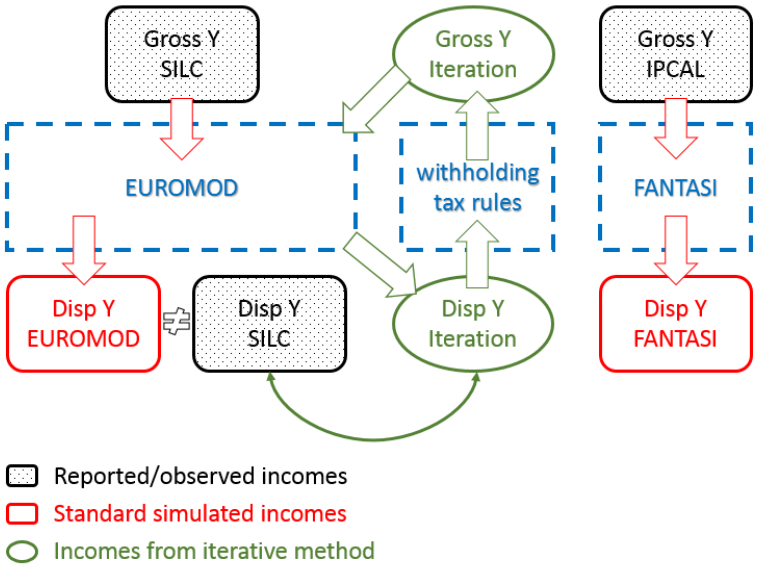
3 More information about the Flemosi project can be found in Decancq et al (2012) or at www.flemosi.be.

section 3 and reveals that gross incomes in SILC are *underreported* compared to the ones in the administrative IPCAL data.

The rest of the paper explores one of several possible routes to adapt or 'update' gross incomes used in EUROMOD. It is based on the assumption that SILC respondents might have better knowledge of their net incomes (as reported in SILC) than of their gross ones. In that case it seems natural to start from these reported net – or disposable - incomes to reconstruct 'corrected' gross incomes which are consistent with the reported net incomes. In Figure 1 we illustrate how this is done by applying an iterative method developed in Immervoll and O'Donoghue (2001) to calculate new gross incomes based on the net incomes as reported in SILC. We explain this iterative method in detail in section 4.

In grey we indicate the observed or registered incomes. For gross incomes (top of the figure) we have two sources: SILC on the one hand, and IPCAL on the other hand. At the bottom of the figure we only mention disposable incomes from SILC as observed (hence in grey). We are able to calculate disposable incomes on the basis of the IPCAL-data by means of a specific micro-simulation model FANTASI. Yet, here we will focus on the validation of gross incomes, and hence not use this microsimulation model FANTASI in this paper. The validation consists of two parts: first we compare gross incomes with gross incomes in SILC. Secondly, we use the difference we observe between the disposable income observed in SILC and the disposable income simulated by EUROMOD, to test one procedure to correct or update gross incomes in SILC. This is indicated in the middle of Figure 1. We start from the reported disposable income (in the form of separate income components which are reported as net incomes), and reconstruct the corresponding gross incomes. We then run EUROMOD iteratively to find gross incomes which produce disposable incomes as close as possible to the observed net incomes.

FIGURE 1: OVERVIEW OF AVAILABLE DATA AND ITERATIVE PROCEDURE TO CALIBRATE GROSS INCOMES



Once these new 'calibrated' gross incomes are obtained, we re-run EUROMOD and compare disposable household income, poverty rates and inequality indicators between the directly reported ones of SILC, and the simulated ones of EUROMOD. Finally, in section 5, we also compare the distribution of 'corrected' gross incomes with the gross incomes from the administrative dataset of IPCAL to check whether the procedure improves the correspondence between SILC and IPCAL datasets.

The comparison in that final section 5 naturally points towards other possibilities to investigate the quality of gross incomes in SILC. Since, we do have an exact match between SILC and IPCAL-information, we could as well *replace* the SILC-information on e.g. gross labour incomes, by the IPCAL-information. We could also insert additional information on fiscal expenditures into the EUROMOD dataset, and as such produce a more accurate calculation of personal income tax liabilities.[4] Since, for that purpose we would have to expand EUROMOD with new information and also new routines, we leave this for future research. In this paper we confine the exercise to an application of the iterative procedure of Immervoll and O'Donoghue (2001), and consider this paper as reporting on a first step in a broader investigation of SILC gross incomes.

[4] The fact that we *know* that important fiscal expenditures are missing in the EUROMOD-calculation of disposable incomes, also weakens the argument in favour of the iterative procedure announced above, if we assume that SILC-respondents report their disposable income taking into account the effect of fiscal expenditures.

2. MATCHING BE-SILC 2010 WITH IPCAL

The Survey on Income and Living Conditions (SILC), a micro data set with a representative sample of private households, is the standard dataset for distributional and poverty analysis in the EU. It combines socio-demographic information with an array of income components at the level of the individual and the household. The Belgian version of EU-SILC is more detailed than the European version which has been aggregated and harmonized by EUROSTAT to ease European cross-country comparisons. For Belgium, the dataset contains a representative sample of the population covering over 14,000 individuals living in about 6,000 households.

The IPCAL database is an administrative dataset which contains the fiscal forms of the Belgian citizens, with all variables relevant for the calculation of personal income taxes. Within the FLEMOSI-project we obtained the subset of the IPCAL-dataset which contains the fiscal forms for the BE-SILC-2010 individuals. This was delivered as a unique identifier for each individual in both datasets. Both datasets contain information on incomes earned in 2009.

Table 1 lists the number of observations that have (not) been matched. For EUROMOD, we use the EUROMOD base dataset as our starting point which contains information on 14,700 individuals, including children (first row in Table 1). Evidently, for most children we cannot find a corresponding tax return file as they are 'fiscally dependent' and have no income to declare. The p-file of SILC (second row in Table 1) which contains information, including gross incomes, for all individuals aged 16 or more, has 11,816 observations. Since these are all potential tax units, this is the relevant subpopulation to look for a corresponding record in IPCAL. The IPCAL data has 11,792 observations.[5] Of the 11,816 individuals in the p-file of SILC 11,656 observations have a corresponding record in the IPCAL data, which can be considered a very good match given that the p-file might still contain fiscally dependent children. The remaining 136 observations in IPCAL for which no corresponding record in the SILC could be identified are discarded from any further analyses.

5 It should be noted that the dataset we obtained actually contains 18,471 observations, 6,679 of which have a missing value for variable rb030, the id-variable in SILC, and hence cannot be matched. 752 of them have total gross income equal to zero while 5,927, or almost 89% of this subset, have positive total gross income. It is unclear why these are in the dataset. They are deleted for further analysis.

TABLE 1: NUMBER OF MATCHED OBSERVATIONS

Dataset	Matched	Not Matched	Total number of observations
EUROMOD	11,656	3,044	14,700
SILC (P-file)	11,656	160	11,816
IPCAL	11,656	136	11,792

Given that 11,656 individuals of the EUROMOD base dataset have a corresponding tax return, 3,044 of the total of 14,700 SILC-individuals are either fiscally dependent or their record is missing in the dataset with tax returns. Table 2 shows the age distribution of these 3,044 individuals in the EUROMOD database for whom no tax return data was found.

TABLE 2: AGE CLASSES OF INDIVIDUALS IN EUROMOD BASE DATASET FOR WHOM NO TAX RETURN DATA WERE FOUND IN IPCAL-DATASET

Age	Number of Observations	Percent of Subpopulation
<=25	2,911	95.6
25-65	120	3.9
>65	13	0.5
All	3,044	100.0

Nearly 96% of those individuals are younger than 25 years old. Of the 3,044 individuals for whom no tax return was found, 2,780 are individuals who have no tax return, no partner and for whom we do know the identification number of the mother and/or father. We can assume that these 2,780 individuals are 'potentially fiscally dependent' children. The 264 remaining individuals may still be dependent on other persons and therefore need not be considered as problematic at this stage.

In Table 3 we show the total number of actual tax returns as well as the total per region. In Belgium a married or legally cohabitating[6] couple has to fill in one single tax form. Therefore each tax form has two columns, one for the head and one for his or her partner. In the data the distinction is made by referring to those persons as respectively the "A"-person (filling out the first column) and the "B"-person (filling out the second column). The number of actual tax returns is thus obtained by counting the number of "A"-persons in the data. We also included the weighted totals as well as the official statistics on tax returns. The official statistics also include 'zero incomes', i.e. tax returns with zero net taxable income.

6 'Legally cohabitating' means that both partners have signed an official document before a local official, stating they live together.

TABLE 3: ACTUAL TAX RETURNS IN SILC AND IPCAL– INCOMES 2009

	Belgium	Flanders	Wallonia	Brussels
Total (SILC)	6,498,733	3,706,889	2,130,309	661,535
% of total (SILC)		57.0	32.8	10.2
Total (IPCAL)	6,731,398	3,867,099	2,182,727	681,572
% of total (IPCAL)		57.5	32.4	10.1
Official Statistics[7]	6,771,747	3,870,356	2,107,508	793,883
% of total (official)		57.2	31.1	11.7
Ratio of SILC to official	0.96	0.96	1.01	0.83
Ratio of IPCAL to official	0.99	1.00	1.04	0.86

As can be seen from the bottom two rows, the relative distribution among the regions quite closely mirrors the official distribution. Not surprisingly, the IPCAL data are in general closer to the official statistics than the SILC data, notwithstanding that the latter are quite close as well with an overall underestimation of the number of tax returns of around 4%. In the case of the IPCAL data – and this holds true for the SILC data as well – we see a considerable underestimation for the Brussels region, 14% and 17% for IPCAL and SILC respectively. The number of tax returns in the Walloon region is overestimated in both datasets but more so in the IPCAL data (4%) than in the SILC data (1%). For Flanders the IPCAL data show an almost perfect score with a ratio of 0.99 while the underestimation is around 4% in the SILC data. Except for the Brussels region, we can conclude that the number of tax returns is captured quite well in both datasets.

Finally, in Table 4 we show the distribution of tax units per sociological household. Almost 70% of households consist of only one tax unit. This group includes single (parent) households, as well as married and legally cohabitating couples. The second largest group is that of households consisting of two fiscal units, which are mainly unmarried couples living together and without a legal contract of cohabitation, that are treated as separate fiscal units. About 8% of sociological households house three or more fiscal units.

[7] See http://statbel.fgov.be/nl/statistieken/cijfers/arbeid_leven/fisc/

TABLE 4: NUMBER OF TAX RETURNS PER HOUSEHOLD IN BELGIUM

	not weighted		Weighted	
number of tax units	**Number**	**Percentage**	**number**	**Percentage**
0	69	1.1	53,813	1.2
1	4,134	67.4	3,228,976	69.0
2	1,437	23.4	1,030,091	22.0
3	390	6.4	281,893	6.0
4	82	1.3	68,354	1.5
5 or more	20	0.3	17,079	0.4
Total	6,132		4,680,206	

3. COMPARING GROSS INCOMES IN IPCAL AND SILC

In this section we directly compare the gross income information in IPCAL with the one in SILC. In a first subsection the focus will be on gross income at the household level, while in a second subsection attention will be shifted to incomes at the individual level. All amounts are annual amounts in Euro.

3.1 Household gross income

In the SILC-data, household gross income is captured by one variable (HY010). We start from this variable to compare household gross income between SILC and IPCAL. Some adjustments are necessary however, since not all income components of HY010 are included in a gross income concept for tax purposes. Child allowances and social assistance, for example, are tax exempt. In the first column of Table 5, we list the income components, found in variable HY010, in the SILC data. In the second column, with a '+' or '-', we indicate whether or not the specific component has been withheld in the definition of gross household income for comparison with the administrative fiscal data. The last column provides a description of the variables in question. Variable PY080 (pension from private pension plans) is not included in HY010, but we have included it here because these amounts have to be declared on the tax form. Further adjustments were made to account for the fact that personal social insurance contributions are included in the gross income variable in SILC but not in IPCAL. Income from (self-)employment and pensions has been adjusted by a factor reflecting the social insurance contributions. For labour market income (employee and self-employed) 13.07% and for pensions 3.55% was deducted to arrive at comparable gross amounts found in the administrative data.[8]

8 For self-employed the 13.07% deduction is a rather crude generalisation as the rules are more refined for income from self-employment. We also make no distinction between employees and civil servants, and for employees we did not take into account the 'work bonus', which is a reduction of social contributions for low wages.

As for the administrative data (IPCAL), we use all codes that are available, i.e. that were requested in the data demand, to calculate gross income. Here we add all incomes from all members of the sociological household as identified in SILC. It is possible, and even common, to have more than one fiscal unit in the same sociological household. For the comparison in this subsection we add all gross incomes of the different fiscal units together up to the level of the sociological household.

TABLE 5: COMPONENTS OF HOUSEHOLD GROSS INCOME (HY010) INCLUDED OR EXCLUDED FOR COMPARISON WITH FISCAL GROSS INCOME

Income component	included (+) or excluded (-)	Description
HY040G	+	income from rental of a property or land
HY050G	-	family/children related allowances
HY060G	-	social exclusion not elsewhere classified
HY070G	-	housing allowances
HY080G	+	regular inter-household cash transfer received
HY090G	+	interests, dividends, profit from capital investment in unincorporated businesses
HY110G	+	income received by people aged under 16
PY010G	+	employee cash or near cash income
PY021G	-	company car
PY050G	+	cash benefits or losses from self-employment
PY090G	+	unemployment benefits
PY100G	+	old age benefits
PY110G	+	survivor benefits
PY120G	+	sickness benefits
PY130G	+	disability benefits
PY140G	-	education-related allowances
(PY080G)	+	pension from individual private plans

3.1.1 Total gross household income

In the tables below we show the differences between gross incomes in IPCAL and in SILC, unless otherwise stated. Monetary differences are always the amount in IPCAL (administrative data) minus the amount in SILC. Percentage differences are with respect to the gross amounts in SILC (= ((IPCAL – SILC) / SILC) * 100). Thus, a positive (negative) monetary difference implies the amount in the administrative data is larger (smaller) than in the SILC-data. Similarly for positive (negative) percentage differences: a positive (negative) percentage amounts to a gross income amount in IPCAL that is larger (smaller) than in SILC. We mostly show the mean difference and a summary of the dispersion of the differences by means of the quantile values at some selected

percentiles (1%, 5%, 50%, and 95%) of the distribution of the differences, as well as the standard deviation of the differences.

Table 6 displays this information for the difference in total household gross income. We show the results for Belgium as well as for the three regions. Average gross household income in IPCAL amounts to €44,692 per year, compared to a much lower €40,021 in SILC. That means that, on average, the difference equals €4,700, or expressed as a percentage of the average gross income in SILC, IPCAL exceeds the SILC-average by 11.7%. Individual differences show a large dispersion: a standard deviation of more than €31,000. There are also many households for which gross income in SILC is larger than the gross income constructed from the fiscal information: for 33% of all households the difference is negative. Since on average the IPCAL average gross income exceeds the SILC one, the positive differences outweigh the negative differences. For the Brussels region the average difference is considerably less than for the other regions but the dispersion is much higher as shown by the standard deviation that is nearly 40% higher than in Flanders and even close to 70% higher than in Wallonia.

TABLE 6: AVERAGE DIFFERENCE IN TOTAL GROSS HOUSEHOLD INCOME IN EURO: IPCAL VS SILC

	Belgium	**Flanders**	**Wallonia**	**Brussels**
mean income in € per year				
IPCAL	44,693	48,131	42,284	36,711
SILC	40,021	42,973	36,982	35,646
difference [IPCAL-SILC] in € per year:				
Mean in €	4,672	5,158	5,302	1,065
(in %)	(11.7%)	(12.0%)	(14.3%)	(3.0%)
percentile 1	-54,216	-46,941	-37,369	-91,890
percentile 5	-14,400	-13,659	-12,601	-25,088
percentile 50 (median)	1,916	2,023	2,200	981
percentile 95	30,025	32,004	30,520	23,740
standard deviation	31,284	31,006	25,500	43,325
Number of observations	6,095	3,269	2,031	795

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

Figure 2 and Figure 3 visualise the information on the two distributions more in detail. Figure 2 reveals that, except for the very low incomes (i.e. below €5000 per year), the frequency is lower in IPCAL than in SILC for incomes below about €50000. We find more higher gross incomes in IPCAL than in SILC. The distribution of absolute differences in Figure 3 is quite symmetric, but with some more mass on the positive side. Certainly for the large differences (more than €20000), we have much more positive ones (IPCAL exceeds SILC) than the reverse.

FIGURE 2: DISTRIBUTION OF HOUSEHOLD GROSS INCOME IN IPCAL AND IN SILC (INCOMES 2009)

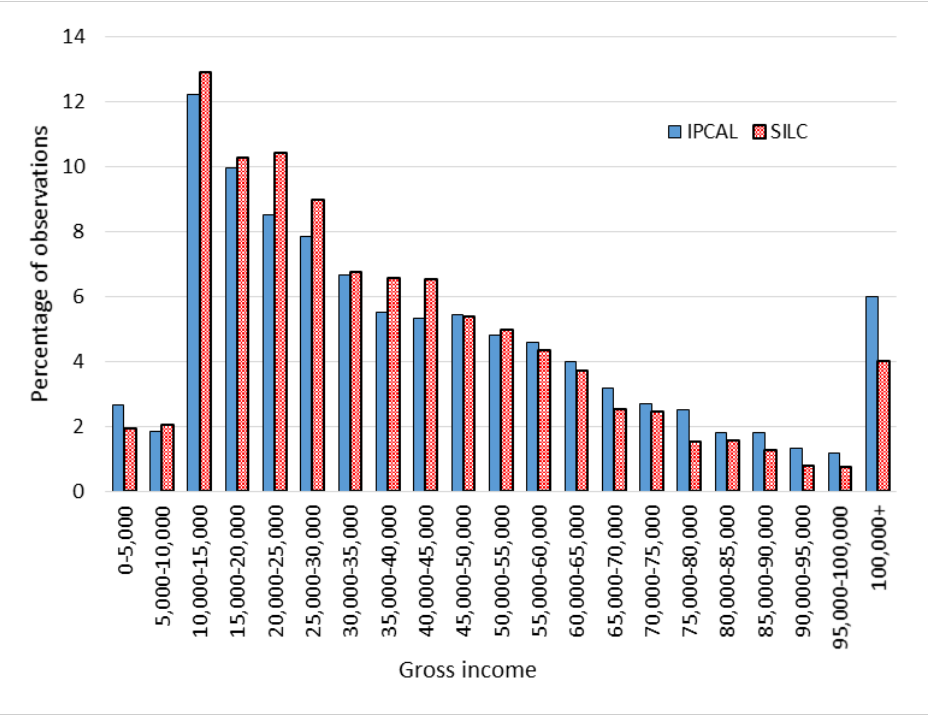
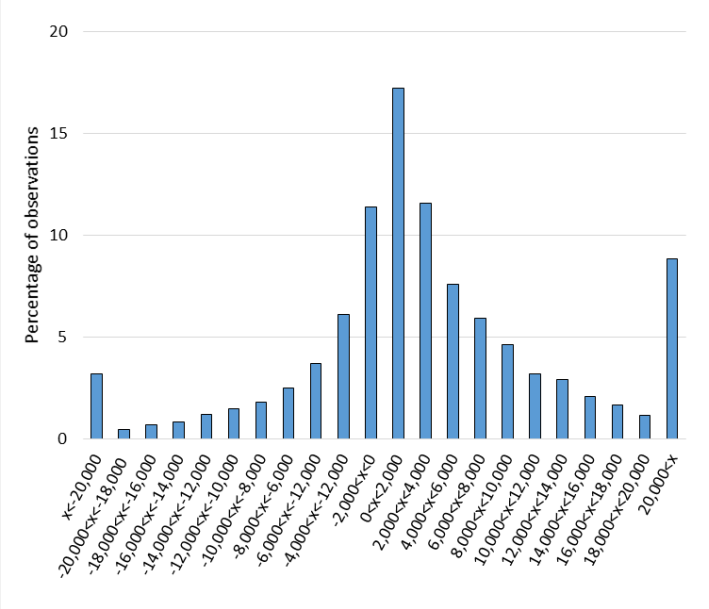


FIGURE 3: DISTRIBUTION OF DIFFERENCES IN HOUSEHOLD GROSS INCOME BETWEEN IPCAL AND SILC (IPCAL-SILC; INCOMES 2009)



Finally, Table 7 tries to answer the question whether the difference in gross incomes between the two datasets is more than a mere translation of the distribution, but also leads to serious re-

rankings of households in the distribution. The table cross-tabulates the position of households in the two respective distributions. We construct deciles based on total gross household in SILC and IPCAL and compare whether there are large differences in decile classification for households. The answer is clearly negative: most of the households are clustered around the diagonal, implying that they stay close to their original decile based on SILC data. In fact, for all deciles, more than 80% of households remain within one deviation from their original decile in SILC. For the first two and last two deciles the percentage of households that 'stay close' is around 89%. While some households do make large moves, the overall picture is one of relative stability.

TABLE 7: TRANSITION MATRIX FOR DECILES OF GROSS HOUSEHOLD INCOME: IPCAL VS SILC

deciles of gross household income in IPCAL	**deciles of gross household income in SILC**										
	1	2	3	4	5	6	7	8	9	10	**total**
1	396	126	50	10	6	5	3	5	5	4	**610**
2	150	302	96	30	16	8	2	3	0	2	**609**
3	27	142	301	89	36	8	1	3	1	2	**610**
4	12	19	107	284	135	35	7	4	2	4	**609**
5	9	11	29	123	250	122	40	14	3	9	**610**
6	4	4	12	33	115	255	128	30	17	11	**609**
7	4	1	5	13	24	114	262	138	41	8	**610**
8	2	2	4	20	11	23	105	267	150	25	**609**
9	4	1	4	2	13	17	41	99	292	137	**610**
10	2	1	2	5	4	22	21	46	99	407	**609**
Total	**610**	**609**	**610**	**609**	**610**	**609**	**610**	**609**	**610**	**609**	**6,095**

In general the fact that SILC gross incomes are *lower* than the ones declared in the fiscal forms, comes as a surprise to us. A priori we would have thought that tax evasion and fraud, combined with trustful revelation in survey settings, would have led to the opposite result. It is of course possible that the administrative fiscal data contains incomes which are delayed, and which should have been received in previous years but have actually been paid out in the year of data collection. But, if they have actually been received in 2009, then they should also be reported in the survey data as having been received in 2009. A preliminary conclusion is then that some suspicion about the quality of gross incomes in SILC is warranted. Of course, this first finding calls for more detailed research to find explanations for this difference. That is why we will decompose the difference in two directions. In the next subsections we look at the differences in gross income between SILC and IPCAL for *different income components*, but we stay at the household level. This will allow a more detailed view into the components that may be driving the differences shown in Table 6. In

section 3.2 we then leave the household level, and analyze more in detail the differences at the individual level.

3.1.2 Gross household labour market income

In Table 8 we show the difference between gross household *labour market income* in the tax return data (IPCAL) and the survey data (SILC). We find that the mean difference is quite substantial (€4592), meaning that gross labour incomes declared on the tax form exceed the SILC registered labour incomes by 16.8%. There is no significant difference in this deviation across regions. The dispersion, as measured by the standard deviation, is highest in Brussels and lowest in Wallonia. Further research is needed to explain this difference.

TABLE 8: AVERAGE DIFFERENCE IN GROSS HOUSEHOLD LABOUR MARKET INCOME: IPCAL VS SILC

	Belgium	**Flanders**	**Wallonia**	**Brussels**
mean income in € per year				
IPCAL	32,005	34,977	29,557	26,040
SILC	27,413	29,992	25,198	22,466
difference [IPCAL-SILC] in € per year:				
Mean in € **(in %)**	4,592 (16.8%)	4,985 (16.6%)	4,359 (17.3%)	3,574 (15.9%)
percentile 1	-27,656	-25,549	-28,272	-38,613
percentile 5	-9,910	-10,762	-9,348	-9,270
percentile 50 (median)	0	20	0	0
percentile 95	26,257	27,814	27,140	20,274
standard deviation	22,806	22,328	19,338	31,338
Number of observations	6,095	3,269	2,031	795

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

3.1.3 Gross household replacement income

Table 9 summarizes the results for gross household replacement income, which includes unemployment benefits, public pension benefits, and sickness and disability benefits. The mean differences are considerably less for replacement income than they are for total household income and labour market income. In the Brussels region the average difference is 'only' 400 Euro but the standard deviation is higher than in the other regions. Also here we notice that, on average, reported income in SILC is lower than fiscally declared income in IPCAL.

TABLE 9: AVERAGE DIFFERENCE IN GROSS HOUSEHOLD REPLACEMENT INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	11,550	11,943	11,734	9,469
SILC	10,338	10,805	10,085	9,069
difference [IPCAL-SILC] in € per year:				
Mean in € **(in %)**	1,212 (11.7%)	1,138 (10.5%)	1,649 (16.4%)	400 (4.4%)
percentile 1	-15,166	-15,166	-14,716	-22,170
percentile 5	-3,589	-3,379	-3,786	-3,674
percentile 50 (median)	155	257	253	0
percentile 95	9,725	9,432	10,779	8,036
standard deviation	20,258	20,827	16,002	26,616
Number of observations	6,095	3,269	2,031	795

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

3.1.4 Gross household income from (financial) assets

Income from (financial) assets includes mainly interests received and dividends. Table 10 shows the results for this income component at the household level. Since dividend and interest income is generally taxed at the source and does not need to be declared on the tax form, we know that this income component is 'underreported' in the administrative tax data. Yet, for some taxpayers it can be advantageous to have income from (financial) assets globalized with other income components and have them taxed jointly with the general personal income tax rate schedule. It is up to the taxpayer to decide whether or not to declare "income from (financial) assets"-items.[9]

[9] If they do declare non-compulsory items and it would turn out to be disadvantageous, the tax authorities will automatically apply the most favourable system for the taxpayer.

TABLE 10: AVERAGE DIFFERENCE IN GROSS HOUSEHOLD INCOME FROM (FINANCIAL) ASSETS: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	94	127	17	154
SILC	928	1,092	428	1,529
difference [IPCAL-SILC] in € per year:				
Mean in € (in %)	-834 (-73.3%)	-965 (-88.4%)	-411 (-96.0%)	-1,375 (-89.9%)
percentile 1	-20,000	-20,000	-5,952	-27,000
percentile 5	-2,700	-4,000	-1,950	-6,500
percentile 50 (median)	-125	-500	-5	-19
percentile 95	0	0	0	0
standard deviation	4,950	5,140	2,212	8,123
Number of observations	6,095	3,269	2,031	795

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

The non-compulsory reporting of income from financial assets in IPCAL naturally leads to the finding that this income component is higher in SILC than in IPCAL. Yet, we find that the average differences are surprisingly low. But, the dispersion is relatively high.

3.1.5 Gross household income from real estate

In the SILC data, income from real estate is described as income from rental of a property or land and thus applies mainly to rental income. In the personal income tax, the principal income from real estate that is taxable is the 'cadastral income', rents are not taxed. Cadastral incomes have not been updated since the 1970's and for tax purposes only a yearly updated index factor is applied. That means that for many (even most) properties, this 'cadastral income' is a rather arbitrary amount, with no close connection any more to the market value of the property. Moreover since 2005, this cadastral income is no longer registered in the fiscal data. We therefore conclude that the content of the variables is not really comparable between IPCAL and SILC, and hence a comparison is not very instructive.

3.2 Individual gross income components

In the previous subsection, we described gross income differences at the level of the household. We now turn to an analysis of income components at the *individual* level in an attempt to further clear out the differences in gross incomes in the two datasets.

Not unexpectedly, Table 11 confirms the conclusion from the previous subsection that, compared to the administrative data in IPCAL, SILC underreports gross individual incomes. The average difference for individual total gross income between IPCAL and SILC is €2,870 a year for

Belgium and is slightly higher in Wallonia and Flanders, while considerably lower in Brussels. The lower average difference in Brussels comes with a higher standard deviation however.

To investigate whether the 'underestimation' of gross incomes in SILC, compared to gross incomes, also holds for income components which are clearly defined and registered, such as employee income, we now zoom in on individual income components.

TABLE 11: AVERAGE DIFFERENCE IN TOTAL GROSS INDIVIDUAL INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
	mean income in € per year			
IPCAL	22,611	23,798	21,425	20,430
SILC	19,741	20,845	18,450	18,246
	difference [IPCAL-SILC] in € per year:			
Mean in € **(in %)**	2,870 (14.5%)	2,953 (12.2%)	2,975 (16.1%)	2,184 (12.0%)
percentile 1	-24,129	-23,778	-21,725	-34,457
percentile 5	-7,864	-7,752	-7,600	-9,804
percentile 50 (median)	880	998	782	514
percentile 95	16,677	17,723	16,445	14,841
standard deviation	21,582	21,377	17,526	30,909
Number of observations	11,642	6,394	3,879	1,369

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

3.2.1 Individual gross employee income

In this section we show results for those individuals aged 16 or more, whose sole source of income is from salaried employment. The income sources are identified by SILC data as mentioned above. If in the survey data an individual's personal total gross income equals his or her gross income from salaried employment, the individual is defined as an employee.[10]

10 This does not preclude the individual from having other income, such as income from real estate, or income from (financial) assets. These are income concepts that are recorded at the household level in SILC and excluded from the comparisons at the individual level.

TABLE 12: AVERAGE DIFFERENCE IN TOTAL GROSS INDIVIDUAL EMPLOYEE INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	30,918	31,788	29,651	30,023
SILC	29,342	30,140	28,083	28,790
difference [IPCAL-SILC] in € per year:				
Mean in € **(in %)**	1,576 (5.4%)	1,648 (5.5%)	1,568 (5.6%)	1,233 (4.3%)
percentile 1	-29,040	-24,359	-30,705	-46,421
percentile 5	-10,313	-10,049	-10,172	-12,126
percentile 50 (median)	1,760	1,767	1,756	1,741
percentile 95	13,152	12,515	13,821	13,731
standard deviation	8,840	8,267	9,089	10,787
Number of observations	4,544	2,607	1,431	506

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

Even for this well-defined and well registered gross income component, the finding is persistent: individual gross employee income is lower in SILC than the one declared in the fiscal form for the same period. The difference amounts to €1576 per year (or 5.4% of SILC income). We repeat that it is the topic of further research to try to explain this unexpected difference. Is it due to misreporting, miscomprehension, forgetfulness, or distrust toward 'authorities' and other privacy concerns? Or does it have to do with the specific way in which missing information on gross incomes has been imputed in the SILC-survey?

3.2.2 Individual gross self-employment income

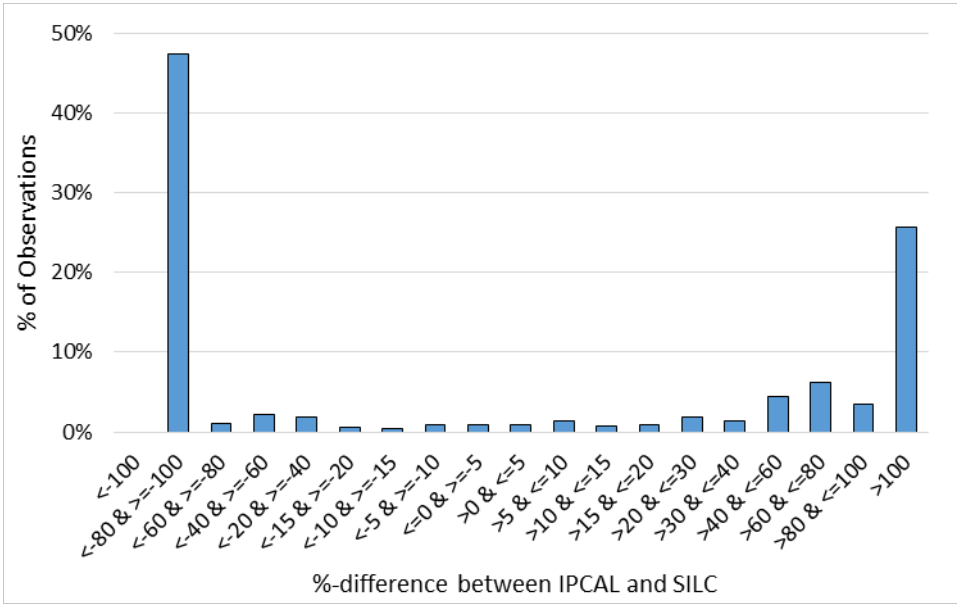
Table 13 shows average differences for self-employment income between IPCAL and SILC. Even for this income component, quite sensible to tax evasion and/or fraud, the fiscally declared gross income exceeds, on average the SILC registered one by €5241 of 20.5%. This case however, illustrates nicely that the average conceals a lot of the variation in the differences for individual tax payers. In Figure 4 we show the histogram of the percentage difference between the value in IPCAL and SILC (as % of SILC). The mode of the distribution is to the extreme left of the graph, indicating that for most individuals *SILC-incomes are higher than the incomes declared in IPCAL*: the deviation (IPCAL minus SILC) is *negative* and exceeds 100%. The fact that the average difference is nevertheless positive, reveals that the numerous negative differences must be compensated by still larger positive monetary in the cases where IPCAL income is larger than the SILC one.

TABLE 13: AVERAGE DIFFERENCE IN TOTAL GROSS INDIVIDUAL SELF-EMPLOYMENT INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	30,865	30,670	33,462	25,682
SILC	25,624	26,093	25,609	23,683
difference [IPCAL-SILC] in € per year:				
Mean in € **(in %)**	5,241 (20.5%)	4,577 (17.5%)	7,853 (30.7%)	1,999 (8.4%)
percentile 1	-104,316	-104,316	-93,884	-62,590
percentile 5	-46,142	-48,718	-43,813	-40,676
percentile 50 (median)	-3,537	-4,451	-450	-12,938
percentile 95	82,639	96,990	71,364	52,451
standard deviation	64,158	62,190	55,460	87,696
Number of observations	550	308	169	73

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

FIGURE 4: HISTOGRAM OF PERCENTAGE DIFFERENCE IN GROSS INDIVIDUAL SELF-EMPLOYMENT INCOME



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

Table 13 and Figure 4 clearly illustrate the potential for future analyses with these data. Compared to the administratively 'correct' data, the results show that many gross incomes for self-employed are 'overestimated' in SILC. This implies that tax liabilities based on those incomes might be overestimated by EUROMOD. Combined with the assumption that the

registered gross incomes in the survey data (SILC) are likely to be more accurate (due to underreporting in the administrative data) this certainly points into the direction of an underestimation of disposable incomes, the cornerstone of welfare and distributional analysis. In that case one could opt for a mixture of information from both sources using gross incomes from survey data and taxes paid from administrative data, thereby correcting disposable incomes to the upside and potentially eliminating (part of) the problem with poverty estimation.

3.2.3 Individual gross pension income

Table 14 shows that pension incomes are the first income component for which we find lower incomes in the fiscal data than in the survey data. On average the SILC-pension is €1338 or 7.9% higher than the IPCAL one. The distributional information however also shows that for more than half of the pensioners, the difference goes in the other direction, with IPCAL-values exceeding the SILC ones.

TABLE 14: AVERAGE DIFFERENCE IN TOTAL GROSS INDIVIDUAL PENSION INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	15,608	15,374	15,430	17,657
SILC	16,946	16,536	16,553	20,810
difference [IPCAL-SILC] in € per year:				
Mean in €	-1,338	-1,162	-1,123	-3,153
(in %)	(7.9%)	(7.0%)	(6.8%)	(15.2%)
percentile 1	-22,394	-22,235	-20,833	-56,134
percentile 5	-13,478	-13,959	-13,199	-12,172
percentile 50 (median)	178	167	150	435
percentile 95	7,806	6,335	8,673	14,233
standard deviation	12,530	7,505	6,886	33,929
Number of observations	2,207	1,291	707	209

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

3.2.4 Individual gross unemployment income

The same finding as for pension incomes, holds for the other major replacement income: unemployment benefits. In Table 15 the unemployment benefit registered in the IPCAL-data exceeds the survey average by €1276 or 10.7%. However, more research is needed to exploit the large variation in individual differences.

TABLE 15: AVERAGE DIFFERENCE IN TOTAL GROSS INDIVIDUAL UNEMPLOYMENT INCOME: IPCAL VS SILC

	Belgium	Flanders	Wallonia	Brussels
mean income in € per year				
IPCAL	10,686	11,877	9878	9715
SILC	11,962	13,113	11,277	10,759
difference [IPCAL-SILC] in € per year:				
Mean in € **(in %)**	-1,276 (-10.7%)	-1,236 (9.4%)	-1,399 (-12.4%)	-1,044 (9.7%)
percentile 1	-29,112	-29,112	-25,200	-2,880
percentile 5	-12,936	-11,537	-16,564	-10,905
percentile 50 (median)	4	55	-18	-50
percentile 95	3,641	4,528	3,147	3,391
standard deviation	6,534	6,237	6,725	6,830
Number of observations	688	287	294	107

Notes: the % difference is calculated as (IPCAL-SILC)/SILC

4. CONSTRUCTING UPDATED GROSS INCOMES

In order to try to remedy both the presupposed underestimation of disposable household income in EUROMOD and the underreporting of gross incomes in SILC in comparison to the IPCAL data, we can choose among different options. One possibility would be to assume that the administrative gross incomes are the 'correct' ones, and hence replace (some) gross incomes in SILC, by the information from IPCAL. A run of EUROMOD on these corrected gross incomes would then produce an updated disposable income distribution, which can be compared with the distribution of reported disposable incomes in SILC. This investigation is planned for future research.

In this section we report on another possibility, to wit the application of an iterative method developed by Immervoll and O'Donoghue (2001). A microsimulation model establishes a relationship between gross incomes and net (or disposable incomes) for a given tax-benefit structure. Immervoll and O'Donoghue (2001) invert this relationship by calculating gross incomes which correspond to given disposable incomes. Since the complexity of a real-world tax-benefit system makes it impossible to invert the relationship analytically, one searches for the gross incomes iteratively. Make an informed guess of gross income, run the microsimulation model to calculate the corresponding net income, and compare this net income with the reported net income. Based on that difference, adjust the first guess of gross income. In this way one produces a gross income distribution which, for a given tax benefit structure and for a given parameterisation of this structure in a tax-benefit model, is consistent with observed or reported disposable incomes. In this case, the assumptions are: 1) we accept that reported net incomes in SILC are more reliable than gross incomes, and 2) we take the accuracy of the microsimulation model, linking gross and net incomes, for granted. This iterative procedure is

described in section 4.1. In section 4.2 we then compare these new gross incomes with the ones reported in SILC.

4.1 Description of the iterative procedure

Since we use the Belgian SILC 2010, with incomes corresponding to 2009 we apply the 2009 policies as simulated in EUROMOD. We estimate or update gross incomes for seven different income components. For each of these components the SILC registration was done on a monthly basis: wage earned by an employee (py010g), income earned by a self-employed person (py050g), unemployment benefit (py090g), retirement pension (py100g), survivor pension (py110g), health benefit (py120g) and disability benefit (py140g).

The iterative method as presented by Immervoll & O'Donoghue (2001) is used to calculate new gross incomes starting from the net incomes as reported in the Belgium SILC 2010. This calculation takes place in two steps. Starting from the reported net incomes, we first calculate the amount of taxes paid, hereby making use of the withholding income tax schedule as simulated within EUROMOD. We use the withholding income tax (and not the final income tax) because we expect that individuals in BE-SILC report their monthly net income equal to the monthly amount they receive on their bank account, which is gross income minus social security contributions and minus withholding income taxes. We assume that a SILC-respondent does not take into account the final tax settlement at the end of the fiscal year when reporting his or her monthly net income. In that case, one should use the withholding income tax schedule and not the standard income tax schedule in EUROMOD, to calculate the original gross income. Another advantage of opting for the withholding tax, is that under this scheme all income components are taxed separately. This simplifies the calculation of gross incomes for people who combine different incomes in the same month. In a second step, we use the sum of the net incomes and withholding taxes to estimate the amount of social security contributions paid. Also this amount is calculated separately for each income component. The 'updated' gross incomes are then the sum of originally reported net incomes, newly calculated withholding taxes, and newly calculated social security contributions. Below we will use an asterisk (*) if we refer to variables with updated values. Since we to run the iterative method for each income component separately, total gross income is the sum of the different gross income components for those individuals who combine two or more income components.

The basic withholding tax spine as programmed in EUROMOD was useful for only four of the seven income components: wage earned by an employee (py010g), income earned by a self-employed person (py050g), retirement pension (py100g) and survivor pension (py110g). For the other three income components - unemployment benefit (py090g), health benefit (py120g) and disability benefit (py140g) -, special rules concerning the calculation of the withholding income tax apply. We have extended EUROMOD with these special rules:

- A withholding income tax of 10.09% is levied on monthly unemployment benefits (py090g). However, different categories of unemployed people are exempted from this

payment. Not enough information is available to disentangle the different categories in the BE-SILC.[11] We assume that only the unemployed who combine their unemployment benefit with an income from employment or self-employment have to pay the withholding income tax of 10.09% on their unemployment benefits. The other categories of unemployed are exempted from paying withholding income taxes.

- Persons receiving a sickness benefit (py120g) pay a withholding tax of 11.11%. When receiving a disability benefit (py140g), the withholding tax equals zero.

Let us now look into detail at the iterative procedure. In a first step, the personal income tax liability is calculated, starting from the reported net income of a specific income component:

Iteration procedure:

1. Set gross income (*grossY*) equal to original net income (*orig_netY*), as reported in the BE-SILC data, for that specific income component.

2. Use *grossY* to calculate simulated net income (*netY = grossY – withholding income tax*) through the withholding income tax schedule in EUROMOD for that specific income component.

3. Calculate the difference *dif = netY – orig_netY*.

4. Update gross income for this component by replacing *grossY = grossY - dif* for this income component. If the updated net income exceeds reported net income, *dif>0*, and we will adjust the previous guess of gross income downwards. The reverse applies if updated net income is lower than reported net income. In that case *dif<0*, and we adjust the guess of gross income upwards.

5. Repeat steps 2-4 until convergence.

These iterations continue until the absolute value of the difference (*dif*) between *netY* and *orig_netY* is smaller than €25. Only when this is the case for all household members[12], the family is removed from the iteration and the updated gross income (*grossY**) of each household member is written away in a new data file. After 23 iterations, *grossY** could be calculated in this way for 98.8% of the cases (14,520 of a total of 14,700 cases). That means that for nearly all individuals we could reconstruct gross income components such that, when applying the

11 For example detailed information about the unemployment duration is needed, which is not available within the BE-SILC database.

12 We drop individuals from our iteration once a correct gross income is calculated for all household members. Reason for this is that the amount of withholding tax paid by an individual also depends on the presence and income of a partner and the number of children in the household.

withholding tax scheme of EUROMOD, the resulting net income comes close to the reported net income in SILC within a margin of €25.

The main reason why the iterative procedure did not converge for the remaining 1.2% (or 180 cases) of the individuals was due to the way in which EUROMOD assigns the head of the household. In EUROMOD the household head is determined as the household member with the highest income. If there is a tie in terms of incomes, the oldest is selected to be the household head. All households for which the procedure did not converge, were households where, for a given income component, the two adults reported nearly identical net incomes (for example both partners earn a monthly net employee income of €1500). In that case, in the first iteration, the oldest person will be selected household head. Several tax credits in the withholding income tax are determined at the household level and hence are assigned to the head of the household. This reduces the withholding tax of the household head, which results in a higher net income. This updated net income will probably exceed original net income. Therefore, in step 4 of the first iteration, we will adjust gross income downwards. The partner of this person will face the opposite: he or she does not receive the tax credits determined at the household level. This causes the simulated net income to be lower than the reported original net income, and consequently we update gross income of the partner upwards before the start of the second iteration. As a result, the assignment of whom of the two partners is household head switches in the second iteration. But the story is repeated across different iterations and this explains why we the procedure does not converge for these households.

Given the cause of the problem, the solution is straightforward: we keep the household head fixed in each iteration. The iterative procedure described above remained almost the same. The only difference is that we lowered the gross income of the partner so that it is exactly one euro below the gross income of household head in each iteration. By doing this we are able to calculate an updated *grossY** for the remaining 180 cases of our dataset.

In the second step of the iterative procedure, we calculate the social security contributions, starting from the values of *grossY**.

Iteration procedure:

1. Set gross income (*grossY*) equal to *grossY**.
2. Use *grossY* to calculate simulated net income (*netY*) through the social security schedule in EUROMOD.
3. Calculate the difference *dif = netY – grossY**
4. Replace *grossY = grossY - dif*
5. Repeat steps 2-4 until convergence.

This iteration will continue until the absolute value of the difference (*dif*) is smaller than €5. Only when this is the case for all household members, the family drops out of the loop and the original gross income (*orig_grossY*) of every household member is written away in a new data file. Thanks to the fact that the calculation of social security contributions is purely individual in Belgium, we obtained convergence for every person in the dataset after maximum 5 iterations.

4.2 Comparison of reported and updated gross income components in SILC

4.2.1 Individual gross employee income

In Table 16, we compare the updated gross income for employees (py010g* in column 3) with the reported ones in BE-SILC 2010 (py010g in column 2). We give both the absolute difference in euro (column 4) and the percentage adjustment (column 5). We do this by decile of equivalised disposable household income. In the two rightmost columns we sketch the progressivity of the implied tax rates, defined as the difference between gross incomes and net incomes. Column 6 shows the implied tax rate when reported gross incomes are used. In column 7 we use the updated gross incomes.

TABLE 16: GROSS MONTHLY INCOME FROM EMPLOYMENT, REPORTED GROSS INCOME (PY010G) VS NEW UPDATED GROSS INCOME (PY010G*), BE-SILC2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Decile	**reported net income (py010n)**	**reported gross income (py010g)**	**updated gross income (py010g*)**	**absolute difference (3)-(2) in €**	**percentage difference (3)-(2) in % of (2)**	**implicit tax rate (2)-(1) in % of (2)**	**updated implicit tax rate (3)-(1) in % of (2)**
1	274	358	288	-70	-19.6	23.5	4.9
2	863	1,136	1,001	-135	-11.9	24.0	13.8
3	1,219	1,705	1,604	-101	-5.9	28.5	24.0
4	1,439	2,060	2,017	-43	-2.1	30.1	28.7
5	1,602	2,343	2,409	66	2.8	31.6	33.5
6	1,754	2,573	2,733	160	6.2	31.8	35.8
7	1,922	2,879	3,126	247	8.6	33.2	38.5
8	2,125	3,265	3,587	322	9.9	34.9	40.8
9	2,434	3,848	4,316	468	12.2	36.7	43.6
10	3,558	5,723	7,292	1,569	27.4	37.8	51.2
Total	1,718	2,588	2,836	248	9.6	33.6	39.4

Notes: 5618 observations;
deciles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

On average gross monthly income from employment is adjusted upwards with nearly €250, or about 10%. But the adjustment varies widely across the income distribution. For the bottom four deciles, gross employment incomes have to be adjusted downwards. For the upper half of the income distribution, we have to revise gross incomes upwards. In the highest decile e.g., the difference between py010g* and py010g is, on average, more than 1,500 euro or 27%.

If we buy the assumption that SILC-respondents report their net income more accurately, and that they perceive the question as gauging their net earnings after withholding tax, then this implies that respondents in the bottom deciles who also report a gross income *overestimate* the withholding tax and social security contributions paid they have paid. On the contrary, in this interpretation, individuals in higher deciles seem to *underestimate* the taxes and contributions paid. Since we have to revise their gross employment income upwards, the implied amount of taxes will, for a given net income, be higher.

Anyhow, the last two columns show how important the adjustment of gross incomes is when one uses the data to draw conclusions about the progressivity of the tax system. Implictly, accepting the correction or update of gross incomes as described above, and leaving the disposable incomes untouched boils down to the introduction of a way more progressive tax system (in this case on employment income). The implicit tax rate in the top decile increases from 38 to 51%.

4.2.2 Individual gross self-employment income

Table 17 shows the adjustment for gross incomes from self-employment. Across the whole income distribution, we have to update the gross incomes from self-employment (py050g*) to match the reported disposable incomes. In the top decile we even have to add more than €4,300 or nearly 70%. In this case, it seems quite evident that the assumption of an accurate calculation of tax liabilities by EUROMOD is not really tenable. The upward adjustment of gross incomes will certainly also have to do with the fact that for self-employed, tax liabilities might be overestimated by EUROMOD.

TABLE 17: GROSS MONTHLY INCOME FROM SELF-EMPLOYMENT, REPORTED GROSS INCOMES (PY050G) VS UPDATED GROSS INCOMES (PY050G*), BE-SILC 2010

Decile	(1) **reported net income (py050n)**	(2) **reported gross income (py050g)**	(3) **updated gross income (py050g*)**	(4) **absolute difference (3)-(2) in €**	(5) **percentage difference (3)-(2) in % of (2)**	(6) **implicit tax rate (2)-(1) in % of (2)**	(7) **updated implicit tax rate (3)-(1) in % of (2)**
1	100	121	358	237	195.9	17.4	72.1
2	499	665	822	157	23.6	25.0	39.3
3	885	1,113	1,354	241	21.7	20.5	34.6
4	1,199	1,534	1,885	351	22.9	21.8	36.4
5	1,436	1,820	2,325	505	27.7	21.1	38.2
6	1,627	2,149	2,954	805	37.5	24.3	44.9
7	1,868	2,413	3,456	1,043	43.2	22.6	45.9
8	2,310	2,953	4,588	1,635	55.4	21.8	49.7
9	2,877	3,878	6,005	2,127	54.8	25.8	52.1
10	4,834	6,361	10,717	4,356	68.5	24.0	54.9
Total	1,703	2,212	3,298	1,086	49.1	23.0	48.4

Notes: 700 observations;
deciles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.
The high tax rate in the first decile is explained by the fact that self-employed people have to pay a minimum social security contribution of 2,601.36 euro per year in 2009, independent of their income (only if self-employment is the main activity, if not lower contributions have to be paid). If e.g. someone reports a net income from self-employment of 100 euro per month , we know that he had to pay a social security contribution of 216.78 euro per month (2,601.36 /12) so his gross income was than equal to 316.78 euro.

4.2.3 Individual gross pension income

Looking at the old age pensions (Table 18), the updated gross incomes (py100g*) are lower in the bottom eight income deciles compared to the registered ones (py100g). This picture changes in the two highest deciles. The explanation for the low implicit tax rate in the bottom part of the distribution is twofold: old age pensioners are eligible for a substantial tax credits (2,202 euro per year in 2009) and they do not have to pay social insurance contributions if their monthly gross income is lower than 1,281 euro per month in 2009[13].

13 If they earn more, they have to pay a limited social insurance contribution of 3.55% of their total gross income.

TABLE 18: GROSS MONTHLY INCOME FROM AN OLD-AGE PENSION, REPORTED GROSS INCOMES (PY100G) VS UPDATED GROSS INCOMES (PY100G*), BE-SILC 2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Decile	**reported net income (py100n)**	**reported gross income (py100g)**	**updated gross income (py100g*)**	**absolute difference (3)-(2) in €**	**percentage difference (3)-(2) in % of (2)**	**implicit tax rate (2)-(1) in % of (2)**	**updated implicit tax rate (3)-(1) in % of (2)**
1	280	302	280	-22	-7.3	7.3	0.0
2	798	839	798	-41	-4.9	4.9	0.0
3	961	1,014	961	-53	-5.2	5.2	0.0
4	1,038	1,108	1,039	-69	-6.2	6.3	0.1
5	1,115	1,204	1,119	-85	-7.1	7.4	0.4
6	1,212	1,344	1,262	-82	-6.1	9.8	4.0
7	1,317	1,507	1,427	-80	-5.3	12.6	7.7
8	1,484	1,778	1,767	-11	-0.6	16.5	16.0
9	1,704	2,093	2,273	180	8.6	18.6	25.0
10	2,738	3,277	4,726	1,449	44.2	16.4	42.1
Total	1,259	1,438	1,557	119	8.3	12.4	19.1

Notes: 2423 observations;
deciles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

The same picture appears when looking at the survivor pensions (Table 19).

TABLE 19: GROSS MONTHLY INCOME FROM A SURVIVOR PENSION, REPORTED GROSS INCOMES (PY110G) VS UPDATED GROSS INCOMES (PY110G*), BE-SILC 2010

Decile	(1) **reported net income (py110n)**	(2) **reported gross income (py110g)**	(3) **updated gross income (py110g*)**	(4) **absolute difference (3)-(2) in €**	(5) **percentage difference (3)-(2) in % of (2)**	(6) **implicit tax rate (2)-(1) in % of (2)**	(7) **updated implicit tax rate (3)-(1) in % of (2)**
1	743	760	743	-17	-2.2	2.2	0.0
2	1,110	1,161	1,123	-38	-3.3	4.4	1.2
3	1,467	1,652	1,666	14	0.8	11.2	11.9
Total	1,099	1,183	1,169	-14	-1.2	7.1	6.0

Notes: 103 observations;
Because of the limited number of observations with a survivor pension (n = 103), we decided to look at income tertiles (based on total equivalised disposable household income) instead of income deciles.
Tertiles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

4.2.4 Individual gross unemployment income, health and disability benefits

In Table 20 (unemployment benefit), Table 21 (health benefit) and Table 22 (disability benefit), we look at the cases with special rules concerning the calculating of the withholding income tax (see above). Looking at the unemployment benefit, we assume that only the unemployed who combine their unemployment benefit with an income from employment or self-employment have to pay a withholding tax of 10.09%. No social security contribution has to be paid on unemployment benefits.

In this case, the adjustment which has to be made is quite small (1.1%). Hence, the regressive pattern of taxation of unemployment benefits also stays intact. Individuals with a low unemployment benefit often combine the benefit with income from employment, and thus have to pay a withholding income tax on their unemployment benefit. Individuals with a higher (often full time) unemployment benefit do not combine this income with an income from employment and do not pay a withholding income tax.

Individuals receiving a sickness benefit have to pay a withholding tax of 11.11% and also limited social security contributions. Updated gross incomes (py120g*) are about 10% higher than the net incomes (py120n) in all income tertiles.

Individuals receiving a disability benefit (py140g) don't have to pay a withholding income tax or social security contributions. Naturally, updated gross income (py140g*) equals original net income (py140n) in the different income tertiles. The old gross incomes (py140g) produced some weird results for which no explanation has been found.

TABLE 20: GROSS MONTHLY INCOME FROM UNEMPLOYMENT BENEFIT, REPORTED GROSS INCOMES (PY090G) VS UPDATED GROSS INCOMES (PY090G*), BE-SILC 2010

Decile	(1) reported net income (py090n)	(2) reported gross income (py090g)	(3) updated gross income (py090g*)	(4) absolute difference (3)-(2) in €	(5) percentage difference (3)-(2) in % of (2)	(6) implicit tax rate (2)-(1) in % of (2)	(7) updated implicit tax rate (3)-(1) in % of (2)
1	61	66	66	0	0.0	7.6	7.6
2	126	136	138	2	1.5	7.4	8.7
3	184	193	199	6	3.1	4.7	7.5
4	277	290	296	6	2.1	4.5	6.4
5	369	387	392	5	1.3	4.7	5.9
6	491	503	522	19	3.8	2.4	5.9
7	703	729	731	2	0.3	3.6	3.8
8	866	882	893	11	1.2	1.8	3.0
9	963	979	977	-2	-0.2	1.6	1.4
10	1,293	1,313	1,333	20	1.5	1.5	3.0
Total	525	540	546	6	1.1	2.8	3.8

Notes: 1186 observations;
deciles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

TABLE 21: GROSS MONTHLY INCOME FROM A HEALTH BENEFIT, REPORTED GROSS INCOMES (PY120G) VS UPDATED GROSS INCOMES (PY120G*), BE-SILC 2010

Decile	(1) reported net income (py120n)	(2) reported gross income (py120g)	(3) updated gross income (py120g*)	(4) absolute difference (3)-(2) in €	(5) percentage difference (3)-(2) in % of (2)	(6) implicit tax rate (2)-(1) in % of (2)	(7) updated implicit tax rate (3)-(1) in % of (2)
1	110	119	122	3	2.5	7.6	9.8
2	383	406	425	19	4.7	5.7	9.9
3	1,006	1,055	1,118	63	6.0	4.6	10.0
Total	496	523	551	28	5.4	5.2	10.0

Notes: 182 observations;
Because of the limited number of observations, we decided to look at income tertiles (based on total equivalised disposable household income) instead of income deciles.
Tertiles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

TABLE 22: GROSS MONTHLY INCOME FROM A DISABILITY BENEFIT, REPORTED GROSS INCOMES (PY140G) VS UPDATED GROSS INCOMES (PY140G*), BE-SILC 2010

Decile	(1) **reported net income (py140n)**	(2) **reported gross income (py140g)**	(3) **updated gross income (py140g*)**	(4) **absolute difference (3)-(2) in €**	(5) **percentage difference (3)-(2) in % of (2)**	(6) **implicit tax rate (2)-(1) in % of (2)**	(7) **updated implicit tax rate (3)-(1) in % of (2)**
1	14	50	14	-36	-72.0	72.0	0.0
2	40	61	40	-21	-34.4	34.4	0.0
3	175	23	177	154	669.6	-660.9	1.1
Total	76	44	76	32	72.7	-72.7	0.0

Notes: 182 observations;
Because of the limited number of observations, we decided to look at income tertiles (based on total equivalised disposable household income) instead of income deciles.
Tertiles based on equivalised disposable household income.
The implicit tax rates in columns (6) and (7) are calculated on the basis of the decile averages in columns (1) to (3), and are not the averages of individual tax rates.

4.3 Poverty and inequality outcomes with new gross incomes

In this section we investigate the impact on poverty and inequality when using the update gross incomes as input for the calculation of disposable income in EUROMOD. It might come as a surprise that a calibration of gross incomes on reported net incomes, would not lead to an exact replication of the poverty and inequality measures on these reported net incomes. There are several reasons for a remaining gap between poverty incidence as measured on reported disposable incomes, and measured on disposable incomes simulated by EUROMOD. First, we remind that we apply the procedure of updating gross incomes at the level of singular income components. Secondly, we use the withholding tax for the calibration, but the final tax settlement may of course differ, and hence produce a different disposable income.

Table 23 displays poverty rates calculated for different poverty thresholds and using three different income concepts: reported net income in SILC, EUROMOD simulated disposable income, based on SILC gross incomes, and EUROMOD simulated disposable income, based on the updated gross incomes.

TABLE 23: POVERTY RATES (% OF INDIVIDUALS) AT DIFFERENT POVERTY THRESHOLDS FOR THREE DIFFERENT INCOME CONCEPTS

	(1)	(2)	(3)	(4)	(5)
	Reported SILC incomes	**EUROMOD using SILC-gross incomes**	**Ratio (2)/(1)**	**EUROMOD using updated gross incomes**	**Ratio (4)/(1)**
			below 40% of median equivalent income		
All	4.13	1.91	0.46	3.89	0.94
Females	4.46	1.89	0.42	3.91	0.88
Males	3.78	1.93	0.51	3.88	1.03
			below 50% of median equivalent income		
All	7.78	4.87	0.63	7.91	1.02
Females	8.08	5.01	0.62	8.06	1.00
Males	7.46	4.70	0.63	7.75	1.04
			below 60% of median equivalent income		
All	14.56	10.25	0.70	15.41	1.06
Females	15.18	10.93	0.72	16.33	1.08
Males	13.92	9.55	0.69	14.47	1.04
			below 70% of median equivalent income		
All	23.75	19.16	0.81	25.31	1.07
Females	25.57	20.62	0.81	26.93	1.05
Males	21.88	17.65	0.81	23.64	1.08

Using the old gross incomes, we notice a (severe) underestimation of the poverty rates compared to the poverty rates based on the net incomes as reported in the SILC. The gap decreases (or the ratio comes closer to 1) the higher the poverty line is defined, but a considerable gap of about 20% remains even at the 70% threshold. Updating gross incomes along the lines explained above, reduces the underestimation of the poverty rate at the 40% threshold to a 6%-point gap. With higher poverty lines, the updated gross incomes even (slightly) overestimate the poverty incidence compared to the estimate based on reported SILC incomes (ranging from 2% using the 50% threshold to 7% using the 70% threshold).

Table 24 shows that the discrepancy between the poverty figures (measured as 60% of median equivalised disposable household income) based on simulated EUROMOD incomes and reported SILC incomes is not evenly distributed over age categories. Using the reported gross incomes, we notice a serious underestimation of poverty in EUROMOD relative to the SILC figures up to 42% for children, while in EUROMOD poverty among the older working cohorts and pensioners deviates only with 21%. Using the updated gross incomes, there is still an underestimation of poverty in EUROMOD in comparison with SILC figures when looking at children and young adults. But after the adjustment, EUROMOD overestimates poverty rates in comparison to SILC for older working cohorts and pensioners, with respectively 34% and 21%.

TABLE 24: POVERTY RATES BY AGE GROUPS FOR THREE DIFFERENT INCOME CONCEPTS

Age group:	(1) Reported SILC incomes	(2) EUROMOD using SILC-gross incomes	(3) Ratio (2)/(1)	(4) EUROMOD using updated gross incomes	(5) Ratio (4)/(1)
0-15	18.46	10.63	0.58	15.07	0.82
16-29	14.30	10.16	0.71	13.37	0.93
30-44	11.39	7.66	0.67	10.92	0.96
45-64	11.62	9.01	0.78	15.54	1.34
65+	19.48	15.35	0.79	23.62	1.21

Note: poverty line is 60% of median equivalised disposable household income

In addition to the underestimation of poverty figures by EUROMOD, when using SILC gross incomes, EUROMOD also underestimates inequality indicators. In Table 25 we show how both the Gini-indicator and the S80/S20 measure are lower when using the reported EUROMOD gross incomes. Equivalised average income per decile reveals that this underestimation of inequality is mainly driven by divergence in the ends of the income distribution. Disposable income in the first decile is on average 14% higher when simulated by EUROMOD compared to the reported SILC incomes. In the 10^{th} decile however, disposable income in EUROMOD is on average 14% lower in comparison to the reported SILC incomes.

Using the updated gross incomes both the Gini-indicator and the S80/S20 measure are now (slightly) higher when simulated by EUROMOD compared to the results based on the reported SILC incomes. For all deciles average equivalised disposable income simulated by EUROMOD is now quite close to the one reported by SILC incomes.

TABLE 25: INEQUALITY INDICATORS AND AVERAGE EQUIVALISED INCOME PER DECILE

	(1)	(2)	(3)	(4)	(5)
	Reported SILC incomes	**EUROMOD using SILC-gross incomes**	**Ratio (2)/(1)**	**EUROMOD using updated gross incomes**	**Ratio (4)/(1)**
GINI	0.26	0.22	0.85	0.27	1.04
Income quintile ratio (S80/S20)	3.57	3.11	0.87	3.99	1.12
Average income per decile of equivalised disposable income (reported)					
1	7,696	8,748	1.14	7,445	0.97
2	11,717	12,407	1.06	11,692	1.00
3	13,980	14,298	1.02	13,668	0.98
4	16,183	16,184	1.00	15,727	0.97
5	18,384	18,087	0.98	18,280	0.99
6	20,625	19,867	0.96	20,823	1.01
7	23,110	21,945	0.95	23,487	1.02
8	26,085	24,430	0.94	26,680	1.02
9	30,166	27,862	0.92	31,075	1.03
10	44,100	38,067	0.86	45,297	1.03
Mean income (unequivalised)	40,903	38,396	0.94	40,905	1.00
Mean income (equivalised)	21,201	20,187	0.95	21,413	1.01
Median income (equivalised)	19,469	18,927	0.97	19,588	1.01

We conclude that it is possible to bring poverty and inequality indicators produced by EUROMOD closer to the external statistics, based on reported net incomes in SILC. Using the updated gross incomes as input for the EUROMOD calculations produces estimates of disposable income which are close to the reported SILC ones for all income deciles. Also the poverty rates and inequality indicators are closer to the ones based on the reported SILC incomes. For some specific age and income groups, we now tend to slightly overestimate the poverty rates and inequality indicators in comparison to the poverty rates based on the reported SILC incomes.

5. ARE NEW CALCULATED GROSS INCOMES CLOSER TO GROSS INCOMES FROM THE IPCAL-DATASET?

In the final part of this paper, we compare the updated gross incomes again with the IPCAL data and check whether using these new gross incomes (in comparison to the reported ones) improves the correspondence between the SILC and the IPCAL datasets. Gross individual incomes from the IPCAL dataset are respectively compared with gross individual incomes from

the reported SILC dataset ("SILC reported") and with the updated gross incomes ("SILC adjusted").

5.1 Individual total gross income

In Table 26 summary statistics are shown for total individual gross income. The results shown are for the entire population. The average difference for individual gross income is substantially lower than that for total household gross income at €2,870 a year for SILC-reported and €1,173 for the SILC-adjusted data. The latter might imply a better match with the administrative data but it comes at the cost of a considerably higher dispersion as shown by the more than 27% increase in standard deviation for SILC-adjusted data.

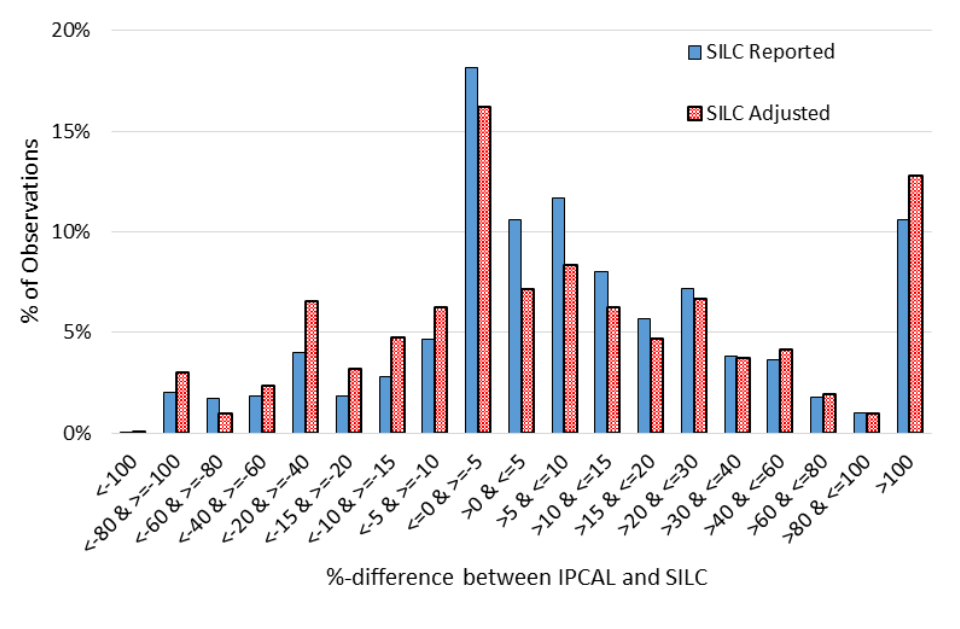
TABLE 26: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS INCOME: IPCAL VS SILC AND SILC-ADJUSTED (IN EURO'S)

	IPCAL-SILC reported	IPCAL-SILC adjusted
mean difference (in €)	2,870	1,173
percentile 1	-24,129	-39,859
percentile 5	-7,864	-12,886
percentile 50 (median)	880	511
percentile 95	16,677	15,231
standard deviation	21,582	27,521
total obs.	11,642	11,642

Figure 5 shows the changes in the distribution of percentage deviations between SILC and IPCAL. It seems as if the adjustment process has reduced the number of observations where IPCAL gross incomes exceed the SILC reported ones (except for largest deviations); whereas the number of negative differences (IPCAL below SILC) has increased.

Table 26 and Figure 5 above showed results for the entire sample of individuals who had a matching record in the IPCAL data. In the next subsections we will look at subsamples based on the income concept being analysed.

FIGURE 5: DISTRIBUTION OF PERCENTAGE DIFFERENCES BETWEEN IPCAL AND REPORTED AND ADJUSTED INCOMES FROM SILC FOR TOTAL GROSS INCOME OF INDIVIDUALS



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

5.2 Individual gross employee income

In this section we show results for individuals aged 16 or more whose sole source of income is from salaried employment. The income sources are identified by SILC data as mentioned above: if in the survey data an individual's personal total gross income equals his or her gross income from salaried employment, the individual is said to be an employee.[14]

The mean monetary differences in Table 27are similar, though somewhat smaller in absolute terms when using original SILC data versus adjusted data. While the average difference was positive for the original data, it actually becomes negative after gross incomes have been recalibrated, implying that original SILC employee income now rises above the IPCAL counterpart. This apparently results in larger average differences for the observations where survey income is larger than administrative income. The iterative methodology uses the withholding tax scheme rather than the entire tax-benefit system to impute gross labour market incomes. Given that most of the tax and income deductions not accounted for in the withholding tax scheme are also not modelled in EUROMOD, the difference must come from

14 This does not preclude the individual from having other income, such as income from real estate, or income from (financial) assets. These are income concepts that are recorded at the household level in SILC and excluded from the comparisons at the individual level.

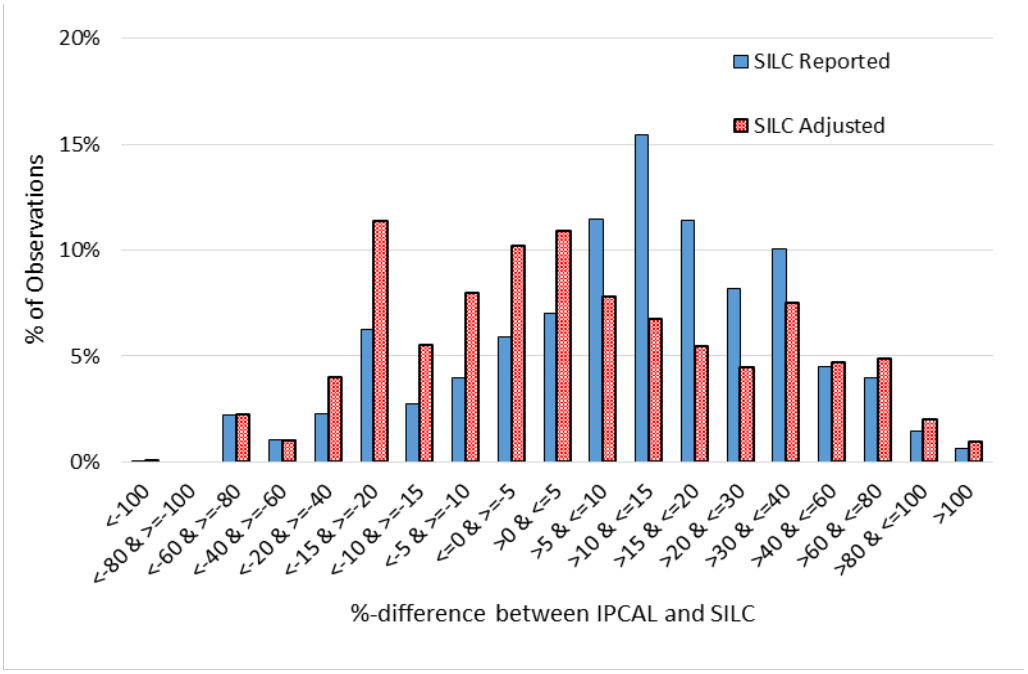
other sources. One of those might be the higher tax rates applied in the withholding tax scheme as compared to the final personal income tax rate schedule. For a net-to-gross income imputation, higher tax rates will result in higher gross incomes, ceteris paribus.

TABLE 27: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS EMPLOYEE INCOME IN EURO AND PERCENTAGE: IPCAL VS SILC AND SILC-ADJUSTED

	IPCAL-SILC reported	IPCAL-SILC adjusted
mean difference (in €)	1,576	-1,871
percentile 1	-29,040	-54,298
percentile 5	-10,313	-18,411
percentile 50 (median)	1,760	-411
percentile 95	13,152	11,005
standard deviation	8,840	13,043
total obs.	4544	4395

The distributions of percentage differences in Figure 6 illustrates that the adjustment is not symmetric. The relative frequency of individuals where IPCAL gross incomes are lower than the SILC ones increases considerably.

FIGURE 6: DISTRIBUTION OF PERCENTAGE DIFFERENCES BETWEEN IPCAL AND REPORTED AND ADJUSTED EMPLOYMENT INCOMES FROM SILC



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

5.3 Individual gross self-employment income

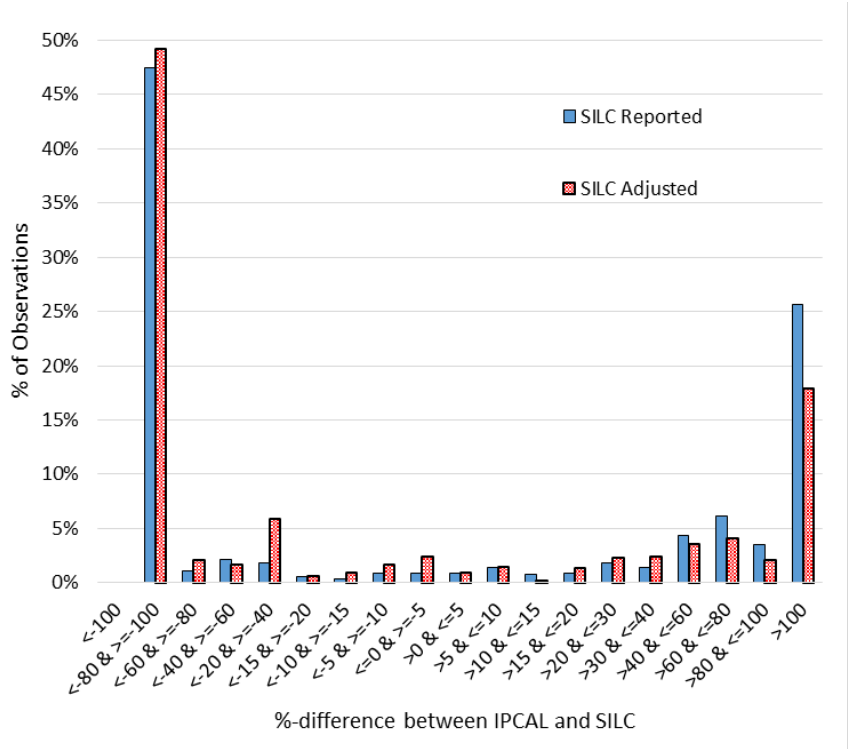
Table 28 shows that for self-employment income, the average difference with IPCAL was much smaller when one uses the original SILC-data , than with the adjusted SILC-data. Also the dispersion increases with the adjusted data. The peculiarity of self-employment income becomes clear when one notices that differences are quite substantial for some observations. About 1% of the individuals have gross income from self-employment which, in the administrative data, is more than €100,000 lower than in the original survey data. This quantile value for the largest negative difference now even doubles to more than €200,000 in the adjusted survey data.

TABLE 28: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS SELF-EMPLOYMENT INCOME IN EURO AND PERCENTAGE: IPCAL VS SILC AND SILC-ADJUSTED

	IPCAL-SILC reported	IPCAL-SILC adjusted
mean difference (in €)	5,241	-9,211
percentile 1	-104,316	-206,990
percentile 5	-46,142	-79,409
percentile 50 (median)	-3,537	-11,440
percentile 95	82,639	70,204
standard deviation	64,158	73,061
total obs.	550	532

Figure 7 illustrates that the re-calibration by means of the iterative procedure does not really change the original picture that, for self-employment income, there are both very large positive, and very large negative differences between IPCAL and SILC.

FIGURE 7: DISTRIBUTION OF PERCENTAGE DIFFERENCES BETWEEN IPCAL AND REPORTED AND ADJUSTED SELF-EMPLOYMENT INCOMES FROM SILC



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

5.4 Individual gross pension income

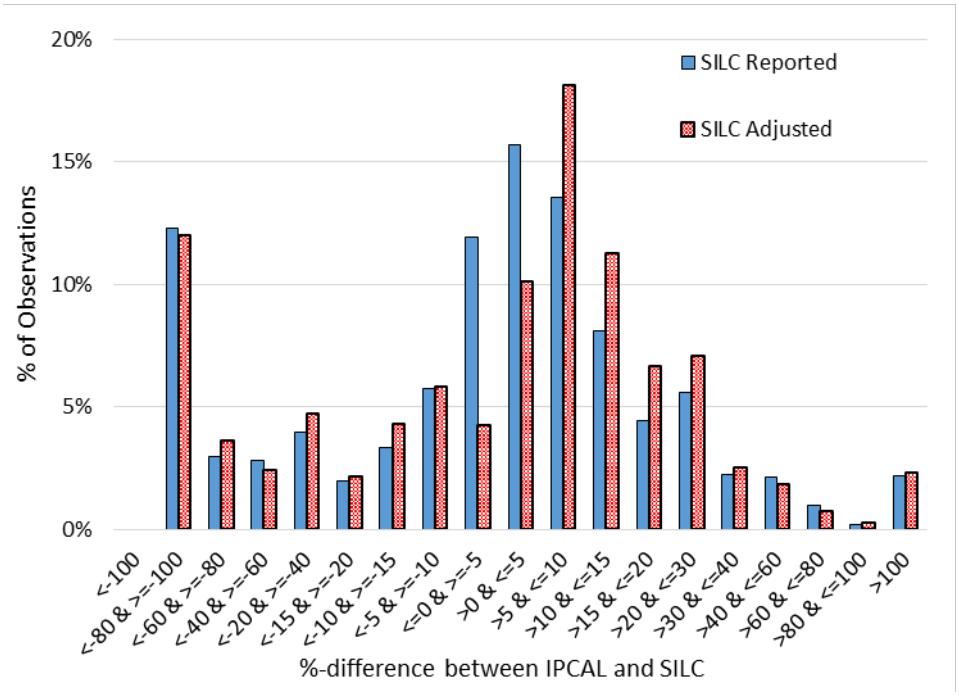
As for pension incomes, one can see from Table 29 that the average gross pension income in the original SILC data was, on average, closer to the administrative gross pension income than after the adjustment. The dispersion in the adjusted data stands at more than double that in the original SILC data. While the median gross pension income difference is positive, the average is clearly negative. This result is driven by the relatively large concentration of individuals at the negative extremes where the monetary differences outweigh the fact that more than half of the individuals have a gross pension income in IPCAL that is larger than in (adjusted) SILC.

TABLE 29: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS PENSION INCOME IN EURO AND PERCENTAGE: IPCAL VS SILC AND SILC-ADJUSTED

	IPCAL-SILC reported	IPCAL-SILC adjusted
mean difference (in €)	-1,338	-2,439
percentile 1	-22,394	-33,833
percentile 5	-13,478	-13,974
percentile 50 (median)	178	539
percentile 95	7,806	5,934
standard deviation	12,530	27,637
total obs.	2,207	2,249

Figure 8 shows how the adjustment procedure has mainly shifted away mass from the smaller positive differences, to much larger ones. Probably, where the IPCAL gross pension income already exceeded the SILC reported one, the gross pension income in SILC has been revised downwards.

FIGURE 8: DISTRIBUTION OF PERCENTAGE DIFFERENCE BETWEEN IPCAL AND REPORTED AND ADJUSTED PENSION INCOMES FROM SILC



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

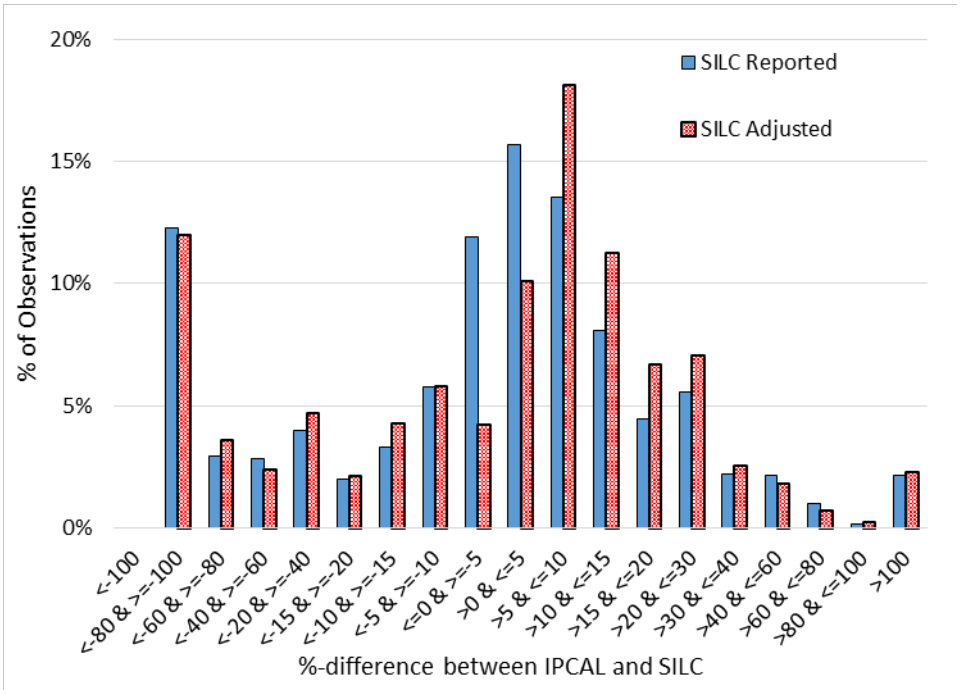
5.5 Individual gross survivor pension income

A comparison of survivor pension incomes can be found inTable 30. Here average income differences are quite similar in the two SILC datasets (original and adjusted). The dispersion and the extremes are again more outspoken in the adjusted data than in the original SILC data. Figure 9 is very similar to Figure 8.

TABLE 30: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS SURVIVOR PENSION INCOME IN EURO AND PERCENTAGE: IPCAL VS SILC AND SILC-ADJUSTED

	IPCAL-SILC reported	IPCAL-SILC adjusted
mean difference (in €)	-1,004	-1,094
percentile 1	-24,000	-36,253
percentile 5	-15,215	-14,964
percentile 50 (median)	8	595
percentile 95	3,848	4,614
standard deviation	5,567	7,187
total obs.	72	70

FIGURE 9: DISTRIBUTION OF PERCENTAGE DIFFERENCES BETWEEN IPCAL AND REPORTED AND ADJUSTED SURVIVOR PENSION INCOMES FROM SILC



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

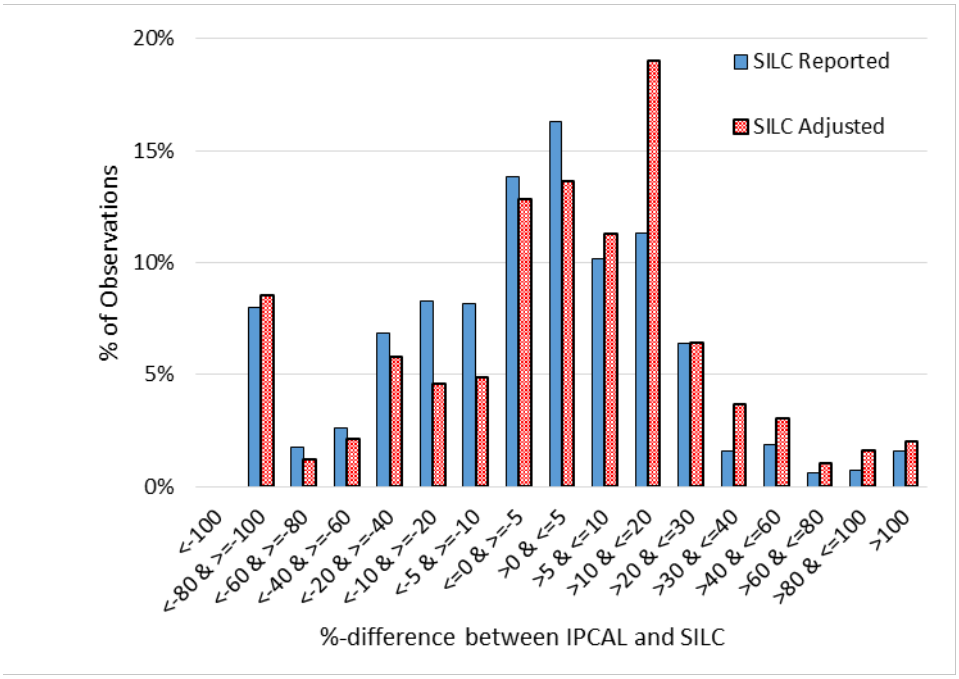
5.6 Individual gross unemployment income

Finally, we show the effect of the adjustment of unemployment income in Table 31. The average difference between gross unemployment income in IPCAL and the survey data is substantially reduced after the recalibration of gross unemployment income. Again the median difference is positive and much larger in the adjusted data than in original SILC data, while the average difference is negative. Large negative extremes seem to drive this result. Indeed, around 8% of individuals have gross unemployment income in the (adjusted) survey data which is more than 95% higher than in the administrative data. At the other extreme this percentage drops to around 2% with a positive difference of more than 100%, obviously not enough to cancel the negative outliers. From the median positive difference we know that more than half of the observations have gross unemployment income in the administrative data that is larger than in the (adjusted) survey data and a large part of these are concentrated in the center of the percentage difference distribution.

TABLE 31: AVERAGE DIFFERENCE IN INDIVIDUAL TOTAL GROSS UNEMPLOYMENT INCOME IN EURO AND PERCENTAGE: IPCAL VS SILC AND SILC-ADJUSTED

	IPCAL-SILC reported	**IPCAL-SILC adjusted**
mean difference (in €)	-1,276	-419
percentile 1	-29,112	-20,940
percentile 5	-12,936	-12,600
percentile 50 (median)	4	327
percentile 95	3,641	4,619
standard deviation	6,534	5,568
total obs.	688	654

FIGURE 10: DISTRIBUTION OF PERCENTAGE DIFFERENCES BETWEEN IPCAL AND REPORTED AND ADJUSTED UNEMPLOYMENT INCOMES FROM SILC



Note: the percentage difference is calculated with respect to the value of reported income in SILC.

6. CONCLUSION

In this paper, we have tried to shed light on the similarity between gross incomes in administrative data versus survey data. We did this for a unique sample of individuals for whom we had both administrative as well as survey data. For each of the individuals in the SILC data, information was requested from administrative tax data (IPCAL), resulting in a dataset with both survey data and tax return data for the same individuals.

This combined dataset gives a unique opportunity to analyse administrative versus survey income data. Whereas the exercise is useful and informative in and of itself, one of the main underlying reasons was also that it was found that poverty and inequality statistics for Belgium calculated with EUROMOD did not correspond with official statistics. One can think of multiple causes for this discrepancy, one of which is that gross incomes in survey data are incorrectly reported. From the tables in the text it is safe to say that gross income in administrative tax data is generally higher than in survey data and this holds true for most of the separate income components analysed. These differences can be substantial, certainly for specific income components, such as self-employed income.

To mitigate this 'underreporting of gross incomes' in the SILC, we applied an iterative procedure for a net-to-gross imputation based on EUROMOD, assuming that net/disposable incomes reported in SILC are more accurate than gross incomes. Belgium is a good test case for this iterative procedure for two reasons. Firstly, Belgium has a compulsory monthly withholding income tax. Thus we can use the withholding income tax schedule as implemented within EUROMOD to calculate a new gross income for each individual in our dataset. Secondly, the different income components within the withholding income tax in Belgium are taxed separately, which makes it a lot easier to calculate new gross incomes for different income components. In theory, it should also be possible to implement the same method using the end of the year taxation. But in practice, we expect that the joint taxation of the different income components would make it more difficult to calculate the new gross income components for each household member in the same iteration.

When using the updated gross incomes in EUROMOD, we get poverty and inequality estimates which correspond better to official statistics. Meanwhile, comparing the recalibrated gross incomes with the IPCAL data yields mixed results. The differences in gross incomes that exist between survey data and administrative data do not become smaller.

Therefore, further analysis into the possible causes of differences in gross incomes between administrative and survey data is certainly warranted. Insights from this may lead to an improvement of imputation methods. One could consider using administrative data directly in EUROMOD, both at the level of gross incomes and at the level of tax liabilities calculated by the tax authorities and present in the IPCAL-dataset. This could shed light on the accuracy of the personal income tax calculation in EUROMOD, given that many detailed aspects of the personal income tax system, such as fiscal expenditures, are not modelled in EUROMOD. We leave this for future research.

7. REFERENCES

Decancq, K. & Decoster, A. & Spiritus, K. & Verbist, G. (2012), *Mefisto: a new micro-simulation model for Flanders*. Flemosi Discussion paper 14.

Decoster A. & De Swerdt, K. (2013), *Fantasi: een microsimulatiemodel voor de personenbelasting op de IPCAL-data.* Steunpunt Fiscaliteit en Begroting, Spoor A3b1, working paper.

Hufkens, T. & Spiritus, K. & Vanhille, J. (2013), *EUROMOD country report. Belgium: 2009-2012.*

Immervoll, H. & O'Donoghue, C. (2001), *Imputation of gross amounts from net incomes in household surveys. An application using EUROMOD.* Essex: EUROMOD working paper no. EM1/01.