

Steering Committee BELMOD

29 October 2020

## Importance of Administrative Data

Comparing the personal income tax in survey-based Euromod with an administrative data-based model for Belgium

A. Decoster, L. De Sadeleer, T. Vanheukelom and S. Van Houtven

Department of Economics - KU Leuven

# 1. introduction – research question

- increased **access to** and **focus on** administrative data
  - specialized data sources (e.g. finance and social security administrations)
  - cross-over between survey and administrative data (e.g. plans EU-SILC)
  - detailed population data
- but, hurdles (= **costs**) to increased use of administrative data
  - data access: application often more cumbersome, access conditions stringent
  - internal logic of data: administrative not economic, sociological etc.
  - **dependent on current legislation**
    - important variables for distributional or behavioral analyses missing (sociological hh, education level, etc.)
    - changing over time
  - **large investment costs** in MSM on administrative data (model modifications)
- question in this paper: what about the **benefits**?

# 1. introduction – what are the benefits?

To guide future decisions on investing in administrative data, we:

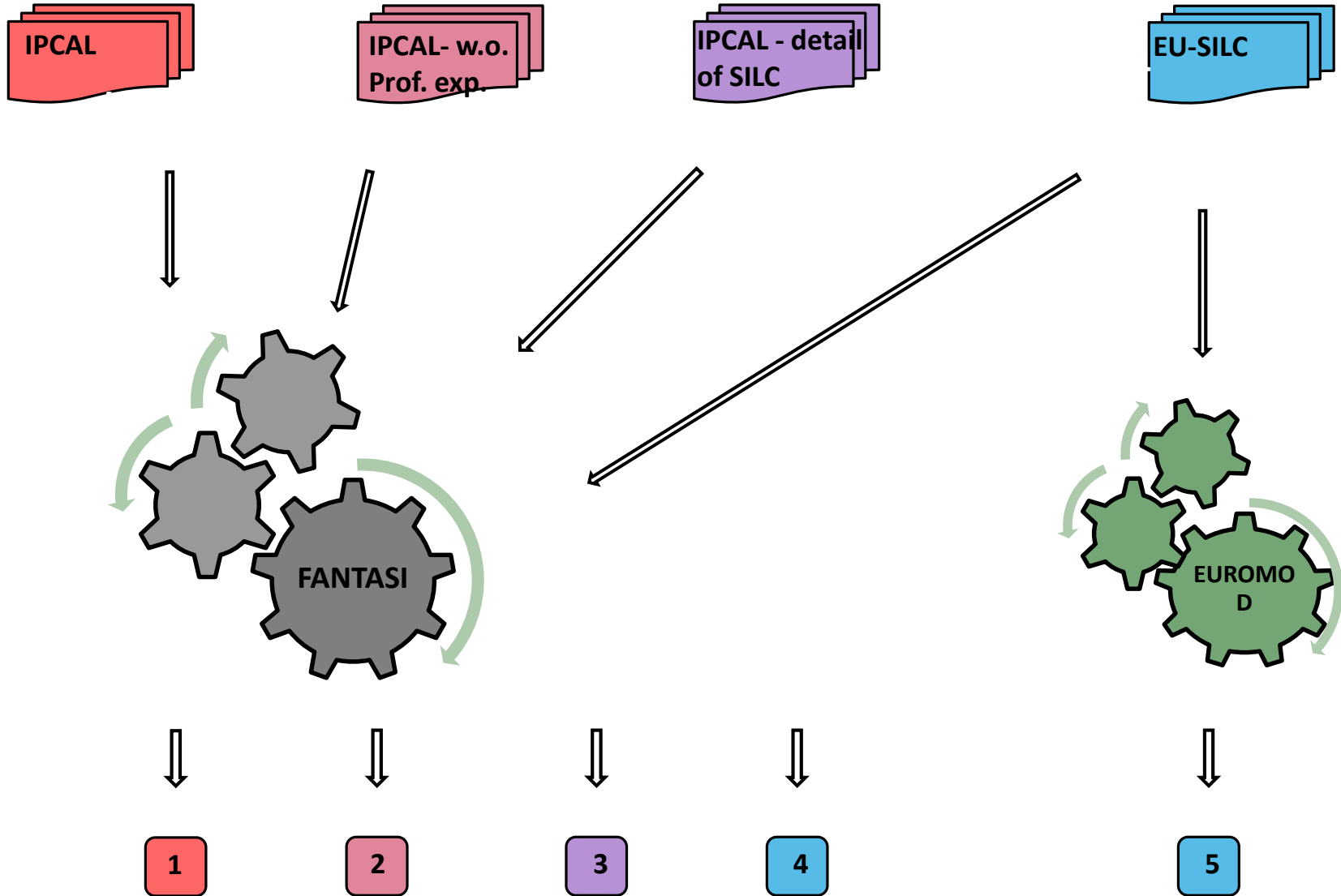
- elicit/quantify the benefits from moving to administrative data
- answer whether these benefits come from the **data** itself:
  - better (more complete) coverage of population (even when using a sample)  
→ longer tails, data on hard-to-survey groups
  - no errors due to self-reporting (! ‘errors’ = deviation of admin reality)
- or whether these benefits come from the **detail** in the data:
  - income split in different components (= taxable bases)
  - => exact computation of taxes and benefits possible (! ‘exact’ = admin reality)
  - which necessitates an investment in a detailed MSM-model
- answer in the context of Belgian personal income tax (and **only** PIT)
  - comparison of two data-sources EM-SILC and IPCAL
  - using two microsimulation models EUROMOD and FANTASI
  - for Flemish population (extension to Belgian population: under construction)

1. introduction
2. methodology
3. data and the Belgian personal income tax
4. results
5. conclusions

## 2. methodology – decomposition analysis

- Use of two microsimulation models
  - EUROMOD
    - Pan-European tax-benefit model
    - Runs on EU-SILC
    - We will make use of Belgian personal income tax module
  - FANTASI
    - Developed and maintained by collaboration of Department of Economics KU Leuven and the Department of Finance and Budget of the Flemish government
    - Detailed model of the personal income tax
    - Runs on Belgian (for now only Flemish) administrative data from tax receipts

# 2. methodology – decomposition analysis



1. introduction
2. methodology
3. data and the Belgian personal income tax
  1. EM-SILC & IPCAL
  2. Belgian personal income tax
4. results
5. conclusions

## 3.1 data: EM-SILC and IPCAL

### ■ EM-SILC

- SILC: Survey on Income and Living Conditions
- UDB-SILC: Harmonized over European countries, and input-dataset for EUROMOD
- EM-SILC: The output (and input) of EUROMOD
  
- we use SILC 2018, with reported incomes of 2017, and simulated values from EUROMOD-year 2017, sample size: +/- 14 000
- Self-reported incomes

### ■ IPCAL

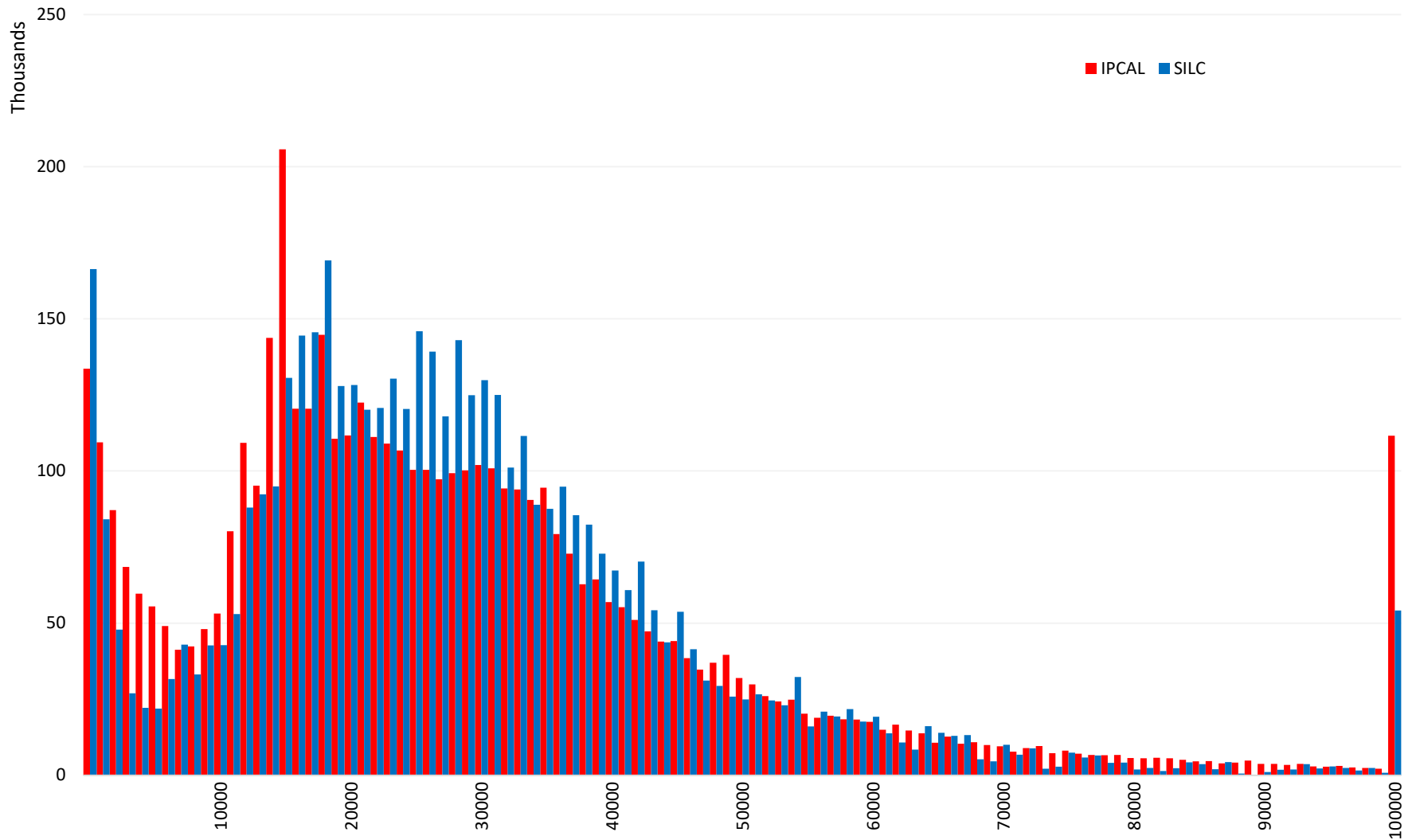
- ‘Impôt des Personnes Physiques Calculé’
- input for the microsimulation model for the personal income tax: FANTASI
  
- income year 2017 (tax year 2018), sample (151000)
- reported data from personal income tax forms
- includes calculations from the administration of FPS Finances



## 3.1 data: EM-SILC and IPCAL – income concept: GTI

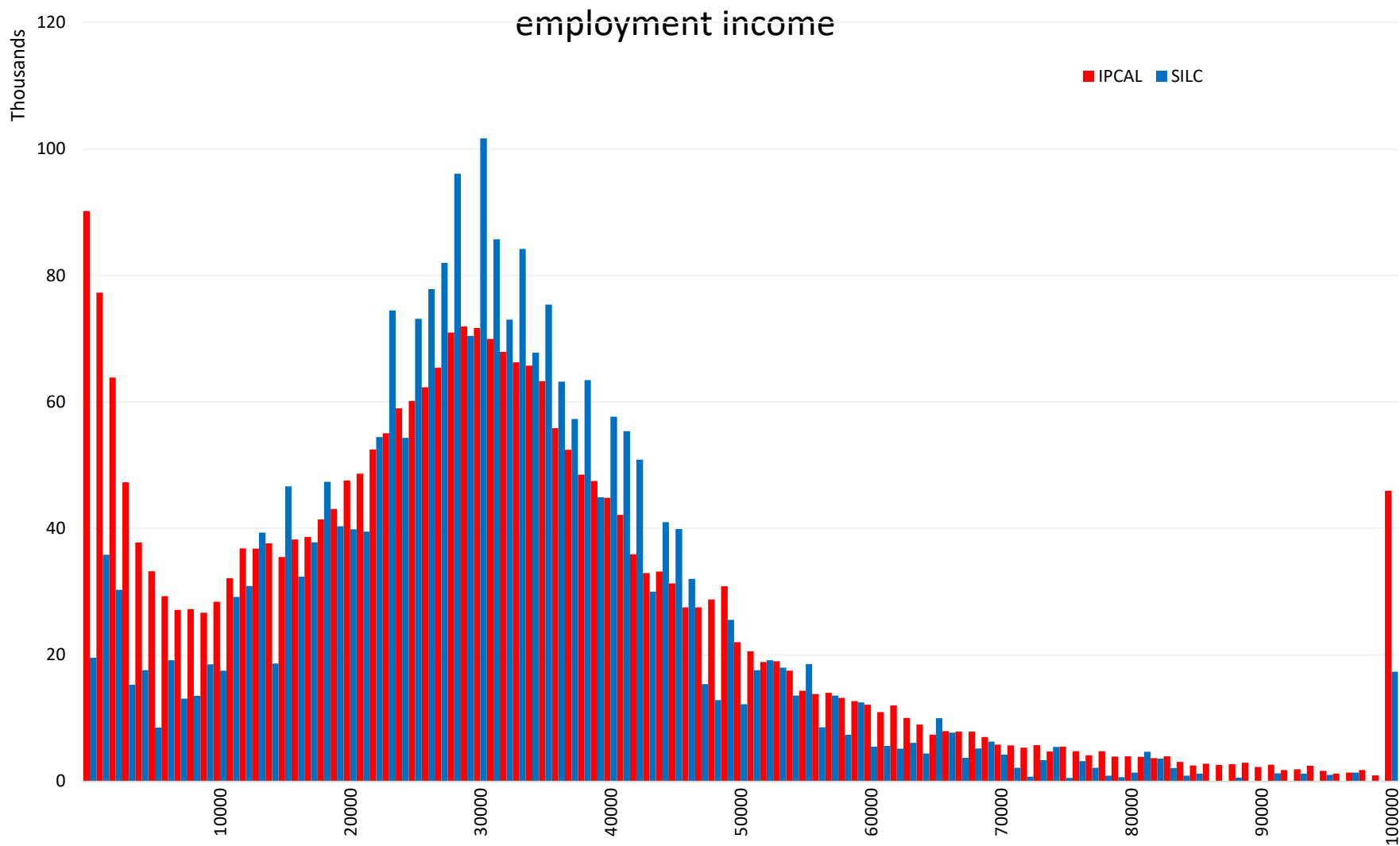
- **after** social security contributions
- **including** professional expenses (if possible), consisting of
  - employment income
  - self-employment income
  - unemployment benefits
  - public pensions
  - early-retirement income
  - sickness and disability
  - private pensions
  - property income
  - investment income & income from financial assets
  - income from private transfers (mainly maintenance allowances)
- not perfectly comparable
  - some issues with isolating reported expenses
  - liberating withholding tax (income is not stated on tax forms)

# 3.1 data: EM-SILC and IPCAL – gross taxable income distribution



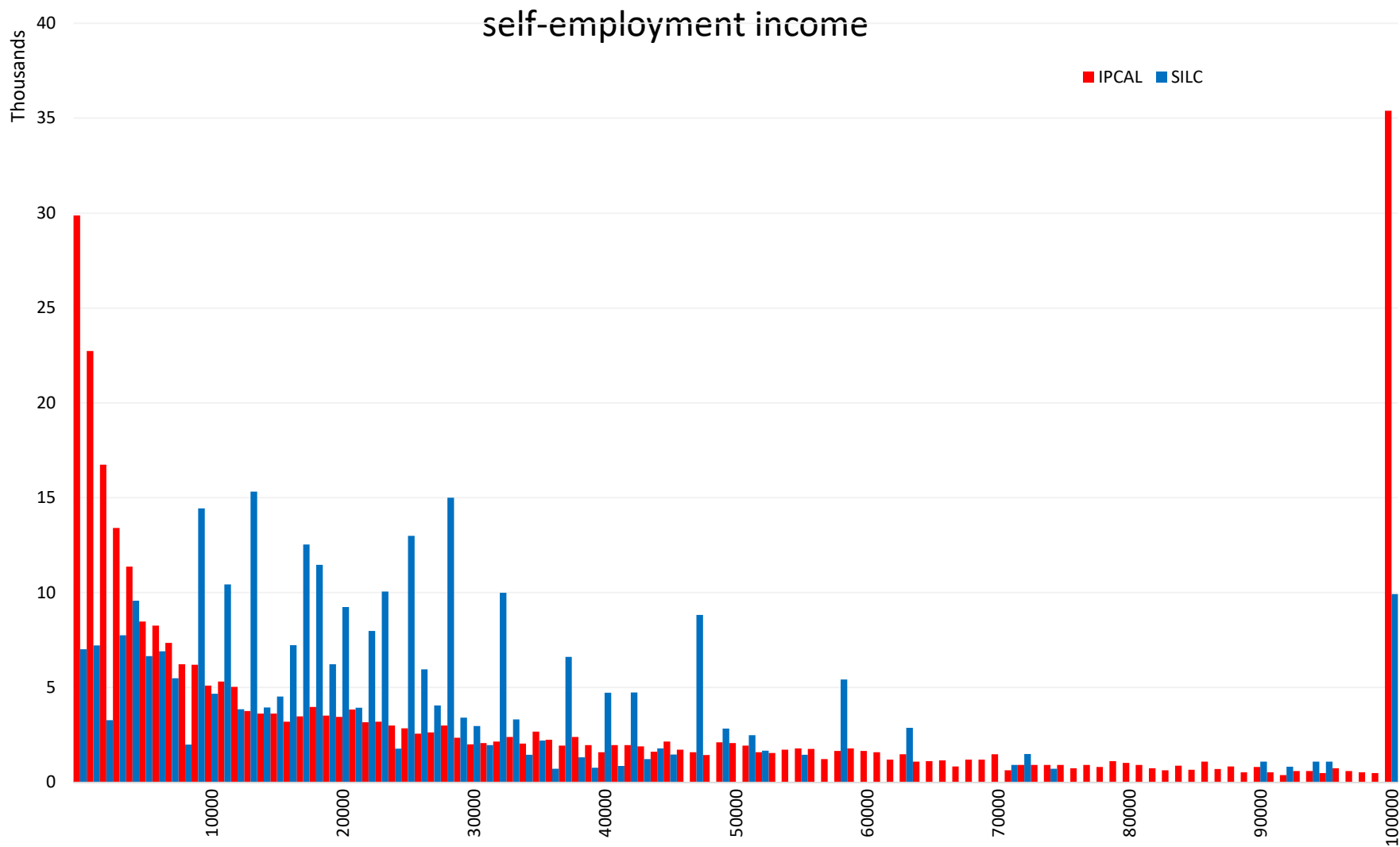
Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 3.1 data: EM-SILC and IPCAL – comparing income distributions

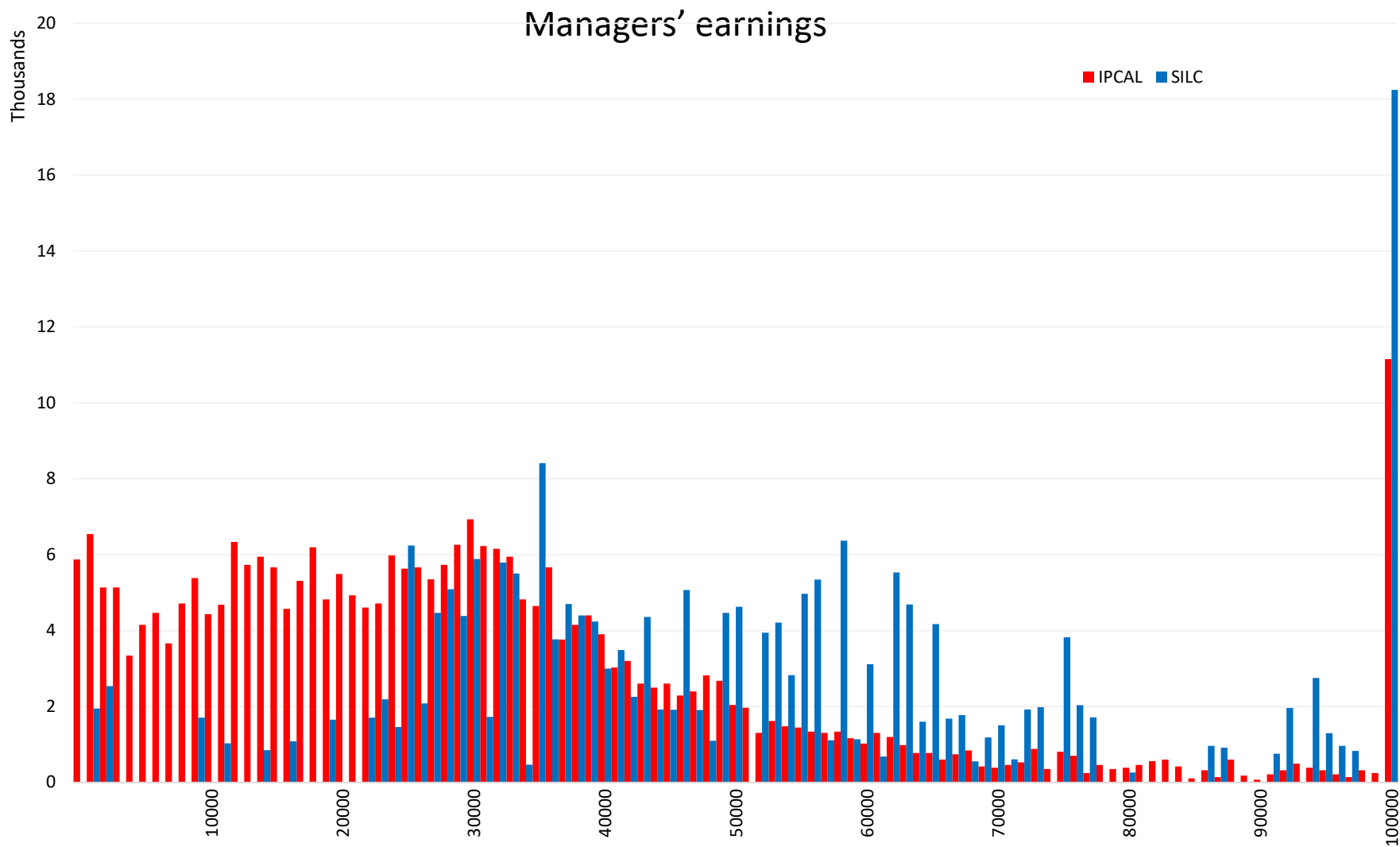


Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 3.1 data: EM-SILC and IPCAL – comparing income distributions

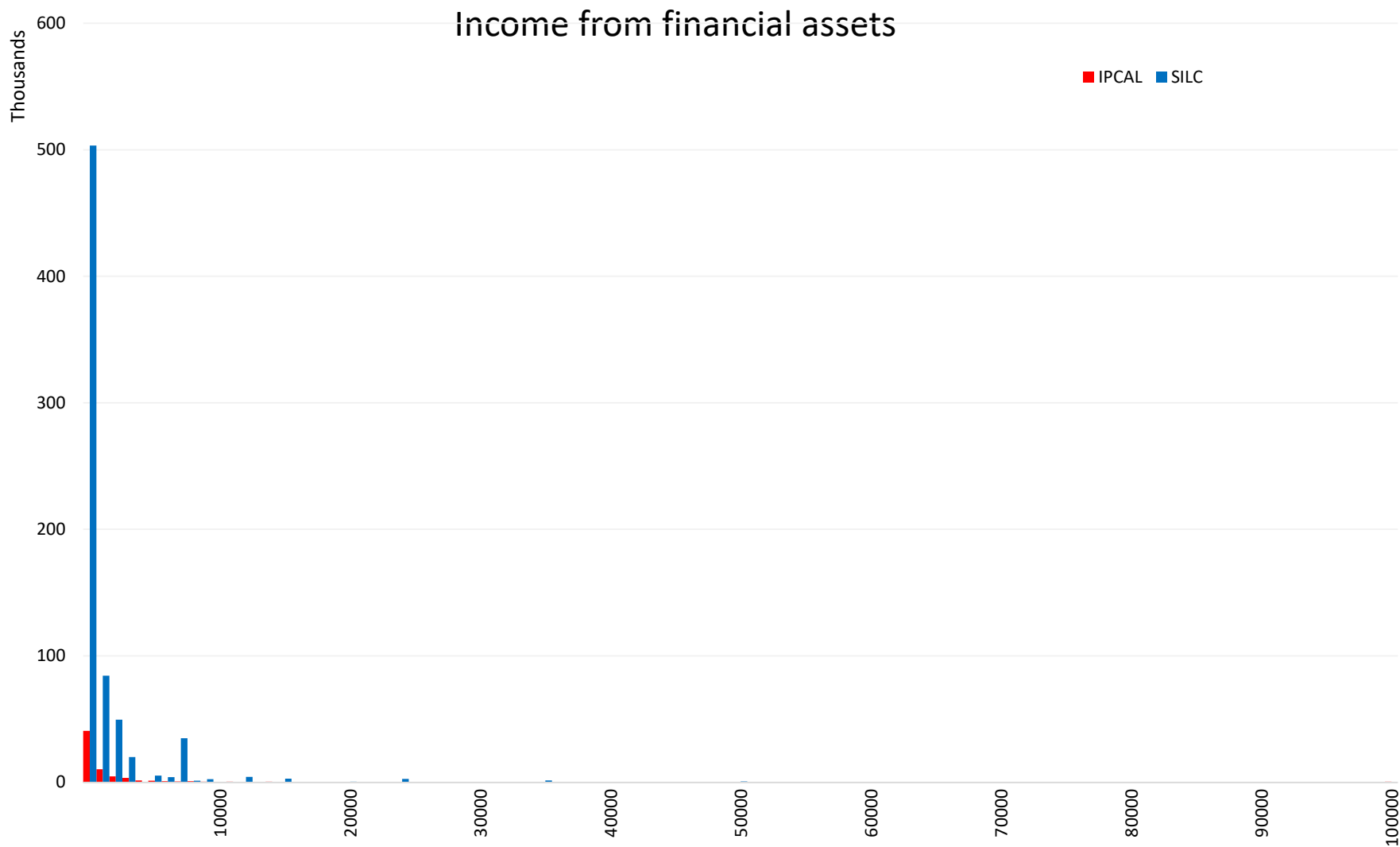


Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI



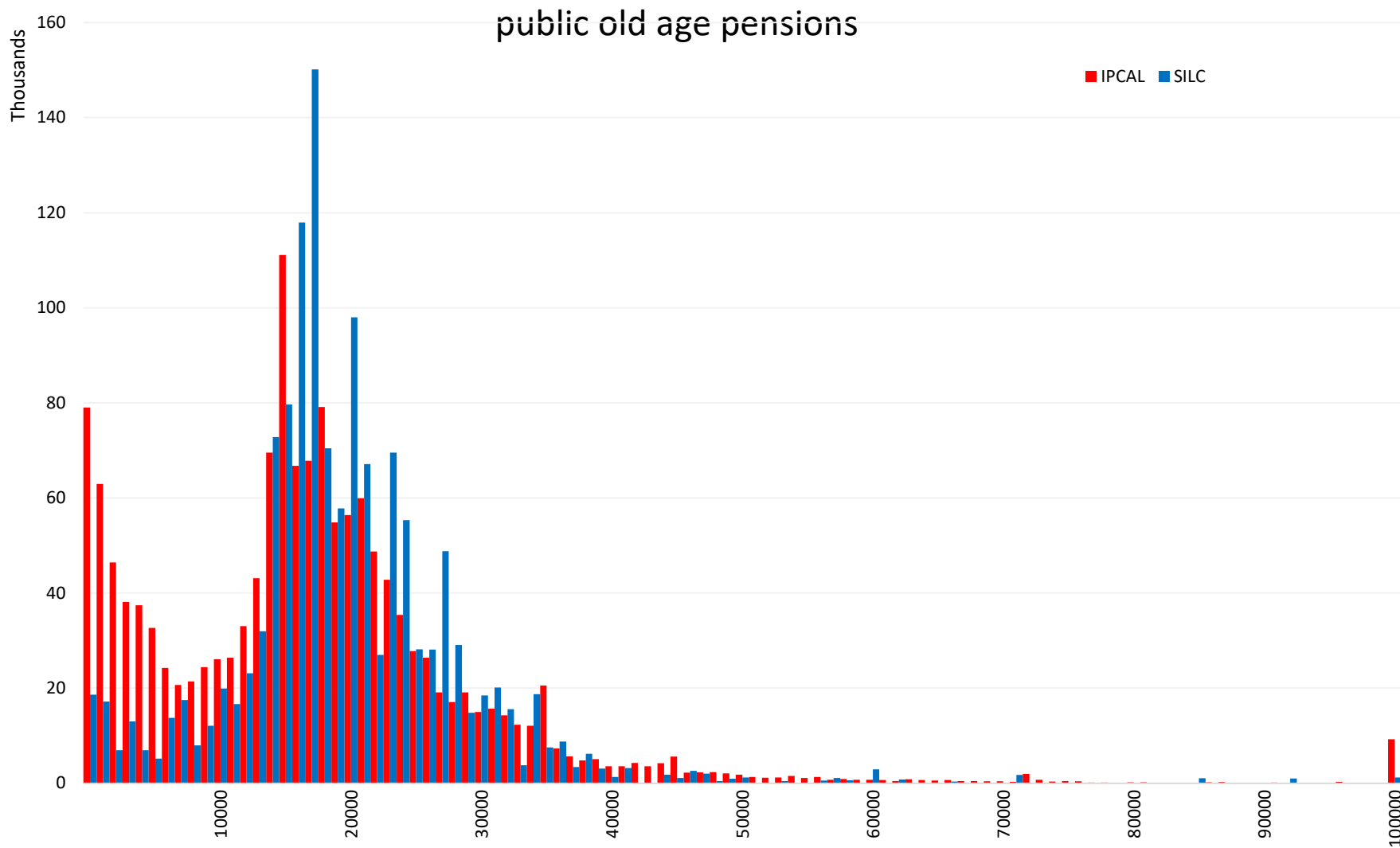
Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 3.1 data: EM-SILC and IPCAL – comparing income distributions

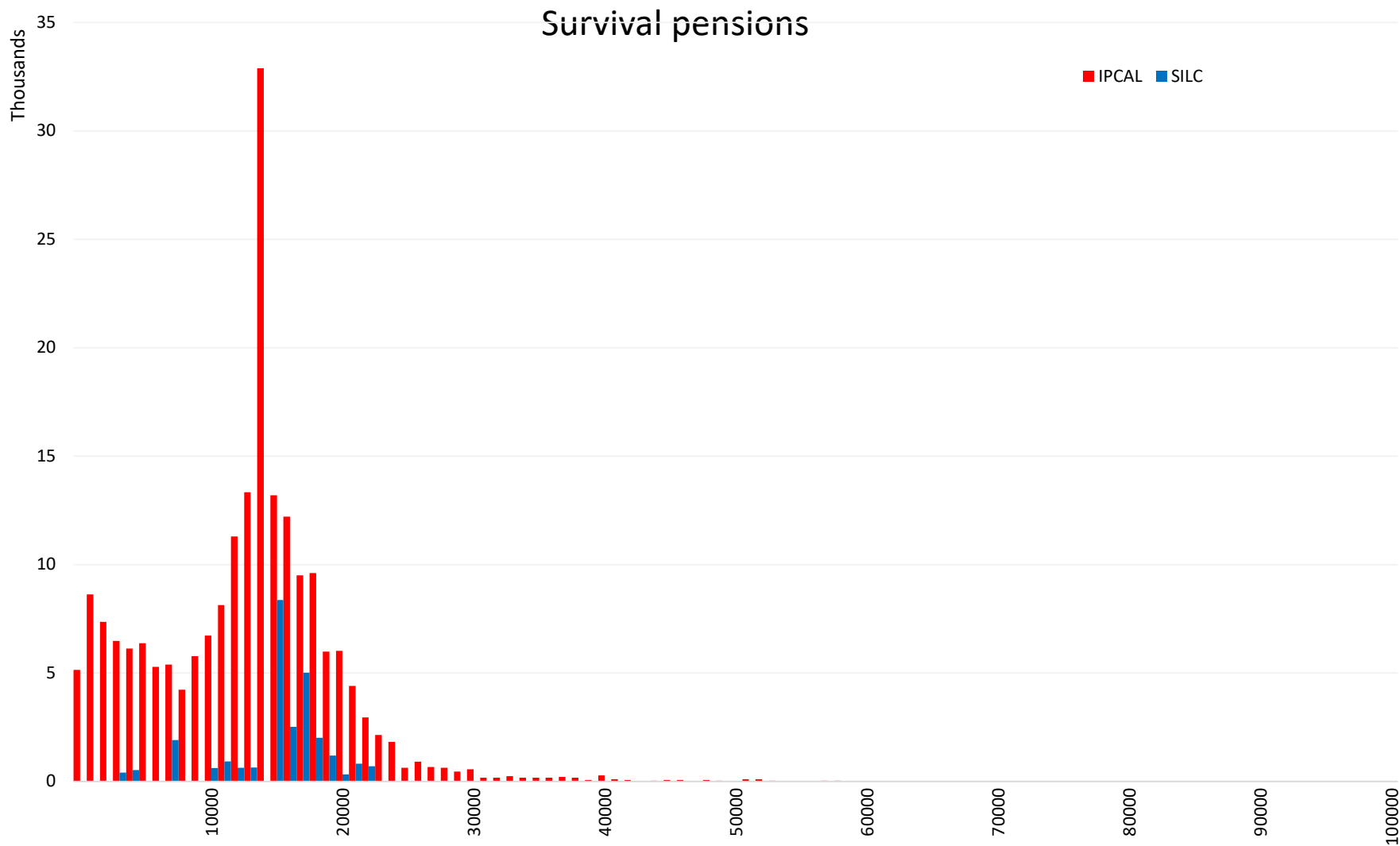


Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 3.1 data: EM-SILC and IPCAL – comparing income distributions



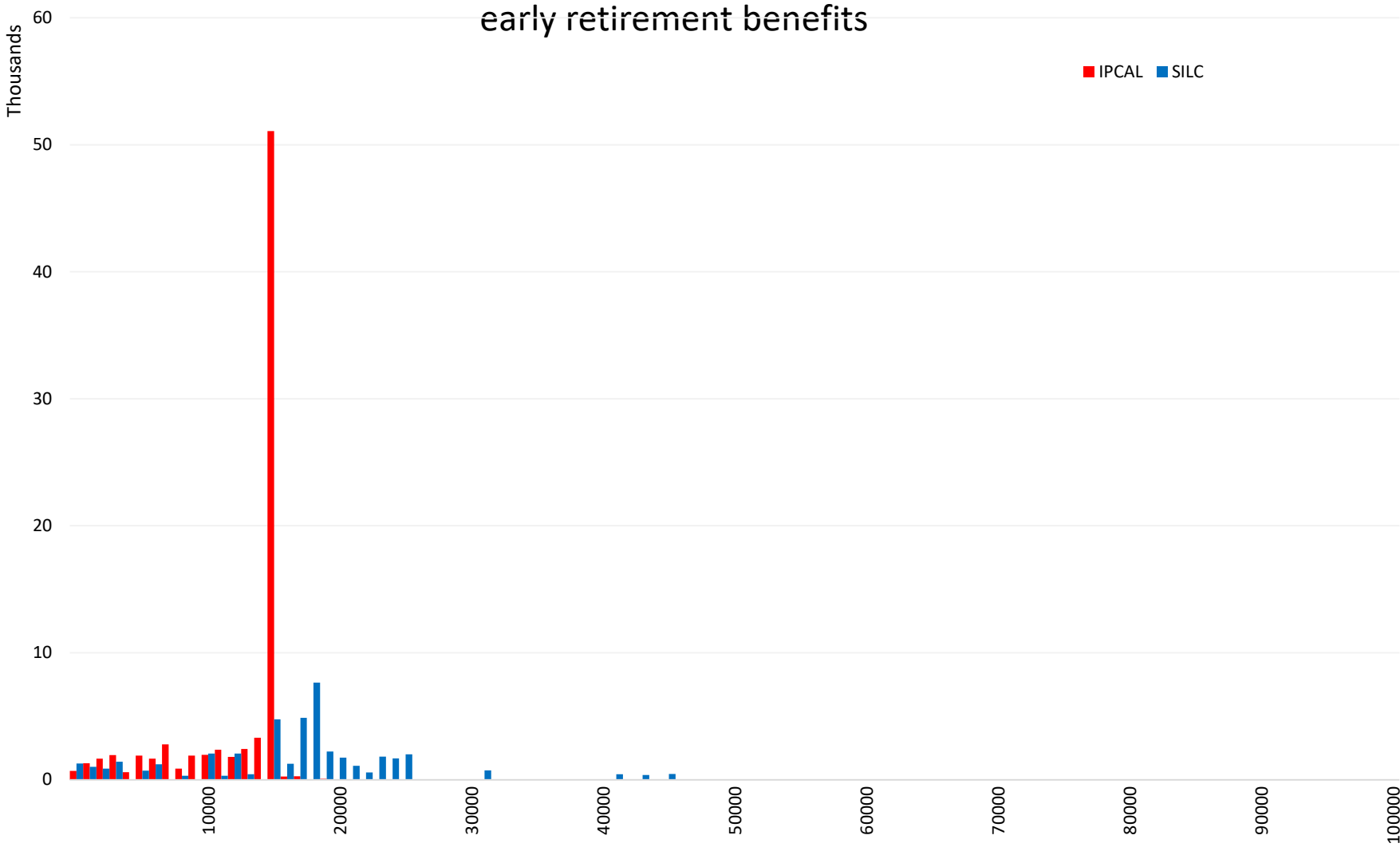
Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

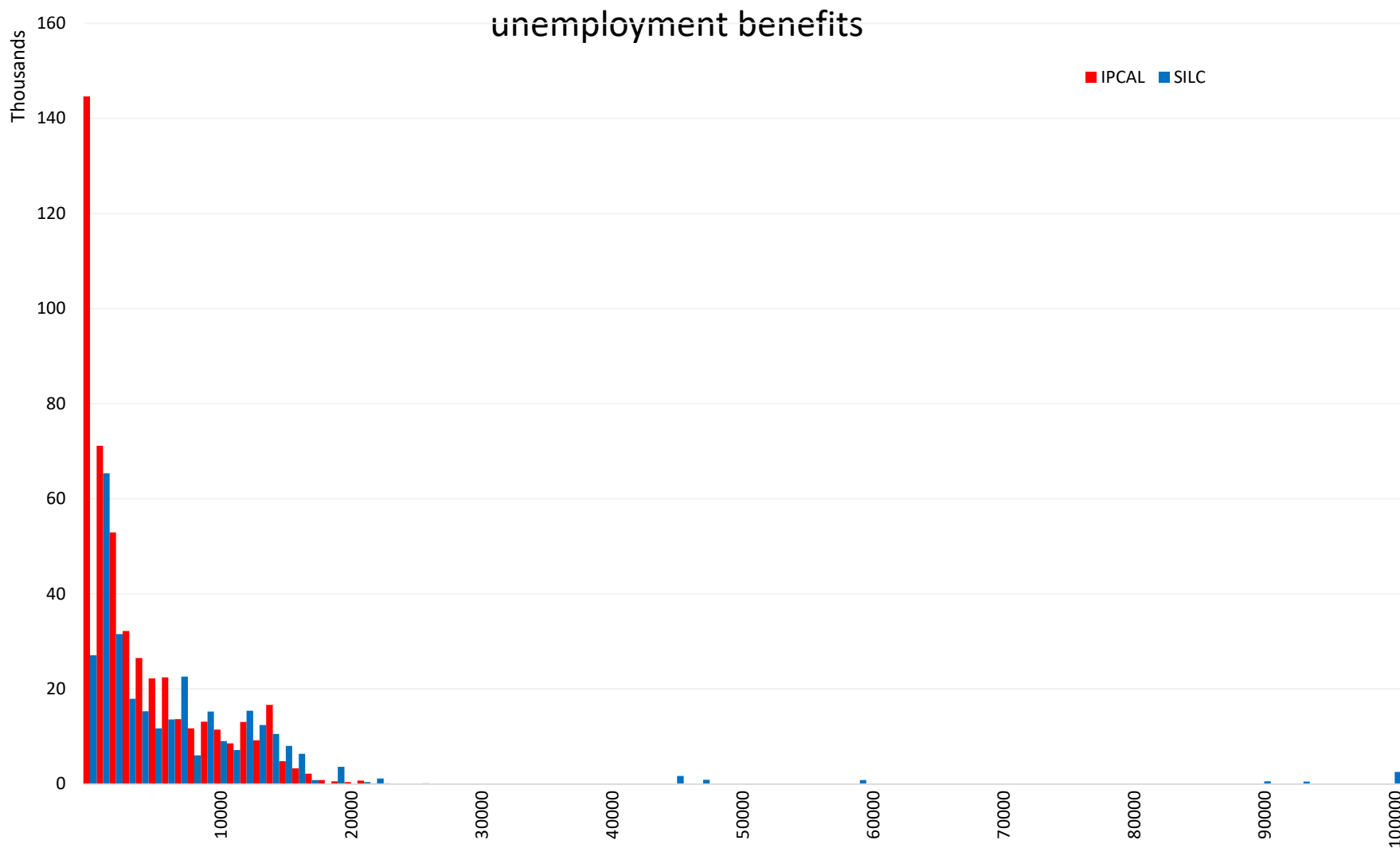


# 3.1 data: EM-SILC and IPCAL – comparing income distributions

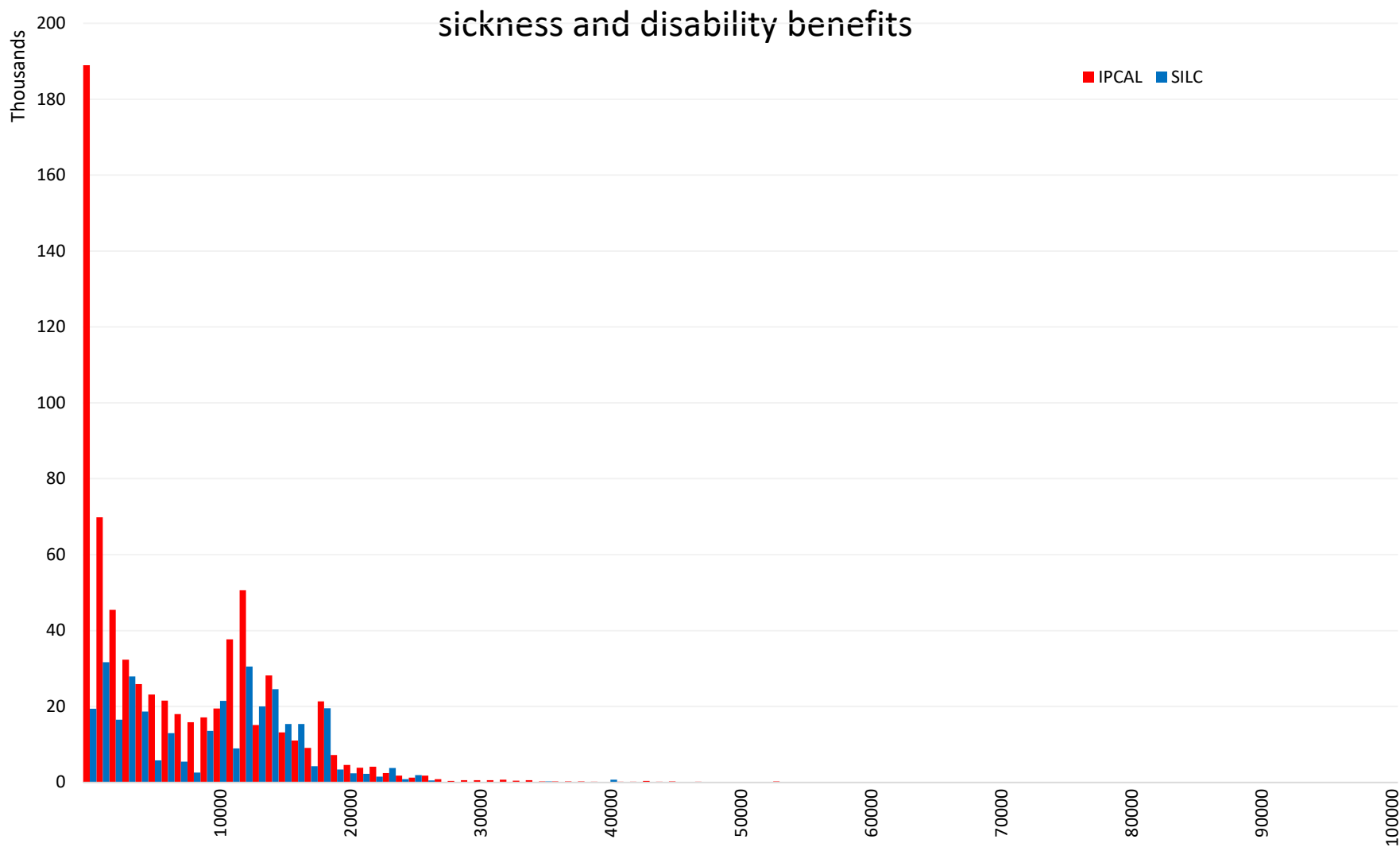


Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 3.1 data: EM-SILC and IPCAL – comparing income distributions



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

## frequency in thousands by socio-econ status

	Total	early retired	employee	manager	mixed	pensioned	residual	self-employed	unemployed
IPCAL	5327,1	56,6	2237,2	219,0	534,5	1269,2	782,9	141,4	86,4
SILC	5293,8	29,7	2270,2	195,7	287,1	1286,4	885,8	235,2	103,8

## average income in thousands by socio-econ status

	Total	early retired	employee	manager	mixed	pensioned	residual	self-employed	unemployed
IPCAL	28,4	15,2	36,2	40,9	40,3	19,1	3,6	78,2	11,4
SILC	27,8	21,2	35,4	58,9	40,6	22,8	4,5	36,6	12,9

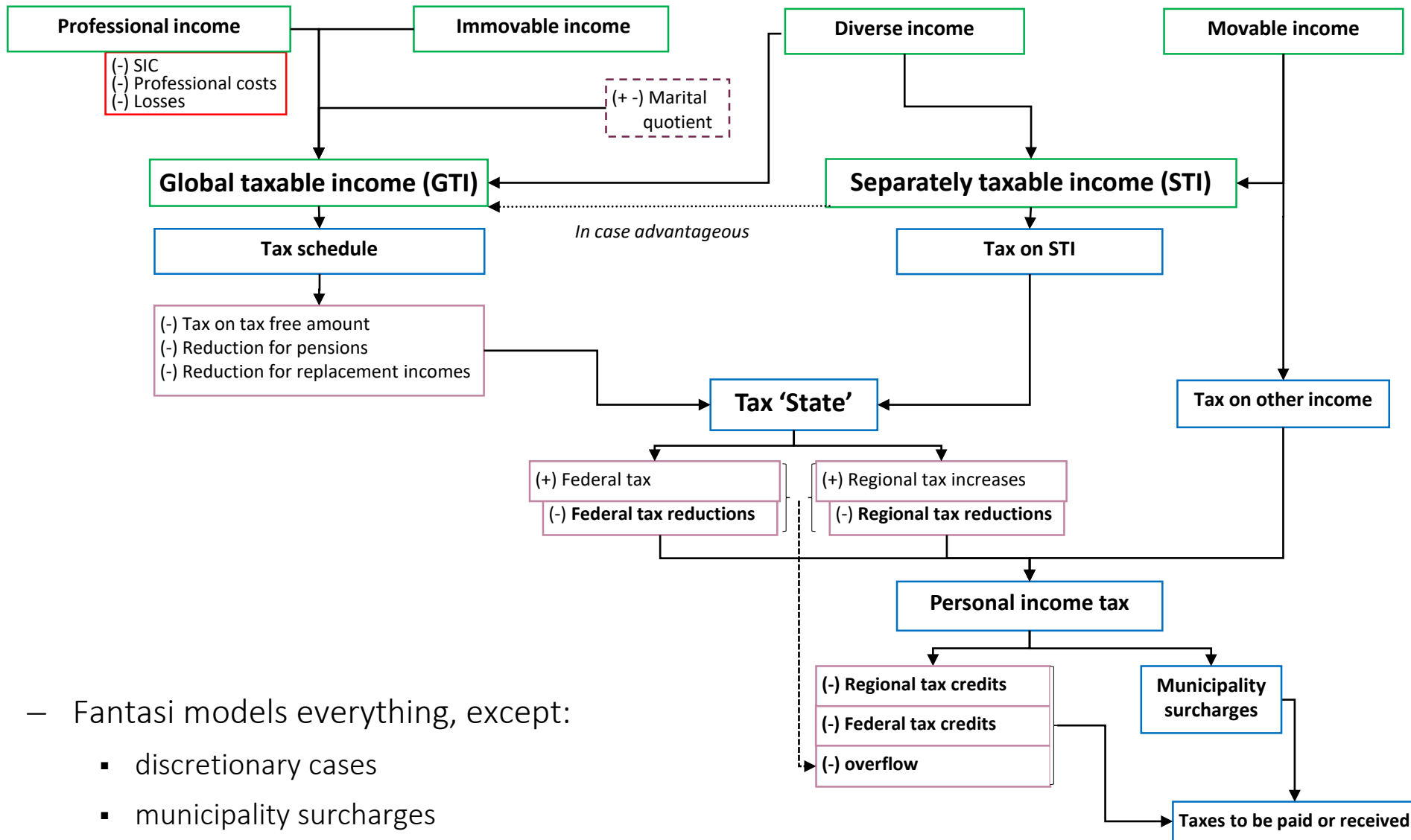
## taxable income in millions by socio-econ status

	Total	early retired	employee	manager	mixed	pensioned	residual	self-employed	unemployed
IPCAL	151553,2	862,7	81051,9	8952,9	21545,7	24240,0	2855,4	11059,8	984,7
SILC	144066,8	619,8	78584,7	11290,1	11519,7	28822,2	3379,7	8537,2	1313,3

Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

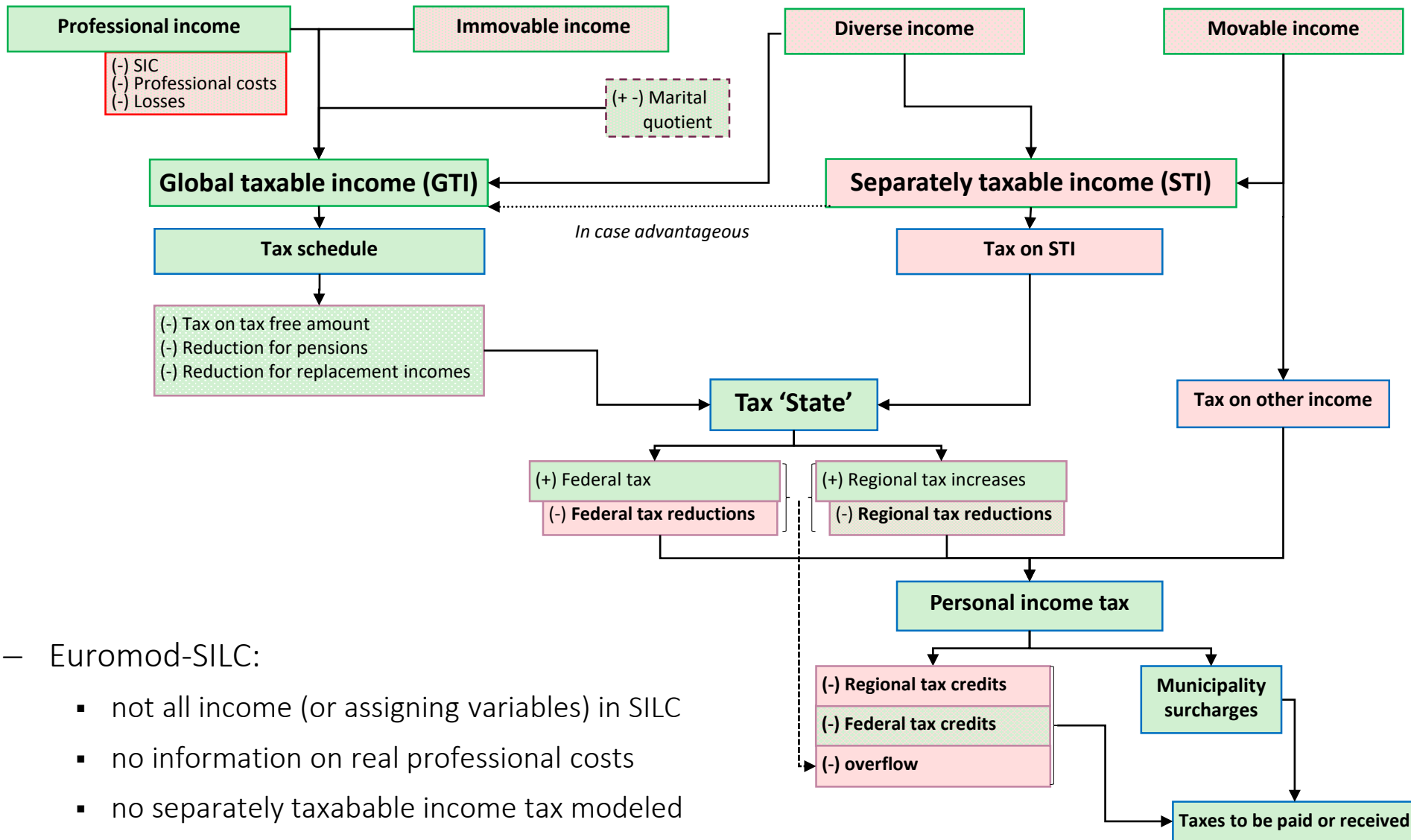
1. introduction
2. methodology
3. data and the Belgian personal income tax
  1. EM-SILC & IPCAL
  2. Belgian personal income tax
4. results
5. conclusions

# 3.2 Belgian personal income tax



- Fantasi models everything, except:
  - discretionary cases
  - municipality surcharges

# 3.2 Belgian personal income tax



- Euromod-SILC:
  - not all income (or assigning variables) in SILC
  - no information on real professional costs
  - no separately taxabable income tax modeled
  - only part of reductions & credits (not sufficient information in SILC)

1. introduction
2. methodology
3. data and the Belgian personal income tax
4. **results**
  1. taxes paid
    1. **average taxes paid**
    2. total taxes paid
  2. income after tax
  3. ATR
  4. summary statistics
  5. effect of a reform
5. conclusions

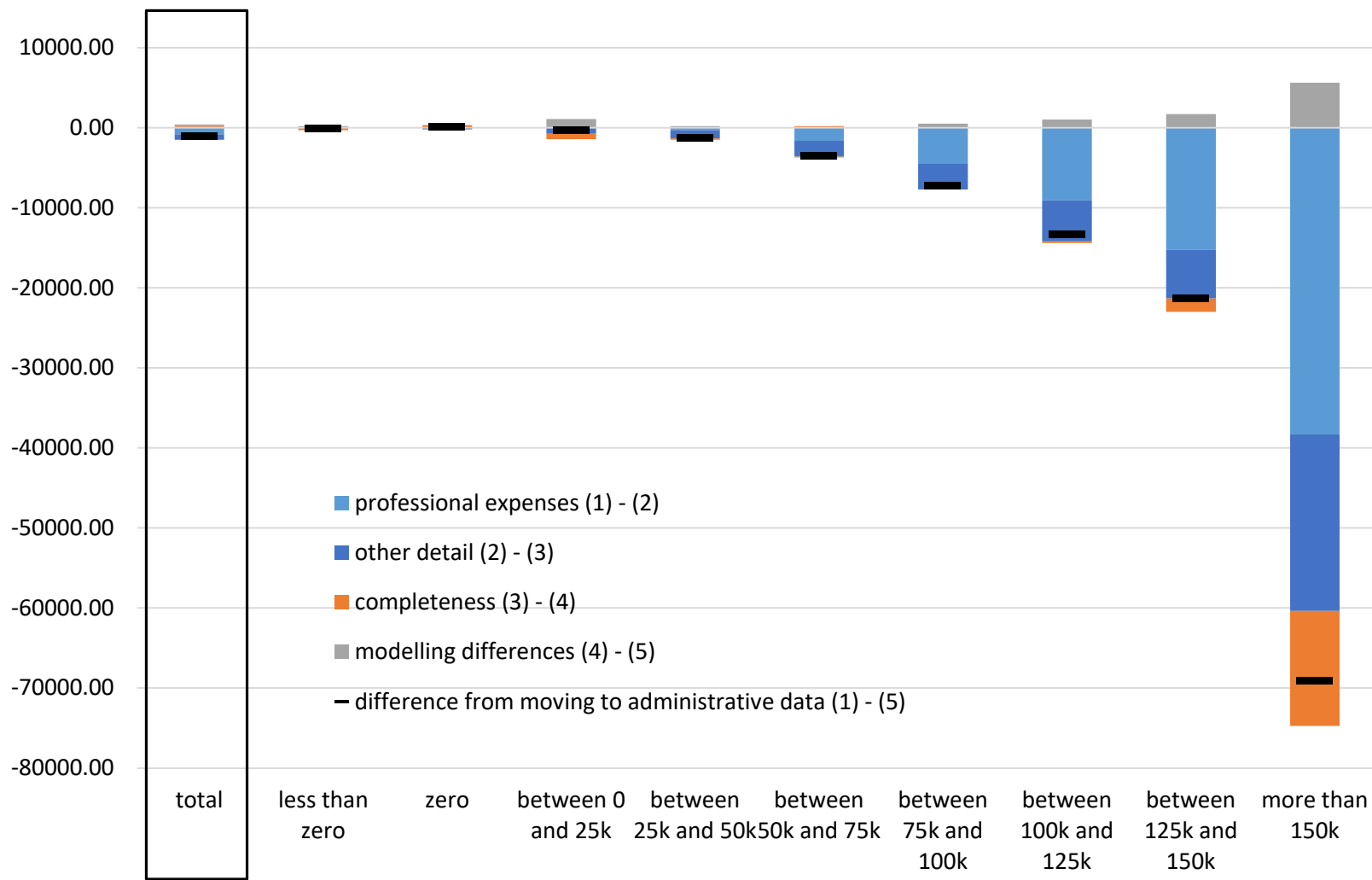


# 4.1.1 decomposition – average taxes paid by income class (€/yr)

model and data	total	less than zero	zero	between 0 and 25k	between 25k and 50k	between 50k and 75k	between 75k and 100k	between 100k and 125k	between 125k and 150k	more than 150k
Fantasi - IPCAL (1)	5470,56	-109,60	164,48	901,22	6910,89	16179,29	24626,84	31196,23	37004,43	60589,21
Fan. - IPCAL w.o. prof. exp. (2)	6280,02	-96,84	176,73	933,60	7203,07	17816,26	29068,06	40261,43	52244,81	98907,42
an. - IPCAL with SILC-detail (3)	6960,06	-93,02	355,08	1659,03	8237,17	19721,88	32351,49	45391,99	58318,97	120960,86
Fantasi - EU-SILC (4)	6753,05	193,53	48,42	2328,08	8394,86	19510,14	32319,38	45579,93	59988,10	135314,34
Euromod - EU-SILC (5)	6539,04	0,00	41,04	1219,80	8180,56	19681,58	31847,40	44549,46	58290,60	129669,87
difference from moving to administrative data (1) - (5)	-1068,49	-109,60	123,44	-318,58	-1269,67	-3502,29	-7220,56	-13353,23	-21286,17	-69080,65
prof. expenses (1) - (2)	-809,46	-12,76	-12,25	-32,38	-292,18	-1636,97	-4441,22	-9065,19	-15240,38	-38318,20
other detail (2) - (3)	-680,04	-3,81	-178,35	-725,44	-1034,10	-1905,62	-3283,43	-5130,56	-6074,16	-22053,45
completeness (3) - (4)	207,01	-286,55	306,66	-669,05	-157,69	211,74	32,12	-187,94	-1669,13	-14353,47
mod. differences (4) - (5)	214,01	193,53	7,38	1108,29	214,30	-171,44	471,97	1030,47	1697,50	5644,47

Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 4.1.1 decomposition – average taxes paid by income class (€/yr)



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 4.1.1 decomposition – average taxes paid by age group (€/yr)

model and data	total	younger than 21	between 21 and 40	between 41 and 60	between 61 and 80	older than 81
Fantasi - IPCAL (1)	5470,56	164,94	5199,35	8252,19	3428,49	1762,91
Fantasi - IPCAL w.o. prof. exp. (2)	6280,02	187,98	5994,96	9625,90	3762,36	1778,30
Fantasi - IPCAL with SILC-detail (3)	6960,06	217,47	6751,49	10598,55	3961,09	2754,96
Fantasi - EU-SILC (4)	6753,05	145,18	6282,36	10048,44	5129,73	3824,92
Euromod - EU-SILC (5)	6539,04	222,88	6742,93	10379,05	3447,93	1952,45
difference from moving to administrative data (1) - (5)	-1068,49	-57,94	-1543,58	-2126,86	-19,44	-189,54
professional expenses (1) - (2)	-809,46	-23,04	-795,60	-1373,70	-333,87	-15,39
other detail (2) - (3)	-680,04	-29,49	-756,54	-972,65	-198,73	-976,65
completeness (3) - (4)	207,01	72,29	469,13	550,11	-1168,64	-1069,97
modelling differences (4) - (5)	214,01	-77,70	-460,57	-330,61	1681,80	1872,47

Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

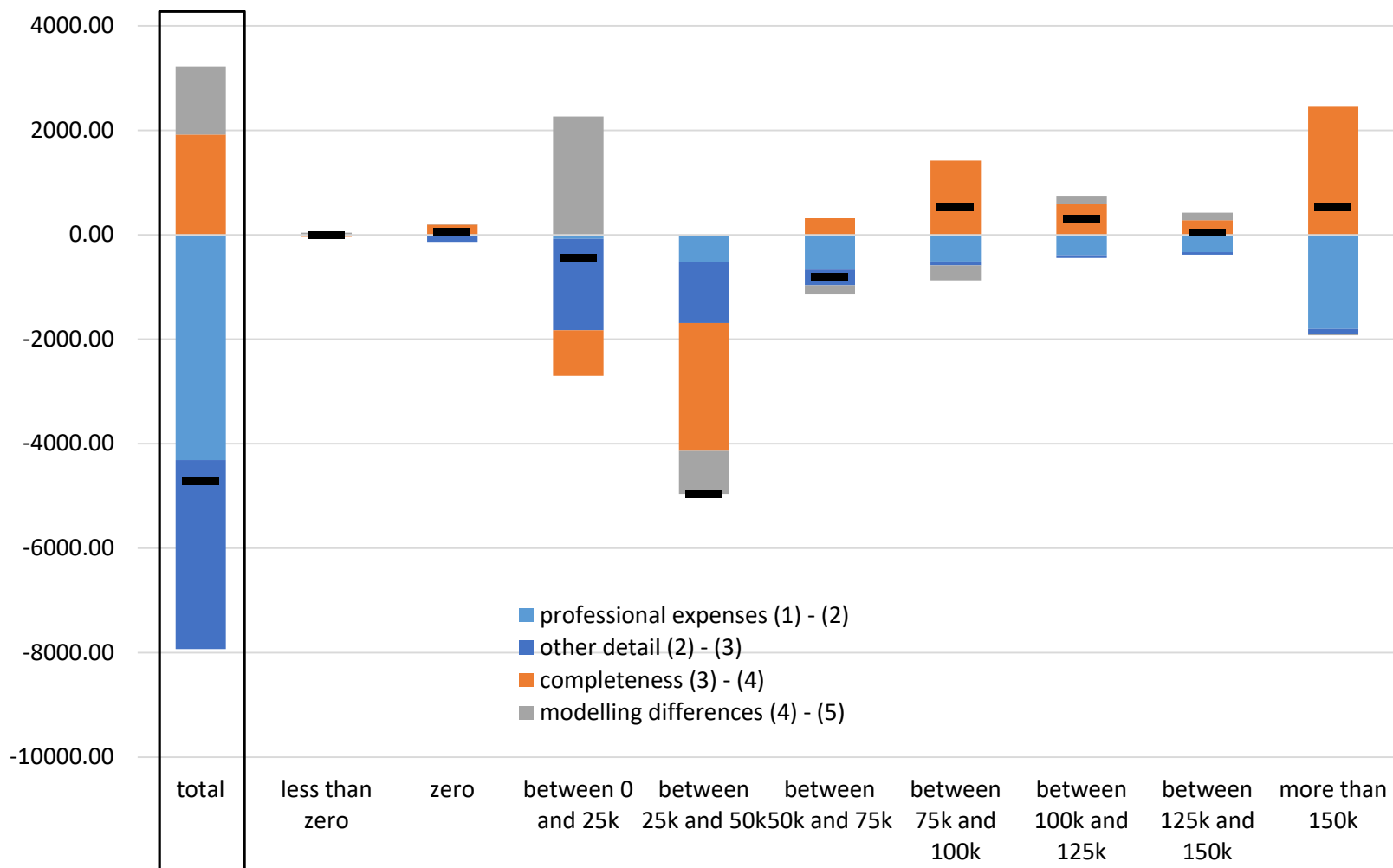
# 4.1.1 decomposition – average taxes paid by income source (€/yr)

model and data	total	early retired	employee	manager	mixed	pensioned	residual	self-employed	unempl.
Fantasi - IPCAL (1)	5470,56	1537,81	8244,90	11103,12	6879,44	2601,63	330,02	6624,86	52,09
Fantasi - IPCAL w.o. prof. exp. (2)	6280,02	1558,21	8279,96	11167,73	8712,55	2605,75	344,14	29410,29	56,94
Fantasi - IPCAL with SILC-detail (3)	6960,06	3192,33	9112,31	11320,40	9447,04	3489,56	669,20	30356,73	792,42
Fantasi - EU-SILC (4)	6753,05	4081,53	8152,48	17355,77	13925,20	4908,10	598,38	14045,44	1100,46
Euromod - EU-SILC (5)	6539,04	2157,21	9230,86	21299,08	11867,72	3020,06	306,36	9297,30	799,79
difference from moving to administrative data (1) - (5)	-1068,49	-619,40	-985,96	-10195,96	-4988,29	-418,43	23,66	-2672,44	-747,71
professional expenses (1) - (2)	-809,46	-20,40	-35,06	-64,61	-1833,12	-4,12	-14,13	-22785,42	-4,85
other detail (2) - (3)	-680,04	-1634,12	-832,35	-152,67	-734,49	-883,81	-325,05	-946,44	-735,48
completeness (3) - (4)	207,01	-889,19	959,83	-6035,37	-4478,16	-1418,54	70,82	16311,29	-308,04
modelling differences (4) - (5)	214,01	1924,32	-1078,38	-3943,31	2057,48	1888,04	292,02	4748,14	300,67

Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

1. introduction
2. methodology
3. data and the Belgian personal income tax
4. **results**
  1. taxes paid
    1. average taxes paid
    2. **total taxes paid**
  2. income after tax
  3. ATR
  4. summary statistics
  5. effect of a reform
5. conclusions

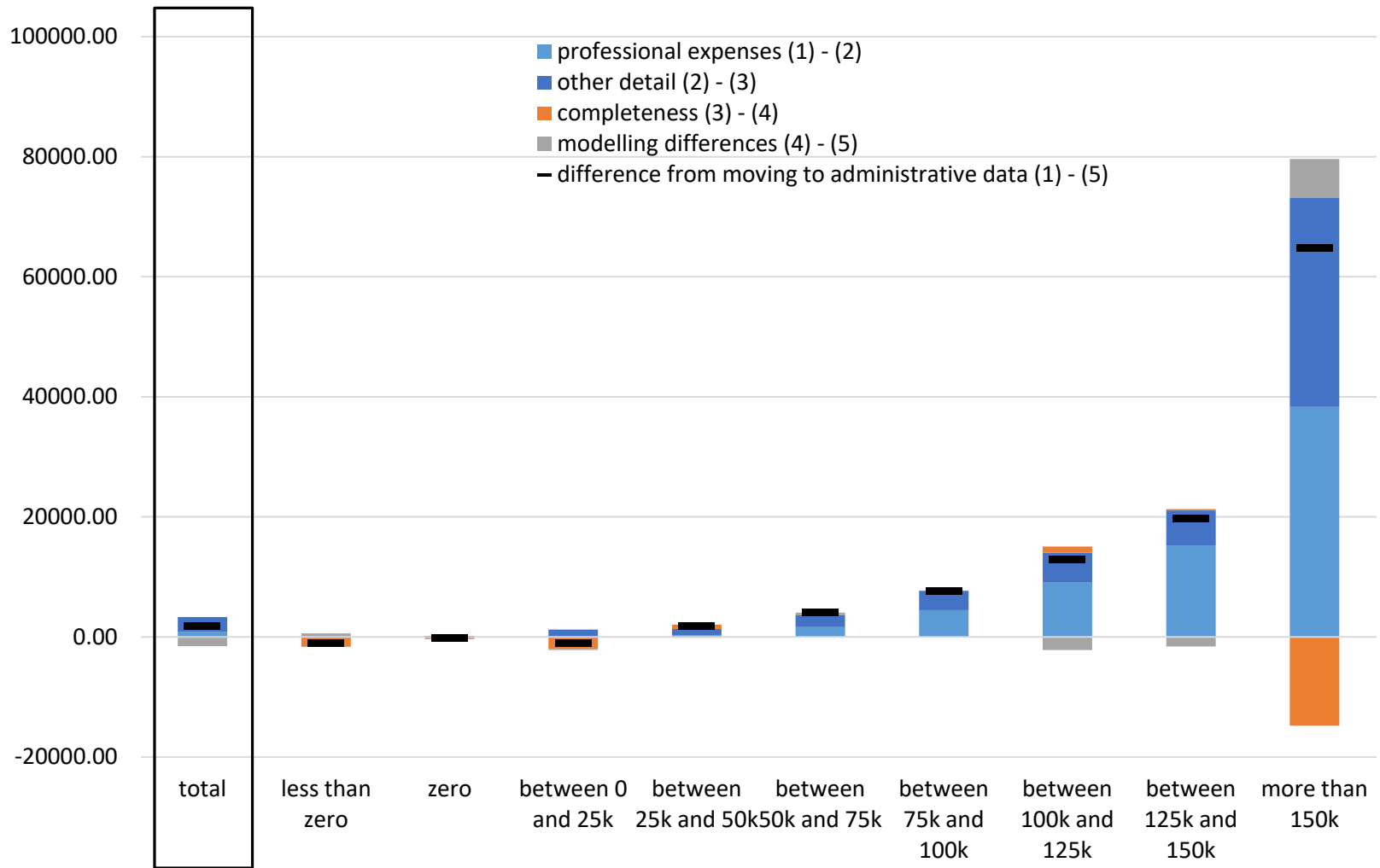
# 4.1.1 decomposition – total taxes paid in income source (€/yr)



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

1. introduction
2. methodology
3. data and the Belgian personal income tax
4. **results**
  1. taxes paid
  2. **income after tax**
  3. ATR
  4. summary statistics
  5. effect of a reform
5. conclusions

# 4.2 decomposition – income after tax by income class (€K/year)

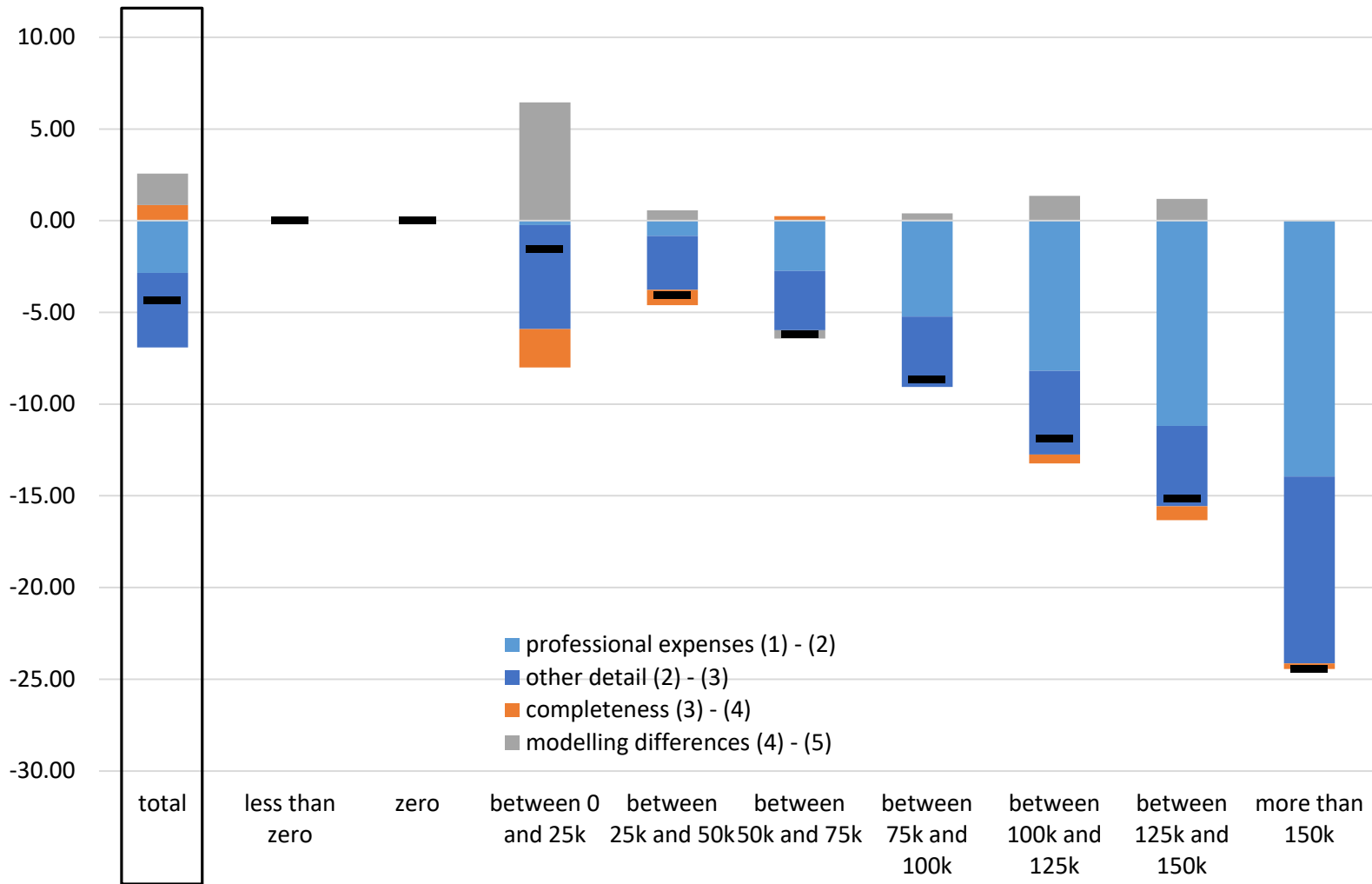


Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI



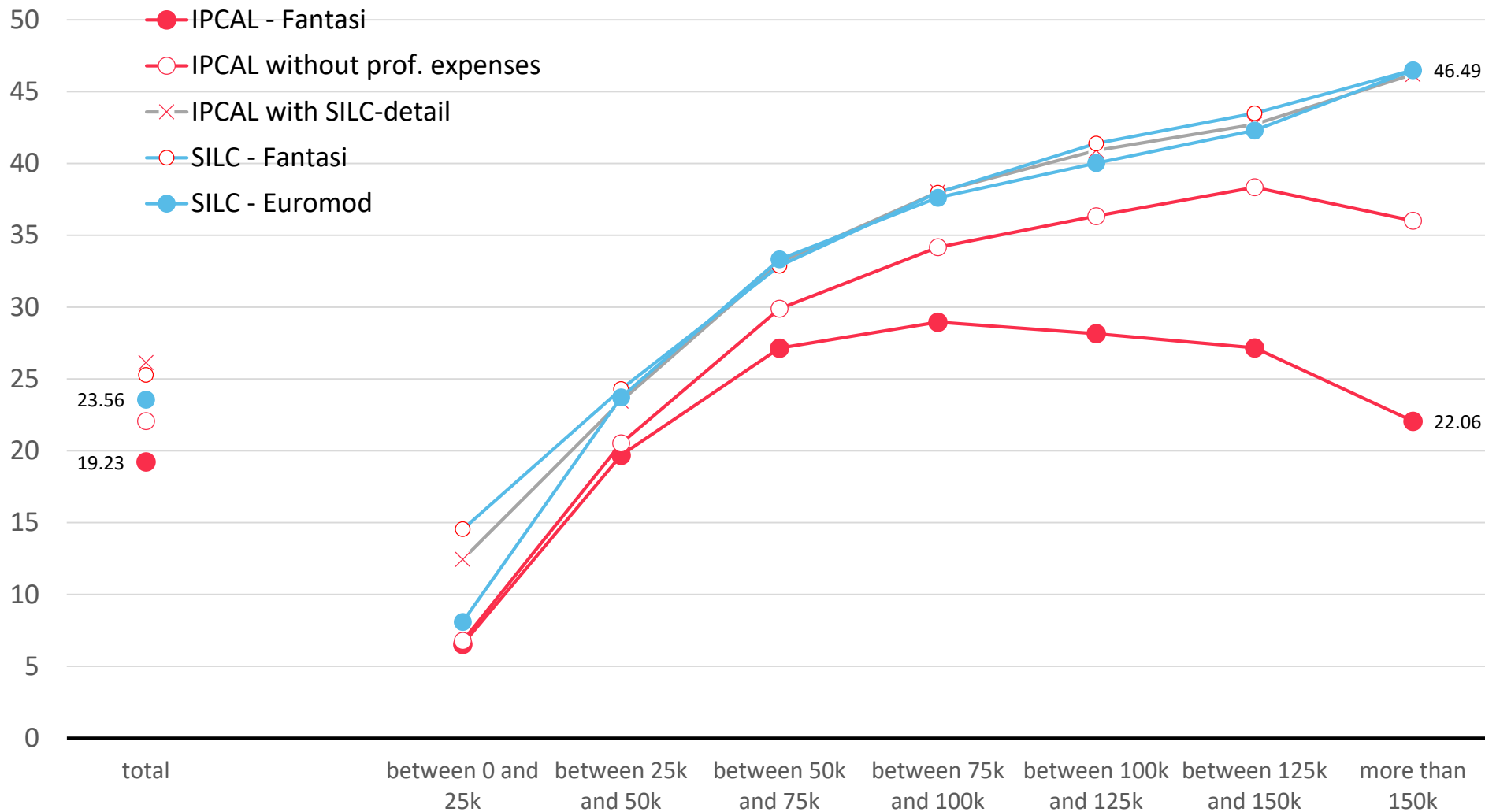
1. introduction
2. methodology
3. data and the Belgian personal income tax
4. **results**
  1. taxes paid
  2. income after tax
  3. **ATR**
  4. summary statistics
  5. effect of a reform
5. conclusions

# 4.3 decomposition – average tax rate income class



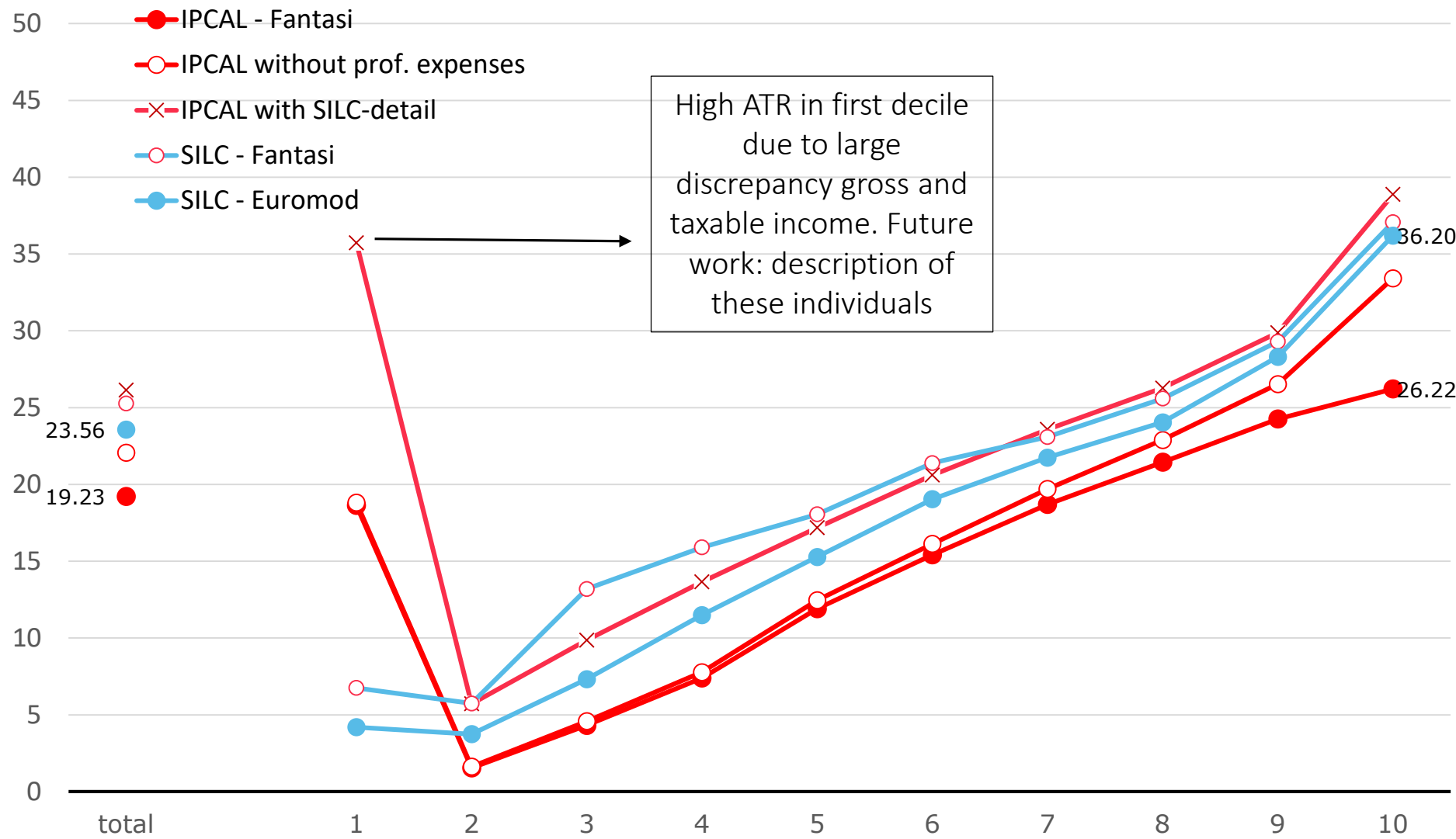
Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 4.3 decomposition – average tax rate by income class



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

# 4.3 decomposition – average tax rate by deciles



Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

1. introduction
2. methodology
3. data and the Belgian personal income tax
4. **results**
  1. taxes paid
  2. income after tax
  3. ATR
  4. **summary statistics**
  5. effect of a reform
5. conclusions

	Gini pre-tax income	Gini after-tax income	progressivity
Fantasi - IPCAL (1)	46,98	43,07	17,89
Fantasi - IPCAL w.o. prof. exp. (2)	46,98	41,38	20,89
Fantasi - IPCAL with SILC-detail (3)	48,75	43,47	15,75
Fantasi - EU-SILC (4)	41,53	36,77	15,09
Euromod - EU-SILC (5)	39,36	32,93	21,87
difference from moving to administrative data (1) - (5)	7,62	10,14	-3,98
professional expenses (1) - (2)	0,00	1,69	-3,00
other detail (2) - (3)	-1,77	-2,10	5,14
completeness (3) - (4)	7,22	6,71	0,66
modelling differences (4) - (5)	2,17	3,84	-6,78

Source: own calculations based on IPCAL (Statbel) and SILC (Eurostat), and microsimulation models EUROMOD and FANTASI

1. introduction
2. data and the Belgian personal income tax
3. methodology
4. decomposition
5. **conclusions**

## 5. does benefit of move to admin data justify the cost?

- **yes**, ..., admin data capture upper tail of income distribution much better
  - important for revenue estimate, ...
  - ... and for distributional analysis
- it is the **detail** of admin data which is important
  - for precise tax liabilities (and hence after tax incomes) of upper tail
  - reason: high incomes often complex aggregates, with specific tax treatment and professional expenses
  - => feeding SILC with admin data will *not* help if same level of aggregation is maintained as in current SILC;
  - Initiatives, such as BELMOD, are necessary to capture the full advantages of administrative data
- “completeness” (= both misreporting and population diff.) also important
  - in aggregates like total tax revenues and summary statistics as the Gini



## 5. does benefit of move to admin data justify the cost?

- also at the bottom: significant differences
  - but less clear whether low incomes in admin correspond to socio-econ reality
  - less relevant for revenue
  
- results are
  - country specific, since dependent on tax legislation
  - dependent on the statistic used (→ deciles, fixed classes ...)
  
- Important advantage of admin data
  - Specific policy question
  - Detail in admin data allows identification of specific groups
  - Especially relevant in shock as Corona-crisis