Using fiscal data to estimate the evolution of top income shares in Belgium

A. Decoster
Department of Economics - KU Leuven

Thanks to Liebrecht De Sadeleer, Koen Dedobbeleer, Sebastiaan Maes, Lien Tam Co, Stijn Van Houtven and Toon Vanheukelom for help with the fiscal data and the development of the microsimulation model FANTASI, running on these data. The usual disclaimer applies.
roadmap

1. contribution of CRESUS-project to (first) estimate of top 1% income share
2. what was the problem/challenge?
3. results
4. conclusions
1. 2017: Belgium missing in WID-database

Top 1% national income share

before CRESUS-project: Belgium missing in WID-dataset
1. CRESUS-project: added Belgium in the WID-database
2. what was the problem/challenge?

- no lack of studies on inequality in Belgium
- no lack of pronounced statements in public debate:
  - Helft inkomen voor 20 procent rijksten
  - but (academic) work mostly based on
    - distribution of disposable income (after taxes/transfers)
    - retrieved from surveys (EU-SILC)
  - WID: revival of Kuznets-methodology
    - long run comparisons
    - based on administrative data (fiscal data)
    - income before taxes
    - corrected for changes in enrolled population and missing income
    - linked to National Accounts (DINA-project)
2. what was the problem/challenge?

- in this project: we tried to comply with international practice (read: WID)
  - for evolution of income shares top 1%, top 10%
  - which has now been enlarged to DINA-approach

- Belgium: *does* have published fiscal data, but:
  - NTI: Net Taxable Income (i.e. after deductions)
    - zero NTI’s are removed from tables (# is mentioned)
    - anyhow dependent on legislation, and administrative practice (e.g. enrolment)
    - illustration:
2. what was the problem/challenge? # of NTI=0 records

![Bar graph showing # of NTI=0 records from 1973 to 2015.](image)

- 1973: 8455
- 1975: 81891
- 1979: 175015
- 1983: 473982
- 1987: 664674
- 1991: 541355
- 1995: 590240
- 2001: 664674
- 2005: 541355
- 2009: 590240
- 2013: 590240
- 2015: 590240

What was the problem/challenge?
2. what was the problem? income growth D1 of published NTI
2. what was the problem? income growth P100 of published NTI

![Graph showing income growth P100 of published NTI from 1974 to 2015. The years 1984 and 1999 have the highest values at 16.1 and 0.1, respectively. The graph also shows some years with negative values, notably 1984 with -4.4.}
2. What was the problem? Gini published nti (0 NTI’s removed)
2. what was the problem/challenge?

- administrative data compared to EU-SILC-survey data: no magic bullet
  - data access: application often more cumbersome, access conditions stringent
  - internal logic of data: administrative not economic, sociological etc.
  - dependent on legislation and hence changing over time
  - important variables for distributional or behavioral analyses missing
    (sociological household, education level, etc.)
  - ... other pro’s and cons

- but:
  - more complete coverage of population:
    - longer tails
    - data on hard-to-survey groups
  - no ‘errors’ due to self-reporting (‘errors’ = deviation of administrative reality)

- illustration with micro-data IPCAL versus EU-SILC (Income Year 2014)
  - only for Flemish population (comparable to admin data: older than 16 or TI≠0)
2. what was the problem/challenge?
### 3.1 Data: EM-SILC & IPCAL – Distribution GTI in Income Brackets

#### Frequency in Thousands

<table>
<thead>
<tr>
<th>Data</th>
<th>Total</th>
<th>Less than zero</th>
<th>Zero</th>
<th>Between 0 and 25k</th>
<th>Between 25k and 50k</th>
<th>Between 50k and 75k</th>
<th>Between 75k and 100k</th>
<th>Between 100k and 125k</th>
<th>Between 125k and 150k</th>
<th>More than 150k</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPCAL</td>
<td>5,277.1</td>
<td>2.4</td>
<td>535.8</td>
<td>2,590.1</td>
<td>1,616.9</td>
<td>357.6</td>
<td>86.3</td>
<td>36.8</td>
<td>16.2</td>
<td>35.0</td>
</tr>
<tr>
<td>SILC</td>
<td>5,172.7</td>
<td>4.4</td>
<td>478.3</td>
<td>2,353.4</td>
<td>1,932.8</td>
<td>309.0</td>
<td>54.5</td>
<td>21.3</td>
<td>10.1</td>
<td>9.0</td>
</tr>
</tbody>
</table>

#### Average Income in Thousands

<table>
<thead>
<tr>
<th>Data</th>
<th>Total</th>
<th>Less than zero</th>
<th>Zero</th>
<th>Between 0 and 25k</th>
<th>Between 25k and 50k</th>
<th>Between 50k and 75k</th>
<th>Between 75k and 100k</th>
<th>Between 100k and 125k</th>
<th>Between 125k and 150k</th>
<th>More than 150k</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPCAL</td>
<td>25.4</td>
<td>-3.1</td>
<td>0.0</td>
<td>12.9</td>
<td>34.7</td>
<td>59.3</td>
<td>85.0</td>
<td>110.5</td>
<td>135.4</td>
<td>274.9</td>
</tr>
<tr>
<td>SILC</td>
<td>25.3</td>
<td>-1.1</td>
<td>0.0</td>
<td>14.4</td>
<td>34.3</td>
<td>58.7</td>
<td>85.4</td>
<td>111.1</td>
<td>136.2</td>
<td>263.0</td>
</tr>
</tbody>
</table>

#### Total Income in Thousands

<table>
<thead>
<tr>
<th>Data</th>
<th>Total</th>
<th>Less than zero</th>
<th>Zero</th>
<th>Between 0 and 25k</th>
<th>Between 25k and 50k</th>
<th>Between 50k and 75k</th>
<th>Between 75k and 100k</th>
<th>Between 100k and 125k</th>
<th>Between 125k and 150k</th>
<th>More than 150k</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPCAL</td>
<td>133,851,112</td>
<td>-7,426</td>
<td>1,252</td>
<td>33,298,711</td>
<td>56,139,767</td>
<td>21,200,858</td>
<td>7,339,055</td>
<td>4,067,402</td>
<td>2,195,975</td>
<td>9,615,518</td>
</tr>
<tr>
<td>SILC</td>
<td>128,846,097</td>
<td>-4,508</td>
<td>1,539</td>
<td>33,971,046</td>
<td>66,155,417</td>
<td>18,078,168</td>
<td>4,641,322</td>
<td>2,366,883</td>
<td>1,358,757</td>
<td>2,277,474</td>
</tr>
</tbody>
</table>
2. what we did in this project

1. reconstruct **gross taxable incomes** (GTI) instead of NTI
   (with FOD Economie – Lien Tam Co)

2. **adjustments** in accordance with **exogeneous reference totals** for...
   - population – correction for non-filers to reach population reference
     - adding zeroes at bottom up to population reference
     - this shifts the quantile values where top income groups start
     - => hazardous to calculate overall inequality measures, but top 1% ± ok
   - gross household income – correction for missing income
     - construction of income reference from national accounts (NA)
     - to account for missing capital income
       (e.g. liberating withholding tax, but also retained earnings, etc...)

3. to calculate income shares of **top 1%** and **top 10%** of GTI-distribution
2. what we did in this project

- Population of fiscal households
- Actual # of tax returns (including 0 NTI's)
- IPCAL population (excluding 0 NTI's)
- 70% of population

Graph showing population growth from 1990 to 2012, with different categories and actual data points.
3. results of step 1 and 2a

1. reconstruct gross taxable incomes (GTI) instead of NTI

2. adjustments in accordance with exogeneous reference totals for...
   - population – correction for non-filers to reach population reference
     - adding zeroes at bottom up to population reference
     - this shifts the quantile values where top income groups start
   - gross household income – correction for missing income
     - construction of income reference from national accounts (NA)
     - to account for missing capital income (e.g. liberating withholding tax, but also retained earnings, etc...)
3. results - share D10 before and after correction for 0’s and GTI

Chart showing the comparison between Published NTI, corrected for 0’s, Constructed GTI, corrected for 0’s, and Published NTI.

- Published NTI:
  - 1977: 35.2
  - 2015: 31.7

- Constructed GTI, corrected for 0’s:
  - 2015: 34.2

- Published NTI, corrected for 0’s:
  - 2015: 34.2
3. results - share P100 before and after correction for 0’s and GTI

- Published NTI: 7.7, 8.2, 8.3, 9.8
- Constructed GTI, corrected for 0’s: 7.7, 8.2, 8.3, 9.8

Graph showing the share P100 over the years from 1977 to 2015 for both published NTI and constructed GTI with and without correction for 0’s.
3. missing income

- construction *Income Control* in (3): starting from (1) and itemizing (2)
- has been compared with GTI in (6)
- to estimate ‘*missing income*’ in fiscal data
3. income reference in % of GDP and PSPTI

- **Income Control as % of Net Balance Primary Incomes in S14, 98.6**
- **Income Control as % of NNI, 86.8**
- **Income Control as % of Personal sector Pre-Tax Income, 79.7**
- **Income Control as % of GDP; 81.3**

![Graph showing income control as a percentage of Net Balance Primary Incomes, NNI, Personal sector Pre-Tax Income, and GDP between 1977 and 2014. The values range from 66.8% to 99.7%.](image-url)
3. missing income: NTI & GTI as % of Income Control in NA

- NTI as % of IC-NA: 57.3
- GTI as % of IC-NA: 66.2

<table>
<thead>
<tr>
<th>Year</th>
<th>NTI as % of IC-NA</th>
<th>GTI as % of IC-NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>1978</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>1979</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>1980</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>1981</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>1982</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>1983</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>1984</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>1985</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

Data points:
- NTI: 57.3, 84.9
- GTI: 66.2, 71.9
3. missing income by income component (in % of IC-NA)

- **Gross Taxable Income**, 33.8%
- **Property**, 63.5%
- **Replacement**, 45.0%
- **Labour**, 7.5%
- **Fin. Assets**, 99.3%
3. missing income

- 2013: 15% of IC-NA missing in tax files; down from 34% in 1990
  - how reliable is ↓ of IC-NA itself? to be investigated further

- how to allocate this ‘missing income’ (MI) across distribution?
  - without additional information, three options to allocate MI:
    1. allocate all MI outside top income groups => lower bound
    2. allocate proportionally: previous results unchanged
    3. allocate all MI to top income groups => upper bound

- future: use HFCS to impute income from FA and property income
  - but only one year (2010)
  - no returns by asset; only aggregate return and portfolio composition
3. share D10 after allocation of missing income

![Graph showing the share D10 after allocation of missing income over years from 1977 to 2013. The graph compares GTI Upper bound, GTI, Published NTI, and GTI Lower bound. The values are as follows:

- GTI Upper bound: 45.0
- GTI: 35.2
- Published NTI: 31.6
- GTI Lower bound: 29.9

The graph shows a downward trend for all three indicators, with the Published NTI showing the least variability.]

Published NTI: 31.6

GTI Lower bound: 29.9

GTI Upper bound: 45.0

GTI: 35.2
3 comparison of D10 internationally
3 share P100 after allocation of missing income
3 comparison of P100 internationally
4. summary

- evidence for increasing Belgian top income shares in last 25 years: poor

- but... very preliminary first step and caution is needed
  - 97% of NA-total of financial assets is missing
  - NA: must be understood better and not limit income to sector S14
    - incorporated income
    - retained earnings
  - growth rates of income from property and FA in NA: puzzling

- administrative data:
  - certainly valuable complementary info to surveys
  - accessability should be enhanced (micro-data)
4. to do’s

- impute capital income in fiscal data
  - based on HFCS 2010, 2014 en 2017
  - with statistical matching techniques
  - distribution of missing income from Property and FA over percentiles

- fill in other (bottom) parts of distribution (eg. replacement income, ...)
  - which boils down to: DINA-approach
  - allowing growth incidence curves over longer time period

- stretch analysis further back in time
  - based on fiscal data
  - reconstructed national accounts