A register-based input database for EUROMOD micro-simulation model

BaNoCoSS 2019
16-20 June

Kaja Söstra

19.06.2019
What is EUROMOD?

- Tax-benefit microsimulation model for the EU
- Enables researchers and policy analysts to calculate the effects of taxes and benefits on household incomes and work incentives for the population of each country and for the EU as a whole
- Developed by Institute for Social and Economic Research, University of Essex
- More information: [https://www.euromod.ac.uk/](https://www.euromod.ac.uk/)
Example: Policy effects in 2017-2018, %

Source: Masso et al. (2018)
EUROMOD structure

Policy rules (national legislation, proposals, ideas)

Base inputs

Register-based input database

Original micro-data (mainly SILC data)

EUROMOD software

User interface (program)

EUROMOD micro-data

Input data (text files)

Output data (text files)

Tax-benefit routines (‘policy parameters’) (xml-files)

Calculation engine (program)

Tools for analysis and extensions (plug-ins, add-ons)

Statistical software package (e.g. Stata)
Triggers for register-based input

- Strong interest to use EUROMOD model by Estonian analysts for analysing the impact of changes of taxes or benefits

- Problems:
  - Accuracy – EU-SILC sample data is used for input
  - Scope and level of policy data – information about small groups is missing due to the sample
  - Time-lag for input data – 2016 data added in 2018

- Parliament elections in March 2019 – need to analyse the impact of party’s promises
Prerequisites

- Experience and methodology for using register data:
  - Income variables in SILC are mainly based on the register data since 2014
  - Methodology developments for register-based census can be used for other register-based projects – residency index, partnership index, algorithms for calculating census variables
- Register data was available in SE, no new data requests from registers
- Human resources – one full-time working person
- Cooperation with EUROMOD national team

19.06.2019
Steps of project

- Joint project with Center for Policy Studies PRAXIS, financed by Ministry of Economics
- Feasibility study – February-June 2018
- Constructing input micro-data based on registers covering income and taxes of whole population in 2017 – October 2018 – January 2019
- Validation and amending the micro-data – February – April 2019
- Use by specialists of Ministry of Finance – April 2019
## Content of dataset

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number of variables</th>
<th>Examples of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>socio-demographic identifiers</td>
<td>24</td>
<td>sex, age, place of residence, citizenship, marital status, disability</td>
</tr>
<tr>
<td>labour market</td>
<td>18</td>
<td>activity status, number of months spent in full/part-time work, economic activity, number of months in retirement</td>
</tr>
<tr>
<td>income</td>
<td>34</td>
<td>employee income, income from self-employment, number of month receiving employee income</td>
</tr>
<tr>
<td>benefits</td>
<td>34</td>
<td>education-related allowances, unemployment benefits, sickness benefits, disability benefits, old-age benefits</td>
</tr>
<tr>
<td>taxes</td>
<td>11</td>
<td>Personal income tax, property tax</td>
</tr>
<tr>
<td>expenses</td>
<td>5</td>
<td>mortgage payment : interests</td>
</tr>
<tr>
<td>assets</td>
<td>9</td>
<td>number of rooms available to the household, car ownership</td>
</tr>
</tbody>
</table>
Registers used for constructing dataset

- Address Data System
- Commercial Register
- Estonian Education Information System
- State Register of Construction Works
- Estonian Unemployment Information System
- Health Insurance Information System
- Register of Mandatory Funded Pension
- Land Register
- National Defence Obligation Register
- Register of Taxable Persons
- Estonian National Pension Insurance Register
- Population Register
- Social Services and Benefits Registry
- Register of Employment
- Estonian Traffic Register
Constructing variables – activity status

- **Questionnaire-based**
  - Your main activity status – full/part-time employee, self-employed, unemployed, student, retired, disabled etc

- **Register-based**
Problems

- Metadata of register data is missing or incomplete
- Missing data, erroneous values in register data
- Incomplete data from registers – missing units
- It is not possible to construct all variables based on register – use of company car, housing costs, municipality benefits
- Registered addresses differ from real address -> register-based households differ systematically from real situation

19.06.2019
Validation of the dataset

- Dataset was macro-validated using different independent data sources:
  - SILC data for comparing the distribution of income
  - LFS for the number of employed persons
  - Statistics on benefits – the total sum and the number of recipients
  - Other relevant statistics
Size of household

- **REG 2017**
- **SILC 2017**
- **Census 2011**

<table>
<thead>
<tr>
<th>Size of Household</th>
<th>REG 2017</th>
<th>SILC 2017</th>
<th>Census 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 persons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 persons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 persons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ persons</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Distribution of disposable income by deciles, %
Impact of extreme values of income to Gini coefficient

- Full sample
- Income > 0

- Full population
- Income censored 50,000 EUR

- SILC
- Income < 50,000 EUR
- 0 <= income < 50,000 EUR
- 0 < income < 50,000 EUR

- REG
- Income < 50,000 EUR
- 0 <= income < 50,000 EUR
- 0 < income < 50,000 EUR

Gini coefficient
Difference between EUROMOD dataset and comparison statistics, %

<table>
<thead>
<tr>
<th>Category</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of employed persons</td>
<td>3.5%</td>
</tr>
<tr>
<td>Number of registered unemployed</td>
<td>-3%</td>
</tr>
<tr>
<td>Total income from employment</td>
<td>-4%</td>
</tr>
<tr>
<td>Old age pension</td>
<td>-2%</td>
</tr>
<tr>
<td>Parental benefits</td>
<td>-6%</td>
</tr>
<tr>
<td>Disability benefits</td>
<td>-8%</td>
</tr>
<tr>
<td>Sickness benefits</td>
<td>4%</td>
</tr>
</tbody>
</table>

19.06.2019
# EU-SILC vs register based input

<table>
<thead>
<tr>
<th></th>
<th>EU-SILC/Estonian-SILC</th>
<th>Administrative registers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population coverage</strong></td>
<td>Sample of ca 15 thousand persons</td>
<td>Whole population (1.3 mln)</td>
</tr>
<tr>
<td><strong>Time lag</strong></td>
<td>2-3 years</td>
<td>1-2 months</td>
</tr>
<tr>
<td><strong>Panel data features</strong></td>
<td>Partial (rotating panel)</td>
<td>Complete</td>
</tr>
<tr>
<td><strong>Residency &amp; household structure</strong></td>
<td>Complete</td>
<td>Partial</td>
</tr>
<tr>
<td><strong>Socio-demographics</strong></td>
<td>Mainly current (at time of interview)</td>
<td>Annual or monthly</td>
</tr>
<tr>
<td><strong>Incomes</strong></td>
<td>Previous calendar year (mainly register-based), non-cash incomes</td>
<td>Annual or monthly (100% register-based), only cash incomes</td>
</tr>
<tr>
<td><strong>Tax deductions</strong></td>
<td>-</td>
<td>Detailed (annual)</td>
</tr>
<tr>
<td><strong>Consumption</strong></td>
<td>Partial (housing, child care)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Assets</strong></td>
<td>Partial (housing, cars)</td>
<td>Detailed (domestic only)</td>
</tr>
<tr>
<td><strong>Quality problems</strong></td>
<td>Non-response, measurement errors</td>
<td>Incomplete or out-of-date information, income underreporting</td>
</tr>
</tbody>
</table>

*Stat.ee* 10.05.2018

---

1.3 mln

---

**EU-SILC vs register based input**

<table>
<thead>
<tr>
<th></th>
<th>EU-SILC/Estonian-SILC</th>
<th>Administrative registers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population coverage</strong></td>
<td>Sample of ca 15 thousand persons</td>
<td>Whole population (1.3 mln)</td>
</tr>
<tr>
<td><strong>Time lag</strong></td>
<td>2-3 years</td>
<td>1-2 months</td>
</tr>
<tr>
<td><strong>Panel data features</strong></td>
<td>Partial (rotating panel)</td>
<td>Complete</td>
</tr>
<tr>
<td><strong>Residency &amp; household structure</strong></td>
<td>Complete</td>
<td>Partial</td>
</tr>
<tr>
<td><strong>Socio-demographics</strong></td>
<td>Mainly current (at time of interview)</td>
<td>Annual or monthly</td>
</tr>
<tr>
<td><strong>Incomes</strong></td>
<td>Previous calendar year (mainly register-based), non-cash incomes</td>
<td>Annual or monthly (100% register-based), only cash incomes</td>
</tr>
<tr>
<td><strong>Tax deductions</strong></td>
<td>-</td>
<td>Detailed (annual)</td>
</tr>
<tr>
<td><strong>Consumption</strong></td>
<td>Partial (housing, child care)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Assets</strong></td>
<td>Partial (housing, cars)</td>
<td>Detailed (domestic only)</td>
</tr>
<tr>
<td><strong>Quality problems</strong></td>
<td>Non-response, measurement errors</td>
<td>Incomplete or out-of-date information, income underreporting</td>
</tr>
</tbody>
</table>

*Stat.ee* 10.05.2018

---

1.3 mln
Additional statistics for small regions
Family benefits by locality, euro
Conclusions and future plans

- The first trial for constructing the dataset of income of whole population was successful.
- Dataset can be used for EUROMOD model and other purposes.
- Next steps:
  - to construct datasets for 2018 and 2013-2016
  - to develop monthly updates of dataset
References


19.06.2019
Thank you for attention!