1. Modelling framework at the CBR
2. Administrative data: access & management
3. Input database for the model TATRASK: generation process & challenges
4. Microsimulation model TATRASK: simulated policies
Council for Budget Responsibility (CBR) in Slovakia

- independent authority for monitoring and evaluation of Slovak public finance
- established by the Constitutional Act on Fiscal Responsibility in 2011

Main tasks of the CBR include
- evaluating long-term sustainability of public finance
- assessment of legislative proposals
- assessment of tax and spending forecasts in the budgetary process

Independent modelling framework
- essential requirement for the CBR tools
Impact assessment models in CBR

SIMTASK – microsimulation model of tax & transfer system (static)
• runs on survey data EU-SILC
• underlying part of labour-supply module, indirect tax-tool and micro-macro module for long-run effects (all developed in CBR)
• developed in CBR since 2012
• simplified version available online http://simtask.rozpoctovarada.sk

TATRASK (Taxes and Transfers in Slovakia) – microsimulation model of tax & transfer system
• new model using admin data from different sources (static)
• objective is to evaluate effects of changes in taxes & transfers (distributional, budgetary)
• technical support in development provided by the JRC Sevilla (finalized in 2022)
• still work in progress
Input datasets for model TATRASK – the core

(1) Central Registry of Inhabitants (CRoI)
   • represents national population (available years 2019, 2022)
   • demographic information: age, gender, relationship and marital information, residency
     input dataset for simulation of households / families

(2) Social Insurance Agency (SIA)
   • pensions (old age, disability, survival)
   • contributory benefits (sickness, maternity, unemployment)
     contributory benefits are simulated using information on contribution history
   • paid social insurance contributions
     labour income is inferred from paid contributions

(3) Central Office of Labour and Family (CoLSaF)
   • family benefits (child benefits, parental allowance etc.)
   • social assistance (material need benefit, monetary compensations for disabilities etc.)
   • registered unemployed and participation in active labour market policies
Input datasets for model TATRASK – the core

(1) Central Registry of Inhabitants (CRoI)
• represents national population (available years 2019, 2022)
• demographic information: age, gender, relationship and marital information, residency
  input dataset for simulation of households / families

(2) Social Insurance Agency (SIA)
• pensions (old age, disability, survival)
• contributory benefits (sickness, maternity, unemployment)
  contributory benefits are simulated using information on contribution history
• paid social insurance contributions
  labour income is inferred from paid contributions

(3) Central Office of Labour and Family (CoLSaF)
• family benefits (child benefits, parental allowance etc.)
• social assistance (material need benefit, monetary compensations for disabilities etc.)
• registered unemployed and participation in active labour market policies

(4) Health insurance companies (HIC)
• information on all insured persons (also by the state, usually with no labour income)
• paid health insurance contributions
  additional information on labour income (self-employed, armed forces)
Input datasets for model TATRASK – additional sources

Access already granted or agreements expected to be signed soon for the additional data

(5) Tax returns
  • information on all types of taxable income
    additional information on non-labour income, income of people working abroad

(6) Registry of students
  • data on full-time students in Slovakia, from elementary schools to universities
    student status needed for the definition of dependent child
Input database of model TATRASK

Preparation steps

Data anonymization
linking individuals across datasets with a unique random identifier

Monthly structure of the input dataset (balanced panel)
every person has exactly 12 rows representing data for every month of the year

Time consistency of data - accrual principle is applied
• incomes reported in the month they are earned/eligible
• transfers reported in the month they are paid (one month later) – paid transfers are shifted one month back

Challenges
• Underlying database for the model - definition of relevant population
• Definition of households/families for the simulations
TATRASK – simulation policies

Model TATRASK is work in progress

- model is running and computing first partial results
- not ready to use in policy simulations
- simulations are computed on monthly basis
- new simulations: contributory benefits

<table>
<thead>
<tr>
<th>Simulations of the tax &amp; transfer system</th>
<th>Model TATRASK (admin data)</th>
<th>Model SIMTASK (survey data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal income tax</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Social insurance contrib. (all components)</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Health insurance contrib. (all components)</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Contributory benefits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternity benefit</td>
<td>simulated</td>
<td>partially simulated*</td>
</tr>
<tr>
<td>Sickness benefit</td>
<td>simulated</td>
<td>not simulated</td>
</tr>
<tr>
<td>Nursing benefit</td>
<td>simulated</td>
<td>not simulated</td>
</tr>
<tr>
<td>Pregnancy benefit</td>
<td>simulated</td>
<td>partially simulated*</td>
</tr>
<tr>
<td>Unemployment insurance benefit</td>
<td>simulated</td>
<td>partially simulated*</td>
</tr>
<tr>
<td>State social support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child birth allowance</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Child benefit and surcharge</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Multiple birth allowance</td>
<td>simulated</td>
<td>not simulated</td>
</tr>
<tr>
<td>Parental allowance</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Social assistance benefits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material need benefit</td>
<td>simulated</td>
<td>simulated</td>
</tr>
<tr>
<td>Special allowance</td>
<td>simulated</td>
<td>not simulated</td>
</tr>
<tr>
<td>Compensations of severe disability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cash benefit for personal assistance</td>
<td>partially simulated*</td>
<td>not simulated</td>
</tr>
</tbody>
</table>

* partially simulated = eligibility is taken from underlying data
Simulation results (preliminary)

Ratios of simulated values by TATRASK/SIMTASK to official statistics (aggregate yearly volume in 2019). The closer the ratio is to value 1, the better simulation fits the official statistics.
Summary & future steps

Ongoing development of TATRASK model
• inclusion of additional data sources (still missing some income sources)
• updating the input database to match the policy year
• updating/new simulation policies

Open issues
• relevant population – which persons should be included in the final database
• model should evaluate correctly
  income distribution impact – assessment of financial situation in families
  fiscal impact – all tax liabilities paid and transfer eligibilities
Thank you for your attention!
Background slides
Input database of TATRASK

Final input dataset includes (as of today)

- 70 mil. rows of observation (monthly structure)
- 180 variables
- 42 GB in size

Data processing model structure
Relevant population

Individuals to be included in the final input database:
working (paying taxes) in Slovakia and eligible for transfers

Main objective of the model: provide distributional analysis
correct information on income is substantial, if income is missing means-tested transfers will be overestimated

Relevant population based on merged data from CRoI & income databases

<table>
<thead>
<tr>
<th>Main CRoI file (in mil. persons, 2019)</th>
<th>8.13</th>
</tr>
</thead>
<tbody>
<tr>
<td>- people with no income information in SIA &amp; HIC</td>
<td>-1.48</td>
</tr>
<tr>
<td>- cleaning of CRoI (died before start of ref year, over 110y. etc.)</td>
<td>-1.41</td>
</tr>
<tr>
<td>Relevant population (merged CRoI, SIA &amp; HIC databases)</td>
<td>5.24</td>
</tr>
<tr>
<td>note: people with income in SIA &amp; HIC, but not present in CRoI</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Census 2021 residence of inhabitants
• 5.45 mil. permanent
• 5.23 mil. actual
Households & families

Final input database created from admin data:
• information on relationship (parentship/partnership) & residence of persons
• no information on dwelling households

Challenges in creating households & families in the final dataset:

**Family** - persons in a relationship (husband/wife), parents & children up to 25 years
• assumption: unmarried couples with common children and the same residence address are considered as family