Microsimulation model

Martín Cicowiez         Marco V Sánchez
(CEDLAS-UNLP)         (UN-DESA)

Presentation for the Final Workshop of the Project
“Strengthening Macroeconomic and Social Policy
Coherence through Integrated Macro-Micro
Modelling”, Amman, 3-5 June 2014

CGE models and inequality

• A typical CGE model possesses one or more groups of
  “representative households”
  – Only between-group income distribution
  – Omits within-group income distribution
  – Both types of income distribution affect total income
    distribution (and poverty outcomes!)
  – Also, how do we know which workers are more likely to
    change position in the labour market?
    • e.g.: if, as a result of a policy simulation, the
      unemployment rate increases: who is expected to lose
      her/his job?
  – In MAMS: The goal of reducing extreme poverty (MDG 1) is
    not targeted in the same way as the other MDGs
A microsimulation model can be used: what is it?

- Model that uses information at the level of microeconomic individual agents (individuals, households, firms).
- As such, it permits to evaluate effects of policy or other shocks on those individual agents
  - Often through ex-ante evaluations

A microsimulation model can be used: what is it? – cont.

- Typically it requires micro-data from a household survey covering:
  - socio-economic characteristics of individuals
  - labour-market status and labour incomes
  - household spending
- Changes in budget constraints are simulated
- Works at the partial equilibrium level
- It implies no simultaneous modelling of prices, wages, or macro processes
Top-down macro-micro approach

- CGE (macro) simulation results taken and applied to the full distribution as given by a micro data set (i.e., the household survey)
  - Assumption: there are no further feedback effects
- Macro modelling: provides simulation results on employment, wages, non-labour factor income, etc.
- Micro modelling: permits to transform macro modelling results into results given by micro-data, using the household survey

Top-down modelling approach
Alternative approaches

• Parametric
  – generally imply using a system of equations that determines occupational choice, returns to labour and human capital, consumer prices and other household (individual) income components

• Non-parametric
  – generally imply seeking individuals with similar characteristics to simulate certain change (for example, a change in labour income for an individual that moves from unemployment to employment)
  – occupational shifts may be proxied by a random selection procedure within a segmented labour-market structure

Per capita household income

• Generally, the top-down approach is used to, after simulated (counterfactual) changes, generate a simulated per capita household income defined as follows:

\[
ypc_h = \frac{1}{n_h} \left[ \sum_{i=1}^{n_h} yp_{hi} + yq_h \right]
\]

where,
• \( n_h \) = size of household \( h \)
• \( yp_{hi} \) = labour income of member \( i \) of household \( h \)
• \( yq_h \) = sum of all non-labour incomes of the household
What if poverty is calculated based on consumption?

• ¿Use poverty incidence indicators based on income anyway?
• Alternative 1: in order to match base-year poverty incidence as measured from both income and consumption, there may be two options:
  – recalculate per capita income to match it with per capita consumption
  – recalculate the poverty lines
• Alternative 2: use consumption that changes when income changes (based on marginal propensity to consume by product).

Non-parametric microsimulations: where do we start?

• The method analyses effects of a change in the labour market structure on poverty and inequality.
• It is a counterfactual analysis: what would poverty and inequality indicators look like had the labour market structure be different from the observed one?
  – For example, one could use the labour market structure as recorded in MAMS instead of the observed labour market structure as recorded in a given household survey.
Non-parametric microsimulations: where do we start? – cont.

- The economically active population (EAP) is split up into $j$ groups according to
  - sex (2; but not in MAMS)
  - level of skills (for example, 3 in MAMS)
  - (other individual’s characteristics?)
- The occupied population is split up into $k$ groups according to
  - sector of employment
  - occupational category
  - (individual’s characteristics)
- The counterfactual structure of the labour market can be defined arbitrarily or as a result of a macro-simulation.

### Classification of population in working age ($j$ groups = 3)

<table>
<thead>
<tr>
<th>Participation status</th>
<th>Unskilled</th>
<th>Semi-skilled</th>
<th>Skilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Employed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Un-employed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification of employed population
(EXAMPLE: j groups = 3; k groups = 3; 9 labour categories)

<table>
<thead>
<tr>
<th>Sector</th>
<th>Unskilled</th>
<th>Semi-skilled</th>
<th>Skilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modelling of the labour market

- The labour market structure $\lambda$ is a function of the following variables:
  \[
  \lambda = \lambda (P, U, S, O, W_1, W_2, M)
  \]
  - $P$ - participation rates for labour type $j$
  - $U$ - unemployment rate for labour type $j$
  - $S$ - employment structure by production sector
  - $O$ - employment structure by occupational category
  - $W_1$ – remuneration structure by sector
  - $W_2$ – overall average remuneration
  - $M$ - composition of employment by individual’s characteristics (e.g., skill level)
How does it work?

• A random number is assigned to each person at working age
• Population at working age is ordered according to:
  – activity condition (active vs inactive),
  – employment condition (employed vs unemployed)
  – economic sector
  – occupational category
  – education level
  – random numbers

How does it work? – cont.

New labour market structure $\lambda^*$

– Individuals become active/inactive, employed/unemployed, change their occupational position and/or level of education
  • How? See example in next slide.
– Income (YPI) is assigned to all those individuals who, according to $\lambda^*$, become employed, or change their occupational position and/or level of education
  • How? A non-parametric (random) process is also employed to assigned incomes
– Income of all those individuals that become unemployed or inactive are set equal to zero
Example: effect of a change in the unemployment rate of skilled men workers (N=100)

<table>
<thead>
<tr>
<th></th>
<th>Simulation 1</th>
<th>Simulation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Unemployment rate falls to 6%</td>
<td>90 Unchanged</td>
<td>90 Unchanged</td>
</tr>
<tr>
<td></td>
<td>The first 4 unemployed become employed</td>
<td>The last 2 employed become unemployed</td>
</tr>
<tr>
<td>Unemployed</td>
<td>10</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>6 Unchanged</td>
<td>2 Unemployed</td>
</tr>
</tbody>
</table>

How does it work? - cont.

- Same randomized procedure applied for other shifts:
  - active vs inactive (P)
  - employed vs unemployed (U)
  - employment by sector (S), occupational category (O), kills (M)

- Wages
  - Changes in $W_1$: all YPs within each of the 9 labour categories are multiplied by an adjustment factor, maintaining the overall average wage/labour income level fixed
  - Changes in $W_2$: all YPs are multiplied by an adjustment factor to reflect the change in the average wage/labour income

- Simulated YPs are used to compute household incomes per capita (YPC).

- New inequality indicators (Gini, Theil, etc.) and poverty indicators (for alternative poverty lines) are computed
Key assumptions

- We do not need a full model of the labour market – there are only individuals changing their j or k groups

- A randomized process is applied to simulate the effects of changes in the labour-market structure
  - It assumes that, on average, the effect of the random changes correctly reflects the impact of the actual changes in the labour market

- Because of the introduction of a process of random assignation, the micro-simulations are repeated a large number of times in Monte Carlo fashion this allows constructing 95 per cent confidence intervals for the indices of inequality and poverty

In summary:

- From CGE model, changes in the labour market structure are applied (individually or sequentially) to micro data, affecting the overall income distribution:

\[ \lambda^* = \lambda^*(p^*, U^*, S^*, O^*, W^*_{1}, W^*_{2}, M^*) \]

- Who moves? Determined through a random process which generates a new income distribution

- Micro-simulations are repeated many times in Monte Carlo fashion to compute confidence intervals for inequality and poverty indicators that are statistically significant
Advantages vs disadvantages

Advantages:
• Enables to analyse the impact of a wide range of labour-market parameters, individually or sequentially
• Shows separate and combined effects of each type of labour market shift (e.g. unemployment change, wage change, etc.) on poverty and inequality outcomes
• It does not demand econometric estimation

Possible disadvantages:
• Behaviour is not modelled
• Results in sequential application may depend on the order in which the sequence of labour-market parameter changes is applied ("path dependence")

In a dynamic setting

• Observed survey data may only be available for the base year and perhaps a few years beyond that, but certainly not for the forward simulation period.
• Only one household survey is used. The year for which this was conducted becomes the base year of the microsimulations.
  – It may be different from the base year of the CGE model; generally more recent.
• For lack of additional modelling of demographic shifts and labour participation, some changes in the population structure as a result of migration, for example, are not taken into account.
• For variables of the labour market taken from the CGE model scenarios, changes for each year are calculated relative to the base year of the microsimulations. These changes are used to conduct the microsimulation as described.
Example of sequential and cumulative effects for changes in the labour-market parameters under a baseline scenario

<table>
<thead>
<tr>
<th>Year</th>
<th>Total poverty incidence (% of population)</th>
<th>Extreme poverty incidence (% of population)</th>
<th>Gini coefficient for labour income</th>
<th>Gini coefficient for per-capita household income</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S</td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S + O</td>
<td></td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
</tr>
<tr>
<td>U + S + O + W₁</td>
<td></td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂</td>
<td></td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂ + M</td>
<td></td>
<td>20.7</td>
<td>4.3</td>
<td>0.461</td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>20.6</td>
<td>4.3</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S</td>
<td>20.6</td>
<td>4.3</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S + O</td>
<td></td>
<td>20.6</td>
<td>4.3</td>
<td>0.461</td>
</tr>
<tr>
<td>U + S + O + W₁</td>
<td></td>
<td>19.8</td>
<td>4.1</td>
<td>0.456</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂</td>
<td></td>
<td>19.6</td>
<td>4.1</td>
<td>0.456</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂ + M</td>
<td></td>
<td>19.5</td>
<td>4.1</td>
<td>0.456</td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>20.5</td>
<td>4.2</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S</td>
<td>20.5</td>
<td>4.2</td>
<td>0.461</td>
<td>0.497</td>
</tr>
<tr>
<td>U + S + O</td>
<td></td>
<td>20.4</td>
<td>4.2</td>
<td>0.461</td>
</tr>
<tr>
<td>U + S + O + W₁</td>
<td></td>
<td>18.1</td>
<td>3.8</td>
<td>0.447</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂</td>
<td></td>
<td>16.6</td>
<td>3.6</td>
<td>0.447</td>
</tr>
<tr>
<td>U + S + O + W₁ + W₂ + M</td>
<td></td>
<td>16.5</td>
<td>3.6</td>
<td>0.447</td>
</tr>
</tbody>
</table>

Source: CGE model and microsimulation results for Costa Rica.

Changes in non-labour incomes

- Changes in non-labour incomes such as government transfers and remittances from abroad can be taken into account.

- Computed from the CGE model scenarios for each year relative to the base year of the microsimulations.

- These changes are proportionally scaled up or down to re-estimate the new household income distribution.
Changes in non-labour incomes

- Changes in non-labour incomes such as government transfers and remittances from abroad can be taken into account.
- Computed from the CGE model scenarios for each year relative to the base year of the microsimulations.
- These changes are proportionally scaled up or down to re-estimate the new household income distribution.

Decomposing non-labour incomes

- Recall:

\[ y_{pc} = \frac{1}{n_h} \left[ \sum_{i=1}^{n_h} y_{p_{hi}} + y_{q_h} \right] \]

where,
- \( n_h \) = size of household \( h \)
- \( y_{p_{hi}} \) = labour income of member \( i \) of household \( h \)
- \( y_{q_h} \) = sum of all non-labour incomes of the household

Let us assume that \( y_{q_h} = \text{transfgov} + \text{remittances} + \text{residual} \)

where, \textit{residual} allows to match changes in \( y_{pc} \) (with respect to MAMS) given the changes in \( y_{p_{hi}}, \text{transfgov} \) and \textit{remittances}
Residual Effect

- Final step in the microsimulation model: adjusts the average income per capita for each representative household in line with MAMS results.
- In order to adjust the micro data such that the percentage change in the household per capita income matches the change in household per capita income in MAMS simulations, the following adjustment is carried out for each individual belonging to the representative household $h$:

$$ypc_{h, resid} = \frac{ypc_{h, trrow}}{\mu_{h, trrow}} \mu_{h, obs} (1 + \Delta ypc_h^*)$$

Residual Effect – cont.

where

- $ypc_{h, resid}$ = household per capita income for RA $h$ in “residual” effect
- $ypc_{h, trrow}$ = household per capita income for RA $h$ in “trnsfr RoW” effect
- $\mu_{h, trrow}$ = mean of $ypc(h, resid)$
- $\mu_{h, obs}$ = mean of observed household per capita income
- $\Delta ypc_h^*$ = change in household per capita income from MAMS simulation
Residual Effect – cont.

• Thus, the “Residual Effect” implicitly accounts for changes in all items not previously considered (i.e., non-labor and non-transfer incomes) such as natural resource and capital rents
  – those income sources enter the household budget constraint in MAMS and thus have an income effect on consumption