Using micro-simulation in EU-SILC for early estimates of income: strengths and limitations

Vítor JUNQUEIRA[1]

INE – Statistics Portugal

Abstract. This paper aims to promote the debate on the production of early estimates of indicators such as the At-Risk-Of-Poverty through the use of micro-simulation techniques by the national statistics offices. It provides a brief presentation of a simple micro-simulation model, along with its results, limitations and strengths.

1. Introduction

The European Union Statistics on Income and Living Conditions (EU-SILC) are collected every year, providing all the data needed in the calculation of indicators such as the 'At-Risk-Of-Poverty (AROP) rate'[2] or the Europe 2020 Strategy's 'Population at Risk of Poverty or Exclusion'. Over the past years, most EU-SILC national teams manage to reduce the gap between the release date and the income reference year of such indicators. Further progress is still possible and desirable, especially with the implementation of early estimates procedures which may contribute to major gains in timeliness at some (affordable or not) loss of accuracy.

2. Urgency for greater timeliness

More timely available data from income surveys like EU-SILC has always been a demand from policy makers and researchers due to the important role of such indicators as the At-Risk-of-Poverty Rate (AROP rate) to monitor the evolution of society, the impact of public policies and, in particular, the effectiveness of anti-poverty policies and measures. It has become even more important over the past years of great instability on income distribution due both to the economic crisis and to the several changes in social and fiscal policies, including austerity measures. The instability over income distribution may also generate "methodological traps" that make harder for the public in general and even the untrained policy makers to understand the potentially ambiguous evolution of the AROP rate[3].

Timeliness has also been a highly ranked priority in the agendas of the NSIs, with most of them consistently stretching the gap between the income reference year and the release of first estimates of indicators. In Portugal alone, this gap has been greatly reduced from two years (745 days) observed for EU-SILC 2003 to the 395 days achieved with EU-SILC 2014 or the targeted 352 days for EU-SILC 2015 (for the first time, the AROP rate will hopefully be released in the same year as of the survey). At the European level, income data for all the countries may become two years out of date, with profound implications for the usefulness of the EU-SILC instrument for policy purposes (Eurostat, 2015).

[1] The author is grateful to Helena Cordeiro, Emília Saleiro, Carlos Farinha Rodrigues and Eduarda Góis at Statistics Portugal for their valuable comments and suggestions. This short paper relies essentially in previous work (Junqueira, 2015), as presented in the EU-SILC conference in Lisbon (October 2014) and in the European Micro-simulation Modelling for Policymaking conference in Brussels (May 2015).

[2] The share of persons with an equivalent disposable income below the risk-of-poverty threshold, which is set at 60 % of the national median equivalent disposable income (after social transfers).

[3] Briefly, that is the situation where a general drop in the society incomes may not necessarily lead to an increase in poverty rate, as commonly expected and other occasional indicators would illustrate. This happens because the poverty threshold also drops, occasionally reducing the number of people at risk of poverty. Although reflecting the "relative" nature of the AROP rate it is not easily understood by everyone.

3. Using micro-simulation in the production of early estimates

As seen above, progress in timeliness has been notorious. And there is still room for improvement. Although it may prove to be impossible to continue to stretch much further the gap between the reference year and the date the indicators are released[4], some efforts are being taken outside NSIs and Eurostat in order to obtain complementary and timelier data. This has been one of the main drivers for micro-simulation models such as EUROMOD in the past decades. The term "nowcasting", a contraption of "now" and "forecasting", has become increasingly popular in income distribution studies.

Even a simple micro-simulation model as the one used for the exercise here presented may provide good results on this subject. This exercise was done exclusively at Statistics Portugal and relied essentially on EU-SILC data for inputs.

3.1. Methodology in brief[5]

As for most countries, the Portuguese EU-SILC for the year *n* provides information on income from the year *n-1*. For a fewer group of countries, EU-SILC also gathers information concerning the current year *n*, such as *gross monthly earnings for employees* or *self-defined current economic status*[6]. By combining both past and current information, the model provides some insight on what may have been the evolution of the status in employment of the individual. Then, with some degree of simplification, the model estimates the individual's income for the current year. Examples:

a) If someone has no income from employment in year *n-1* and g*ross monthly earnings for employees* are now reporting positive, the model assumes a transition from unemployment to employment. The simulated income for the year *n* will consider an entire year of employment income by extrapolating the *gross monthly earnings for employees*.
b) *Au contraire*, if someone had income from employment in the year *n-1* and now reports no *gross monthly earnings for employees* while answering *unemployed* to the *elf-defined current economic status* question, the model assumes the opposite transition and removes employment in the simulated incomes for the year *n*. Then, the model simulates an amount for unemployment benefit to be considered for the entire year (or less than one year, according to the recipient's age).

These are just two of the many transitions that the model simulates based upon information collected exclusively from a single EU-SILC operation/year[7]. It is a short term model designed for simulating transitions in employment between two years and – as in other static micro-simulation models – it also neglects any demographic change (people will not grow older, there are no births nor deaths, households keep their familial structure and weights are not recalculated). It does not consider transitions between employment and retirement or new entries into the active population.

Along with the simulation of transitions, the model also deals with most of the changes in policies which had direct impacts on the income distribution. Being able to assess such impacts is crucial for countries like Portugal, where in recent years several changes in policies at the social and the fiscal fields

[4] It shouldn't go unnoticed that using the previous year as the income reference year constitutes itself a clear limitation for improving timeliness.
[5] For a more detailed description of the methodology, see (Junqueira, 2015).
[6] Variable PY200G is only available for nine countries: Austria, Greece, Hungary, Ireland, Italy, Poland, Portugal, Spain and the United Kingdom.
[7] Basically, only five variables are considered: income from employment in *n-1*, income from self-employment in *n-1*, income from unemployment benefits in *n-1*, gross monthly earnings for employees in *n* and self-defined current economic status in *n*.

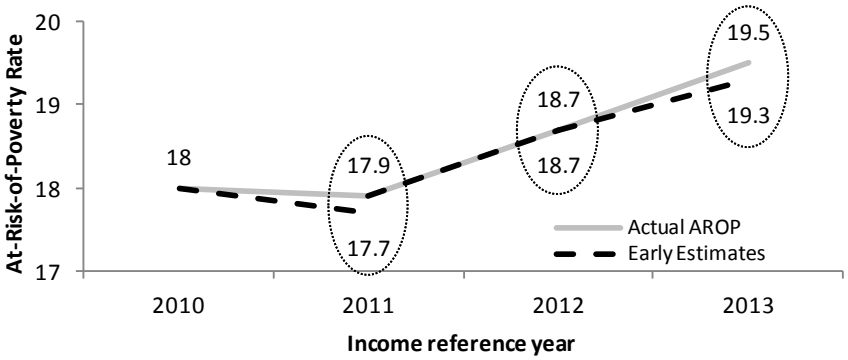
took place. Depending of the year at stake, the model simulates at the micro level the impacts of policies changes such as the cuts in public wages, the cuts in pension income or the significant changes in withholding taxes on wages and pensions.

Apart from the transitions and policies changes, the model also updates incomes according to different factors associated to each type of income: for instance, wages are updated according to nationwide average wage variation, while other variables may be updated according to the Consumer Price Index, etc.

3.2. The good results

This exercise has been done for EU-SILC 2011 to 2013 datasets, thus providing "what would have been early estimates" for the AROP rates concerning income for years 2011 to 2013, respectively, i.e., those officially provided by EU-SILC 2012 to 2014, respectively. The following chart shows the evolution of the official AROP indicator (main curve) as compared to what were the "early estimates" for the following year as calculated in each operation (dotted curves).

Chart #1: AROP rate: actual vs. early estimates



Source: EU-SILC 2011-2014 and Junqueira (2015)

The chart shows that "early estimates" produced upon each EU-SILC dataset were quite similar to the actual AROP rates computed on every following dataset:

(i) For instance, EU-SILC 2011 dataset, which had been used to compute 2010's rate of 18%, was also useful to the simulation of 2011's early estimate, 17.7, which compares to the 17.9 actual rate published only one year later when EU-SILC 2012 was finalised.

(ii) The model was then applied again, with the due modifications (in policy changes, for instance) within the 2012 dataset. The result was even closer, as the "early estimate" was exactly the same as the official rate that EU-SILC 2013 showed a year later, 18.7.

(iii) Finally, the early estimate using the model on EU-SILC 2013 was 19.3 compared to the actual 19.5 revealed by EU-SILC 2014.

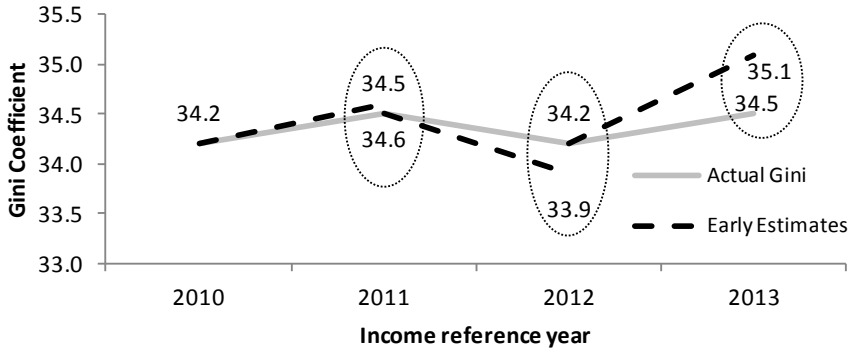
3.3. The not so good results

The resulting AROP rates seem quite satisfactory when compared to the actual values that were to be known later in time, as seen above. Nonetheless, some caution is advised on reading these apparently satisfactory results. Micro-simulation exercises like this can be helpful on determining trends and on supplying resources to understand how multiple factors can influence the evolution of the AROP rate as it is commonly designed. Still, it should not be expected a complete match between the simulated outcome and the actual values to be published in the near future. There is, as shown above, a significant number of

simplifying assumptions. There are demographic changes along with other external factors that are not accounted by the model. All this combined should provide deviations up to a certain amount on the simulated results as compared to actual data gathered by the regular EU-SILC operation.

The model seems capable of generating a new pseudo-distribution and then accounting for the number of people whose incomes fall below a certain line directly derived from the median of the distribution (this is basically the definition of the AROP rate). Is the new pseudo-distribution fully comparable to the actual one, i.e., the one that should be represented by the following year's survey? Chart #2 shows it is not always the case.

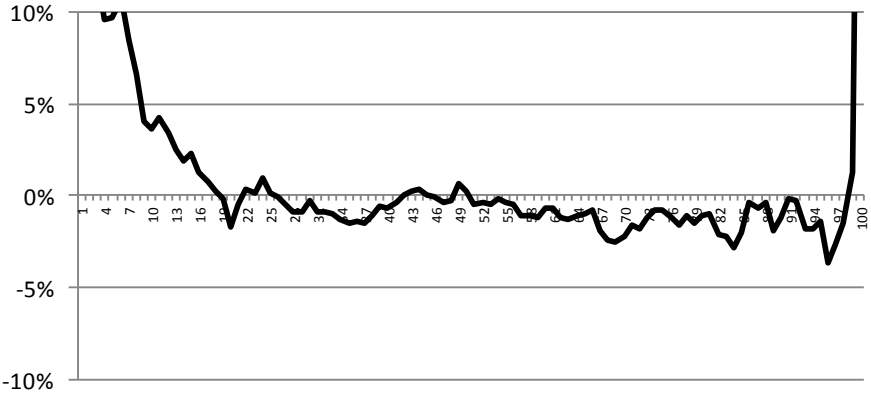
Chart #2: Gini coefficient: actual vs. early estimates



Source: EU-SILC 2011-2014 and author's own calculations

The chart above is drawn in the same fashion as chart #1, but instead of the AROP rate it shows the comparison between actual and simulated values for the Gini coefficient, an indicator used in the measurement of inequalities which uses the broad spectrum of the distribution. Although the directions match the real trend, the exercise no longer seems to produce such good results as before. Some of the limitations above listed may explain this semi-failure. For instance, the simplifying assumption in the extrapolation of the *gross monthly earnings for employees* for the entire year for someone who was previously unemployed may prove to be too generous when compared to reality, where there may be other situations in-between (or even more generous transitions). These occasional "too much" or "too less" transformations will affect the simulated distribution, mainly in the edges, as shown in chart #3.

Chart #3: Percentiles of equivalent income (individuals): actual vs. early estimates ratios



Note: This chart represents the change in % of each percentile from actual to simulated amounts. A positive % (e.g., curve above the 0% line) means the simulation amount is higher by x% than the actual amount.
Source: EU-SILC 2011-2014 and author's own calculations

4. Conclusions, open questions and proposals for the future

- The exercise shows that micro-simulation is a powerful tool at the disposal of the research community which can also be used by the NSIs and Eurostat. By assembling information concerning different moments in time but collected from the same statistical operation, this exercise was able to produce short term anticipations of the AROP rate, potentially responding to the increasing demand for timeliness.
- An early estimate of the poverty rate for 2015 derived from this model could be given only a few weeks after the official rate for 2014 was published. From a wider perspective, a broader range of indicators such as the GDP growth (flash estimate), the unemployment rate and the AROP rate (early estimate) concerning the year *n* could all be made available to the public and the policy-makers during the first quarter of year *n*+1, which would represent an important milestone for timely available statistics.
- Besides the production of early estimates, this exercise may also facilitate stakeholders understanding occasional "methodological traps" concerning the AROP rate, especially in times of instability such as the present for countries facing austerity measures.
- This methodology can easily be replicated by other NSIs, as long as the national EU-SILC gathers information related to the current year such as *gross monthly earnings for employees*, possibly serving as a solution to overcome difficulties in observing the timelines required by the regulation.
- This exercise also shows that the anticipation of the AROP rate and other indicators could significantly benefit from a revision in EU-SILC questions concerning current income and status in employment[8]. For instance, *gross monthly earnings for employees* could be available to a larger number of countries as well as other main sources of income concerning the survey's 'current moment' such as self employed income or social transfers.
- Nonetheless, caution is highly advised, as there are always limitations associated with simple or complex micro-simulation models that may put *nowcasts*, early estimates or other anticipations at risk. In the end, trust and confidence in statistics production might be put at risk. Clearly, there is a trade-off between timeliness and accuracy which has to be discussed and better understood both at the national and European levels.
- Given the potential improvements to timeliness that early estimates and *nowcasting* can provide and the important questions that arise from their use, should it be specifically addressed in the Framework Regulation on Integrated European Social Statistics under current discussion?

We have aimed to encourage and contribute to the debate given the importance for analysis and decision-making of early indications onto so important aspects of today´s society.

References:

Atkinson, T., Cantillon, B., Marlier, E., & Nolan, B. (2002). Social Indicators: The EU and Social Inclusion. Oxford: Oxford University Press.

Eurostat (2011), EU-SILC 065 (2011 operation) Description of Target Variables. Version May 2011. Eurostat.

Eurostat (2012), EU-SILC 065 (2012 operation) Description of Target Variables. Version May 2013. Eurostat.

[8] Actually, Eurostat opened recently an invitation to tender concerning this field of research (Eurostat, 2015).

Eurostat (2013), EU-SILC 065 (2013 operation) Description of Target Variables. Version September 2012. Eurostat.

Eurostat (2014), EU-SILC 065 (2014 operation) Description of Target Variables. Version March 2015. Eurostat.

Eurostat (2015), Invitation to tender for the supply of statistical services – Methodological work for the development of flash estimates, aiming at now casting income based indicators, ESTAT/F/2015/014, Eurostat.

Forster, M. (1994), Measurement of Low Incomes and Poverty in A Perspective of International Comparisons, OECD Labour Market and Social Policy Occasional Papers, No. 14, OECD Publishing

Immervoll, H., C. O'Donoghue & H. Sutherland (1999), An Introduction to EUROMOD, EUROMOD Working Paper no. 0/99, Department of Applied Economics, University of Cambridge.

Junqueira, V. (2015), "Using micro-simulation in EU-SILC to compute At-Risk-of-Poverty rate early estimates" (Paper presented at the "European Micro-simulation Modelling for Policymaking" conference in Brussels, May 2015).

Leventi, C., Navicke, J., Rastrigina, O. & Sutherland, H., Nowcasting risk of poverty and income distribution in the EU in 2013, EUROMOD Working Paper no. 11/14, ISER, University of Essex.

Rodrigues, C.F., Junqueira, V. & Figueiras, R. (2013), EUROMOD Country Report – Portugal 2009-2013, ISER, University of Essex, www.iser.essex.ac.uk/files/euromod/country-reports/Year5/CR_PT_Y5_Jan2014_final_12032014.pdf

Rodrigues, C.F., (2007), Distribuição do Rendimento, Desigualdade e Pobreza – Portugal nos Anos 90, (col Económicas, Série II, nº 5), Coimbra, Almedina.