

*Flash estimates
of income inequalities and poverty indicators for 2019 (FE 2019)
Methodological note*

October 2020

Acknowledgements:

This methodological document is largely based on the research paper prepared for the IARIW 34th General Conference, Dresden, Germany, August 21-27, 2016: "Flash estimates of income distribution indicators for the European Union: methods, assessment and future prospects" by Leulescu, A.*, Leventi, C.**, Lopez-Vilaplana, C.*, Rastrigina, O.**, and Sutherland, H.** from Eurostat* and ISER, University of Essex**.

This paper has been further updated by Eurostat to take into account the most recent developments in 2018-2019.

The microsimulation results in this report represent a major output from the cooperation between Eurostat and the Institute for Social and Economic Research (ISER) in University of Essex. They rely on the use of EUROMOD, the European Union tax-benefit microsimulation model, managed, maintained and developed by ISER. Eurostat would like to thank the EUROMOD team in ISER for their fruitful cooperation, in particular, the Director of EUROMOD, Matteo Richiardi; the Executive Director, Jack Kneeshaw and Katrin Gasior, the Senior Researcher coordinating our cooperation. Many thanks to Holly Sutherland and Olga Rastrigina for their key support in the early stages of the project.

Eurostat

Aura LEULESCU

Cristina LOPEZ VILAPLANA

Bogdan MICU

Mihaela AGAFITEI

Andrea GALLELLI

With the support of SOGETI service contract "Methodological work for the development of flash estimates":

- *Adriano DI GUGLIELMO*
- *Francesca DEMONTIS*

Many thanks to Jean-Louis Mercy for his comments and support and to Emilio di Meglio, Karoly Laszlo Felkl and Didier Dupré for their support with EU-SILC data.

Table of Contents

Summary overview: methodologies used for FE 2019	4
Methodology applied for microsimulation	7
1. Introduction	7
2. Data context.....	7
3. Methodology.....	7
a. Changes in population characteristics including labour market.....	8
b. Updating non-simulated income sources	10
c. Simulating changes in tax-benefit policies.....	11
d. Alignment and discrepancies	12
4. Complementary simulations	12
5. Conclusions and further work.....	12
Current income	14
Appendix: Methodology applied for Macro-Economic Time Series Model.....	15
1. Data Context	15
2. Model Building.....	15
3. Current Models	16
4. Limitations and further work	18
References	19

Summary overview: methodologies used for FE 2019

Providing timelier social statistics – especially indicators on income poverty and inequality – is a priority for the Commission and the European Statistical System.

Indicators on poverty and income inequality are based on EU statistics on income and living conditions ([EU-SILC](#)). These indicators represent an essential tool to prepare the [European Semester](#) (the annual cycle of economic policy coordination between EU countries) and to monitor progress towards the [Europe 2020](#) poverty and social exclusion target.

In 2020, EU-SILC income indicators for 2018 (SILC 2019) will be available for all countries by autumn, which is late for the EU's policy agenda. Efforts for improving the timeliness of EU-SILC data are ongoing but the collection and processing of EU-SILC data based on both survey and administrative sources, will always have a certain time lag.

A new approach was therefore proposed, which consists in the development of [flash estimates](#). Eurostat started the developments needed and flash estimates were published for the first time in 2017 for the income year 2016. This document presents the methods used for producing FE for the income year 2019.

Flash estimates have already been developed at EU level in relation to macro-indicators such as early releases of the GDP growth¹ and inflation rate². However, in our case the focus is on the distributional changes and this implies the use of models that allow the estimation of the entire distribution and capture the complex interaction of a large number of various past and present events, such as the effects of economic and monetary policies, the implementation of social reforms or shifts in macroeconomic circumstances or demographic changes.

Three main approaches were tested in the frame of the flash estimates for income and poverty indicators project: (1) Microsimulation; (2) Current income and (3) Macroeconomic time series modelling (METS). It is important to mention that a variety of models within these approaches were tested, tailored to each country situation and the most robust methodology given current circumstances was selected. The publication as experimental statistics puts the basis for receiving feedback from users and the research community and further improving the methodologies used for the flash estimates.

The main methodology used for most countries is **Microsimulation**. It relies on EUROMOD (Sutherland, H. and Figari, F. (2013), the European Union tax-benefit microsimulation model, managed, maintained and developed by the Institute for Social and Economic Research (ISER) at the University of Essex and the Joint Research Centre (JRC) in the European Commission, in collaboration with national teams from the EU Member States. For the purposes of the flash estimates exercise standard EUROMOD policy simulation routines are enhanced with additional adjustments to the input data to take into account changes in the population structure, the evolution of employment and main indexation factors. The microsimulation approach in the frame of the flash estimates exercise is based on previous work done by ISER, University of Essex (Rastrigina, O., Leventi, C., Vujackov S. and Sutherland, H., 2016) and is being further developed by Eurostat in collaboration

¹ https://ec.europa.eu/eurostat/statistics-explained/index.php/Preliminary_GDP_flash_estimate_in_30_days_for_Europe

² https://ec.europa.eu/eurostat/statistics-explained/index.php/Inflation_in_the_euro_area#Flash_estimate_and_full_HICP_data

with them and the Task Force on “Flash estimates on income distribution”. Several Member States (FR, IT, PT, SE) are also applying this methodology for developing flash estimates.

For a second set of countries the flash estimates are based on national sources:

- For Romania, flash estimates are based on **current income information** collected in HBS³ (Household Budget Survey-RO). This differs from traditional EU-SILC income indicators as information is collected via a small set of questions that refer to the current reference period (e.g. current month).
- For the Netherlands⁴, provisional national register data were used. For Sweden, a national microsimulation model was used.
- For Sweden, a national microsimulation model was used.

Finally macro-economic time series modelling (**METS**) were tested but not used anymore from 2017 flash estimates released in September 2018. Following further analysis of the performance and the consultation of both users and Member States microsimulation was selected for all countries where national sources are not available.

Table 1 summarises the methodological approach chosen by country for the production of FE 2019. In general, empirical results for microsimulation models have proved to be better than the alternative macroeconomic models tested. In addition, microsimulation is the preferred approach for both main users in the Commission and the National Statistical Institutes given the possibilities for further detailed analyses (i.e. by socio economic groups) and for linking estimated changes with policy changes.

Table 1. Methodological approach by country

Methodological approach		Countries
Microsimulation	Labour transitions	BG, DE, DK, FI, LT, PT, MT, SK
	Calibration	BE, CZ, CY, EE, EL, ES, FR*, HR, HU, LU, LV, AT, PL, SI
	National model	SE
Current income		RO
National register based provisional data		NL

***Pending national estimates to be released by the end of November**

By September 2020, Eurostat has produced flash estimates based on microsimulation for 22 Member States. For NL, SE and RO, flash estimates are based on national sources. The table below summarises the methodologies used. Within the microsimulation approach, two different models have been used for the purpose of updating the labour and demographic characteristics: labour market

³ <http://statistici.insse.ro/shop/index.jsp?page=tempo2&lang=en&context=20>

⁴ For the Netherlands, the definition of equivalised income is almost equal to the EU-SILC definition except for the inter-household transfers which are not included. The inter-household transfers form only a small part of the total income, so the deciles in both statistics are quite comparable. In general, inter-household transfers are paid by the higher income groups, so the upper deciles may be somewhat actually lower in EU-SILC compared to the national income statistics.

transitions at the level of the individuals or calibration techniques. Table 1 summarises the main model used by country:

It is important to note that we cannot expect these models to capture perfectly changes in EU-SILC. For microsimulation, potential discrepancies⁵ between the flash estimates based on microsimulation and EU-SILC indicators, can come from various sources:

1. Inconsistencies between EU-SILC and other auxiliary sources used in the estimation process (e.g. EU-LFS in the evolution of employment)
2. Model error for labour changes or uprating factors for market incomes and public pensions
3. Policy effects are simulated under certain assumptions (i.e. only in a few cases the non-take up and tax evasion is modelled).

For current income, differences may appear due to the different reference periods and the more synthesised list of questions that often lead to an underestimation of the income.

For METS, two key drawbacks of the approach need to be pointed out. The first is the fact that direct estimation (without the distribution modelling step) opens the possibility of inconsistent indicators, i.e. ones that do not reflect the same underlying income distribution, or that jointly reflect a distribution that doesn't display the key features of the observed EU income distributions. The second is due to the focus on the nowcasting abilities of the models, which makes building the explanatory narrative less straightforward, without precluding it.

In the next section, more details on the main methodology used for the flash estimates are provided. The last section briefly describes the current income data. Even if not used anymore for the flash estimates, the METS methodology can be found in the appendix.

⁵ Based on detailed country reports for EUROMOD

Methodology applied for microsimulation

1. Introduction

The methodological approach presented in this report is based on microsimulation techniques used in combination with more up-to-date statistics from LFS and other national sources. It aims at developing a generic approach that can be applied to all EU countries in a straightforward, flexible and transparent way. By doing so, it ensures the comparability and consistency of the methodology both across countries and through time.

Microsimulation models have been widely used for assessing the distributional impact of current and future tax-benefit policy reforms, as well as the impact of the evolution of market incomes, changes in the labour market and in the demographic structure of the population⁶. Using microsimulation techniques based on representative household data enables changes in the distribution of market income to be distinguished and the effects of the tax-benefit system to be identified taking into account the complex ways in which these factors interact with each other (Peichl, 2008; Immervoll et al., 2006). Combined macro-micro modelling has also been used for analysing the impact of macroeconomic policies and shocks on poverty and income distribution.⁷

2. Data context

Microsimulation techniques rely on the EUROMOD model combined with the latest EU-SILC users' database (UDB) microdata file and/or national SILC microdata⁸ available at the time of production.⁹ In particular, for flash estimates 2019, EU-SILC 2018 microdata is used for all countries. The main auxiliary source used for labour evolution and demographics in the target year is EU Labour Force Survey (LFS)¹⁰.

3. Methodology

In order to produce flash estimates for income indicators, the microsimulation approach is used to update the structure of a micro dataset to account for changes to the main components of income variables over time. This is based on the following stages: (a) adjustment for changes to the

⁶ Some examples include Brewer et al. (2013) for the UK, Keane et al. (2013) for Ireland, Brandolini et al. (2013) for Italy, Matsaganis & Leventi (2014) for Greece and Narayan & Sánchez-Páramo (2012) for Bangladesh, Mexico, Philippines and Poland.

⁷ A detailed review is provided in Bourguignon et al. (2008) and Essama-Nssah (2005). See also Figari et al. (2015) for a discussion.

⁸ UDB EU-SILC 2018-1: BE BG CY CZ DE DK EE EL ES HR IT LV LT HU PL PT SI FI
UDB EU-SILC 2018-2: MT

In addition, for CZ EE EL LV LT LU PL SI, additional national SILC variables were also used
National SILC 2017: IT AT SK

⁹EU-SILC 2018 UDB. In the meantime EU-SILC 2019 is available for most countries but not yet the UDB and the EUROMOD input file

¹⁰ <http://ec.europa.eu/eurostat/web/lfs/data/database>

demographic structure of the population and labour market characteristics (e.g. labour transitions¹¹ or calibration techniques¹²) with EU-LFS; (b) uprating the level of market income components; and (c) changes in taxes and benefits due to policy reforms via a tax-benefit model at EU level (EUROMOD) (O'Donoghue and Loughrey, 2014).

The remaining of this section explains each of these stages in detail.

a. Changes in population characteristics including labour market

There are two main approaches to take into account changes in population characteristics: static and dynamic. The static approach is based on reweighting (or calibration). It consists in the derivation of a new vector of sample weights that brings the marginal distributions from the base year for a set of main socio-economic variables (e.g. age, labour, gender) to the level of the target year. In the dynamic process individual trajectories are modelled and individuals in the sample undergo transitions. The paper presents the two approaches as both are used depending on the country. The main auxiliary source of information used to obtain the population characteristics in the target year is the Labour Force Survey (LFS) statistics. LFS micro-level statistics for year N are usually available in April N+1. This allows the production of flash estimates for year N based on the updated structure for labour and demographics.

Modelling labour market transitions

The dynamic approach to take into account changes in population characteristics is based on modelling net employment transitions. It accounts for changes in labour market characteristics, while other population characteristics (such as demographics) are kept constant. In order to carry a more detailed analysis, Eurostat tested for the first time during the FE 2019 cycle the use of a new variable: the labour status, that split in employed, self-employed and unemployed the active population. Keeping in mind this change for FE2019, the estimation part is basically identical to the previous year cycle and done in two steps:

- Changes in employment are modelled by explicitly simulating transitions between labour market states (Figari et al., 2011; Fernandez Salgado et al., 2013; Avram et al., 2011). Two types of transitions are modelled: (i) from non-employment into employment (employee or self-employment) and (ii) from employment into short-term/long-term unemployment (or inactivity). Observations are selected for transitions based on their conditional probabilities of being employed rather than being unemployed or inactive. A logit model is used for estimating these probabilities for working age (16-64) individuals in the EUROMOD input data.

Explanatory variables include age, marital status, education level, country of birth, employment status of partner, unemployment spells of other household members, household size, number of children and their age, home ownership, region of residence and urban (or rural) location.

¹¹ Rastrigina, O., Leventi, C., Vujackov S. and Sutherland, H. (2016) Nowcasting: estimating developments in median household income and risk of poverty in 2014 and 2015, Research Note 1/2015, Social Situation Monitor, European Commission.

¹² Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

- Income is adjusted for those observations that are subject to transitions. In particular, employment and self-employment income is set to zero for individuals moving out of employment. For individuals moving into employment (self-employment), employment income is modelled and imputed from similar individuals in the sample via mixed matching methods. An individual with a similar age, educational profile, region will be used to impute the employment income, number of hours worked and the economic sector. Unemployment benefits are simulated for those moving out of employment in case they are eligible for such benefits according to the country rules.
- Detailed marginal distributions of employment figures by age, gender and economic sector based on LFS are used for deciding the final number of individual making the transition to employment.

A detailed discussion of this approach can be found also in Navicke et al. (2013) and Rastrigina et al. (2016). In comparison with their approach, further improvements were implemented such as the imputation of models via similar individuals.

Reweighting

The static approach to account for changes in population characteristics is based on reweighting. This consists in the derivation of a new vector of sample weights in order to meet control totals for the policy simulation year for a set of main socio-demographic variables (Immervoll et al., 2005). For more details on calibration techniques please see also Deville, and Särndal (1992).

The variables that are more likely to impact the income distribution over time are related to labour market information, including part time and temporary contracts. It includes also more detailed breakdowns of the number employed by age and economic sector. Other relevant controlled characteristics include number of persons by age and gender groups, household size and number of dependent children, region and degree of urbanisation. The reweighting was done at household level.

The target distributions of the relevant variables are obtained from LFS. However, the initial distributions observed in EU-SILC and LFS are not always consistent for the same year. Therefore, the calibration is based on adjusted margins, based on adding up the percentage change from LFS to the EU-SILC base year. Hence, the adjustment reflects the change to a more recent structure, while systematic source inconsistencies are ruled out.

Labour transitions versus reweighting

In comparison with the labour transitions, static ageing allows controlling for a wider range of population characteristics including retirement and demographic changes. However, there are limitations as it only adjusts the structure of the population to some marginal distributions so it may perform worse in times of rapid economic changes. For example, reweighting cannot capture if individuals entering a particular state have characteristics completely different from the characteristics of the people observed in that state in the base year. Dekkers and Liégeois (2012) compare static and dynamic ageing of microsimulation ageing under a number of headings. Static ageing is typically faster and it allows controlling for both demographic and labour changes but can be more difficult to communicate with stakeholders and is general better suited for short term projections. As for the flash estimates the lag is usually of two years it can be assumed that static ageing could be sufficient. Labour transitions could be extended to demographic characteristics and has fewer risks to distort other distributions but is more complex to put in practice. Empirically, the tests show that the results are country dependent and for the moment both methodologies are applied.

b. Updating non-simulated income sources

After adjusting the input data for changes in the population characteristics, the next step is to update non-simulated income beyond the income data reference period. This approach applies uprating coefficients to market incomes¹³ and non-simulated social benefits (or taxes). The coefficients are based on more timely data sources from the target year, which reflect indexation rules specific for each country or average changes by income component. Two approaches are tested in the paper.

EUROMOD uprating factors

EUROMOD contains uprating factors based on available administrative or survey statistics. Country-specific uprating factors are derived for each income source, reflecting either statutory rules (such as indexation rules) or the change in the average amount per recipient between the income data reference period and the target year. The latter is preferred for the nowcasting exercise, especially for pensions. The evolution of average pensions can capture important changes in the population of pensioners (e.g. inflow of newly retired pensioners with higher average pensions). However, currently the information available for uprating wages in EUROMOD is often very aggregated so in most countries the uprating factors applied at individual level is just the average; in 7 countries is differentiated for the private/public sector (Cyprus, Germany, Greece, Spain, Italy, Latvia and Portugal) and for Germany, Greece and Poland we have more detailed uprating factors by sector.

Labour Cost Index for uprating wages and salaries

In order to capture differential growth rates in employment income, Eurostat tested for the first time during the FE 2017 cycle the use of uprating factors disaggregated by economic activity and/or by economic sector if such information is available. Labour Cost Index (LCI) measures short-term trends in "average hourly labour costs", defined as (total) labour costs divided by the corresponding number of hours worked.. Data is available by detailed economic sector (NACE groups, B-S) and it is used for the main economic sectors (industry (B-E), construction (F), services (G-J, K-N) and mainly non-business economy (O, P-S)). Tailored breakdowns by country were applied if the economic sector has a large share and there are very different growth rates inside a specific group. This allows introducing a more differentiated uprating of wages and salaries across the income distribution. The main limitation is that the LCI doesn't discount the compositional effect derived from a change in the composition of employment within an economic sector. This means that, for instance, the LCI may increase due to the redundancies of low paid workers within one sector. The rest of market incomes are uprated using the EUROMOD uprating factors.

Statutory minimum wages and self-employment income

A further focus on wages has been introduced in flash estimates 2019¹⁴, consisting in the uprating of employees' wages below a statutory national minimum wage. Consistently with the literature on minimum wages (Brandolini et al 2010; Fernández-Macías and Vacas-Soriano 2016), the proportion of employees with salaries below the national statutory minimum wage, is calculated in terms of monthly full-time equivalent gross wage. The individual annual gross earnings are divided by the

¹³ Market incomes are wages and salaries, self-employment income, property income, income from capital, etc.

¹⁴ This adjustment was applied to ES

number of months in full-time work in the same reference period, plus the number of months in part-time jobs. Part-time work is further adjusted by the sex-specific ratio between the median of part-time and full-time hours of work.

The estimation of the income from self-employment has been conducted following a country-specific approach. For countries that have shown an evolution of the indicator particularly consistent with other sources, a proxy variable has been chosen as updating factor. Research conducted at Eurostat (Eurostat 2018) has shown that ‘Gross mixed income’ from the National Accounts (ESA 2010), is the item closest to the self-employment income in the EU-SILC, with reference to definition and empirical evolution. The year on year change of this proxy variable has been selected to update self-employment income.

For countries that have shown a less consistent evolution with other sources, a model based estimation has been conducted. In particular, regression methods have been applied to compute the year on year change in the self-employment income (dependent variable) from a set of available covariates, such as GDP and wages indicators from the National Accounts, the Labour Cost Index and other labour indicators. A stepwise linear regression has been used to select the best predictors.

c. Simulating changes in tax-benefit policies

After updating market income and other non-simulated income sources, we simulate tax-benefit policies for each year from the base year up to the target year.

EUROMOD is used to simulate changes in the income distribution within the period of analysis. Income elements simulated by the model include universal and targeted cash benefits, social insurance contributions and personal direct taxes. Income elements that cannot be simulated mostly concern benefits for which entitlement is based on previous contribution history (e.g. pensions) or unobserved characteristics (e.g. disability benefits). These are read from the data and updated according to statutory rules (such as indexation rules) or changes in their average levels over time (e.g. for pensions for: Bulgaria, Germany, Estonia, Spain, France, Croatia, Italy, Latvia, Malta, Poland, Portugal). Both contributory and non-contributory unemployment benefits are simulated in the model; but e.g. severance payments are not. Detailed information on EUROMOD and its applications can be found in Sutherland & Figari (2013).

All simulations are carried out on the basis of the tax-benefit rules in place on the 30th June of the given policy year. The exceptions to this rule are Estonia (in 2013 and 2017), Greece (in 2014-2019), Lithuania (in 2017), Portugal (in 2012), Slovakia (in 2019), Spain (in 2018), and the United Kingdom (in 2016-2017), where policy changes after the 30th of June were taken into account to better match the annual income observed in the EU-SILC data. In order to enhance the credibility of estimates, an effort has been made to address issues such as tax evasion (e.g. in Bulgaria, Greece, Italy and Lithuania) and benefit non-take-up (e.g. in Belgium, Estonia, France, Greece, Spain, Croatia, Ireland, Latvia, Finland, Slovenia and the United Kingdom)¹⁵. However, such adjustments are not possible to implement in all countries due to data limitations.¹⁶

¹⁵ <https://www.euromod.ac.uk/sites/default/files/working-papers/em5-18.pdf>

¹⁶ Detailed information on the scope of simulations, updating factors, non-take-up and tax evasion adjustments is provided in the EUROMOD Country Reports (see: <https://www.iser.essex.ac.uk/euromod/resources-for-euromod-users/country-reports>).

d. Alignment and discrepancies

The last methodological step involves an attempt to account for differences between EUROMOD and EU-SILC estimates of household income in the data reference year. The main reasons for these discrepancies are related to the precision of simulations when information in the EU-SILC data is limited, issues of benefit non-take-up and tax evasion, under-reporting of income components, and small differences in income concepts and definitions. These discrepancies are expected to improve with new SILC data collected on disaggregated benefits, further checks of the modelling for non-take up and net to gross adjustments.

In order to account for these differences, an alignment factor is calculated for each household. The factor is equal to the absolute or the percentual difference between the value of the equivalised household disposable income in EU-SILC 2018 and EUROMOD. The alignment can also be done just for specific income components or for total income, depending on the country and the specific type of inconsistencies that can emerge. For consistency reasons, the same household specific factor is applied to all later policy years. This is based on the assumption that the discrepancy between EUROMOD and EU-SILC estimates based on the same input file remains stable over time. Further work needs to investigate the plausibility of these assumptions and the effect on the final results.

4. Complementary simulations

An additional feature of the microsimulation methodology is that it allows simulating complementary counterfactual scenarios where only some parts of the model are updated (e.g., what would be the impact of policy changes in 2019 if labour market adjustments had remained constant?). Changes in 2018-2019 were calculated taking into account *only* policy changes in 2018-2019 that impact social benefits and taxes (labour market characteristics and wages and self-employment income are as in 2018).

These complementary estimates are used to assess the plausibility of the estimated change and are included in the main report in order to provide more information and clarity on the assumption of the model and the transparency of the process.

5. Conclusions and further work

Microsimulation proved to be the most accurate methodology for producing the flash estimates on income distribution and poverty. It takes into account both market developments but also changes in policies so it can estimate different changes across the income distribution. It also showed in these first years of publication that it can capture important changes in social benefits and taxes and their subsequent effects on poverty indicators.

It is important to raise also some limitations related to the methodology described:

- Inconsistencies between EU-SILC and external sources or lack of detailed data concerning the evolution of specific income components such as income from property or self-employment;

- There are still limitations related to the use of either reweighting or labour transitions in the context of nowcasting. The first method only adjusts the structure of the population to marginal distributions it may perform worse in times of rapid economic changes. However, further improvements were done in using calibration by important sub-groups (e.g. number of employees by economic sectors; age groups). As concerns the dynamic ageing, for the moment the transitions are done only for the working age population and related to the employment status. For more details see Leulescu et al. (2016); Rastrigina et al. (2016).
- There are sometimes large differences between simulated and observed income in EU-SILC (Rastrigina et al., 2016) which need further investigation in order to improve the consistency of the flash estimates with EU-SILC values for specific income components.

Further developments could be envisaged to address these limitations:

- use of more recent EU-SILC files for microsimulation so that to minimize the impact of revisions and breaks in series but as well to improve the model for reweighting or labour transitions;
- further assess and improve the discrepancies between EU-SILC and EUROMOD simulated benefits and taxes;

Finally, while there are still limitations in the current methodology and its ability to replicate changes in SILC, it can provide an early indication concerning the direction and magnitude of the change, including the effects due to change in employment, in uprating factors and impact of policies.

Current income

Eurostat collected current income data from 11 countries during last year and tested its use for producing flash estimates retrospectively (2012-2018). These flash estimates entered the quality assessment framework together with the other approaches (microsimulation, macro-economic time series modelling). So far the current income was used for the cycle FE 2019 only for RO.

The degree of heterogeneity among countries is rather high in terms of the income components considered, the question design (e.g. allowing or not the use of income bands), the level of analysis (individual vs. household income), or the type of income measured (gross or net), among others aspects.

Data was further processed in Eurostat with the imputation of exact values when income bands were collected, consistency checks and several models tested. The results showed that:

- Levels are generally underestimated in all countries;
- Yearly changes for deciles are often captured in terms of direction and trends but not the exact magnitude, for AROP the results are mixed;
- Current income as such was not performing better than microsimulation in most of the countries analysed

For RO, FE 2019 are based on HBS data¹⁷. The Household Budget Survey (HBS) is organized as a continuous quarterly survey over a period of three consecutive months, based on a sample of 9504 permanent dwellings, divided into monthly independent sub-samples of 3168 permanent dwellings (per year the sample cover 38016 households). Response rate is around 80% -85%. The survey covered people with permanent residence in Romania, members of households in all counties and in Bucharest. Main variables collected are expenditures, incomes, endowment with durable goods and other demographic variables. Data are collected by face-to-face interview and self-registration for the diary. The support of data collection is the household questionnaires (CG) and the household diary (JG). The reference period for the data registration in the survey questionnaire and household diary is the calendar month (from the first to the last day of the month).

¹⁷ <http://statistici.insse.ro/shop/index.jsp?page=tempo2&lang=en&context=20>
<http://colectaredate.insse.ro/metadata/viewStatisticalResearch.htm?locale=en&researchId=4356>

Appendix: Methodology applied for Macro-Economic Time Series Model

1. Data Context

- The reference (official) values of the indicators are **sample-based**; this adds a major source of uncertainty, and complicates the calculation of the prediction intervals.
- **Short data series** (yearly values, not reliable / comparable before 2003).
- Inequality indicators (AROP, QSR) are **very stable** (very few significant YoY changes).
- For some countries, the economic crisis seems to have triggered a **change in the relationship** between indicators and the macroeconomic variables.

2. Model Building

Neither cross-sectional nor pooled models were considered (despite the possible advantage of multiplication of data points), because we do not investigate differences between countries or the shared sources of variation in the indicators.

The indicators are nowcasted directly, instead of nowcasting the entire distribution and subsequently deriving the indicators.

Another element in model selection is that the users have expressed a strong preference for simple models that can be explained easily and intuitively¹⁸ in terms of sources for the expected change. Given the need to link the estimated change to specific economic and social factors, univariate models like ARIMA or exponential smoothing were not included (despite their robustness and stability¹⁹).

Predictors

The main criteria for choosing predictors are "quality of the association between the predictors and the response, data quality, and availability of the predictors at the time of prediction, known as ex-ante availability."²⁰ Given this objective, an exhaustive list of predictors (shown in the next paragraph) has been created. On the other hand, we need to take into account that a high number of predictors could most likely bring to some statistical issues such as parsimony, multicollinearity and/or overfitting of the models, that is the reason why we decided to use some automated procedures in order to understand which of the pool of candidate variables should be included in our models.

The goal of variable selection becomes one of parsimony: achieve a balance between simplicity (as few predictors as possible) and fit (as many predictors as needed).

Finally, to reach the above mentioned goal, three different automated variable selection methods have been used: a) [Stepwise Regression](#) b) [General Linear Models \(GLM\) Selection](#) c) [Lasso Selection](#)

¹⁸ See also [Handbook on Rapid Estimates](#) (p. 484): "The model should be interpretable: a model should be explicable, and understandable from the economic point of view [...] the link between the target variable and the explanatory variables should have a clear economic interpretation and should ideally be persistent with time."

¹⁹ [Handbook on Rapid Estimates](#), p. 485.

²⁰ [To Explain or to Predict?](#), p. 298

[\(LASSO\)](#); these methods were also accompanied by country specific checks for a tailored selection of predictors based on theoretical grounds.

3. Current Models

Variables

Variables enter the model as year-on-year (YoY) change, instead of yearly values; these transformations might generate stationary series of the dependent variables.

Dependent variables: YoY change in the indicators; percentage change for deciles.

Predictors: country-specific, selected via the following automated procedures

- **STEPWISE:** focused on maximizing R^2 (SAS procedure & settings: [proc REG](#) + SELECTION=STEPWISE).
- **LASSO:** focused on minimizing prediction error (SAS procedure & settings: [proc GLMSELECT](#) + SELECTION= [LASSO](#) choose=press stop=press [LSCOEFFS](#). [Cross validation](#) has been used for estimating prediction error).
- **GLM:** focused on minimizing prediction error (SAS procedure & settings: [proc GLMSELECT](#) + SELECTION= [STEPWISE](#) select=press choose=press stop=press. [Cross validation](#) has been used for estimating prediction error).

As previously mentioned, the selected predictors should have the potential of being used for building a compelling explanatory narrative. Consequently, predictors were chosen on the basis of their theoretical relation to income as precursors and components (wages, social benefits, employment status), or as proxies (consumption, savings).

In this context proxies are variables that do not have a causal relationship to income, but a correlational one, and should be seen as stand-ins for impacting factors not included in the model; therefore their contribution to the dependent's value (expressed by the impact coefficient) should be read as "contribution of other factors not included in the model that are observable as the respective marker variable".

Inflation (HICP) is also included because it is an [indicator of macroeconomic stability](#), and has redistributive [effects](#). Inflation could also indicate to what extent the observed change in decile thresholds is real or just nominal.

<i>Potential predictors</i>	
GDP	Gross domestic product at market prices, chain-linked volumes (NA : B1GQ)
GDI	Gross Disposable Income (NA : B6G)
HICP	Yearly Harmonised Indices of Consumer Prices [2015 = 100]
Activity Rate	Active population
Part-time employment	Part-time employment contracts
Temporary employment	Temporary employment contracts
Unemployment	Unemployed persons – annual average

Long-term unemployment	Long-term unemployment (unemployed persons since 12 months or more)
Employment	Total employment (resident population concept - LFS)
Compensation	Compensation of employees (NA : D1)
HH Savings	Gross household saving rate (NA : SRG_S14_S15)
HH Consumption	Final consumption expenditure of households (NA : P31_S14)
Wages and Salaries	Wages and salaries (NA : D11)
Social Benefits	Social Benefits other than social transfers in kind (NA : D62)
Taxes	Current taxes on income, wealth, etc. (NA : D5)
GNI*	Modified gross national income at current market prices

Methods used

- Multivariate Linear Regression (SAS: [REG](#))
- SUR: Seemingly Unrelated Regressions (SAS: [SYSLIN](#))

The Seemingly Unrelated Regressions (SUR) model is a special case of the generalized linear regression model where a linear system of equations is taken into account in order to improve the efficiency of the estimation. The relations within the system are given by the correlation of the error terms across the equations.

The SUR model can be further generalized into the simultaneous equations model (SEM), where the right-hand side predictors are allowed to be the endogenous variables, while SUR contain only exogenous predictors.

The optimal estimator for SUR model is the generalized least squares GLS estimator in two steps. This method of estimation is also called feasible generalized least squares ([FGLS](#)).

- regARIMA: regression models with ARIMA errors (SAS: [AUTOREG](#))

If working with time series data, to use ordinary regression analysis might be not the best option. One key assumption of ordinary regression is that the errors are independent of each other, while with time series data, the ordinary regression residuals are usually correlated over time.

Violation of the independent errors assumption has three important consequences for ordinary regression. First, statistical tests of the significance of the parameters and the confidence limits for the predicted values are not correct. Second, the estimates of the regression coefficients are not as efficient as they would be if the autocorrelation was taken into account. Third, since the ordinary regression residuals are not independent, they contain information that can be used to improve the prediction of future values.

The AUTOREG procedure solves this problem by augmenting the regression model with an autoregressive model for the random error, thereby accounting for the autocorrelation of the errors.

Then regARIMA creates a regression model with ARIMA (p, D, q) time series errors to maintain the sensitivity interpretation of regression coefficients, where p is the autoregressive degree, D is the differencing degree, and q is the moving average degree.

The RegARIMA model assumes that the regression variables influence the time series errors concurrently.

Three different estimation methods are available:

1) The *Yule-Walker* method can be considered as generalized least squares using the OLS residuals to estimate the covariances across observations, and Judge et al. (1985) use the term estimated generalized least squares (EGLS) for this method. For a first-order AR process, which corresponds to our case, the Yule-Walker estimates are often termed Prais-Winsten estimates. There are variations to these methods that use different estimators of the autocorrelations or the autoregressive parameters.

2) The *maximum likelihood* method.

3) The *unconditional least squares* (ULS) method, which minimizes the error sum of squares for all observations, is referred to as the nonlinear least squares (NLS) method by Spitzer (1979).

4. Limitations and further work

- The main limitation is due to the very few *data points in our time series*. Using least squares estimation, or some other non-regularized estimation method, it is possible to estimate a model only if you have more observations than parameters. However, there is no guarantee that a fitted model will be any good for forecasting, especially when the data are noisy. The shortness of our time series, which therefore might impact the model stability, allows us to only take into consideration a certain category of forecast techniques, excluding more complex analysis. On the other hand, very simple time series analysis such as the exponential smoothing have been carried out and used like benchmarks models in order to validate our flash estimates.
- Lack of distributional information that has a direct impact on indicators like AROP.
- Direct estimation (without the distribution modelling step) opens the possibility of inconsistent indicators, i.e., ones that do not reflect the same underlying income distribution, or that jointly reflect a distribution that doesn't display the key features of the observed EU income distributions.
- It is more difficult for macro models to explain and build a plausibility argumentation related to the evolution in the income distribution.

References

- Avram, S., Sutherland, H., Tasseva, I. and Tumino, A. (2011) Income protection and poverty risk for the unemployed in Europe, Research Note 1/2011 of the European Observatory on the Social Situation and Demography, European Commission.
- Bourguignon, F., Bussolo, M. and Da Silva, L.P. (2008) The impact of macro-economic policies on poverty and income distribution Macro-Micro Evaluation Techniques and Tools, The World Bank and Palgrave-Macmillan, New York.
- Brandolini, A., D'Amuri, F. and Faiella, I. (2013) "Country case study – Italy", Chapter 5 in Jenkins et al, The Great Recession and the Distribution of Household Income, Oxford: Oxford University Press.
- Brandolini, A., Rosolia, A., Torrini, R. (2010) "The distribution of employees' labour earnings in the European union: Data, concepts and first results", in: Atkinson, AB., Marlier, E. (eds) Income and Living Conditions in Europe. Luxembourg: Publications Office of the European Union: 265–287.
- Brewer, M., Browne, J., Hood, A., Joyce, R., Sibieta, L. (2013) The Short- and Medium-Term Impacts of the Recession on the UK Income Distribution, Fiscal Studies, 34(2): 179–201.
- Cohen, R. (2006) Introducing the GLMSELECT Procedure for Model Selection. Proceedings of SAS Global Forum 2006 Conference. Cary, NC: SAS Institute Inc.
- Dekkers, Gijs, 2012, The simulation properties of microsimulation models with static and dynamic ageing – a guide into choosing one type of model over the other. Mimeo
- Dekkers, Gijs and Philippe Liégeois (2012). The (dis)advantages of dynamic and static microsimulation. Mimeo Sept 14th 2012
- Deville, J.-C. (1999). Estimation de variance pour des statistiques et des estimateurs complexes : linéarisation et techniques des résidus. Techniques d'enquête, 25, 219-230.
- Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87, 376-382.
- Essama-Nssah, B. (2005) The Poverty and Distributional Impact of Macroeconomic Shocks and Policies: A Review of Modeling Approaches, World Bank Policy Research Working Paper 3682, Washington, DC.
- EUROMOD (2020) Effects of tax-benefit policy changes across the income distributions of the EU-28 countries: 2018-2019, EUROMOD Working Paper Series EM15/20, Institute for Social and Economic Research, University of Essex
- European system of accounts :ESA 2010. (2013). Publications Office. <https://doi.org/10.2785/16644>
- Fernandez Salgado M., Figari, F., Sutherland, H. and Tumino, A. (2014) Welfare compensation for unemployment in the Great Recession, The Review of Income and Wealth, 60(S1): 177–204.
- Fernandez-Macias, E., Vacas-Soriano, C. (2016) A coordinated European Union minimum wage policy?. European Journal of Industrial Relations, 22(2): 97-113.
- Figari, F., Iacovou, M., Skew, A. and Sutherland, H. (2012) Approximations to the truth: comparing survey and microsimulation approaches to measuring income for social indicators, Social Indicators Research, 105(3): 387-407.
- Figari, F., A. Paulus and H. Sutherland (2015) "Microsimulation and Policy Analysis" in A.B. Atkinson and F. Bourguignon (Eds.) Handbook of Income Distribution, Vol 2B. Amsterdam: Elsevier, pp. 2141-2221. ISBN 978-0-444-59430-3
- Figari, F., Salvatori, A. and Sutherland, H. (2011) Economic downturn and stress testing European welfare systems, Research in Labor Economics, 32: 257-286.
- Funda Gunes, Model Selection for Linear Models with SAS/STAT Software.
- Gasior, K. and Rastrigina, O. (2017) Nowcasting: timely indicators for monitoring risk of poverty in 2014-2016, Research Note 1/2016, Social Situation Monitor, European Commission
- Harrell, F. (2001) Regression Modelling Strategies. New York: Springer-Verlag.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009) The Elements of Statistical Learning : Data Mining, Inference, and Prediction. New York: Springer-Verlag.

- Immervoll, H., Lindström, K., Mustonen, E., Riihelä, M., & Viitamäki, H. (2005c). Static data “ageing” techniques. Accounting for population changes in tax-benefit microsimulation. EUROMOD Working Paper, EM7/05.
- Immervoll, H., Levy, H., Lietz, C., Mantovani, D. and Sutherland, H. (2006) The sensitivity of poverty rates in the European Union to macro-level changes, *Cambridge Journal of Economics*, 30: 181-199.
- Jara, X.H. and Leventi, C. (2014) Baseline results from the EU27 EUROMOD (2009-2013), EUROMOD Working Paper 18/14, Colchester: Institute for Social and Economic Research, University of Essex.
- Keane C., Callan, T., Savage, M., Walsh, J.R. and Timoney, K. (2013) Identifying Policy Impacts in the Crisis: Microsimulation Evidence on Tax and Welfare, *Journal of the Statistical and Social Inquiry Society of Ireland*, XLII: 1-14.
- Leulescu, A., Leventi, C., Lopez-Vilaplana, C., Rastrigina, O., and Sutherland, H. (2016) Flash estimates of income distribution indicators for the European Union: methods, assessment and future prospects). 34th General Conference, Dresden, Germany
- Makovec, M. and Tammik, M. (2017) Baseline results from the EU28 EUROMOD (2011-2016), EUROMOD Working Paper 6/17, Colchester: Institute for Social and Economic Research, University of Essex.
- Matsaganis, M. and Leventi C. (2014) Poverty and inequality during the Great Recession in Greece, *Political Studies Review*, 12: 209 -223.
- Moomaw, R.L and Olson, K. W. (2006) *Economics and Contemporary Issues*. South-Western College Pub
- Narayan, A. and Sánchez-Páramo C. (2012) Knowing, when you don't know: using microsimulation models to assess the poverty and distributional impacts of macroeconomic shocks, The World Bank, Washington DC.
- Navicke, J., Rastrigina, O. and Sutherland, H. (2013) Nowcasting Indicators of Poverty Risk in the European Union: A Microsimulation Approach, *Social Indicators Research*, 119(1): 101-119.
- O'Donoghue, C., Loughrey, J. (2014) Nowcasting in Microsimulation Models: A Methodological Survey, *Journal of Artificial Societies and Social Simulation* 17 (4) 12.
- Peichl, A. (2008) The Benefits of Linking CGE and Microsimulation Models: Evidence from a Flat Tax Analysis, IZA Discussion Paper No. 3715.
- Rastrigina, O., Leventi, C., Vujackov S. and Sutherland, H. (2016) Nowcasting: estimating developments in median household income and risk of poverty in 2014 and 2015, Research Note 1/2015, Social Situation Monitor, European Commission.
- Rastrigina, O., Leventi, C. and Sutherland, H. (2015) Nowcasting: estimating developments in the risk of poverty and income distribution in 2013 and 2014, Research Note 1/2014, Social Situation Monitor, European Commission.
- Reconciliation of European Union statistics on income and living conditions (EU-SILC) data with national accounts: 2019 edition. (2019). Publications Office. <https://doi.org/10.2785/489634>
- Sanchez-Robles, B. (1998) Macroeconomic stability and economic growth: the case of Spain. *Applied Economics Letters*, 5:9, 587-591
- Shmueli, G. (2010) To Explain or to Predict? *Statistical Science* 25, no. 3, 289-310.
- Studenmund, A. H. (2017) *Using Econometrics: A Practical Guide*. Prentice-Hall
- Sutherland, H. and Figari, F. (2013) EUROMOD: the European Union tax-benefit microsimulation model, *International Journal of Microsimulation*, 6(1): 4-26.
- Tammik, M. (2018) Baseline results from the EU28 EUROMOD: 2014-2017, EUROMOD Working Paper 5/18, Institute for Social and Economic Research, University of Essex.
- Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267-288.
- United Nations Statistics Division (October 2016) *Handbook on Rapid Estimates*.
- Zou, H. 2006. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101, 1418-1429
- Yamano, T. (2009) *Lecture Notes on Advanced Econometrics: GLS, WLS and FGLS*.