A quality assessment of flash estimates for the income distribution

Aura Leulescu[1], Daniel Koch[2], Bogdan Micu[3], Cristina Lopez-Vilaplana[4]

[1] *Eurostat , Luxembourg, Luxembourg; aura.leulescu@ec.europa.eu*
[2] *Sogeti, Luxembourg, Luxembourg; Daniel.KOCH@ext.ec.europa.eu*
[3] *Eurostat , Luxembourg, Luxembourg; Bogdan-Alexandru.MICU@ec.europa.eu*
[4] *Eurostat , Luxembourg, Luxembourg; Cristina.LOPEZ-VILAPLANA@ec.europa.eu*

Abstract

EU-SILC indicators on poverty and income inequalities are an important part of the toolkit for the European Semester which is the yearly cycle of economic policy coordination among EU member states. However, income year N is only available in the autumn N+2 which comes too late for the policy agenda. In order to be able to provide earlier information, one approach pursued in Eurostat is the development of flash estimates of the income distribution.

Several strands are investigated including: re-weighting procedures and modelling techniques in order to account for demographic and labour market changes captured in more timely sources such as LFS, and the use of the microsimulation models that replicate at individual/household level the effects of the different taxation regimes.

The flash estimates are therefore model-based and rely on micro-data. The quality framework is based on an (1) ex-ante quality assurance via consistency analysis of auxiliary data sources and intermediary quality checks and (2) a quality assessment via back-testing (measuring the model's ability to forecast the past) supplemented by the development of a quality measure that quantifies the uncertainty of the model based estimates.

Keywords: flash estimates, nowcasting, income distribution, poverty

1. General Framework

1.1. Objectives

Policy makers in the European Union (EU) have an increasing demand for a social scoreboard to better monitor the changes in social conditions, especially during periods of economic crisis. EU-SILC indicators on poverty and income inequalities are a key part of the toolkit for the European Semester, the yearly cycle of economic policy coordination among EU member states. However, income data for year N is only available in the autumn N+2 which comes too late for the policy agenda.

The strategy for providing more timely data on income as far as possible is based on two pillars:

- ✓ Flash estimates about income N if possible in June N+1, that would be available to prepare and to start the European Semester and provide data to MIP in autumn N+1
- ✓ Final EU-SILC data (or at least data more robust than flash estimates) on income N during the European Semester (end N+1 / early N+2)

The flash estimates on income distribution and poverty should[1]:

- ✓ refer to a past yearly reference period (year N);
- ✓ refer to a set of distributional indicators for equivalised disposable income (such as median, quintiles, at-risk-of-poverty rate, Gini coefficient, S80/S20). Therefore, in this case, we aim to estimate the whole distribution of equivalised household disposable income (at household /individual level) that would enable us to compute inequality indicators and estimate the change at different points of the distribution;

[1] http://unstats.un.org/unsd/nationalaccount/workshops/2009/ottawa/AC188-S31a.PDF

- ✓ are based on an information set that includes the latest income data available from EU-SILC (income N-1 or N-2 even), plus more timely auxiliary information from the reference period (year N) such as LFS, National accounts etc.;
- ✓ are based on a set of statistical techniques, such as calibration, modelling, extrapolation that are not traditionally used in the calculation of social statistics indicators.
- ✓ are assessed based on a specific quality framework.

The development in Eurostat of flash estimates for income distribution changes is based on two main methodological approaches. Method 1 is essentially based on microsimulation and aims to estimate household and individual level distributive effects. The second is a meso-level approach as it aims to estimate the reconstitute the distribution via a set of distributional elements (quantiles and parameters).

This paper provides a quality framework for assessing the results with a view in a first place to choosing the best method and auxiliary information for the production of flash estimates. Secondly the quality assessment will need to validate the results in order to decide if the estimated indicators are reliable enough to be used in the European Semester.

As part of the assessment procedure, we look at: (1) consistency of trends in auxiliary sources (2) distribution tests for the evolution of main income components, and (3) performance of the model for past data and (4) possible quality measures that take into account different sources of uncertainty for our model based estimates. The quality assessment is conducted for all indicators (median income, income deciles / quintiles, AROP, Gini) but it includes also tests for the income distribution.

1.2. Methodological approaches

Flash estimates have already been developed at EU level in relation to macroeconomic indicators such as early releases of the GDP growth and inflation rate[2]. However, in the case of

[2]http://ec.europa.eu/eurostat/statistics-explained/index.php/Inflation_%E2%80%93_methodology_of_the_euro_area_flash_estimate

income and poverty the focus is on the estimation of the whole income distribution and specific indicators. There are two main approaches that are currently being tested in the frame of the flash estimates for income and poverty indicators project:

- ✓ Evolution of income components-microsimulation
- ✓ Parametric quantile approach

The first approach is in line with current practices in different Member States and it aims to micro-simulate income changes at individual/household level within EU-SILC. This is done via the decomposition of the household disposable income in some main components which are 'updated' separately in terms of both number of recipients (presence of income) and levels of income (e.g. changes to average wage). The literature on microsimulation identifies a number of stages for modelling changes to income distribution over time: adjustment for changes to the demographic structure of the population; adjustment for changes to the presence of income sources such as employment rates; uprating the level of income components (indexation) and tax benefit changes due to policy reforms/update (O'Donoghue and Loughrey, 2014). An essential step of this approach consists in the microsimulation of social benefits and taxes in order to calculate policy effects via EUROMOD. This is a model already put in place at EU level for the microsimulation of effects of different policies. However, its application in the frame of flash estimates requires an 'update' of the socio-deomographic structure and market level incomes to the target year. In terms of statistical methods there are two main 'adjustment' techniques: static and dynamic ageing. Static ageing is based on reweighting and 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-demographic variables (age, labour, gender) to the level of the target year. For dynamic ageing individual trajectories are modelled and individuals in the sample undergo transitions. The second step consists in the indexation of the monetary values of different income components to fit developments from auxiliary sources (such as National Accounts).

The second approach plans to use the parametric estimation of the income distribution, where the main target indicators are derived from the GB2 (generalized beta distribution of the second kind) curve fitted to the empirical income distribution. The starting point is the work done in the context of the AMELI[3] project (Advanced Methodology for European Laeken Indicators). The parametric approach involves nowcasting the entire income distribution using the relationship between its parameters and macroeconomic variables. It relies to a large extent on data available in Eurostat.

2. Evolution of income components- microsimulation approach

In order to produce flash estimates for income indicators, the microsimulation approach 'updates' the structure of a microdataset to account for changes to the main components of the income over time.

In order to formalize this approach we would say that, in general terms, disposable income can be defined as follows:

$$DINC_{H,t} = \sum_{i=1 \dots M} \{ Y_{i,t}(\ ) \times I_{i,t}(\ ) \} - T(\ ) + B(\ )$$

Where

$Y_{i,t}(\ )$ - Level of Market Income

$I_{i,t}(\ )$ - Presence of Market Income

---

[3] https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP2-D2.1-20110409.pdf

$T(\ ), B(\ )$ - Tax, Benefit system

For income components 1…M

The basic assumption behind this framework is that the updated auxiliary information would then lead to new income indicators via the 'uprating' of different income components both in terms of level and number of recipients. These 'uprated' components will be re-assembled in the equivalised household disposable income (at microlevel).

The general framework is composed of three main stages: 1) Adjustments for changes to the demographic structure of the population and for changes to the presence of income sources such as changes in the employment rate via re-weighting with LFS data; 2) updating the evolution of different income components via uprating factors and 3) the micro-simulation of taxes and benefits in EUROMOD. The results of this paper are based on the latest EUROMOD input files available (EU-SILC 2012).

3. Parametric approach

This section presents an overview of a parametric approach to the production of Flash Estimates of the income distribution. The indicators are produced for the years before the current year for which directly observed income data is not available.

The approach is "parametric" because we use a powerful tool for describing the income distribution: the generalized beta distribution of the second kind or GB2 distribution (MacDonald and Xu, 1995).

Using the GB2 we can:

- describe the entire income distribution using only 4 parameters
- reconstitute the income distribution starting from these parameters, or from distribution elements (e.g. quantiles)
- calculate directly the monetary Laeken indicators (AROP, ARPT, Gini, QSR) from the distribution parameters

Three different options were explored for the production of flash estimates

- ✓ Extrapolation of the trends in the time series of each of the GB2 parameters
- ✓ Econometric modelling of each of the GB2 parameters
- ✓ Econometric modelling of the income centiles. This approach can be described as a Parametric, Quantile-based model using Macro variables (PQM) and it was chosen for further development.

The PQM approach involves the following steps:

I. Determine the quantiles (income thresholds separating, for example, the lowest 1% from the rest) of the observed income distribution

II. Model the time series for each quantile using macro variables as covariates

III. Produce estimates of the quantiles for the target year(s)

IV. Calculate the indicators from either (a) the quantiles estimated by the model or (b) the GB2 curve reconstituted from these quantiles

4. Quality framework

4.1.1. Quality assurance: Consistency analysis of auxiliary sources

A backward assessment of consistency was performed for the auxiliary information to be used in the estimation process (LFS, Labour Cost Index and National Accounts). This was done for 5 years (SILC 2009-2014) and for all countries. The conclusions of this stage were:

- ⇒ For the labour update, LFS can be used with the mention that some adjustments need to be made for specific countries in order to level out in the calibration the 'source effects'.
- ⇒ For uprating income components, National Accounts or Labour Cost Index (LCI) can provide consistent information for wages and salaries. For property and self-employment income, the consistency is rather low.
- ⇒ EUROMOD simulates the information on types and levels of social benefits and taxes. As mentioned before there are several transformations to the SILC input dataset and

thus further inconsistencies can arise. A correction factor is usually applied for the base year to account for differences between SILC and EUROMOD.

4.1.2. Quality assurance: Intermediary quality checks

As the microsimulation relies on 3 main stages for this approach, additional intermediary quality checks were implemented.

- ***Assessing the update of labour information***

The first step in assessing quality is to assess how the new labour distribution 'updated' based on LFS compares to SILC trends for previous years. We have chosen to quantify similarity of two distributions based on the Hellinger distance (HD). A value of 0 indicates a perfect similarity between two probabilistic distributions, whereas a value of 1 indicates a total discrepancy[4].Results show that in general that distributions after calibration on household level are rather close (<2%), while for labour status at individual level we still have discrepancies (the average for 15 countries is around 5%.).

- ***Assessing the update of main income components***

The estimated income distribution of the either uprated or micro-simulated *income components* is compared with the corresponding SILC income distribution. The analysis is performed for point estimates (e.g. the median and the median of quintiles) and for parts/the whole income distribution of the income components (e.g. tests for assessing if the difference between two distributions is statistically significant, as the Kolmogorov-Smirnov test on the maximum vertical distance between the two cumulative distribution functions). First tests showed that the uprating for wages and salaries gives in general better results, while the 'uprated' self-employment income can be rather different from the target one.

[4] In general a Hellinger distance smaller than 5% is considered acceptable.

4.2. Quality assessment of flash estimates for income distribution

As part of the assessment procedure, we look at the: (1) performance of the model for historical data; (2) quality measures for the different flash estimates.

4.2.1. Performance of the different models against past data

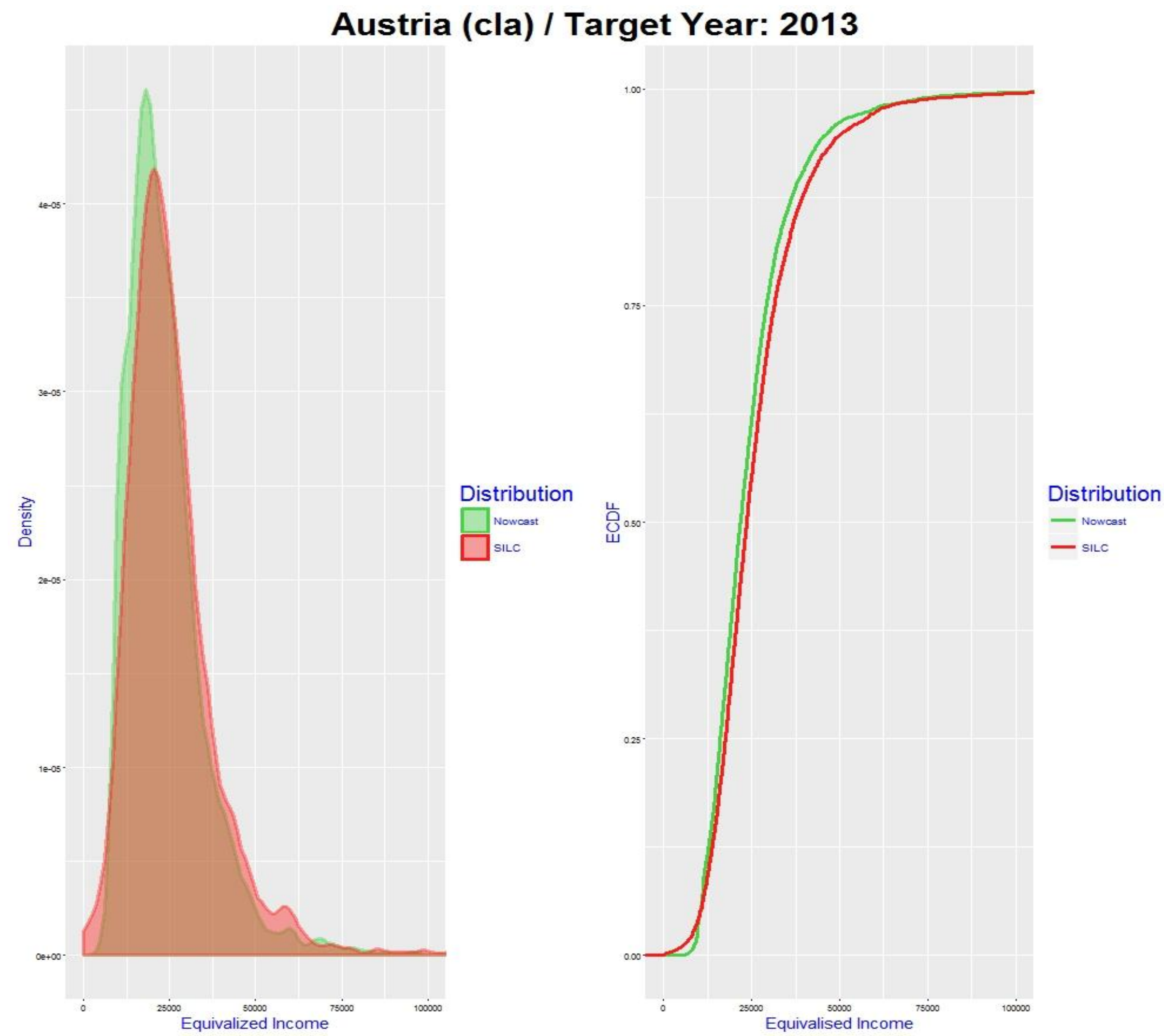
Two performance metrics are used: **accuracy** (based on the mean absolute percentage error, MAPE) and **consistency** (extent to which the year-on-year rates of change are similar across the time series). Accuracy and consistency involve subtraction from 1 in order to have higher values indicating better performance.

$$accuracy = 1 - MAPE = 1 - average_{i=1}^{n}\left(abs\left(\frac{EST_i}{REF_i} - 1\right)\right)$$

$$consistency = 1 - average_{i=1}^{n}\left(abs\left(\frac{EST_i}{EST_{i-1}} - \frac{REF_i}{REF_{i-1}}\right)\right)$$

The two performance metrics allow performing an extensive comparative analysis across countries, years and methods and to synthesize a large amount of information. However, more in depth investigation and corroboration with other performance checks is needed to provide an assessment concerning the quality of the flash estimates.

It is important to note that in both nowcasting approaches presented earlier, that is the microsimulation approach (see Section 2) and the parametric approach (see Section 3), the income indicators are not modelled directly but merely derived from the nowcasted income distribution. Hence, instead of directly evaluating the performance of the nowcasted indicators we might evaluate the performance of the latter by measuring the goodness-of-fit of the nowcasted income distribution with respect to the sample-based income distribution derived from the SILC dataset at the target year. The measure of performance in this kind of quality assessment procedure is given by the p-value of some non-parametric distributional test such as for example the Kolmogorov-Smirnov test (Massey, 1951) or the Anderson-Darling test (Anderson and Darling, 1954). You can see below the graph for the density plot and the cumulative distribution plots of the sample-based income distribution for income year 2013 and the associated nowcast obtained using the microsimulation approach.



Some general conclusions can be drawn for the two methodologies after the performance analysis:

- The two performance metrics were used for an extensive analysis of results for all countries in the case of the PQM approach and for ten countries for the microsimulation approach. In aggregated amounts, across all countries and methods, performance metrics for all indicators are quite large, all of them exceeding 90. This can give a first basis for comparison of different methods but it is more difficult to evaluate the results in absolute terms: it is often the case that a measure of 95% is not good enough. The thresholds for deciding if the indicators are good enough still need further investigation on a case by case basis.

- In general for the microsimulation approach the results seem better for consistency than accuracy so we are better in predicting changes rather than absolute values.
- For AROP, the flash estimates performance is overall lower than the other indicators.
- There is no one method better for all indicators and all years and this raises further questions for the method selection. There would be two steps to address this issue 1) quantify the loss of accuracy taking into account in a more systematic way all indicators but also model uncertainty/stability and in the next chapter a proposal is made in this direction 2) consider several methods together and make an estimate based on their convergence and/or combination (e.g. weighted according to their performance with past data).

4.2.2. Uncertainty measurement

Even if the results of the estimation process are expressed as absolute values for a set of indicators related to the income distribution, the primary interest for users and policy makers lies often in the changes or trends from one time period to another: e.g. were there changes in the at-risk-of poverty and are these changes statistically significant?

Therefore an important element in the quality assessment is the estimation of uncertainty related to the flash estimates. The uncertainty needs to be integrated in the quality framework in two stages:

1) In a first stage in the analysis of past performance in order to select the best method(s).
2) In a second stage it can provide a measure of quality for the flash estimate in the target year under some assumptions.

The importance of including variance analysis in the quality framework is straightforward when analyzing the trends of the income indicators in the SILC data. For example, an analysis of the at-risk-of-poverty changes for the last years shows that only 14 countries have a significant change for income years 2013/2012. In general for assessing if the change observed in EU-SILC between two reference periods is significant Eurostat is currently relying on the multivariate regression approach developed by Berger and Priam (2016). This takes also into account the fact that the comparison is based on two waves of cross –sectional data that are partially overlapping. An alternative method was tried to assess changes in the income distributions based on the Kolmogorov Smirnov test. This confirmed that indicators in several countries tend to be rather stable.

An important element of uncertainty in flash estimates is related to the sampling variance coming from EU-SILC. However, in order to obtain a measure of quality for flash estimates we need to integrate two other sources of uncertainty:

- ✓ Estimating model bias based on a simplified assumption of time-stationarity. Thus the systematic error might be estimated by computing the average of the differences between the flash estimate and the indicator derived from the SILC sample of each year (considered unbiased). Estimates for the target year can then be adjusted based on the past performance of the model.

- ✓ Developing estimates for model variance using simulation methods and a risk function associated not only with one indicator but with the set of estimates that quantifies the expected loss in accuracy in the form of a weighted sum of covariances of the set of estimated parameters. This would allow also using different weights for each of the estimates in order to take into account different scales or in order to give higher importance to a specific subset of target indicators.

Ongoing work has focused on developing the framework for the risk function as a way to quantify and compare different methodologies in a more systematic way. Risk functions play a major role in the model selection procedures. Model selection procedures rank different nowcasting approaches according to the value of their risk function and pick the estimation approach having the smallest value for the risk function or it can be used for weighting estimates coming from different methods.

Under the assumption that the sample-based estimates are unbiased estimates of the corresponding income indicators, we then have that the risk function can be written as

$$\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j E(\tilde{\rho}_i - \rho_i)(\tilde{\rho}_j - \rho_j) + m_i m_j (E(\hat{\rho}_i) - \rho_i)(E(\hat{\rho}_j) - \rho_j) + m_i m_j E(\hat{\rho}_i - E(\hat{\rho}_i))\left(\hat{\rho}_j - E(\hat{\rho}_j)\right)$$

Where $\tilde{\rho}_i$ to be the sample based estimator of ρ_i derived directly from the corresponding SILC dataset with $i = 1, \dots, N$.Thus, the risk function can be decomposed into three separate parts with $\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j E(\tilde{\rho}_i - \rho_i)E(\tilde{\rho}_j - \rho_j)$ being the quadratic sum of the covariances of the estimated income indicators derived directly from the SILC datasets, $\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j E\ (\hat{\rho}_i -$ *Eρi(ρj−Eρj)* representing the quadratic sum of covariances of the flash estimates and $\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j (E(\hat{\rho}_i) - \rho_i))(E(\hat{\rho}_j) - \rho_j)$ denoting the quadratic sum of cross-bias terms of the flash estimates.

However, a risk function is a theoretical expression depending on unknown parameters taking the form of first and second moments of the sample-based and the nowcasted estimators at the target year. Thus, in order to allow for risk functions to be used in the context of a model selection algorithm, these unknown parameters need to be replaced by estimates.

To do this we first replace the unknown income indicators by the corresponding sample based estimate derived directly from the respective SILC dataset at the target year. The second moments of the SILC-based estimators are derived on the basis of a bootstrapping procedure consisting in drawing a large number (M) of subsamples of the SILC dataset at the target year. For each of these subsamples we then compute the corresponding set of indicators providing us with M observations for each of the indicators. The variances and covariances are then computed using these M observations.

In the parametric approach the expected values and the covariances of the flash estimates are obtained on the basis of a dynamic factor model (Banbura et al., 2013) of the income percentiles which is estimated on the basis of an iterative optimization algorithm combining Kalman filter/smoother (Kalman, 1960) and the expectation-maximization (Shumway and Stoffer, 1982) approach using historical SILC data and past and present data of a collection of more timely explanatory variables.

In the remaining part of this section we are going to illustrate how the risk function can be used to assess and compare the performances of flash estimates provided by different modelling approaches. One of the major issues in the field of factor modelling is to identify the number of factors to be included into the model framework. One possibility to fix the number of factors is to compute the risk function for different factor models which only differ in their respective number of factors. The optimal factor model and hence the optimal number of factors is then given by the model which minimizes the risk function.

The table below shows the values of the risk function for three different factor models with one, two or three factors respectively. Given that the second moments of the sample estimates are identical for each of the factor models, the latter do not intervene in the model selection procedure.

	Single-Factor	Two-Factor	Three Factor
$\sum_{i=1}^{N} m_i m_j E(\hat{\rho}_i - E(\hat{\rho}_i))(\hat{\rho}_j - E(\hat{\rho}_j))$	0.0070	0.0045	0.0041
$\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j E(\hat{\rho}_i - E(\hat{\rho}_i))(\hat{\rho}_j - E(\hat{\rho}_j))$	0.0083	0.0359	0.0412
Risk Function	0.0153	0.0404	0.0453

Given the relatively small number of parameters of the single-factor model specification, the latter is characterized by a large degree of robustness relative to the other model specifications reflecting in a low value of the aggregated model-based second moments and eventually in a low value of the corresponding risk function.

This illustrates how the risk function can be used in the model selection taking into account the whole set of indicators, the variance associated with the point estimate and the bias. In a second stage, it can also provide a quality measure for the flash estimates of the target year under the assumptions of stationarity.

5. Conclusions

Therefore, the quality assessment for flash estimates on the income distribution has two main stages:

- The first one focused mainly on choosing the best methodology and sources or a combination of these for flash estimate based mainly on assessment of past performance. Overall, the first steps in the quality assessment focused on assessing the plausibility of **point estimates** based on intermediary checks, performance for the past data based on different metrics.
- The second stage will be mainly focused on estimating a quality measure based on the chosen model(s), its performance with past data and the estimation of uncertainty. This would result in the estimation of a prediction interval that might also require weighting of different methods.

The last point on **uncertainty measurement** raises not only a question of quality but also of communication. In some cases the time series in SILC are rather stable so flash estimates, even if small changes are registered, can give only a general message of stability. In other cases the direction and magnitude of change might be significant and more precise figures and a more clear quantification of change can be presented.

6. References

Anderson, T. W. and Darling D. A. (1954), A test of goodness-of-fit, Journal of the American Statistical Association, vol. 49, no. 268, pp. 765-769.

Banbura, M., Giannone, D., Modugno, M. and Rechlin, L. (2013), Now-casting and the real-time data flow, ECB Working Paper Series, no. 1564, pp. 1-53.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, Transactions of the ASME-Journal of Basic Engineering, no. 8, pp. 35-45.

MacDonald, J. B. and Xu, Y. J. (1995), A generalization of the beta distribution with applications, Journal of Econometrics, vol. 66, no. 1-2, pp. 133-152.

Massey, J. (1951), The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, vol. 46, no. 253, pp. 68-78.

O'Donoghue, C., Loughrey, J. (2014) Nowcasting in Microsimulation Models: A Methodological Survey, Journal of Artificial Societies and Social Simulation 17 (4) 12

Shumway, R. H. and Stoffer, D. S. (1982), An approach to time series smoothing and forecasting using the EM algorithm, Journal of Time Series Analysis, vol. 3, no. 4, pp. 253-264.