WORKING PAPER 3-19

This project is
co-funded by the
European Union

Medium-term projection for Belgium of the at-risk-of-poverty and social exclusion indicators based on EU-SILC

February 2019

Dekkers Gijs, gd@plan.be
Tarantchenko Ekaterina, te@plan.be
Van den Bosch Karel, kvdb@plan.be

Avenue des Arts 47-49 – Kunstlaan 47-49
1000 Brussels

E-mail: contact@plan.be
http://www.plan.be

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

url: http://www.plan.be
e-mail: contact@plan.be

Responsible publisher: Philippe Donnay
Legal Deposit: D/2019/7433/11

Federal Planning Bureau
Avenue des Arts - Kunstlaan 47-49, 1000 Brussels
phone: +32-2-5077311
fax: +32-2-5077373
e-mail: contact@plan.be
http://www.plan.be

Medium-term projection for Belgium of the at-risk-of-poverty and social exclusion indicators based on EU-SILC

January 2019

Dekkers Gijs, Tarantchenko Ekaterina, Van den Bosch Karel
gd@plan.be, te@plan.be, kvdb@plan.be

Abstract - The Federal Planning Bureau has developed within the Nowcasting project a dynamic microsimulation model for nowcasting and medium-term forecasting (currently up to 2020) of indicators of poverty and social exclusion. The model uses the EU-SILC 2014 cross-sectional dataset – complemented by auxiliary information – as input data to simulate micro-datasets for each of the subsequent projection years. From these, estimates of the at-risk-of-poverty rate (AROP), the very low work intensity rate (VLWI) and the index of severe material deprivation (SMD) can be calculated; that is, all three components of the overarching AROPE ("at-risk-of-poverty or social exclusion rate") indicator.

This Working Paper begins by describing the ambitions of this project in more detail. The data and methodology are presented afterwards, highlighting an innovative approach that allows an individual to have multiple socio-economic states during the year. The results for the indicators of poverty and social exclusion are presented in some detail. Finally, it surveys the main challenges and limitations and explores some avenues for future improvements of the model.

This research project was jointly funded by the European Commission and the Belgian Federal Public Service Social Security (FPS SS) through the NOWCASTING project VS/2015/0179. The contents of this note reflect only the views of the authors and the Commission nor the FPS SS are responsible for any use that may be made of the information it contains.

Jel Classification - C53, H31, I32
Keywords - Nowcasting, Microsimulation, At-risk-of-poverty rate, Very low work intensity, Severe material deprivation, Belgium

Table of contents

List of tables

List of graphs

Executive summary

The Federal Planning Bureau has developed within the Nowcasting project a dynamic microsimulation model for nowcasting and medium-term forecasts (currently up to 2020) of indicators of poverty and social exclusion. The main motivation for the Nowcasting project is that these indicators become available with a delay, which is an important obstacle for the use of these indicators in a policy context. The model is mainly based on the EU-SILC 2014 cross-sectional dataset, complemented by more detailed variables from the Belgian questionnaire. This dataset, combined with semi-aggregate projections by the Federal Planning Bureau of demographic and labour-market developments, is used as input data for the model to simulate micro-datasets for each of the subsequent projection years. From these data, estimates of the at-risk-of-poverty rate (AROP), the very low work intensity rate (VLWI) and the index of severe material deprivation (SMD) can be calculated; that is, all three components of the overarching AROPE ("at-risk-of-poverty or social exclusion rate") indicator. In addition, the Gini coefficient of income inequality is also projected. These simulated indicators of poverty and social exclusion can be calculated both for the whole population and for any sub-group. Projections of many other indicators in the portfolio of EU social indicators that are based on EU-SILC could also be derived from the simulated data. The advantage of a microsimulation model is that it allows to assess the impact of policy reforms and of economic and demographic developments on these indicators of poverty and social exclusion.

Key messages of this project are that nowcasting and medium-term forecasting are now possible using a fully dynamic microsimulation model. The provisional results of the model suggest that the overall poverty risk would remain stable, but that of the 65+ subpopulation would decrease over time, while that of the younger population would show a small increase. Furthermore, the increase of overall inequality would come to a halt and the level of inequality would become more stable. Finally, the very low work intensity rate would continue its decrease, driven by the continuing increase of the employment rate among the working-age population.

The current version of the model is still experimental and in need of some adjustments. Also, it should be updated to the latest available wave of the EU-SILC data. The model, when adapted and updated, would allow Belgium to provide more up-to-date and projective information on indicators of poverty and social exclusion to the EU level for international comparisons and analysis, among other things for the preparation of the European Semester and for the evaluation of the EU2020 programme. The model itself might be put at the disposal of other Belgian agencies or international partners. Finally, some methodological advancements made in the development of the Nowcasting model are of more general relevance and will be disseminated to the international community of developers of microsimulation models.

The Nowcasting project was jointly funded by the European Commission (project VS/2015/0179) and the Federal Public Service Social Security.

Synthèse

Le Bureau fédéral du Plan a développé, dans le cadre du projet Nowcasting, un modèle de microsimulation dynamique pour la prévision immédiate (nowcasting) et les prévisions à moyen terme (actuellement jusqu'à 2020) des indicateurs de pauvreté et d'exclusion sociale. La motivation principale du projet Nowcasting est que ces indicateurs deviennent disponibles avec un certain retard, ce qui constitue un obstacle important à leur utilisation dans un contexte politique. Le modèle est basé principalement sur l'ensemble de données transversales EU-SILC 2014 complétées par des variables plus détaillées du questionnaire belge. L'ensemble de données ainsi constituées, combinées avec les projections semi-agrégées du Bureau fédéral du Plan sur les évolutions démographiques et du marché du travail, sont utilisées comme données d'input par le modèle afin de simuler un set de micro-données pour chacune des années de projection. Ces micro-données sont ensuit utilisées pour calculer les estimations du taux de risque de pauvreté ("at-risk-of-poverty rate", AROP), du taux de très faible intensité de travail ("very low work intensity rate", VLWI) et du taux de privation matérielle sévère (SMD) ; c'est-à-dire les trois composantes de l'indicateur global AROPE ("at-risk-of-poverty or social exclusion rate"). En outre, le coefficient de Gini de l'inégalité des revenus est également projeté. Ces indicateurs simulés de pauvreté et d'exclusion sociale peuvent être calculés à la fois pour l'ensemble de la population et par sous-groupes. Les données simulées peuvent également servir à obtenir les projections de nombreux autres indicateurs du portefeuille d'indicateurs sociaux de l'UE basés sur les statistiques EU-SILC.

L'avantage d'un modèle de microsimulation est qu'il permet d'évaluer l'impact des réformes politiques et des changements économiques et démographiques sur ces indicateurs de pauvreté et d'exclusion sociale.

Les messages clés de ce projet sont les suivants : la prévision immédiate et la prévision à moyen terme sont maintenant possibles en utilisant un modèle de microsimulation dynamique. Les résultats provisoires du modèle suggèrent que le risque global de pauvreté resterait stable, mais que celui de la sous-population des 65 ans et plus diminuerait avec le temps, tandis que celui de la population plus jeune augmenterait légèrement. En outre, les résultats montrent une légère croissance d'inégalité des revenus pour l'ensemble de la population dans les premières années mais cette croissance s'estomperait ensuite et l'inégalité deviendrait plus stable. Enfin, l'indicateur VLWI baisserait sur l'ensemble des années simulées. Cette baisse serait alimentée par l'augmentation continue du taux d'emploi de la population en âge de travailler.

La version actuelle du modèle est expérimentale, à ce stade, et nécessite quelques ajustements. De plus, elle devrait être mise à jour avec la dernière vague disponible de données EU-SILC. Une fois adapté et mis à jour, ce modèle permettrait à la Belgique de fournir des informations sur les indicateurs de pauvreté et d'exclusion sociale plus actualisées et plus prospectives au niveau de l'UE pour des comparaisons et analyses internationales, notamment pour la préparation du Sommet Européen et pour l'évaluation de la stratégie EU2020. Le modèle lui-même pourra être mis à la disposition d'autres institutions belges ou partenaires internationaux. Enfin, certains avancements méthodologiques réalisés dans le cadre du développement du modèle de prévision immédiate (Nowcasting) ont un intérêt plus général et seront diffusés auprès de la communauté internationale des développeurs de modèles de microsimulation.

Le projet Nowcasting a été financé conjointement par la Commission Européenne (projet VS/2015/0179) et le Service Public Fédéral Sécurité Sociale.

Synthese

Het Federaal Planbureau ontwikkelde binnen het Nowcasting project een dynamisch microsimulatiemodel voor nowcasting en middellange-termijn projecties (tot 2020) van indicatoren van armoede en sociale uitsluiting. De voornaamste aanleiding voor dit project is de vertraging waarmee deze indicatoren ter beschikking komen, hetgeen een belangrijke hinderpaal is voor hun gebruik in een beleidscontext. Het model is vooral gebaseerd is op de EU-SILC 2014 cross-sectionele dataset en aangevuld met meer gedetailleerde informatie uit de Belgische bevraging. Deze dataset, in combinatie met projecties door het Federaal Planbureau van demografische variabelen en van de arbeidsmarkt, dient als input voor het model dat golven van de micro-data voor de volgende jaren simuleert. Met deze gegevens kunnen schattingen van het armoederisico ("at-risk-of-poverty rate", AROP), de "very low work intensity rate" (VLWI) en de "severe material deprivation rate" (SMD) worden berekend; dit zijn de drie componenten van de overkoepelende AROPE ("at-risk-of-poverty or social exclusion rate") indicator. Daarnaast wordt ook de Gini coëfficiënt van de inkomensongelijkheid geprojecteerd. Deze gesimuleerde indicatoren van armoede en sociale uitsluiting kunnen worden berekend voor de bevolking als geheel, en ook voor alle mogelijke subgroepen. Ook projecties van andere indicatoren uit de portfolio van sociale indicatoren van de EU, voorzover deze gebaseerd zijn op EU-SILC, kunnen uit de gesimuleerde data afgeleid worden.

Het voordeel van een microsimulatiemodel is dat het een inschatting mogelijk maakt van de effecten van beleidshervormingen en van sociaal-economische en demografische ontwikkelingen op indicatoren van armoede en sociale uitsluiting.

De conclusies van dit project zijn de volgende: ten eerste is gebleken dat het produceren van nowcasting en middellange-termijn projecties mogelijk is met een dynamisch microsimulatiemodel. Ten tweede suggereren de voorlopige resultaten dat het algemene armoederisico ruwweg stabiel zou blijven, waarbij het voor de 65+ zou afnemen terwijl het voor de actieve bevolking licht zou toenemen. Ten derde zou de algemene inkomensongelijkheid niet verder toenemen maar zich stabiliseren. Tenslotte, de VLWI zou verder afnemen, vooral onder invloed van de toenemende activiteitsgraad onder de bevolking op actieve leeftijd.

De huidige versie van het model is nog onvoltooid, en kan nog verbeterd worden. Ook zou het geüpdated moeten worden naar de laatst beschikbare golf van de EU-SILC gegevens. Het aangepaste en geüpdatete model zou België in staat stellen om projecties te maken van de indicatoren van armoede en sociale uitsluiting en deze te gebruiken voor comparatieve analyses op Europees niveau, onder meer in de voorbereiding van het Europese Semester en ter evaluatie van het EU2020-programma. Het model kan zelf ook ter beschikking worden gesteld aan andere Belgische instellingen of internationale partners. Enkele methodologische innovaties in het kader van dit model hebben een ruimere relevantie, en zullen worden gecommuniceerd naar de internationale gemeenschap van bouwers van microsimulatiemodellen.

Het Nowcasting project werd gefinancierd door de Europese Commissie (project VS/2015/0179) samen met de FOD Sociale Zekerheid.

1. Introduction

Since the launch of the Europe 2020 strategy in 2010, the statistics on poverty and social inclusion in the EU have gained in importance. This is not surprising given the important role of these indicators in the preparation of the European Semester (the annual cycle of economic policy coordination between EU countries) (European Commission, 2018b). Moreover, one of the headline targets in the EU2020 strategy is to reduce by at least 20 million by 2020 the number of people living at risk of poverty or social exclusion. This target is monitored through the at-risk-of poverty or social exclusion (AROPE) rate that is composed of three components: (a) monetary poverty or the at-risk-of poverty rate (AROP); (b) severe material deprivation (SMD) and (c) very low work intensity (VLWI). Those three components are derived from the European Union Statistics on Income and Living Conditions (EU-SILC), which is the main source of comparable statistics on income, social exclusion and living conditions at EU level.

EU-SILC data are collected at the household and the individual level, covering demographic and labour-market characteristics as well as detailed income information. In Belgium, EU-SILC data are collected by Statbel in an annual survey of about 6,000 households (or 11,000 individuals).[1] A major drawback of EU-SILC, however, is that the data become available with a delay. This delay has been substantially reduced in the past couple of years and more efforts are ongoing in that direction. Currently, the data become available in May of year N+1, i.e. 5 months after the year N of data collection. However, due to the complexity of the data collection and its processing, there will always be a certain delay in the production of EU-SILC data. Moreover, all income variables refer to N-1, i.e. the year preceding the one of data collection. Providing timelier statistics on income poverty and inequality is essential in evaluating the effectiveness of the Union's policies and programmes and measuring the progress towards the Europe 2020 poverty target. Therefore, it is important to supplement the official statistics with timelier indicators.

The aim of this project is to resolve the problem of the delay in the EU-SILC results by constructing a so-called "nowcasting" model based on the dynamic microsimulation approach. Nowcasting refers to a contraction of "now" and "forecasting" and involves projection of current indicators based on the last available observed data that are updated to the present time using macro-level indicators of the trends in incomes and on the labour market. However, the model developed in the present project goes beyond the "simple" nowcasting exercise and also provides short- and medium-term projections of the EU-SILC indicators. This is the second aim of the project. So, nowcasting here implies the projection of socio-economic developments in the proximate past *and* in the near future. The proposed nowcasting model delivers estimates of the at-risk-of-poverty rate (AROP) and the very low work intensity rate (VLWI). A regression model makes it possible to also project the index of severe material deprivation (SMD), so that all three components of the overarching at-risk-of-poverty-or-social-exclusion (AROPE) indicator are covered. The advantage of such a model (i.e. based on microsimulation techniques) is that it allows to assess the impact of policy reforms and other economic developments on these indicators of poverty and social exclusion, both for the whole population and for any sub-group. The model will allow

1 See https://statbel.fgov.be/en/themes/households/poverty-and-living-conditions/risk-poverty-or-social-exclusion#documents

Belgium to provide more up-to-date and projective information based on the EU-SILC to the EU level for international comparisons and analysis, among other things in the context of the EU2020 programme.

An earlier projection of the population at risk of poverty or social exclusion up to 2030 was published by the Federal Planning Bureau in 2016 (Frère, 2016). This projection integrates into a coherent framework various data which influence the evolution of the size of this population. This projection is made on the macro-level, which implies that breakdowns by population subgroups are generally not possible.

The indicators presented in this paper are provisional and should be interpreted with caution. The purpose of the developed model is to provide nowcasted estimates as close as possible to the EU-SILC survey. However, the model cannot be expected to capture the changes in the EU-SILC survey perfectly and a comparison of the projected indicators with the EU-SILC based observations reveals some deviations. The accuracy of the nowcasted indicators depends on various parameters, among them the differences between EU-SILC and exogenous parameters used in the model (e.g. earnings growth rate or changes in employment rate), assumptions underlying the model and estimations that may fail to properly capture the observed behaviour. Notwithstanding these possible deviations and limitations of the model (that is subject to further improvement), the nowcasted estimates can still be used as an indication of the direction of current and near-future changes.

The model has been developed within the Nowcasting project, which was jointly funded by the European Commission (project VS/2015/0179) and the Federal Public Service Social Security.

The remainder of this paper is organised as follows. Section 2 discusses the nowcasting methodology developed by Eurostat using EUROMOD and compares it with the model presented in this report. Section 3 describes the data used to implement the Nowcasting model. Section 4 presents the Nowcasting model and its various modules. The results are presented in section 5. The same section also includes simulation results for a hypothetical policy variant, illustrating an alternative way of using the model. Finally, section 6 is devoted to the challenges and limitations of the project and section 7 concludes.

2. Nowcasting by Eurostat, using EUROMOD

'Nowcasting', a contraction of 'now' and 'forecasting', is the projection of economic and other developments for the recent past, the present and the near future. (Banbura et al., 2013). Nowcasting models based on microsimulation techniques have been recently developed to analyse the effect of current public policies on socio-economic indicators. This interest is driven by micro-data usually becoming available with time delay. Nowcasting then "updates" data from the recent past to the present, before real data are available directly from the source.

Because indicators on poverty and income inequality play an essential role in the European semester and in the EU2020 goals, there is a manifest need for more timely information about these indicators than is available through the source database EU-SILC. The latter requires time for collecting and processing, which results in time delay. For this reason, Eurostat is developing 'flash estimates' for these indicators, which were published for the first time in 2017 for the income year 2016. The second wave of these flash estimates, which are still experimental, was published in October 2018 for the income year 2017 (European Commission, 2018a).

The flash estimates are not based on a single methodology. For each country, the most suitable method is used, given the information available and in consultation with national authorities. Also, methods are constantly revised within a Quality Assessment Framework (European Commission, 2018b). For some countries, current income information from EU-SILC or from other sources was used for the latest flash estimates. Earlier, macroeconomic time series modelling was tested, but not used anymore for the 2017 flash estimates. This approach models indicators of poverty and income inequality on the aggregate level, without recourse to microdata. Because of the scarcity of data points and other reasons, the results were considered insufficiently robust.

The main method used for most countries, including Belgium, is microsimulation, relying on EUROMOD. For several years, the EUROMOD team has worked on adapting the EUROMOD model for use in nowcasting (European Commission, 2010, 2013; Leventi et al., 2014; Rastrigina et al., 2015). For this purpose, the EUROMOD model *strictu sensu* (i.e. the tax-and-benefit simulation routines) was enhanced with adjustments to the input data. Those adjustments take into account changes in the population structure, the evolution of employment and the growth of the main income components.

The EUROMOD nowcasting process is implemented in two steps. The first step consists in updating the labour and demographic characteristics. This is done by using two different models, that boil down to the use of two key techniques in microsimulation – dynamic and static ageing (Dekkers, 2015a). For most countries (including Belgium), a static approach is used, where changes in the population are taken into account through a process of reweighting. A new vector of the sample weights is derived to change the marginal distributions from the base year to those in the target year. The reweighted distributions include the number of employed, the age structure, household size and the number of dependent children. The main source for the target marginal distributions is the Labour Force Survey. The reweighting is done at the household level.

In the second model, which is being used for 8 of the 24 member states, changes in employment are modelled at the level of the individuals by explicitly simulating two kinds of transitions: from non-employment into employment and from employment into unemployment (or inactivity). People that transit into employment will have wages imputed from an individual with similar characteristics, such as age, educational attainment level and region. Unemployment benefits are simulated for those moving out of employment, if they are eligible. Detailed employment figures by age, gender and economic sector based on the Labor Force Survey are used for deciding the final number of individuals making the transitions. The number of cases that are selected to go through a transition corresponds to the net yearly change in employment levels by age group, gender and education (a total of 18 strata) as shown in the Labour Force Survey (Rastrigina et al., 2015).

Eurostat and the EUROMOD team do not favour one of these technics over the other, but note that both have advantages and disadvantages (European Commission, 2018b; Navicke et al., 2014; cf. Dekkers, 2015a). A disadvantage of the static reweighting method is that it makes the implicit assumption that the characteristics of the reweighted groups remains unchanged, which may be unrealistic. For example, during an economic crisis, the characteristics of those that move into unemployment may be rather different from those that were already unemployed. Simulating transitions from employment to unemployment may be better at capturing real developments in such circumstances. Also, reweighting may result in excessively high weights for some observations. On the other hand, reweighting is simpler to put in practice.

The second step in the EUROMOD nowcasting process consists of updating all non-simulated incomes to the target year, using uprating coefficients that reflect either statutory rules, or observed changes in average incomes by income source. As far as is feasible, the updating factors for employment income are disaggregated by economic activity or economic sector. Furthermore, some benefits and taxes are simulated, using the regulations in effect in the target year. Finally, an alignment factor is calculated for each household, to correct for differences between the income as observed in EU-SILC 2016 and as estimated in EUROMOD. This factor is then applied in all later years.

3. Data

The currently developed version of the Nowcasting model is based on the Belgian EU-SILC 2014 cross-sectional dataset, as well as the EU-SILC 2007-2014 longitudinal data. Both types of data are provided by Statbel. In addition to conducting the Belgian survey, Statbel also checks and corrects the collected data for outliers and missing values and reorganizes data into EU-SILC variables to enable comparisons between the European countries and regions. Nowcasting mostly uses these reorganized variables but complements them with more detailed information from the questionnaire on different social security schemes (*e.g.* distinction between retirement pensions, survival pensions and guaranteed minimum income for elderly; or distinction between regular unemployment benefits and unemployment benefits with employer's supplement). The cross-sectional dataset is used as input data for the model to produce synthetic datasets for each of the following projected years. The longitudinal version of EU-SILC is mostly used for estimation of parameters in the dynamic simulation model. More precisely, the estimations of the labour market transitions (probability to be in a particular labour market state in year t given the labour market state in year t-1), wages, working hours and pensions are based on the longitudinal EU-SILC data. This falls outside the scope of the current note.[2]

A cleaning process is performed on the base dataset before the data is imported into the Nowcasting model. Most data inconsistencies are solved in the cleaning process. Some inconsistencies could not be solved, and the Nowcasting model therefore was adapted accordingly. The two major unsolved inconsistencies are the following:

- Some individuals are reported to be married but the identifier of their partners is missing. In many cases, we need to know partner's characteristics such as his labour market state or income (*e.g.* for the simulation of social assistance benefits). If the partner's identifier is missing, this would lead to inconsistencies within the model. In such cases, we assume that the person is single for practical reasons (*i.e.* the person is probably not living together with his/her partner).
- Some key variables have missing values (*e.g.* level of education, labour market state, white- or blue-collar identifier for wage earners, etc.). In most microsimulation models, these missing values are imputed. We decided to keep missing values for some of these variables (*e.g.* labour market state)[3] and replace missing values for other variables (*e.g.* level of education). The decision not to impute missing values is based on the purpose of the project, which is to create synthetic waves of the initial dataset. That is, we want to be as close as possible to the initial data that allows for missing values. On the other hand, labour, education and health information in EU-SILC is only collected for persons aged 16 and over. Hence and by design of the survey, the level of education is mostly missing for people aged below 16. This is a problem, because those younger than 16 at the outset over time become 16 and older. Hence those in this group gradually might move out of education and into one of the other labour market states, which means that the missing education level becomes a problem.

---

2 Appendix 9.4 includes a technical note on the mentioned estimations used within the Nowcasting model.

3 To be more precise, the model allows the *possibility* to observe missing labour market states. To complete the missing information in the reported current labour market state, we mostly used the information on the reported state in December of the income reference period and the type of reported income. In the currently used dataset (EU-SILC 2014), all missing values could thus be imputed. It is nevertheless possible that future input datasets based on the following EU-SILC waves contain labour market state with missing values. In this case, the model will take these missing values into account.

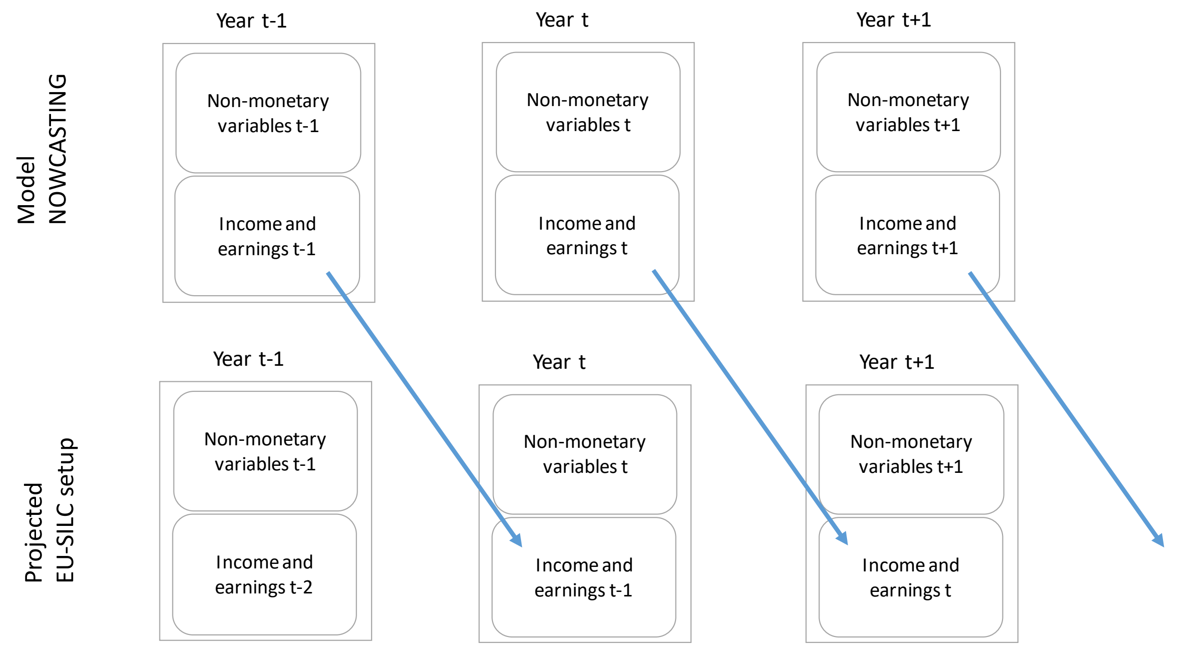
Hence the need to fill the missing values for the level of education to simulate labour market transitions because the transition probabilities depend on the level of education. The missing values are imputed randomly consistently to the age of the individual and observed frequencies within each level of education.

An important adaption to the database is the presentation of income information. In the EU-SILC survey, most information except the income data in the EU-SILC wave of year *t* is provided for the calendar year *t*, while the income data is reported for the previous calendar year *t*-1.[4] This is shown in the lower panel of Graph 1. Thus, for example, the EU-SILC of wave 2014 asks the respondents about the labour market state that they currently occupy, and about the earnings or incomes they received in 2013. In contrast, the Nowcasting model simulates the incomes that pertain to the labour market state in the same year - if an individual in the model is working, then she receives an income from work. This is shown in the upper panel of Graph 1.

The purpose of the Nowcasting model is to nowcast the EU-SILC data. Hence the structure of the simulated results must be changed in order for this time lag in the presentation of earnings and incomes to be present. This is done as shown in Graph 1 below. First, the Nowcasting model updates the incomes from year *t*-1 to *t*. This is done by carrying over the incomes from *t*-1 to *t* if the income source matches the current labour market state. If an individual's labour market state changed between *t*-1 and *t* (i.e. income source from *t*-1 does not match with the labour market state at *t*), then we simulate the new incomes at *t*. For the following (simulated) years, Nowcasting simulates all the variables simultaneously (labour market state, incomes, etc.) for a given period. Finally, after simulation, simulated incomes of each previous year *t*-1 are linked to the non-monetary simulation results (such as labour market state) of year *t*, as denoted by the arrows in Graph 1. Thus, in the results reported in the reminder of this note, one should keep in mind this EU-SILC structure of the data.

4 An exception to this is the labour market state reported for each month of the income reference period, i.e. calendar year *t*-1.

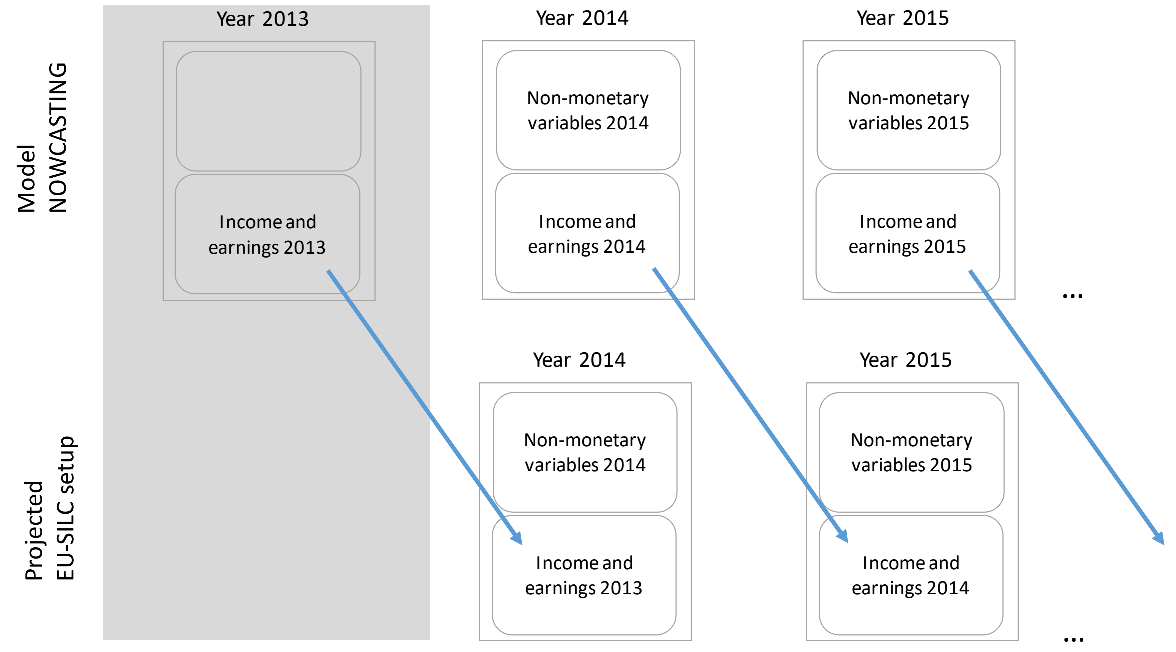
Graph 1 Replicating the structure of the EU-SILC (general setup)



Graph 2 illustrates how the EU-SILC structure is replicated in the model using the current input data of 2014 EU-SILC wave. First, in the input year 2014, most non-monetary variables are observed.[5] Based on these observed data, Nowcasting simulates incomes and earnings for 2014. Starting 2015, the model simulates both non-monetary and income variables pertaining to the same year. This is shown in the upper panel of Graph 2. Second, after all the data has been simulated, the incomes from *t*-1 are carried over to *t*. This means that for 2014, the output of the model includes (mostly) observed non-monetary data from 2014 and observed incomes from 2013, thus basically the input data. The output data for 2015 includes simulated non-monetary variables from 2015 and simulated incomes from 2014. This projected EU-SILC setup is displayed in the lower panel of Graph 2.

[5] Due to the previously mentioned inconsistencies, certain missing information is simulated. In addition, all the persons aged 65 or older are considered as retired in the Nowcasting model and their observed labour market state is adapted accordingly if necessary. In the currently developed version of the model, the labour market state has been changed to retirement for around 3% of 65+ population (based on weighted data).

Graph 2 Replicating the structure of the EU-SILC (example)



An important difference between EU-SILC and simulated data is that EU-SILC is a rotating panel, while the simulated data are close to a full panel. In EU-SILC, one fourth of the sample is renewed every year. In contrast, in Nowcasting, the largest part of the population observed in the initial dataset remains in all the simulation years. A unique exception to this are immigration/emigration and fertility/mortality changes in the data. Therefore, although the EU-SILC sample remains representative of the population in all years, larger sampling errors may occur while measuring annual net changes and trends than if it was a full panel.

Furthermore, note that the Nowcasting model is in constant prices with all incomes in all simulated years expressed in prices of the base year 2014.

4. Methodology

4.1. General approach

The model developed in this project is a fully dynamic longitudinal-ageing microsimulation model (Li et al., 2014), where the characteristics of individuals change following probabilistic or deterministic processes. Dekkers and Van den Bosch (2016) show that dynamic microsimulation models are not uncommon for policy analysis of especially pensions in European member states, but these simulations often focus on the long run. Thus, our method falls within the "evolution of income components-microsimulation" approach used in flash estimates of the at-risk-of-poverty rate by Eurostat (European Commission, 2018b).

The model is developed and run using the LIAM II software[6], a specialized tool for microsimulation.

Like the Eurostat-approach to produce flash estimates, the dynamic microsimulation model developed in the NOWCASTING project also uses auxiliary data to take account of demographic and economic developments in the short run. However, unlike flash estimates that use calibration techniques[7], our model relies on alignment techniques (Li and O'Donoghue, 2014) which use the individuals' estimated risk of any transition happening (say, finding a job if one is unemployed, or divorcing if one is married). It makes sure that the proportion of individuals in any state (e.g. working, or being unemployed) in the simulation matches an externally given proportion, often disaggregated by age and gender. We will come back to the alignment techniques used within the model later in section 4.3.

Overall, the Nowcasting model presented in this report simulates 1) demographic developments (fertility, mortality, immigration and emigration), 2) household formation and dissolution (marriage and cohabitation, divorce and separation, household-leaving by young adults), and 3) labour market and social security transitions. The latter point covers transitions between employment, unemployment, disability, unemployment with company allowance, inactivity (including full-time education) and, finally, retirement. Furthermore, within the labour market, transitions are simulated between types of work (private sector, public sector, self-employment). All transitions are based on individual risk profiles reflected by either reduced-form logistic regressions or continuous uniform distribution, that depend on the previous state of the individual. Through alignment procedures (Li and O'Donoghue, 2014), these are combined with semi-aggregate short-run demographic and labour market simulations by the MALTESE system of models of the Federal Planning Bureau (FPB), originated from the HERMES model (see section 4.3 for a discussion). See Dekkers, Inagaki and Desmet (2012) for a more extensive discussion of simulation of transitions with alignment, and Federal Planning Bureau (2017) for a short description of the MALTESE model.

Furthermore, demographic transitions, specifically household formation (marriage, cohabitation) and dissolution (divorce, separation) are all aligned to the LIPRO-projections by the FPB (Vandresse et al., 2018).

6 http://liam2.plan.be/pages/about.html# . See also de Menten et al. (2014)

7 For some European countries, flash estimates are based on labour transitions derived at the individual level (European Commission, 2018). For Belgium however, and most other European countries, calibration/reweighting techniques are used.

The other aspects of time mentioned in European Commission (2018b) are the development of earnings and social security incomes. The changes of the rules of the various social security schemes and the gross-net trajectory are also simulated.

The Nowcasting model is an annual discrete time model and in the above characteristics, it is not unlike other dynamic microsimulation models. This approach is however innovative in the sense that other discrete time microsimulation models usually assume a unique person-year socio-economic state with individuals remaining in the same socio-economic state during the whole year. This means that if an individual changes her socio-economic state between year t-1 and year t, the change is assumed to occur at the beginning of the year t. In contrast, the model presented here allows an individual to change its socio-economic state at any time during the year, and therefore to serially occupy 2 different states over the course of the year.

The EU-SILC data include weights, that correct for differences in the probability to be selected due to the sampling design and for differential non-response, and also are calibrated so that sample totals match population numbers by age, gender and province[8]. Unfortunately, weights as such cannot be used in the present version of LIAM II. For this reason, these weights are rounded to whole numbers, and the records for each household are copied as many times as those numbers indicate. In other words, if we designate the weight for each household h by w_h, then each household is transformed into w_h 'clones' of itself (Dekkers and Cumpston, 2012). This procedure is called 'expansion', and the expanded sample yields the same results as the original weighted sample, except for rounding errors.

In the demographic module (see section 4.2.1), birth, death, immigration and emigration are simulated in a way which ensures that the projected population, in terms of the distribution by age and gender, conforms to the population projections or observations. Thus, no further reweighting is needed.

4.2. Modules

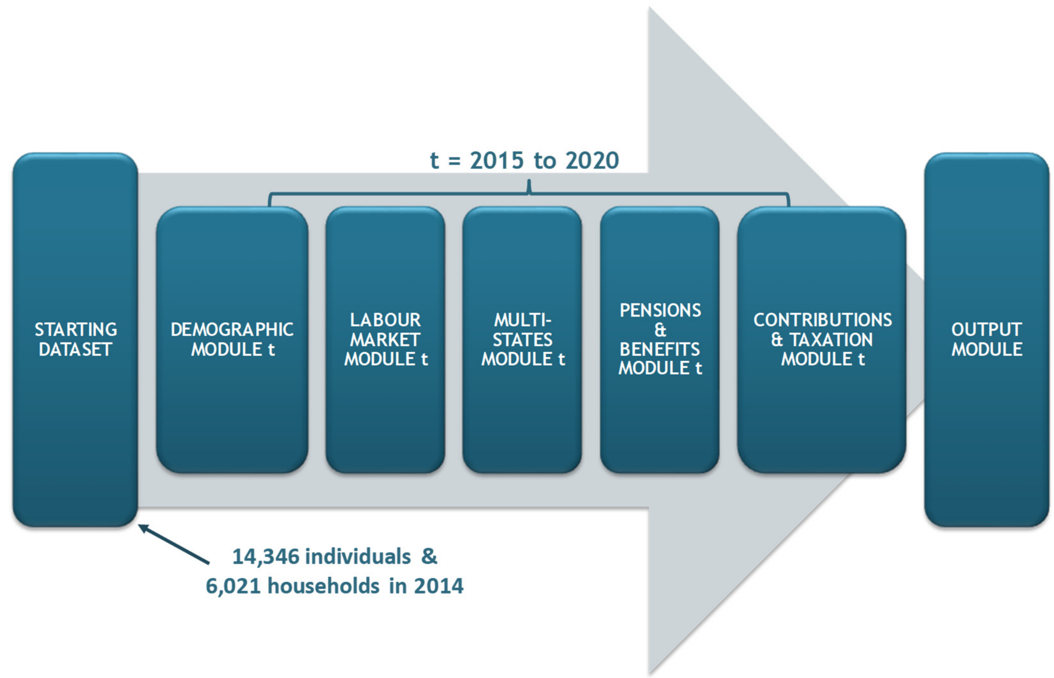
The Nowcasting model consists of many processes, and related processes are arranged in modules. Furthermore, adopting a discrete-time approach means that any object (be it an individual or a household) can face one particular "risk" only. Being simultaneously subject to multiple risks, for example the risk of catching a cold and being in a car crash, is typically for an environment where time is continuous. Hence, in a discrete-time model, recursive simulation is used, by which events are simulated through processes one at a time and in a fixed and particular order for a given period (Dekkers and Van den Bosch, 2016). This order is relevant because if the simulation of event *a* precedes that of event *b*, the outcome of event *a* might affect the probabilities of event *b* happening in the same period.

Processes are therefore set in an (ultimately arbitrary) order, under the assumption that earlier processes affect, through changing the state of the individual, later processes within the same period, but not vice versa. Typically, this order starts with the processes underlying the demographic events, and grouped in a demographic module, followed by the labour market processes. Last come those simulations that are a function of events, but that do not simulate any events themselves; this includes the simulation of

8 SILC Quality Report, https://statbel.fgov.be/nl/themas/huishoudens/armoede-en-levensomstandigheden/armoederisico/plus

earnings, social security benefits, net incomes, and the like. The modules of the Nowcasting model and order in which they are executed are presented in Graph 3.

Graph 3 Order of modules in the Nowcasting model



Before turning to the more detailed discussion of the various modules of the Nowcasting model, one fundamental issue pertaining to the alignment procedure needs to be highlighted. Many of these alignment tables that the model automatically adheres to, are expressed in incidence rates, or levels. Thus, these projections may dictate certain proportions of various labour market and social security states, e.g. the unemployment rate. However, in the first simulation year the unemployment rate specified by the alignment table may not coincide with the observed unemployment rate in the SILC. Discrepancies can be due to sample error, another definition in the questionnaire, or the lack of any definition, and respondent error. Imposing the unemployment rate as observed in the alignment tables would lead to an unrealistic jump in the projection results. In order to prevent this, and using the implicit assumption that any differences in definition will affect the incidence level but not its growth rate over time, all of the alignment tables that describe incidence levels are transformed in growth rates and then brought to the level observed in the starting data set.

4.2.1. Demographic module

This module consists of the processes underlying birth, survival or mortality, education, immigration/emigration and household formation and dissolution.

The first processes that are simulated in any typical period, i.e. the processes that are the earliest in the recursive simulation, describe immigration and emigration. Immigration and emigration are modelled separately, because the age-gender profile of immigrants and emigrants is very different. It is technically straightforward to simulate individuals to migrate 'on their own': one just makes a clone of a selected existing immigrant to represent the new immigrant. But this approach produces biased simulations of household size and characteristics, because many individuals – especially children – enter or leave the country not alone, but as part of a household. This technique of individual migration is not satisfying since many poverty and social inclusion EU2020 indicators are based on household characteristics and income. We therefore follow the "donor approach" (Duleep and Dowha, 2008), which involves using the pageant algorithm (Chénard, 2000) to sample from the dataset of individuals grouped in households, while aligning to historical immigration statistics and official projections on the level of individuals (Vandresse et al., 2018), and up to 2020. These are available by age, gender and nationality (Belgian/other). Put differently, given that we know how many individuals of various gender-age combinations will enter the country in any prospective year, then the model will select donor households in the sample to meet this required number of individuals, and clone them to represent new immigrants. In the case of emigration, the "donor sample" is the sample available in the model at any future point in time. In the case of immigration, the donor sample for people of non-Belgian nationality is the observed dataset of immigrants prior to simulation. This cloning procedure of individuals subject to alignment has been developed in the framework of the model building toolbox LIAM II and a technical description can be found in Dekkers (2015b).

The second process describes the event of death. Given auxiliary demographic mortality rates to age and gender, each individual faces a certain risk of dying. If this is the case, then the spouse of the individual becomes widowed, the individual is removed and the links between this individual and the remaining individuals are broken.

Next follows the process of giving birth. Given auxiliary demographic fertility rates to age and period, a surviving woman has a probability of giving birth. If this is the case, then a zero-year old individual is created and linked to the mother and, by assumption, her partner.

Finally, various processes describe household formation and dissolution. These are aligned to corrected 2017 LIPRO projections (Vandresse et al., 2018). The household dissolution processes include divorce (in case of married partners), separation (in case of unmarried cohabiting partners) and a process of young adults leaving the nest – somewhat normatively called the "get a life" process. In the case of divorce or separation, the events are simulated for the female partner. These simulations are based on aligned logistic regressions, whereby the probability depends on covariates including the presence of children, the duration of the relation, whether the woman and her partner are working, and – in the case of divorce – the age differential between the partners. If divorce or separation indeed comes to pass, then the male partner moves out of the existing households and gets a household of his own.

The "get-a-life" routine pertains to those young working individuals who are between the ages of 20 and 29, not in a marriage or cohabitation, and who still live with their parents. Unless they find a partner and move out to live with their partner, some of them will be randomly selected to move out of their parents' household and live alone. This random selection is subject to alignment to reflect the input data. Therefore, only a part of all the young adults who fall under the "get-a-life" condition will be selected to move out.

4.2.2. Labour market module

The labour market block of Nowcasting essentially consists of a number of modules that fall in two groups: the first are those that simulate labour market transitions, while the second is the earnings module. The structure of these processes is in broad lines comparable to the long-run administrative data-based model MIDAS Belgium (Microsimulation for the Development of Adequacy and Sustainability, see Dekkers *et al.*, 2010), but the logistic regressions used to simulate the individual risk patterns obviously differ.

Labour market transitions

The probability of remaining in work while one was in work in the previous period is essentially an aligned random process. However, civil servants who are not eligible to retirement remain in their current state by default. The probability of entering work while not being in work in the previous period is also subject to alignment, but here the a priori selection probabilities (i.e. the individual "risks") are simulated through separate logistic regression models for men and women. Individuals who are 50 years and over and who are out of employment (that is unemployment with company allowance, unemployed older worker, disabled, or homemaker) are assumed to remain out of the working state.

Thus, all individuals who work have a relative "risk" of remaining in work, while all individuals who do not work, and who are not in one of the so-called absorbing states, have a risk of entering the working state. Next, all these eligible individuals are ranked using these a priori selection risks, and the actual number of people selected to be or remain in work while being not in work is simulated through an alignment process, using the same auxiliary data from the semi-aggregate model MALTESE as is used in MIDAS. See Dekkers *et al.*, 2015, for an extensive discussion.

Next comes the simulation of the events of remaining in -or moving into- the states within the category of workers. As is the case with MIDAS, the Nowcasting model discerns between four types of workers: those in the private sector, in the public sector (but not as civil servants), civil servants, and self-employed. Like the working states previously explained, those are all simulated through a process of alignment by sorting, where the individual selection risk (separated by gender and previously occupied socio-economic state) is simulated through a range of logistic regression models. The only exception is the selection probability for being a civil servant, given that one works in the public sector. This is basically a random process with a higher probability for those who were civil servants in the previous period.

See Appendix 4 and specifically part section 2 for a more detailed discussion.

Earnings

The second part of the labour market block of Nowcasting is the earnings module that simulates income from work and weekly working hours. For wage earners and civil servants, the simulations are based on estimations. The hourly wage (in logarithm) and weekly hours worked (in logarithm) are simulated separately for full-time and part-time workers. The procedure is implemented in two steps. First, we simulate who is working full-/part-time using the estimated parameters from a logit regression. The parameters are obtained by estimating the probability of working part-time among workers in the longitudinal dataset. Second, we determine the hourly wage and weekly hours worked. The wage and hours are simulated using the estimated parameters from a simultaneous equations model. More precisely, given that both variables are likely to be endogenous (income depends on hours worked and vice-versa), they are estimated simultaneously. We use a simultaneous equations model (estimated separately for men and women), where hours appear as an independent variable in the wage equation and vice-versa.

For self-employed people, we first simulate who is working full-/part-time using the same principle as for wage earners and civil servants. Then, the hourly income and weekly hours worked are determined by randomly assigning a class of income and hours using observed frequencies. This procedure is performed for four sub-groups separately: men working part-time, men working full-time, women working part-time and women working full-time. Individuals who remain self-employed keep the full-/part-time status and hours from the previous period but their hourly income is simulated.[9]

See section 3 of Appendix 4 for a more detailed discussion of how earnings, hours and self-employed income are simulated.

4.2.3. Pensions and benefits module

This module includes the unemployment benefits, disability benefits, unemployment with company supplement benefits, retirement and survival pensions, income guaranty for elderly, social assistance and family allowances. In general, all individuals observed with a benefit in the starting dataset keep the same (uprated) amount if they stay in that particular state. For those who enter a state in which they are entitled to a social security benefit, the benefit is simulated based on the available information and the parameters in effect in the simulation year. This is the case for unemployment benefits, unemployment with company allowance, and disability benefits. The exception are the retirement and survival pensions. To properly simulate those, we should require a lot of retrospective information on the individuals in the starting dataset: their full previous career, including earnings, whether the minimum right is applied, and other information. Nothing of this is at our disposal, so we cannot simulate the pension benefit as done in MIDAS. Hence, we estimate pensions of new beneficiaries based on regression models. This is discussed in detail in section 4 of Appendix 4 to this report. The estimated pension benefits are then uprated so that all pension benefits (observed and simulated) follow the average observed growth rate of pension benefits.

People aged 65 or older with insufficient means of living are eligible for the income guarantee for elderly (IGO/GRAPA). This benefit is means-tested using the income of all adult household members, but excluding the incomes of first-degree descendants (children) as well as ascendants (parents). The eligibility for IGO/GRAPA is tested when the potential claimant reaches the age of 65, and takes into account the income of the claimant and his/her spouse. Older people over 65 who receive the IGO/GRAPA in the starting data keep this benefit during the entire simulation period. People aged below 65 with insufficient means of living are generally eligible for social assistance. These benefits are means-tested at the family level within the household. We take into account the income of the claimant and her spouse,

[9] The observed data reveal that part-time indicator and hours worked seem to be constant if the individual remains in self-employment, while the variation in self-employment income is more random between two consecutive years.

as well as their first-degree ascendants and descendants by blood or alliance (e.g. stepchildren) who live in the same household. In contrast to the income guarantee for the elderly, however, not all persons selected by the means-test receive the benefit, otherwise the number of beneficiaries would be overestimated. This overestimation is due to 1) non-take-up and 2) possibly incomplete information on incomes in the data. To take this into account, the final simulated share of social assistance beneficiaries is aligned to the one observed in 2014 in EU-SILC, with priority given to persons who were already receiving the benefit in 2014.

These social security benefits, as well as earnings, taxes and other sources of household income are core components of the equivalised disposable income that is presented in section 4.2.5. This latter is the income concept underlying key EU2020 indicators, such as income poverty and inequality. It essentially is the disposable total net income of the household, which is corrected for differences in the size and composition of the household, using the so-called modified OECD equivalence scale (see appendix A1 for more details).[10] Note that some sources of income are not simulated but rather kept constant within the model, because we lack the information or theoretical underpinning to simulate them. This among other things includes individual pension benefits from the 2nd and 3rd pension pillars, company cars and education-related allowances. On the household level, it includes net received rents, housing allowances (hy070g), regular inter-household cash transfers received. A more detailed list of simulated variables and variables that are kept constant is presented in appendix A2. The appendix also reports names of variables in EU-SILC used to compute the equivalised disposable income (hy020) and variables used to create input data for Nowcasting.

4.2.4. Multi-states module

Nowcasting is an annual discrete time model with potential transitions between different socio-economic states assumed occurring once a year. This seemingly strong assumption is not too far from reality as according to the EU-SILC data, most individuals have no transitions at all during the year (91% in 2014) and those who have, generally change their state only once a year (more than 2/3 of individuals with at least 1 transition in 2014). It is further assumed that an individual has no other sources of income other than the one that corresponds to the simulated state.[11] We allow the transitions to take place at any month of the year. This is done by simulating two durations (in months) for each simulated period for all persons. The first duration represents the number of months an individual remains in the state occupied in the previous year (=first part of the current year), while the second duration gives the number of months spent in the current state (=second part of the current year). We start by simulating the first duration (in the state from the previous period) based on the observed frequencies from the starting dataset for individuals who in fact experienced a transition.[12] An exception is made for entry into retirement, which is based on the date of birth. The second duration is obtained by subtracting the first duration from 12.

---

10 This modified OECD scale assigns a value of 1 to the first household member aged 14 and older, 0.5 to each additional member aged 14 and older and 0.3 to each child aged under 14.

11 There is an exception for survival pension and social assistance or 'life wage', which are not linked to any specific state and can be combined with other sources of income.

12 More precisely, we use the information in EU-SILC on the monthly reported socio-economic states from the income reference period.

Once both durations are simulated, the incomes from work and social security benefits are adapted proportionally to the durations in the states pertaining to these incomes and benefits. This means that at most one change of state per individual and per year are possible, with the resulting 2 different sources of income at most. The model thus rules out the possibility of secondary incomes received simultaneously with those related to the socio-economic state.

4.2.5. Contributions and taxation module

The next step is to derive net amounts from the gross earnings and social security benefits. This module allows to simulate the after-final-tax personal incomes (income from work, benefits and pensions) by following the rules and parameters of the actual annual taxation system. However, the resulting simulated net amounts are on average much lower than the net amounts observed in EU-SILC leading to "over-taxation" in Nowcasting (see the validation report by Dekkers and Tarantchenko, 2018). This is due to several reasons.

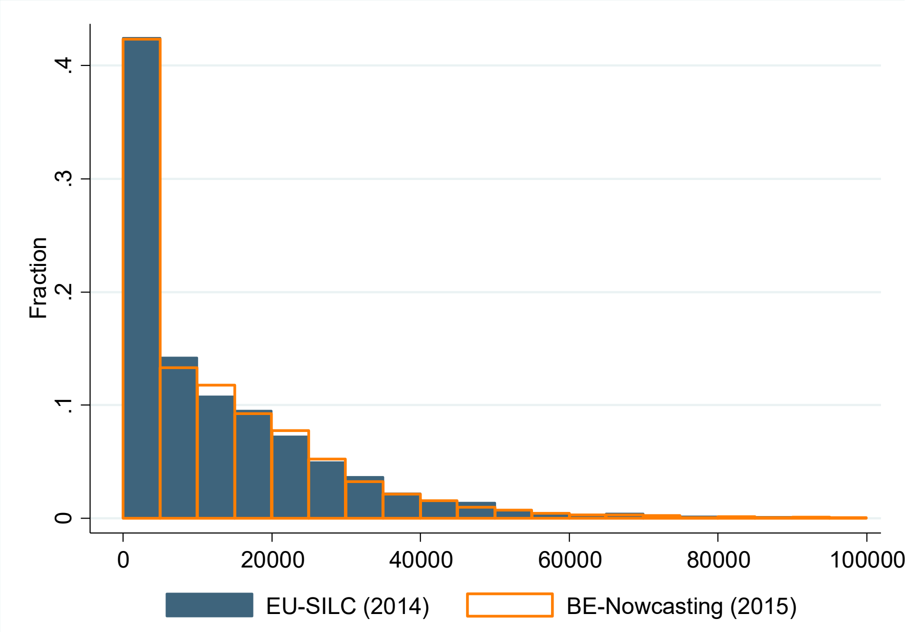
First, the observed net amounts in EU-SILC are those received by the individuals in the course of the year and are thus subject to the withholding tax ("bedrijfsvoorheffing/précompte professionnel") only, in contrast to the simulated final tax. Second, we do not simulate tax evasion nor ad-hoc tax reductions, for example that result from gifts or 2nd pillar pension savings. Third, given that EU-SILC is survey data, it is possible that some amounts are misreported. It is also possible that people have better knowledge of their net than of their gross income. Fourth and finally, the taxation module of the Nowcasting model does not consider some income sources and expenditures (e.g. real estate, real professional expenses, received rents, etc.).

To resolve these issues and improve the fit between the simulated and observed net incomes and taxes paid, we implemented an additional module that adjusts the simulated net incomes to cover for the difference between the final tax and the withholding tax. Adding this adjustment to the "after-final-tax net results" of the model would result in a proxy of the "after withholding-tax net results" that would be more consistent with the observed EU-SILC data.

Moreover, in addition to the adjustment component, the model simulates a proxy of the tax settlement to mimic the reported tax settlement that people receive or pay in the reference income year but that is related to the tax declaration from one year earlier (thus, for the incomes received 2 years prior to the income reference period). This amount has to be simulated, since the tax settlement is included in the EU-SILC definition of disposable income. See section 5 of Appendix 4 for a more detailed discussion of these issues.

The results of the fiscal module are displayed in Graph 4, which compares the distribution of taxes and contributions paid that are observed in the input data (EU-SILC 2014) and simulated in the first simulation period (BE-Nowcasting 2015). The differences are small and not systematic.

Graph 4 **Distribution of tax on income and social security contributions paid by households in EU-SILC and BE-Nowcasting**
in Euro



4.2.6. Material deprivation module

The severe material deprivation indicator SMD is an indicator that reflects the inability to afford a number of items that are required to lead an adequate life. This indicator measures the part of the population that cannot afford at least four out of the following nine items:

- to pay their rent, mortgage or utility bills;
- to keep their home adequately warm;
- to face unexpected expenses;
- to eat meat or proteins regularly;
- to go on holiday;
- a television set;
- a washing machine;
- a car;
- a telephone.

The Nowcasting model only projects household composition, labour market states, hours of work and incomes including earnings. It does not include consumption nor any of the information required to simulate whether households obtain or loose these items, and it is therefore impossible for the model to project SMD directly. We therefore opted for a somewhat experimental indirect approach where the information available (including on the VLWI and AROP) is used to 'statistically explain' the occurrence of SMD. Only variables that the Nowcasting model projects for future years can be included, which excludes a number of potentially important variables, e.g. housing tenure. The regression model is not supposed to reflect causal effects.[13] The main concern is to project the proportion of households in SMD, while conserving the correlations between AROP, VLWI and SMD, in order to be able to produce a valid projection of the overarching indicator AROPE. A lower correlation between the components of AROPE would imply less overlap, and therefore the projections would result in a larger proportion of the population in AROPE than would otherwise be the case.

Since the deprivation questions are asked only once for each household, and the SMD is a household level variable, the model is estimated on the household level, using household-level variables[14]. In order to maximize the number of observations at our disposal, the cross-sectional EU-SILC data for Belgium between 2004 and 2014 were pooled. Since the Very Low Work Intensity indicator VLWI is only calculated for the population of 60 and younger, a separate regression is run for the households where the head is 60 years or older, and VLWI, as well as the number of children, is excluded as a covariate in that model. The years of data collection are included as a linear variable, as well as in squared and cubic form, to allow for non-linear trends over time. The results are presented in Appendix 6. Since the estimates serve only to feed the nowcasting and projection model, and have little substantive interest, we refrain from a discussion of these results. In any case, due to the large number of inter-action effects, they are difficult to interpret.

Given projections of the AROP, the VLWI and the SMD, we have the three indicators that together determine the overarching at-risk-of-poverty-or-social-exclusion (AROPE) indicator. This is the proportion of individuals that live in households that are either at risk of poverty (AROP), or – if the individual is younger than 60 – have low work intensity (VLWI), or that are in a situation of severe material deprivation (SMD).

4.2.7. Output module

The output of Nowcasting is essentially a panel dataset with one observation per individual and per simulation year. Thus, the dataset is exported in the form of (one or more) CSV files which can then be imported in a statistical program like Stata to produce aggregate indicators such as the at-risk-of-poverty rate (AROP) the very low work intensity (VLWI) and the severe material deprivation (SMD) rate.

[13] For a thorough analysis of the determinants of material deprivation using EU-SILC see Guio et al. (2012). See also Dekkers (2008); who proposes a dynamic approach where the occurrence of financial poverty at any point is used to explain the hazard of falling into material deprivation at a later point in time.

[14] The VLWI will have different values for members of the same household in cases where a 60+ person lives together with younger people. In those cases we use the value of the VLWI of the oldest member as representative of that of the household.

4.3. Alignment

As with MIDAS, the current model used for nowcasting and short-term projections makes extensive use of alignment to i) be consistent with projected or observed trends in labour markets and social-security incidence rates, ii) to follow projected growth rates of average earnings and pension benefits and iii) include parameters of the various social security systems and the tax system. These are the "channels of consistency" (Dekkers et al., 2012; see Dekkers et al., 2015 for a more detailed discussion) between the microsimulation model and auxiliary information, in this case observed information for the years 2015 to 2017 and short-term and middle-term projections and/or hypotheses for the years 2018 to 2020.

The source of most of the projected auxiliary information are macro-economic projections by the Federal Planning Bureau. The macroeconomic projections consist of various steps, from the short to the long run, and each step takes the information from the previous step into account. The short-term projections are produced with the model MODTRIM three times per year, and the first of these are published usually early February. These make the projections to the final quarter of that year. The middle-term projections are produced with the model HERMES and are published early June, and have a simulation horizon of 6 years. The HERMES projections are calibrated for the short-run to the MODTRIM projections of February. The long-run socio-demographic and budgetary projections, specifically on the budgetary impact of aging, are produced by the model system MALTESE and usually published late June. These are calibrated to HERMES projections for the middle-term (and so indirectly to MODTRIM projections for the short run). We refer below to MALTESE, as this is the immediate source of the alignment tables that are used in the Nowcasting model. However, it should be kept in mind that for the years for the simulation years of Nowcasting, the models HERMES and MODTRIM are the original sources of the macro-economic projections to which the simulations are aligned.

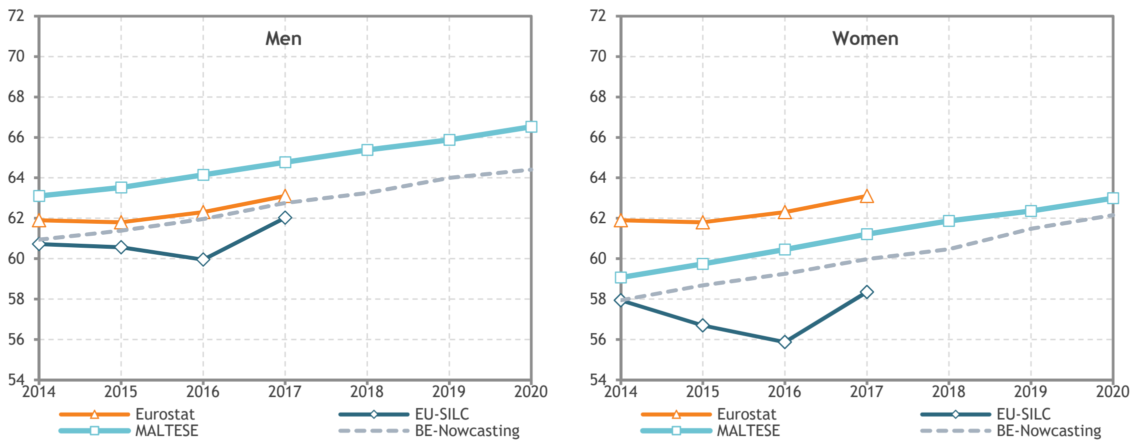
4.3.1. The first channel of consistency: state alignment

The Nowcasting model uses alignment through sorting to replicate incidence rates of labour market states (working, working in the private and public sector, working as self-employed) and social security beneficiary states (unemployed, disabled, unemployed with company allowance, retired). These are observed (up to 2017) and thereafter projected by age and gender by the semi-aggregate model MALTESE/HERMES. Furthermore, the model uses given fertility and mortality rates, combined with alignment techniques, to model changes in civil state (between married, cohabiting and single). In addition, immigration and emigration are simulated through alignment of number of persons that are "translated" to the household level (see Dekkers, 2015b, for a more detailed discussion). The observed and projected demographic figures are available for each age-gender combination, and stem from the FPB demographic individuals and households' projections (Vandresse et al., 2018).

However, contrary to the alignment techniques in the MIDAS model based on administrative data, the survey data EU-SILC may cause important differences in the incidence levels between the exogenous alignment tables and the observed proportions. These differences would result in very important changes in the first year of simulation, to "jump" from the latter to the former. This will greatly affect simulation results. One way to prevent that from happening is by proportionally uprate the alignment tables to the incidence rates observed in the starting dataset. Here we make the implicit assumption that whatever causes the difference between the observed incidence levels and the alignment proportions, does not affect the *intertemporal change* of the latter.

Appendix A3 to this report summarizes the information used in the various alignment processes of the Nowcasting model. As an illustration, the below Graph 5 presents the employment rate of the Nowcasting model and compares it with the observed LFS employment rate (denoted "Eurostat"), the MALTESE employment rate and, finally, the observed employment rate in the EU-SILC data. It illustrates the principles behind alignment. In Graph 5, the MALTESE employment rates are represented by the light-blue square line. As can be seen from the graph, the employment rate in EU-SILC in 2014, our starting year, is lower than the one reported for MALTESE. At that point in time, EU-SILC (dark-blue diamond line) and Nowcasting (short-dashed grey line) employment rates coincide as they are both based on the observed data. To take the discrepancy between EU-SILC and MALTESE rates into account, the latter are proportionally adjusted within the model to the incidence rates observed in the starting dataset. This adjustment is assumed constant throughout all the simulated years. Starting from 2015, the resulting nowcasted employment rates are projected by age and gender using the adjusted employment rates from MALTESE. They are very close to the Eurostat LFS-based employment rates in 2015-2017 for men, but somewhat low for women. The development of the EU-SILC employment rate is more erratic due to i) the relatively low number of observations, ii) LFS is conducted quarterly, while the EU-SILC survey frequency is once a year, iii) definitions differ considerably, with LFS following ILO guidelines, while the labour market state in EU-SILC is self-assessed, and iv) the minimum age of an employed person in EU-SILC is 16.

Graph 5 Employment rate in Eurostat, EU-SILC, MALTESE and BE-Nowcasting for ages 15-64
in %



Notes: Nowcasted estimates are aligned to corrected MALTESE projections (correction is operated in 2014 and assumed constant for the following years).
Source: Eurostat database, LFS annual series (lfsa_ergan) Own calculations based on EU-SILC microdata (2014-2017 waves). FPB, MALTESE projections.

In general, the fit between the nowcasted and EU-SILC employment rates is not perfect, especially for women. However as intended, the model captures the dynamics in the MALTESE employment rates very closely, and even those from Eurostat starting 2015. We will show later in Graph 23 and footnote 19 that the fit of the simulation results is considerably better, especially for women, when the comparison is limited to the working age population (i.e. ages 18-59, excluding students of age 18-24).

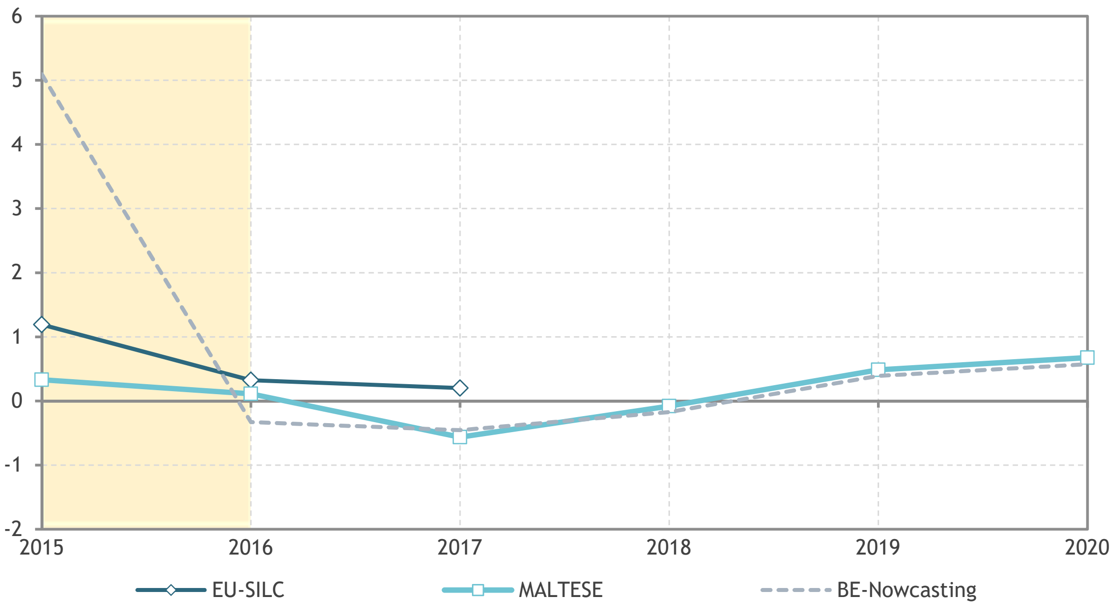
4.3.2. The second channel of consistency: alignment of the growth rates of incomes from work

The growth rate of observed and simulated income from work follows the observed and short-term projected growth rates of incomes in each of the employment states (private sector employees, public sector employees excluding civil servants, civil servants and self-employed). Uprating in this case is done separately by gender.

Graph 6 shows the growth rates of all the incomes from work in EU-SILC, Nowcasting and MALTESE. Because of the alignment to MALTESE, the almost perfect fit between the Nowcasting results and the MALTESE results from 2016 on is not a coincidence. However, the results for 2015 do reflect some of the choices that have been made in the modelling process, and more specifically the removal of all secondary revenues or earnings.

First of all, the Nowcasting model follows the EU-SILC structure, with incomes pertaining to the previous year. Thus, in 2015 we actually compare the growth rate between incomes from 2013 and 2014. The reported high rate is due to removed additional and secondary incomes and earnings in 2014. For example, if somebody who has been an employee for a full year has for whatever reason also revenues from independent work, then those are removed. As these secondary amounts are often (very) low, this drives up the average incomes per person and results in a high growth rate. In the year 2016 (comparing incomes between 2015 and 2014), the simulated growth rate is lower than that from the auxiliary source. This is because there are more transitions from and into employment in 2015, when the labour market module kicks in, than in 2014. This results in lower annual earnings for 2015. After 2016, the situation stabilises and the two growth rates, MALTESE and Nowcasting, are a near-perfect match by design.

Graph 6 Real growth rate of average gross employment income in EU-SILC, MALTESE and BE-Nowcasting
in %



Notes: All averages are obtained using strictly positive gross incomes from work. Nowcasted estimates are aligned to MALTESE projections. Shaded area represents transition from EU-SILC to the first simulated year where all secondary incomes and earnings are removed.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves). FPB, MALTESE projections.

4.3.3. The third channel of consistency: alignment of social security benefits

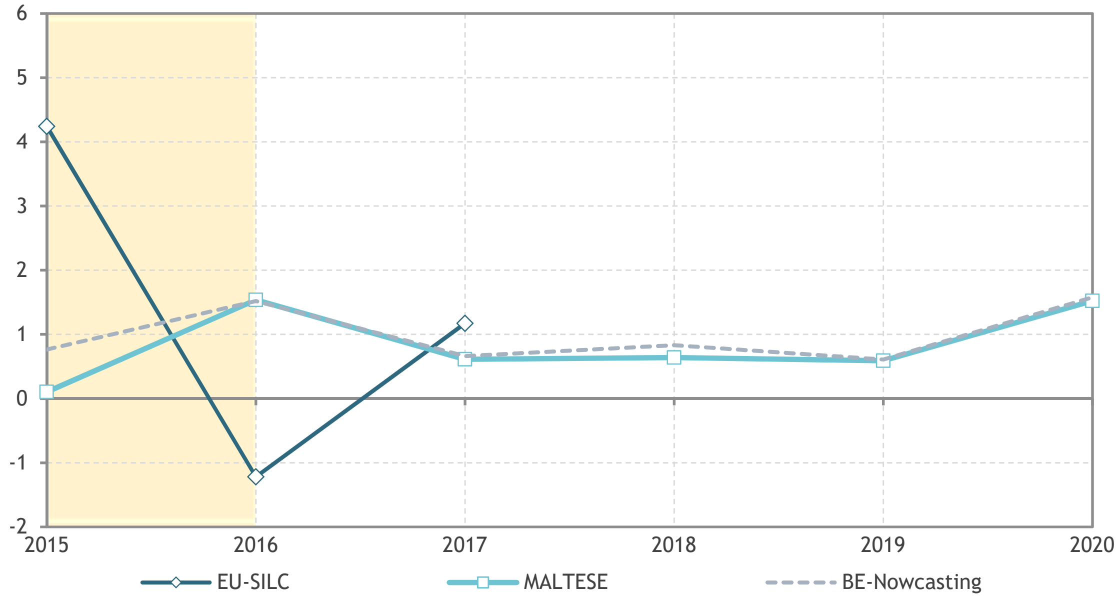
In general, social security benefits are not aligned: if an individual becomes eligible for a benefit at time t then the benefit will be the result of (projected) parameters in combination with the simulated previous earnings. The parameters are minimum benefits, floors and ceilings in the fiscal and various social security benefits that reflect observed changes and short-term projected values. Once an individual is in a state where he or she receives a certain benefit, then the benefit is the one-period uprated value of the benefit in the previous period[15]. The uprating or real-term indexation of these amounts reflects the hypotheses of the Study Committee for Ageing (High Council of Finances, 2018, 14). The exception are the retirement and survival pensions. First, the on-going pension benefits (i.e. benefits paid out to those who are observed in retirement or are widowed in the starting dataset or who were retired or widowed in the previous year) are increased in real terms by 0.5% per year, according to the hypotheses[16]. Then, newly simulated benefits (i.e. for those moving into retirement or widowhood) are adapted as to bring the growth rate of all pension benefits to the observed or assumed real growth rates from MALTESE. Put differently, in the simulation of pension benefits at the moment of retirement (or the loss of the spouse) the alignment process kicks in and changes the average retirement and survival pensions for the new beneficiaries. But once the benefit is set, it does not change anymore, save for the yearly 0.5% increases. Any alternative method where all benefits would be corrected to achieve the target growth rate would result in important in- and decreases of pensions of those already retired, would be unrealistic.

Graph 7 shows the growth rates of average retirement and survival pension benefits in the Nowcasting model and compares them to the MALTESE growth rates and the actual growth rates in the EU-SILC. As in the case of the alignment of incomes from work, the growth rate of simulated pensions is close to that of the auxiliary source except for 2015. The difference in 2015 illustrates the limits of the selected approach, as it puts the full impact of the alignment process on the new pensioners only. That is, the growth rates of pension benefits might be heavily affected by the sometimes low number of individuals who go into retirement or become widowed. If the number of new pensioners or surviving spouses is low, then the growth rate correction required to meet the MALTESE growth rates can be too much to be realistic. As a result, the correction will not be implemented. This is exactly what is going on in 2015. The target growth rate of pension benefits is close to zero, while on-going pension benefits are uprated with a positive rate. Thus, to reach the desired growth rate, the pension benefits of new retirees or surviving partners should be low enough to "compensate" the positive growth rate of the on-going pension benefits. However, as the number of new beneficiaries is too low (the number of new surviving partners is actually zero), the correction would be so strong that the level of the simulated benefits would be negative. Hence the alignment was not implemented, and the growth rate of total pensions remains higher than the MALTESE target. This is why, like in the previous graph, the results for 2015 are not comparable to the projections thereafter.

[15] The unemployment benefit is an exception to this rule as it decreases with the duration in unemployment.

[16] These hypotheses also specify that minimum pensions are uprated by 1% per year. However, in the EU-SILC data it is not possible to identify pensioners who benefit from a minimum pension.

Graph 7 Real growth rate of average gross retirement and survival pensions in EU-SILC, MALTESE and BE-Nowcasting
in %



Notes: All averages are obtained using strictly positive gross pensions. Nowcasted estimates are aligned to MALTESE projections. Shaded area represents transition from EU-SILC to the first simulated year where all secondary incomes and earnings are removed.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves). FPB, MALTESE projections.

In short, these four channels of alignment make the Nowcasting model (as) consistent (as possible) with observed trends and short-term projections by the FPB.

5. Results

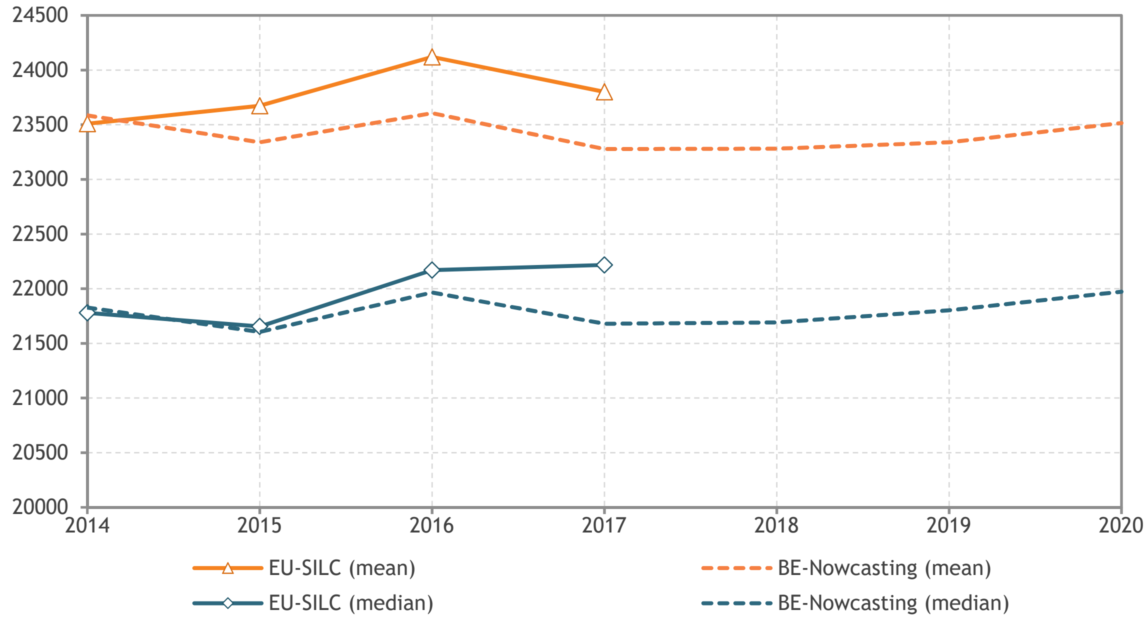
In this chapter we present the simulation results. First, we show the projections of mean and median equivalent income, since the AROP and the inequality indicators are based on this income concept. In the second section, we discuss the main indicator of income adequacy, the at-risk-of-poverty rate. The third section is dedicated to the inequality indicators. The very low work intensity (VLWI) indicator and the severe material deprivation (SMD) rate are the subject of the fourth and fifth sections, respectively. Section five also presents the projection of the overarching indicator AROPE (at-risk-of-poverty-and-social-exclusion rate). In section six we compare our results with the Eurostat flash estimates and EUROMOD nowcasts. Finally, section seven shows that the Nowcasting model can also be used to perform simulations of proposals for policy changes. All results are expressed in 2014 prices.

5.1. Mean and median equivalent household income

The at-risk-of-poverty rate (AROP), as well as the indicators of inequality and redistribution rely on the concept of the equivalised disposable income, so we discuss the simulation results for this concept first. Equivalent disposable income is constructed by adding up all incomes and earnings for all members of the household. These summed incomes are then corrected for differences in the size and composition of the household by dividing them by an equivalence scale, in this case the modified OECD scale.[17] Thus, the mean and median equivalent incomes presented in Graph 8 and Graph 9 are affected by possible simulation errors in incomes, labour market states, size and composition of the households (including -of course- immigration and emigration) – in short, in the entire model. This is also why it is not easy to establish what causes discrepancies between the observed and simulated results. All in all, however, the fit of the results is quite good. Both the mean and the median are a bit too low, which might also be the results of the Nowcasting model not simulating “secondary” incomes and benefits, but they follow the observed developments quite closely and the development in projection appears reasonable. The fit and dynamics of the simulated median with the observed one is especially close until 2016. In 2017 however, the nowcasted and observed medians go in the opposite directions, which widens the gap between the two sources.

[17] Note that Nowcasting again follows the EU-SILC structure by using the previous calendar year as income reference period, with age calculated at the end of the income reference period.

Graph 8 Mean and median equivalised disposable income in EU-SILC and BE-Nowcasting
in Euro

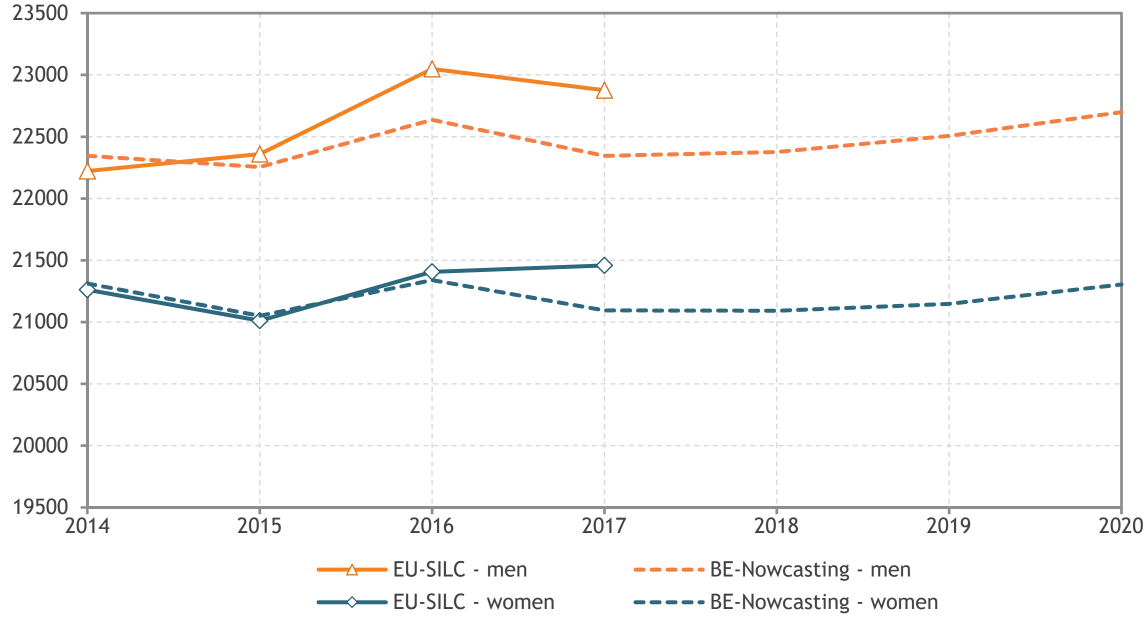


Notes: Numbers are based on incomes expressed in 2014 prices.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Note that the fit is better for median equivalent incomes than for the mean. This suggests that the differences are the result of issues in the extreme ends of the distribution (where the number of individuals and households is lower).

The results presented in Graph 9 separate the medians by gender. Although the fit is less good for men than for women, the overall conclusion remains that the dynamics are close to the observed EU-SILC. The discrepancies between EU-SILC and Nowcasting in 2014 are due to slightly different techniques used in the computation of the total disposable household income. In EU-SILC, it is based on the single variable representing the total disposable income of a household that is available in the data. In Nowcasting, the total disposable household income is calculated by us, using the information in EU-SILC and the Belgian questionnaire. Moreover and most importantly, the weighting might differ due to adaptations required by the Nowcasting model. In the model, the weights are not directly simulated but rather the initial dataset is first expanded using the observed weight (see section 4.1), resulting in rounding errors, which might affect the statistics. The simulations (demographic and labour market) are then performed on the expanded dataset.

Graph 9 Median equivalised disposable income in EU-SILC and BE-Nowcasting, by gender
in Euro



Notes: Numbers are based on incomes expressed in 2014 prices.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

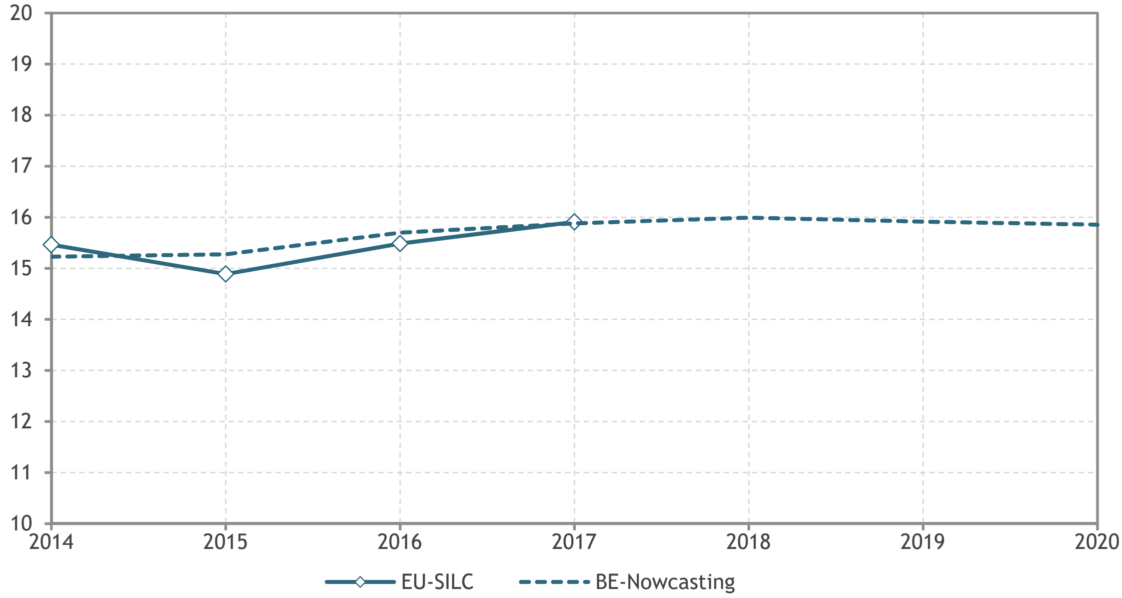
5.2. The at-risk-of-poverty rate

Next, we turn to the main indicator projected by the Nowcasting model, being the at-risk-of-poverty rate (henceforth AROP). This indicator is derived by establishing a threshold value, in this case equal to 60% of the overall median equivalent income. Each individual who lives in a household where the equivalent income is lower than this threshold, is considered to have a higher poverty risk because of her lower income compared to the others. Thus, the simulation results of the AROP are affected by the simulation errors in the overall median (shown in Graph 8), because that affects the poverty threshold. However, beyond that, the AROP is affected by discrepancies in the distribution of equivalent income around the threshold between and within groups (such as gender). Thus, especially for smaller subgroups, the simulation results can differ from the observed results, and it is often very difficult to identify what causes these differences.

The first Graph 10 shows the overall AROP incidence rate or the overall poverty risk. This is the first of the three social indicators for the EU2020 "Poverty and Social Exclusion target".[18] The fit of the model is very good, and the development of the poverty risk seems reasonable, at least starting 2015. The evolution of the simulated indicator between 2014 and 2015 differs from EU-SILC. A slight increase is observed for the simulated AROP, while EU-SILC AROP decreases.

[18] See Eurostat (2018).

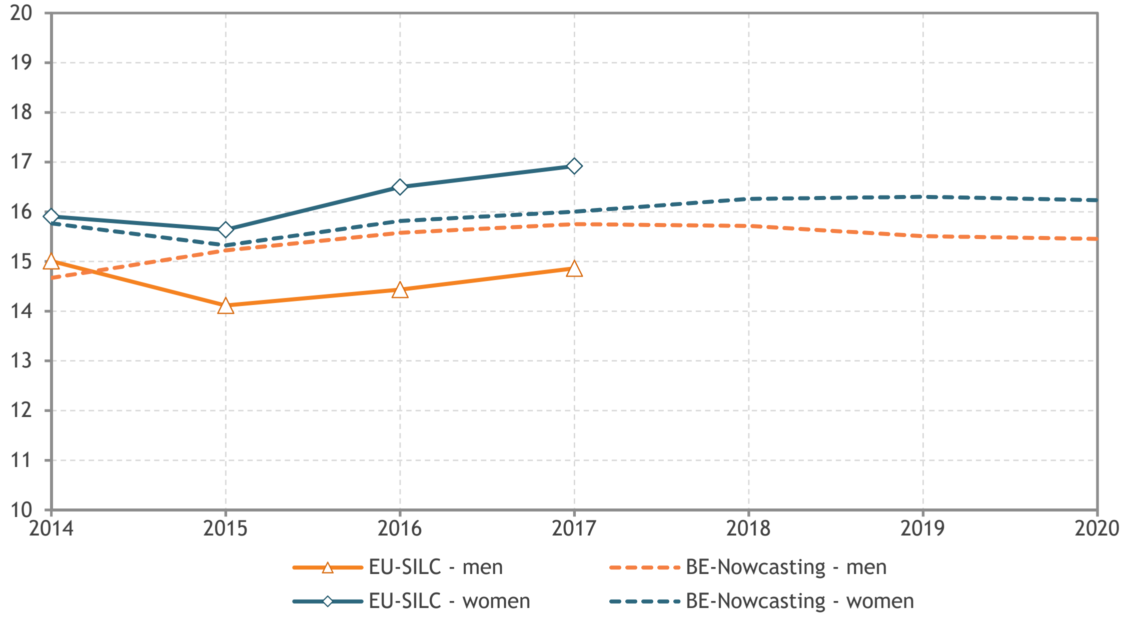
Graph 10 At-risk-of-poverty rates in EU-SILC and BE-Nowcasting (threshold: 60% of median)
in %



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 11 shows the poverty risks by gender. The difference in the dynamics of the overall AROP between Nowcasting and EU-SILC in 2014 and 2015 is driven by the AROP of men. The changes go in the opposite direction compared to the median equivalent income of men displayed in Graph 7. Although, there is no a direct link between both indicators, the median shows that the share of men with lower income increases in Nowcasting, and so does their probability of being below the poverty threshold. The opposite is true for EU-SILC, where their lower probability of being below the poverty threshold in 2015 is further decreased by the drop in the poverty threshold computed from the overall median equivalent income. Starting 2015, the AROP of men in Nowcasting is higher than in EU-SILC but both increase until 2017. For women, there are also clear differences in the AROP level between EU-SILC and Nowcasting but the changes in the indicator go in the same direction. For both sources, the AROP first decreases between 2014 and 2015 and then increases until 2017.

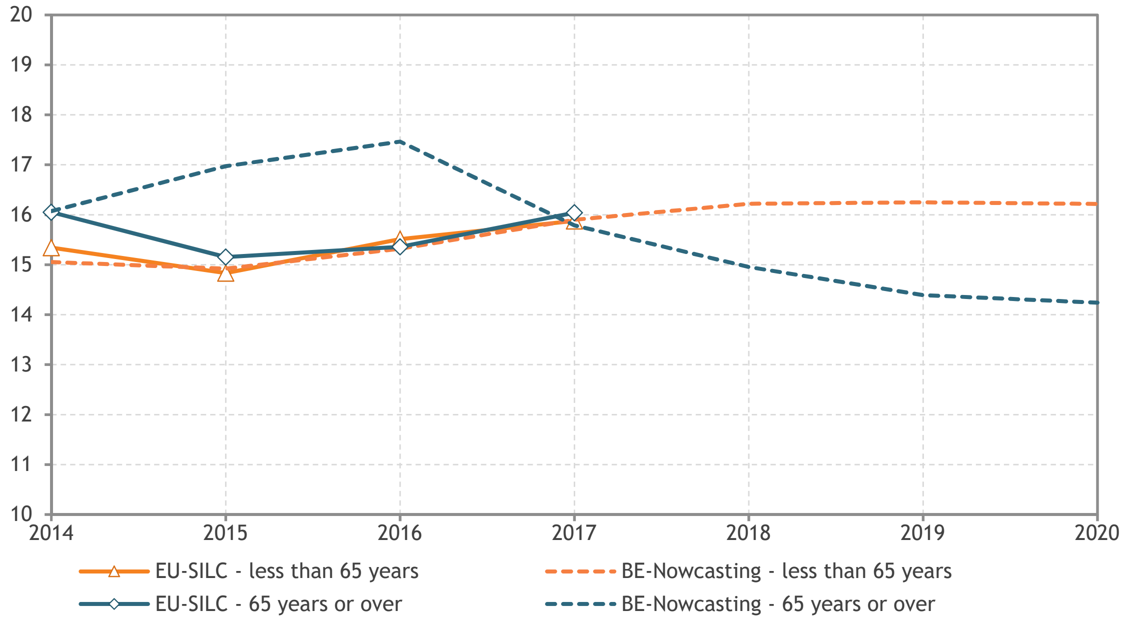
Graph 11 At-risk-of-poverty rates in EU-SILC and BE-Nowcasting, by gender (threshold: 60% of median)
in %



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Next, we turn to Graph 12 which presents the observed and simulated poverty risks among the elderly (aged 65 or more) and non-elderly (aged less than 65).

Graph 12 At-risk-of-poverty rates in EU-SILC and BE-Nowcasting, by age (threshold: 60% of median)
in %

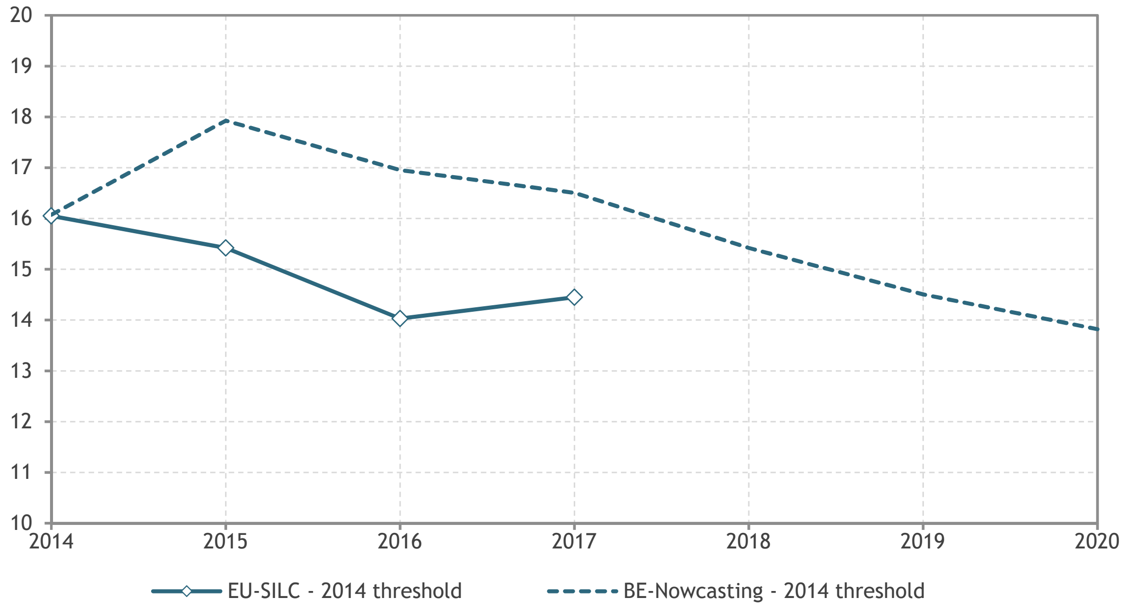


Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

The agreement between Nowcasting and EU-SILC regarding the poverty risk of non-elderly is very good, but this is not the case for the elderly. The simulated poverty risk among the 65+ exceeds considerably the actual poverty risk in 2015 and 2016. This is due to a constant increase in the simulated AROP between 2014 and 2016, while the actual AROP first decreases in 2015 and then increases in 2016. After

2016 the simulated AROP for the 65+ displays a downward trend, first dropping considerably to reach a level slightly below the observed one in 2017 and then keeping a steadier pace until 2020. The difference in the simulated and observed trends might be due to the fluctuations in the at-risk-of-poverty threshold across observation years, that may be partly due to sampling error. Graph 13 shows what happens if we maintain a constant poverty threshold set at 60% of the median equivalised income observed in 2014. In that case, there is still a sharp increase in the simulated AROP for the 65+ in 2015, while the observed AROP decreases. However, starting in 2016, the difference between simulated and observed AROP remains more or less constant. Note that with a constant poverty threshold, the drop in the simulated AROP observed in Graph 12 happens one year earlier, i.e. in the first simulation year. In light of this evidence, it seems that the differences between the observed and simulated data for 65+ originate from the transition from 2014 to 2015.

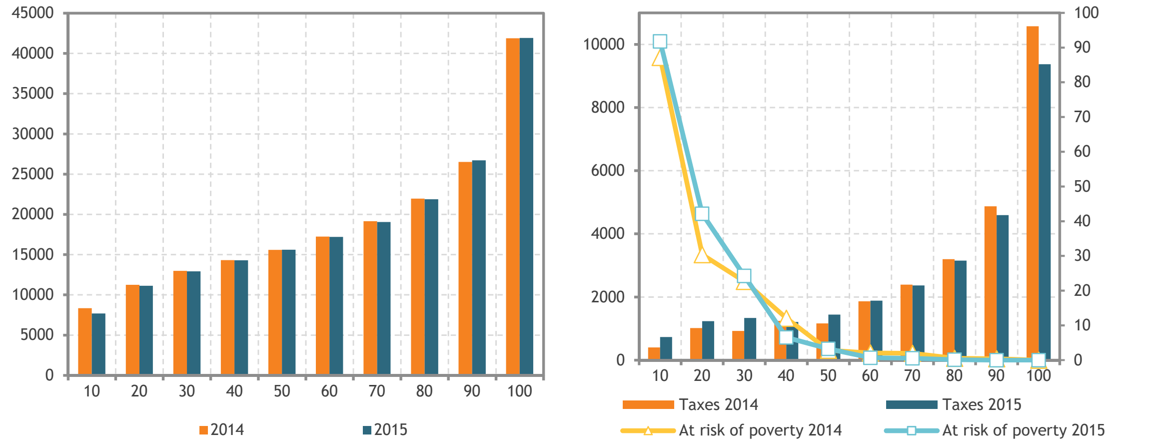
Graph 13 At-risk-of-poverty rates in EU-SILC and BE-Nowcasting for 65 years or over (threshold: 60% of median in 2014)
in %



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 14 analyses the distribution of gross household income and taxes paid by the households for the 65+ in 2014 and 2015 in the simulation data. The distribution is described by income deciles that are formed based on the gross household income divided by the household size. Each household member thus has the same per capita gross income. After the income deciles are formed, we compute the average gross per capita income and per capita taxes (total taxes paid by a household divided by household size) within each decile. The left panel of the graph shows that, except for the first (lowest) and second deciles, the average gross per capita income changes little between 2014 and 2015. In contrast, the average per capita taxes (left axis of the right panel) are higher in 2015 for the first five deciles where the poverty risk is the highest (right axis of the right panel). The right panel also shows that persons in the richest deciles (9th and 10th) see their taxes decrease in 2015. It seems that Nowcasting overestimates taxes of the poorest and underestimates those of the richest as compared to the observed data, which results in an increase in the simulated AROP rate in 2015. After this initial adjustment between 2014 and 2015, the situation stabilises.

Graph 14 Gross household income (left panel, in Euro) and taxes (left axis of right panel, in Euro) and AROP rate (right axis of right panel, in %) by income decile in 2014 and 2015, 65 years or over (BE-Nowcasting)



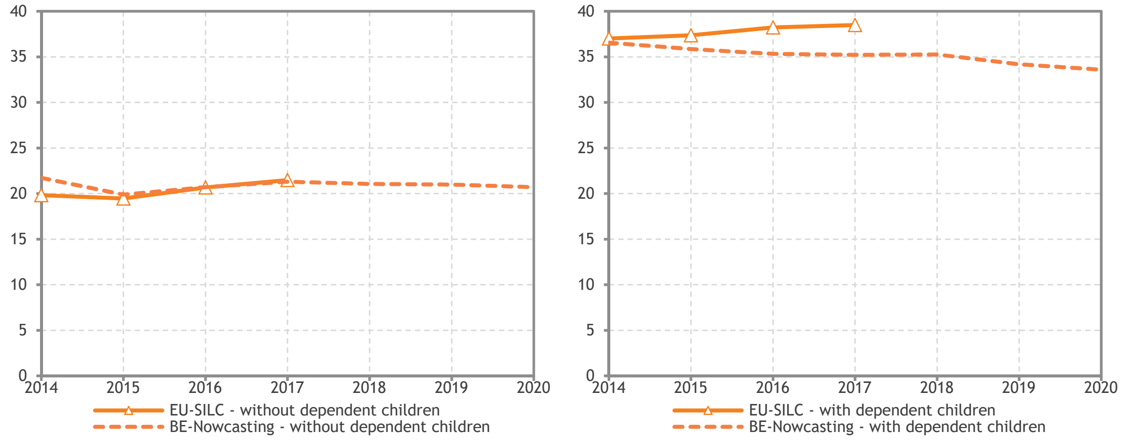
Note: Income deciles are determined based on the gross household income per household member (i.e. gross household income divided by household size, no equivalence scale used).

The decrease in the gross equivalent income between 2014 and 2015 in the first two deciles is due to two particularities of the Nowcasting model: the presence of new surviving partners without income and the removal of secondary incomes. In reality as well as in the Nowcasting model, those who become a surviving partner in any year receive a surviving partner pension benefit. However, since the Nowcasting model follows the EU-SILC structure, this benefit only becomes visible one year later. Inversely stated, assume a married woman who has no income of her own but who is not poor because of the income of her partner. Furthermore, assume that her partner dies near the beginning of 2015, and that she starts to receive a survivor's pension benefit. In 2015 her deceased husband's income received during 2014 will not be included in her household income (as he is no longer there), and the survivor's benefit will only appear in 2016, resulting in zero household income in 2015 for the survivor. Hence, she will be in the first decile depicted in the left panel of Graph 14, which will result in lower average gross equivalent income. Furthermore, as said earlier in this report, all secondary incomes have been removed in the model starting 2015, which is not the case in 2014 as it is mostly based on the observed data. Hence, for example, those pensioners that have additional incomes will see those removed and their gross equivalent income will be driven down.

The decrease of the poverty risk among the 65+ from 2016 on is by itself not unexpected and is in line with earlier projections over the longer run using MIDAS (see Dekkers et al., 2015).

Graph 15 and Graph 16 display the poverty risks by household type. This is interesting because the household type is not directly related to income, labour market state or poverty rate in the Nowcasting model. Any relation between the two is therefore indirect, for example via labour market participation, via earnings differentials or selective mating. The Graph 15 shows the poverty risk among single adults without (left panel) and with (right panel) dependent children. A significant proportion in the group without children are single pensioners. Starting off somewhat too high in 2014 – a result which is explained in the beginning of the section – the simulated poverty risk quickly converges to the observed one and then follows a comparable development with a good fit. According to the simulations, the poverty risk would gradually decline over time. The right panel shows a remarkable difference between the development of the actual and simulated at-risk-of-poverty rates for single adults that live with dependent children. Starting at a similar level, the simulated poverty risk of single parents declines, whereas the observed poverty risk increases. It requires a more detailed analysis to assess the cause of this divergence.

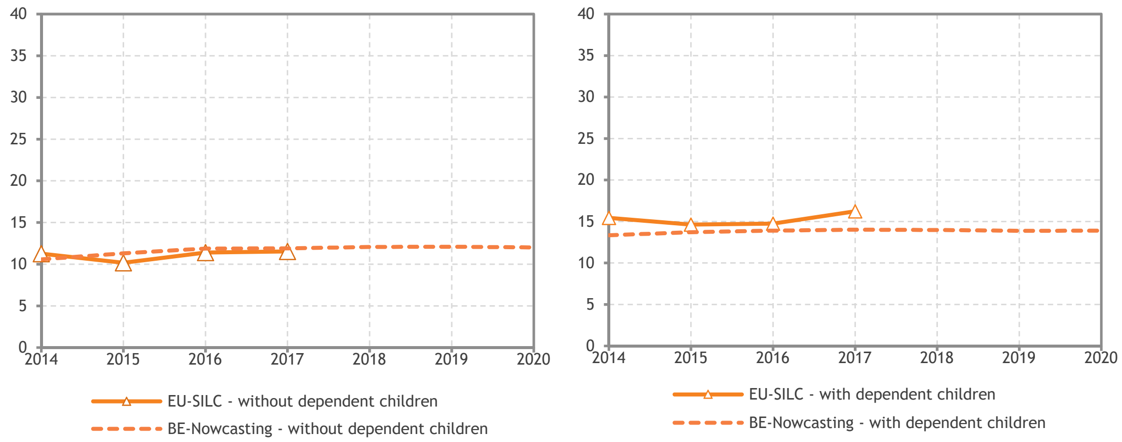
Graph 15 AROP among adults living in single-adult households without and with dependent children



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

The results presented in Graph 16 show that the fit of the poverty risk among adults living in a household with at least one other adult, but without children (left panel), is good in level as well as in development. Finally, the right panel of the same graph shows that the poverty risk of adults who live in a household with at least one other adult and one or more dependent children is somewhat underestimated by the model.

Graph 16 AROP among adults living in multiple-adult households without and with dependent children



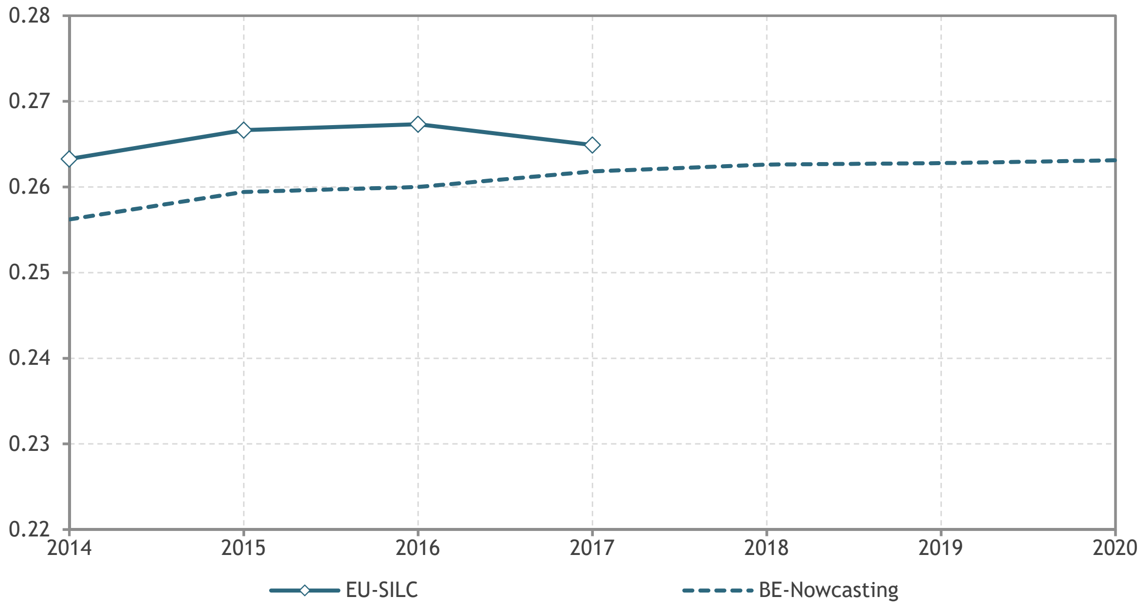
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

5.3. Inequality

The indicator used here, as well as in the EU system of social indicators, to measure inequality is the Gini coefficient. A Gini coefficient of zero expresses perfect equality, where everyone has the same income. A Gini coefficient of 1 represents maximum inequality, where one person in the population has all income. We also consider the distribution of equivalent disposable income by percentiles, to explain differences between the simulated and observed trends in the Gini coefficient.

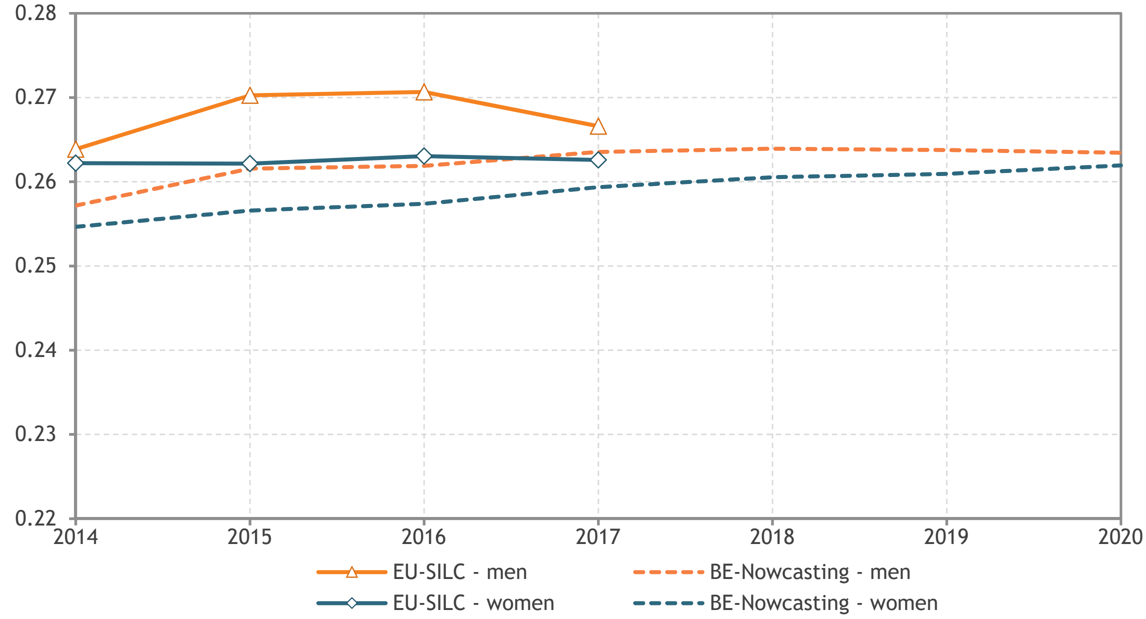
Graph 17 shows that the model underestimates somewhat the overall inequality reflected by the Gini. Graph 18 indicates that this is the case for both men and women. Clearly the model underestimates the variation of incomes, which could be the result of the (omitted) secondary incomes and their relationship with observed incomes. However, in both graphs the overall trends seem reasonable and in line with the observed data.

Graph 17 Gini coefficients



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 18 Gini coefficients, by gender

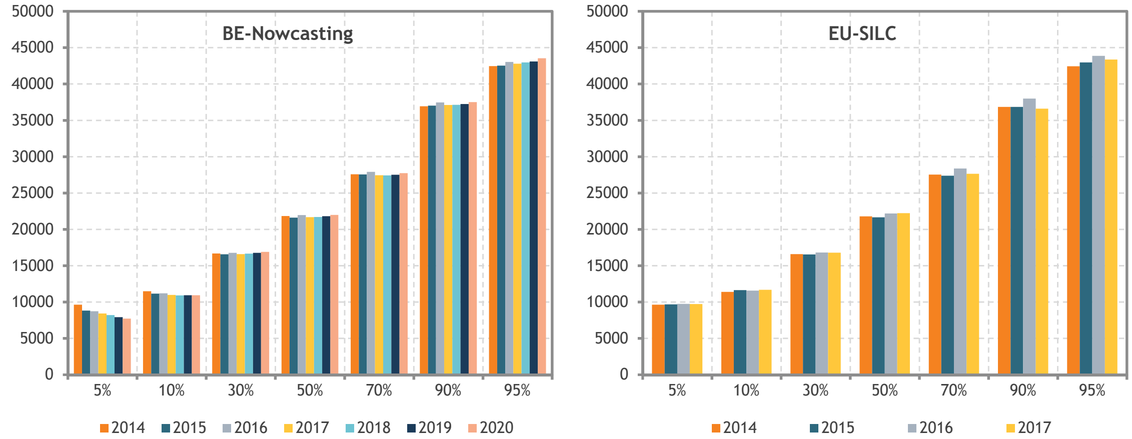


Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Disaggregating the income distribution by percentiles, as is done in Graph 19, makes a more detailed comparison possible. This comparison reveals a mixed picture. The nowcasting model appears to display the same patterns as the observed EU-SILC data does in the middle and upper percentiles. The average equivalent income of the 95 percentile increases both in the model and in the EU-SILC until 2016. The results for 2016 stand out in both cases, which is also the case for the lower deciles from the 30th decile on. However, the Nowcasting model shows a decline in the equivalent income of the 5 and 10 percentiles between 2014 and 2017, while it increases in the observed SILC data. The 5 percentile continues to decline until 2020 in Nowcasting. It appears that Nowcasting underestimates the incomes of the lowest-income households and that this underestimation increases with time. This comes mostly from younger generations as shown in Graph 20 and Graph 21.

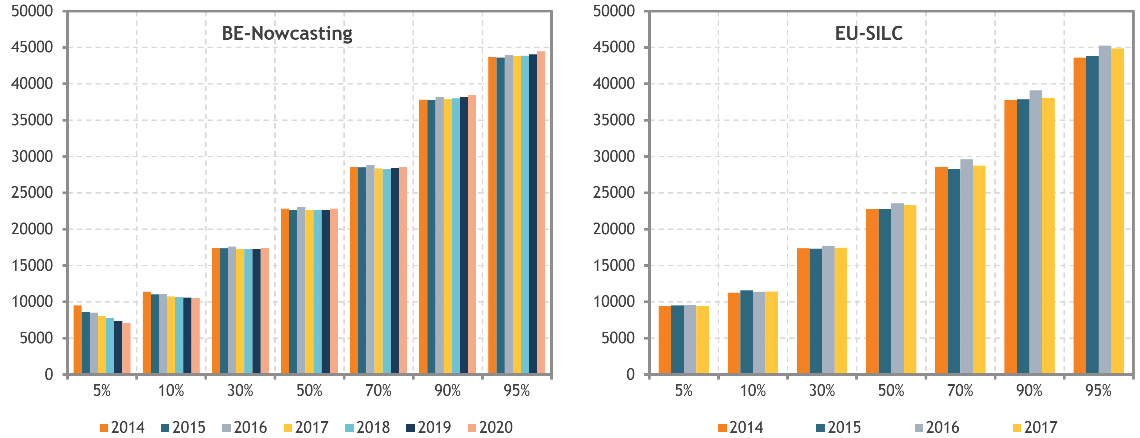
Like in the previous graphs, the comparison between Nowcasting and EU-SILC in Graph 21 shows that the Nowcasting model captures the broad trends in inequality among 65+ fairly well. The average equivalent income of most percentiles increases, but this trend is less outspoken for the higher percentile (95 and more).

Graph 19 Equivalised disposable income in BE-Nowcasting and EU-SILC by percentiles
in Euro



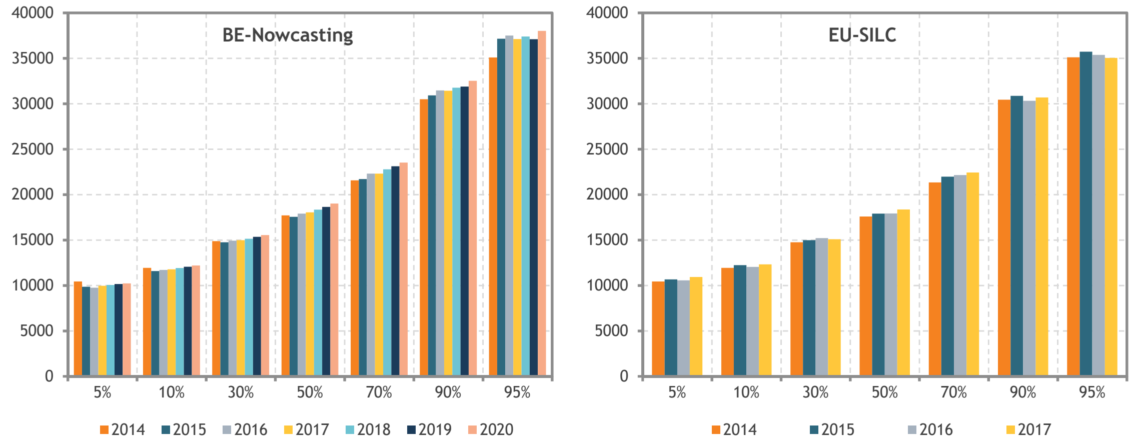
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 20 Equivalised disposable income in BE-Nowcasting and EU-SILC for less than 65 years, by percentiles
in Euro



Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 21 Equivalised disposable income in BE-Nowcasting and EU-SILC for 65 years or over, by percentiles
in Euro



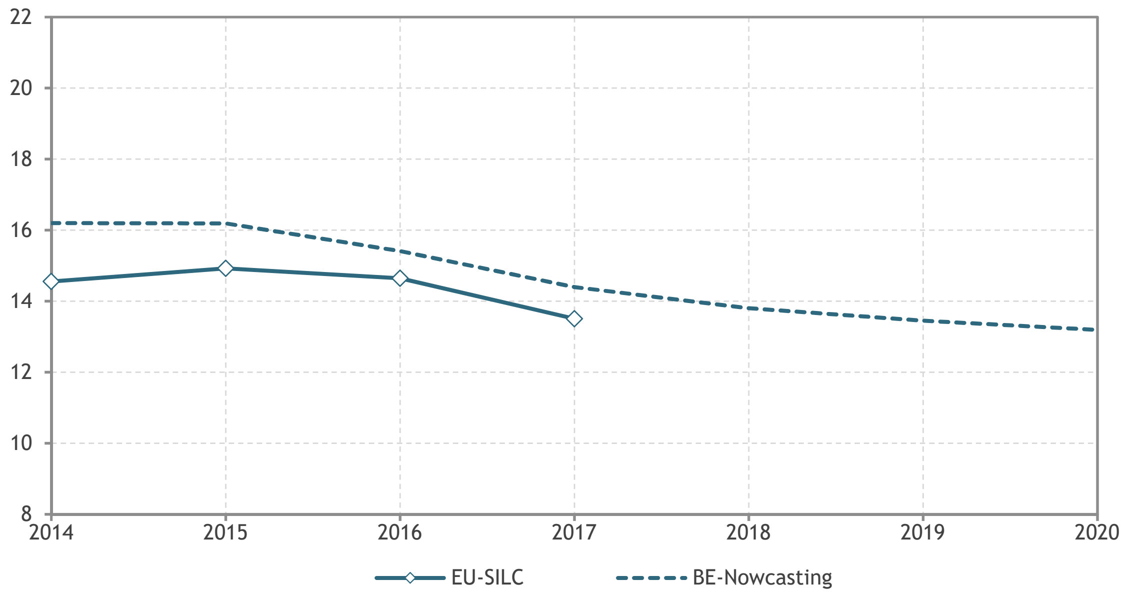
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

5.4. Very low work intensity

We turn now to the very low work intensity indicator (henceforth VLWI), which is defined as "People aged 0-59, living in households, where working-age adults (ages 18-59, excluding students aged 18-24) work less than 20% of their total work potential during the past year" (Eurostat, 2018). Whether someone works depends on an alignment process in combination with an estimated behavioural equation while the hours worked are simulated without alignment and based on a reduced form estimated model. Furthermore, the VLWI status depends on the household composition. Results are therefore sensitive to possible mismatches in the simulation of household formation and dissolution in the demographic module.

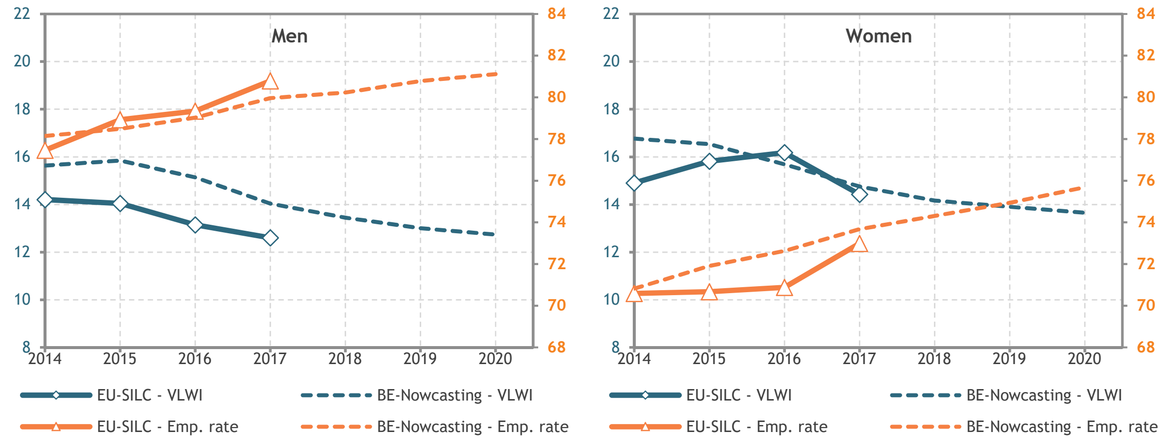
Graph 22 shows that the difference between the actual and simulated VLWI is limited and fairly constant over time, so that the simulated VLWI shows the same development as its observed counterpart.

Graph 22 People living in a household with very low work intensity
in %



Notes: EU-SILC numbers are derived following Stata do-file by Goedemé and Montaigne, 2010. Similar strategy is used to obtain nowcasted numbers.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

Graph 23 People living in a household with very low work intensity (VLWI, left axis) and employment rate for working age adults (Emp. rate, right axis), by gender
in %



Notes: EU-SILC numbers are derived following Stata do-file by Goedemé and Montaigne, 2010. A Similar strategy is used to obtain nowcasted numbers.
Source: Own calculations based on EU-SILC microdata (2014-2017 waves).

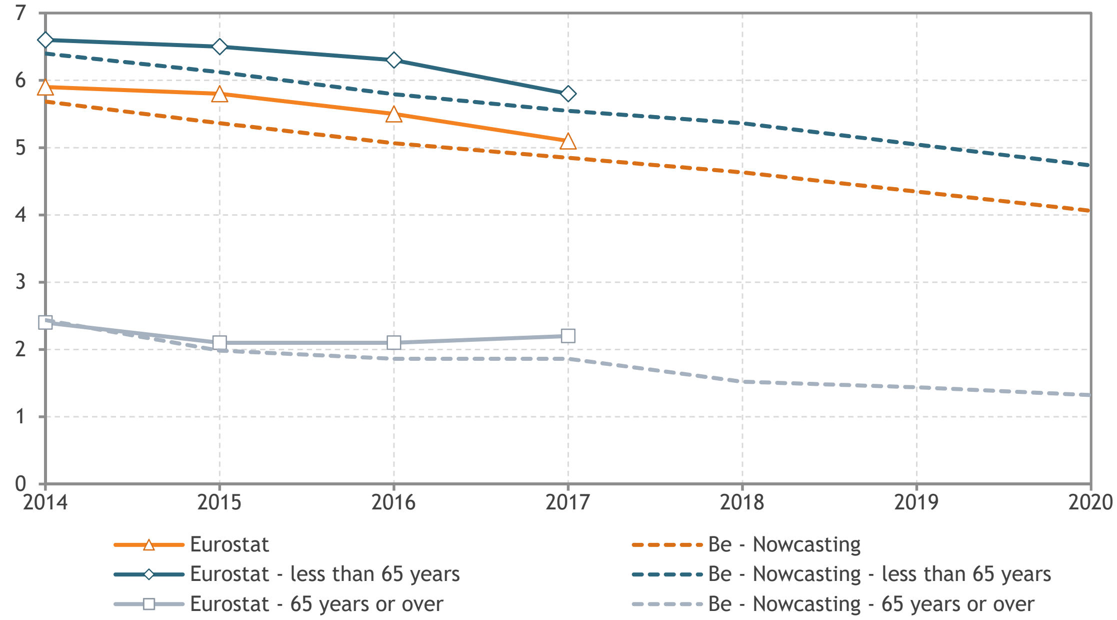
As mentioned, the VLWI depends on the proportion of people who work as well as the hours that they work compared to their work potential. The above Graph 23 compares the development of the VLWI of men and women to the employment rate of men and women of working age, respectively[19]. As expected, this graph suggests that the VLWI is heavily affected by the employment rate, and therefore by the MALTESE alignment tables that the Nowcasting model uses. If the employment rate among the working-age population increases, then this is accompanied with a decrease of the VLWI.

5.5. The severe material deprivation rate and the at-risk-of-poverty-and-social-exclusion rate

To simulate the severe material deprivation rate (SMD) in the Nowcasting model, we use the estimated parameters presented in section 4.2.6 and the simulated AROP and VLWI indicators and other relevant simulated household variables. The resulting household indicator for SMD is then brought to the level of the individual in the production of output. The results are presented in Graph 24.

[19] Note that the employment rates in Graph 5 and this Graph 23 differ in the group that they pertain to. The former pertains to the entire group aged between 15 and 65 whereas the latter pertains to the same category that is used for the calculation of the VLWI. These "working age adults" are those aged 18-24 (excluding students) and 25-59. Note that the fit of the simulated employment rates in Graph 23 is considerably better than those in Graph 5 – for men, but especially for women. This suggests that the problems in the fit of the former are concentrated in the 60-64 age group. This is hardly surprising, as the Nowcasting model does not have a retirement decision module as such. Rather, it is a function of the alignment tables of the other states: if an individual is potentially eligible to retirement (i.e. if he or she meets the age and career requirements for early retirement) and if one ceases to be in work and one does not go into unemployment, disability or unemployment with company allowance, then one goes into retirement. Thus, possible simulation errors in the other states together have their inverse in the simulation of those who are in work (or not) between 60 and 65.

Graph 24 Severe Material Deprivation (SMD) rate for total population and by age group: Eurostat data (Eurostat 2018c; 2018d - 2014-2017) and Nowcasting simulation results (2014-2020)
in %

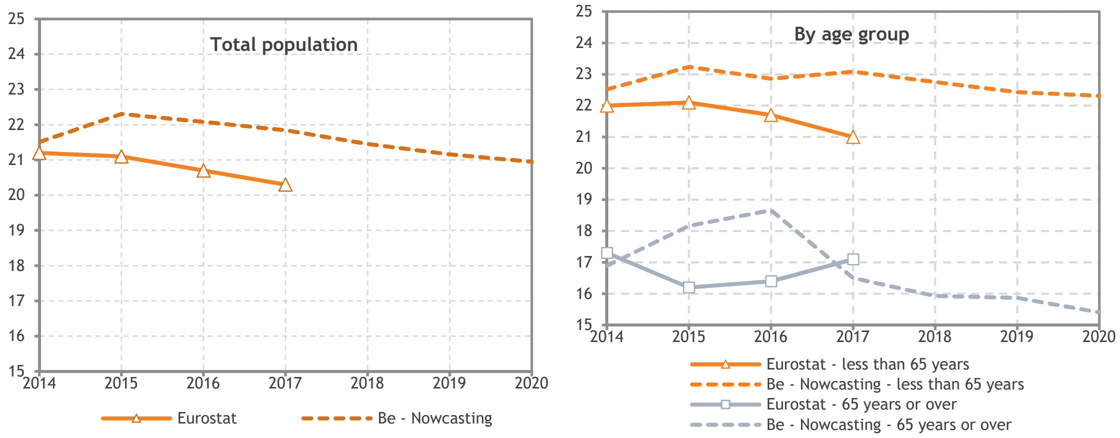


The fit of the simulated SMD is quite good; the difference between the official and simulated indicator for the years 2014-2017 is, on average 0.33 percentage-points, ranging between 0.22 in 2014 and 0.44 in 2015. Both simulated and observed (Eurostat) indicators indicate a downward trend of SMD between 2014 and 2017. This is comparable to the difference between the actual and simulated results for the group younger than 65, which is 0.33 percentage-points on average and ranges between 0.2 percentage-points in 2014 and 0.5 percentage-points in 2016. For the group aged 65 and older, the fit is even better: on average 0.2 percentage-points, ranging between virtually no difference in 2014 and -0.34 percentage-points in 2017. There is a slight difference in the trends, where the stabilisation and minor increase of the SMD rate for the elderly population between 2015 and 2017 is less visible in the simulation results. Overall, the conclusion would be that the current declining trends in the SMD rate would continue for the younger population, whereas the apparent stabilisation for the elderly would come to an end and a further decrease would set in.

Now, we turn to the overarching at-risk-of-poverty-or-social-exclusion (AROPE) indicator (Graph 25). As explained earlier, this is a composite of the other three social indicators: a person is AROPE if he or she is in at least one of the following situations: being at-risk-of-poverty (AROP), living in a household with very low work intensity (VLWI) or severe material deprivation (SMD). The difference between the simulated and observed AROPE rates is higher than that of the SMD, when measured in percentage points. It is 1.1 percentage-points on average, and ranges between 0.3 in 2014 and 1.5 in 2017. Furthermore, the observed AROPE shows a slight decrease from 2014 on, while the simulated version rises throughout 2014 and 2015 and starts to decrease only from 2016 on. Considering the simulation results for AROP (Graph 10), the VLWI (Graph 22) and the SMD (Graph 24), it seems that the overestimation of the AROPE proportion is mainly the results of the fact that the simulated VLWI is in all years higher than the observed VLWI.

The fit of the AROPE simulation for the population younger than 65 is comparable to that of the population as a whole (1.2 percentage-points on average, ranging between 0.5 percentage-points in 2014 and 2.1 percentage-points in 2017). The fit of the AROPE projection for the population of 65 and older is clearly affected by the simulation error in the AROP of that group. The fit is on average good (0.8 percentage-points; ranging between -0.4 percentage-points in 2014 and 2.3 percentage-points in 2016), but there is an obvious issue in the projection of the trend in AROPE of this group: the observed AROPE declines from 17.3 in 2014 to 16.2 in 2015, then remains stable and increases back to 17.1 in 2017. This is not the case for the simulated AROPE for the elderly population, which would increase up to 18.7 percent in 2016, followed by a rather strong decrease in 2017. From 2018 on, there would be a more gradual decline.

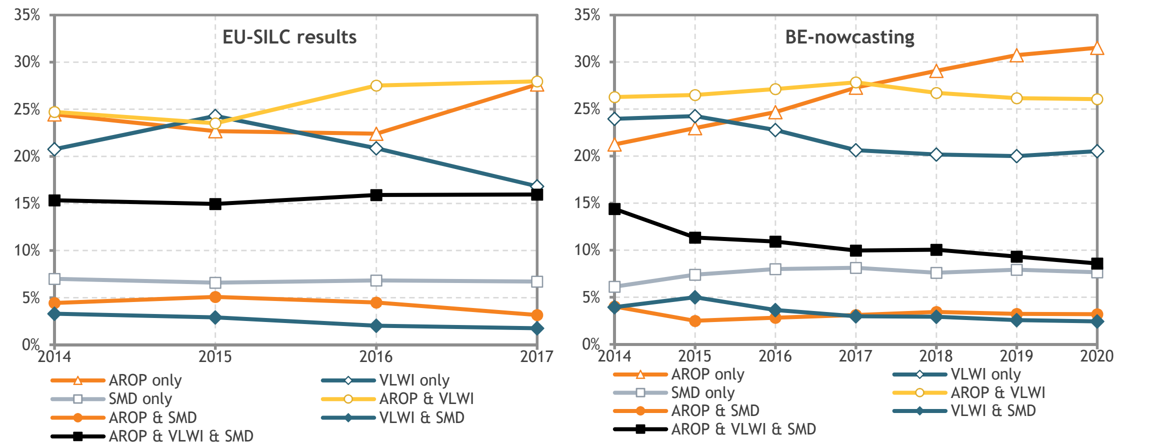
Graph 25 At Risk of Poverty or Social Exclusion (AROPE) rate: Eurostat data (Eurostat 2018d - 2014-2017) and Nowcasting simulation results (2014-2020)
in %



Overall, the conclusion of the comparison between the simulated and observed AROPE indicators is that the level of the simulated AROPE is somewhat too high, while the decrease over time of latter is smaller than that of the observed AROPE. The main reason for the difference in the level is the overestimation of the VLWI by the Nowcasting model (see section 5.4). It is less clear why the trend in AROPE deviates, since the Nowcasting projections of the underlying indicators AROP, VLWI and SMD coincided with, or were parallel to, the observed trends. A possible reason is a decreasing degree of overlap between the underlying indicators: everything else equal, less overlap would still result in a higher AROPE rate. Below we look in some detail at the extent of overlap of the observed and projected indicators.

Graph 26 shows the percentages among the respondents less than 60 years of age in AROPE who score only on VLWI, AROP or SMD, on combinations of two indicators (AROP and VLWI; SMD and VLWI; and AROP and SMD) and, finally, on all three indicators (AROP, VLWI and SMD). We start by discussing the actual overlaps, as observed in the SILC data and shown in the left panel of Graph 26. The degree of overlap is determined, on the one hand, by the rates for the three indicators separately and their evolution (for AROP and VLWI these are around two-thirds of the population in AROPE; for SMD it is around 30%), and on the other hand by the correlations between the indicators. The correlation between AROP and VLWI is increasing somewhat over time from 0.53 in 2014 to 0.58 in 2017; the correlations between SMD and both AROP and VLWI are smaller and constant over time in the range 0.35-0.40[20]. The proportion with only AROP increases slightly over time, from 25% in 2014 to 28% in 2017; the overlap between AROP and VLWI follows a similar trajectory. As a consequence of the increase in that overlap, the proportion with only VLWI declines after 2015. Remarkably, the decline in the proportion in SMD among the population below 60 is accompanied by a drop in the proportions who combine SMD with AROP or VLWI, while the proportion with only SMD remains fairly constant. A virtually constant percentage of people below 60 has an arrear on all three indicators (around 15% of those in AROPE; around 3.5% of the complete population below 60).

Graph 26 Overlap between AROP, VLWI and SMD among persons aged < 60 in AROPE: EU-SILC and BE-Nowcasting results



The Nowcasting simulation results shown in the right panel of Graph 26 are quite similar to the observed proportions, especially in the first years. The trends are somewhat different, though. The proportion with only AROP increases more strongly in the simulation than in observation, while the proportion with only VLWI declines less, and the overlap between the simulations of AROP and VLWI does not increase like it does for the observed indicators. Perhaps most strikingly, the simulated proportion with arrears on all three indicators declines by 5.8 percentage-points (of all people below 60 in AROPE). At the same time, the percentage with an arrear on only one indicator increases from 51% in 2014 to 60% in 2020; such an increase is not observed in the EU-SILC results. The decline in the degree of overlap is part of the reason for the smaller decrease in the simulated AROPE rate, relative to the observed one. Clearly, the Nowcasting model is not able to replicate the correlations between the indicators completely. Such correlations are notoriously difficult to model correctly; they may even be due to variables that are unobserved in the starting data (e.g. parental background).

An important policy conclusion follows from the Nowcasting simulations. Even if the observed AROPE would continue to decline at the simulated speed (though starting from the lower observed level), this decrease would not be strong enough to reach the EU2020 target. This conclusion was also drawn by SPF SS (2018, 15) and Frère (2016, 22), namely that the EU2020 target on the Belgian level for the reduction of the number of people in poverty or social exclusion is unlikely to be met.

[20] Correlations calculated within the whole sample, not just within the group with AROPE.

5.6. Comparison with Eurostat flash estimates, EUROMOD results and a FPB projection

Eurostat publishes (experimental) flash estimates for AROP and a few other income-based indicators, but not for VLWI, nor SMD, and so also not for AROPE (see Chapter 2). In order to emphasize the uncertainty surrounding these estimates, Eurostat communicates these estimates in a peculiar way, using what it calls the Rounded Uncertainty Interval (RUI) (European Commission, 2018a). This gives an indication of the magnitude and the direction of the expected change. It is emphasized that the centre of the RUI does not coincide with the point estimate, though it is close. For Belgium in 2017, the RUI for the change (with respect to 2016) runs from -1.5 to 1.0 percentage-points, and the RUI for the level is 14.4% to 16.9% The midpoints of these intervals are respectively -0.2 and 15.7%. (Note that these figures refer to income year 2017, so the actual observations corresponding to these estimates will be derived from EU-SILC 2018.) The corresponding estimates from Nowcasting are +0.1 and 16.0%[21]. These nowcasting estimates are of course quite close; both indicate stability in the overall AROP.

A more interesting comparison is with the EUROMOD nowcasting results published by the EUROMOD team. Unfortunately, early nowcasting papers did not include estimates for Belgium; the earliest (and also latest) results for Belgium can be found in Gasior and Rastrigina (2017). These nowcasting estimates have the advantage, though, that they are broken down by gender and age category. For Belgium their nowcasting model uses EU-SILC 2012 as base data. Belgium is one of the countries for which nowcasted AROP rates are considered to be relatively accurate by the authors, based on a comparison of EUROMOD nowcasting estimates with EU-SILC observations for the years 2011-2014. Table 1 presents these results and compares them with the ones obtained with our Nowcasting model and with the EU-SILC observations.

Both our model (referred in the table as BE-Nowcasting) and EUROMOD underestimate the increase in median household equivalent income during the period 2014-2016, though EUROMOD more so than Nowcasting[22]. The at-risk-of-poverty threshold of course rises at the same rate as median equivalent income, and this may be the reason why the actual change in the AROP rate for the total population between these years is about twice as large as the nowcasted rates. By contrast to the EUROMOD estimates, Nowcasting is correct in predicting that the poverty risk would increase more for women than for men. An important difference appears with respect to the 65+, compared with the population aged 0-64. For the latter group, the Nowcasting estimates are virtually spot-on, while EUROMOD nowcasts of the increase in AROP are too low. By contrast, the EUROMOD estimate of the rise in the poverty risk of the elderly is somewhat too high, while Nowcasting predicts incorrectly a strong decrease for this group.

21 Cf. Graph 8. Note that in that graph, data-points are labelled by EU-SILC year, not by income year. So 16.0 is the Nowcasting estimate for EU-SILC 2018, income year 2017.

22 For these years, the macro-economic figures which are used both by EUROMOD and Nowcasting are observations, not projections. Therefore, the discrepancy between the nowcasting estimates and the EU-SILC observations may well be due to sampling error. The 95% reliability interval for median income in the Belgian EU-SILC varies across years between 2 percent and 4 percent of the point estimate.

Table 1 Comparison of Nowcasting results with EUROMOD nowcasting estimates, 2015-2017 (1)

		Estimated change EUROMOD	Actual change 2015-17 BE-Nowcasting	Observed level 2017	
Median income (2)		1.9	2.8	5.2	22,783
AROP rates (3)	Total population	0.5	0.6	**1.0**	15.9
	Men	0.5	0.5	**0.8**	14.9
	Women	0.5	0.7	1.3	16.9
	Age 0-17	0.4		1.3	15.0
	Age 18-64	0.4		0.6	18.6
	Age 0-64		1.0	1.1	15.9
	Age 65+	1.1	-1.2	0.8	16.0

(1) The years refer to the EU-SILC data collection years (as in the rest of this report). So, the estimates are in fact for the previous year. By contrast, in EUROMOD publications, the years refer to the income year.

(2) Median household equivalent income; percentage change, in current prices.

(3) Percentage-point change.

EUROMOD has also made nowcasting estimates of the very low work intensity for a total of 12 EU countries, unfortunately not including Belgium, for the period 2009-2013 (Rastrigina et al., 2015). The EUROMOD team concluded that the nowcasted very low work intensity estimates seem to align well with the Eurostat values in most cases.

An earlier study published by the Federal Planning Bureau (Frère, 2016) projected the population at risk of poverty and social exclusion in Belgium until 2030y. The projections were realised using aggregate-level information available up to July 2016 and under the assumption of no change in policy. This study projected a decrease in the AROP rate from 14.9 percent in 2015 to about 13 percent in 2020, which contrasts with the slight increase observed in EU-SILC (and nowcasted by our model) for the period 2015-2017, and the virtually stable AROP rate projected by our model for the years 2017-2020. In addition, Frère (2016) expects a slight drop in the percentage of the population not at risk of poverty but with a low work intensity. As a result, the projected AROPE rate drops from 21 percent in 2015 to about 20 percent in 2017, continuing to about 18 percent in 2020. For the period 2015-2017 this agrees with the EU-SILC observations, while our model projects a smaller decrease[23].

5.7. Policy simulations

In this section we illustrate the possibilities that Nowcasting offers by simulating a simple variant. This is not to be interpreted as real policy proposal, but rather as illustrations.

The scenario concerns a simple increase of the progressivity of the fiscal system. We maintain the income tax brackets, but the marginal tax rates of the lower brackets decrease whereas those of the highest brackets increase. These marginal tax rates are shown in Table 2.

[23] The projection of AROPE in fact implies projections of the VLWI and the SMD, but these are not shown in Frère (2016).

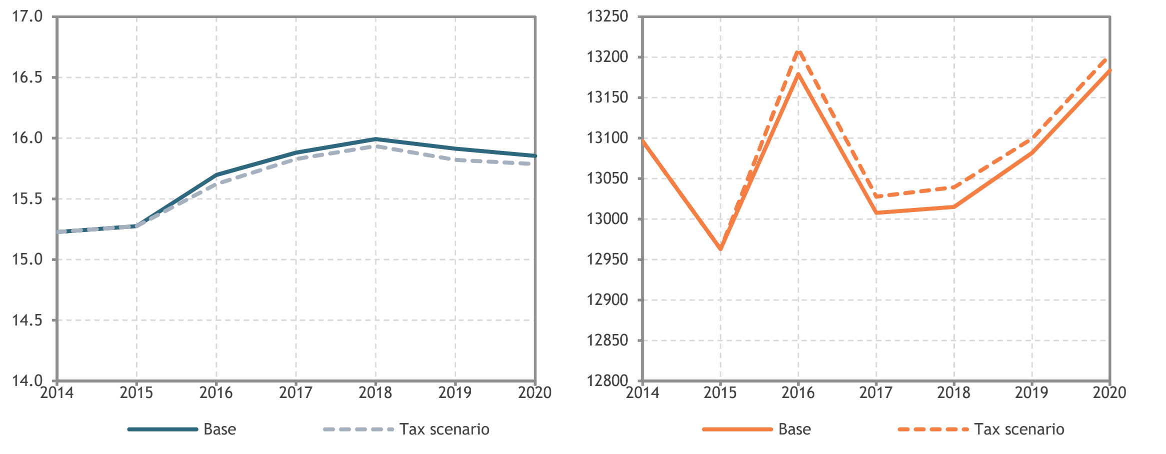
Table 2 Marginal tax rates

Income tax bracket	Base scenario	Tax scenario (2015 and later)
1st bracket	25%	20%
2nd bracket	30%	25%
3rd bracket	40%	40%
4th bracket	45%	48%
5th bracket	50%	50%

The marginal tax rates to the lowest two brackets decrease by 5%-points. The next marginal tax rate remains the same whereas the one-but highest marginal tax rate increases by 3%-point. The tax rate that pertains to the highest tax bracket remains unchanged. Hence, we here have a redistribution from the higher (but not highest) incomes to the lowest incomes, and we can therefore expect the poverty risks to go down. The tax rates in this technical variant have been chosen in such a way that total tax receipts do not change relative to the base scenario.

The results in Graph 27 confirm our expectations. The increasing progressivity of taxes causes the poverty rate of the population as a whole to decrease. Note that, due to the EU-SILC-structure of the simulation results, the impact of this measure implemented in 2015 reveals itself only from 2016 onwards (i.e. 2015 incomes). The impact of this measure is small, yet one should realise that it is the combination of a direct and an indirect effect. Because of the higher net incomes and the low end of the income distribution, the poverty risk increases because fewer people will be in a household where equivalent income is lower than the poverty threshold. However, the variant causes median income and therefore the poverty threshold itself, to go up. This cet. par. increases the poverty risk again, while however increasing the average income of the poor. Hence, the direct effect causes the poverty risk to decrease while the indirect causes it to increase again. The former still outweighs the latter, but the difference is small.

Graph 27 At-risk-of-poverty rate (left panel; threshold: 60% of median, in %) and poverty threshold (right panel; €): impact of change of the marginal tax rates



6. Challenges, limitations and updates of the model

6.1. Challenges

A first challenge to this project was, and remains, that some of the results of any dynamic microsimulation model are quite vulnerable to assumptions, modelling approaches and the sensitivity of the indicators to small changes. For example, the AROP of the elderly population is very sensitive, because any change can have both direct and indirect effects, combined with a relatively large group of individuals are clustered around the poverty threshold. Another example is the relation of the risk of unemployment of one household member to whether the other household members are, or are not, working. A small change in this correlation, or in the size and composition of households can have very strong impacts on the VLWI. The simulation of indicators or relations that are inherently nonlinear is a strong point of microsimulation, but also a challenge.

A second challenge pertains to the nature of the data. On the one hand, Nowcasting, like most microsimulation models, assumes a somewhat standard framework where for example a married person is always observed with his or her spouse or people who are retired cannot receive unemployment benefits. In contrast, EU-SILC is not bound by such rules and therefore may contain information that is not always consistent with the model (or even with actual benefit rules), for example respondents that report retirement as their main activity status, while also reporting receiving unemployment benefits. It is thus difficult for a model to replicate such data. Consequently, the model was continually adapted to take into account the uncovered particularities of the data. A series of tests and evaluations were conducted at each step of the model development process to assess the accuracy of the simulated results.

Moreover, during the development of the model and following the discussions with and suggestions of the Steering Group, many changes and additions were implemented within the model. For example, it was necessary to allow for changes in household composition and dissolution in the demographic module, as well as a net income correction (“adjustment component” and “proxy of tax settlement”) within the taxation module, which was not anticipated at the beginning of the project. Also, the short- and medium-term-nature of the model sometimes required different choices from those taken and defensible for the long-term model MIDAS. Thus, a multi-states module was added to Nowcasting that allows the transition between labour market states to take place at any month of the year, and not just at the start of the year.

6.2. Limitations

The current project also has its limitations, some of which are fundamental to the methodology applied, while others point to parts of the model that could be improved.

First of all, in some ways the model is not finished and will require some additional work to improve the fit between the observed trends and produced simulations. The most obvious example of this is the AROP of the elderly population.

Second, we would have liked to proceed with the inclusion of the structural labour model RURO (Random Utility – Random Opportunity, see Capéau *et al.,* 2016) in the Nowcasting model. This would have had the advantage that intensive and/or extensive labour market decisions would have become endogenous. So a typical policy measure would have had an impact on labour market participation and could have higher-order effects on the VLWI and the AROP. We however stand behind our decision to give priority to methodological issues that were deemed more useful in creating the model. These included the simulation of the "multi-states module". It also included the simulation of the "adjustment component" as well as the "proxy of tax supplement" in the taxation module. Finally, it included the handling of missing values and the proportional calibration of the alignment tables to the observed incidence rates.

Third, some sources of income where kept constant, because we lack the information or theoretical underpinning to simulate them. Although not all of these incomes are necessarily equally important in terms of sample-wide averages, they surely have a profound impact on the poverty risks of specific groups in society and it therefore would have been interesting to simulate them in the model.

Finally, the model, operating in a somehow standard framework as other microsimulation models, cannot fully reproduce the observed data and relies on a certain number of assumptions. Among these assumptions is the way missing information in the input dataset is treated (i.e. imputing values or keeping them missing). It is yet unclear whether the choices made in this respect are the optimal ones.

6.3. Updating the Nowcasting model

The updating of the Nowcasting model to a new EU-SILC dataset is done in four steps. First, the EU-SILC data must undergo a treatment process to prepare an input dataset for the Nowcasting. The procedure is already coded in Stata. If the variables and their values in the survey did not change since 2014 wave (same data coding), it is simply matter of running the existing Stata programmes on the new wave. Some checks are however warranted to validate the resulting input dataset for the Nowcasting. Once the input data is created, it must be transferred into the format used for the Liam II software in which Nowcasting is run, and expanded using the frequency weights available in EU-SILC.

Second, all the external alignment parameters and parameters underlying policy rules must be updated. The latter parameters describe the policy rules in effect, while the former are provided by FPB macro models MODTRIM, HERMES and MALTESE at regular intervals (see section 4.3). The Nowcasting model also relies on the internal alignment with parameters obtained using the input data itself. This is the case, for example, for the share of beneficiaries of social assistance. The model limits the number of beneficiaries within each of the simulation years based on the observed figures in the input data. Without this procedure, the simulated number will be too high. The parameters for this internal alignment are obtained using the same Stata programmes as those used to create the input dataset.

Third, the model itself must be updated to take into account recent changes in policy rules that go beyond a simple adjustment of the parameters (e.g. a unique replacement rate for unemployment benefits regardless the duration of unemployment), the current year of the input data, the change in input/output file names, paths and number of simulated periods. Moreover, we would recommend updating the inference-based parameters used in the model, e.g. those used in behavioural equations. The current parameters may no longer accurately represent the individuals' behaviour in the future. It is especially relevant in times of a changing institutional or economic context or important structural changes. Finally, a series of tests and evaluations should be conducted to validate the simulated data.

7. Conclusions

The model developed in this exercise is a fully dynamic longitudinal-ageing microsimulation model (Li *et al.*, 2014), where the characteristics of individuals change following probabilistic or deterministic processes.

Compared to already existing nowcasting estimates for Belgium from the static EUROMOD model, the Nowcasting model presented in this report has the added value that it produces estimates for the very low work intensity rate (VLWI) and the severe deprivation rate (SMD), in addition to the at-risk-of-poverty rate (AROP), so the overarching indicator at-risk-of-poverty-and-social-exclusion rate (AROPE) is also projected. Moreover, the Nowcasting projections extend to the near future (currently to 2020), while the last year in the EUROMOD nowcasts is in fact the previous year. Methodologically, Nowcasting is distinguished from EUROMOD nowcasting in the development of a fully dynamic model to project demographic and labour-market changes. Theoretically, dynamic modelling has a number of advantages over static ageing techniques and may produce more realistic results in some circumstances. In practice, however, the application of dynamic models requires a number of simplifications (e.g. not simulating secondary incomes) that – for the short-term perspective of nowcasting – may cancel out those advantages.

While the Nowcasting results turn out to be – for years for which we have EU-SILC data – in general reasonably close to the observations, they are not clearly superior to those of EUROMOD, and for some categories, in particular the 65+, they are obviously in need of improvement before the model can be used for policy-oriented purposes. The model should better replicate observed data, and future improvements should focus on reducing differences between observed and simulated data. In order to achieve this, fundamental choices, such as dynamic or static ageing, may be less important than seemingly more mundane issues, e.g. to project secondary incomes, and to correctly simulate taxes. It might also be helpful to consider some of the techniques used in nowcasting with EUROMOD. Also, the model should be updated to the latest available wave of EU-SILC data.

The achievements of this project are, first, that it has been demonstrated that short-term simulations (nowcasting and forecasting) of the EU-SILC are possible, using a full-dynamic microsimulation model developed in LIAM2. In this, the methodology developed might be an alternative for current nowcasting techniques, that allows for detailed policy analysis and the production of detailed simulated EU-SILC datasets. From these datasets, estimates of the main social indicators used in the EU2020 programme (the at-risk-of-poverty rate, the very low work intensity rate, the severe material deprivation rate, and the overarching indicator at-risk-of-poverty rate-and-social exclusion rate) for Belgium have been calculated. Projections of many other indicators in the portfolio of EU social indicators (European Commission, 2015) that are based on EU-SILC could also be derived from the simulated data.

Secondly, some innovative approaches and methodological advancements were made in the development of the Nowcasting model. These are possibly relevant to the developers of microsimulation models elsewhere, and we will therefore search for ways to communicate these to them. Third, provided conditions for data-protection are met, the simulated micro-data could be made available to others, to perform additional analyses.

Fourth and finally, the model itself, when adapted and finalized, might be put at the disposal of other federal agencies or international partners. Indeed, since the Nowcasting model is based largely (but not entirely) on the Belgian EU-SILC, it should be possible to transfer it to other European Member states within a reasonable time frame. We will consider the possibility to put the code of the model at the disposal of research teams in other EU member states that already are participating in the "MIDAS network"[24]. We will also search actively for other possibilities to find new member states that might be interested in collaborating with us to adopt the Nowcasting model. This will allow a more comprehensive use of the EU-SILC in the context of the Europe 2020 strategy (and beyond), through comparative research in various member states. The work done on long-term projections in the context of the SPC WG AGE has clearly demonstrated this value added, and we are convinced that it will also prove useful for nowcasting issues and short-term projections.

24 The "MIDAS network" is an informal network of researchers or research teams that have in the past joined forces in adapting the dynamic microsimulation model MIDAS to other member states. This includes LISER (Luxembourg), a research team led by professor Amilcar Moreira (University of Lisbon, Portugal), and the National Treasury in Hungary.

8. References

Bánbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013), *Now-Casting and the Real-Time Data Flow*, European Central Bank Working Papers Series no 1564, July 2013.

Capéau, Bart, Decoster, André, and Dekkers, Gijs (2016), "Estimating and Simulating with a Random Utility Random Opportunity Model of Job Choice. Presentation and application to Belgium", International Journal of Microsimulation, 9(2), 144-191.

Chènard, Denis (2000), Individual alignment and group processing: an application to migration processes in DYNACAN. In Mitton, L., Sutherland, H. and Weeks, M. (eds), Microsimulation Modelling for PolicyAnalysis, Cambridge University Press, Cambridge.

Dekkers, Gijs, Raphaël Desmet, and Greet De Vil (2010), The long-term adequacy of the Belgian public pension system: An analysis based on the MIDAS model, Working Paper 10-10, May 2010, Brussels: Federaal Planbureau.
http://www.plan.be/publications/publication_det.php?lang=nl&TM=30&IS=63&KeyPub=946 [26/10/2018].

Dekkers, Gijs, Seiichi Inagaki, and Raphaël Desmet (2012), Dynamic Microsimulation Modeling for Policy Support: An Application to Belgium and Possibilities for Japan. Review of Socionetwork Strategies, 6(2), pp 31-47.

Dekkers, G. (2015a), "The simulation properties of microsimulation models with static and dynamic ageing – a brief guide into choosing one type of model over the other". *International Journal of Microsimulation*, 8 (1), pp. 97-109.

Dekkers, Gijs (2015b), On the modelling of immigration and emigration using LIAM2. FPB Note LIAM2-11155. Brussels: Federal Planning Bureau. https://www.researchgate.net/publication/286869730_On_the_modelling_of_immigration_and_emigration_using_LIAM2 [29/08/2018].

Dekkers, Gijs, and Cumpston, Richard (2012), On weights in dynamic-ageing microsimulation models. *International Journal of Microsimulation*, 5 (2), pp. 56-65.

Dekkers, Gijs, Raphaël Desmet, Nicole Fasquelle, and Saskia Weemaes (2015), The social and budgetary impacts of recent social security reform in Belgium. In Ioana Salagean, Catalina Lomos & Anne Hartung, "The young and the elderly at risk: Individual outcomes and contemporary policy challenges in European societies", Intersentia. ISBN 978-1-78068-343-0. Chapter 6, pp. 129-158.

Dekkers, Gijs, and Karel Van Den Bosch (2016), Prospective microsimulation of pensions in European Member States. In: Dekkers, Gijs, and József Mészáros (Eds), 2016, Applications of microsimulation modelling, Budapest: Társadalombiztosítási Könyvtár. 13-33, ISBN 978-963-693-766-9, 13-33.

Dekkers, Gijs, and Ekaterina Tarantchenko (2018), The Nowcasting model – First results. Paper presented to the Steering Committee of the Nowcasting Project. Unpublished paper Federal Planning Bureau. January 2018.

de Menten, Gaëtan, Gijs Dekkers, Geert Bryon, Philippe Liégeois, and Cathal O'Donoghue (2014), LIAM2: a new open source development tool for discrete-time dynamic microsimulation models, *Journal of Artificial Societies and Simulation*, 17(3)9. http://jasss.soc.surrey.ac.uk/17/3/9.html.

Duleep, H.O., and Dowha (2008), Adding immigrants to microsimulation models. *Social Security Bulletin*. 68(1): 51-66.

European Commission (2010), *Using EUROMOD to nowcast poverty risk in the European Union*. Luxembourg: Publications Office of the European Union. KS-RA-13-010-EN-N. https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-RA-13-010.

European Commission (2013), *Using EUROMOD to nowcast poverty risk in the European Union, 2013 edition*. Luxembourg: Publications Office of the European Union. KS-RA-13-010-EN-N. https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-RA-13-010.

European Commission (2015), *Portfolio of EU social indicators for the monitoring of progress towards the EU objectives for social protection and social Inclusion, 2015 update*. Luxembourg: Publications Office of the European Union.

European Commission (2018a), Flash estimates of income inequalities and poverty indicators for 2017 (FE 2017): Experimental results., Directorate F: Social statistics, Unit F-1: Social indicators; methodology and development; relations with users. Update - October 2018, https://ec.europa.eu/eurostat/documents/7894008/8256843/Flash-estimates-of-income-inequalities-and-poverty-indicators-experimental-results-2017.pdf [16/11/2018].

European Commission (2018b), Flash estimates of income inequalities and poverty indicators for 2017 (FE 2017): methodological note. Technical Report EUROSTAT, Directorate F: Social statistics, Unit F-1: Social indicators; methodology and development; relations with users. September 2017 https://ec.europa.eu/eurostat/documents/7894008/8256843/Methodological-note-2017.pdf [26/10/2018].

Eurostat (2018), Archive: Material deprivation and low work intensity statistics. Eurostat Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Material_deprivation_and_low_work_intensity_statistics&oldid=110770 [17/10/2018].

European Commission, 2018, Glossary: At risk of poverty or social exclusion (AROPE). Eurostat Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:At_risk_of_poverty_or_social_exclusion_(AROPE) [22/11/2018]

Federal Planning Bureau (2017), Economic Policy Committee's Ageing Working Group. Belgium: Country Fiche 2017. Paper retrieved from the website of the Federal Planning Bureau. https://www.plan.be/admin/uploaded/201805251442290.AWG_Country_fiche_Belgium_2017_11597.pdf.

Federal Public Service Social Security (SPF SS) (2018), The Evolution of the Social Situation and Social Protection in Belgium 2018 – 'increasing pressure on social protection adequacy'. Report retrieved from the website of the SPF SS. https://socialsecurity.belgium.be/sites/default/files/analysis-social-situation-and-protection-belgium-2018-en_2.pdf. [23/11/2018]

Frère, J.-M. (2016), *The population at risk of poverty or social exclusion in Belgium - projection until 2030*. Federal Planning Bureau, Working Paper 12-16, November 2016. https://www.plan.be/admin/uploaded/201611280930210.WP_1612_11326_E.pdf

Gasior, K., and Rastrigina, O. (2017), *Nowcasting: timely indicators for monitoring risk of poverty in 2014-2016*, EUROMOD Working Paper Series EM7/17.

Goedemé, T., and Montaigne, F. (2010), The computation of the Europe 2020 poverty indicators in Stata, Antwerp; Herman Deleeck Centre for Social Policy.

Guio, A.-C., Gordon, D., and Marlier, E. (2012), Measuring Material Deprivation in the EU: Indicators for the Whole Population and Child-Specific Indicators. Eurostat Methodologies and Working Papers, Luxembourg: Publications Office of the European Union.

Leventi, C., Navicke, J., Rastrigina, O. and Sutherland, H. (2014), *Nowcasting risk of poverty and income distribution in the EU in 2013*, EUROMOD Working Paper Series EM11/14.

Li, Jinjing, Cathal O'Donoghue, and Gijs Dekkers (2014), Dynamic Microsimulation. In: Cathal O'Donoghue (Ed). Handbook of Microsimulation. Emerald Insight. Contributions in Economic Analysis series.

Li, Jinjing, and Cathal O'Donoghue (2014), An overview of binary alignment methods in Microsimulation. In: Gijs Dekkers. Marcia Keegan and Cathal O'Donoghue (Eds), 2014. New Pathways in Microsimulation. Farnham: Ashgate Publishing Limited. Chapter 14, pp. 217-232.

Navicke, J., Rastrigina, O., and Sutherland, H. (2014), "Nowcasting indicators of poverty risk in the European Union: a microsimulation approach", *Social Indicators Research*. vol. 119, pp. 101-119.

Philippe Liégeois, and Gijs Dekkers (2014), Combining EUROMOD and LIAM tools for the development of dynamic cross-sectional microsimulation models: a sneak preview. In: Gijs Dekkers, Marcia Keegan and Cathal O'Donoghue (Eds), 2014, New Pathways in Microsimulation. Farnham: Ashgate Publishing Limited, Chapter 13, pp. 203-216.

Rastrigina O., Leventi, C., and Sutherland, H. (2015) "Nowcasting risk of poverty and low work intensity in Europe". EUROMOD Working Paper EM9/15. Colchester: ISER, University of Essex.

Vandresse, Marie, Johan Duyk, and Jean-Marc Paul (2018), Demografische vooruitzichten 2017-2070 - Bevolking en huishoudens. Paper retrieved from the website of the Federal Planning Bureau. https://www.plan.be/admin/uploaded/201802221219560.FOR_POP1770_11621_N.pdf.

9. Appendices

9.1. Appendix 1: Equivalised disposable income

$$Equivalised\ disposable\ income = \frac{Household\ disposable\ income}{Equivalised\ scale}$$

With

$$\begin{aligned} &Household\ disposable\ income \\ &\quad = Gross\ income\ components\ at\ household\ level \\ &\quad + \sum_{i=1}^{h}(Gross\ income\ components\ at\ individual\ level) \\ &\quad - Taxes\ and\ transfers\ paid\ at\ household\ level \\ &\quad - \sum_{i=1}^{h}(Taxes\ and\ transfers\ paid\ at\ individual\ level) \end{aligned}$$

$$Equivalised\ scale = 1 + (nadult14 - 1) * 0.5 + nchild14 * 0.3$$

Where h equals the total number of persons within the household, *nadult14* represents the number of persons aged +14 within the household and *nchild14* represents the number of persons aged 0-13 within the household.

9.2. Appendix 2: List of simulated and projected income components in the nowcasting model that are part of the equivalised disposable income

Variable description	Variable name in EU-SILC	Variable used in BE-Nowcasting	Simulated/projected in BE-Nowcasting
Gross income components at individual level			
Gross employee cash or near cash income	py010g	py010g	Simulated
Company car	py021g	py021g	Projected
Income from self-employment	py050g	py050g	Simulated
Retirement pension	py100g	i102b_b * i102b_m + i102d_b * i102d_m + i102e_b * i102e_m	Simulated
Income guaranty for elderly		i102c_b * i102c_m	Simulated
Survival pension	py110g	i102a_b * i102a_m	Simulated
Pension from private plans (2nd and 3rd pillars)	py080g	i110 + i113 + i104 + i107	Projected
Unemployment benefits	py090g	i98a_b * i98a_m + i98b_b * i98b_m + i98c_b * i98c_m + i98d_b * i98d_m + i98e_b * i98e_m + i98h_b * i98h_m + i98f_b * i98f_m + i98g_b * i98g_m	Simulated
Unemployment with company suppl.		i99 * i101	Simulated
Sickness and disability benefits	py120g + py130g	i115a_b * i115a_m + i115c_b * i115c_m + i115b_b * i115b_m + i115d_b * i115d_m + i115e_b * i115e_m + i115f_b * i115f_m + i115g_b * i115g_m + i115h_b * i115h_m + i115i_b * i115i_m + i115j_b * i115j_m	Simulated
Education-related allowances	py140g	py140g	Projected
Social assistance allowances		i184b * i186	Simulated
Gross income components at household level			
Income from rental of a property or land	hy040g	hy040g	Projected
Family/children related allowances	hy050g	hy050g	Simulated
Social assistance allowances	hy060g		
Housing allowances	hy070g	hy070g	Projected
Regular inter-household cash transfers received	hy080g	hy080g	Projected
Interests, dividends, profit from capital investments in unincorporated business	hy090g	hy090g	Projected
Income received by people aged under 16	hy110g	hy110g	No (observed in 3 households in 2014)
Taxes and transfers paid at individual level			
Tax on income and social insurance contributions		Gross* – Net**	Simulated
Tax supplement for incomes received in t-3		i130 – i132	Simulated
Taxes and transfers paid at household level			
Regular taxes on wealth	hy120g	hy120g	Projected
Regular inter-household cash transfer paid	hy130g	hy130g	Projected
Tax on income and social insurance contributions (including tax supplement)	hy140g		

Note: Gross* represents the sum of the variables listed in the column "Variable used in BE-Nowcasting" for gross income components at the individual and household level.

Net** is the net equivalent of Gross**, using net amounts.

9.3. Appendix 3: Alignment in the first channel of consistency

The below list shows the information used in the alignment in the first channel of consistency of the Nowcasting model.

- Immigrating individuals, Belgian nationality, to age and gender (numbers[25])
- Immigrating individuals, Other nationality, to age and gender (numbers[24])
- Emigrating individuals, Belgian nationality, to age and gender (numbers[24])
- Emigrating individuals, Other nationality, to age and gender (numbers[24])
- Fertility rate to age (probability[25])
- Mortality rate to age and gender (probability[25])
- Living with a partner, to 5-year age group and gender (probability[26])
- Living with a partner, to 5-year age group and gender (probability[25])
- Registered cohabitation, to 5-year age group and gender (probability[25])
- Divorce rate, to 5-year age group and gender (probability[25])
- Separation rate, to 5-year age group and gender (probability[25])

Labour market states (idem)

- Total population, 5-year age group and gender (numbers[24])
- Active population, 5-year age group and gender (numbers[24])
- Unemployed population, to 5-year age group and gender (numbers[24])
- Self-employed population, to 5-year age group and gender (numbers[24])
- Unemployed population, to 5-year age group and gender (numbers[24])
- Public sector employee, to 5-year age group and gender (numbers[24])
- Civil servant, to 5-year age group and gender (numbers[24])
- Unemployment with Company Allowance, to 5-year age group and gender (numbers[24])
- Disabled, to 5-year age group and gender (numbers[24])

Note that the above alignment tables for the labour market states are expressed in numbers of individuals, but are transformed in proportions through the confrontation with the population numbers. Other relevant labour market states are derived from them, and in the same order as the decision tree for the simulation of the various labour market states. For example, the proportion of people in work is derived by (active population - unemployed population) / total population, and the proportion of employees among the working population is then 1 - (self-employed / (active population - unemployed population)).

---

[25] These actual or projected numbers are downscaled to correct for the difference between the Belgian population and the size of the expanded EU-SILC sample.

[26] Proportionally uprated to the incidence rates observed in the starting dataset.

9.4. Appendix 4: The estimation of labour market transitions and other equations for the Nowcasting model

The estimations of the labour market transitions (probability to be in a particular labour market state in t given the labour market state in t-1), wages, working hours and pensions are based on the longitudinal SILC data. The estimations used in taxation module are based on the cross-sectional SILC data of 2014. Stata do-files used:

- Treatment of the data: 'Nowcasting_4_createlongitudinal_SILC.do'
- Labour market transitions: 'Nowcasting_5_estimate_transitions.do'
- Income from work, working hours and pensions: 'Nowcasting_6_estimate_earnings_hours.do'
- Estimates used in the taxation module: 'Nowcasting_9_estimate_tax_difference.do'

This note first describes the treatment of the longitudinal SILC data and the final sample selection. We then describe four methodological approaches that are used to estimate 1) labour market transitions; 2) income from work and working hours; 3) pensions and 4) tax adjustment and tax supplement.

9.4.1. Preparing longitudinal data

- Input datasets: all individual and household files from longitudinal EU-SILC as well as cross-sectional datasets that were generated in 'Nowcasting_3_startdata_SILC.do'.
- Program: 'Nowcasting_4_createlongitudinal_SILC.do'
- Output dataset: '…\nowcasting\SILC\results\long_estim_2years.dta'

We need a sufficiently large panel of individuals observed for at least 2 consecutive years. Currently, we have 2007-2014 longitudinal waves with individual and household information for the period 2004-2014. This subsection provides a description of how the data is treated.

9.4.1.1. Linking Longitudinal data across waves

a. Successive waves

Each successive wave contains information on t till t-3 and is updated compared to the previous wave. For instance, the wave 2007 contains information on years 2004-2007, the wave 2008 contains information on 2005-2008 that is updated (for 2005-2007) compared to the wave 2007, etc. This means that some variables for a particular year can change from one wave to another (e.g. the date of birth of an individual for 2007 can be different if he is observed in the wave 2007 or 2008). The updated information is verified and validated and thus should be more reliable. Keep in mind that the update is done for each new longitudinal wave and does not go backwards. Thus, if several waves are linked together, it is better to keep the information from the most recent wave. Once the redundant observations are dropped, it could still happen that for the same wave and same year, an individual is observed twice. This is due to the change of the household (e.g. children leaving their parents) in which case the person appears twice: once with the old household identifier and once with a new one. Keep the line with the new identifier. If the above steps are followed, you end up with a panel of individuals observed 1 to 4 years over a period that starts in 2004 and ends with the most recent year available (currently 2014).

Citizenship and country of birth

The variables indicating the citizenship and country of birth are only available since 2014 wave.

b. Weights

There are 4 individual weight variables: RB060, RB062, RB063 and RB064. RB060 is the base weight. RB062 is used to compare the year of the wave to the previous year. That is, the wave 2008 can be used to compare the year 2008 to 2007 but not 2007 to 2006. To compare 2007 to 2006 you need the wave 2007. RB063 is used to compare the year of the wave to the year t-2 (e.g. for the wave 2007, compare 2007 to 2005). RB064 is used to compare the year of the wave to the year t-3 (e.g. for the wave 2007, compare 2007 to 2004). The choice of the weight variable depends on the analysis. If the aim is to analyse the transitions from one labour market state to another within 1-year interval when all the waves are combined, then one should choose RB062. Be aware that you can do that only for the years 2006 and the following. That is, the first wave is 2007 that can be used to only compare year 2006 to 2007 but not 2005 to 2006 (and neither 2004 to 2005). To compare year 2007 to 2008 take RB062 from the wave 2008. To compare 2008 to 2009 take RB062 from the wave 2009, etc.

9.4.1.2. Linking cross-sectional data across years

To estimate various equations, we need some variables that are only available in the cross-sectional files (e.g. workstate, citizenship, country of birth, etc) that we created in 'Nowcasting_3_startdata_SILC.do' (the data file is '…\nowcasting\SILC\results\midas_silc.dta'). Until 2014, the two types of data, longitudinal and cross-sectional, were produced independently with their own identifiers and could not be linked directly. Starting 2014, the individual and household identifiers are the same in both types of data. However, for the waves up to 2014, Statistics Belgium (FPS Economy) provided us with a file that contains the cross-sectional identifiers for each wave linked to the longitudinal identifier. These allow to link the observations from several cross-sectional years for the same individual. But these links are not always perfect and could result in linking wrong individuals from one year to another or changing the identifier for the same individual. To correct this, there is no other way as to consider each individual separately. Given that this is a very time-consuming procedure (and we are not sure to correct all the anomalies), we did not do it. However, some small corrections have been done:

- The file with links may contain 2 observations (rows) for the same longitudinal identifier for the same wave. There are 2 possibilities. First, an individual may change the household in which case a new household identifier is created. In that case, just drop the row with the old household identifier. Second, the same longitudinal identifier appears for 2 different people (for one of the cross-sectional waves). There are 6 observations in that case (3 longitudinal identifiers repeated twice). You can look at the cross-sectional waves to determine which observation should be dropped.
- When the links are applied to the cross-sectional waves, some observations remain unlinked (waves 2009-2011). Moreover, several tests show that gender and date of birth vary for the same individual across the years. The change in the date of birth can be explained, at least partly, by the updating of

the data (see subsection 9.4.1.1). While the change in the gender is probably due to the incorrect links provided in the links file.

9.4.1.3. Linking longitudinal with cross-sectional files

The variables transferred from linked cross-sectional waves to longitudinal waves are: workstate (labour market state), collar (blue-/white-collar status), parttime (indicator for working part-time), citizen (citizenship) and cofbirth (country of birth). Some observations from the linked cross-sectional waves (panel A) are missing in the panel of longitudinal waves (panel B) and vice-versa. In the first case (individuals from panel A are missing in the panel B), some individuals are observed up to 3 years. I do not know why it is the case yet (imperfect links for cross-sectional waves? But the number of observations in that case is too large compared to the missing observations in the panel A but present in the panel B). The second case (individuals from the panel B are missing in the panel A) is probably due to the imperfect links. Given that we need the information from both panels, we keep only the observations that are present in both panels (panel C).

Some additional variables are created that are equivalent to the variables generated for the input file for the Nowcasting model (i.e. based on the last cross-sectional wave).

Remarks:

– Some missing values for workstate were completed: if the individual has an income from work and workstate(t-1) = 1,2,3 or 4 then workstate = workstate(t-1).
– Some missing values for education (educ) were completed: if a person is observed for several periods but his education is missing for the 2^{nd}, 3^{rd} or 4^{th} period then replace the missing with the value of the previous period. The command also replaces the value of education in period t with that of period t-1 if the individual had higher level of education in t-1 (only for 2^{nd}, 3^{rd} or 4^{th} period).

  bys rb030_long (rb010): replace educ = educ[_n-1] if _n > 1 & educ < educ[_n-1]

  The missings are replaced with 0 before this code is applied and reverted to missings after the code.

Incomes are expressed in constant prices (of year 2015) using the CPI index. Unlike all the other variables that reflect current situation of a person (labour market state, working hours, family situation, etc.), reported incomes pertain to the previous year. The index takes that into account (see 'Price_index.xls"). For example, the index applied to the incomes reported in wave 2014 is that from 2013, given that those incomes were received in 2013. These incomes are then expressed in prices of 2015. The transformed variables are employees' earnings (earny) and self-employment income (selfearny).

9.4.1.4. Selection of the common sample for estimations based on longitudinal data

Once the additional information from the cross-sectional data was added to the panel of longitudinal waves (panel C), we can proceed to select the desired sample of individuals whose labour market transitions will be analysed. The sample consists of individuals with non-missing socio-economic state and who are observed for at least 2 consecutive years. More precisely:

a) Drop observations with missing workstate

```
keep if workstate ~= -999 & !missing(workstate)
```

b) Keep individuals with at least 2 consecutive years

```
tsset rb030_long rb010
tsspell, f(L.rb010 == .)
bys rb030_long _spell: keep if _N > 1
```

9.4.2. Labour market transitions

- Input dataset: '…\nowcasting\SILC\results\long_estim_2years.dta' generated in 'Nowcasting_4_createlongitudinal_SILC.do'.
- Program: 'Nowcasting_5_estimate_transitions.do'
- Output results: 'Nowcasting_LMequations_allstatesLogit_noweights_final.log'

9.4.2.1. Estimation sample

The general condition for the estimations is being aged 15-60 with non-missing education (we use education as explanatory variable). We create a new variable workstate_1 that represents the lag of the variable workstate. This variable is used to estimate labour market transitions conditional on the previous labour market state. By construction, workstate_1 is missing for the first observed period. We then keep the observations with non-missing workstate_1. Thus, a person who was observed for 2 consecutive periods appears only once, but we keep the information on her labour market state for both periods.

9.4.2.2. Methodology

The aim is to estimate the probability of being in a particular state conditional on the state in the previous period. The model takes the following form:

$$Prob(y_{i,t} = 1 \,|y_{i,t-1}, X_{i,t}) = G(X_{i,t}\beta) \qquad i = 1, \dots, N \; and \; t = 2004, \dots, 2014,$$

where $G(X_{i,t}\beta)$ is a logistic function[27]. The model is estimated separately in each of the 9 cases described below:

A) Probability of being in work in t conditional on:

1) being in the absorbing states[28] in t-1

2) being in education in t-1 and not being in education in t

---

[27] $G(X_{i,t}\beta) = \frac{e^{X_{i,t}\beta}}{1+e^{X_{i,t}\beta}}$

[28] The absorbing states include unemployment (workstate = 6), disability (workstate = 7), unemployment with company supplement (workstate = 8), retirement (workstate = 9) and other inactive (workstate = 10).

B) Probability of being employee in t[29] conditional on being in work in t and:

3) being employee in t-1

4) being self-employed in t-1

5) not being in work in t-1

C) Probability of being employee in public sector in t conditional on being employee in t and:

6) being wage earner in public sector in t-1

7) not being employee in public sector in t-1 -> i.e. being wage earner in private sector or self-employed

8) not being in work in t-1

D) Probability of being unemployed in t conditional on:

9) being wage earner in t-1 and not in work in t and not receiving conventional early leave benefits in t and not being disabled in t

The conditioning is explained by a step-by-step alignment. We first determine the number of people who work (step A). Second, among the people who work, we determine who works as employee (step B). The remaining part works as self-employed. Third, among the employees, we determine who works in the public sector (step C). The remaining part of the employees work in the private sector (i.e. wage earners in the private sector). Finally, among those who were selected as employees in the public sector, we determine who is civil servant. The final number of civil servants is adjusted randomly through the alignment, with a priority given to those who were civil servants in t-1. Furthermore, the number of unemployed is determined through alignment, using the estimated probability of being unemployed (step D).

Depending on the estimated equation, $X_{i,t}\beta$ includes all or a subset of the following explicative variables: dummies for the observation year, dummies for 1, 2 or +3 children aged 0-17 within the household (0 children is defined as reference category), level of education, dummies for being born in a foreign country and having foreign citizenship and their cross-product, dummies for blue-collar wage earners and white-collar wage earners (civil servants are defined as reference category), dummy for partner, dummy for partner in work in t-1, second degree polynomial in age and in person's and her partner's earnings in t-1. All the equations include year dummies. For the estimation purposes, all the partner's characteristics are put to 0 if there is no partner (i.e. partner's identifier is missing or the person has no partner). If the partner is present but his characteristics are missing, we leave them as missings. The earnings used in the estimations are divided by 1000.

[29] Employees are working people except the self-employed (i.e. wage earners from private and public sectors, as well as civil servants).

The model is estimated with clustered standard errors at the individual level.[30] The estimation results are in line with Pooled Ordinary Least Squares (POLS) in terms of the sign and statistical significance. The final estimation results are reported in 'Nowcasting_LMequations_allstatesLogit_noweights_final.log'. We tried various specifications (including/excluding some explanatory variables, performing joint F-test on excluded variables). Only final estimations are reported in the program and log file. If certain reported variables appear not statistically significant at the individual level, the F-test rejects their joint significance. Thus, there is multicollinearity issue.

9.4.3. Earnings, self-employment income and working hours

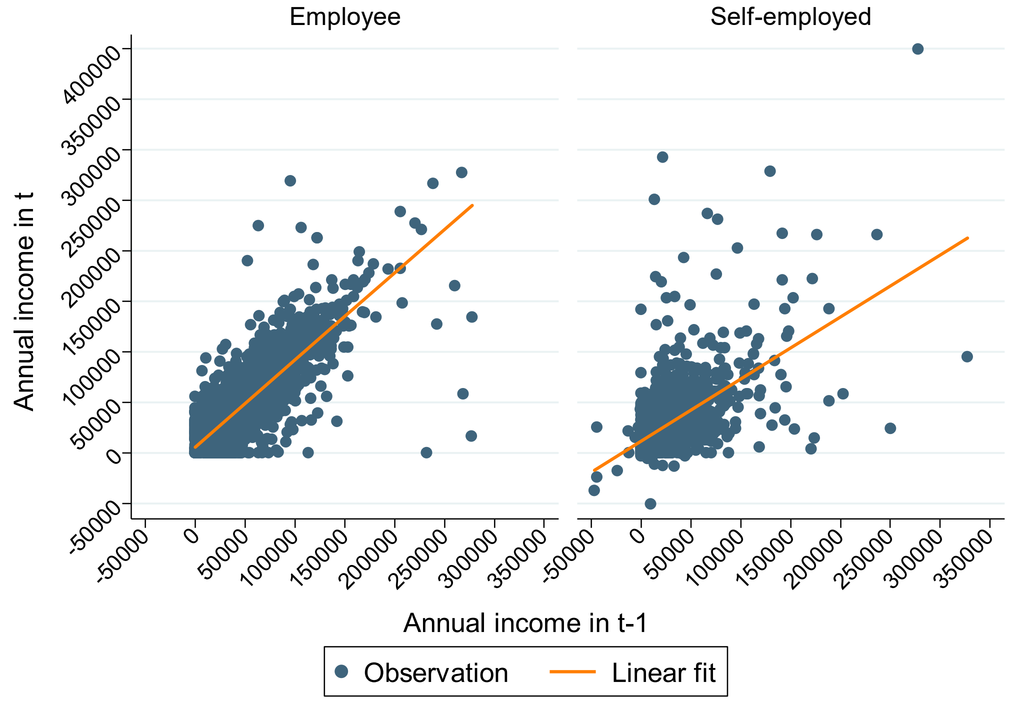
9.4.3.1. Analysis of the income from work and working hours for employees and self-employed

In general, regardless of the income source, the Nowcasting model simulates 'new' incomes for individuals whose labour market state changed with respect to the previous year. As for the individuals whose labour market state did not change, they mainly keep the income from the previous year. Usually, this makes sense for all incomes except the income from self-employment which is more unstable. In this section we analyse the income of self-employed and compare its "behaviour" to the income of employees. This analysis should help deciding whether we should use the same strategy that is used for employees or take a different approach. The first approach consists in the same idea as for employees, which is to estimate a model that will be used to predict working hours and wages for new employees (those who remain employees mainly keep the same working hours and wages). The second approach would be to choose randomly an income/working hours pair in the observed distribution. This second approach does not guarantee the same level of income/hours from self-employment as in the previous year for individuals who remain self-employed. If the income distribution is more or less random from one year to another, the second approach is a better choice. But if the level of income remains equivalent from one year to another, the first approach is more suitable.

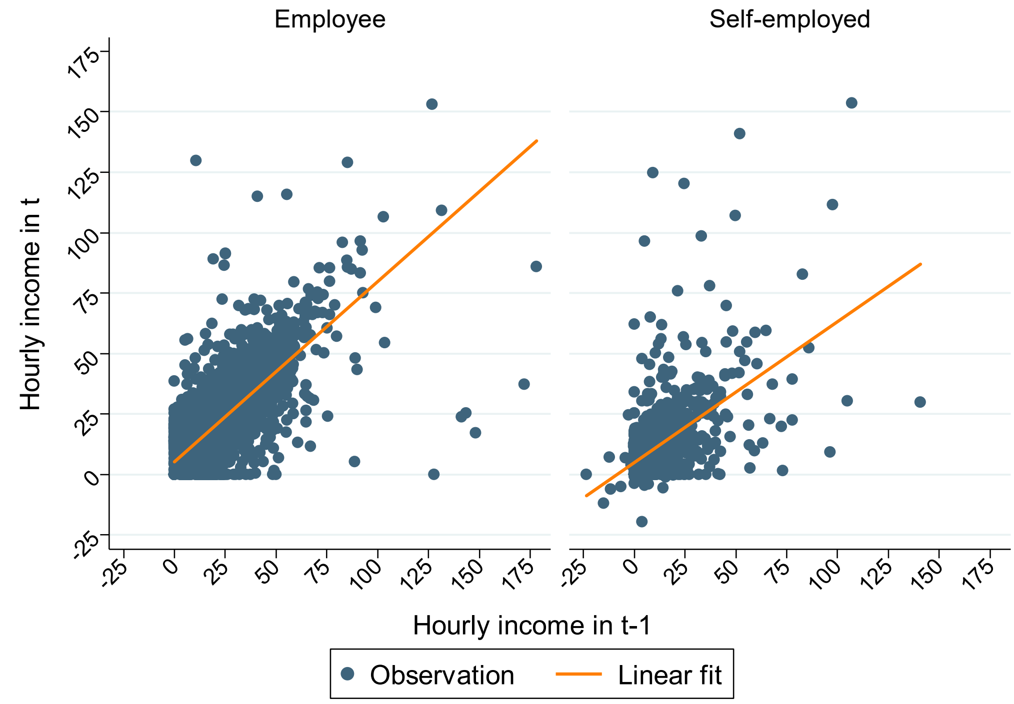
Graphs A1 and A2 show positive relation between the previous period's income and current income for both employees and self-employed. However, the dispersion for self-employed is more important. As Graph A3 shows, the relation between the previous and current income for self-employed is not 1 to 1 and is more unpredictable. Graph A4 shows strong relationship between hours from the previous and current year for both employees and self-employed. The part-time indicator also remains the same between two consecutive years in 93% of the cases for both employees and self-employed.

[30] There is serial correlation in the error term because the same individual can be observed during several time periods. This means that the residuals are not independent for the same individual. Clustered standard errors take into account this correlation (as well as heteroscedasticity).

Graph A1 Comparison of the current annual income to the previous period's annual income for all current employees and self-employed

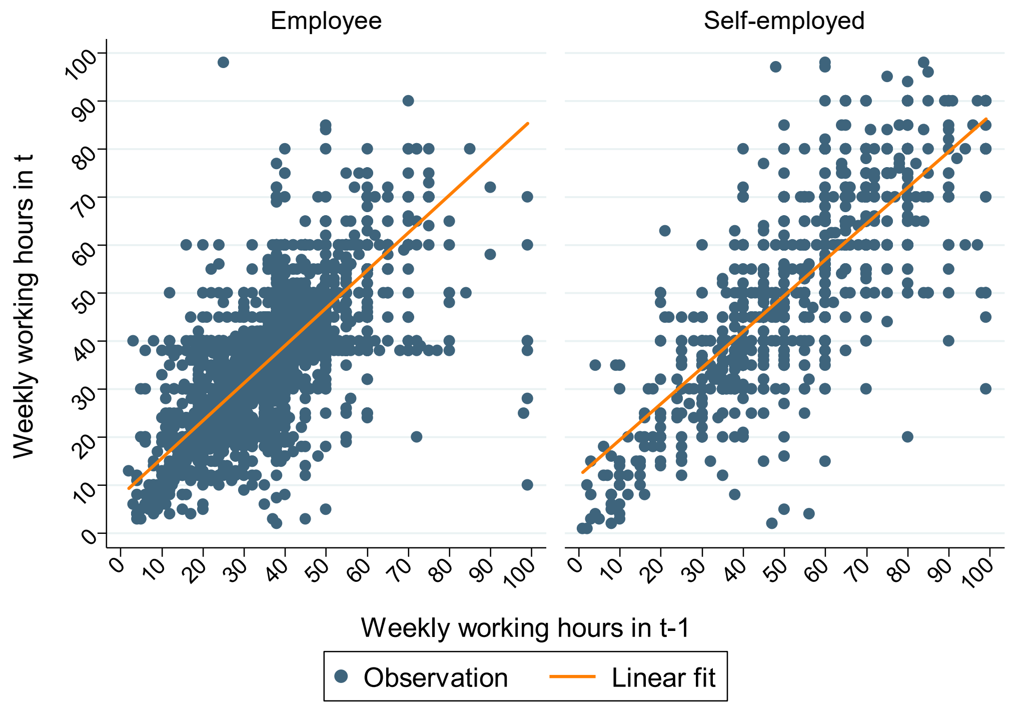


Graph A2 Comparison of the current hourly income to the previous period's hourly income for all current employees and self-employed



Graph A3 Comparison of the current hourly income to the previous period's hourly income for current employees and self-employed with non-zero previous period's income

Graph A4 Comparison of the current weekly working hours to the previous period's weekly working hours for current employees and self-employed



9.4.3.2. Estimating earnings and working hours for employees

– Input dataset: '…\nowcasting\SILC\results\long_estim_2years.dta' generated in 'Nowcasting_4_createlongitudinal_SILC.do'
– Program: 'Nowcasting_6_estimate_earnings_hours.do'
– Output results: 'Nowcasting_Long_wages_hours.log'

a. Estimation sample

As already mentioned, each SILC wave t reports information on year t (e.g. self-defined current economic status, family situation, household composition, working hours…) and year t-1 (e.g. incomes, number of months in unemployment…). Logically, we should regress working hours on earnings from the same year. Therefore, we add the earnings from wave t+1 to the wave t. Which means that the last wave is dropped from the estimations.[31] The remaining sample selection is as follows:

– Being wage earner or civil servant (workstate = 1, 2 or 3)
– Working hours should be strictly positive and non-missing
– Earnings should be strictly positive and non-missing
– Part-time indicator should be non-missing
– Being aged 15-59[32]
– Education should be non-missing

b. Methodology

The methodology consists of two steps. First, we estimate the probability of working part-time that is used to simulate who works part-/full-time. In a second step, we estimate earning and working hours simultaneously for 4 subgroups of individuals: men working full-time, men working part-time, women working full-time and women working part-time.

Step 1: Estimating the probability of working part-time

The probability of working part-time is estimated through the following model (where individual *i* and time *t* subscripts are omitted for simplicity)

$$Prob(y = 1\ |X) = G(X\beta)$$

where $G(X\beta)$ is a logistic function. The model is estimated separately for men and women, with standard errors clustered at the individual level. $X\beta$ represents a set of exogenous covariates (excluding working hours and earnings) such as year dummies, dummies for blue-collars in public sector, white-collars in public sector, blue-collars in private sector, white-collars in private sector (civil servants are

31 We could have added the hours from wave t to earnings in wave t-1. However, in that case we need to transfer all the other characteristics from the wave t to t-1 or assume that these characteristics did not change between the two waves, which is a too strong assumption.
32 The compulsory education ends at age 18 (or as soon as a pupil obtains the diploma of secondary education) but as of age 15, pupils can do part-time learning and working.

defined as reference category), (second degree) polynomial in age and experience, dummies for 1, 2 or +3 children aged 0-17 within the household (0 children is defined as reference category), level of education, dummies for being born in a foreign country and having foreign citizenship and their cross-product and dummy for partner.

The estimation results are used in simulations to predict the work intensity of individuals (part- or full-time). This latter procedure is directly implemented in LIAM2 and consists in selecting the workers with the highest predicted probability of working part-time. The final number of part-time workers depends on the alignment figures obtained from the last cross-sectional wave.

Step 2: Estimating earnings and working hours

Earnings and working hours are estimated through Simultaneous Equations Model (SEM) where both earnings and working hours are considered endogenous (earnings in function of working hours and vice-versa). A SEM is a structural model that takes the following form (where individual i and time t subscripts are omitted for simplicity):

$$\begin{cases} y_1 = X\beta_1 + Z_1\gamma_1 + y_2\alpha_1 + u_1 & (1) \\ y_2 = X\beta_2 + Z_2\gamma_2 + y_1\alpha_2 + u_2 & (2) \end{cases}$$

Where X is a vector of exogeneous variables common to y_1 and y_2, $Z_1$is a vector of exogenous variables appearing only in (1) and Z_2 is a vector of exogenous variables appearing only in (2). The identification requirement for a 2-equations SEM is that each equation must contain at least one exogenous variable that does not appear in the other equation. We use two-stage least squares (2SLS) to estimate each equation, with all the exogenous variables from both equations as instruments in the first stage. More precisely, for equation (1) we first regress y_2 on all the exogenous variables that appear in both equations (1) and (2) (i.e. X, Z_1 and Z_2) and predict $\hat{y}_2$ (first stage). Then, equation (1) is estimated with $\hat{y}_2$ instead of y_2 (second stage). The same principle applies to equation (2).

If one is not interested in the parameters of equations (1) and (2) but rather in the prediction of y_1 and y_2, which is our case, then estimating the reduced form of both equations by OLS is sufficient. The predicted values are the same as through the estimation of equations (1) and (2) by 2SLS or even the first stage from the 2SLS. That is, the reduced form for y_1 is a linear function of X, Z_1 and Z_2. The reduced form for y_2 is also a linear function of X, Z_1 and Z_2. This is exactly the same as estimating the first stage in 2SLS. Or if, in equation (1) in the second stage of 2SLS, we replace $\hat{y}_2$ by [$y_2 -$ residuals from the first stage of 2SLS], with y_2 expressed as a linear function of X, Z_1 and Z_2, we end up regressing y_1 on X, Z_1 and Z_2. Which is again equivalent to estimating the reduced form of equation (1).

Given that it is simpler to use OLS estimates in LIAM2 and we are not interested in the parameters of the structural model but rather the predictions, we proceed as follows:

a) Estimate the reduced form for earnings by regressing y_2 on all the exogenous variables that appear in both equations (1) and (2) (i.e. X, Z_1 and Z_2) and predict $\hat{y}_2$.

b) Estimate equation (1) with $\hat{y}_2$ instead of y_2. Both estimations are done by OLS and can be directly used in LIAM2.

In simulations we proceed as follows:

c) Predict earnings and residuals using estimated parameters from the reduced form equation for earnings in a). Simulated earnings = predicted earnings + earnings residuals. Basically, in the base year simulated earnings are the same as observed ones for individuals with reported positive values. Individuals with missing earnings will be given simulated earnings equal to predicted value based on individual characteristics, plus a residual randomly drawn from the distribution of estimated earnings residuals.

d) Predict hours using estimated parameters in b), with simulated earnings as control variable that replaces $\hat{y}_2$ and obtain residuals. Simulated hours = predicted hours + hours residuals. Similar to earnings, in the base year simulated hours are the same as observed ones for individuals with reported positive values. Individuals with missing hours will be given simulated hours equal to predicted value based on individual characteristics, plus a residual randomly drawn from the distribution of estimated hours residuals. Note that this step is equivalent, in terms of the expected value, to predicting hours using estimated parameters in b), with ***predicted*** earnings as control variable instead of simulated ones.

We estimated several equations using various combinations of instruments and exogenous variables. The common and instrument variables reported in the log files seem to be the best choice given the endogeneity and overidentification tests results (the tips on the implementation of these tests in Stata can be found in the Appendix 5 of this document). All the tests were performed on the 2SLS estimates. We also tried various specifications, including/excluding some explanatory variables. A joint F-test was performed each time on excluded variables. As a result, earnings and working hours are expressed in logarithms. The common variables included in X are year dummies, dummies for blue-collars in public sector, white-collars in public sector, blue-collars in private sector, white-collars in private sector (civil servants are defined as reference category), dummy for partner, dummy for partner in work, dummies for 1, 2 and +3 children aged 0-17 within the household (0 children is defined as reference category), dummies for being born in a foreign country and having foreign citizenship and their cross-product, level of education. The common list of variables may be shorter for some of the 4 subgroups of individuals (men working full-time, men working part-time, women working full-time, women working part-time) as we exclude non-relevant exogenous variables. There are also some problems of multicollinearity. Thus, if the remaining variables appear non-significant at the individual level, they are jointly significant, which explains their presence in the final equations. The instruments for log(hours) are (second degree) polynomial in age and dummies for 1, 2 or +3 children aged 0-17 within the household. The instruments for log(earnings) include (second degree) polynomial in number of years of experience. All the estimations are done with standard errors clustered at the individual level.

9.4.3.3. Determining income and working hours for self-employed

- Input dataset: '…\nowcasting\SILC\results\long_estim_2years.dta' generated in 'Nowcasting_4_createlongitudinal_SILC.do'.
- Program: 'Nowcasting_6_estimate_earnings_hours.do'
- Output results: 'Nowcasting_Long_wages_hours.log'

a. Estimation sample

As for employees, self-employment incomes are transferred from wave t+1 to the wave t, which results in a removal of the last wave. The remaining sample selection is as follows:

- Being self-employed (workstate = 4)
- Working hours should be strictly positive and non-missing
- Earnings should be non-missing (but can be null)
- Part-time indicator should be non-missing
- Being aged 18-59
- 17 observations with extreme hourly income are excluded (they represent 0.6% of the remaining sample)

b. Methodology

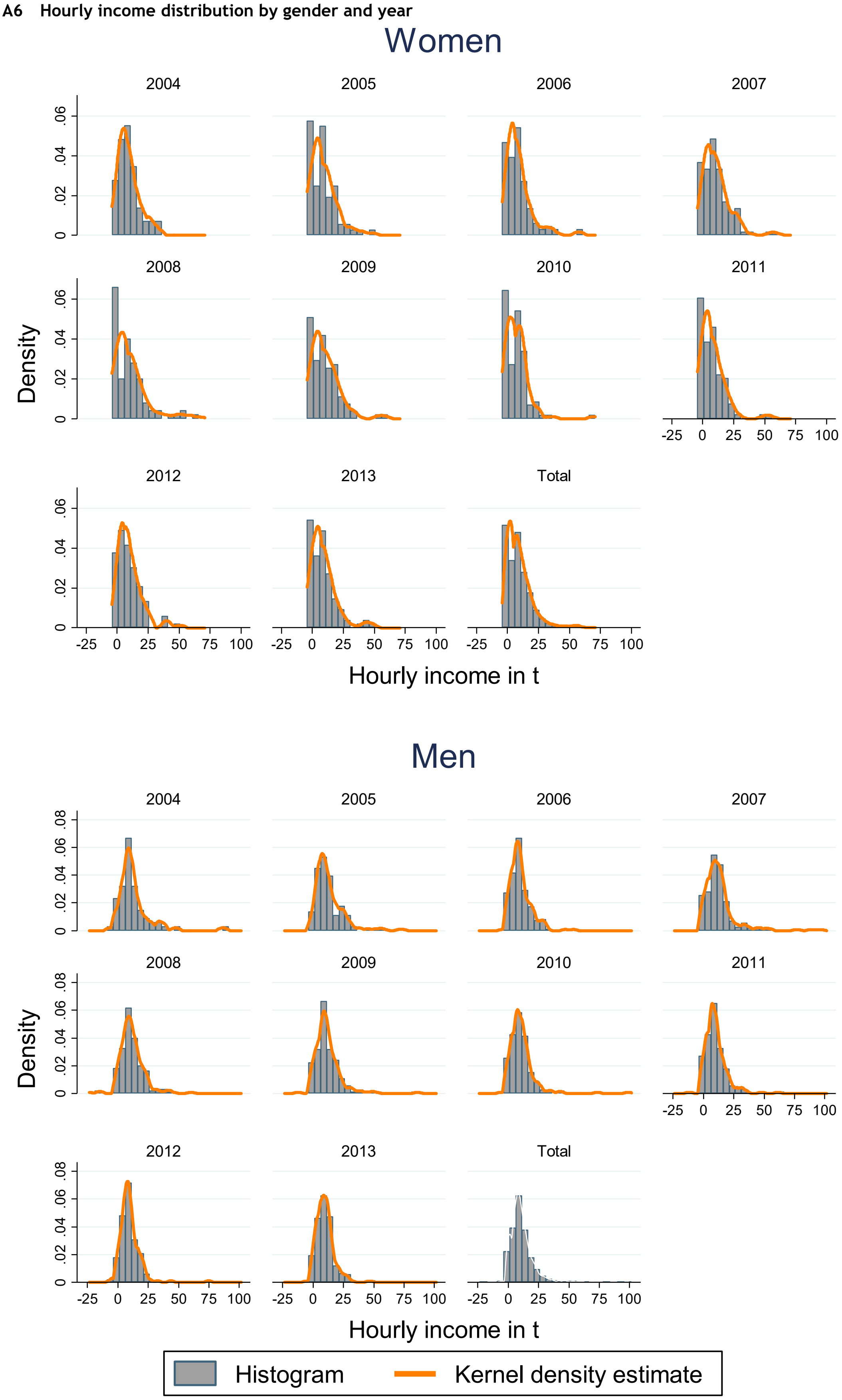
The methodology for self-employed also consists of two steps. First, similarly to employees, we estimate the probability of working part-time that is used to simulate who works part-/full-time. The specification is similar to that for employees, except that men and women share the same equation and gender dummy is added as explanatory variable. The final number of self-employed working part-time is obtained using the predicted probabilities and alignment figures (those are the observed frequencies in the last cross-sectional wave).

In a second step, we choose income and working hours from the observed distributions of four subpopulations of individuals distinguished by gender and work intensity (part-/full-time). More specifically, we separate each the observed working hours and incomes in different classes and compute the fraction of individuals observed in each class. Although the principle is similar for both variables, the procedure itself is slightly different. We start with working hours which is a discontinuous variable with some values observed more frequently than others (e.g. 38, 40, 45, 50). Thus, first, we determine who works exactly 10, 15, 20, 30, 35, 38, 40, 45, 50, 55, 60, 65, 70, 80 or 90 hours. These represent different classes, each including the observed fraction of individuals. Individuals whose working hours are outside of these main classes are split between classes 0 and. The class 0 includes those with less than 40 working hours, while the class 1 has individuals with more than 40 working hours1 (hours distribution within these two classes is more uniform, see "remaining working hours" in Graph A5). All the classes are determined by gender and work intensity. In the Nowcasting model, working hours are simulated only for new self-employed and those who were already self-employed, keep the same working hours as in the previous period. The number of working hours for new self-employed is simulated according to the observed fractions in each class (given the distribution of working hours reported from the previous period for those who remain self-employed). Individuals with simulated class 0 (or 1) are assigned a random number of working hours between 1 and 39 (41 and 99).

Graph A5 Weekly working hours distribution by gender and part-time indicator (whole period 2004-2013)

Regarding self-employment income, we first compute hourly income and determine the observed fractions of individuals with positive, negative and zero hourly income by gender. Those with positive and negative hourly incomes are then divided into classes of €5. The observed hourly income distribution is displayed in Graph A6, by gender and year. Around 20% of women have €0 hourly income, while this percentage is much lower for men, less than 7%. The observed fractions in each class are then used in the Nowcasting model to simulate hourly income. More precisely, we first choose randomly who will have negative, zero and positive income using the observed fractions. Individuals with non-zero income are then attributed a €5-class randomly according to the observed fractions in each class. The resulting simulated income is then increased by a random number from a uniform distribution on]0;5[. Unlike working hours, self-employment income is simulated for all the self-employed (new and those who were already self-employed in the previous period) because of its more unpredictable nature.

Graph A6 Hourly income distribution by gender and year



Notes: The reported year refers to the tax year (when incomes were actually received). Width of bins is €5.

9.4.4. Pensions

– Input dataset: '…\nowcasting\SILC\results\long_estim_2years.dta' generated in 'Nowcasting _4_createlongitudinal_SILC.do'

– Program: 'Nowcasting_6_estimate_earnings_hours.do'

– Output results: 'Nowcasting_Long_income_pensions.log'

a. Estimation sample

Here again incomes are transferred from wave t+1 to the wave t, which results in a removal of the last wave. The remaining sample selection is as follows:

– Being retired (workstate = 9) or other inactive aged at least 65 (workstate = 10 and age >= 65)

– Retirement pensions should be non-missing (but can be null)

– Being aged 60-69

– Education and experience should be non-missing

– Keep only observations with known pension scheme (wage earner, civil servant or self-employed)

– Drop married persons whose spouse is present but spouse's income is missing

b. Methodology

Many retired couples in wage earners or self-employed schemes benefit from the household pension supplement which is paid to the spouse with the highest pension (if the supplement is more generous than the lowest pension). If the spouse with the lowest pension is in civil servants' scheme, then the supplement is reduced by the amount of this pension. Otherwise, the spouse with the lowest pension gets 0. Women are generally the ones with lowest accumulated pension benefits and are more frequently observed to give up their pension in favour of more generous household pension supplement. Therefore, the methodology consists in two steps: we first determine whose pension will be simulated and then the amount of pension.

Thus, first we estimate the probability to have zero retirement pension, for married women not in civil servant scheme. As compared to women, almost all married men have positive retirement pension. We thus assume that for men the probability of having zero amount of pension benefits is 0. The estimated probability for women is used in the Nowcasting model to select, through alignment, married women whose pension benefits will be set to 0 (women with highest estimated probability are selected).[33] The probability of having zero retirement pension is estimated through the following model (where individual *i* and time *t* subscripts are omitted for simplicity)

$$Prob(y = 1 \mid X) = G(X\beta)$$

where $G(X\beta)$ is a logistic function. The model is estimated with standard errors clustered at the individual level. $X\beta$ represents a set of exogenous covariates that are year dummies, (second degree)

[33] Alignment represents the fraction observed in the cross-sectional wave 2014

polynomial in woman's and her husband's age and experience, dummy for retired husband and dummies for husband being in wage earners and self-employed schemes (civil servants' scheme is defined as reference category).

In a second step, we estimate retirement pension for 5 different groups of individuals (all with strictly positive pensions):

- Civil servants
- Non-married wage earners and self-employed
- Married wage earners and self-employed whose spouse receives pension in civil servants' scheme
- Married wage earners and self-employed whose spouse has no income (we assume that those persons receive pension at the household rate)
- Married wage earners and self-employed whose spouse either receives pension but not in civil servants' scheme or does not receive pension but has other income (we assume that those persons receive pension at the single rate).

The equations are estimated by OLS with retirement pensions expressed in logarithm. One of the key factors that explains the amount of retirement pension is the number of years on employment or on replacement income. SILC data contains 2 variables that could approximate this indicator and be taken as explicative variable: work experience and the age at which a person began his first regular job. The latter can be used to compute the total number of years on employment or on replacement income, assuming that the person was in work or received a replacement income since he began his first regular job. In that case, we just take the difference between the current age (limited to 65 for those aged +65) and the age of the first regular job. However, (the polynomial of) this variable appears not statistically significant in most of the equations. If instead we take work experience, it is highly significant. We keep the latter as proxy to the number of years that determine the pension amount. Overall, the estimated equations include all or a subset of the following explicative variables: year dummies, level of education, dummies for being born in a foreign country and having foreign citizenship, (second degree polynomial in) experience, gender, marital status, husband's level of education and experience, dummy for self-employed, dummy for being aged at least 65 and age. The equations may also include cross-products of some of the listed variables.

9.4.5. Estimates used in the taxation module

- Input dataset: '…\nowcasting\SILC\results\Inputdata_Nowcasting.dta' generated in 'Nowcasting _3_startdata_SILC.do'
- Program: 'Nowcasting_9_estimate_tax_difference.do'
- Output results: 'Nowcasting_Cross_net_correction.log'

The estimations are used for two purposes. First, the taxation module implemented in the Nowcasting model generates simulated net incomes that are in average substantially lower than the observed net incomes. This results in overestimation of the simulated average income tax and social security contributions. To address this issue, we estimate the difference between the observed and simulated net amounts for the same period. This allows to adjust net incomes simulated for the next periods so that they are close to the incomes that would be observed in the future SILC waves. Second, the observed data includes information on tax supplement that individuals paid or received the same year for the incomes from t-2. This tax supplement is included in the computation of the equivalised disposable income and thus, needs to be simulated. Therefore, we estimate a regression model where the amount of tax supplement is taken as dependent variable.

a. Estimation sample

The sample is drawn from 2014 cross-sectional wave which is merged with net incomes simulated by the Nowcasting model for the same year. Before we proceed with sample selection, we compute gross total observed, as well as net total observed and simulated individual incomes that include all individual sources of income except social assistance.[34]

Estimating the adjustment of net incomes

The sample is limited to individuals who have positive amounts for the three computed incomes. We also exclude individuals with missing labour market state as it will be used as one of the explicative variables in the estimations.

Estimating tax supplement

Here we take individuals aged at least 18 and remove those with missing labour market state.

b. Methodology

Estimating the adjustment of net incomes

Many individuals are observed with gross incomes equal to their net counterparts. Those are mostly (but not always) low incomes. Thus, in a first step we estimate the probability of having gross income higher than net income for men and women separately (individual *i* subscript is omitted for simplicity):

$$Prob(y = 1 \,|X) = G(X\beta)$$

where $G(X\beta)$ is a logistic function and y takes value of 1 if gross income is higher than net income. $X\beta$ represents a set of covariates that include all or subset of variables (depending on the subsample of men or women): (second degree) polynomial in age, dummy for married or legally cohabiting, dummies for labour market states, logarithm of gross income, cross-product of labour market state dummies with logarithm of gross income, dummy for partner with positive gross income, logarithm of partner's gross income, level of education, dummies for being born in a foreign country and having foreign citizenship and dummies for 1, 2 or +3 children aged 0-17 within the household. In the Nowcasting model, individuals with highest predicted probability are selected through alignment (using proportions that are observed in the 2014 cross-sectional wave).

[34] Social assistance is exempt from income tax and social contributions.

In a second step, we estimate the difference between the observed and simulated net incomes (expressed in thousands) using OLS method. The estimates are obtained for men and women separately, with the following explicative variables specified for both subsamples: (third degree) polynomial for age, dummy for married or legally cohabiting, dummies for labour market states, logarithm of gross income, cross-product of labour market state dummies with logarithm of gross income, dummy for partner with positive gross income, logarithm of partner's gross income, dummies for being born in a foreign country and having foreign citizenship and dummies for 1, 2 or +3 children aged 0-17 within the household.

Estimating tax supplement

Tax supplement is reported in 60% of the cases. Thus, we first need to determine who will get simulated tax supplement. We do this by estimating the following model for men and women separately (individual i subscript is omitted for simplicity):

$$Prob(y = 1 \,|X) = G(X\beta)$$

where $G(X\beta)$ is a logistic function and y takes value of 1 for non-zero reported tax supplement. $X\beta$ represents a set of covariates that are: (third degree) polynomial for age, dummy for married or legally cohabiting, dummies for labour market states, dummy for partner with positive gross income, (second degree) polynomial for person's and her partner's incomes (expressed in thousands), level of education, dummies for being born in a foreign country and having foreign citizenship and dummies for 1, 2 or +3 children aged 0-17 within the household. In the Nowcasting model we use again the predicted probabilities and alignment (using proportions that are observed in the 2014 cross-sectional wave) to select those who will get simulated tax supplement.

Second, we estimate tax supplement through OLS method, by gender, with tax supplement expressed in thousands. The specified covariates are the same as in the logistic function that predicts the probability of having non-zero tax supplement.

9.5. Appendix 5: Exogeneity test

Use *ivreg2* command with *endog*(specify variable to test) to perform the endogeneity test. The hypothesis we test is (i.e. under the null) whether the specified variable can be treated as exogenous. A rejection of the null means that the variables should be treated as endogenous. Example: in a regression of hours on earnings (+ other exogenous variables) we would like to test whether earnings can be treated as exogenous. We thus add the option *endog*(earnings) to the hours equation estimated using *ivreg2*. A statistically significant statistic tells us that we should treat earnings as exogenous.

Example: ivreg2 hours exogvars (lnearny = instruments), endog(lnearny)

Overidentification test:

ivreg2 reports this test automatically (see Hansen J statistic reported after the regression). 'The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation… A rejection casts doubt on the validity of the instruments' (from help for ivreg2 in Stata).

9.6. Appendix 6: Results of logistic regression of severe material deprivation

EU-SILC 2004-2014, Belgium

	Head aged < 60	Head aged ≥60
Head age	0.0846	-0.0822
Head age square	-0.0012	0.0003
AROP (base: No)		
Yes	1.1361	0.5572
Gender head (base: Female)		
Male	-0.2009	-0.0510
AROP and Head Male	0.2234	0.2455
Basic Activity Status Head (base: Working)		
Unemployed	1.4847	1.3450
(Early) retired	-0.5199	0.6869
Other inactive	1.1929	1.3897
Region (base: Brussels)		
Flanders	-1.2697	-2.4937
Wallonia	-0.4053	-0.9975
Unemployed and Flanders	0.2146	1.4196
Unemployed and Wallonia	0.0512	0.3875
(Early) retired and Flanders	0.4013	0.6580
(Early) retired and Wallonia	0.8959	0.3567
Other inactive and Flanders	0.0657	1.0082
Other inactive and Wallonia	0.2493	-0.0543
Unemployed and AROP	-0.1775	-0.3231
(Early) retired and AROP	0.3591	0.2988
Other inactive and AROP	-0.0137	0.1692
Very Low Work Intensity (VLWI) (Base: No)		
Yes	1.2295	
Unemployed and VLWI	-0.7661	
(Early) retired and VLWI	-0.3269	
Other inactive and VLWI	-0.3725	
Male and LWI	0.0614	
Number of Children in the household (base: None)		
One	-0.1312	
Two	-0.2557	
Three	0.0656	
At least four	0.0447	
Adults other than head or spouse present in the household (base: No)		
Yes	0.2327	0.5469
One child and other adult	0.2602	
Two children and other adult	0.6815	
Three children and other adult	0.5184	
At least four children and other adult	0.3887	
Other adult and VLWI	-0.5781	
Other adult and AROP	0.1534	-0.2081
Flanders and AROP	0.1804	0.1437
Wallonie and AROP	-0.0196	0.2270
AROP and VLWI	-0.4569	
Head of household has a partner (base: No)		
Yes	-0.2472	-1.4333
Head has a partner and AROP	0.1040	0.4488
Head has a partner and VLWI	0.0862	
At least one person other than the head of the household works (base: No)		
Yes	-0.5598	-1.0345
One person works and VLWI	0.5319	
One person works and head has a partner	-0.4967	0.4513
Other adult and head has a partner	-0.0974	0.5529
Flanders and head has a partner	-0.3025	-0.5382
Wallonia and head has a partner	-0.3243	-0.5237
Trend	-0.0717	-0.0131
Trend square	0.0085	-0.0028
Trend cube	-0.0004	
Intercept	-3.4471	1.8969
Observations	36992	22635
LR X^2(degrees of freedom)	5151130.50 (50)	1114466.87 (32)
Pseudo R^2	0.3146	0.1885

All covariates are statistically significant at the 1% level. Trend is expressed as 2005 ≤ year ≤ 2014. Trend is expressed as 2005 ≤ year ≤ 2014.