CeMPA Working Paper Series

CeMPA WP 05/22

LABSim: A dynamic life course model of individual life course trajectories for Italy

Matteo Richiardi

Patryk Bronka

March 2022

CeMPA
Centre for Microsimulation and Policy Analysis

LABSim: A dynamic life course model of individual life course trajectories for Italy

Matteo Richiardi and Patryk Bronka

Centre for Microsimulation and Policy Analysis

First version: September 2021

This version: December 2021

1. Introduction

LABSim is a rich dynamic microsimulation model of individual and household life course events, which means that it simulates individual units over time and allows for individual characteristics to be changed according to the processes specified within the model. One of the key innovations in LABSim is its linkage with EUROMOD, a static tax-benefit microsimulation model used to evaluate the immediate distributional impact of policy changes (the "morning after" effect). The static model allows ex-post evaluation of policies, as well as ex-ante evaluation of hypothetical policy changes. When combined with the dynamic model, the policies are applied to an evolving population and evaluated over time.

Users of LABSim can construct as many policy scenarios as they wish by using a bespoke simplified EUROMOD graphical interface, directly called from within the LABSim application.[1] They can then specify which policy scenario to use in each simulated year.

The integration of a static tax-benefit calculator within a dynamic context makes LABSim quite rare in the literature. Indeed, to the best of our knowledge there are only two other models which to some extents are comparable to LABSim. One is LINDA (van de Ven, 2018), developed at the National Institute of Social and Economic Research on UK data, while another one is T-DYMM, the dynamic model developed by the Italian Treasury. LINDA is characterised by a sophisticated modelling of the individual intertemporal decision-making process (labour supply and savings) but covers fewer dimensions than LABSim. As for T-DYMM, very little is known apart from a 2011 project report (Ministry of Treasury, 2011). Looking at the available documentation it appears that the model does not include a health dimension, which is a significant source of life course risk. Internal sources suggest that the model is currently under a major redevelopment.

The structure of this note is as follows. Section 2 describes the inputs required by the model, while Section 3 discusses the model structure and the specification of each process. Section 4 presents validation statistics and discusses model uncertainty. Session 5 analyses some of the model outputs, while Section 6 concludes.

[1] Advanced users can construct the policy scenarios directly from the standard EUROMOD interface, allowing for more flexibility in the creation of the scenario.

2. Input data

LABSim uses three types of data as input:

i) The **initial population** to be evolved over time.
ii) **Donor populations** from the static tax-benefit model (EUROMOD) to provide data on the effects of particular policy schedules. Each year, simulated individuals and households are statistically matched to individuals and households from the appropriate EUROMOD donor population in order to transform gross incomes into net incomes, as specified by the policy schedule.
iii) **Estimated parameters** for the processes modelled in the simulation (further described in the "Processes modelled" section).

The **initial population** comes from EUROMOD input data for 2017, itself derived from IT-SILC, the Survey of Income and Living Conditions. Multiple initial populations are provided coming from different waves of IT-SILC (from 2011 to 2017), to allow the model to start in different years in the past for validation purposes. When users select an initial population that is not the most recent available, validation statistics are displayed comparing the outcome of the simulation with the available data.

The **donor populations** are created by the model itself from the EUROMOD output files (one output file for every different policy system considered in the model), by combining policy-invariant characteristics of donors (for example, gender) with policy-dependent variables (for example, disposable income). Only a single set of policies can be applied in any single year, although the same set of policies can also be applied to multiple years. The simulation allows the policy schedule to be easily modified between runs. For example, in one run of the model, one set of policies can apply in 2011, and in another run a second set of modified policies can be applied to the same year.

Parameters have been estimated on the longitudinal version of IT-SILC for the years 2008-2017 (see Appendix 1).[2]

3. Model structure

LABSim is coded in JAS-mine, a Java-based simulation platform explicitly conceived for dynamic microsimulation and agent-based modelling (Richiardi and Richardson, 2017). As such, LABSim implements a Model-Collector-Observer structure. The Model creates and manages objects and relationships between them and defines the order of events that take place in the simulation. The Collector collects the data from simulated objects and computes statistics, both for use by the simulation and for analysis of the model outcomes after the simulation has completed. The Observer allows the user to inspect the simulation in real time and monitor several pre-defined outcome variables.

Simulated objects are maintained in an underlying database, which can be explored using a *database explorer* tool embedded in the simulation software. The simulated data can also be exported to a set

[2] Education processes do not control for income and also make use of the additional years 2005, 2006, 2007, where income information was not collected.

of csv files (individuals, benefit units, households, aggregate statistics) that create a "synthetic", forward-looking, panel dataset that can be analysed in standard statistical software.

In LABSim, individuals are structured in benefit units (for fiscal purposes), and benefit units are structured in households.[3] The output produced by the model therefore consists of SQL database tables and / or CSV files at the individual, benefit unit, and household level, which can be linked through unique identifiers. The output files contain the values of simulated variables for each individual unit in each year of the simulation, effectively producing a "synthetic" panel dataset.

The model is based on a conditional independence assumption: all processes are modelled as independent; however, they are based on lagged variables determined by other processes.

We use a partial equilibrium model of labour supply, which means that we model labour supply (worker side of the market) but not labour demand (firm side of the market) .[4]

The processes are ordered as in Figure 1; however, as the simulation and the estimated processes are sampled at yearly frequency, the sequence of events within each simulated arbitrary is arbitrary.[5]

LABSim is composed of seven different modules: (i) Demography, (ii) Education, (iii) Health, (iv) Household composition, (v) Non-labour income, (vi) Labour supply, and (vii) Consumption. Each module is in turn composed of different processes or sub-modules, for example ageing process in the demographic module, or a wage setting process in the labour supply module. In each period, agents first go through the ageing process, followed by the population alignment process, which adjusts the population structure to official projections by gender, region, and age. Then, the education module determines whether students should remain in education, or – for individuals who are no longer in education – re-enter education. Students are assumed not to work and therefore do not enter the labour supply module. Individuals who leave education have their level of education determined (for those who returned to education, their level of education can only go up) and can become employed. The health module calculates an individual's continuous health score, and evaluates whether the individual is long-term sick or disabled (in which case, he / she is not at risk of work). Next, in the household composition module, adult children who still live with their parents can leave the parental home, and couples are formed matching the level of homophily between partners observed in the data. Females in couples can then give birth to a child, as determined by the fertility process. Fertility is modelled at the individual level, with an option of alignment to the fertility rate implied by the population projections. Individuals then enter the labour supply module, in which a) their potential wage is calculated using a Heckman-corrected wage equation, b) the closest matching EUROMOD

[3] Naturally, there can be households formed by a single benefit unit, and benefit units composed by a single individual.

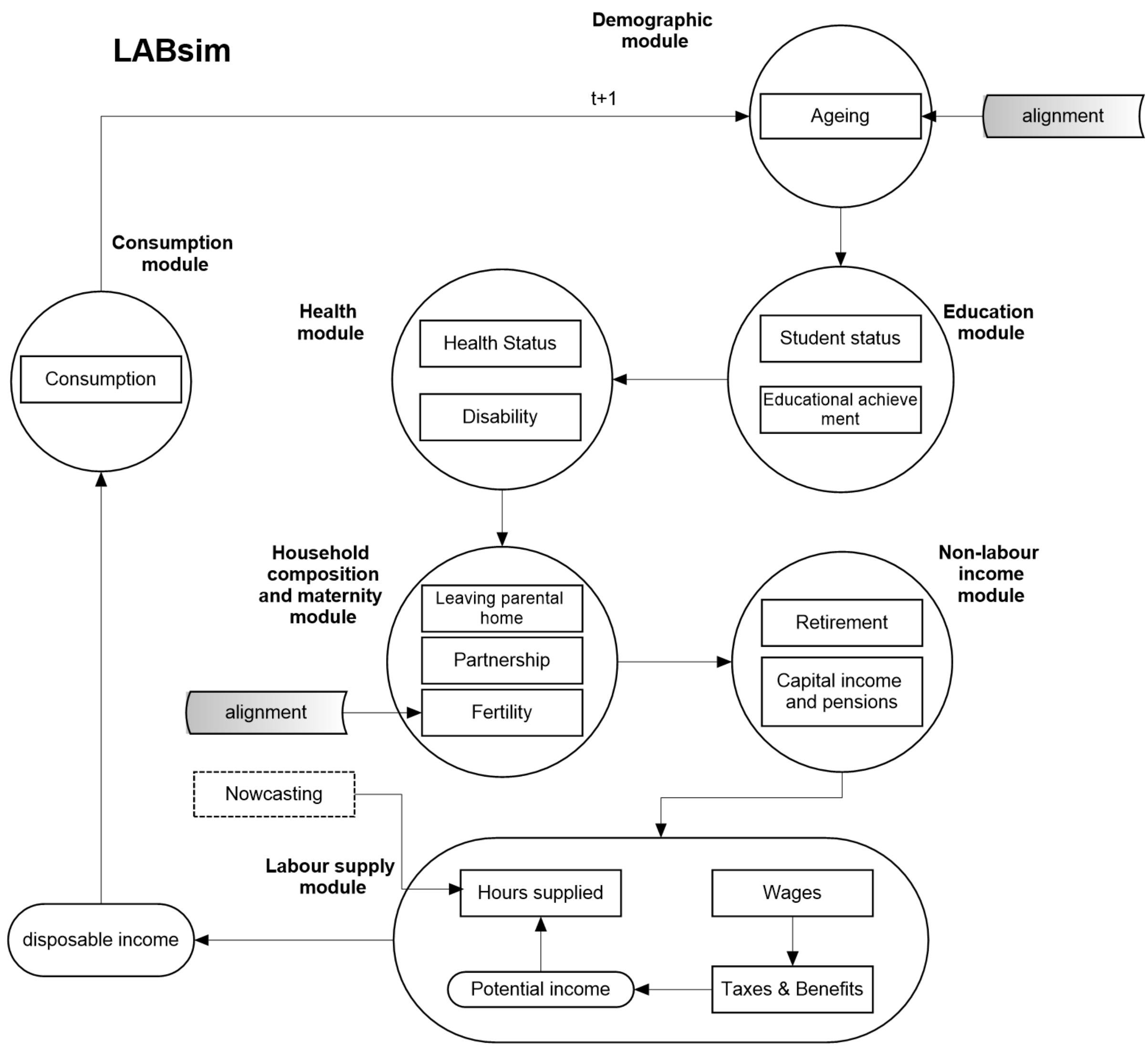
[4] Partial equilibrium analysis is a way of obtaining an estimate of the impact of a change in the economy which does not require a solution of a general equilibrium system. "Partial" refers to the fact that some things in the model are held constant and the demand side of the labour market (firms) can accommodate all workers, who choose the hours of work, given hourly wage determined by a reduced-form model, to maximise their utility. One important consequence of this would be lack of unemployment in this type of models, as "working" 0 hours would be a choice.

[5] The order of the events in the simulation has however some technical implications with respect to how lagged individual characteristics, which are included as controls in most processes, are measured. These lagged values will be the value of the variable in the previous year of the simulation, if that variable has been already updated, or the currently stored value of the variable, if that variable has not been updated yet.

household (in terms of key characteristics relevant for the tax and benefit policies, such as health, number of children, region, and age) is selected and the net income calculated, c) the utility-maximising choice of number of hours of work supplied by the members of the household is determined using a structural labour supply model, whose parameters were estimated on the EUROMOD input data. That determines household's actual disposable income. Finally, a simple consumption module transforms disposable income into consumption by applying an homogenous saving rate, calibrated on the data. The same saving rate is also used when calculating capital income.

Figure 1 shows the structure and order of processes modelled in LABSim. Each year, the simulation begins with the *Demographic module* (which models ageing, leaving parental home, and retirement decision) and ends with the calculation of *disposable income / consumption*. In this section the main processes are described in more detail.

Figure 1. Structure and order of processes modelled in LABSim



The processes included in LABSim are listed in Table 1, and described in more details below.

Table 1. List of estimated processes.

Type	Process name	Content
Education	E1a	Probability of being in education for those in continuous education (who have always been in education without interruptions).
	E1b	Probability of being in education for those not in continuous education.
	E2	Level of education for those leaving education for the first time.
Health	H1a	Self-rated health status for those in continuous education.
	H1b	Self-rated health status for those not in continuous education (out of education or returned having left education in the past).
	H2b	Probability of becoming long-term sick or disabled for those not in continuous education.
Household composition	P1	Probability of leaving the parental home for those who left education. (Students stay in the parental home).
	U1	Probability of entering a partnership for those not in continuous education.
	U2	Probability of partnership break-up for those not in continuous education.
	F1	Probability of giving birth to a child.
Non-labour income	I1a	Probability of receiving capital income for those in continuous education.
	I1b	Probability of receiving capital income for those not in continuous education.
	I2a	Amount of capital income for those in continuous education.
	I2b	Amount of capital income for those not in continuous education.
	R1a	Probability of retiring for single individuals.
	R1b	Probability of retiring for partnered individuals.
	I3	Amount of pension income.
Labour supply	W1a	Heckman corrected wage equation, females
	W1b	Heckman corrected wage equation, males
	LS1	Hours worked, single males
	LS2	Hours worked, single females
	LS3	Hours worked, single male adult children
	LS4	Hours worked, single female adult children
	LS5	Hours worked, males with dependent partner
	LS6	Hours worked, females with dependent partner
	LS7	Hours worked, couples

Some specifications distinguish between students who never experienced a break in their education ("in continuous education"), and individuals who are not in continuous education, that is they are either not in education, or they returned to education after a break. This is because the level of education is assigned only when individuals leave education for the first time, so this variable is not defined for students in continuous education. Processes distinguishing between individuals in continuous education and individuals not in continuous education generally differ only because the latter control for the level of education, while the first don't.

3.1 Demographic module

(A1) Ageing

Every simulated year, the age of individuals increases by 1. Population alignment is then performed to adjust the number of individuals by age, gender, and region to past data and future projections from the Italian statistical agency (Istat). In case where there are too many individuals in a given age-gender-region cell, individuals are removed from the simulation at random. If there are too few individuals, new individuals are created by cloning existing individuals ("donors") from the same age-gender-region cell at random. If no suitable donors are found, the age requirement is gradually relaxed by +/- 1, until a match is found. Alignment is performed at the individual level; however, an attempt is made to recreate the household characteristics of the donors, by finding suitable partners for the clones (if the donors were partnered), and assigning cloned children to cloned mothers (if the donors had children).

3.2 Education module

(E1) Student status

Individuals aged between 16 and 29 who have always been in education consider leaving school with a probability determined by a probit model conditional on sex, age, age squared, mother's and father's education, region, and year. Individuals who are 30 years old and still in education are forced to leave education.

Individuals aged 16 – 35 who are not students can re-enter education with a probability determined by a probit model conditional on sex, age, age squared, lagged level of education, lagged employment status, lagged number of children in the household, lagged number of children aged 0-2 in the household, mother's and father's level of education, region, and year.

Students are not allowed to work. Those who returned to education can leave in any subsequent year, with the possibility of increasing their level of education.

(E2) Educational achievement

Individuals leaving education (as determined in Process E1 above) have a level of education set based on a multinomial probit model conditional on sex, age, age squared, mother's and father's level of education, region, and year. If they returned to education, their level of education cannot decrease.

3.3 Health module

(H1) Self-rated health

The overall health status is based on the self-rated health (5 categories, from Poor to Excellent). Level of health status is determined according to the outcome of a weighted least squares regression, where the level of health is assumed to be linear. The prediction is conditional on sex, age, age squared, lagged household income quintile, lagged health status, region, and year for those in continuous education, and additionally on the level of education, lagged employment status, and lagged household composition for those not in education.

(H2) Disability

Any individual aged 16 and above who is not in continuous education can become disabled or long-term sick with probability given by a probit model conditional on the health status, sex, age, age squared, level of education, lagged household income quintile, lagged self-rated health status, lagged disability status, lagged household composition, region, and year.

3.4 Household composition and maternity module

(P1) Leaving parental home

Individuals who become 18 years old set up new benefit units and consider leaving their parental home, if not in education. The probability of leaving home is based on a probit model conditional on sex, age, age squared, level of education, lagged employment status, lagged household income quintile, region, and year (reflecting time trend observed in the data). Individuals who stay at home become *adult children* and can leave home in any subsequent year.

(U1/U2) Partnership formation

Individuals above 18 who do not have a partner enter a partnership based on the outcome of a probit model conditional on: i) sex, age, age squared, lagged household income quintile, lagged number of children, lagged number of children aged 0 – 2, lagged self-rated health status, region, and year if they are in continuous education, or ii) level of education and lagged employment status in addition to the variables listed in i) if they are not students or are students who have returned to school.

Partnership dissolution is modelled at the benefit unit level with the probability determined by a probit model conditional on female's age, age squared, level of education, lagged personal gross non-benefit income and its square, lagged number of children, lagged number of children aged 0 – 2, lagged self-rated health status, lagged level of education of the spouse, lagged self-rated health status of the spouse, lagged difference between own and spouse's gross, non-benefit income, lagged duration of partnership in years, lagged difference between own and spouse's age, lagged household composition, lagged own and spouse's employment status, region, and year.

Individuals who decide to enter a partnership are matched using either a parametric or non-parametric process. In the (default) parametric process, males are matched with females with a probability corresponding to a matching score, calculated in the following way:

$matching\ score = earningsScore^2 + ageScore^2$, where

$earningsScore = (male's\ potential\ earnings - female's\ potential\ earnings) - female's\ desired\ difference\ in\ potential\ earnings$, and

$ageScore = (male's\ age - female's\ age) - male's\ desired\ difference\ in\ age.$

The non-parametric process aims to replicate the distribution of matches observed in the data between different types of individuals, where a type is defined as a combination of sex, region, education level, and age. The distribution of matches between different types is adjusted using an iterative proportional fitting procedure constrained by the marginal frequencies of each type observed in the simulation.

(F1) Fertility

Females aged 18 to 44 who have a partner can give birth to a child with a probability determined by a probit model conditional on: i) age, age squared, the overall fertility rate of the Italian population (in a given year, projected for future years), lagged household income quintile, lagged number of children, lagged number of children aged 0 – 2, lagged health status, and lagged partnership status for those aged 18 to 29 who were in continuous education, and ii) lagged employment status, level of e ducation, and region for those who were not in continuous education. The inclusion of the overall fertility rate implies this is a model of differential fertility, where the overall change in fertility projected by the statistical authority is distributed across individuals according to their observable characteristics.

3.5 Non-labour income module

(I1/I2) Capital income

Individuals above 16 years old receive non-employment non-benefit income (capital income) determined by an outcome of a weighted least squares regression, conditional on i) sex, age, age squared, lagged health status, lagged gross employment income, lagged capital income, region, and year if they are in continuous education, or ii) (in addition to variables used in i)) level of education, lagged employment status, and lagged household composition.

(R1/I3) Retirement and pensions

Individuals above 50 years old consider retirement, with the probability based on a probit model estimated separately for couples and singles. These control for sex, age, age squared, level of education, dummy indicating if individual is above state pension age (allowing for past an d planned changes in the state pension age), lagged employment status, lagged household income quintile, lagged disability status, dummy indicating if spouse is above state pension age, lagged employment status of the spouse, interaction term between reaching pension age and lagged employment status, and year. For single individuals, the variables relating to the status of the spouse are omitted from the model. Retired individuals do not work and retirement is considered an absorbing state (retired individuals cannot return to work). Retirement before reaching state pension age is allowed, to match the observed distribution of age at retirement.

Retired individuals receive pension income determined by an outcome of a weighted least squares regression conditional on sex, age, age squared, level of education, lagged employment status, lagged household composition, lagged health status, lagged gross employment income, lagged capital income, region, and year.

3.6 Labour supply and labour income module

(LS1-7) Structural labour supply module

To simulate labour supply, we use a random utility labour supply model. This is a static labour supply model, in a sense that labour demand and wage rates are assumed fixed, but a detailed representation of the individual and household budget constraint is considered. Individuals maximise their utility from income and leisure over a restricted number of alternatives (5 for individuals, and 25 for couples as utility is determined at the benefit unit level). Utility is uniperiodal (although inter-temporal considerations might be captured by time varying characteristics such as age). The model is unitary, which means that decision making is at the level of the benefit unit – for example, one single choice involving labour supply decisions for each partner is made by couples, in order to maximise their joint utility. The utility of the benefit unit is calculated using a quadratic utility function with fixed costs and depends on disposable benefit unit income, and on the number of hours worked by each benefit unit component. Labour market income is computed, for each of the alternatives and for each individual component of the benefit unit, by multiplying the number of hours for a predicted hourly wage, as estimated by means of a Heckman-corrected wage equation. Individual labour market income is then transformed into disposable benefit unit income by following a procedure described below, in the "Disposable income and consumption" subsection.

The labour supply model is described in more details in Richiardi and He (2021).

3.7 Disposable income and consumption

Disposable income is calculated after the labour supply module, by multiplying the benefit unit's gross income (both labour income and non-labour, non-benefit income) by a ratio of disposable income to gross earnings of a closest matching *donor.* Donor benefit units are obtained from the EUROMOD population, and their disposable to gross income ratio depends on the tax-benefit schedule in place in a given year (specified and adjustable at the beginning of the simulation). Simulated benefit units are matched to donor benefit units on a number of key characteristics: labour supply hours of adult members of the benefit unit, health status, number of dependent children, region of residence, and age of adult members. We first attempt to find an exact match on all characteristics, but if there are no matching observations, we relax the requirement of exact matching characteristics one-by-one (starting with age), instead replacing it with a minimum distance procedure. For example, if no exact match can be found on all key characteristics, we will look for an exact match on labour supply, health, number of children and region, and attempt to minimise the difference in age between the simulated and donor benefit unit.

Yearly equivalised disposable income is calculated by adjusting the sum of monthly disposable income by benefit unit composition by equivalised weight, calculated using the OECD-modified equivalence scale. Yearly equivalised consumption is equal to yearly equivalised disposable income for retired individuals, and to (1 - saving rate)*yearly equivalised disposable income otherwise. By default, the saving rate is calibrated to its long-term average. However, users can modify this parameter. Changes in this parameter will be reflected in a proportional adjustment of the predicted capital income.

4. Uncertainty assessment and validation

Stochasticity is incorporated in LABSim in the following way: i) for linear models, a Gaussian random number, multiplied by the residual standard deviation of the regression is added to the calculated score[6]; ii) for logit and probit models, the probability of an event is calculated by drawing a Boolean whose value is true with probability equal to the logit or probit transforms of the linear regression score of the corresponding model; iii) in case of multinomial logit and probit models - used to

[6] The score is simply the sum of each coefficient multiplied by the value of the corresponding covariate.

determine the outcome taken from a finite set of possible outcomes - the logistic or probit distributions transform the linear regression scores for N-1 outcomes compared to a reference group which has a score of 0. From this, relative probabilities of outcomes are created, which can then be sampled to determine which of the N outcomes occurs.

To address the issue of parameter uncertainty, regression coefficients of the model can be bootstrapped. Bootstrapping involves sampling the set of regression coefficients of a regression object from a multivariate normal distribution whose vector of expected values (means) are the set of regression coefficients estimated from the data, with the covariance matrix derived from the statistical error of the estimates.[7] With bootstrapping on, each simulation run uses values of the parameters randomly drawn from the joint distribution of the estimated coefficients, for each process (as detailed in the estimated variance-covariance matrix). Hence, unless the seed of the random number generator controlling the stochastic elements in the simulation is kept fixed across simulation runs, each run will produce slightly different results, allowing for an assessment of the uncertainty coming from estimation sample size errors.

The model can be initiated using initial populations extracted from different waves of IT-SILC data. Normally, users would select the latest initial population available (currently, 2017). However, selecting earlier initial populations allows to use the period of overlap between the simulated and observed data for validation purposes. Two approaches can be used for validation. According to a first approach, the parameters remain unchanged at the value estimated using the whole available period of observation, irrespective of the initial population used. The alternative is to estimate the parameters using data only up to the chosen initial population, which poses a stronger validation test. However, given the model complexity, even the *fixed parameters* approach is challenging, as there is nothing that guarantees that when all the (separately estimated) processes are put together the model outputs remain on track with the observations. Moreover, restricting the estimation sample to a shorter period would seriously undermine our ability to detect meaningful time trends, which ultimately prompts us to opt for the fixed parameters approach.

Figures 2-11 show some validation statistics, initiating the model with the 2011 population and comparing the simulated statistics with the same statistics computed on IT-SILC data for the period 2011-2017. In all the graphs, the solid lines represent model outcome, while the dashed lines represent observed statistics. These validation statistics are computed on a single run of the model with a simulated population of 10,000 individuals, without bootstrapping of the coefficients.[8] Overall, the model is able to reproduce the observed data remarkably well.

[7] For a methodological discussion of uncertainty in microsimulation models, see Appendix 2.
[8]

Figure 2. Validation: Educational levels

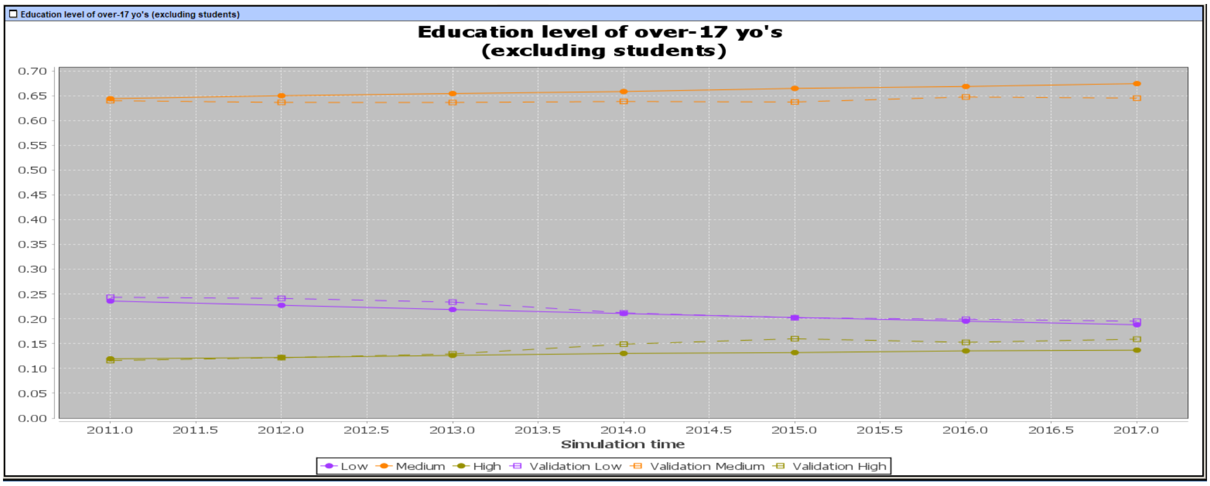


Figure 3. Validation: Proportion of students by age

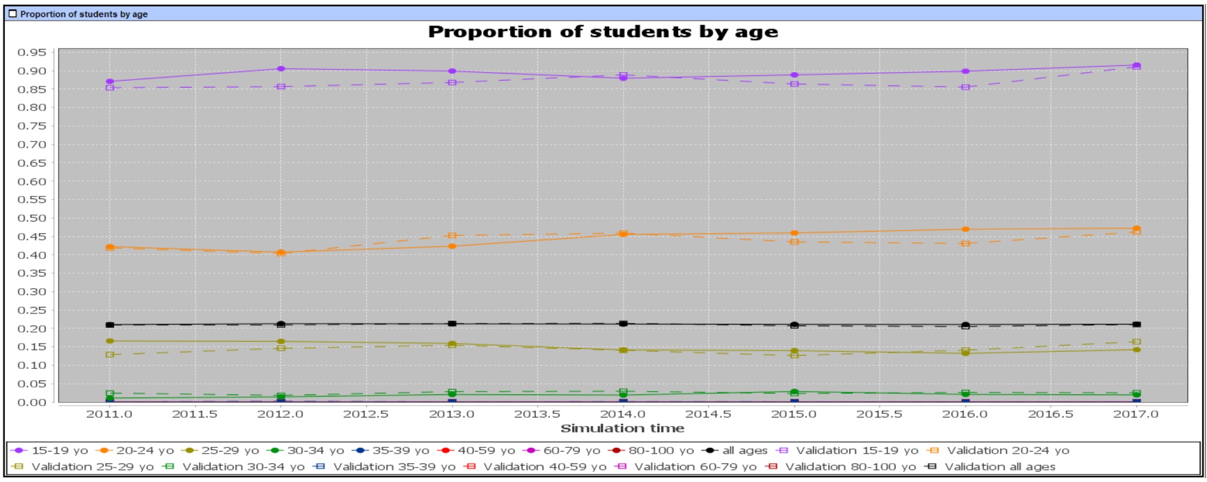


Figure 4. Validation: Proportion of students by region

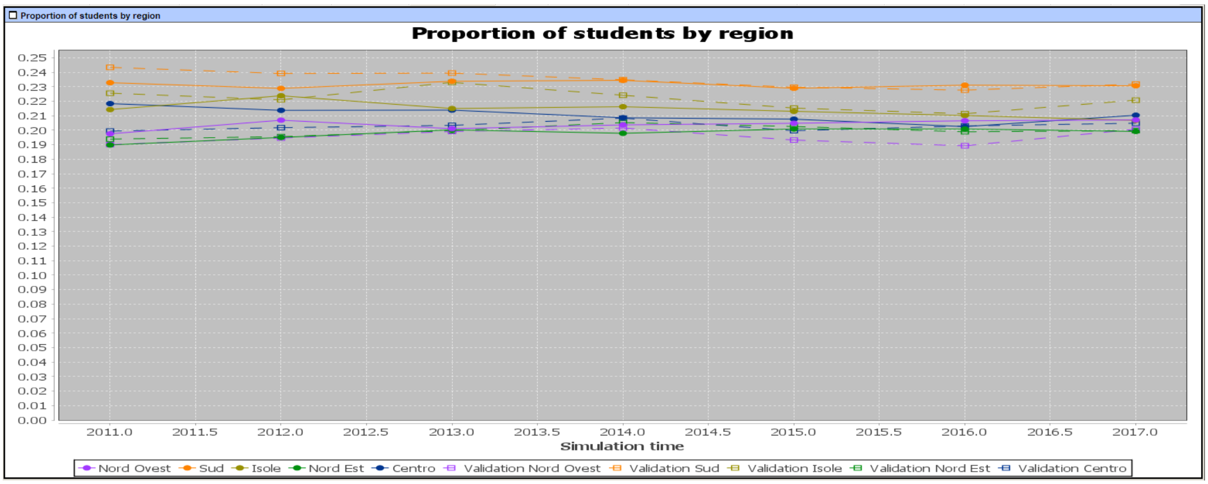


Figure 5. Validation: Proportion of cohabiting individuals

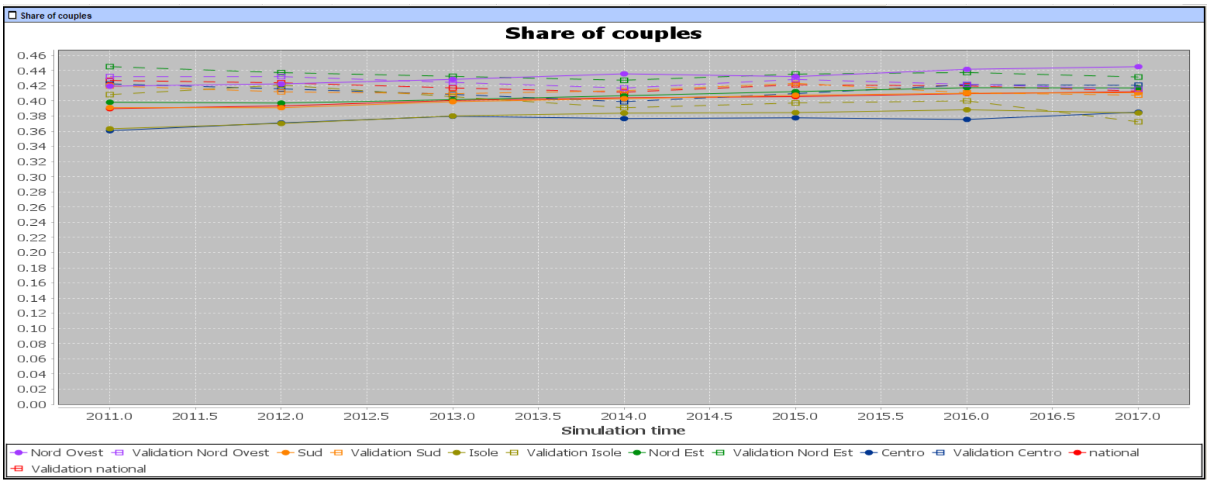


Figure 6. Validation: Health status

Figure 7. Validation: Disability status

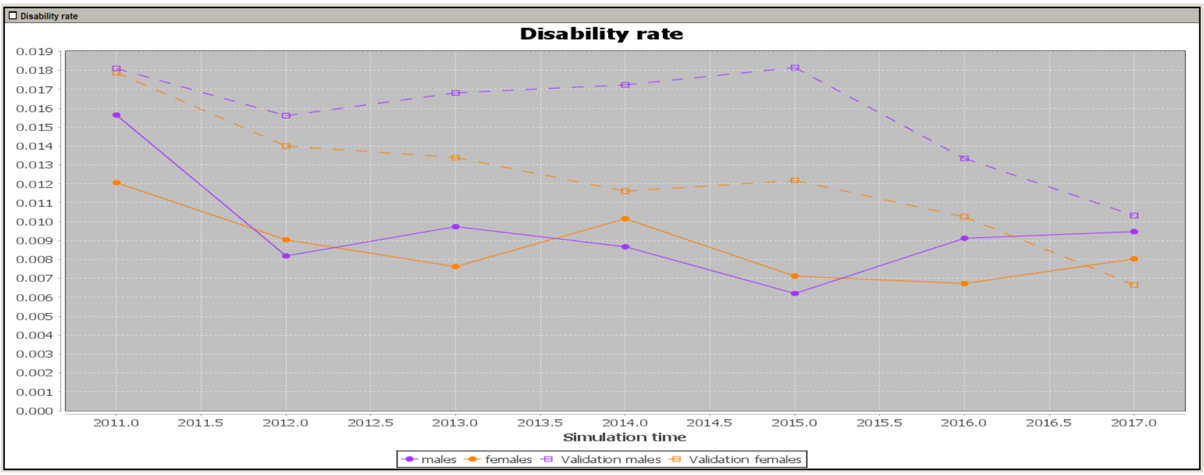


Figure 8. Validation: Activity status

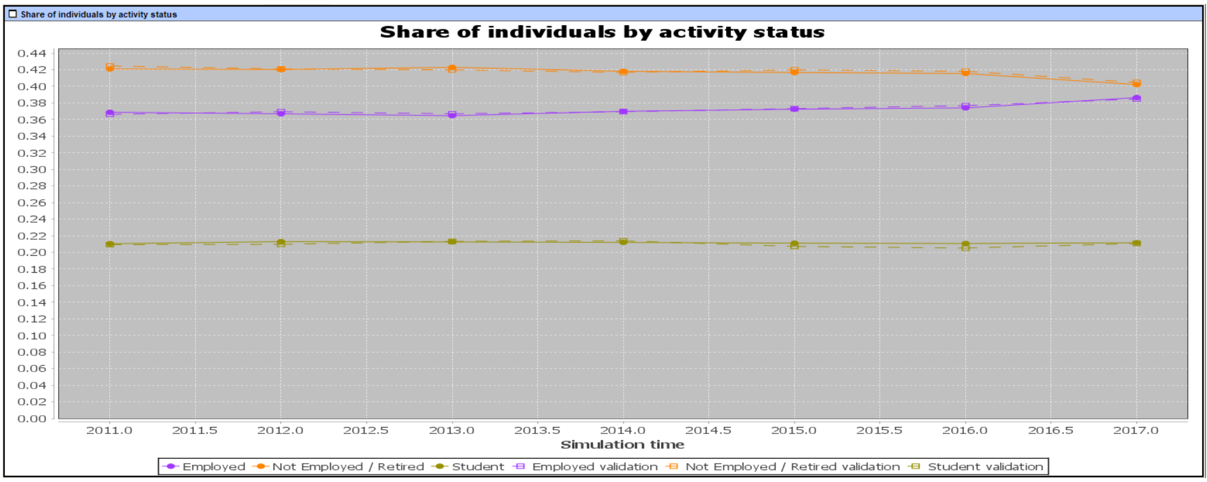


Figure 9. Validation: Employment rate by gender

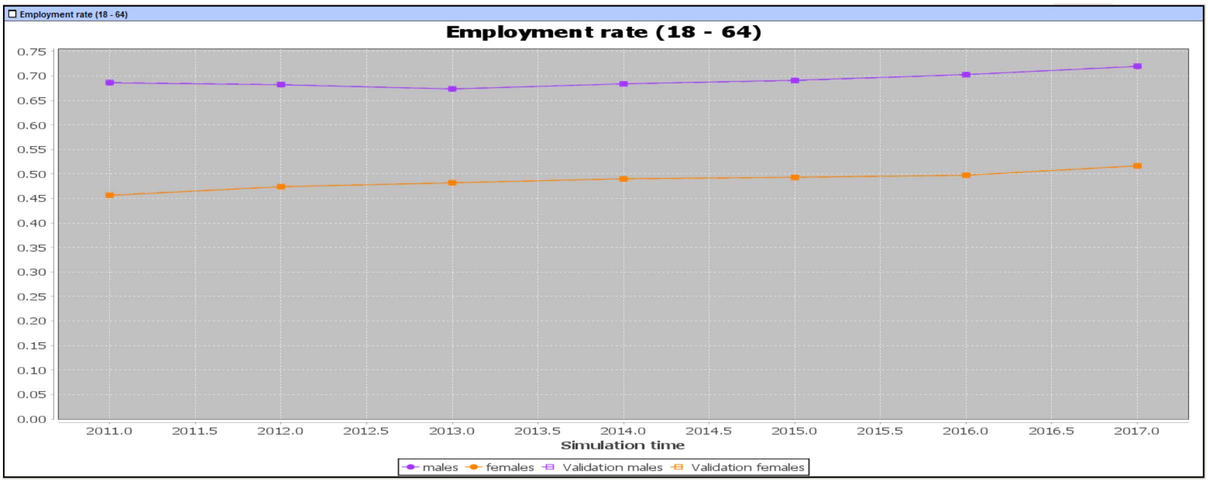


Figure 10. Validation: Employment rate by age

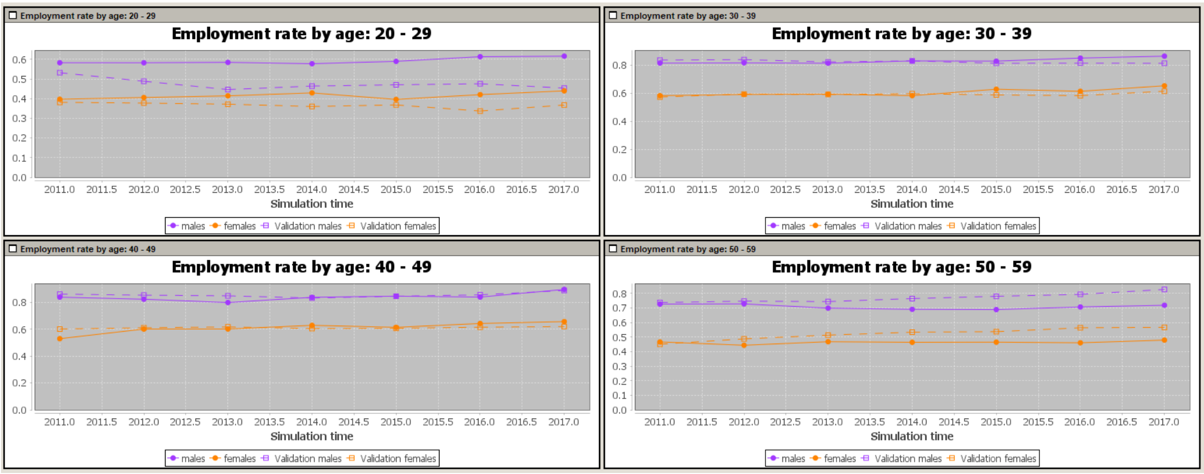
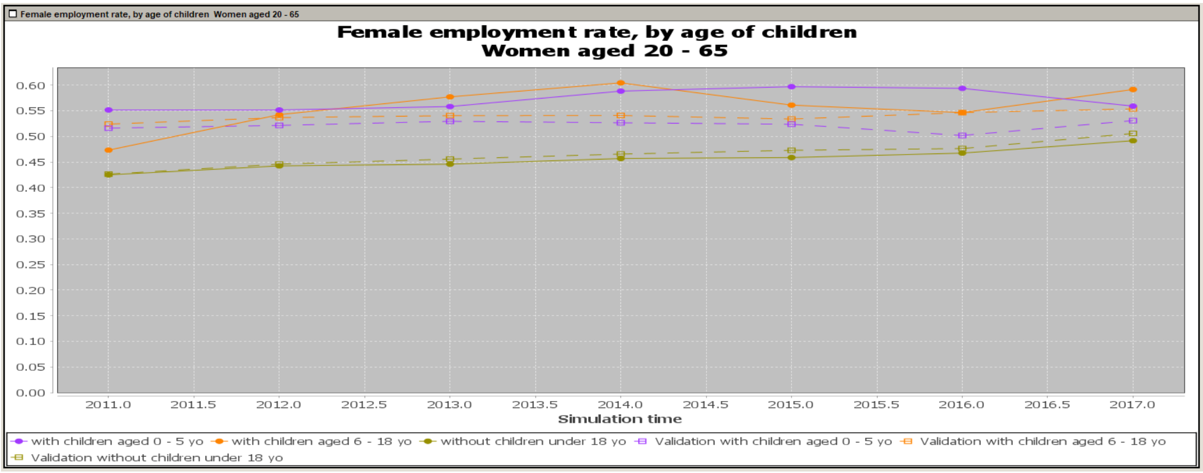


Figure 11. Validation: Female employment rate by age of children



5. Simulation results

In presenting the model results, it must be stressed that dynamic microsimulation models aim not at predicting the future, but rather at highlighting a direction in which the different forces at play in the model, as estimated in the data, are pulling. Each force might pull in a different direction, and different forces might interact with each other, either through reinforcing or attenuating mechanisms. The overall interplay of all the forces considered, and their overall effects in the medium and longer term do not show up in shorter term analyses, when each process is considered separately. The microsimulation model acts therefore as a synthesis of the available evidence. In the real world, other factors will surely come to play that are not considered in the model. Moreover, changes in behavioural patterns and in policies will surely be triggered, possibly by the same dynamics at play in the model. What the model does is identifying overall trends and criticalities, which might call for further examination and analysis.

5.1 Population ageing

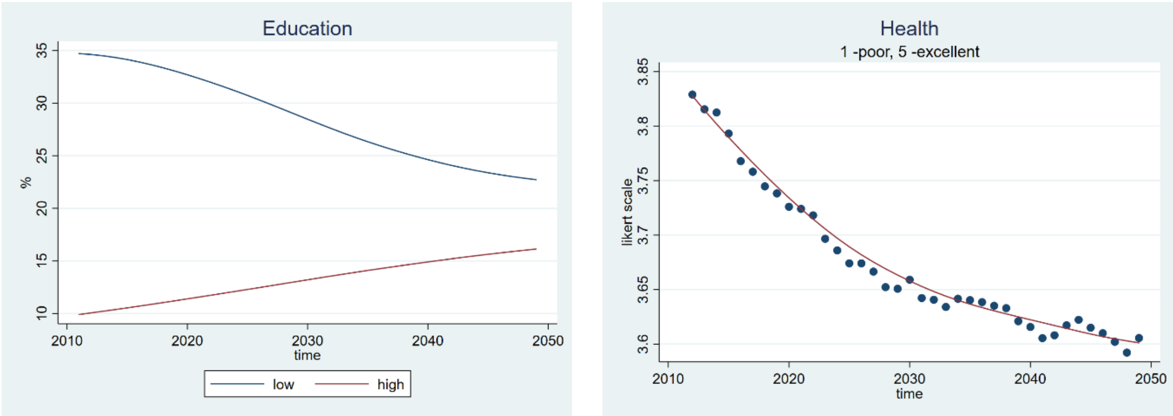
The model replicates the demographic transition described in the official population projections, with an old age dependency ratio (people aged 65+ divided by people aged 18-64) passing 60% in the 2040s (Figure 12).

Figure 12. Model output: Population structure

5.2 Education and health

The model predicts an increase of the share of people with high education (Figure 13, left panel), and an even more pronounced decrease in the share of people with low education. This is due to two factors: the passing away of older generations characterised by lower educational levels, and the continuation of the current trend towards increasing levels of education.

Figure 13. Model output: Education and health

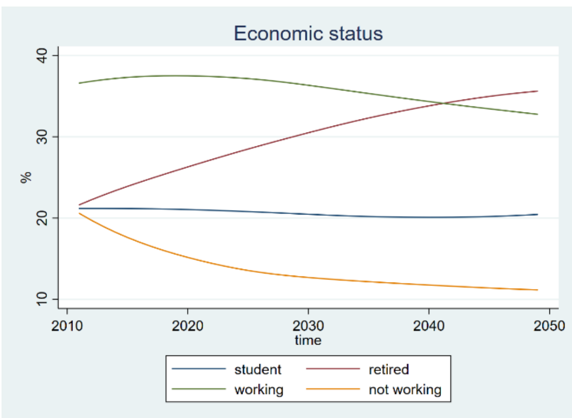


At the same time, population ageing is a powerful force that will drive down population health (Figure 13, right panel), thus posing difficult challenges for the health system and more broadly the welfare state.

5.3 Employment and income

Figure 14 describes the evolution of the labour force, distinguishing between the employed, not employed, retired and student population. The reduction in the labour force is striking, mainly driven by the increase in pensioners. On the other hand, the share of inactive working age population (not working and not retired) goes steadily down.

Figure 14. Model output: Economic status



The current version of the model contains no growth. Incomes are expressed in real terms (2015 levels) and the wage structure is constant. However, the projected evolution of the characteristics of the population, and in particular population ageing and the increase in the levels of education, are pushing incomes up, both market incomes (Figure 15, left panel) and equivalised disposable income (right panel). To be noted, the evolution of equivalised disposable income is also driven by changes in household structure.

Figure 15. Model output: Education and health

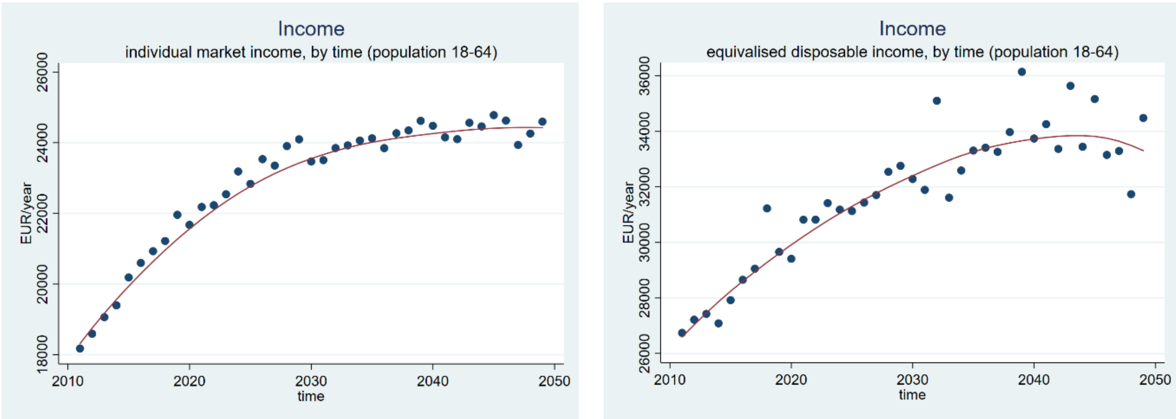
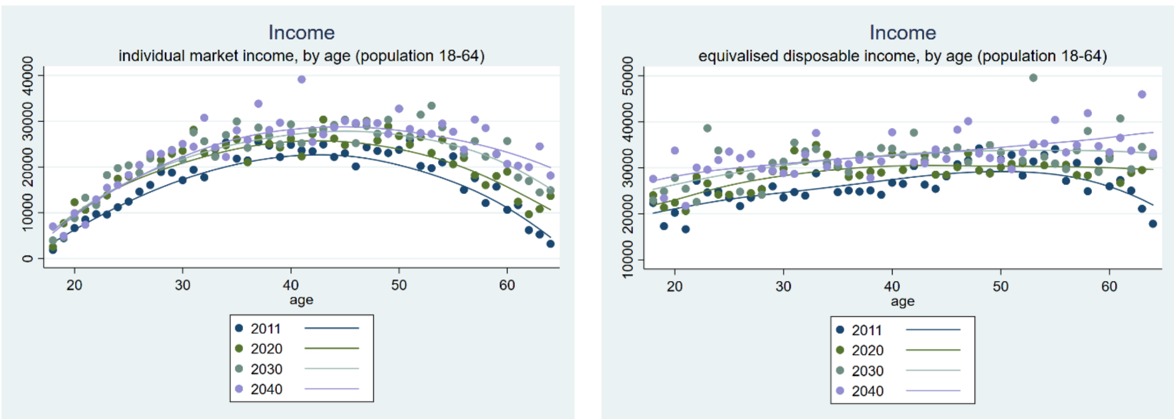


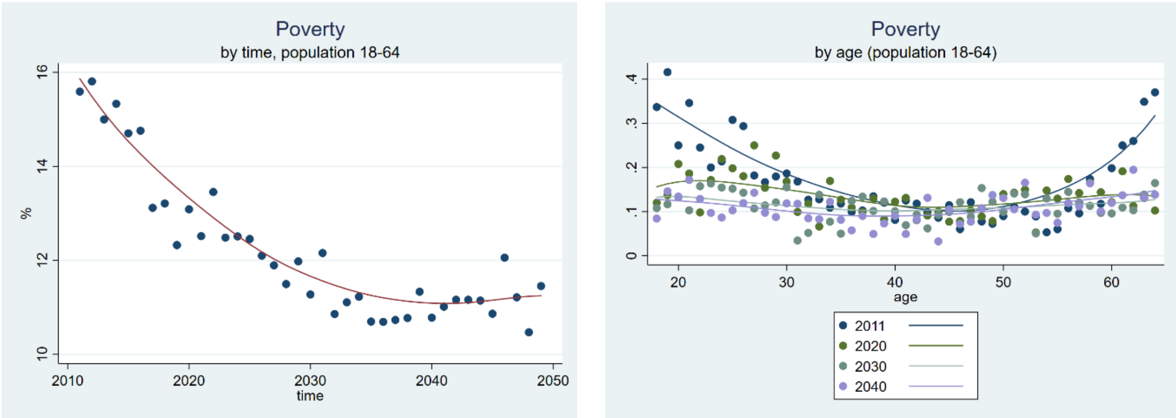
Figure 16 shows the evolution of age profiles for market and equivalised incomes. Both are shifting upwards over time, coherently with the overall increase in real incomes. However, the age profiles in equivalised income are less pronounced, due to the impact of household structure (bigger family sizes in prime age).

Figure 16. Model output: Age and time income profiles



For the same reasons why incomes grow, the model predicts poverty to reduce. Poverty age profiles also become flatter (Figure 17), a pattern attributable in the reduction of bigger size households.

Figure 17. Model output: Age and time poverty profiles

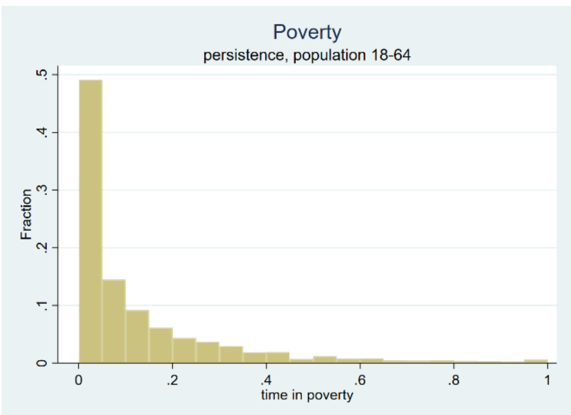


Having an artificial panel covering the future life course of all the individuals in the initial population allows analyses that are normally precluded by the longitudinal size of the available panel data (e.g. 4 years only for EU-SILC).

As an example of such analyses, in the population 18-65, the tax-benefit system causes a reduction of 10 Gini points in the Gini coefficient - from 0.51 (market incomes) to 0.41 (equivalised disposable income). Also, the longitudinal dimension of the microsimulation allows to compute a measure of inequality in future earnings. In the population 18-64, the Gini coefficient for the present value of future market incomes is 0.44, smaller than that of cross-sectional incomes. The Gini coefficient for the present value of future equivalised disposable income is 0.35.

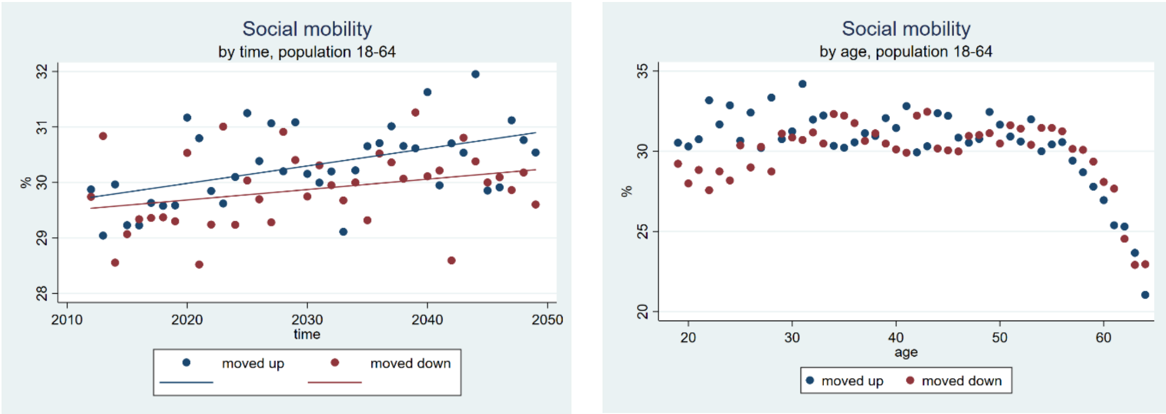
Another statistics that can be computed on the artificial panel is the distribution of future time in poverty (Figure 18).

Figure 18. Model output: Time in poverty



Finally, Figure 19 describes the evolution of social mobility, defined as the probability of moving up or down between quintiles of equivalised income.

Figure 19. Model output: Social mobility



The model predicts a slight increase in social mobility over time. This is due to a higher mobility associated to higher educational levels, as well as a trend towards increasing fragmentation in working and family lives, only partially compensated by population ageing and the lower mobility associated to older ages (Figure 19, right panel).

6. Application: Economic insecurity

As a first application of the model, we compute the measure of economic security proposed in Richiardi and He (2020a). This refers to a normative evaluation of the expected stream of resources available to any individual over the course of his or her residual lifetime, appropriately discounted for family composition and time, and is defined as

$$S_i = \frac{\int_t \int_z W(z)\delta^t f_{i,t}(z) p_{i,t} dz dt}{\int_t \delta^t p_{i,t} dt} \tag{1}$$

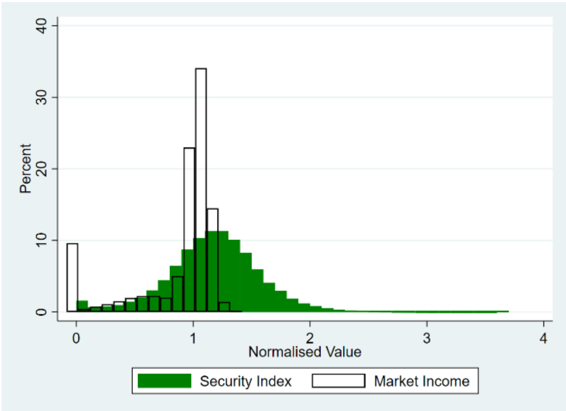
where $W(z)$ is the welfare associated to the resources available to individual i at time t, with $W' > 0$ and $W'' > 0$ (concave function), δ is the discount factor, $f_{i,t}(z)$ is the density at which resources are available, and $p_{i,t}$ is the probability that individual i is alive at time t.

We operationalise this measure considering equivalised consumption as resources z, and a welfare function of the form

$$\mathrm{W}(z) = \sqrt{z} \tag{2}$$

which implies that volatility in resources is penalised (see Richiardi and He, 2020a for a discussion). Further, as a simplification we compute the integrals, for each individual, over the first 10 simulated years only, rather than the whole residual lifetime. This measure of security is then normalised by the median level of security in the population in the initial period of the simulation, so that individuals with a level of security above 1 are in the top half of the distribution of security, and individuals with a level of security below 1 are in the bottom half. Figure 20 depicts the distribution of security vis-a'-vis the distribution of market income (the latter is also normalised by its median).

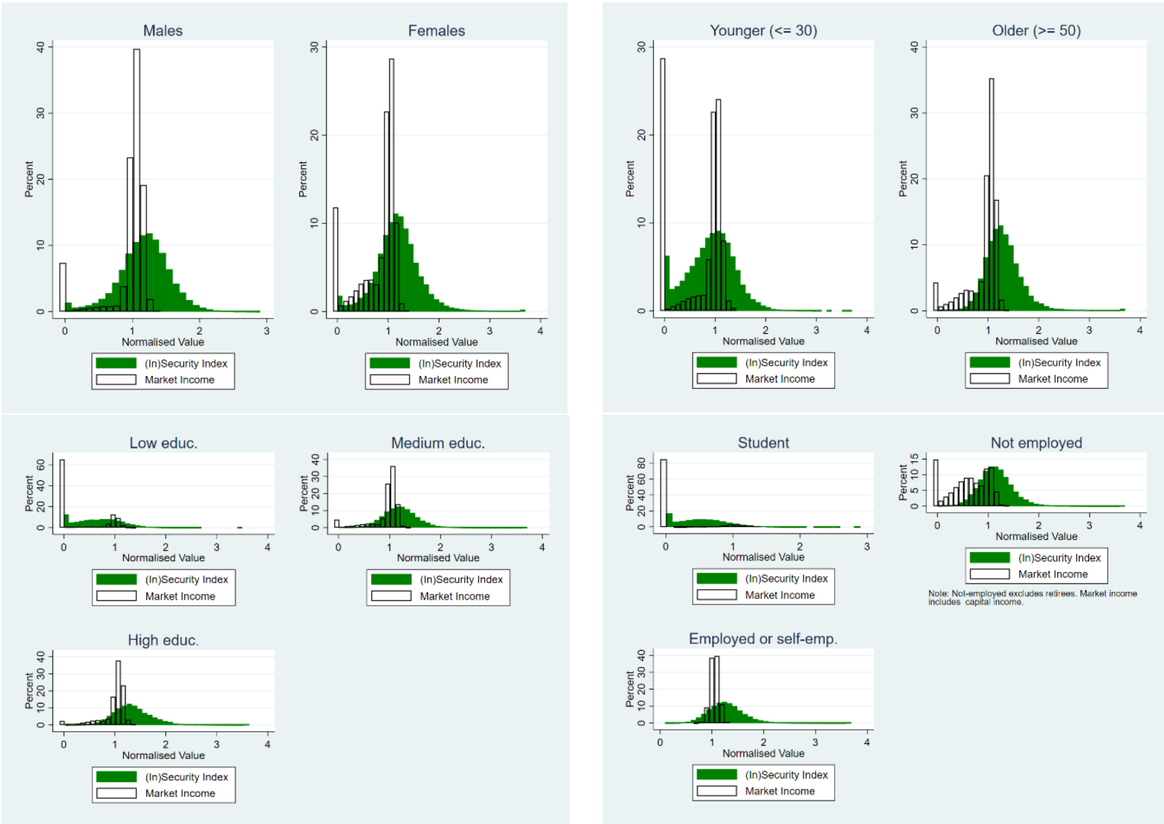
Figure 20. Model output: Distribution of economic security and market income



We can see that security is more evenly distributed than market income. This equalising effect is brought about by four factors: (i) our measure of security is a time average of available resources, so there is regression to the mean, (ii) resources are computed post- taxes and benefits, (iii) they are equivalised within the household, hence allowing for resource pooling, and (iv) they are evaluated using a convex welfare function, which reduces the impact of occurrences of extreme levels.

Figure 21 shows the distribution of economic security and market income by a number of individual characteristics (gender, age, education and economic activity). In all groups security is more equally distributed than market incomes. To be noted, economic security is more or less equally distributed in women and men. On the other hand, economic security is higher and more equally distributed among the elderly (as opposed to young individuals), and among individuals with medium and high education (as opposed to individuals with low education). Importantly, this approach to measuring economic security allows to attach a level of security to individuals who are normally excluded from the analysis using other measures based on current and past income/wealth (Richiardi and He, 2020b), and in particular students. Although students might not have an income today, their future prospects should be taken into account when discussing their level of security, or lack thereof.

Figure 21. Model output: Distribution of economic security and market income by gender, age, education and economic activity



7. Conclusions

In this note we have described the structure and estimation for Italy of LABSim, an integrated static-dynamic microsimulation model of individual life course events. We also presented some validation results, and an illustrative description of some of the results produced by the model, including an innovative measure of economic security that allows to take into account individual future prospects. Our results allow to characterise the implications over the medium and long-term of trends and dynamics currently at play in the Italian society. Much more can be extracted from the model. In particular, the integration of the static tax-benefit calculator EUROMOD within the dynamic structure of LABSim allows to analyse in details the short-, medium- and long-term effects of fiscal policies, and explore the challenges to the welfare state in the context of rapid population ageing and changing family dynamics.

Appendix 1. Estimates

Table A1. Process E1a: Probability of being in education. Sample: Individuals aged 16-29 in continuous education.[9]

In education	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	-0.15	0.00
Age	-0.52	0.00
Age Squared	0.01	0.00
Mother's Education Level (Ref = High)		
Medium	-0.26	0.00
Low	-0.24	0.00
Father's Education Level (Ref = High)		
Medium	-0.33	0.00
Low	-0.50	0.00
Region (Ref = Central Italy)		
Northwest Italy	-0.07	0.06
Northeast Italy	-0.09	0.02
South Italy	-0.05	0.19
Insular Italy	-0.08	0.12
Time	-0.02	0.00
Constant	41.15	0.00
Number of obs	30,781	
Wald chi2(12)	1049.77	
Prob > chi2	0.0000	
Pseudo R2	0.0728	
Log pseudolikelihood	-48052757.00	

[9] Here and below, “in continuous education” means having always been in education without interruptions.

Table A2. Process E1b: Probability of being in education. Sample: Individuals aged 16-35 not in continuous education.

In education	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.09	0.29
Age	0.36	0.01
Age Squared	-0.01	0.00
Lagged Level of Education (Ref = High)		
Medium	-0.43	0.00
Low	-0.84	0.00
Lagged Labour Force Status (Ref = Employed)		
Student	3.17	0.00
Not Employed	0.25	0.03
Lagged Number of Children in the Household	0.02	0.92
Mother's Education Level (Ref = High)		
Medium	0.56	0.01
Low	0.65	0.01
Father's Education Level (Ref = High)		
Medium	0.00	0.99
Low	0.11	0.55
Region (Ref = Central Italy)		
Northwest Italy	-0.16	0.17
Northeast Italy	-0.25	0.03
South Italy	-0.16	0.20
Insular Italy	-0.33	0.03
Time	0.04	0.00
Constant	-93.71	0.00
Number of obs	22,673	
Wald chi2(17)	901.12	
Prob > chi2	0.0000	
Pseudo R2	0.5441	
Log pseudolikelihood	-4896413.70	

Table A3. Process E2: Educational attainment. Sample: Respondents from Process 1a who have left education.

Educational Attainment	***Coef.***	***P>z***
High		
Gender (Ref = Women)		
Men	-0.13	0.01
Age	3.29	0.00
Age Squared	-0.06	0.00
Mother's Education Level (Ref = High)		
Medium	-0.12	0.09
Low	0.00	0.97
Father's Education Level (Ref = High)		
Medium	-0.19	0.00
Low	-0.54	0.00
Region (Ref = Central Italy)		
Northwest Italy	0.44	0.00
Northeast Italy	0.57	0.00
South Italy	-0.11	0.09
Insular Italy	-0.09	0.33
Time	0.04	0.00
Constant	-117.81	0.00
Medium (Reference)		
Low		
Gender (Ref = Women)		
Men	0.30	0.00
Age	-0.75	0.00
Age Squared	0.02	0.00
Mother's Education Level (Ref = High)		
Medium	0.00	0.98
Low	0.46	0.02
Father's Education Level (Ref = High)		
Medium	0.31	0.04
Low	0.30	0.15
Region (Ref = Central Italy)		
Northwest Italy	0.16	0.23
Northeast Italy	0.42	0.00
South Italy	0.18	0.16
Insular Italy	0.50	0.00
Time	-0.04	0.00
Constant	87.40	0.00
Number of obs =	30,752	
Wald chi2(34)	1671.80	
Prob > chi2	0.0000	
Log pseudolikelihood	-35924765.00	

Table A4. Process H1a: Self-rated Health Status. Sample: Individuals aged 16-29 in continuous education.

Self-rated health status	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.04	0.00
Age	0.03	0.26
Age Squared	0.00	0.21
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	-0.01	0.56
3rd Quintile	-0.02	0.57
4th Quintile	0.02	0.60
5th Quintile	0.02	0.40
Lagged self-rated health status	0.36	0.00
Region (Ref = Central)		
Northwest	-0.03	0.05
Northeast	-0.02	0.21
South	0.05	0.00
Insular	0.01	0.63
Year	0.00	0.07
Constant	10.19	0.02
Number of obs	12,074	
F(13, 17132)	52.34	
Prob > F	0.0000	
R2	0.1410	

Table A5. Process H1b: Self-rated Health Status. Sample: Individuals aged 16 or older not in continuous education.

Self-rated health status	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.02	0.00
Age	0.00	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	-0.05	0.00
Low	-0.15	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.03	0.02
Not Employed	-0.02	0.00
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	0.03	0.00
3rd Quintile	0.04	0.00
4th Quintile	0.08	0.00
5th Quintile	0.12	0.00
Lagged self-rated health status	0.45	0.00
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	0.00	0.46
Single with No Children	-0.04	0.00
Single with Children	-0.04	0.01
Region (Ref = Central)		
Northwest	-0.02	0.00
Northeast	0.01	0.28
South	-0.02	0.00
Insular	-0.02	0.00
Year	0.01	0.00
Constant	-23.62	0.00
Number of obs	209,998	
F(20, 201782)	4569.76	
Prob > F	0.0000	
R2	0.4506	
Root MSE	0.64691	

Table A6. Process H2b: Probability of becoming long-term sick or disabled. Sample: Individuals aged 16 or older not in continuous education.

Long-term Sick or Disabled	***Coef.***	***P>z***
Self-rated Health Status	-0.44	0.00
Gender (Ref = Women)		
Men	0.16	0.00
Age	0.05	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	0.13	0.03
Low	0.25	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.45	0.00
Not Employed	0.52	0.00
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	-0.23	0.00
3rd Quintile	-0.19	0.00
4th Quintile	-0.32	0.00
5th Quintile	-0.28	0.00
Lagged self-rated health status	-0.14	0.00
Lagged long-term sick/disabled status	1.67	0.00
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	-0.14	0.00
Single with No Children	0.24	0.00
Single with Children	-0.05	0.71
Region (Ref = Central)		
Northwest	-0.07	0.04
Northeast	-0.06	0.12
South	-0.06	0.12
Insular	-0.11	0.02
Year	0.00	0.88
Constant	-0.35	0.97
Number of obs	209,998	
Wald chi2(22)	6319.70	
Prob > chi2	0.0000	
Pseudo R2	0.4423	
Log pseudolikelihood	-51058027.00	

Table A7. Process P1a: Probability of leaving the parental home. Sample: All non-student respondents living with a parent.

Leaving parental home	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.46	0.00
Age	0.07	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	-0.08	0.00
Low	-0.20	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.12	0.12
Disabled/Long-term Sick	-0.21	0.03
Not Employed	-0.05	0.05
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	-0.03	0.46
3rd Quintile	-0.05	0.16
4th Quintile	-0.08	0.02
5th Quintile	-0.14	0.00
Region (Ref = Central Italy)		
Northwest Italy	-0.06	0.03
Northeast Italy	-0.04	0.19
South Italy	-0.06	0.06
Insular Italy	-0.12	0.00
Time	-0.04	0.00
Constant	73.01	0.00
Number of obs	217,515	
Wald chi2(17)	1563.16	
Prob > chi2	0.0000	
Pseudo R2	0.0718	
Log pseudolikelihood	72109188.00	

Table A8. Process U1: Probability of entering a partnership. Sample: All single respondents aged 16 and older, not in continuous education.

Entering a Partnership	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.16	0.00
Age	-0.10	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	-0.28	0.00
Low	-0.45	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.16	0.28
Not Employed	-0.19	0.00
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	-0.41	0.00
3rd Quintile	-0.43	0.00
4th Quintile	-0.52	0.00
5th Quintile	-0.62	0.00
Lagged Number of Children in the Household	-0.57	0.00
Lagged Number of Children Aged 0-2 in the Household	-0.24	0.00
Lagged self-rated health status	0.01	0.55
Region (Ref = Central)		
Northwest	-0.03	0.49
Northeast	-0.10	0.02
South	-0.26	0.00
Insular	-0.23	0.00
Year	-0.03	0.00
Constant	65.49	0.00
Number of obs	139,537	
Wald chi2(19)	1633.07	
Prob > chi2	0.0000	
Pseudo R2	0.2220	
Log pseudolikelihood	-33993950.00	

Table A9. Process U2: Probability of partnership break-up. Sample: Female member of a couple aged 16 or older, not in continuous education.

Exiting a Partnership	***Coef.***	***P>z***
Age	-0.05	0.04
Age Squared	0.00	0.30
Educational Attainment (Ref = High)		
Medium	0.27	0.10
Low	-0.14	0.56
Lagged Personal Gross Non-benefit Income	-0.05	0.06
Lagged Personal Gross Non-benefit Income Squared	0.00	0.08
Lagged Number of Children in the Household	-0.25	0.02
Lagged Number of Children Aged 0-2 in the Household	0.22	0.10
Lagged Self-rated Health Status	0.08	0.28
Lagged Spousal Educational Attainment (Ref = High)		
Medium	-0.12	0.49
Low	0.01	0.96
Lagged Spouse Self-rated Health Status	-0.10	0.19
Lagged Own and Spouse Personal Gross Non-benefit Income Difference	0.02	0.19
Lagged Own and Spouse Age Difference	0.00	0.66
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	0.26	0.15
Single with No Children		
Single with Children		
Lagged Own and Spouse Employment Status (Ref = Both Employed)		
Employed and Spouse Not Employed	0.16	0.34
Not Employed and Spouse Employed	0.00	0.99
Both Not employed	0.23	0.23
Region (Ref = Central)		
Northwest	-0.29	0.02
Northeast	-0.34	0.00
South	-0.45	0.00
Insular	-0.55	0.01
Year	0.10	0.00
Constant	-203.51	0.00
Number of obs	38,200	
Wald chi2(20)	123.91	
Prob > chi2	0.0000	
Pseudo R2	0.1125	
Log pseudolikelihood	-1140670.20	

Table A10. Process F1: Probability of giving birth to a child. Sample: Women aged 18-44.

Having a Child	***Coef.***	***P>z***
Age	0.29	0.00
Age Squared	-0.01	0.00
Educational Attainment (Ref = High)		
Medium	-0.06	0.14
Low	-0.06	0.56
UK Fertility Rate	0.03	0.00
Lagged Employment Status (Ref = Employed)		
Student	-0.42	0.00
Not Employed	0.01	0.86
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	0.06	0.31
3rd Quintile	0.07	0.25
4th Quintile	0.16	0.01
5th Quintile	0.33	0.00
Lagged Number of Children in the Household	-0.19	0.00
Lagged Number of Children Aged 0-2 in the Household	0.24	0.00
Lagged Self-rated Health Status	0.01	0.68
Lagged Partnership Status (Ref = Married)		
Single	-0.95	0.00
Previously Partnered	-0.45	0.00
Region (Ref = Central)		
Northwest	-0.10	0.05
Northeast	-0.01	0.90
South	-0.15	0.00
Insular	-0.04	0.58
Constant	-6.63	0.00
Number of obs	42,141	
Wald chi2(20)	977.41	
Prob > chi2	0.0000	
Pseudo R2	0.1632	
Log pseudolikelihood	-29199165.00	

Table A11. Process I1a: Probability of receiving capital income. Sample: All respondents aged 16-29 in continuous education.

Receiving capital income	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	-0.08	0.16
Age	0.27	0.03
Age Squared	-0.01	0.08
Lagged self-rated health status	0.02	0.72
Lagged Employment Income	-0.01	0.42
Lagged capital income	0.51	0.00
Region (Ref = Central)		
Northwest	0.24	0.00
Northeast	0.29	0.00
South	-0.25	0.00
Insular	-0.19	0.07
Year	0.06	0.00
Constant	-118.01	0.00
Number of obs	12,276	
Wald chi2(11)	1859.55	
Prob > chi2	0.0000	
Pseudo R2	0.276	

Table A12. Process I1b. Probability of receiving capital income. Sample: Respondents 16 or older not in continuous education.

Receiving capital income	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.00	0.91
Age	0.00	0.55
Age Squared	0.00	0.04
Educational Attainment (Ref = High)		
Medium	-0.31	0.00
Low	-0.49	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.01	0.90
Not Employed	-0.12	0.00
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	-0.11	0.00
Single with No Children	-0.26	0.00
Single with Children	-0.47	0.00
Lagged self-rated health status	0.09	0.00
Lagged Employment Income	0.01	0.00
Lagged capital income	0.46	0.00
2 period lagged Employment Income	0.00	0.46
2 period lagged capital income	0.20	0.00
Region (Ref = Central)		
Northwest	0.09	0.00
Northeast	0.21	0.00
South	-0.29	0.00
Insular	-0.48	0.00
Year	0.13	0.00
Constant	-254.23	0.00
Number of obs	102,120	
Wald chi2(20)	16113.66	
Prob > chi2	0.0000	
Pseudo R2	0.2936	

Table A13. Process I2a. Amount of capital income. Sample: All respondents aged 16-29 in continuous education and receiving positive capital income.

Amount of capital income	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.02	0.64
Age	0.25	0.02
Age Squared	-0.01	0.03
Lagged self-rated health status	0.03	0.53
Lagged Employment Income	0.00	0.92
Lagged capital income	0.44	0.00
Region (Ref = Central)		
Northwest	0.11	0.11
Northeast	-0.10	0.15
South	0.09	0.21
Insular	-0.10	0.34
Year	-0.07	0.00
Constant	150.57	0.00
Number of obs	7,197	
F(11, 4738)	180.27	
Prob > F	0.0000	
R-squared	0.3777	
RMSE	1.5127	

Table A14. Process I2b: Amount of capital income. Sample: Respondents 16 or older not in continuous education and receiving positive capital income.

Amount of capital income	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.01	0.35
Age	0.00	0.89
Age Squared	0.00	0.15
Educational Attainment (Ref = High)		
Medium	-0.20	0.00
Low	-0.38	0.00
Lagged Employment Status (Ref = Employed)		
Student	0.37	0.00
Not Employed	0.10	0.00
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	-0.09	0.00
Single with No Children	0.26	0.00
Single with Children	0.09	0.26
Lagged self-rated health status	0.04	0.00
Lagged Employment Income	0.01	0.00
Lagged capital income	0.37	0.00
2 period lagged Employment Income	-0.01	0.02
2 period lagged capital income	0.18	0.00
Region (Ref = Central)		
Northwest	0.06	0.00
Northeast	-0.08	0.00
South	0.11	0.00
Insular	0.04	0.23
Year	-0.06	0.00
Constant	129.00	0.00
Number of obs	58,923	
F(20, 38498)	1385.25	
Prob > F	0.0000	
R-squared	0.4495	
RMSE	1.4134	

Table A15. Process R1a: Probability of retiring, singles. Sample: Non-partnered individuals aged 50+ who are not yet retired.

Probability of retiring	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.27	0.00
Age	0.40	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	0.07	0.25
Low	-0.03	0.72
Pension Age (Ref=Not Reached Pension Age)		
Pension Age	0.28	0.00
Lagged Employment Status (Ref = Employed)		
Not Employed	0.51	0.00
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	0.23	0.00
3rd Quintile	0.24	0.00
4th Quintile	0.37	0.00
5th Quintile	0.32	0.00
Lagged Disability Status (Ref = No)		
Disabled	0.18	0.00
Region (Ref = Central Italy)		
Northwest Italy	0.12	0.01
Northeast Italy	0.08	0.05
South Italy	-0.17	0.00
Insular Italy	-0.24	0.00
Time	0.02	0.00
Constant	-51.17	0.00
Number of obs	67,035	
Wald chi2(17)	1127.24	
Prob > chi2	0.0000	
Pseudo R2	0.4001	
Log pseudolikelihood	-40423877.00	

Table A16. Process R1b: Probability of retiring, partnered. Sample: Partnered individuals aged 50+ who are not yet retired.

Probability of retiring	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.57	0.00
Age	0.53	0.00
Age Squared	0.00	0.00
Educational Attainment (Ref = High)		
Medium	0.09	0.05
Low	0.07	0.15
Pension Age (Ref=Not Reached Pension Age)		
Pension Age	0.29	0.00
Lagged Employment Status (Ref = Employed)		
Not Employed	0.61	0.00
Lagged Household Income Quintile (Ref = 1st Quintile)		
2nd Quintile	0.00	0.97
3rd Quintile	0.07	0.14
4th Quintile	0.20	0.00
5th Quintile	0.29	0.00
Lagged Disability Status (Ref = No)		
Disabled	0.54	0.00
Spouse Pension Age (Ref=Not Reached Pension Age)		
Pension Age	0.10	0.01
Lagged Spouse Employment Status (Ref = Employed)		
Student		
Not Employed	0.10	0.00
Lagged Spouse Disability Status (Ref = No)		
Disabled	0.25	0.01
Region (Ref = Central Italy)		
Northwest Italy	0.15	0.00
Northeast Italy	0.13	0.00
South Italy	-0.10	0.00
Insular Italy	-0.19	0.00
Pension Age and Lagged Employment Status Interaction (Ref= Not Reached Pension Age and Employed)		
Pesion Age and Not Employed	-0.46	0.00
Time	-0.02	0.00
Constant	19.80	0.02
Number of obs	45,161	
Wald chi2(21)	3273.92	
Prob > chi2	0.0000	
Pseudo R2	0.2102	
Log pseudolikelihood	-63252546.00	

Table A17. Process I3: Amount of pension income. Sample: Individuals aged 50 or older who are retired.

Amount of pension income	***Coef.***	***P>z***
Gender (Ref = Women)		
Men	0.15	0.00
Age	0.12	0.01
Age Squared	0.00	0.01
Educational Attainment (Ref = High)		
Medium	-0.14	0.00
Low	-0.25	0.00
Lagged Employment Status (Ref = Employed)		
Student	-0.29	0.03
Not Employed	-0.28	0.00
Lagged Household composition (Ref = Couples with No Children)		
Couples with Children	0.00	0.98
Single with No Children	-0.09	0.00
Single with Children	-0.14	0.82
Lagged self-rated health status	0.05	0.00
Lagged Employment Income	0.04	0.00
Lagged capital income	0.57	0.00
2 period lagged Employment Income	0.02	0.00
2 period lagged capital income	0.06	0.00
Region (Ref = Central)		
Northwest	0.02	0.41
Northeast	0.01	0.68
South	0.01	0.81
Insular	-0.01	0.77
Year	-0.05	0.00
Constant	91.85	0.00
Number of obs	27,966	
F(20, 17381)	335.77	
Prob > F	0.0000	
R-squared	0.5917	
RMSE	1.1671	

The estimates for the labour supply model are separately described in Richiardi and He (2021).

Appendix 2. Uncertainty assessment

Uncertainty analysis

Uncertainty regarding a model's projections can arise from a variety of reasons (Bilcke et al., 2011; Creedy et al., 2007). In particular, sources of uncertainty are generally distinguished in (i) input data, for instance due to sampling errors in the initial population, (ii) model structure, that is the validity of the general modelling approach used (also called "methodological uncertainty"), (iii) model specification, which concerns the choice of the covariates and the functional forms used, and in particular the crucial assumption that any regularity observed in the data will not break up in the future, (iv) model parameters, pointing to the imprecision of the estimates and/or externally provided parameters, and finally (v) Montecarlo variation of the model output, which originates from the fact that the simulated aggregate quantities are also imprecise estimates of the theoretical aggregate quantities that the model implicitly defines. None of the above sources of uncertainty is generally considered in microsimulation studies, although this is recognised and criticised (see for instance Goedemé et al, 2013). However, "the calculation of confidence intervals around model results that account for all sources of error remains a major challenge" (Mitton et al., 2000).

Generally speaking, source (i) should be limited, due to the use of appropriate input data and sampling weights. Sources (ii)-(iii) are often left unexplored, by making the common assumption that the model is well specified (measures of fit should be reported for each estimated equation to corroborate this hypothesis). Montecarlo variation of the model outcome (source v) can be brought down to negligible by appropriately scaling up simulated population size. The remaining source of uncertainty that needs to be addressed is therefore parameters uncertainty, stemming from sampling errors in estimation (source iv).

There are two approaches that can be used to deal with this uncertainty (Creedy et al., 2007). The first is what we might label "brute force", and prescribes to bootstrap the coefficients of the estimated equations from their estimated distribution (e.g. multivariate normal in case of multinomial probit regressions) with mean equal to the point estimate, and variance-covariance matrix equal to the estimated variance-covariance. Bootstrapping needs to be performed only once, at the beginning of the simulation: the entire simulation is then performed with the bootstrapped values of the coefficients. The second approach provides an approximation by assuming from the onset a normal distribution for the resulting confidence intervals, requiring many fewer draws from the parameter distribution.

LABSim implements the "brute-force" approach, by exploiting the bootstrapping feature of the JAS-mine Regression library within a multi-run implementation: the simulation is run many times (e.g. 1,000 times), each using a different set of coefficients. The result is a distribution of model outcomes, around the central projections obtained with the estimated coefficients.

References

Bilcke J, Beutels P, Brisson M, Jit M (2011). Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide. Medical Decision Making 31(4): 675-692.

Creedy J, Kalb G, Kew H (2007). Confidence intervals for policy reforms in behavioural tax microsimulation modelling. Bulletin of Economic Research 59(1): 37-65.

Goedemé T, Van den Bosch K, Salanauskaite L, Verbist G (2013) Testing the Statistical Significance of Microsimulation Results: A Plea. International Journal of Microsimulation 6(3): 50-77.

Ministry of Treasury, Italian Government and Fondazione Giacomo Brodolini (2011). T-DYMM. Innovative datasets and models for improving welfare policies. Final report. Available at http://www.dt.mef.gov.it/modules/documenti_it/analisi_progammazione/finanziam_comunitari_ricerca/Final-Report.pdf.

Mitton L, Sutherland H, Weeks M (2000). Microsimulation Modelling for Policy Analysis. Challenges and Innovations. Cambridge University Press, Cambridge, UK.

Richiardi M, He Z (2021). No one left behind: The labour supply behaviour of the entire Italian population. Centre for Microsimulation and Policy Analysis, mimeo.

Richiardi M, Richardson RE (2017). JAS-mine: A New Platform for Microsimulation and Agent-Based Modelling. International Journal of Microsimulation, 10(1): 106-134.

van de Ven J (2018). The National Institute model for lifetime income distributional analysis, LINDA. National Institute of Economics and Social Research, mimeo.