Ministry of Economy and Finance
Department of the Treasury

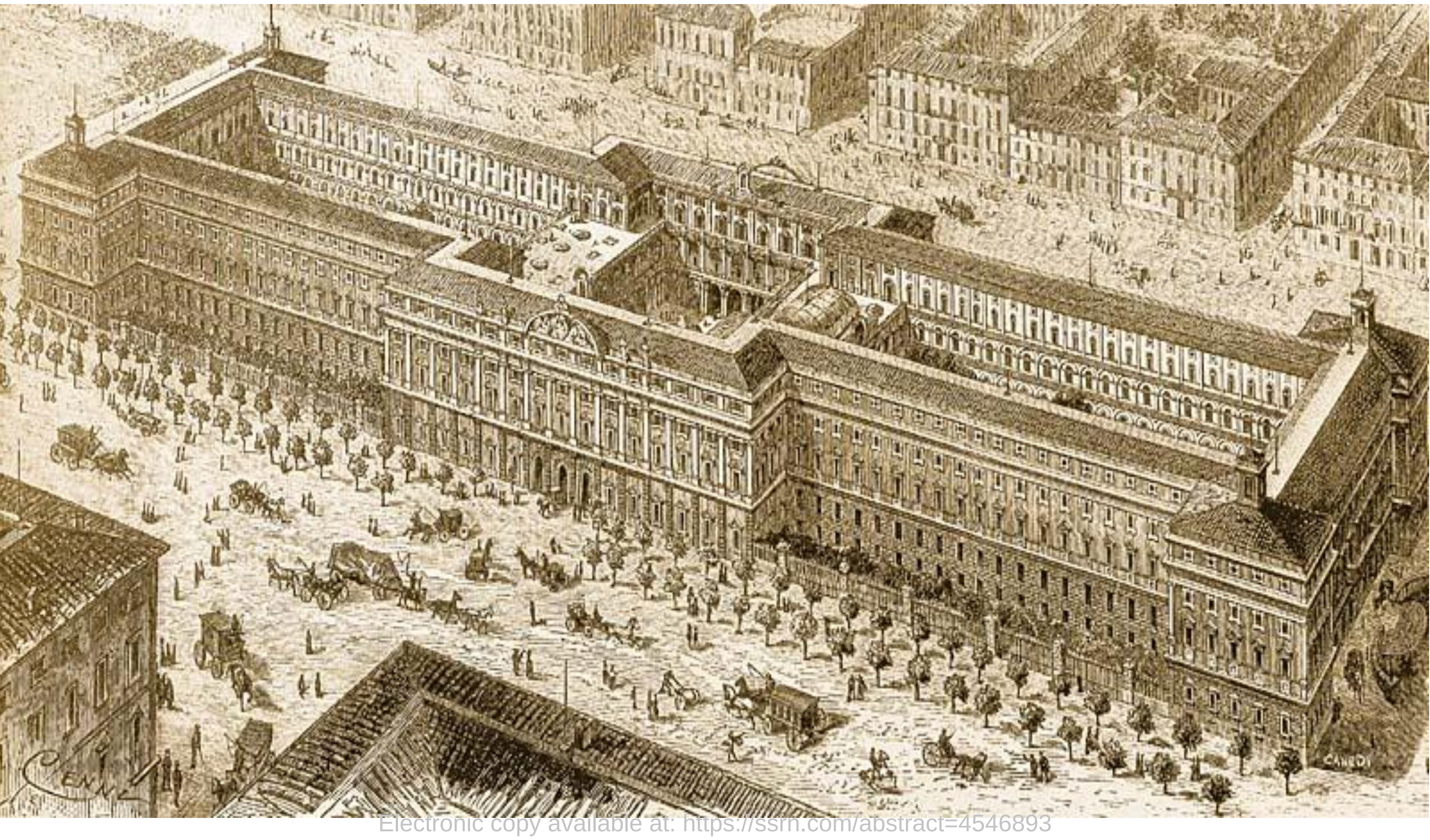
Analisi e Ricerca Economico Finanziaria

Working Papers
N°1 - March 2023

ISSN 1972-411X

The Italian Treasury Dynamic Microsimulation Model (T-DYMM): data, structure and baseline results

Riccardo Conti, Michele Bavaro, Stefano Boscolo, Elena Fabrizi, Chiara Puccioni, Ottavio Ricchi, Simone Tedeschi



Working Papers

CONTENTS

The Italian Treasury Dynamic Microsimulation Model (T-DYMM): data, structure and baseline results

Riccardo Conti[a], *Michele Bavaro*[b], *Stefano Boscolo*[c], *Elena Fabrizi*[d], *Chiara Puccioni*[e], *Ottavio Ricchi*[f], *Simone Tedeschi*[g]

Abstract

The Treasury Dynamic Microsimulation Model (T-DYMM) is a dynamic microsimulation model (DMM) owned by the Department of the Treasury of the Italian Ministry of Economy and Finance. One of the very few DMMs presently operating in Italy, it has been developed through three research projects, spanning from 2009 to 2021. The present article is intended to provide a general and comprehensive description of T-DYMM. The model, based on the AD-SILC dataset, which matches administrative and survey data, delivers long-term projections and is organised in five modules: demographic, labour market, pension, wealth and tax-benefit (the last two contain the most relevant novelties of the present version of T-DYMM). The broad aim of the model is to provide long-term analyses of the Italian social security system, with a focus on pension and social protection adequacy and their distributional characterization.

JEL Classification: C1, C5, C6.

Keywords: dynamic microsimulation, labour market, pensions, wealth, inequality, long-term projections.

Acknowledgements: The research team at the Ministry of Economy and Finance is greatly indebted to our institutional partners: the National Institute for Public Policy Analysis (INAPP) and the Fondazione Giacomo Brodolini (FGB) for research support; the National Social Security Institute (INPS), the Italian National Institute of Statistics (ISTAT), the Department of Finance at the Italian Ministry of Economy and Finance and the Italian Supervisory Authority on Pension Funds (COVIP) for providing essential data, as well as for the support in discerning them; the Directorate-General for Employment, Social Affairs and Inclusion of the European Commission for its continued financial support to the research projects for the development of T-DYMM throughout the years.

---

[a] Sogei S.p.A. Email: rconti@sogei.it

[b] Department of Economics, University of Roma Tre. Email: michele.bavaro@uniroma3.it

[c] Department of Statistical Sciences Paolo Fortunati, University of Bologna; Centre for the Analysis of Public Policies (CAPP), University of Modena and Reggio Emilia. Email: stefano.boscolo5@unibo.it

[d] Department of Political Science, University of Teramo. Email: efabrizi@unite.it

[e] Economic Research Department, Confindustria; Department of Economics and Finance, University of Tor Vergata. Email: c.puccioni@confindustria.it

[f] Italian Ministry of Economy and Finance. Email: ottavio.ricchi@mef.gov.it

[g] Department of Economics and Law, University of Cassino and Southern Lazio. Email: simone.tedeschi@unicas.it

1 Introduction

The Treasury Dynamic Microsimulation Model (T-DYMM) is a dynamic microsimulation model (DMM) owned by the Department of the Treasury of the Italian Ministry of Economy and Finance. One of the very few DMMs presently operating in Italy, it has been developed through a series of three research projects sponsored by the European Commission, spanning from 2009 to 2021. The first release of T-DYMM benefited largely from the experience of MIDAS-IT, a DMM developed by ISAE (*Istituto di studi e analisi economica*), which in turn drew its main features from MIDAS Belgium and the work of Gijs Dekkers, to which T-DYMM is still strongly indebted and connected. The second and third version of T-DYMM have vastly expanded both the model structure and the dataset, as well as moved the model to its current platform, Liam 2.0.[1]

T-DYMM aims at providing long-term analyses of the Italian social security system, with a focus on pension and social protection adequacy and their related distributional effects. Amongst the latest additions to T-DYMM is the expansion of the Tax-Benefit module, which now encompasses the vast majority of the Italian fiscal system, as well as the inclusion of an international Migration submodule and of a Wealth module.[2]

The present paper is intended to provide a general and comprehensive description of the present version of T-DYMM, which we will refer to as T-DYMM 3.0 (third release). In the first Section, we describe the data employed, with appendixes focusing on the steps undertaken to build our dataset, named 'AD-SILC', and to calibrate the starting sample for the simulations. In the following Section, we explore the model structure and the main processes. Finally, we illustrate some key findings of our baseline scenario, the evolution of the sample in its demographic and labour market characteristics, wealth components and estimate poverty and inequality indicators, with a focus on the position of retirees and elderly people. Conducting an in-depth external validation of the model in its baseline version, as we do in a dedicated appendix, is the cornerstone for the sensitivity analyses and policy evaluations that may follow the present introductory paper.

2 Microsimulation models and T-DYMM

In the broad field of economics, microsimulation models are simulation tools based on micro units of data such as individuals, households or firms. The modelling activity consists on the use of computational software in order to capture the impact of changes in (e.g. tax-benefit) policy parameters and/or changes in the behaviour of economic agents (Dekkers, 2014). Unlike macroeconomic models, individual heterogeneity is entirely preserved and modelled.

In the field of microsimulation, dynamic microsimulation models, different from static models, simulate agents' behaviours over time, updating the sample over each period. They can and have been used for a number of different purposes such as: projections of socio-demographic dimensions under current or alternative policies; assessment of sustainability and adequacy of welfare systems; to combine microsimulation with macroeconomic analysis, where dynamic models serve as inputs to CGE models (Colombo, 2009; Buddelmeyer et al., 2011; Peichl, 2015).

The starting point is a representative sample of the population at a given time t, often a mixed dataset combining survey data with administrative and/or census data at the micro level (the case where only hypothetical data are used is more rare). Behaviours are modelled and simulated using regression-based analyses combined with Monte Carlo simulation techniques.[3] Unlike static models, sample evolution takes place by simulating future individual and household characteristics, in a procedure referred to as 'dynamic ageing'.

Using the classifications put forward in O'Donoghue (2001), DMM models can simulate behaviours for the entire population (*population models*) or for a population subgroup (*cohort models*). The difference lies in the fact that cohort models age a specific cohort and follow each individual until death occurs, while population models project all individuals to a given time t in the future.

Dynamic models can also be classified by the way time is treated, either as a discrete or continuous variable. When time is discrete, behaviours are simulated at each period t (usually a year) based on a defined sequential order, and status transitions take place at the start of each new period $t+1$ (e.g. from employed at time t to retired at time $t+1$). When time is continuous, instead, survival functions are used to simulate the timing of events, and status transitions can occur at any time, allowing for a more precise computation of duration. Though more theoretically appealing, the treatment of time as a continuous variable poses a series of challenges, starting with data availability and estimation problems (Li and O'Donoghue, 2014).

Finally, the use of alignment procedures in dynamic microsimulation is worthy of mention. Alignments consist of calibration techniques used to constrain the output of the model to externally derived aggregates. It is a way to incorporate additional information that is not available in the estimation data: the underlying assumption is that the microsimulation

[1] LIAM2 is a tool for the design of microsimulation models developed by the Federal Planning Bureau (Belgium), with testing and funding by CEPS/INSTEAD (Luxembourg) and IGSS (Luxembourg).

[2] A correct simulation of the mechanisms of transmission and accumulation of wealth is crucial in inequality and poverty analyses, given the recent evidence on the Italian context (Acciari et al., 2020).

[3] The estimated probability for event x to occur is compared with a random number extracted from a uniform distribution. If the former is higher than the latter, x takes place.

model is a poor(er) model of the aggregate, but a good model of individual heterogeneity. The number of individuals who experience a specific event x (e.g. giving birth, dying, being employed, etc.) can be aligned to external projections (e.g. fertility rates, mortality rates, employment rates), often specified by age and gender. The selection of individuals that are to experience a given event may be carried out randomly or by ranking them according to a score – what is defined in the literature as sorting-based alignment method (Li and O'Donoghue, 2014). The importance of alignments lies in the limited predictive capability of dynamic models. The essential aim of such models is not to provide forecasts of demographic or macroeconomic variables; rather, these models incorporate demographic and macroeconomic projections and provide insights on aggregate and distributive traits of the simulation sample concerning the labour market, the tax-benefit system, the pension system, etc. Projecting the future characteristics of the simulation sample on the sole basis of the model's equations may lead to totals that do not match with official projections or expected patterns. For these reasons, the use of alignment procedures is generally regarded as a fundamental step towards the construction of a coherent projection of the model's baseline scenario, although the debate on what variables should be aligned and how extensive alignments should be is far from settled (Baekgaard, 2002; Li and O'Donoghue, 2014).

Within this framework of literature, T-DYMM is a DMM that simulates behaviours for the entire population (*population model*), treats time as a discrete variable, employs alignment procedures for a number of dimensions and often uses sorting techniques based on regressions estimated externally.

3 Data

3.1 Microeconomic data

The core of T-DYMM's dataset is composed by linking data from the Italian component of the European Union Statistics on Income and Living Conditions (IT-SILC) survey, delivered for Italy by the Italian National Institute of Statistics (ISTAT), with administrative data from the Italian National Institute of Social Security (INPS). The merging procedure is conducted through individual tax codes (*codici fiscali*) that are subsequently anonymised. We call the merged dataset 'AD-SILC'. AD-SILC is an unbalanced panel dataset that in its current version comprises the information contained in all SILC waves from 2004 to 2017 and in the INPS archives (for the linked individuals). From IT-SILC, we derive longitudinal data (respondents are followed up to 4 years) on socio-economic characteristics for a total of 254,212 individuals[4]; from INPS, we derive longitudinal data on pensions (disability, old-age, survivor, etc.) and working history (occupational status, income evolution, contribution accrual, etc.), for a total of 6,182,926 observations over the 1922-2018 period.[5] In addition to the IT-SILC and INPS data, the latest version of AD-SILC includes information from tax returns and the Cadastre (collected by the Department of Finance of the Italian Ministry of Economy and Finance) for the 2010, 2012, 2014 and 2016 corresponding IT-SILC waves. Furthermore, after applying a specific correction for data on financial wealth to take into account a well-documented under-reporting issue (see Appendix A for a complete explanation), we included, through a statistical matching procedure (described in detail in Appendix B) information from the Survey on Household Income and Wealth (SHIW) conducted by the Bank of Italy. Housing wealth is constructed based on the administrative data from the Cadastre and from tax returns, whereas information on financial wealth and liabilities is retrieved from SHIW. These last two innovations in the data sources allow building a rich dataset on household wealth, setting the base for the Wealth module, that constitutes one of the crucial additions to the latest release of T-DYMM. Tax returns data are also used to correct the information contained in INPS data concerning labour income from self-employment activities (see Appendix C).

AD-SILC can be employed for a number of uses: i) to analyse historical dynamics (e.g., within the labour market); ii) to estimate transition probabilities and determinants of labour income to be included in T-DYMM (these estimates are carried out over a panel version of AD-SILC); iii) to derive the starting sample for the simulations (which run on a single extract of AD-SILC, relative to the 2016 EU-SILC wave linked with all the aforementioned data). Before running the simulations, the starting sample needs to be properly calibrated in order to improve the overall representativeness of a series of dimensions we are interested in. Subsequent to the weight calibration, we expand the starting sample by multiplying individuals by calibrated weights, which is necessary to overtake representativeness issues that emerge when making use of alignment methods in dynamic microsimulation (Dekkers and Cumpston, 2012). We then draw with replacement 100 samples of 100,000 households and select the best-fitting sample with respect to administrative data. As a result, the starting sample contains 238,431 individuals. More information on weight calibration and sample extraction is provided in Appendix D.

3.2 Exogenous data and alignments

Exogenous data are used to align a number of patterns within the simulations. The use of alignment procedures is generally regarded as a fundamental step towards the delivery of a coherent projection, although the debate on what variables

[4]For a detailed illustration of the sample design see: https://www.istat.it/en/archivio/212387

[5]Many individuals in INPS have multiple records per year. In Inps archives we observe an average career of 19.72 years (SD: 13.59)

should be aligned and how extensive alignments should be is far from settled (Baekgaard, 2002; Li and O'Donoghue, 2014). Institutional models such as T-DYMM may wish to ensure that certain demographic or macroeconomic dynamics stay in line with third-party projections and focus (for instance) on distributional concerns. Alignments may be easily modified to simulate sensitivity scenarios.

The main sources for alignments and exogenous variables in T-DYMM are: i) Eurostat demographic projections[6], used to align mortality rate, fertility rate, immigration and emigration by age and gender; ii) European Commission Forecasts and Ageing Report assumptions[7], used to align employment rate, inflation growth, GDP growth, productivity growth, disability rate by age and gender, returns on risk-free assets; iii) Population-level data by the Italian Department of Finance, used to align the number of households paying rent and the total beneficiaries of specific tax expenditures and substitute tax regimes; iv) Population-level data by ISTAT, used to align the probability of leaving the household of origin, the age and country of birth of migrants, educational levels, acquisitions of houses, the propensity to consume, the propensity to marry and to divorce; v) Population-level data by INPS, used to align the occurrence of disability allowances and inability pensions; vi) Population-level data by the Italian Supervisory Authority on Pension funds (COVIP), used to align enrolment rates to private pension plans.

Table 1 lists the main processes in T-DYMM by use of alignment procedures, associating the corresponding module and source. The majority of processes in the Demographic module are aligned, while the opposite is true for all other modules.

Table 1: Aligned processes in T-DYMM

Module	**Process**	**Source**
Demographic module	Fertility, mortality, immigration/emigration flows, education, age of exit from original household, informal/formal marriages, divorces	Eurostat, ISTAT
Labour Market module	Employment rates, quota of permanent public employees, average labour income growth, take-up rate of unemployment benefits	European Commission Forecasts and AWG assumptions, INPS, ISTAT
Pension module	Number of disability pensions, enrolment in private pension plans, retirement through the 'Quota 100' criterion	COVIP, INPS
Wealth module	Houses sold/bought, families that pay rent, average propensity to consume	Bank of Italy, ISTAT, Department of Finance, European Commission Forecasts and AWG assumptions, S&P500, Housing Observatory (OMI)
Tax-Benefit module	Take-up rates of specific social assistance measures, beneficiaries of specific tax expenditures and substitute tax regimes	Department of Finance, INPS

4 Model structure

The present Section illustrates the current structure of the baseline version of T-DYMM. Appendixes E and F illustrate regression results for the most relevant estimations conducted on AD-SILC, which are at the foundation of the behaviours simulated in the model. The modular structure and the set of estimates are what underlies the results presented in Section 5. The units of analysis in the model are individuals and households. Regions and municipalities are not modeled in T-DYMM, hence each individual and household is simply 'Italian'. All legislation simulated in the model is updated to 2021.

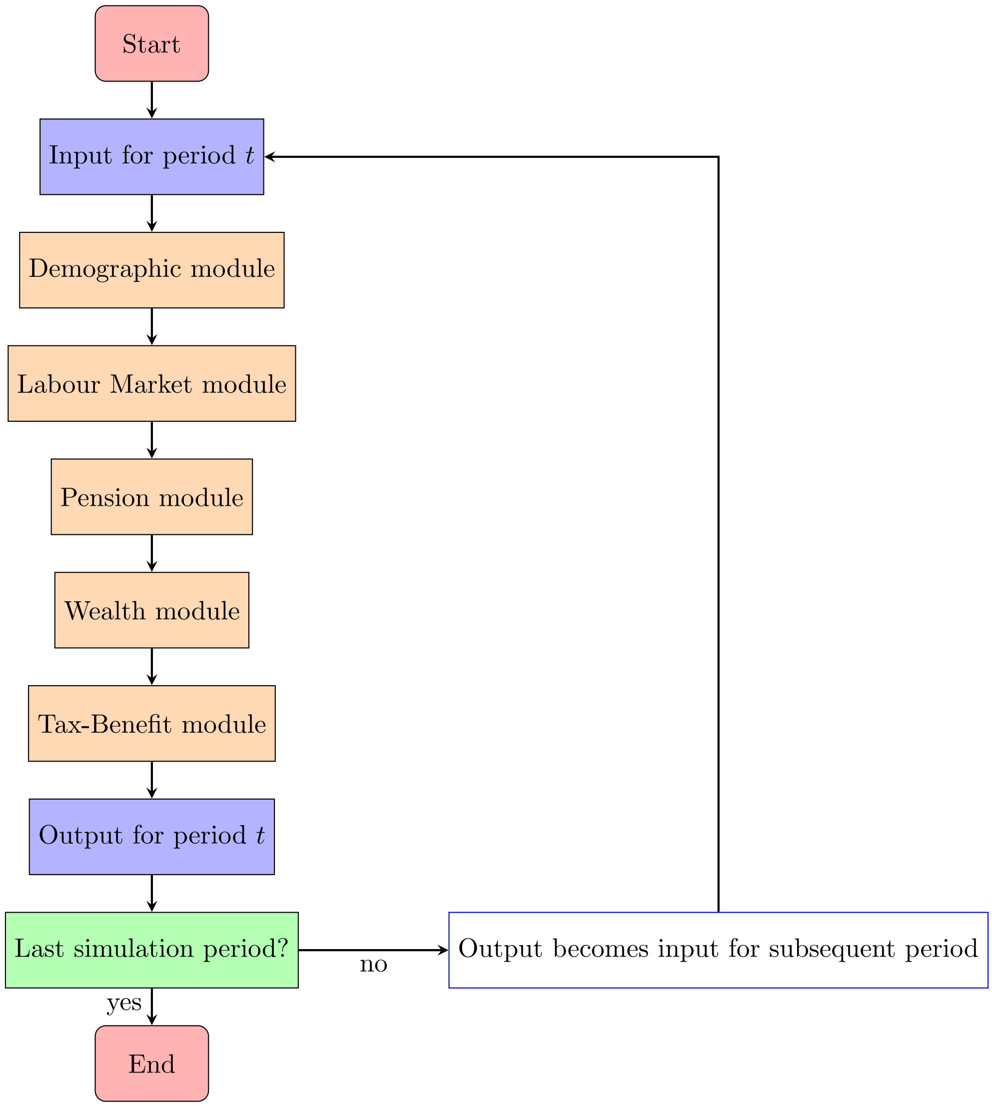
The model is organized in five modules, as shown in Figure 1, and operates sequentially. In the present version of the model, the starting sample is set in 2015 and simulations run on an annual basis from 2016 until 2070 (the projection horizon of the 2021 Ageing Report by the European Commission). The organization in modules is logical and does not

[6]For the current version of T-DYMM, Europop 2019 projections are employed.

[7]For the current version of T-DYMM, the 2022 Spring Forecasts and the assumptions underlying the 2021 Ageing Report from the Ageing Working Group of the Economic Policy Committee are employed.

strictly represent the sequence of processes that the model solves. For instance, the ‘consumption/saving’ process, which by logic belongs to the Wealth module, is solved last, as one can only compute consumption on the basis of a final definition of income (comprehensive of benefits and net of taxes).

Figure 1: Modular structure of T-DYMM



4.1 Demographic module

In the Demographic module, the sample evolves in its components inherent to demographic aspects. Individuals in the sample are born, age, die, migrate, get educated, leave their household of origin, form couples and separate and become disabled. In the following, we shall briefly describe these processes.

Ageing, mortality and fertility
T-DYMM is an annual model. All states are updated annually, starting from ageing. At present, death is assigned randomly according with aligned mortality rates by age and gender[8], following the latest Europop projections.

Fertility rates (for women between 14 and 50, by age) are aligned to the latest Europop projections. Parameters estimated via logit regressions in the AD-SILC dataset distribute the probability to have children across women by civil status, duration of marriage/cohabitation, presence of other children and employment status.

International migration

[8]The team is currently testing the inclusion of heterogeneity in mortality, i.e. distributing average mortality rates (by age and gender) across socio-economic conditions and labour market characteristics.

Considering the well-known scarcity of quality data on the international migration phenomenon[9], we have opted for a rather simplified modelisation of the Migration submodule in T-DYMM, which could serve as a basis for future expansions when further micro data become available. We focus on three essential dimensions to define migrants: age, gender and area of birth (Italy, EU and non-EU). We simulate immigration and emigration separately and follow Chénard (2000) and Dekkers (2015) in implementing a 'cloning procedure' for households using Chénard's Pageant algorithm, which allows selecting households in the model (to either immigrate or emigrate) while ensuring that certain individual characteristics (in our case, age, gender and area of birth) are matched. Inflows and outflows of migrants are aligned to Europop projections; education (for immigrants) and area of birth (for both immigrants and emigrants) are assumed constant (by gender and age group) according to respectively OECD[10] and ISTAT data. Immigrants are included as clones, but lose all the original characteristics of the individuals they are cloned from other than age, gender (which are aligned) and household composition; the implicit assumption is that all immigrants 'start fresh' when they arrive in Italy, carrying no relevant work experience with them and no pension rights. These simplifying assumptions are due to lack of data, though they should not necessarily be looked at as excessively stringent, given the age structure of the immigrant population and the segregational features of the Italian labour market Strozza and De Santis (2017). As no data is available at this point to model emigrants' behaviour once they leave Italy, they are simply deleted from the simulation, i.e. households (and individuals) are not followed in their (possibly) multiple entries/exits. Therefore, while we align the overall flows to Europop projections, we may be overestimating the incidence of the migration phenomenon on a micro basis.

Disability
Each year T-DYMM assigns an individual probability to become disabled ('strongly-limited in daily activities on a long-term basis', according to the EU-SILC definition), based on regression parameters estimated on AD-SILC that highlight the role of education, income and the disability state at time t-1 (disability is highly persistent). Probabilities by gender and age class (nineteen age classes) are aligned to the assumptions underlying the 2021 Ageing Report.

Education
Individuals in the model may hold elementary, lower-secondary, upper-secondary or tertiary education. Following the legislation on compulsory education in Italy, in simulation years T-DYMM assigns lower-secondary as the lowest possible education achievement for individuals who receive their education in Italy.[11] Probabilities to get tertiary education are assigned individually based on estimations run on the AD-SILC dataset. Due to the difficulty of attributing to individuals time-variant characteristics relative to the moment in time when their latest educational level was achieved, the only explanatory variables employed concern parental education, assumed to be time-invariant. Levels of education lower than tertiary are assigned randomly, while all probabilities by gender are aligned to ISTAT data. Individuals receive lower-secondary education at 16, upper-secondary (if entitled) at 19 and tertiary (if entitled) between 21 and 29 years of age according to probabilities derived from *ad hoc* survey data.[12]

Leaving household of origin
Because T-DYMM has the ambition to estimate poverty and distributive dynamics, income variables will have to be estimated at the household level. Therefore, it is crucial that households are properly designed. Each year, young individuals still living with their parents are assigned a random probability to leave and form a new household; the latest ISTAT data on the quota of young people living with their parents are used to align future exit flows. Young people below a given income threshold or suffering from severe disabilities are not allowed to leave the family of origin and live as single-member households.

Cohabitation or marriage and divorce
Each year, single individuals are assigned a probability to form a couple according to estimates on AD-SILC and alignments on ISTAT data. Since the past few years have seen a visible and somewhat uncharacteristic decrease in the propensity to marry, we assume that the number of marriages every 1,000 individuals will start rising again and recover, in 2029, its 2008 value. Following the progression observed in census data, we assume that every four marriages a new informal cohabitation is established. Once individuals to be coupled are selected, they are matched according to a score that takes into account age, education differentials and a dummy returning 1 if both potential partners are employed (we are attempting to reproduce the positive 'assortative mating' behaviours that we observe in the AD-SILC sample). Each year, couples are assigned a probability to divorce or separate according to probabilities estimated on AD-SILC. The overall propensity to divorce/separate is aligned to ISTAT data. The passing of the legislation on the so-called 'fast divorce' (*divorzio breve*, which has sped up divorce procedures) has produced a break in the series in 2015-2016, when the number of yearly divorces doubled compared to previous years. In order to account for that and for the following

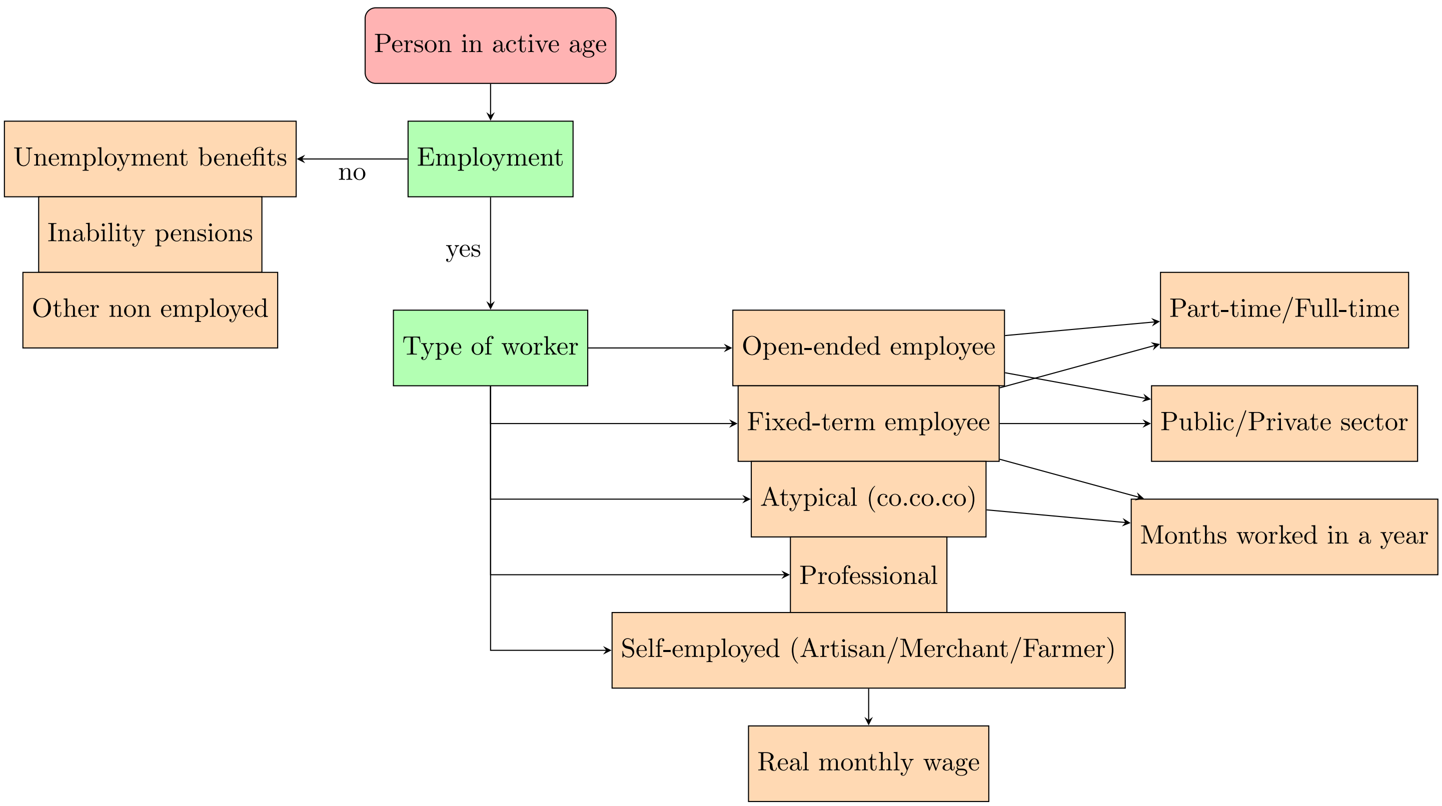
[9] As regions and municipalities are not simulated, the present version of T-DYMM does not include a sub-module on internal migration, only international migration is addressed.
[10] See the Database on Immigrants in OECD Countries.
[11] For immigrants, education achievements are attributed according with OECD data by gender and area of origin.
[12] See Almalaurea, Profilo dei Laureati - Rapporto 2020.

gradual reduction of yearly occurrences in the 2017-2019 period, we assume that the propensity to divorce/separate will keep reducing linearly and, 10 years after the approval of the aforementioned legislation, stabilise at a rate equal to the average of the pre and post-reform values.

4.2 Labour Market module

The Labour Market module assigns individuals to employment, simulates transitions between different employment statuses, imputes the number of months in work within the year and the corresponding level of labour income. The module is based on a sequence of nested choices, as shown in Figure 2, following probabilities estimated through a series of logistic and multinomial logistic equations. Estimates have, in general, been run on individuals recorded in AD-SILC between 15 and 80 years of age, including students and retirees. However, due to the specificity of working pensioners, in some cases, non-retired labour force and pensioners follow separate processes. indeed, to model working pensioners' type of works (10,000 individuals, defined regardless their age), the labour market module uses simplifies regressions to handle the smaller number of observations, in comparison with the total sample AD-SILC.

Figure 2: Structure of the Labour Market module



The Labor Market module is an essential pillar of the microsimulation model because of the implications it has on all the other modules. For example, intermittent careers or low salaries will have an impact not only on the current economic status and on access to unemployment benefits but also on retirement prospects. Different socio-demographic and economic determinants may influence individual careers. Below, we shall illustrate the processes included in the model, while the main regressions estimated using AD-SILC are reported in Appendix E. The first process is the one determining, through a logit regression, whether individuals are employed. As already mentioned, employment rates by gender and age are aligned following the macroeconomic assumptions underlying the 2021 Ageing Report. People who are not employed are instead assigned to 'out-of-work' statuses, which includes individuals receiving inability pensions, unemployment benefits, or those in other forms of unemployment or inactivity.

For those working, a multinomial logit assigns the probabilities of different contractual arrangements. We have grouped individuals in seven working categories using INPS data: open-ended private and public employees, fixed-term private and public employees, professionals, self-employed and atypical workers (i.e. co.co.co.)[13]. In the model, workers are allowed to have one job per year, albeit roughly 10% of the observations in AD-SILC hold multiple job spells. For those individuals, we must identify an annual main employment category. To define the most representative work for a given period, we

[13]See MEF et al. (2020) - chapter 2 for a detailed description of the Italian contractual arrangements covered in the AD-SILC dataset.

compare the multiple jobs performed within a year and order them according with the following criteria: level of earnings, duration of the working relationship, level of social security contributions accrued, latest job position held within the year, level of employment stability (e.g. open-ended contracts are more stable than fixed-term contracts). As previously mentioned, ad-hoc estimates are run to assign working categories to working pensioners. Because of the smaller sample, they are not split by sex and are not allowed to work as public employees, who represent a negligible fraction for working pensioners in the data.

For non-retired employees, a further logit regression determines who works in the public or private sector and who works part-time or full-time. As public employment meets the need to provide public services which in turns depend on the size of the population (at parity of public services provided), the share of open-ended public employees is kept constant at 2015 levels. The relative share of other working categories over total employment in the model is not aligned.

After determining the work type, the following process assigns the number of months worked within the year. Self-employed workers, professionals and permanent employees are assumed to work all year. For the first two categories, because of the nature of the job, AD-SILC data do not allow to obtain a precise estimate of months worked, while for open-ended employees, working for a fraction of the year is not really indicative of a state of precariousness, but rather a result of the fact that workers can be hired or fired at any point within the year, and not necessarily at the beginning. For all other workers (fixed-term employees and atypical workers), a logit model determines who is working for the whole 12 months. If workers are not employed for a whole year, a random-effect model estimates the number of months worked, by gender.

Finally, we estimate monthly labour income, separately for each of the employment categories simulated in the model but grouping together fixed and open-ended contractual arrangements for private and public employees. Keeping fixed-term and open-ended contracts together allows us to analyse separately men and women. The group of public fixed-term male employees is too small to be studied alone and must be combined with either the corresponding female group or with the corresponding open-ended one. We find that the gender component is more relevant for income prospects compared with the distinction between open-ended and fixed-term arrangements. Because the model allows only one type of job per year, wages in the estimate sample are obtained as the sum of the overall labour income earned, attributed wholly to the main employment category assigned for that year. Possible amounts received as indemnities for maternity, sickness or job suspension are included. The atypical group deserves particular attention, because it is an extremely heterogeneous category, with different profiles by gender. Indeed, it typically includes both experienced consultants and young precarious workers.

For the estimation of labour incomes for working pensioners, all employment categories are grouped and only one equation is estimated.

Monthly labour income is simulated in real terms and then aligned to labour productivity growth and the consumer price index, a rather theoretical and optimistic assumption if one looks at the limited wage growth observed in Italy in the last few decades.

Regression approach

Generally speaking, to predict the different labour market outcomes, the explanatory variables to our regression estimates have been selected among the demographic, socio-economic individual or household characteristics and individual working career features. Among all the variables present in the AD-SILC dataset, we only include those for which we can project the evolution over the simulation period. This selection allows to account for the same degree of heterogeneity in both the regressions and the microsimulation model. Whenever the sample size is large enough, men and women working trajectories follow separate models.[14] Furthermore, all models except those concerning monthly wages and months worked are estimated through pooled OLS estimators, even if AD-SILC has a panel structure. In such data, each coefficient encompasses two sources of variation: within-subject variability and between-subject variability. Pooled OLS estimator simply treats within- and between-group variation as the same (i.e. it pools data across waves). This choice is pursued either because we tested that the fraction of variance due to individual effects is close to nil (e.g. in the logit determining who is working) or in order to guarantee a coherent amount of individual heterogeneity both in the regressions and in the model (e.g. in the multinomial logit) (Shmueli, 2010; Martini and Trivellato, 1997).

For months and wages, instead, we use random-effects models, as in this case individual heterogeneity is relevant and can play a role within the microsimulation model. In particular, we use the predicted values of the random effects for in-sample individuals, while for new-born or out-of-sample individuals we impute these values, drawing from a normal distribution with the estimated mean and standard deviation.

Trying more sophisticated techniques for the estimation models was not successful. We verified the possibility of using dynamic models that include the lagged dependent variable as a regressor and of applying a correlated random effect estimator. The first is motivated by the fact that the labour supply behaviour measured at the individual level displays a great deal of persistence (Booth et al., 1999; Francesconi, 2002). The second advancement might better exploit the panel nature of the AD-SILC sample without incurring in some of the pitfalls associated with the random effect models, i.e. relaxing the very restrictive assumption by which unobserved heterogeneity is not correlated with

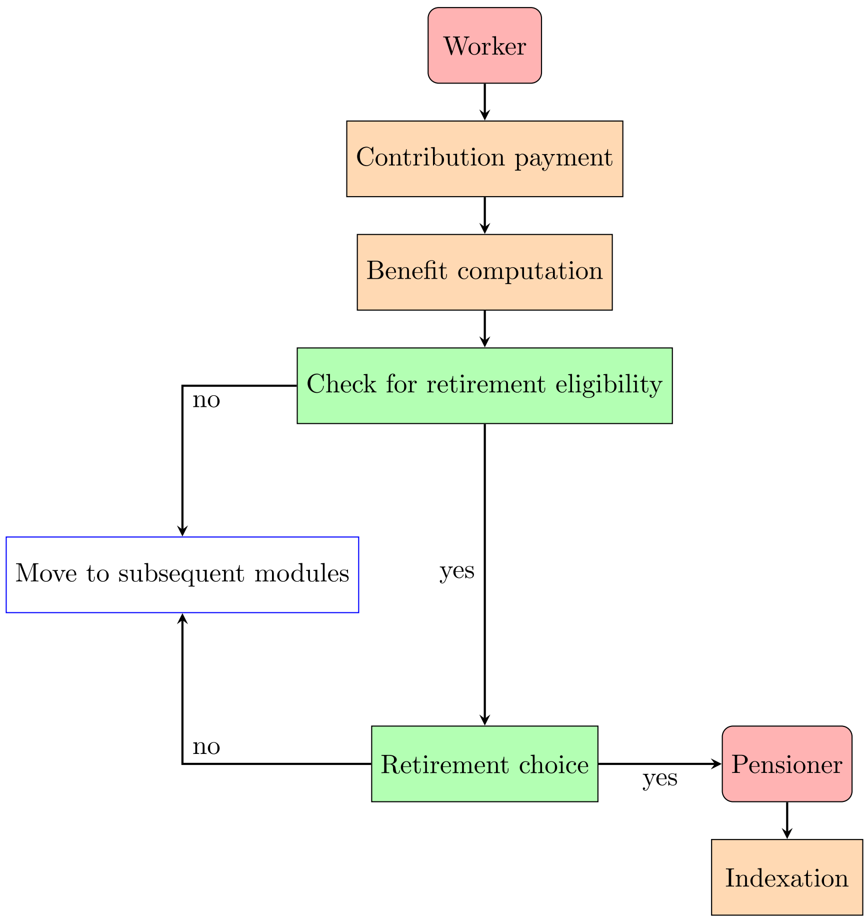
[14] The literature suggests very different employment dynamics by gender (Bertrand, 2020).

the independent variables. Furthermore, we tested Zero truncated Poisson or Zero truncated negative binomial models to fit the asymmetric distribution of months worked, and restricted the range of predicted values to positive numbers. However, all these attempts lead to a misperformance when tested over the projected period, leading to unrealistic and counter-intuitive outcomes in the simulated pattern.

4.3 Pension module

Figure 3 illustrates the essential structure of the Pension module in the latest version of T-DYMM, which largely draws from the experience of previous releases of the model.

Figure 3: Structure of the Pension module, work-related public pensions (1st Pillar)



Workers contribute to the first (public) pension pillar on a mandatory base, with contribution rates set accordingly to the employment category assigned in the Labour Market module. Every year, a potential pension benefit is computed according to the pertinent pension regime. Present contributors to the pension system can be divided into two main categories[15]:

1. '(Pure) NDC' (*contributivo puro*), for workers with no seniority prior to 1996, for whom benefits are entirely calculated according to Notinal Defined Contribution (NDC) rules.[16]

2. 'Mixed' (*misto*), divided into: i) 'Mixed 1995', for workers with less than 18 years of seniority in 1995, for whom benefits are calculated according to the NDC rules *pro rata* for all years of seniority following 1995; ii) 'Mixed 2011', for workers with at least 18 years of seniority in 1995, for whom benefits are calculated according to the NDC rules *pro rata* for all years of seniority following 2011.

While present contributors all compute at least a portion of their pension benefit according to NDC rules, a very consistent portion of present retirees receives a pension that was entirely calculated according to the old Defined Benefit (DB) rules.

[15]Further specification for 'Misto A' and 'Misto B' and for public sector workers are taken into account within the model.
[16]NDC computation rules were first introduced by the so-called 'Dini Reform' in 1995 and then extended to all workers by the so-called 'Fornero Reform' in 2011.

After potential benefits are computed, individuals are checked for retirement eligibility. Table 2 illustrates the various modalities to access retirement in T-DYMM, according to the Italian legislation as of 2021.

Table 2: Eligibility requirements for retirement as simulated in T-DYMM

Criteria	Regime	Requirements	2021
Old age 1	NDC	age	64 years
		seniority	20 years
		amount	2.8* *assegno sociale*
Old age 2	NDC, mixed	age	67 years
	NDC, mixed	seniority	20 years
	NDC	amount	1.5* *assegno sociale*
Old age 3	NDC	age	71 years
		seniority	5 years
Seniority	NDC, mixed	seniority, males	42 years, 6 months
		seniority, females	41 years, 6 months
Seniority - young workers	mixed	seniority	41 years, 12 months accrued before turning 19
Seniority - *Quota 100*	mixed	age	>62 years
		seniority	>38 years

Note: The *assegno sociale* is the social allowance for the elderly. Since 2018, age requirements are aligned to Old age 2. Concerning Old age 2 seniority requirement, 15 years suffice for workers with at least 15 years of seniority as of Dec 31st 1992. For 2022, Seniority - *Quota 102* has replaced *Quota 100* and the age requirement has risen to 64.

Age requirements for 'Old Age 1', 'Old Age 2' and 'Old Age 3' criteria and seniority requirements for 'Seniority' and 'Seniority –young workers' criteria are updated every two years in line with variations in life expectancy at 65 years of age, as established in 2010.[17]. 'Seniority – Quota 100' was introduced in 2019 for the 2019-2021 period, then renewed as 'Quota 102' for 2022 (the age requirement is set at 64, 2 years more than the original 'Quota 100')

In its present version, T-DYMM does not simulate retirement according to the so-called 'APE' criterion, introduced in 2017 and discontinued in 2020 after limited participation, nor the '*Opzione donna*' criterion, by which female workers belonging to the 'Mixed' regime may access retirement many years in advance (58 years of age as of 2021) if they choose to switch entirely to NDC computation rules.

In the previous releases of T-DYMM, retirement decisions were purely deterministic: individuals accessed retirement as soon as they were entitled to. Such an assumption may seem acceptable as of today, as age requirements have raised rapidly in the past few years, especially for women. However, as Notional Defined Contribution (NDC) rules phase in, average pensions are expected to lower and a strong economic incentive to postpone retirement to increase benefits (both by increasing contributions accrued and by reducing life expectancy at retirement) will kick in. By assuming that workers retire as soon as possible, we are implicitly assigning a stronger preference to spending more time in retirement rather than getting a higher benefit to all workers. A choice function that differentiates among different profiles may provide a better representation of reality. The latest version of T-DYMM offers an initial attempt at that.

Workers who meet eligibility requirements for retirement undergo a choice process. The choice function introduced is based on an option value model (Stock and Wise, 1990). For the structure of the model we took inspiration from Van Sonsbeek (2010), but we use the parameters estimated for Italy by Belloni and Alessie (2013). If workers meet the requirements for retirement, they will choose between either ceasing or continuing to work. For each year between the present and the 'maximum retirement age' (the age requirement for the 'Old Age 3' criterion, presently set at 71 and subject to future increases according to changes in life expectancy), utility to access pension in the current period is confronted with the utility obtained by postponing retirement in yearly steps. Future salaries are assumed equal to the latest positive salary, augmented by labour productivity and inflation growth (for each year in the future, growth values in the choice function are assumed equal to the average between period t and t-1). Individuals only consider the option of working a whole year, because in case employment is discontinued, workers will have the possibility to access retirement (seeing as they meet the requirements), regardless of the choice they made beforehand. It is therefore reasonable that individuals shall not consider unemployment risks. For every year in the cycle, there is a different hypothetical retirement age and a different first hypothetical pension, which is computed according to the hypothetical working history accrued on top of the actual working history (the one already known at the time of the decision making). Utilities are computed as a weighted (by survival probability) sum of future salaries and pensions, reevaluated and discounted and net of taxes and contributions. If the utility to access retirement in the current period is higher than in any other option (delay by one year, delay by two years, ..., delay until the maximum retirement age is reached), retirement is accessed. As previously mentioned, regardless of the choice made here, if individuals cannot later find a job in the labour market, they will access

[17]A Decree Law in 2019 has suspended updates to life expectancy variations until 2026.

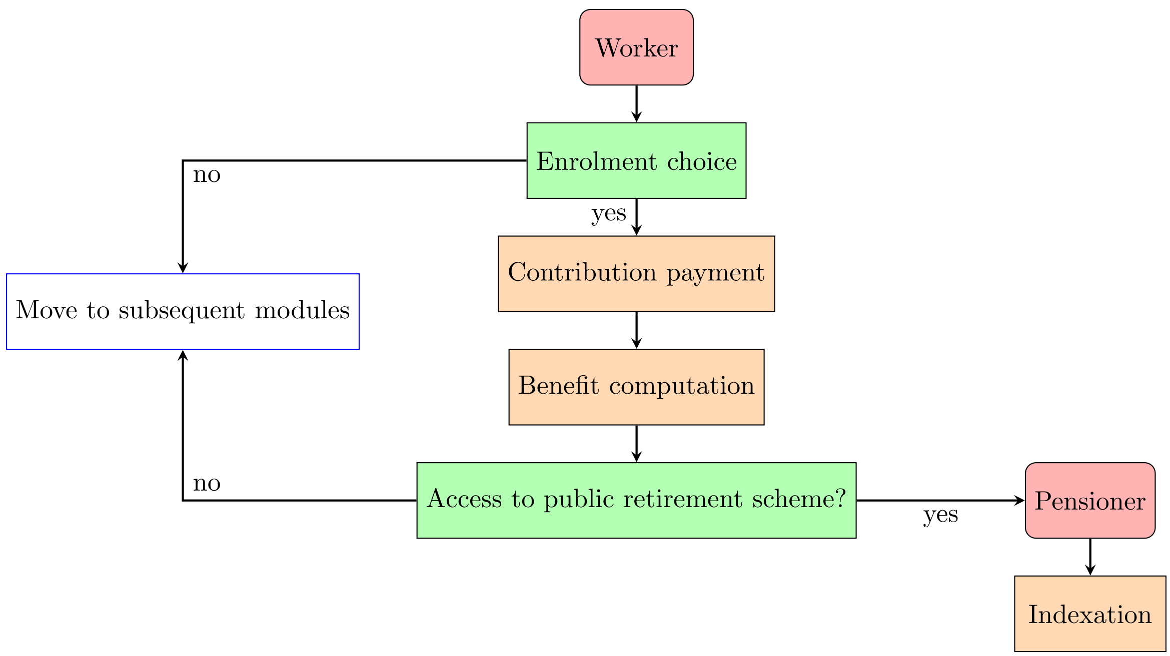
retirement. Parameters exogenous to the model are: i) discount rates (labour productivity and inflation, in our case); ii) risk aversion and iii) leisure preference. Parameters for risk aversion and leisure preference (by gender) are derived from Belloni and Alessie (2013).

Once workers access retirement, their pension is paid out and for the following periods it is indexed to price inflation according to the pertinent legislation, which only allows full indexation to pensions below a certain threshold amount (in 2021, below € 20,000 annually). For the period 2019-2021, an *ad hoc* temporary reduction on pensions above € 100,000 annually (so-called '*pensioni d'oro*') is also in place.

Besides seniority and old-age pensions (work-related pensions), in the Pension module we also simulate the integration to a minimum amount for pensions of workers belonging to the 'Mixed' regime, inability pensions for workers that fall ill or disabled and survivor pensions. For inability pensions, we simulate the legislation put in place in 1984, which introduced the *Assegno ordinario di invalidità*, for severely disabled workers, and the *Pensione di inabilità*, for workers unable to work because of disability. Individual probabilities to receive these benefits are based on regression parameters estimated on AD-SILC, which highlight the high persistency of the phenomenon and the relevance of the disability state (simulated in the Demographic module; see Section 4.1). Probabilities to receive inability pensions are aligned by gender and 5-year age class to INPS data available for the period 2016-2019; beyond 2019, probabilities are projected following the same logic adopted for disability probabilities.

T-DYMM's pension module also comprises a sub-module on private pensions. Figure 4 illustrates its structure.

Figure 4: Structure of the Pension module, private pensions (2nd and 3rd Pillar)



Differently from what happens in the 1st Pillar, workers participate to private plans on a voluntary basis. Individual probabilities are based on regression parameters estimated on AD-SILC, which highlight the role of age, labour income, financial literacy, education, employment category, net wealth and impose a high level of persistence to the phenomenon. The probability to contribute is aligned, irrespective of age and gender, according to data from COVIP; in projection years, it is kept constant to the latest available figures. Contributors to the 2nd pillar (*fondi negoziali*, collective funds) may devolve their TFR (*Trattamento di Fine Rapporto*, end-of-service allowance)[18] and voluntary contributions, while for the 3rd pillar (either *fondi aperti*, open funds, or *piani individuali pensionistici*, individual pension plans) contribution to the fund may vary yearly depending on labour income and net wealth. Investments in the 2nd and 3rd pillar produce returns that are computed following COVIP data for past periods and projections based on the portfolio composition of pension funds and on assumptions on portfolio components for the rest of the simulation period (for a description of the assumptions on various financial assets, see Section 4.4). When individuals access retirement in the public pillar, they are

[18]The TFR is a sort of mandatory severance payment for public and private employees. It acts as a deferred share of wage: TFR contributions are withheld and managed by the employer (the accrual rate is mandated by law and equal to 1.5% plus 75% of inflation growth), who has to pay the accumulated amount to the employee in the event of dismissal or retirement (the TFR may also be anticipated in part under a very limited set of circumstances; T-DYMM only considers the possibility to withdraw up to 70% of the TFR in order to buy a first house). A crucial change in the legislation concerning TFR took place in 2007, when the so-called 'silent consent' formula was introduced: if workers do not explicitly disagree, their TFR flows (not the stock already accrued by firms) are transferred from firms to pension funds. According to the latest available data from COVIP (2018), only about 23% of the overall amount of accrued TFR has been transferred to pension funds.

also assigned an annuity (if any investment is present) from the 2nd and/or 3rd pillar, which is henceforth indexed.

4.4 Wealth module

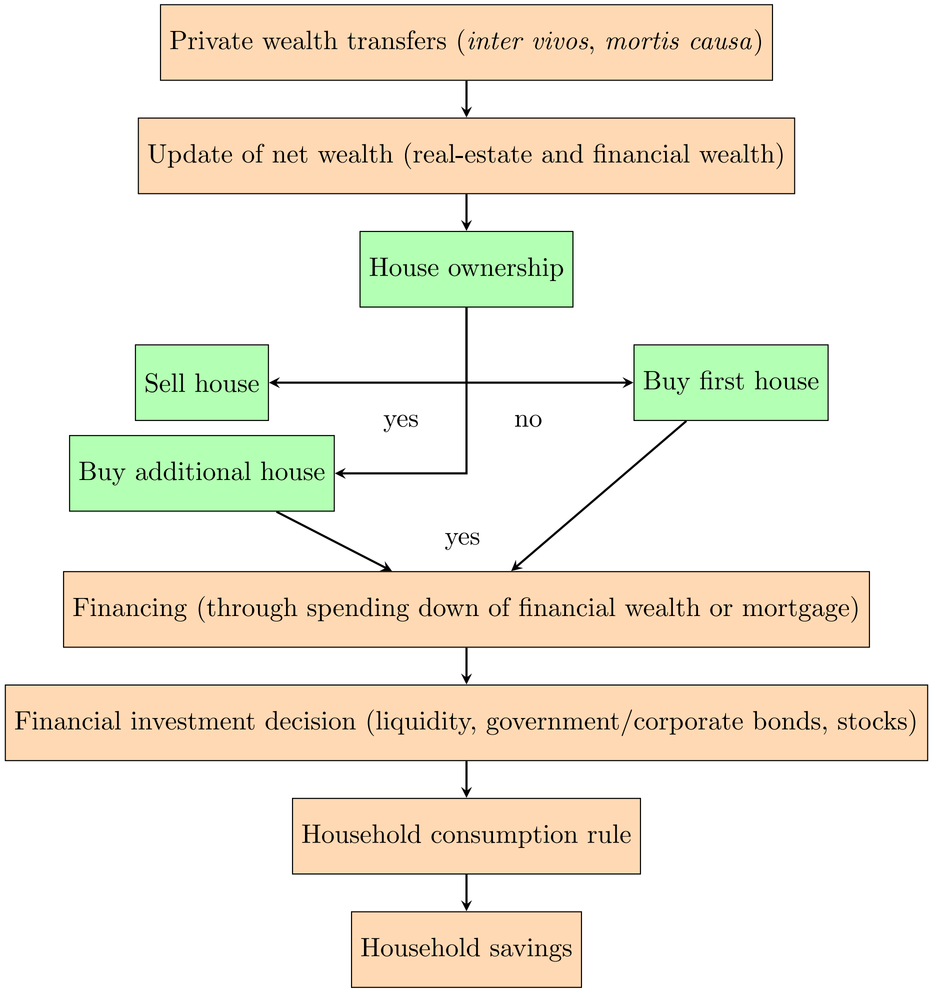
One of the main novelties of T-DYMM 3.0 regards the introduction of a Wealth module that accounts for household wealth dynamics. Modelling private wealth provides a more complete picture of disposable income and households' well-being distribution, also in relation to the adequacy of the social security system. We define net wealth as the sum of real and financial wealth net of liabilities. Houses ownership is the only form of real wealth in our model, while financial wealth is divided into four activities: liquidity, government bonds, corporate bonds and stocks, which correspond to four different risk-return combinations.[19]

The structure of the Wealth module is based on Morciano et al. (2013). It is composed of different processes, which are illustrated in Figure 5. The scheme summarises the multiple processes that are included in the Wealth module in T-DYMM 3.0. The processes in the model are sequential, as presented from top to bottom in the flow-chart, although they are related amongst one-another overtime. For instance, the acquisition of real estate implies a down-spending of financial wealth, the opposite stands for selling. Every step of the Wealth module involves choices made at the household level that are modelled through regressions and alignments (when needed). The estimates adopted in the model are based on SHIW micro data (biennal waves in the 2002-2016 time span). We use discrete choice models (logit) for transitions (for instance buying/selling houses, making/receiving inter-generational transfers, renting second dwellings) and continuous regressions for quantities (either log levels or ratios of income or financial wealth). In Appendix F, we illustrate the regression results underlying some of the estimates included in the Wealth module.[20] The general criteria that guided the selection of the explanatory variables of the estimates are: i) economic relevance of the variables with respect to the dependent variable; ii) significance of the estimated coefficients. The two criteria did not have a hierarchical order but were combined to obtain the most reasonable result in an economic and econometric sense. As mentioned above, alignments are a key part of a dynamic microsimulation model. Even though it is hard to find the right long-run alignment parameters for wealth stocks and related processes, we decided to align some specific processes for which external data sources or projections are available. We use data from ISTAT and the Department of Finance as we will specify in the following.

[19] In the present version of the model, investors do not experience risk, as return vectors contain average values and no idiosyncratic component is simulated.

[20] We show estimates on the probability of investing in financial activities, the household consumption level and the probability of investing in private pensions (II and III pillar).

Figure 5: Structure of the Wealth module



Private wealth transfers
The starting processes involve intergenerational transfers, *inter vivos* (donations) and *mortis causa* (inheritances). Inheritance is driven by demography in the sense that the total amount of transferred wealth equals the wealth of the deceased. Receivers are selected deterministically, as the offspring and/or the married partner of the deacesed, if present in the sample, or probabilistically, through regressions based on SHIW. At the start of the simulation, all individuals living outside of their original households cannot be linked to their parents, hence the need to simulate some inheritances in a probabilistic fashion. *Inter vivos* transfers are based on SHIW data on both the donor and recipient side; the number of households who donate and who receive and the related total amounts are imposed to match each other in terms of totals. Both inheritances and donations are reduced by the inheritance tax, but only if the total amount received exceeds 1 million euros, as foreseen by the Italian legislation.

Wealth updating
The second process is the one updating the amount of yearly wealth. Household savings and the TFR (*Trattamento di Fine Rapporto*, end-of-service allowance) are summed up to the existing financial accrues, and the values of house wealth and financial wealth evolve over time depending on nominal rates of returns. House wealth evolves following inflation, assuming that the real market does not incur in a significant value increase or decrease over time, while income gains follow returns data from OMI (*Osservatorio del Mercato Immobiliare*, the Italian Housing Observatory) and are projected up to 2070 in accordance with the evolution of the implicit interest rate on the Italian public debt (computed in line with the European Commission methodology, following the assumptions underlying the latest Ageing Report from AWG). As for financial wealth, each of its components evolves differently as sketched in Table 3. Liquidity is not updated over time, meaning that its value is eroded in real terms. For the remaining financial investments, except for government bonds, we have divided returns into two components: income and capital gain. The former derives from interests on financial assets, while the latter indicates the evolution of asset values, therefore their setting will affect the evolution of the wealth-to-income ratio in the simulations. Government bonds are assumed to produce no capital gain in real terms (their value

is updated in line with inflation), while their income gain follows the implicit return rate on public debt. For corporate bonds and stocks, we use information from S&P 500 to derive both income and capital gain in historical periods, while for future years we use the projected implicit return rate on public debt for the income component and for the capital gain we attribute an extra yield computed as the historical average return differential with respect to government bonds, kept constant from 2024 (first year beyond the forecast horizon of the latest European Commission forecasts at the time of writing) to 2070. From the latest historical data available (2020) to 2024, a linear convergence is assumed.
Finally, mortgages evolve based on the long-term (10-year Italian government bonds) interest rate for both historical and projected periods. Like the implicit return rate on public debt, the long-term interest rate is also computed following the European Commission methodology and the assumptions underlying the latest Ageing Report from AWG.

Table 3: Projected rate of nominal returns adopted in the wealth module

Wealth type		**2016-20**	**2021-70**
House wealth	Income gain	OMI	Projections based on OMI
Government bonds	Income gain	Implicit rate on debt, AWG	Implicit rate on debt, AWG
Corporate bonds	Income gain	S&P 500	Implicit rate on debt, AWG
	Capital gain	S&P 500	Projections based on S&P 500*
Stocks	Income gain	S&P 500	Implicit rate on debt, AWG
	Capital gain	S&P 500	Mark-up stocks-bonds*
Mortgages		Long-term interest rate, AWG	Long-term interest rate, AWG

Notes: * For the 2021-2023 period a linear convergence is applied.

House ownership
Every household has a probability of buying and selling a house based on regressions estimated on SHIW data. Every year of simulation, the number of houses bought equals the number of houses sold and these values are aligned to the national statistics from ISTAT (historical values for 2016-19, then the 2019 value is kept constant over time). The values of the houses sold are deterministically computed within the model, whereas the values of the houses bought are computed through regressions. However, the total value of houses bought equals the total value sold. House acquisition is financed through down-spending of financial wealth, the accrued TFR (70% of the total, in line with the pertinent legislation) and residually, through mortgages (which are the only form of liability in the model). The values of new mortgages can not exceed 60% of household income. An extra process that is connected to house property is the event of renting, which implies the receipt of a further source of income for owners and an expenditure for households who are not house-owners. The choice whether to rent or not real estate in excess (renting the first house is not allowed in T-DYMM) is modelled with a regression based on AD-SILC data (specifically, the information contained in tax returns allow to study this choice). The households that do not own a house may or may not live in a rented house (the possibility of loan to use is taken into account). To model this circumstance we use a regression based on SHIW data. The amount of rent received and paid are simulated as a share of household income.

Financial investment decision
As said above, T-DYMM simulates four different types of financial activities. Financial investment decisions are modelled in two steps: we first simulate ownership and then the amount owned. We model in a probabilistic fashion the choice of whether to invest in government bonds, corporate bonds or stocks through dynamic regressions based on SHIW data. In order to estimate the dynamic relationship between ownership at time *t* and *t-1* and control for the initial conditions problem, we check for the value of the dichotomous dependent variable in the first year of observation (whether they owned that class of financial activity in 2010) and we average all time-varying variables, following the approach by Wooldridge (2005). In the simulation, we do not use the coefficients for initial conditions and averages, but we consider them as good instruments for improving the precision of coefficients of the lagged dependent variable. The breakdown of total financial wealth into the various asset groups is obtained through regressions based on SHIW data with the ratio of the asset amount over the total as the dependent variable.

Household consumption rule
The last process in the Wealth module is the household consumption decision. This process is one of the most relevant since household savings flow into the household budget the following year in the form of financial wealth. At the end of the simulation period, every household is endowed with an amount of disposable income and the model attributes them a certain level of consumption that may or may not exceed the household disposable income; in the first case, the household will use its financial wealth as a supplementary source to finance its expenditure. Consumption levels are determined through a panel regression based on SHIW data for the period 2002-2016 where the dependent variable is the logarithm of consumption. We adopt a fixed effects estimator, where the estimated correlation between the vector of explanatory variables and the unobserved time-invariant residual is included in the simulation.

The results of the regression estimates highlight an issue related to the difference between micro data and macro aggregates on consumption and savings rate. As well-known in the literature (Cifaldi and Neri, 2013), the discrepancy between the savings rate obtained from SHIW data and the one obtained from National Accounts is large, both in terms of levels and propensities. As a result, our choice is to align the average level of consumption to the national savings rate and keep it constant for all the simulation period. For the initial years of the simulation we use actual data from ISTAT (the savings rate equals 9.8% in 2019). The projection is made using a logarithmic function (the savings rate has decreased to 8.25% by 2070). In other terms, aggregate savings rate is exogenous to the model while it is endogenous at the household level.

4.5 Tax-Benefit module

At the lowest level of the model hierarchy comes the Tax-Benefit module, which simulates taxes paid and benefits granted at the national level (according to the legislation of 2021). T-DYMM does not encompass an internal migration sub-module. Hence, no regional- and municipal-level taxes and transfers are simulated.

On a similar note, we do not simulate COVID-related monetary transfers.[21] The economic crisis following the pandemic has seen the intervention of the government through salary integration, favoring the recourse to labour hoarding behaviours (above all the use of the *Cassa Integrazione Guadagni*, CIG), and through lump-sum transfers to self-employed workers, for an estimated cost of the two measures combined of 6.5 billion euros in March 2020 alone (Figari et al., 2020). As for the former measure, we do not simulate salary integration related to any sort of transitory job inactivity (e.g. suspension or reduction of work activity, maternity, sickness) separately from other labour income, and thus we do not allow the simulation of COVID-related salary integration. As a result and to preserve internal model coherence, we opted for not simulating lump-sum transfers, which could have been possible through randomised assignment only, due to data availability constraints. A further explanation as to why COVID-related measures are not covered in T-DYMM resides in the medium- and long-term focus of our analyses, which makes the simulation of temporary and emergency measures of secondary relevance.

We assume that tax-benefit monetary parameters (e.g. PIT brackets, threshold levels of tax expenditures, benefit amounts, and so on) follow nominal GDP growth starting from 2024, the first year beyond the forecast horizon of the latest European Commission forecasts (2022 Spring Forecasts, at the time of writing).[22]

The module performs the calculation of social insurance contributions (SICs), direct taxes on different income sources (i.e. labour and retirement income, capital income, rental income) and in-cash social transfers, both means-tested and non-means-tested payments. In what follows we provide a brief overview of the module's structure by focusing on its sequence and coverage in terms of simulated measures, as well as methodological implementation.

The starting process of the Tax-Benefit module is the calculation of SICs, which draws largely on the Italian country component of the EUROMOD model, to which reference is made for a more detailed explanation (Sutherland and Figari, 2013). We simulate employer and employee contributions collected for the payment of old-age/seniority, survivor and inability pensions, as well as contributions related to the payment of unemployment benefits, redundancy pay, sickness and maternity pay and family allowances. We also simulate contributions paid by self-employed workers.

Following the sequence of the module, we then move to the computation of proportional taxes (see the full list in Table 4). Even though the personal income tax contributes by far the most to the redistributive effect of the Italian tax-benefit system (Fuest et al., 2010; Boscolo, 2022), proportional taxes have grown significantly in recent years. In fact, a greater share of self-employment income and rental income previously included in the PIT base is now excluded and subject to proportional taxation. Self-employed workers can opt for substitute tax regimes conditional on certain income and organisational criteria, and all individuals, regardless of their working status, can subject rental income from residential properties to more favourable taxation rather than to the personal income tax. In both cases, we select individuals in the simulation by using logistic regression analyses on tax returns micro data for the year 2015 among those who meet statutory requirements based on our data availability. Furthermore, income sources exempt from progressive taxation are relevant when it comes to the calculation of social transfers, given that they are included in the means-testing process.

[21] Salary integration to employees, lump-sum transfers to the low- and medium-income self-employed and to employees working on site in the first month of the pandemic, mortgage suspension for the main residence granted to the self-employed, the so-called *Reddito di Emergenza*, an economic support measure in favor of households in severe difficulties in the period March-May 2021, and other minor national and sub-national measures.

[22] We are aware of the fact that some of these parameters are not automatically linked to GDP growth nor to inflation. However, in the context of baseline projections, assuming that they stay constant in the long run (up to 2070) is not reasonable.

Table 4: SICs and taxes simulated by T-DYMM (order of appearance follows module sequence)

SICs:
employer social insurance contributions
employee social insurance contributions
contributions paid by self-employed workers
Proportional taxes and tax regimes that substitute the personal income tax for:
i. capital income: government/corporate bonds and shares*
ii. private pensions: II and III pillars**
iii. self-employment income subject to *regime fiscale di vantaggio**** or *regime forfetario***
iv. rental income subject to *cedolare secca* (assigned to the head of the household)**
v. productivity bonuses**
Personal income tax (*Imposta sul reddito delle persone fisiche – IRPEF*)****

Notes: * The process 'Financial investment decision' is modelled through dynamic regressions based on SHIW data. More on this in Section 4.4. ** Recipients are aligned to aggregate administrative data in the 2016-2019 interval, while from 2020 onwards we align recipients by taking their external total as of 2019 as reference and update it to: the population growth at the individual level for *ii*; the population growth of self-employed workers for *iii*; the population growth at the household level for *iv*; the population growth of employees for *v*. *** Recipients are bound to gradually diminish to zero under current legislation. We assume that there are no recipients by 2030. **** Recipients of residual tax expenditures are aligned to external totals derived from (weighted) tax returns micro data for the 2015 tax period, annually updated to the population growth of recipients of gross income subject to PIT.

As for the personal income tax, it is worth specifying how deductions and tax credits are calculated. The implemented strategy is in line with previous experiences in microsimulation studies (Albarea et al., 2015). The most sizeable tax expenditures in terms of granted tax relief are fully simulated by the model[23], both with regard to beneficiaries and amount. We determine beneficiaries of residual tax expenditures by using logistic regression analyses on pooled tax returns micro data covering a 7-year interval (2009-2015) and then calibrate amounts with aggregate statistics by income groups. For each calibrated tax expenditure, potential beneficiaries are selected among those with relevant characteristics for eligibility. This procedure allows for a more precise simulation of the personal income tax and overall net liabilities. At the same time, given the growing attention that has been paid to reforming the system of direct taxation in Italy, it contributes to making T-DYMM a reliable tool that could add to the current discussion by focusing on mid- and long-term redistributive effects of proposed tax reforms.

Subsequent to the simulation of SICs and taxes, the module performs the calculation of in-cash benefits as illustrated in Table 5. In its current version, T-DYMM assumes the full take-up rate for all benefits except for disability allowances, minimum income schemes and unemployment benefits. For the latter, recipients are selected according to a score estimated on the probability of being unemployed in the EU-SILC survey. As for minimum income schemes, we randomly choose households among those who meet criteria for eligibility. In both cases, a strong persistency for the state 'recipient' is imposed, i.e. if a person or household satisfies requirements in period *t-1* and receives the benefit, there is a very strong chance they will also receive it in period *t*, if they still meet requirements.

[23] As for deductions, we refer to social insurance contributions paid by self-employed workers, cadastral value of the main residence and contributions to private pension plans; regarding tax credits, the most relevant ones are the tax credit for labour and retirement income and tax credits for dependent family members.

Table 5: In-cash benefits simulated by T-DYMM (order of appearance follows module sequence)

Unemployment benefits (NASpI and DIS-COLL)* **
In-work bonus for employees and atypical workers (*Bonus IRPEF*, which has replaced *Bonus 80 euro*)
Means-tested disability allowances (*Pensione di inabilità agli invalidi civili* up to the Standard Pensionable Age and *Assegno sociale sostitutivo* afterward)***
Non-means-tested disability allowances (*Indennità di accompagnamento* for those aged 18 or above and *Indennità di frequenza* for those aged below 18)***
War pensions and indemnity annuities (*Rendite indennitarie*)
14th month pension****
Social allowance for the elderly and related increases (*Assegno sociale* and *Maggiorazioni sociali*)
Increases to old-age/seniority, survivor and inability integrated pensions (*Maggiorazioni sociali del minimo*)*****
Family allowances for employees' and pensioners' households (*Assegni al nucleo familiare*)******
Newborn bonus (*Bonus bebè*)
Mother bonus (*Bonus mamma domani*, from 2017 onwards)
Minimum income schemes:**
- SIA (*Sostegno all'inclusione attiva*, 2017)
- REI (*Reddito di Inclusione*, 2018)
- RdC (*Reddito di cittadinanza*, from 2019 onwards)

Note: * Unemployment benefits are actually simulated prior to the Tax-Benefit Module, because they are subject to the personal income tax. ** For the first two years of the simulation, administrative totals are employed for the alignments. From the latest available figures onwards, the ratio between actual (administrative data) and potential (obtained from T-DYMM's simulations) recipients is kept constant. *** Average probabilities to receive disability allowances are aligned by gender and 5-year age class to INPS statistics available for the period 2016-2019; beyond 2019, probabilities are projected following the same logic adopted for disability probabilities (which in turn follow the Reference Scenario of the 2021 Ageing Report). **** Recipients in T-DYMM's base year sample will hold these benefits until death. New occurrences are not simulated. ***** Integrations to old-age/seniority, survivor and inability pensions (*Integrazione al trattmento minimo*) are included among pension benefits subject to PIT, and thus not listed in the above table. ****** This measure has recently been replaced by a more comprehensive scheme from 2022 (*Assegno unico e universale*), which is not yet simulated in the current version of the model. The new allowance has also replaced tax credits for dependent children under the age of 21, the newborn bonus, the mother bonus, and family allowances granted at the municipal level.

5 Baseline results

In the present Section, we briefly present illustrative findings from T-DYMM's baseline scenario for the simulation period 2016-2070. The demographic and macroeconomic framework is aligned to the European Commission 2022 Spring Forecasts for Italy and to the baseline scenario underlying the 2021 Ageing Report. In Table 6, the values for the main demographic, macroeconomic and financial aggregates underlying the simulation are reported.

Table 6: Baseline scenario, economic and demographic framework

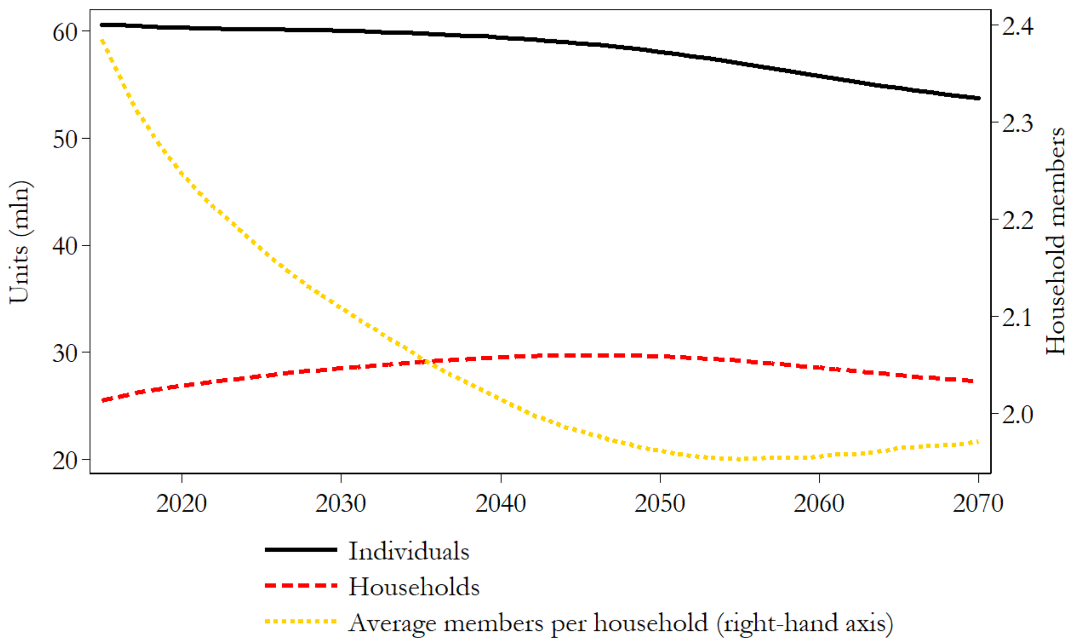
		2020	**2030**	**2040**	**2050**	**2060**	**2070**
Demographic	Life expectancy at 65	21.3	22.2	23.1	24.0	24.8	25.6
	Total Fertility Rate	1.3	1.4	1.4	1.4	1.5	1.5
	Net migration flow	160,697	223,984	217,230	214,328	210,477	206,564
Macroeconomic (%)	Real GDP	-9.0	0.4	1.1	1.5	1.4	1.3
	Labour productivity	0.4	1.3	1.7	1.7	1.6	1.5
	GDP deflator	1.4	2.1	2.0	2.0	2.0	2.0
	Consumer Price Index	-0.1	2.0	2.0	2.0	2.0	2.0
Financial (%)	Total return on government bonds	2.4	2.5	3.2	3.5	3.7	3.7
	Total return on corporate bonds	7.9	2.9	3.6	3.9	4.1	4.1
	Total return on equity	11.9	3.4	4.1	4.4	4.5	4.6
	Return on house wealth	4.2	4.3	5.0	5.3	5.4	5.5

Source: Eurostat, European Commission, OMI, S&P 500

All results in the first years of the simulation may be evaluated alongside administrative population data (when available), but in Appendix G we focus specifically on the external validation of some key outcomes.

While most of the attention in this Section is devoted to adequacy concerns on the side of the pension and welfare systems, we shall first examine how T-DYMM's sample evolves in its demographic structure. In accordance with recent historical data and projections, the sample steadily shrinks over time in terms of individuals (Figure 6). By 2070, the sample reduces by almost 12% compared to 2015, perfectly in line with Eurostat projection for the resident population. In the first years of the simulation, the increase in the propensity to divorce and the reduction in the propensity to form couples observed in the recent data (see Section 4.1) produce an increase in the number of (mostly single) households. Once this process stabilises, the two dynamics for households and individuals embark on a similar pattern. By 2070, households will have increased by almost 7% compared to 2015. In line with the historical trend observed by ISTAT[24], the average household size goes from about 2.4 in 2015 to about 2 in 2070.

Figure 6: Sample evolution, individuals and households



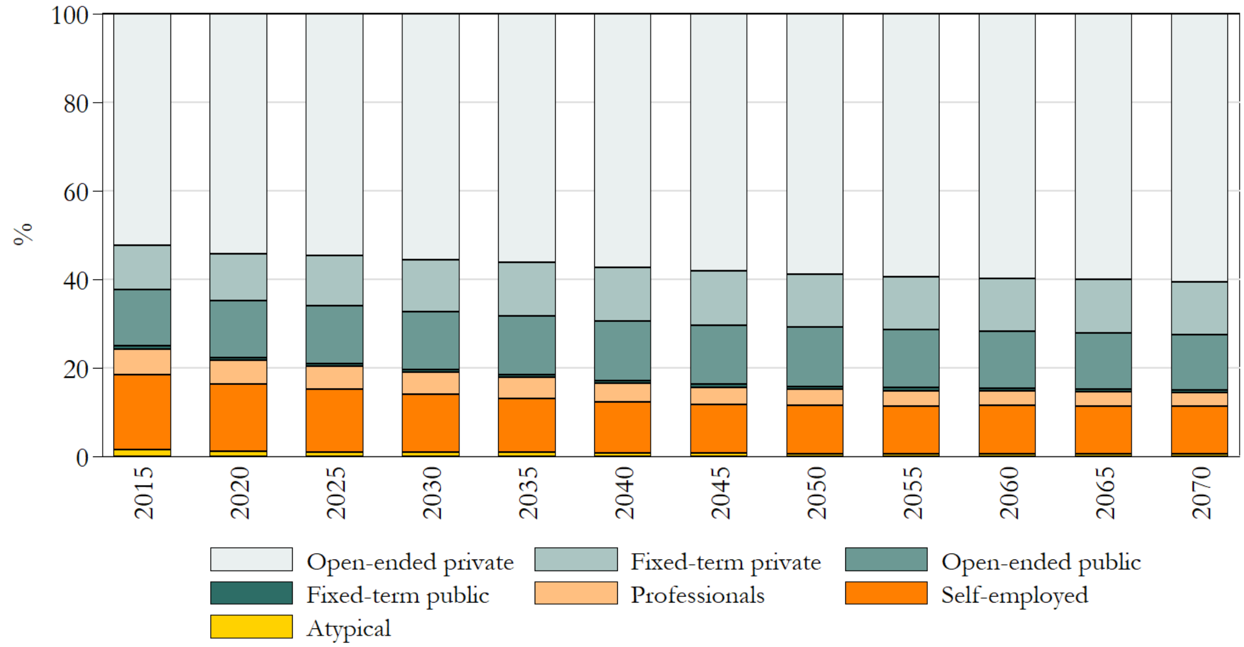
Source: T-DYMM 3.0, authors' elaborations.

Concerning the labour market, a central issue is the composition of employment by work category. Figure 7 illustrates the distribution of work typologies over the simulation period. By far the largest work category is represented by employees with permanent contracts in the private sector. This category absorbs about 52% of total employment at the beginning of the simulation and grows by 8 p.p. over the projection horizon. On the other hand, self-employment becomes less common over time, following a declining trend begun in the 90's that has become more pronounced since the late 2000's.[25]

[24]According to ISTAT, the average number of members per household decreased from 2.7 in 1998-1999 to 2.3 in 2018-2019. See Annuario Statistico Italiano, Popolazione e Famiglie - 2020.

[25]The number of self-employed workers decreased by about 20% from 1990 to 2019. See ISTAT, Flash III Trimestre – Il mercato del lavoro.

Figure 7: Repartition by work category

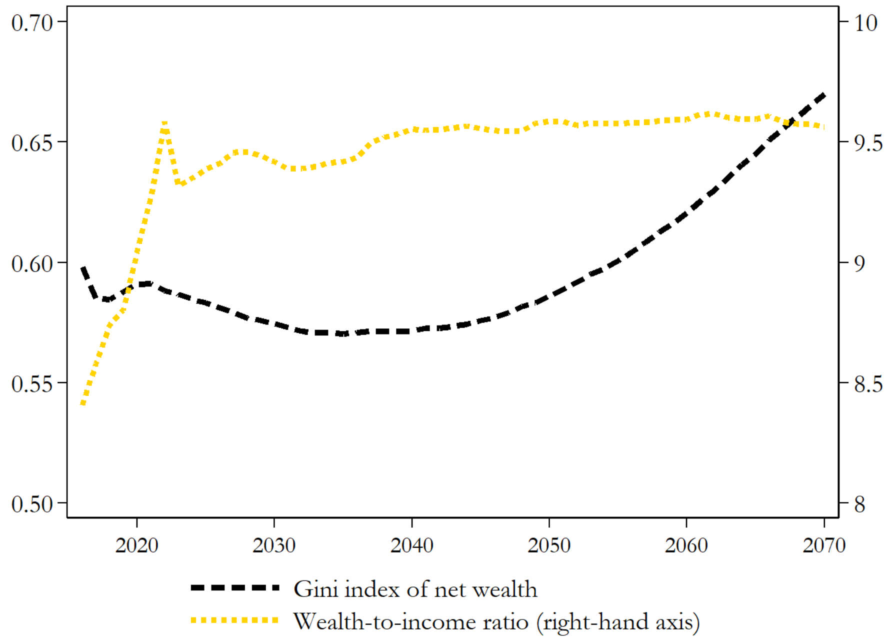


Note: working pensioners are not considered. Source: T-DYMM 3.0, authors' elaborations.

The long-run wealth outcomes are illustrated in Figure 8. We summarise the evolution of wealth by looking at two main indicators, one regarding inequality and the other concerning the stock accumulation. We see from the Gini index of net wealth that the level of inequality is rather stable in the first 40 years of simulation and starts increasing after 2055. This is due to the effect of intergenerational transfers, especially inheritances, whose relevance in explaining the variation in wealth inequality increases over time.[26] Nonetheless, in terms of accumulation and totals we see that the wealth-to-income ratio rises in the historical period (it equals 9.0 in 2020) since the returns rate on financial activities have seen high spikes in these years. The figures are in line with the observed data from ISTAT and Bank of Italy, according to which this ratio equals 8.7 in 2020.[27] Then in the subsequent years (up until 2023) there is still an increase related to the convergence correction applied to the returns rates, that later normalises as we see that the ratio becomes stable over the remaining part of the simulation. The stabilisation mechanism is driven also by the growing role of capital income coming from wealth that operates on the denominator of the ratio, increasing it.

[26] A specific article on the role of intergenerational transfers in shaping wealth distribution using T-DYMM 3.0 is forthcoming.
[27] See ISTAT and Bank of Italy.

Figure 8: Wealth inequality and accumulation

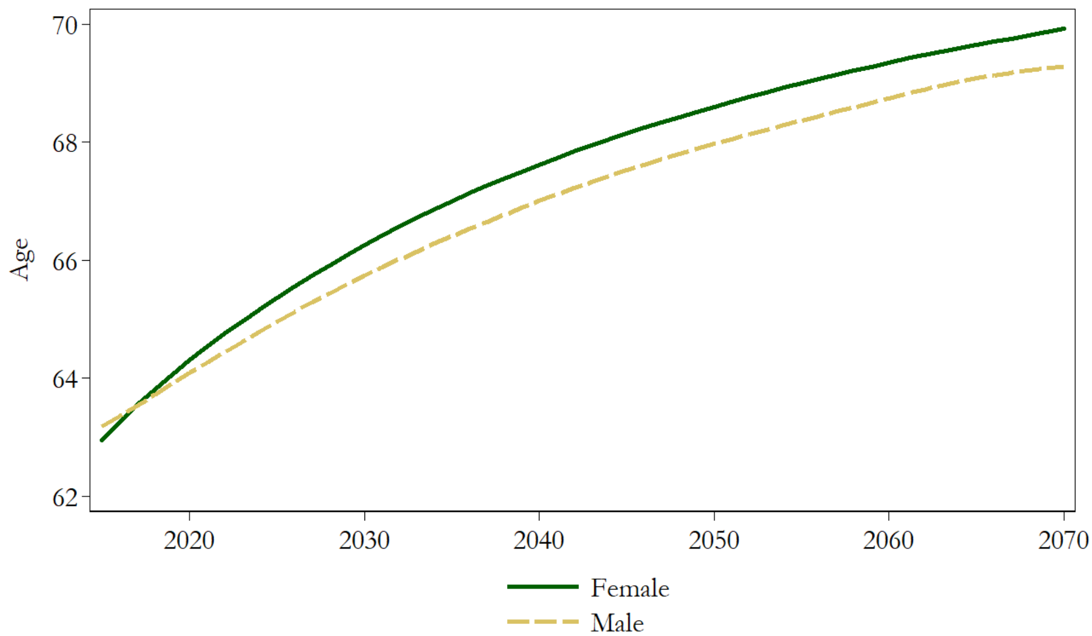


Note: Source: T-DYMM 3.0, authors' elaborations.

T-DYMM has essentially been employed for adequacy assessments of the Italian pension system, hence larger attention should be paid to results stemming from the Pension module. According to T-DYMM 3.0 simulations, average retirement ages would increase by about 5 years for male workers and over 6 years for their female counterparts.[28] Average retirement ages for women equal their male counterparts after the first few years of the simulation and, as a result of discrepancies within the labour market, later surpass it (see Figure 9). Because of increases in retirement ages, the average time spent in retirement is expected to decrease in the long run, especially for women: after a rise of about 2 years in the first two decades of the simulation (due to earlier access to retirement of past workers), retirement duration is expected to lower, from over 25 to 22 for women and from over 21 to less than 20 for men in the 2015-2070 period.

[28]'Old age 3' pensioners are excluded from this statistic in order to facilitate a comparison between beginning and end-of-simulation results. The 'Old age 3' criterion allows retirement to individuals that would have simply lost their contributions under the previous legislation, but pays out very low pensions as a result of the few years spent in the workforce. Furthermore, upon reaching the 'Old age 2' age requirement, poor elderly can always access a social allowance, the *assegno sociale*.

Figure 9: Average retirement age by gender



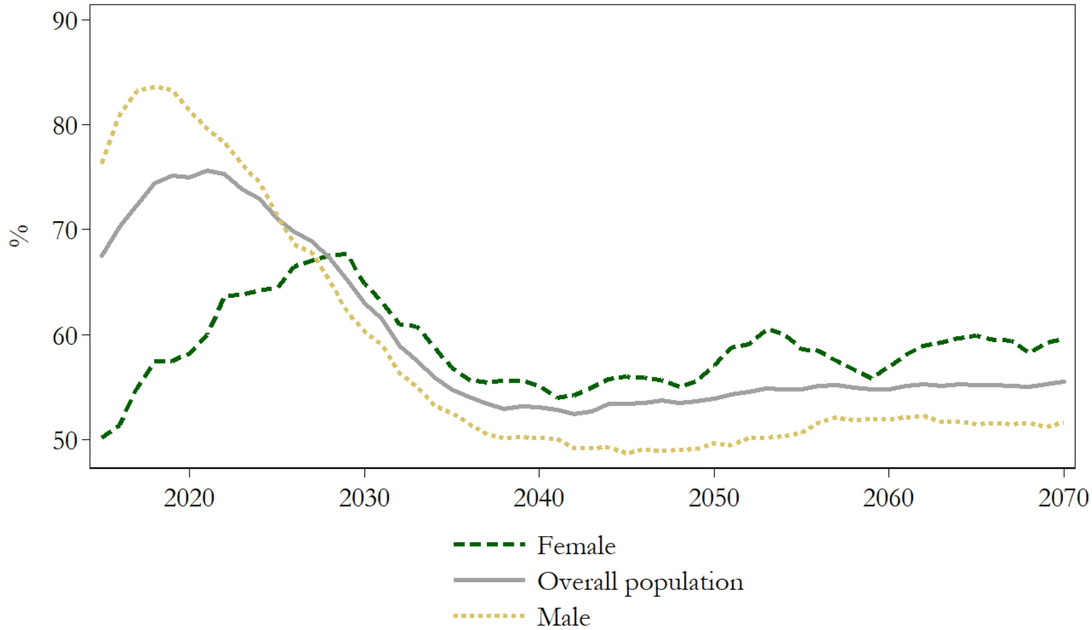
Note: Lowess smoothing. 'Old age 3' pensioners are not included. Source: T-DYMM 3.0, authors' elaborations.

In line with recent trends, the Aggregate Replacement Ratio (ARR)[29] increases by about 10 p.p. in the first decade of the simulation, then decreases and stabilises at a little over 50% after 2040 (Figure 10). If we differentiate by gender, dynamics are opposite in the first 10 years of the simulation: women are still less protected by the pension system than men are, but are projected to recover by 2030 in terms of the ARR. The Gender Gap in Pensions[30] decreases sharply in the first years of the simulation, from 40% in 2016 to about 20% in the 2030s and then stays constant throughout the rest of the simulation period.

[29]The ARR is the ratio between the gross median individual pension income of the population aged 65–74 and the gross median individual labour income of the population aged 50–59. It takes into account old-age/seniority, survivor and inability pensions. The 2021 Adequacy Report from the European Commission (European Commission DG EMPL, 2021) registers a 22 p.p. variation of the indicator in the 2008-2019 period.

[30]The GGP is calculated for persons aged 65-79 as the percentage difference between the average pension for females and the average pension for males. It takes into account old-age/seniority, survivor and inability pensions.

Figure 10: Aggregate Replacement Ratio by gender



Source: T-DYMM 3.0, authors' elaborations.

Table 7 illustrates the condition at retirement by birth cohort (five-year birth cohorts, from 1960 to 1989) in terms of average age and median gross replacement rate (rr). The position of younger cohorts worsens in terms of both pension level and average age at retirement. Results are somewhat affected by the necessary assumption (due to lack of data) that migrant workers do not carry over any pension rights, which could prove to be overly pessimistic. The transition from DB to NDC computation rules is expected to lower average benefits but also lower inequalities among pensioners, as workers with higher incomes will be the ones impacted the most.[31] Table 7 also compares the condition at retirement by birth cohort for the two lowest (first and second) and the two highest (fourth and fifth) income quintiles. The latter are visibly more affected in terms of reduction in pension amounts.

Table 7: Condition at retirement by birth cohort and income quintile

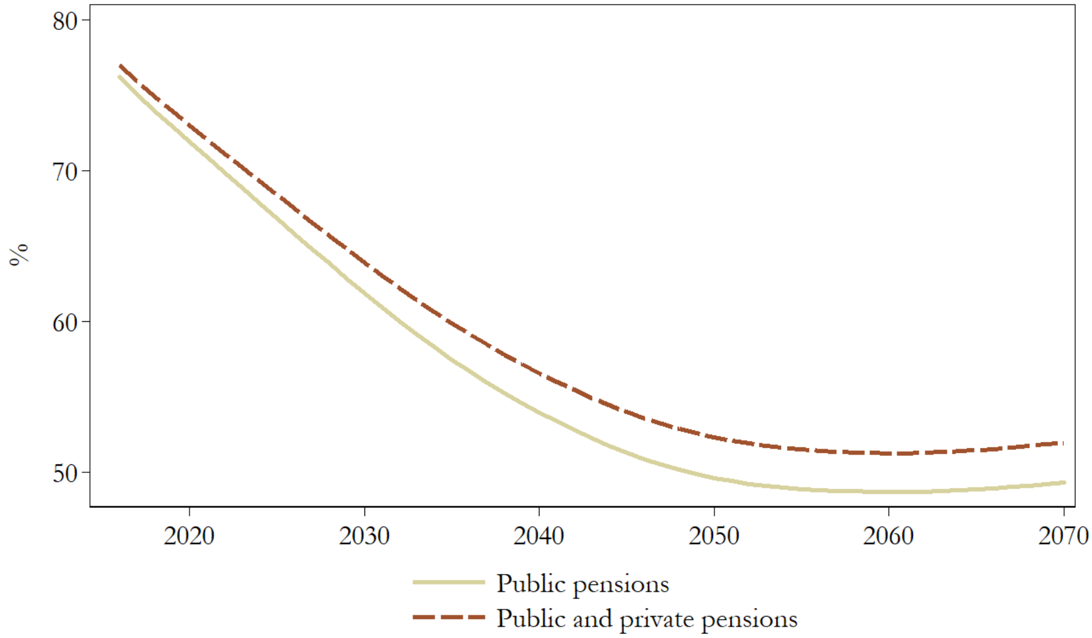
Cohort	**Overall**		**1st/2nd quintile**		**4th/5th quintile**	
	Average age	Median rr	Average age	Median rr	Average age	Median rr
1960-1964	66.1	65.4	67.2	56.7	65.7	67.5
1965-1969	66.6	59.8	67.3	53.5	66.4	61.9
1970-1974	67.1	55.2	68.0	50.3	66.7	56.3
1975-1979	67.6	51.2	68.4	48.0	67.3	52.2
1980-1984	68.0	50.8	68.6	48.0	67.8	51.6
1985-1989	68.7	50.5	69.2	47.0	68.4	51.6

Note: 'Old age 3' pensioners are not included. Source: T-DYMM 3.0, authors' elaborations.

For what concerns complementary pensions plans, as a result of our assumptions on participation rates to private pension pillars (they are kept constant to 2020 values, see Section 4.3), they have a limited effect on overall pension levels. Figure 11 presents the evolution of the replacement rates across the simulation period.

[31] NDC computation rules ensure actuarial neutrality: all contributors earn the same internal rate of return on accrued contribution, while old DB rules favour short and fast-growing careers. While the former can characterise fragile workers, the latter are generally associated with high-earning workers.

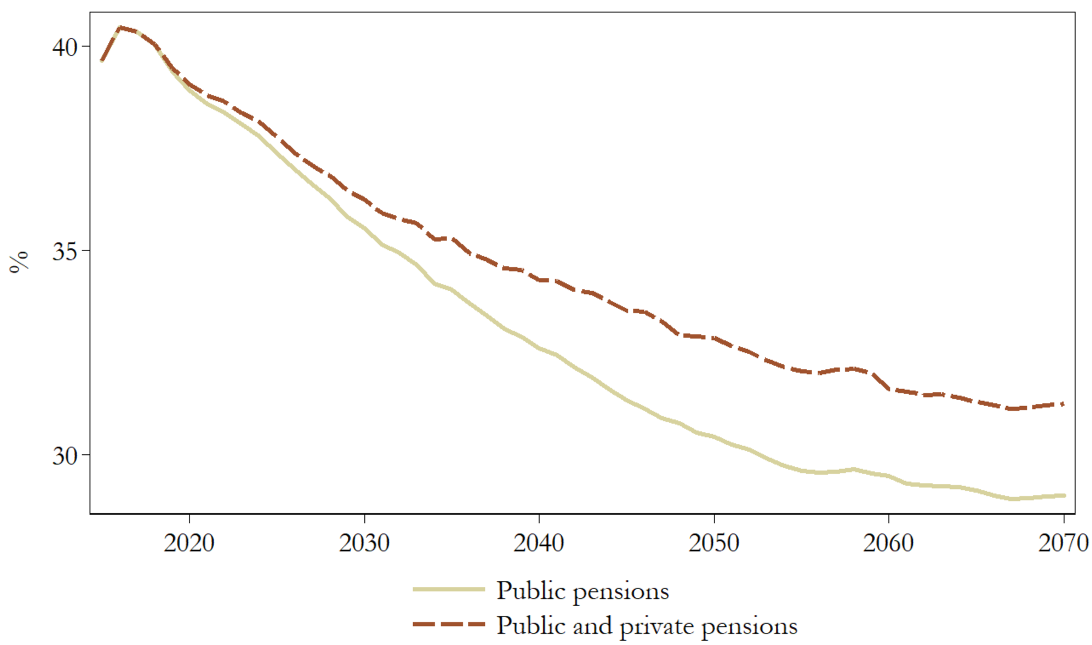
Figure 11: Replacement rate at retirement, public and private pensions



Note: 'Old age 3' pensioners are not included. Source: T-DYMM 3.0, authors' elaborations.

While the impact of private pensions on average pensions is limited, the regressive nature of the system of private pillars is evident if one looks at the evolution of inequality indicators (Figure 12), as workers who have enjoyed longer, more stable careers benefit the most from the chance to enrol in private pension plans. The equalisation (around a lower average public pension) brought about by the NDC rules is somewhat reduced by private pension schemes, as shown by the projected evolution of the Gini index.

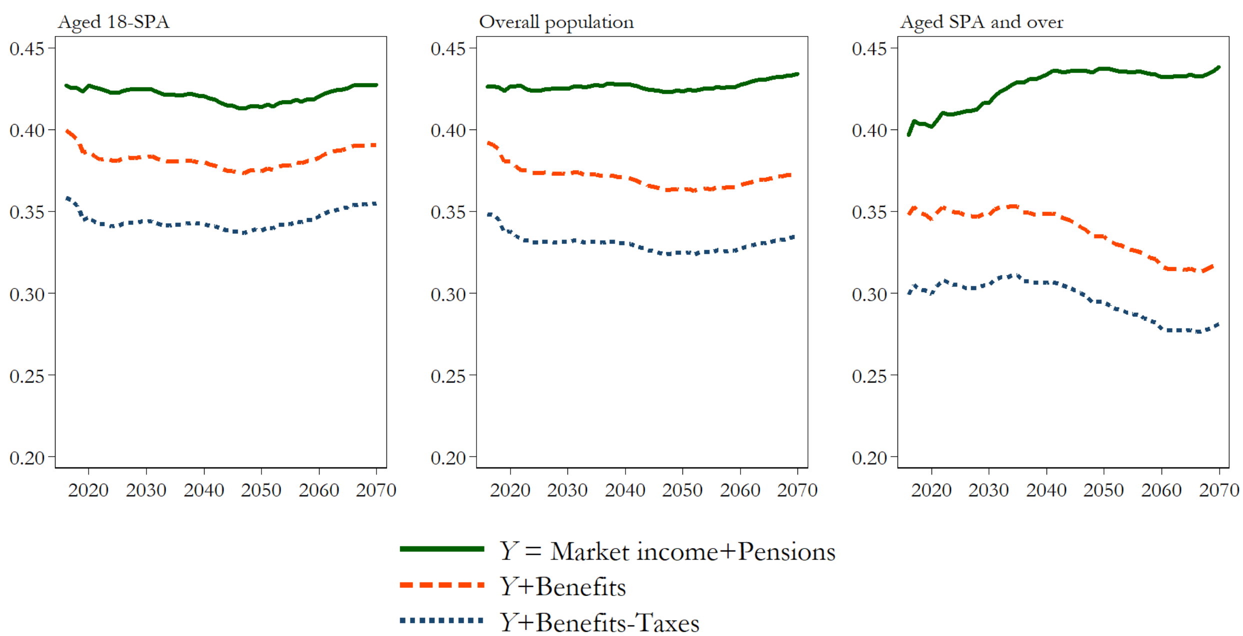
Figure 12: Gini index on stock of old-age/seniority pensioners, public and private pensions



Source: T-DYMM 3.0, authors' elaborations.

The innovations introduced in T-DYMM 3.0, specifically the development of the Tax-Benefit module and of the Wealth module make our DMM fit for inequality and poverty analyses that spread beyond the pension system alone. We shall focus on the medium- and long- term redistributive effect of total transfers and taxes separately, as well as on the projected incidence of poverty. Figure 13 displays trends in income inequality for the overall population and for specific age groups taking the individual as the unit of analysis, while income values are equivalised by using the OECD-modified equivalence scale. We discuss the results based on three income aggregates defined as follows: i) gross income before benefits (Y), which includes labour income (net of social security contributions) and productivity bonuses granted to employees, rental income, financial income, retirement income (old-age/seniority, survivor and inability pensions, including the integration to these pension benefits), and second- and third-pillar private pensions; ii) gross income after benefits (Y+B), which adds to the previous income definition the full list of in-cash benefits reported in Table 5; iii) disposable income (Y+B-T), which subtracts the personal income tax and proportional taxes listed in Table 4 from gross income after benefits. Given the profound changes in the elderly population due to the rapid increase in retirement ages and in employment rates for older workers, in what follows we shall address this category by analysing the position of those with ages equal to the Standard Pensionable Age (SPA) and over. We believe that, especially in the long run, a dynamic definition of the elderly better fits our purposes.

Figure 13: Gini index for different equivalised income definitions by age group



Source: T-DYMM 3.0, authors' elaborations.

Inequality in gross income before benefits does not vary significantly except for the elderly population, for whom inequality first increases up to 2040 and then follows a stable trend.From 2050-2060 onward, we observe a slight increase regardless of the reference population. This is due mainly to the increasing share and concentration of capital income.[32] Inequality in gross income after benefits and disposable income follows a similar trend to that observed for gross income before benefits above all for the overall and young population. Simulated transfers contribute to a greater extent to the reduction of inequality than taxes in absolute terms.[33] This is true above all for the elderly population, where the combined effect of aging and lower pension benefits yields to a redistributive effect of transfers three times higher than that of taxes by the end of the simulation. On the transfers side, what drives the most the reduction in inequality is to be attributed to the role played by the social allowance and disability allowances (both means- and non-means tested measures).[34] A

[32]Capital income, which is here the sum of financial income and rental income, accounted for 8.0% of gross income before benefits in 2016 and steadily increases to 15.8% in 2070. The concentration index of capital income with respect to gross income before benefits goes from 0.64 in 2022 to 0.68 by the end of the simulation.

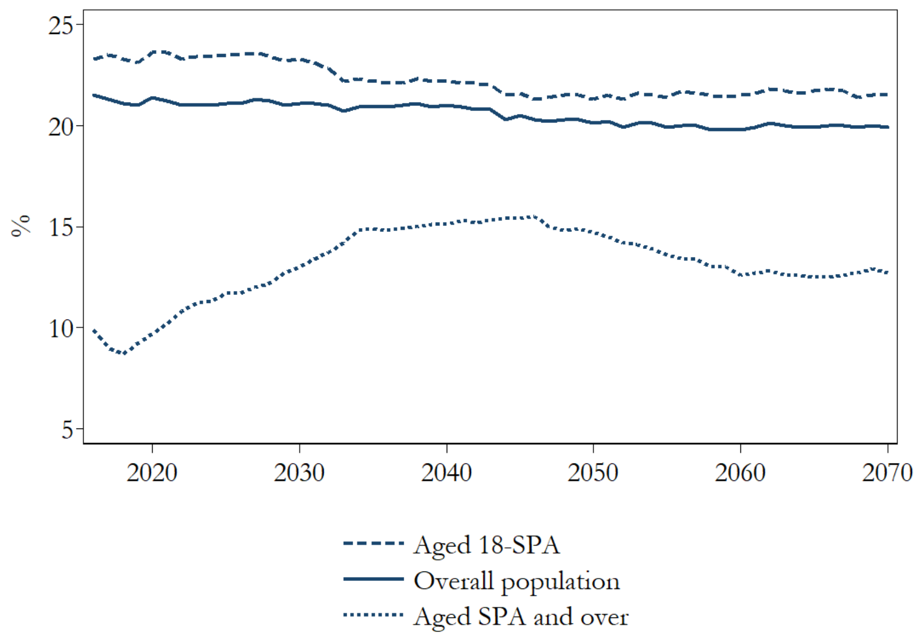
[33]The redistributive effect of transfers (taxes) equals the difference between the Gini index of gross income before benefits (gross income after benefits) and the Gini index of gross income after benefits (disposable income). For the overall population, the redistributive effect of transfers was equal to 0.034 (0.061) in 2016 (2070), while taxes reduced inequality in gross income after benefits by 0.045 (0.039) in 2016 (2070).

[34]The number of recipients amounts respectively to 6.5% (23.2%) and 14.0% (20.3%) of the elderly population in 2016 (2070). Furthermore, disability allowances are more equally distributed by the end of the simulation. The concentration index of disability allowances (social allowance) with respect to gross income before benefits varies from -0.03 (-0.32) in 2016 to -0.12 (-0.31) in 2070. Note that these statistics are computed on non-equivalised values.

large portion of the increase in the social allowance paid out is due to the fact that NDC pensioners are not entitled to the integration to old-age/seniority, survivor and inability pensions, which is included in gross income before benefits and gradually diminish in time. While aligned probabilities to receive disability allowances slightly decrease over time by age class, the quota of individuals aged over 80 (more prone to disability) within the elderly group increases, thus driving up the percentage of recipients. Still referring to the elderly population, the effect of the overall tax-benefit system results in consistently lower levels of inequality in disposable income.

Figure 14 illustrates the evolution of the incidence of poverty for the overall population and for subgroups by age. For the overall population, we observe that the incidence of poverty remains somewhat stable throughout the simulation. We also observe that working-age individuals are more likely to be poor (in relative terms) than the elderly. The gap in incidence levels between age groups narrows down to 2045, then increases but not to the initial level. The inverted U-shaped form of the poverty incidence curve for the elderly is attributable to reasons related both to the labour market and past pension reforms (Conti et al., 2021). The initial steep increase reflects discontinuity in careers and stagnation in labour income affecting Italy in the last 20 years, as well as tightened access to retirement. Pension benefits are then expected to decrease as the transition from the DB system is completed and new pensions are entirely computed according to NDC rules. As a result, the incidence of relative poverty among the elderly slightly decreases back to the level of 2030.

Figure 14: Headcount ratio of equivalised disposable income by age group



Note: the headcount ratio is an indicator of the incidence of poverty and measures the share of the reference population with equivalised disposable income lower than the poverty threshold, which is 60% of the median equivalised disposable income calculated on the overall population. Source: T-DYMM 3.0, authors' elaborations.

6 Conclusions

In this paper, we have illustrated the structure and potential of the dynamic microsimulation model T-DYMM and the distinctive features (sources, construction and content) of its underlying data set, AD-SILC. The latter provides a rich set of information on the socio-economic characteristics of individuals, their working history, their disposable income and their financial and real wealth. AD-SILC includes data from the Statistics on Income and Living Conditions (EU-SILC), delivered in Italy by ISTAT, merged with administrative data provided by INPS. The newly assembled version is characterised by an extended time horizon, now spanning from 2004 to 2017, and includes additional information from tax returns, the cadastre and from the Bank of Italy Survey on Household Income and Wealth (SHIW). The paper provides insight on the merging and matching procedures and on the adjustments operated to better the representativeness of the starting sample for the simulation. The extremely wide content and its longitudinal dimension make AD-SILC fit for analyses on the Italian labour market, tax system and on real and financial wealth. AD-SILC is also at the core of the estimation and the functioning of T-DYMM, a rather comprehensive microsimulation model that allows to undertake analyses on various aspects of the social security system (e.g. pensions, unemployment schemes, minimum income schemes), both on the adequacy and on the expenditure side. The model operates at annual frequency and it is best suited for simulations over the medium and long term. Following an initial section, which frames T-DYMM within the family of microsimulation models, the core of the paper provides a detailed overview of its structure and modules, including the highly relevant new module on household's wealth. Concerning the model structure, each module is schematically presented, starting from the demographic and going through the labour, pension, wealth and concluding with the tax-benefit module. The explanation provides focuses on the main processes and on their interactions, which stem from the sequence of the model solutions. Two appendixes provide information on the most relevant estimated equations embedded in the model. The final section of the paper is devoted to illustrating the baseline simulation, which extends to the year 2070. To a certain extent, results mirror the path of the exogenous variables to which the model is anchored (aligned), most of which reflect the baseline scenario included in the Ageing Report published by the EU Commission. Assumptions mostly concern demography (such as fertility and mortality rates), economic aggregates concerning the labour market (employment rate), growth and inflation, and financial variables (like returns on bonds and shares). Like in any other model, results are conditional on such assumptions.

We explain that the added value of T-DYMM (like for other microsimulation models) is to provide insight on the dispersion around mean values of, e.g., income from pensions, wealth components, resorting to indicators like the Gini coefficient or to other dispersion measures. We illustrate the baseline simulation results with the above qualifications in mind and against the background of an ageing and shrinking population. Concerning retirement patterns, which are traditionally the focus of the model, the main findings are the following. T-DYMM projects an increase of average retirement ages of approximately 5 years (with women soon catching up and then surpassing men) and a decrease in average time spent in retirement. The Aggregate Replacement Rate is expected to diminish until 2040 (much more so for the male population) and then to stabilise at about 50-60%, once the transition to the Notional Defined Contribution system is complete. The reduced level of the average benefit is projected to be matched by a reduction in income dispersion for pensioners, as high-income workers will be the ones most adversely affected by NDC computation rules. While private pension plans, given their limited assumed participation rates, will have a marginal effect in sustaining income replacement at retirement, model simulations show that they will bring about a regressive impact on the projected pattern of inequality indicators. Thanks to the introduction of the Wealth module, alongside the refinement of the Tax-Benefit module, our baseline assesses inequality and poverty relative to a comprehensive definition of household disposable income, which includes income from real and financial wealth, and measures the redistributive effect of transfers and taxes. Inequality for total population does not exhibit pronounced trends; the Gini index is quite steady across the simulation period, especially for active individuals. Results are more nuanced when considering the elderly population, where inequality follows quite different patterns depending on the income measure under observation. Transfers contribute to a large extent to the reduction of inequality, while dispersion is not affected significantly by taxes. We also project the evolution of the incidence of poverty for the overall population and by age-class. Working-age individuals are more likely to be poor (in relative terms) than the elderly, but this gap is somewhat narrowed throughout the simulation period. In the result section, we explain that this pattern is generated by the interaction of different forces simultaneously at play: the functioning of the labor market (and the role of discontinuous careers), the transition to the NDC pension system and the gradual tightening of access to retirement. Baseline results provide the opportunity to perform model validation (which is documented extensively in a dedicated appendix) and to sort out areas of analysis that can be explored thanks to alternative simulations. Forthcoming work with T-DYMM will consider alternative scenarios based on different macroeconomic and demographic assumptions. In addition to that, it is also possible to test the impact on inequality and public finance costs of alternative policies (and policy tools) concerning the pension and the tax-benefit system .

It is to be mentioned, finally, that the modular and flexible structure of the model lends itself to accommodate additional improvements and expansions. A few areas of intervention have already been identified. On a strictly technical level, the team is currently seeking to reduce computational times for simulations, which would allow an examination of the uncertainty in results via bootstrapping methods and alternative random seed extractions. A refinement of retirement choices (encompassing a wider range of variables in the utility function of potential retirees) is underway, as well as attempts to improve the processes of transmission and accumulation of wealth. On a larger scale, finally, in order to better simulate the impact of reforms, T-DYMM could benefit from the inclusion of behavioural responses on the side of labour supply.

References

Acciari, P., Alvaredo, F., and Morelli, S. (2020). The Concentration of Personal Wealth in Italy 1995-2016. *Centre for Studies in Economics and Finance (CSEF) Working Paper* No. 608.

Albarea, A., Bernasconi, M., Di Novi, C., Marenzi, A., Rizzi, D., and Zantomio, F. (2015). Accounting for tax evasion profiles and tax expenditures in microsimulation modelling. The BETAMOD model for personal income taxes in Italy. *International Journal of Microsimulation*, 8(3): 99–136.

Baekgaard, H. (2002). Micro-macro linkage and the alignment of transition processes: some issues, techniques and examples. *National Centre for Social and Economic Modelling (NATSEM) Technical Paper* No. 25.

Battistin, E., Brugiavini, A., Rettore, E., and Weber, G. (2009). The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach. *American Economic Review*, 99(5): 2209–2226.

Bazzoli, M., Di Caro, P., Figari, F., Fiorio, C., and Manzo, M. (2020). Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy. *Italian Department of Finance (DF) Working Papers* No. 8.

Belloni, M. and Alessie, R. (2013). Retirement Choices in Italy: What an Option Value Model Tells Us. *Oxford Bulletin of Economics and Statistics*, 75: 499–527.

Bertrand, M. (2020). Gender in the Twenty-First Century. *AEA Papers and Proceedings*, 110: 1–24.

Bonci, R., Marchese, G., and Neri, A. (2005). La ricchezza finanziaria nei conti finanziari e nell'indagine sui bilanci delle famiglie italiane. *Bank of Italy Temi di Discussione (TD)* No. 565.

Booth, A. L., Jenkins, S. P., and Serrano, C. G. (1999). New Men and New Women? A Comparison of Paid Work Propensities from a Panel Data Perspective. *Oxford Bulletin of Economics and Statistics*, 61(2): 167–197.

Boscolo, S. (2019). Quantifying the Redistributive Effect of the Erosion of the Italian Personal Income Tax Base: A Microsimulation Exercise. *Economia Pubblica*, 2: 39–80.

Boscolo, S. (2022). The contribution of tax-benefit instruments to income redistribution in Italy. *Economia Pubblica*, 2: 181–231.

Brandolini, M., Giarda, E., Moriconi, M., and Loi, M. (2009). Possibili effetti dell'under-reporting sull'analisi della ricchezza finanziaria basata sull'indagine dei bilanci delle famiglie di Banca d'Italia. *Prometeia Rapporto di Previsione* Ch. 8*,* 123-131.

Buddelmeyer, H., Herault, N., Kalb, G., and de Jong, M. (2011). Linking a Microsimulation Model to a Dynamic CGE Model: Climate Change Mitigation Policies and Income Distribution in Australia. *International Journal of Microsimulation*, 5: 40–58.

Chénard, D. (2000). Individual alignment and group processing: an application to migration processes in DYNACAN. In Milton L., Surtherland H. and Weeks M. (Eds.). *Microsimulation modelling for policy analysis: challenges and innovations*, pages 238–247. Cambridge: Cambridge University Press.

Cifaldi, G. and Neri, A. (2013). Asking income and consumption questions in the same survey: what are the risks? *Bank of Italy Temi di Discussione (TD)* No. 908.

Colombo, G. (2009). Linking CGE and Microsimulation Models: A Comparison of Different Approaches. *International Journal of Microsimulation*, 3: 72–91.

Conti, R., Boscolo, S., Desmet, R., Németh-Szűcs, R., Rézmovits, , and Van den Bosch, Karel (2021). *Poverty Risk and Income Inequality for older people in a long-term perspective*. Luxembourg: Publications Office of the European Union.

Creedy, J. and Tuckwell, I. (2004). Reweighting Household Surveys for Tax Microsimulation Modelling: An Application to the New Zeland Household Economic Survey. *Australian Journal of Labour Economics*, 7(1): 71–88.

D'Aurizio, L., Faiella, I., Iezzi, S., and Neri, A. (2006). L'under-reporting della ricchezza finanziaria nell'indagine sui bilanci delle famiglie. *Bank of Italy Temi di Discussione (TD)* No. 610.

Dekkers, G. (2014). 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: 97–109.

Dekkers, G. (2015). On the modelling of immigration and emigration using LIAM2. *Federaal Planbureau* NOTE-LIAM-11155.

Dekkers, G. and Cumpston, R. (2012). On weights in dynamic-ageing microsimulation models. *International Journal of Microsimulation*, 5(2): 59–65.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87(418): 376–382.

European Commission DG EMPL (2021). *2021 pension adequacy report: current and future income adequacy in old age in the EU. Volume 2, Country profiles.* Publications Office of the European Union, Luxembourg,.

Figari, F., Fiorio, C., Gandullia, L., and Montorsi, C. (2020). La resilienza del sistema italiano di protezione sociale all'inizio della crisi COVID-19. *Politica Economica*, XXXVI(1): 3–34.

Francesconi, M. (2002). A Joint Dynamic Model of Fertility and Work of Married Women. *Journal of Labor Economics*, 20(2): 336–380.

Fuest, C., Niehues, J., and Peichl, A. (2010). The Redistributive Effects of Tax Benefit Systems in the Enlarged EU. *Public Finance Review*, 38(4).

Goldman, M. and Kaplan, D. M. (2018). Comparing distributions by multiple testing across quantiles or CDF values. *Journal of Econometrics*, 206(1): 143–166.

Lemaître, G. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13(2): 199–207.

Leuven, E. and Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. https://ideas.repec.org/c/boc/bocode/s432001.html.

Li, J. and O'Donoghue, C. (2014). Evaluating binary alignment methods in microsimulation models. *Journal of Artificial Societies and Social Simulation*, 17(1).

Liégeois, P., Mahdi, Benjelloul, M., Boscolo, S., Calcagno, L., Dekkers, G., and ... Zhan, P. (2021). Spotlight report on dynamic microsimulation. Deliverable 6.9, Leuven, InGRID-2 project 730998 – H2020.

Martini, A. and Trivellato, U. (1997). The role of survey data in microsimulation models for social policy analysis. *Labour*, 11(1): 83–112.

MEF, INAPP, and FGB (2020). T-DYMM Forecast Model Report. MOSPI - Modernizing Social Protection Systems in Italy.

Morciano, M., Mazzaferro, C., Tedeschi, S., and Pisano, E. (2013). Modelling private wealth accumulation and spend-down in the Italian microsimulation model CAPP_DYN: A life-cycle approach. *International Journal of Microsimulation*, 6(2): 76–122.

O'Donoghue, C. (2001). Dynamic microsimulation: A methodological survey. *Brazilian Electronic Journal of Economics*, 4(2).

Pacifico, D. (2014). sreweight: A Stata command to reweight survey data to external totals. *The Stata Journal*, 14(1): 4–21.

Peichl, A. (2015). Linking Microsimulation and CGE Models. *International Journal of Microsimulation*, 9(1): 167–174.

Pisano, E. and Tedeschi, S. (2014). Micro Data Fusion of Italian Expenditures and Incomes Surveys. *Sapienza University of Rome - Department of Public Economics Working Paper* No. 164.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3): 289–310.

Stock, J. H. and Wise, D. A. (1990). Pensions, the Option Value of Work, and Retirement. *Econometrica*, 58(5): 1151–1180.

Strozza, S. and De Santis, G. (2017). *Rapporto sulla popolazione: le molte facce della presenza straniera in Italia*. Bologna: il Mulino.

Sutherland, H. and Figari, F. (2013). EUROMOD: the European Union tax-benefit microsimulation model. *International Journal of Microsimulation*, 6(1): 4–26.

Van Sonsbeek, J.-M. (2010). Micro simulations on the effects of ageing-related policy measures. *Economic Modelling*, 27(5): 968–979.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1): 39–54.

Appendix A Correction for under-reporting of financial wealth

One of the main sources of information on household wealth in Italy is the Survey on Household Income and Wealth (SHIW) held by the Bank of Italy. However, as well known in the literature (Bonci et al., 2005), survey data suffer from under-reporting of financial wealth ownership and volume. As a result, the weighted totals of financial wealth do not match with the amounts reported in National Accounts (NA) statistics. For instance, in 2015, in terms of net wealth (real and financial wealth minus liabilities), the SHIW totals account for about 53% of NA values. In particular, the total amount of financial wealth in Italy was 690 billion according to the SHIW, whereas it amounted to 3,284 billion according to the NA data (excluding the insurance reserves and standard guarantees), for a ratio of one fifth between the former and the latter. For T-DYMM 3.0, we perform a procedure of correction of the initial values of financial wealth and liabilities in order to reduce the relevance of this issue. In this Appendix, we describe the adopted procedure and show some evidence for SHIW 2016.[35]

The correction procedure foresees three steps, as in Boscolo (2019):

1. Correction for ownership of financial instruments, following Brandolini et al. (2009);
2. Attribution of financial wealth amounts to households who are "new" owners;
3. Correction for the amount of financial wealth owned, following D'Aurizio et al. (2006).

In the first step, we proceed to study the probability of owning the specific financial instruments that make up financial wealth (i.e. liquidity, government bonds, corporate bonds, stocks). We run a multinomial model to determine the probability of owning a more or less sophisticated investment portfolio. Then, we use logistic models to analyse the impact of socio-economic and financial variables on the probability of owning a specific instrument.

In Table A1 we summarise the outcomes in terms of ownership. In the first panel (above) we see the number of households detaining each financial activity before and after the correction. The most relevant increase regards government and corporate bonds. In the second panel (below), we illustrate the consistent increase in the number of households owning more than one financial instrument. In particular, before the adjustment we have only 0.71% of households owning all types of financial instruments while this percentage rises to 10.74% afterwards. Overall, the number of households owning financial wealth increases from 6,183 (83% of the SHIW sample) to 6,920 (93%).

Moreover, in Table A2 we breakdown the increase in the ownership of financial activities by socio-demographic characteristics of the household head. We see that the increase in liquidity is more pronounced for women and lowly-educated household heads, while increases in more remunerative financial activities such as bonds and stocks are observed for men, elderly and highly-educated individuals.

Table A1: Financial instruments ownership before and after correction

	SHIW (2016)		**Adjusted SHIW (2016)**	
	No.	%	No.	%
Liquidity	6,156	82.96	6,920	93.26
Government bonds	493	6.64	1,407	18.96
Corporate bonds	432	5.82	1,239	16.70
Stocks	763	10.28	1,355	18.26
No. of financial instruments				
1	6,108	82.32	5,598	75.44
2	989	13.35	580	7.82
3	270	3.64	445	6.00
4	53	0.71	797	10.74
Total	7,420	100.00	7,420	100.00

Source: authors' elaborations based on SHIW data (2016 wave).

[35] Even though the procedure has been applied to all biennal SHIW waves between 2002 and 2016 in order to improve the estimates used in the simulation, we only show results for the 2016 wave since it is the one adopted to attribute financial wealth to the simulation sample.

Table A2: Increase in ownership after correction by socio-demographic characteristics

	Liquidity	**Government bonds**	**Corporate bonds**	**Stocks**
Male	9.76	14.84	13.56	9.72
Female	10.88	9.52	7.90	6.05
< 35	8.82	0.27	1.10	2.41
35-55	10.49	6.42	9.99	7.77
>55	10.32	13.88	12.04	8.50
Below secondary	13.06	8.81	5.59	3.29
Secondary	7.37	15.87	15.30	11.74
Tertiary	5.50	18.85	23.42	19.55

Source: authors' elaborations based on SHIW data (2016 wave)

In the second step, through statistical matching (using a Mahalanobis distance metric) we impute the value of financial activities for households to which ownership of one asset not previously owned was associated. Finally, we use the ratio between 'true' and declared values found in SHIW 2002 by D'Aurizio et al. (2006) to adjust the levels of financial activities. We run regressions with these ratios as dependent variables and households' demographic and socio-economic characteristics as explanatory variables for each instrument to attribute new values of each financial activity to each household that owns any of these activities.

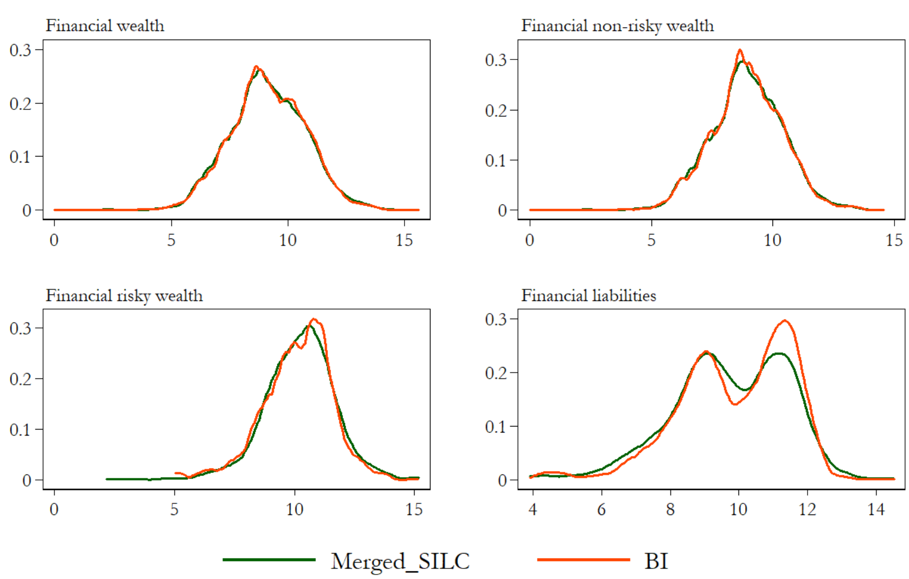
The total amount of financial wealth for adjusted SHIW data amounts to 2,547 billion. The implemented procedure significantly reduces the gap between SHIW totals and the NA, as the adjusted total of financial wealth accounts for 80% of the NA. This result is similar to what is obtained by D'Aurizio et al. (2006), whose correction procedure allows to obtain roughly 85% of the NA totals. Finally, another major difference is found in the household distribution of financial wealth. The correction procedure results in a significant change in the Gini index, which equals 0.762 before the correction and becomes 0.852 afterwards.

Appendix B Statistical matching for integrating financial wealth information on AD-SILC

The link between the (adjusted) SHIW dataset and AD-SILC is achieved through a matching technique performed at the household level. In this procedure, we conceive AD-SILC as the *recipient* sample and SHIW as the *donor* of some missing information. We adopted a propensity score matching technique originally proposed by Rosenbaum and Rubin (1983) and recently used by Pisano and Tedeschi (2014), who combined the SHIW with the Household Budget Survey (HBS) by the Italian National Statistical Institute (ISTAT). The link function is based on a set of common characteristics surveyed in both datasets and properly recodified to make them as homogeneous as possible. As for the matching unit, the common reference is the household. However, the definition of household head is fairly different between the two surveys. To this end, a recodification of the reference person for matching aims is performed, and we assign the role of household head to the highest-earning member. Moreover, in order to control for systematic differences between the two samples and to obtain a more accurate matching, we divide the joint dataset in 45 cells (or strata) obtained by the combination of household real wealth quintiles and nine household typologies (depending on the number of components, number of children and other characteristics of household components). Then, we allow the matching among units conditional on belonging to the same cell. The choice of variables for building the cells was guided by both statistical fit and economic considerations, since a higher correlation between real and financial wealth is found compared to other variables (e.g. income). The matching algorithm (*psmatch2* command in Stata by Leuven and Sianesi (2003)) uses the common information between the two surveys to match the units in the donor and recipient datasets (subject to the cell constraint explained above). The vector of control/balancing variables includes head's gender, age, citizenship, geographical area, educational level, labour income, professional status, sector of employment and qualification, marital status and partner's age, if present, as well as household income, number of income earners and real estate ownership. The variables imported from SHIW are the following: financial assets values (liquidity, government bonds, corporate bonds, stocks), liabilities, number of children outside the family, installments paid and level of savings for 2015

Figure B1 shows the comparison between kernel densities of the (unconditional) distribution from the original SHIW data and the matched AD-SILC file. The four sub-figures refer to financial wealth (overall and broken up by risky and non-risky components) and liabilities. At the lowest level of validity of a data fusion procedure (i.e. preserving marginal distribution), the distributions are supposed to be almost identical to have a good matching. Indeed, we perform a statistical test, inspired by Goldman and Kaplan (2018), that assesses the differences over the whole distribution between the original and synthetic dataset. Table B1 shows that the null hypothesis of equality between the two distributions cannot be rejected at the standard levels for three out of four wealth dimensions considered. Only for liabilities, which even visually do not have a perfect distributive overlap, the null is rejected at 1% level.

Figure B1: Distribution of matched AD-SILC Vs. SHIW financial wealth and liabilities



Source: authors' elaborations based on SHIW and AD-SILC data (2016).

Table B1: Test of equality between distributions, matched AD-SILC and SHIW

	Financial wealth	**Non-risky**	**Risky**	**Liabilities**
p-value	0.430	0.726	0.212	0.017

Source: authors' elaborations based on SHIW and AD-SILC data (2016).

Appendix C Correction of INPS information on self-employed income with tax returns data

As previously mentioned, the AD-SILC dataset is used both to estimate transitions in the labour market and to derive employment and self-employment income levels. INPS archives gather information related to individual working histories, but reported income differs in most cases from declared labour income in tax returns data. Reported income for employees and atypical workers is gross of social insurance contributions (hereinafter SICs) paid by the worker, while income collected in tax returns data is net of any contributions. As for self-employed individuals, we observe that reported earnings in the INPS archives are in line with declared earnings except for specific categories of the self-employed – i.e. individuals with declared earnings below minimum statutory thresholds set for the payment of SICs.

In the Italian tax system, low-earning self-employed workers are required to pay a fixed amount of SICs regardless of the amount of earnings declared for tax purposes (e.g. the contributory rate for craftsmen and traders aged over 21 in 2021 amounts to 24%, which multiplies a threshold $\gamma = 15,953(n/12)$, with n equal to the number of months worked). This means that, for a series of observations in the AD-SILC dataset, the reported income for contributory purposes is systematically greater than the declared income for tax purposes.

Bearing in mind what has been said above, it is easy to understand that the use of earnings from INPS archives for the estimate of self-employed worker's labour income leads to an underestimation of the true poverty conditions among these workers and it is likely to inflate poverty thresholds calculated for the overall population. To tackle this issue, we have imputed the ratio between earnings as declared for tax purposes and earnings as collected in INPS archives (hereinafter indicated with θ) for self-employed workers with reported earnings for contributory purposes exactly equal to observed statutory thresholds in the AD-SILC dataset. We are unable to estimate to what extent declared earnings are affected by tax evasion due to data availability constraints[36], but we want INPS earnings to be a good approximation of administrative earning records for tax purposes. As a result, T-DYMM does not account for evaded income amounts neither includes tax compliance responses to policy changes.

The imputation is performed by using propensity score matching analysis with the Mahalanobis distance metric. Given that θ is known for a set of observations in the AD-SILC dataset, i.e. those observations for whom we observe labour income as declared for tax purposes, that is the 2009, 2011, 2013 and 2015 tax years, we split the AD-SILC dataset in two groups: the donor sub-sample, which gathers observations eligible to act as donors in the imputation of θ; and the treated sub-sample, which is made of observations for whom we do not have tax returns data and therefore θ is missing.[37] We divide the sub-samples in 32 cells and allow the matching only between units of the same cell to preserve the replication of essential characteristics in the imputation process. The division in cells leads to three macro categories: i) 22 cells for craftsmen and traders by SICs-related threshold (from 1 to 11 months worked) and sex, ii) 8 cells for craftsmen and traders by SICs-related threshold (12 months worked), sex and macro area (North West, North East, Middle, South and Islands), and iii) 2 cells for farmers (regardless of the days of contribution accrued) by sex.

Figure C1 shows some preliminary evidence regarding θ for the four years available. We find that roughly 50% of the vast majority of self-employed workers have INPS earnings equal to declared earnings for tax purposes (see A). Furthermore, we notice a considerable incidence of cases with θ equal to or close to zero. Focusing on craftsmen, traders and farmers with INPS earnings equal to minimum statutory thresholds, we see a marked reduction in θ as expected (see B). Finally, we find high consistency between INPS earnings and tax returns data for freelancers (see C), and that is why we have excluded freelancers from the analysis.

Subsequent to the imputation, we obtain self-employment income by multiplying reported income as collected in INPS archives by θ only if $\theta \leq 1$, otherwise we do not adjust INPS earnings. Let us recall that the aim of the analysis is to correct for the underestimation of poverty levels among self-employed workers, and that the specification of the imputation algorithm strictly follows this scope.[38] This can be better appreciated by examining Figure C, where we present kernel density estimates and quantile-quantile plots of θ for the donor and treated sub-groups by category of self-employed workers (see the three macro categories of cells above). The imputed distributions fit with high precision the donor ones, especially in the interval we are interested in, where θ varies from 0 to 1. This is partly due to the econometric specification employed in the imputation procedure and partly to the higher frequency of observations gathered in the left-hand side of the donor distributions ($\theta \leq 1$), which contributes to making the imputation of cases with θ above 1 rather difficult to

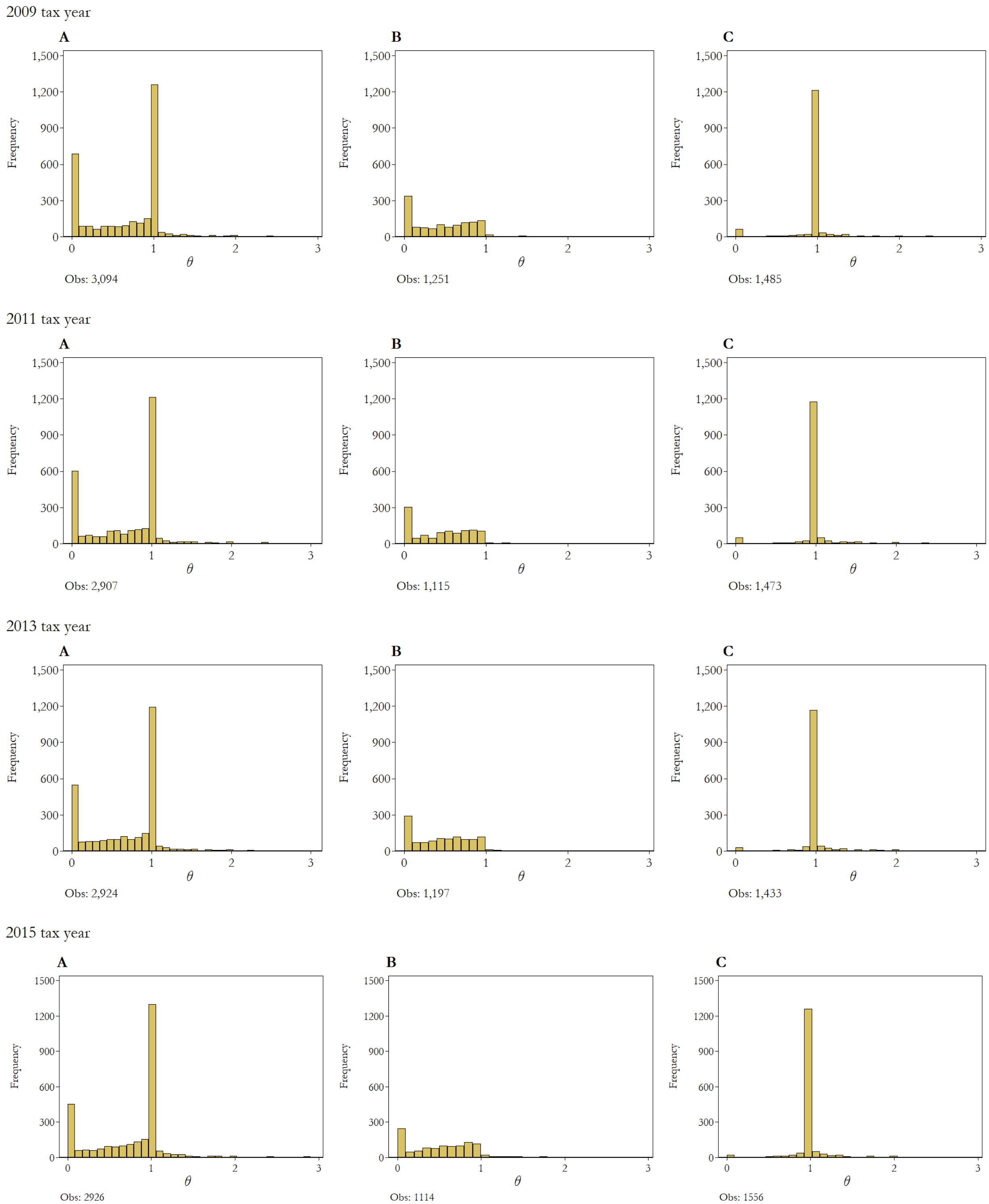
[36]The most recent estimate for the Italian context indicates that self-employed workers do not declare roughly 40% of their income if the analysis is restricted to the self-employed living in households with at least 50% of household income from self-employment (Bazzoli et al., 2020). On a more aggregate level, the average tax evasion rate for the whole population of the self-employed amounts to 24% (Albarea et al., 2015).

[37]As a result, the donor dataset amounts to 5,821 observations, in which 71.2% of individuals are repeated once throughout the 2009-2015 period; 28.7% repeated are twice; 0.1% repeated are three times. As for the treated dataset, we have 23,381 observations, in which 46.7% of individuals are repeated once throughout the 2004-2017 period; 29.5% are repeated twice; 16.1% are repeated three times; 7.7% are repeated 4 times.

[38]As for the covariates used for matching, we have: age and age squared; foreigner/national; macro area of residence; highest educational attainment; marital status; number of consecutive years as craftsman, trader or farmer and its square; to have been a craftsman, trader or farmer in $t-1$; number of days of contribution accrued in $t-1$; to be in the first decile of the IT-SILC equivalised disposable income distribution.

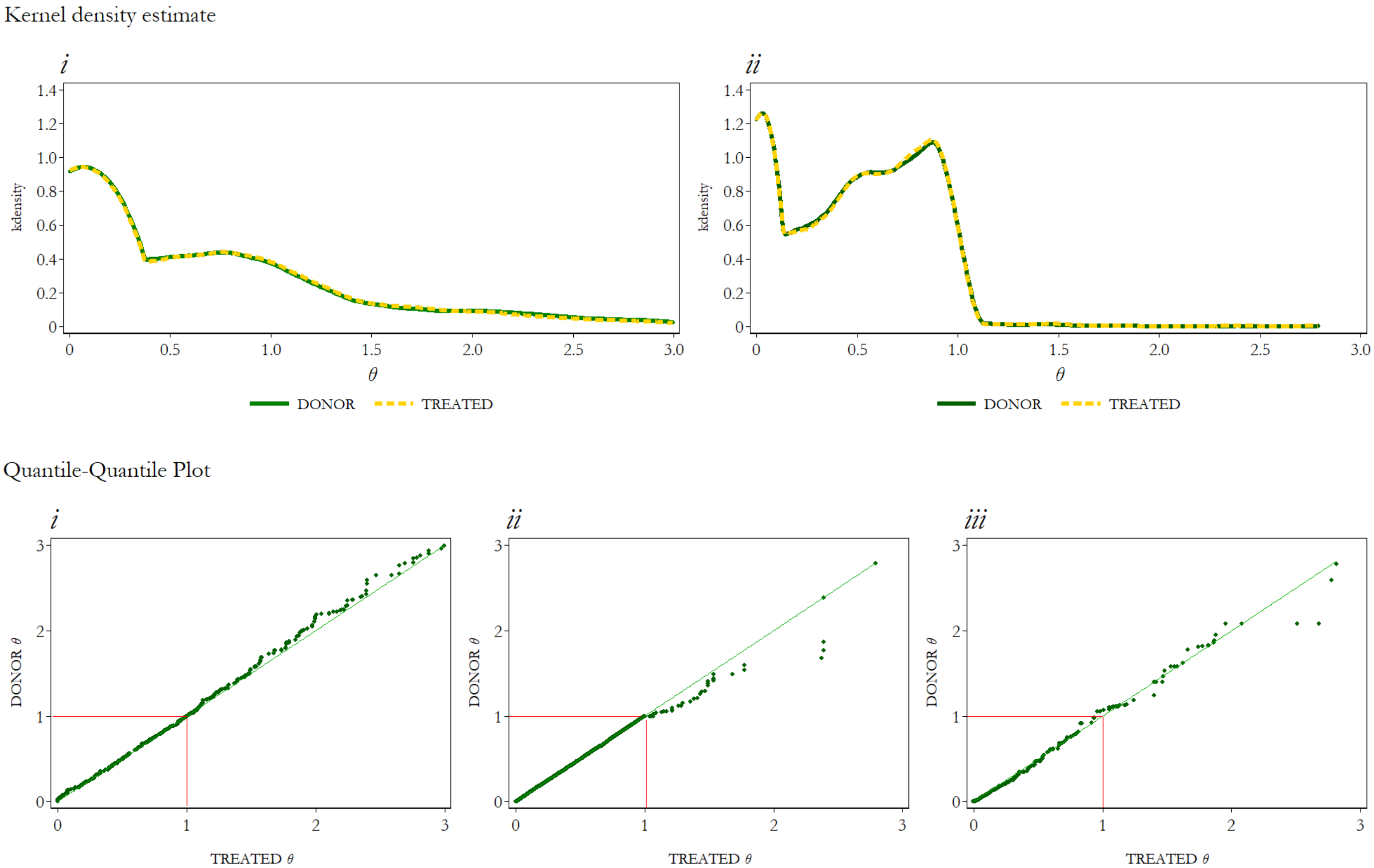
achieve in practice.

Figure C1: Ratio between earnings as declared for tax purposes and earnings as collected in INPS archives by macro category of self-employed workers



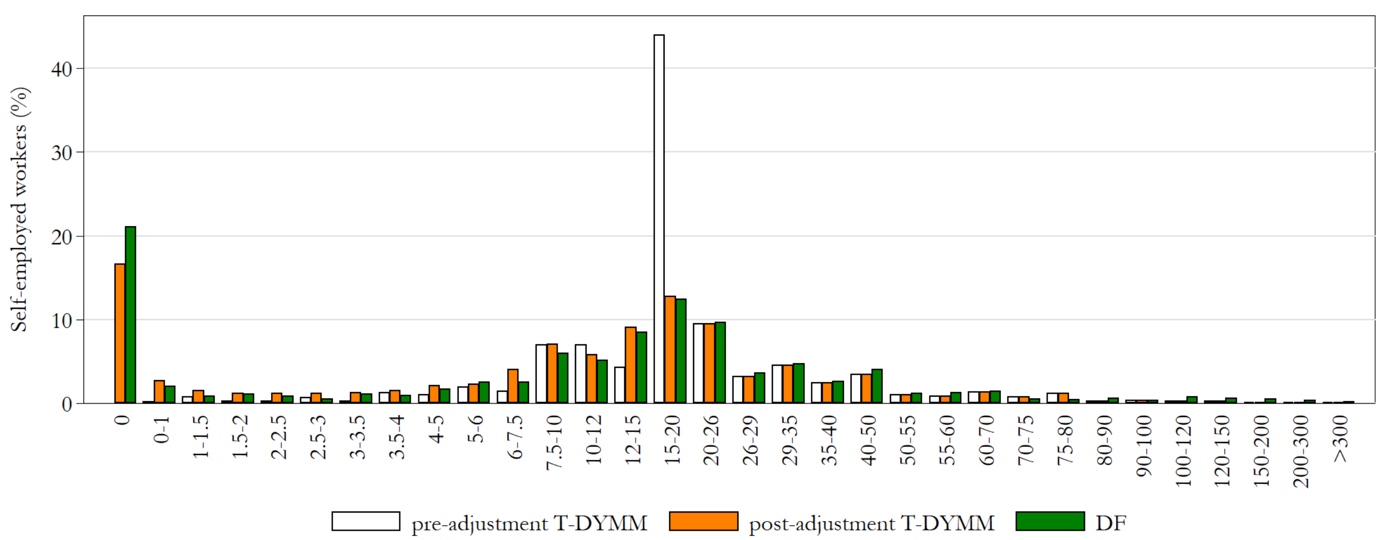
Note: **A**: self-employed workers (i.e. craftsmen, traders, farmers or freelancers) with INPS earnings lower than maximum statutory thresholds (in most cases ranging in the interval 70,000-77,000 euros according to the year considered); **B**: craftsmen, traders and farmers with INPS earnings equal to minimum statutory thresholds; **C**: freelancers with INPS earnings lower than the maximum statutory thresholds. Source: authors' elaborations on AD-SILC 2009, 2011, 2013 and 2015.

Figure C2: Kernel density estimates and Quantile-Quantile Plots of θ by category of self-employed workers



Note: i) craftsmen and traders working from 1 to 11 months; ii) craftsmen and traders working the whole year; iii) farmers. We do not show the kernel density estimate for farmers because of the high incidence of cases with θ close to zero, which affects the rendition of the graph. Source: authors' elaborations on AD-SILC 2004-2017.

Figure C3: Distribution of self-employed workers by self-employment income group (in thousands of euros)



Note: percentages are not weighted and income values refer to price levels in 2015. DF stands for tax return micro data for the 2015 tax period, obtained from the Department of Finance. Pre- and post-adjustment T-DYMM values are obtained from the AD-SILC dataset in the interval 2004-2017. Source: authors' elaborations on AD-SILC 2004-2017.

Figure C3 presents the percentage distribution of self-employed workers before and after the imputation of θ by self-employment income groups. The pre- and post-adjustment AD-SILC distributions are compared with the distribution of self-employed workers from administrative data (DF). First, it should be noted that pre-adjustment values are highly concentrated in the interval € 15,000-20,000. This reflects the high frequency of self-employed individuals working the whole year among those with reported income for contributory purposes equal to observed statutory thresholds. Second, in line with the theoretical framework underpinning the analysis, we do not observe individuals with pre-adjustment values falling into the first income group (i.e. self-employed with negative or zero earnings). Overall, the post-adjustment distribution fits the reference distribution rather accurately, keeping in mind that the imputation exercise affects self-employed workers with earnings included in the interval € 0-20,000.

Appendix D Calibration and extraction of the simulation sample

The use of population-representative samples is the starting point of almost all models for a good depiction of the simulated reality. However, the level of information needed often goes beyond the representativeness provided by sample surveys. One can be interested in studying how a hypothetical reform on a small, specific subgroup of workers would affect them in terms of their group-specific outputs, and how this would then affect overall inequality levels. To do so, the interested sample group must be a good representation of the corresponding one for the reference population. At the same time, microsimulation models can also be used to estimate aggregate and average values, which also requires the appropriate use of weighting techniques.

Given the above, T-DYMM implements a well-established technique for the reweighting of sample units in order to execute our simulations on a dataset that realistically represents the many dimensions we are interested in. The calibration is performed on the IT-SILC weights of the cross-sectional AD-SILC wave for the income year 2015, the latter being the dataset from which we obtain T-DYMM's base year sample.[39] Total frequencies of sample subgroups are calibrated to external totals made available or derived from statistics by the Italian National Institute of Statistics (ISTAT), the statistical office of the European Union (Eurostat), the Italian National Institute of Social Protection (INPS) and the Department of Finance of the Italian Ministry of Economy and Finance (hereinafter DF).

In line with the weights released in the IT-SILC survey, the implemented calibration makes sure that individuals belonging to the same household have the same calibrated weight. We refer to this calibration strategy as 'integrative calibration', which is opposed to the 'independent calibration', where individuals of a same household have different weights after the calibration. From an operating point of view, we arrange the calibration matrix as proposed in Lemaître and Dufour (1987). Supposing that our sample is made of two households ($j = 1, 2$), the first one with two members and the second one with four, whenever one or more individuals belonging to the same household present the q-th calibrated characteristic, we report the ratio between the number of individuals with the q-th characteristic and the number of individuals in the household. In such a way, we use information expressed at the individual level while assuring the equality of the weights for the members of a same household. Therefore, the calibration matrix will be defined as follows:

$$for\, i, j = \begin{bmatrix} i_1, j_1 \\ i_2, j_1 \\ i_1, j_2 \\ i_2, j_2 \\ i_3, j_2 \\ i_4, j_2 \end{bmatrix}, \; Z(q_1, q_2, q_3) = \begin{bmatrix} 1/2 & 1 & 0 \\ 1/2 & 1 & 0 \\ 0 & 1/4 & 1 \\ 0 & 1/4 & 1 \\ 0 & 1/4 & 1 \\ 0 & 1/4 & 1 \end{bmatrix}$$

where $q = (q_1, q_2, q_3)$ are the calibrated characteristics. For example, only one individual out of the four of which household j_2 is composed of presents the characteristic q_2, and therefore each member will report $\frac{1}{4}$ in the calibration matrix.

The number of calibrated characteristics (constraints) in our calibration amounts to 745. More specifically, 391 characteristics are the same ones employed by ISTAT in the calibration of the IT-SILC weights, namely (follows the number of constraints between round brackets): overall population (1); distribution of the population at the macro-regional level by sex and fourteen age groups (140); distribution of the population at the regional level by sex and five age groups (210); distribution of the foreign population at the macro-regional level by sex (10); distribution of the population at the macro-regional level by demographic size of the municipality of residence (30). To these, we add 354 characteristics deemed to be particularly relevant to the scopes of our analyses, namely: distribution of the population by sex and one-year age groups (182); distribution of the foreign population by sex, area of birth and educational attainment (8); distribution of the population by number of household members (6); distribution of individuals with retirement income at the macro-regional level by sex (11); distribution of individuals with retirement income by sex and six age groups (13); distribution of individuals with retirement income by sex and type of pension (20); distribution of retired workers by sex (3); distribution of the population by sex, eight age groups and educational attainment (33); distribution of in-work individuals by area of birth (3); distribution of in-work individuals by employment status (employee vs self-employed) (5); number of individuals with specific sources of income subject to proportional tax regimes (e.g. rental income subject to the *cedolare secca*; self-employment income subject to substitute tax regimes) (7); distribution of employees by type of contract (open-ended vs fixed-term; full-time vs part-time) (6); distribution of individuals with gross income subject to PIT by income group (32); distribution of the population by sex and civil status (8); distribution of the population with severe limitations in activities because of health problems by sex and eight age groups (17). In the calibration algorithm we often include a residual category for each of the marginal distributions above so that the sum of the individuals of a

[39]We refer to the cross-sectional IT-SILC wave for the year 2016 – where income variables refer to the year before the interview, while socio-economic characteristics are collected for the interview year – to which we match data from various administrative sources for the income year 2015 (i.e. INPS archives on working histories and pension amounts, tax returns data, cadastral data) and financial wealth amounts from the SHIW survey for the year 2016. See Section 3 for a description of the AD-SILC dataset. Prior to the calibration, the age of the sample units is brought back to the year before the interview, leading to the exclusion of individuals born in 2016 from the sample.

specific distribution equals the overall population. By doing so we avoid that the calibration of individuals with a certain characteristic involves a loss of representativeness for the individuals who do not possess that trait.

The calibration is carried out thanks to the *sreweight* Stata command (Pacifico, 2014), which replicates the algorithmic solution put forward by Creedy and Tuckwell (2004) in the context of Deville and Särndal (1992)'s raking techniques. The calibration consists in minimising the Lagrangian function that follows:

$$L = G(w,p) + \sum_{q=1}^{n} \lambda_q (t_q - \sum_{i=1}^{k} w_i x_{i,q}) \tag{1}$$

$$G(w,p) = \frac{1}{2} \sum_{i=1}^{k} \frac{(w_i - p_i)^2}{p_i}$$

where w is the calibrated weight; p is the original survey weight; $G(w,p)$ is the chi-squared distance function – note that one can make use of different distance functions, and this choice together with other constraints imposed on the variance of w with respect to p determines the presence of an explicit solution to the minimisation problem in (1); $_q = (_1,_2\ ,\ ,_n\ )$ are the Lagrange multipliers, which can be computed with the Newton's method; $x_{(i,q)}$ is the q-th characteristic object of calibration (translated into either a binary or fractional variable) of the i-th individual; t_q is the external total for the q-th characteristic. The implemented strategy does not guarantee an explicit solution to (1) as the calibrated weights are bounded in the interval $\frac{r_L<w_i}{p_i<r_U}$ for each sample unit i, where $r_L = 0.2$ and $r_U = 4$ are the lower and upper bounds respectively of the ratio between the weights. The algorithm reaches convergence when the difference between the calibrated total and the external one, for each calibrated dimension, is lower than a tolerance level tl chosen by the user (in our case, $tl = 50,000$). Figure D1 reports the ratio between calibrated weights and original weights, while descriptive statistics for both weights are given in the footer of the figure.

Subsequent to the calibration, individuals are duplicated in the sample as many times as their adjusted sample weights require. Then, once the dataset is expanded, householders are selected and sampled with repetition so as to extract one hundred samples of 100,000 householders. Finally, the remaining members of the households are combined with their corresponding householders for each extraction. Among the extracted samples, we select the best-fitting one that minimises the difference between calibrated and external totals. We refer to this latter sample as T-DYMM's base year sample, that is the starting point of all our simulations.

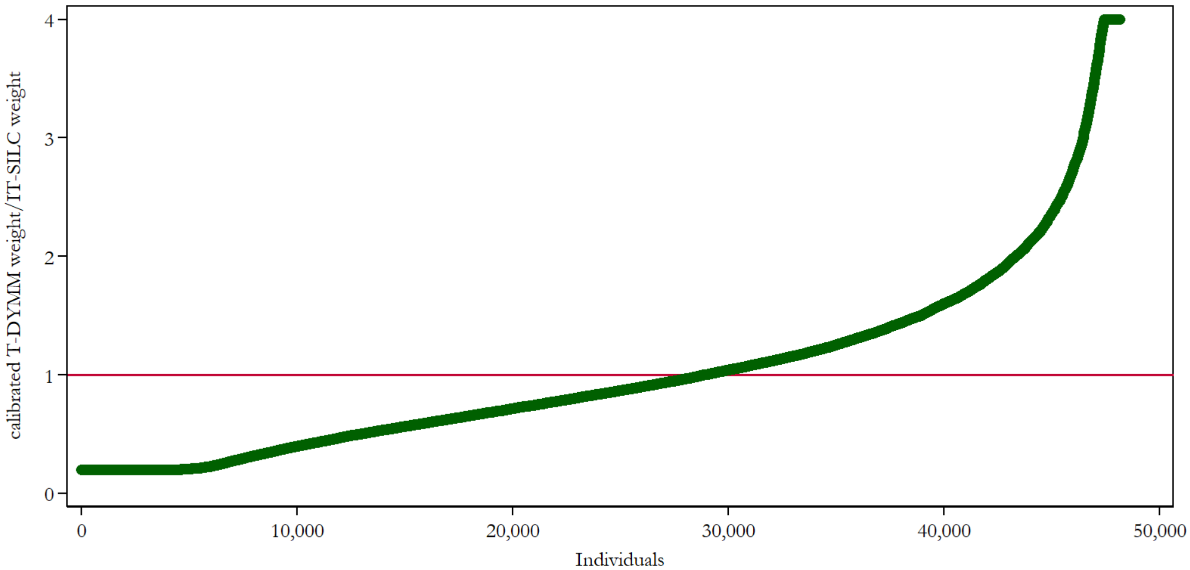
The procedure described here is a common practice in dynamic microsimulation studies and allows the overtaking of the difficulties that modellers meet when using alignment methods, although alternative strategies that do not involve the expansion of the dataset have also been proposed (Dekkers and Cumpston, 2012).

In order to evaluate the implemented procedure, in what follows we compare distributions of T-DYMM's base year sample with totals derived from IT-SILC weights[40] and with external totals. First, it should be noted that representativeness is overall preserved when looking at the same dimensions for which the IT-SILC weights have been originally calibrated (see Figure D2). Second, the calibration and sampling procedure proves to be effective in overcoming the under- and overestimation of single ages that would emerge if we were to use IT-SILC weights (see Figure D3).

Furthermore, we observe marginal improvements for each income group in which the distribution of individuals with gross income subject to PIT is divided (see Figure D4). The calibrated distribution adheres almost perfectly to external totals. Table D1 gathers the results for selected calibrated characteristics and household distributions. From the comparison we conclude that the representativeness of the sample is systematically improved with regards to the many dimensions we calibrate for. As for household-level characteristics not directly included in the calibration algorithm, we see that the implemented procedure preserves rather accurately the representativeness of two key dimensions: the distribution of households at the regional level and the distribution of households by type (e.g. single, single with children, couple, and so on).

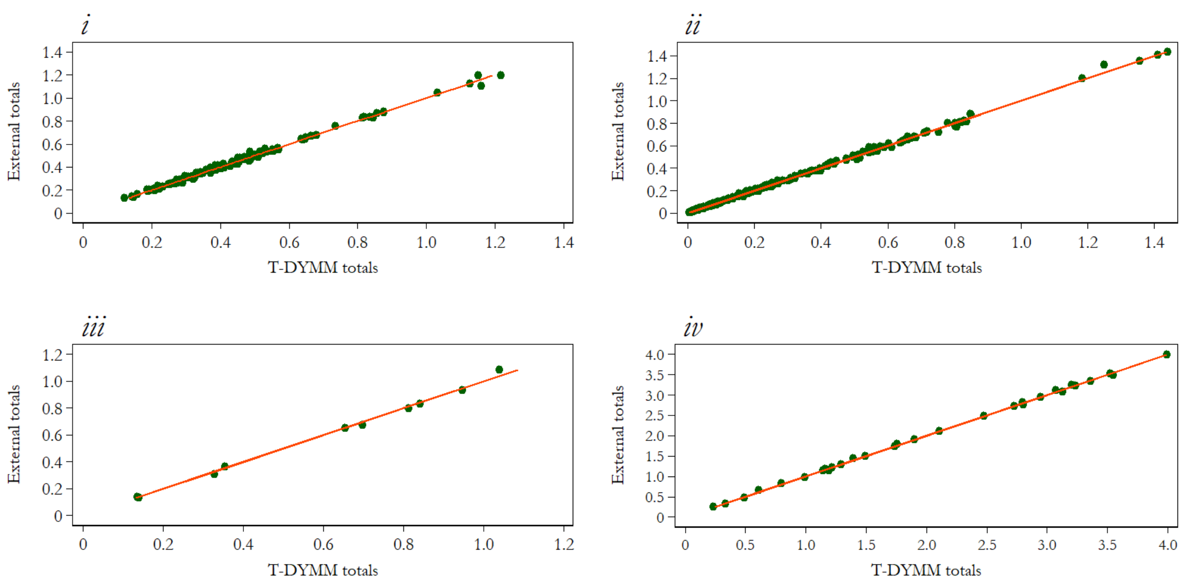
[40]The IT-SILC totals reported in the following statistics are calculated by employing the survey weights of the IT-SILC wave for the 2016 interview year on the AD-SILC wave for the 2015 income year, which is the dataset from which we obtain T-DYMM's base year sample upon completion of the calibration and sampling procedure described above.

Figure D1: Ratio between T-DYMM weights and IT-SILC weights



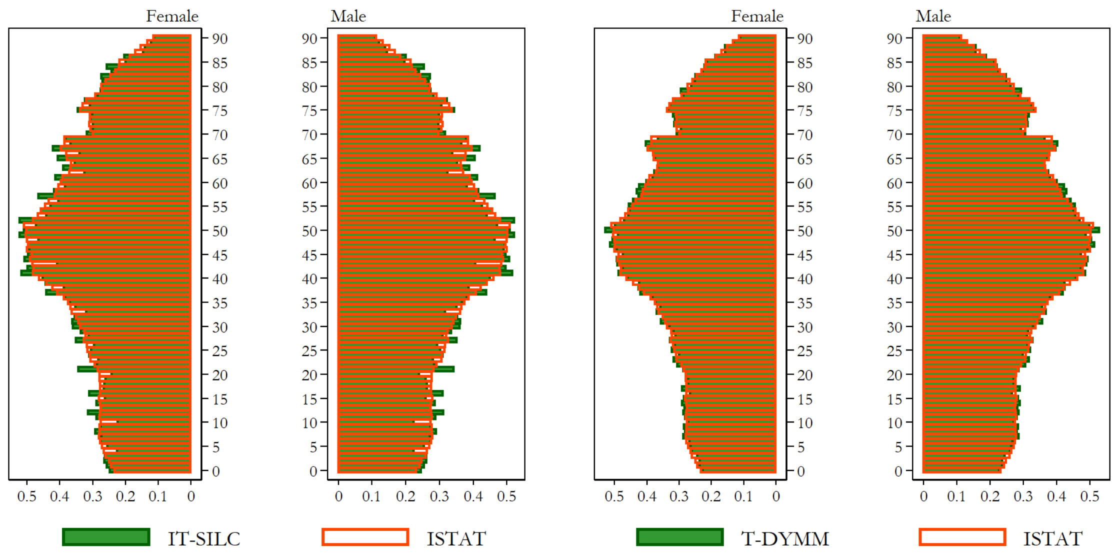
Note: individuals are ordered by increasing values of the ratio. The descriptive statistics for the two sets of weights are the following (first come statistics for calibrated weights, then for IT-SILC weights): mean (1,259.4, 1,251.8); standard deviation (1,562.9, 1,023.7); minimum value (7.9, 34.3); maximum value (23,896.2, 17,338.4); 1st percentile (55.3, 122.6); 99th percentile (7,734.3, 5,073.5). The curve in the figure shows two flattened ends because of the constraints imposed on the minimum and maximum ratios between calibrated weights and original weights. Source: authors' elaborations on AD-SILC 2015.

Figure D2: T-DYMM's base year sample representativeness to distributions employed in the calibration of IT-SILC weights (values in millions of individuals)



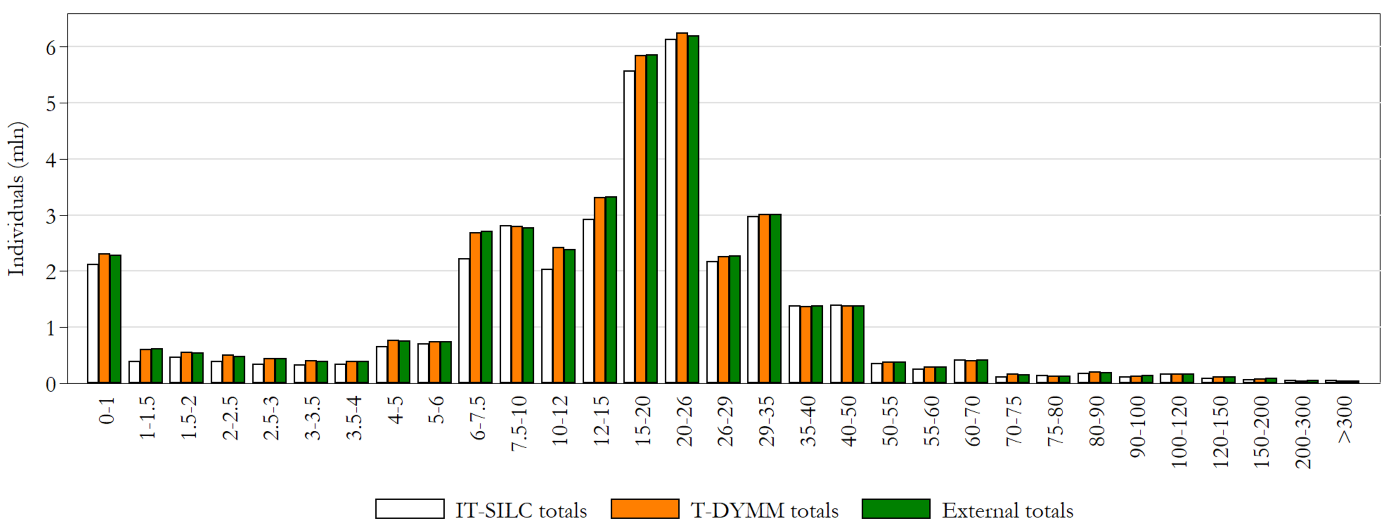
Note: i) distribution of the population at the macro-regional level by sex and fourteen age groups; ii) distribution of the population at the regional level by sex and five age groups; iii) distribution of the foreign population at the macro-regional level by sex; iv) distribution of the population at the macro-regional level by demographic size of the municipality of residence. Source: authors' elaborations on AD-SILC 2015 and ISTAT statistics.

Figure D3: Distribution of the population by sex and one-year age group: comparison between sample and external totals



Note: individuals in millions on the horizontal axes. Source: authors' elaborations on AD-SILC 2015 and ISTAT statistics.

Figure D4: Individuals with PIT gross income by gross income group (in thousands of euros)



Note: gross income is here the sum of PIT gross income and rental income subject to the *cedolare secca*. The total number of individuals with positive gross income subject to PIT amounts to 40.0 million according to official statistics, while IT-SILC and T-DYMM totals sum up to 37.2 and 40.1 million respectively. Source: authors' elaborations on AD-SILC 2015 and DF statistics.

Table D1: The comparison between sample totals and external totals for selected distributions (mostly calibrated)

Variable	IT-SILC totals	T-DYMM totals	External totals	(1)/(3)	(2)/(3)
Individuals	60,323,126	60,631,408	60,665,551	0.994	0.999
Households (in thousands)	25,821	25,429	25,386	1.017	1.002
Ratio individuals/households	2.336	2.384	2.390	0.977	0.998
Foreign population by sex, area of birth and educational attainment:					
Female – EU – Upper secondary	566,866	540,119	532,951	1.064	1.013
Female – EU – Tertiary	140,621	186,397	194,004	0.725	0.961
Female – non-EU – Upper secondary	689,547	706,173	695,296	0.992	1.016
Female – non-EU – Tertiary	311,078	307,186	302,485	1.028	1.016
Male – EU – Upper secondary	287,465	334,650	345,220	0.833	0.969
Male – EU – Tertiary	93,373	61,285	63,090	1.480	0.971
Male – non-EU – Upper secondary	609,550	572,669	573,095	1.064	0.999
Male – non-EU – Tertiary	151,188	205,723	206,235	0.733	0.998
Population by number of household members (in thousands):					
1	8,369	8,032	8,016	1.044	1.002
2	14,332	13,866	13,838	1.036	1.002
3	15,141	15,043	15,111	1.002	0.995
4	16,112	16,075	16,200	0.995	0.992
5	4,890	5,400	5,290	0.924	1.021
6 or more	1,478	2,214	2,211	0.668	1.001
Individuals with retirement income at the macro-regional level by sex:					
North West – Female	2,218,088	2,367,979	2,378,635	0.933	0.996
North East – Female	1,546,322	1,675,793	1,675,812	0.923	1
Middle – Female	1,599,179	1,667,401	1,671,085	0.957	0.998
South – Female	1,701,624	1,753,352	1,766,032	0.964	0.993
Islands – Female	834,713	835,608	832,585	1.003	1.004
North West – Male	1,993,336	2,033,838	2,059,497	0.968	0.988
North East – Male	1,473,823	1,498,042	1,479,615	0.996	1.012
Middle – Male	1,443,607	1,485,073	1,486,315	0.971	0.999
South – Male	1,608,304	1,593,656	1,616,762	0.995	0.986
Islands – Male	796,957	787,546	795,288	1.002	0.990
Individuals with retirement income by sex and six age groups:					
Female – 0-54	564,506	600,895	604,675	0.934	0.994
Female – 55-64	1,023,141	1,067,015	1,072,665	0.954	0.995
Female – 65-69	1,443,927	1,487,870	1,488,105	0.970	1
Female – 70-74	1,208,479	1,276,298	1,269,707	0.952	1.005
Female – 75-79	1,264,050	1,348,517	1,364,252	0.927	0.988
Female – 80 and over	2,395,823	2,519,538	2,524,745	0.949	0.998
Male – 0-54	741,462	694,984	711,781	1.042	0.976
Male – 55-64	1,137,954	1,130,334	1,125,680	1.011	1.004
Male – 65-69	1,606,143	1,605,099	1,607,490	0.999	0.999
Male – 70-74	1,262,353	1,271,212	1,301,312	0.970	0.977
Male – 75-79	1,175,818	1,227,220	1,218,628	0.965	1.007
Male – 80 and over	1,392,296	1,469,307	1,472,586	0.945	0.998
Individuals with retirement income by sex and type of pension:					
Female – Old-age and seniority pensions	4,709,966	5,204,113	5,196,325	0.906	1.001
Female – Inability pensions	565,080	617,424	624,406	0.905	0.989
Female – Survivors' pensions	3,605,291	4,034,618	4,032,187	0.894	1.001
Female – Disability pensions	1,416,051	1,813,111	1,824,118	0.776	0.994
Female – Social pension	518,856	536,305	541,679	0.958	0.990
Male – Old-age and seniority pensions	5,716,676	6,319,952	6,377,756	0.896	0.991

Continued on next page

Table D1 – *Continued from previous page*

Variable	**IT-SILC totals**	**T-DYMM totals**	**External totals**	**(1)/(3)**	**(2)/(3)**
Male – Inability pensions	606,074	642,345	652,098	0.929	0.985
Male – Survivors' pensions	583,017	603,947	617,234	0.945	0.978
Male – Disability pensions	992,888	1,201,027	1,219,673	0.814	0.985
Male – Social pension	266,998	297,015	304,413	0.877	0.976
Retired workers by sex (in thousands):					
Female	196	115	111	1.766	1.036
Male	536	321	331	1.619	0.970
In-work individuals by area of birth (in thousands):					
Italy	20,107	20,132	20,106	1	1.001
EU or non-EU	3,033	2,326	2,359	1.286	0.986
In-work individuals by employment status (in thousands):					
Employees	18,042	17,008	16,988	1.062	1.001
Freelancers	1,161	1,314	1,327	0.875	0.990
Craftsmen, traders and farmers	3,387	3,793	3,801	0.891	0.998
Atypical workers (co.co.co.)	550	344	349	1.576	0.986
Employees by type of contract (in thousands):					
Open-ended	15,290	14,615	14,605	1.047	1.001
Fixed-term	2,752	2,393	2,383	1.155	1.004
Full-time	14,324	13,634	13,642	1.050	0.999
Part-time	3,718	3,373	3,346	1.111	1.008
Population by sex and civil status:					
Female – Single	11,444,708	11,878,040	11,900,653	0.962	0.998
Female – Married or separated	14,719,913	14,669,927	14,683,481	1.002	0.999
Female – Divorced	824,754	862,563	871,345	0.947	0.990
Female – Widow	4,022,839	3,753,115	3,753,751	1.072	1
Male – Single	13,305,714	13,646,650	13,641,747	0.975	1
Male – Married or separated	14,586,291	14,448,378	14,485,092	1.007	1
Male – Divorced	477,936	579,280	584,343	0.818	0.991
Male – Widower	940,970	752,454	745,139	1.263	1.010
Households at the regional level (in thousands):					
Piemonte	2,008	1,963	1,950	1.030	1.007
Valle d'Aosta	61	125	62	0.984	2.016
Liguria	771	788	756	1.020	1.042
Lombardia	4,418	4,299	4,222	1.046	1.018
Bolzano	217	246	215	1.009	1.144
Trento	233	254	228	1.022	1.114
Veneto	2,059	2,022	1,983	1.038	1.020
Friuli-Venezia Giulia	559	622	544	1.028	1.143
Emilia-Romagna	1,993	1,928	1,961	1.016	0.983
Toscana	1,643	1,614	1,636	1.004	0.987
Umbria	383	420	379	1.011	1.108
Marche	644	683	641	1.005	1.066
Lazio	2,631	2,528	2,601	1.012	0.972
Abruzzo	555	532	549	1.011	0.969
Molise	131	167	131	1	1.275
Campania	2,157	2,068	2,173	0.993	0.952
Puglia	1,587	1,549	1,593	0.996	0.972
Basilicata	231	248	238	0.971	1.042
Calabria	800	740	806	0.993	0.918
Sicilia	2,022	1,930	2,022	1	0.955
Sardegna	719	699	696	1.033	1.004

Continued on next page

Table D1 – *Continued from previous page*

Variable	**IT-SILC totals**	**T-DYMM totals**	**External totals**	**(1)/(3)**	**(2)/(3)**
Households by type (in thousands):					
Single	8,369	8,175	8,312	1.007	0.984
Single with children	2,216	2,083	2,360	0.939	0.883
Couple	5,142	4,973	5,110	1.006	0.973
Couple with children	8,692	8,701	8,754	0.993	0.994
Other	1,403	1,494	850	1.651	1.758

Source: authors' elaborations on AD-SILC 2015 and external statistics.

Appendix E Labour Market module regression estimates

Table E1: Probability of being employed

	Male		Female	
	b	se	b	se
extra-EU born	0.318***	(0.042)	0.250***	(0.038)
studying	-1.022***	(0.033)	-1.029***	(0.033)
retired	-1.327***	(0.049)	-1.209***	(0.069)
age	0.383***	(0.012)	0.064***	(0.005)
age^2	-0.010***	(0.000)	-0.001***	(0.000)
age^3	0.000***	(0.000)		
upper sec. degree	0.250***	(0.020)	0.385***	(0.021)
tertiary degree	0.578***	(0.032)	0.850***	(0.031)
disabled	-0.335***	(0.050)	-0.306***	(0.058)
disability pension	-1.048***	(0.100)	-1.030***	(0.096)
disability allowance	-1.021***	(0.125)	-0.864***	(0.131)
invalidity pension	-1.018***	(0.086)	-1.285***	(0.138)
in couple	0.107***	(0.027)	-0.318***	(0.033)
partner working (lag)	0.204***	(0.027)	0.156***	(0.030)
experience	0.085***	(0.003)	0.098***	(0.004)
$experience^2$	-0.001***	(0.000)	-0.002***	(0.000)
last spell duration (out-of-work)	-0.199***	(0.007)	-0.201***	(0.007)
last spell duration (working)	0.027***	(0.001)	0.031***	(0.002)
open-ended private (lag)	3.439***	(0.034)	3.666***	(0.036)
fixed-term private (lag)	2.783***	(0.038)	3.028***	(0.038)
open-ended public (lag)	3.760***	(0.058)	4.506***	(0.063)
fixed-term public (lag)	2.969***	(0.125)	3.549***	(0.084)
professional (lag)	4.470***	(0.101)	4.133***	(0.125)
self-employed (lag)	4.082***	(0.053)	4.415***	(0.068)
atypical (lag)	3.600***	(0.071)	3.144***	(0.070)
children aged 0-6			-0.376***	(0.030)
constant	-5.725***	(0.163)	-2.242***	(0.101)
pseudo-R^2	0.719		0.739	
no. of observations	253,274		250,143	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Experience refers to years of work. Children aged 0-6 is equal to one when children in that age group are present. All other variables except for age and duration in last spell are dummies. Source: authors' elaboration on AD-SILC data.

Table E2: Probability of being employed in each employment category, male

	Fixed-term e.		Professional		Self-employed		Atypical	
	b	se	b	se	b	se	b	se
EU born	0.095	(0.082)	-0.785**	(0.362)	-0.016	(0.148)	-0.218	(0.242)
extra-EU born	0.123**	(0.050)	-0.817***	(0.227)	-0.049	(0.102)	-0.403***	(0.145)
age	-0.024***	(0.009)	0.227***	(0.028)	0.078***	(0.015)	0.097***	(0.019)
age^2	0.000***	(0.000)	-0.002***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)
upper sec. degree	-0.336***	(0.029)	1.407***	(0.148)	0.068	(0.050)	0.392***	(0.072)
tertiary degree	-0.494***	(0.043)	1.902***	(0.162)	-0.275***	(0.076)	0.745***	(0.094)
studying	0.403***	(0.054)	0.156	(0.180)	-0.356***	(0.106)	0.844***	(0.110)
in couple	-0.214***	(0.032)	-0.329***	(0.103)	0.019	(0.056)	-0.063	(0.076)
exp. as open-ended	-0.060***	(0.006)	-0.191***	(0.017)	-0.061***	(0.010)	-0.097***	(0.012)
exp. as fixed-term	0.143***	(0.016)	-0.321***	(0.089)	-0.186***	(0.032)	-0.266***	(0.058)
exp. as atypical	0.012	(0.021)	0.081*	(0.046)	-0.053*	(0.028)	0.392***	(0.024)
exp. as self-employed	0.026***	(0.009)	-0.025	(0.026)	0.174***	(0.009)	0.072***	(0.013)
exp. as professional	-0.068***	(0.020)	0.267***	(0.019)	-0.010	(0.029)	0.020	(0.024)
exp.^{2} as open-ended	0.000	(0.000)	0.003***	(0.001)	0.000	(0.000)	0.001***	(0.000)
exp.^{2} as fixed-term	0.005***	(0.002)	0.019	(0.013)	0.018***	(0.003)	0.017***	(0.006)
exp.^{2} as atypical	-0.003	(0.002)	-0.004	(0.003)	0.005**	(0.002)	-0.013***	(0.002)
exp.^{2} as self-employed	-0.001***	(0.000)	-0.002	(0.001)	-0.004***	(0.000)	-0.002***	(0.000)
exp.^{2} as professional	0.001	(0.001)	-0.006***	(0.001)	-0.001	(0.001)	-0.000	(0.001)
open-ended (lag)	-2.813***	(0.041)	-3.589***	(0.153)	-3.038***	(0.072)	-3.735***	(0.107)
fixed-term (lag)	0.989***	(0.040)	-1.620***	(0.216)	-1.401***	(0.103)	-1.219***	(0.138)
professional (lag)	-0.373**	(0.190)	4.160***	(0.139)	0.083	(0.231)	-0.006	(0.222)
self-employed (lag)	0.226**	(0.097)	0.047	(0.267)	4.861***	(0.082)	1.043***	(0.120)
atypical (lag)	-0.023	(0.107)	0.727***	(0.184)	0.874***	(0.116)	2.752***	(0.097)
constant	0.317*	(0.166)	-8.894***	(0.546)	-3.057***	(0.284)	-5.095***	(0.371)
pseudo-R^2	0.732							
no. of observations	140,331							

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category in the dependent variable is open-ended employee. The omitted category for the nationality is Italian, while for the educational level it is compulsory education. Exp. refers to years of work. All other variables except for age are dummies. Coefficients in units of log odds. Source: authors' elaboration on AD-SILC data.

Table E3: Probability of being employed in each employment category, female

	Fixed-term e.		Professional		Self-employed		Atypical	
	b	se	b	se	b	se	b	se
EU born	-0.045	(0.071)	-0.644***	(0.248)	-0.418**	(0.165)	-0.357**	(0.164)
extra-EU born	-0.147***	(0.054)	-0.747***	(0.231)	-0.339***	(0.125)	-0.475***	(0.131)
age	0.039***	(0.010)	0.249***	(0.036)	0.078***	(0.021)	0.089***	(0.019)
age^2	-0.001***	(0.000)	-0.003***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)
upper sec. degree	-0.310***	(0.032)	0.995***	(0.184)	-0.005	(0.070)	0.172**	(0.073)
tertiary degree	-0.439***	(0.041)	2.117***	(0.185)	-0.604***	(0.098)	0.634***	(0.085)
studying	0.362***	(0.056)	-0.158	(0.179)	-0.252*	(0.137)	0.720***	(0.095)
in couple	0.064*	(0.034)	-0.242**	(0.117)	0.312***	(0.074)	-0.116	(0.071)
children aged 0-6	-0.085**	(0.040)	-0.161	(0.139)	0.160*	(0.085)	-0.187**	(0.086)
exp. as open-ended	-0.044***	(0.006)	-0.122***	(0.022)	-0.028**	(0.013)	-0.048***	(0.014)
exp. as fixed-term	0.149***	(0.015)	-0.351***	(0.094)	-0.310***	(0.042)	-0.249***	(0.049)
exp. as atypical	0.012	(0.020)	0.041	(0.054)	-0.127***	(0.041)	0.353***	(0.027)
exp. as self-employed	0.043***	(0.011)	-0.001	(0.035)	0.160***	(0.013)	0.052***	(0.020)
exp. as professional	-0.020	(0.021)	0.309***	(0.025)	-0.008	(0.052)	-0.031	(0.035)
exp.^{2} as open-ended	-0.000	(0.000)	0.002*	(0.001)	-0.001***	(0.001)	-0.001	(0.001)
exp.^{2} as fixed-term	0.001	(0.001)	0.015	(0.013)	0.025***	(0.004)	0.015***	(0.005)
exp.^{2} as atypical	-0.002	(0.002)	-0.004	(0.004)	0.006*	(0.003)	-0.015***	(0.002)
exp.^{2} as self-employed	-0.002***	(0.000)	-0.000	(0.001)	-0.004***	(0.000)	-0.001*	(0.001)
exp.^{2} as professional	0.001	(0.001)	-0.008***	(0.001)	-0.002	(0.003)	-0.000	(0.001)
open-ended (lag)	-3.319***	(0.043)	-4.417***	(0.201)	-3.419***	(0.096)	-4.158***	(0.109)
fixed-term (lag)	0.990***	(0.041)	-1.562***	(0.206)	-1.179***	(0.130)	-1.056***	(0.110)
professional (lag)	-0.122	(0.192)	4.112***	(0.163)	0.214	(0.337)	0.726***	(0.219)
self-employed (lag)	0.120	(0.130)	0.213	(0.325)	5.234***	(0.113)	0.258	(0.181)
atypical (lag)	-0.233***	(0.090)	0.288*	(0.172)	0.172	(0.160)	2.215***	(0.088)
constant	-0.480***	(0.186)	-7.982***	(0.702)	-3.335***	(0.395)	-4.213***	(0.359)
pseudo-R^2	0.699							
no. of observations	110,553							

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category in the dependent variable is open-ended employee. The omitted category for the nationality is Italian, for the educational level it is compulsory education. Exp. refers to experience in the labour market. All other variables except for age are dummies. Coefficients in units of log odds. Source: authors' elaboration on AD-SILC data.

Table E4: Probability of being employed in each employment category, working pensioners

	Open-ended private		Fixed-term private		Professional		Atypical	
	b	se	b	se	b	se	b	se
partner working	-0.800***	(0.181)	-0.769***	(0.230)	-1.620***	(0.378)	-0.705***	(0.175)
exp. as open-ended private	0.086***	(0.019)	0.061***	(0.023)	-0.031	(0.031)	0.002	(0.021)
exp. as fixed-term private	2.251***	(0.521)	3.048***	(0.536)	2.009***	(0.611)	1.829***	(0.537)
exp. as atypical	-0.017	(0.077)	0.094	(0.079)	0.245***	(0.092)	0.808***	(0.051)
exp. as self-employed	-0.123***	(0.018)	-0.179***	(0.021)	-0.324***	(0.036)	-0.207***	(0.018)
exp. as professional	-0.231	(0.249)	-0.176	(0.254)	0.782***	(0.133)	0.178	(0.128)
exp.^{2} as open-ended private	-0.001**	(0.000)	-0.001**	(0.001)	0.000	(0.001)	-0.000	(0.001)
exp.^{2} as fixed-term private	-0.147***	(0.029)	-0.177***	(0.030)	-0.129***	(0.036)	-0.118***	(0.033)
exp.^{2} as atypical	-0.002	(0.004)	-0.005	(0.005)	-0.013**	(0.005)	-0.032***	(0.003)
exp.^{2} as self-employed	0.001	(0.000)	0.002***	(0.000)	0.004***	(0.001)	0.002***	(0.000)
exp.^{2} as professional	0.006	(0.005)	0.003	(0.005)	-0.014***	(0.003)	-0.002	(0.003)
constant	-0.279	(0.309)	-0.927**	(0.415)	-0.048	(0.445)	-0.182	(0.323)
pseudo-R^2	0.581							
no. of observations	9,627							

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category in the dependent variable is self-employed. Exp. refers to years of work. Partner working is a dummy. Coefficients in units of log odds. Source: authors' elaboration on AD-SILC data.

Table E5: Monthly wages of private employees

	Male		Female	
	b	se	b	se
EU born	-0.033**	(0.013)	-0.235***	(0.014)
extra-EU born	-0.102***	(0.008)	-0.308***	(0.011)
upper sec. degree	0.126***	(0.004)	0.126***	(0.005)
tertiary degree	0.334***	(0.008)	0.277***	(0.008)
children aged 0-3	0.035***	(0.004)	-0.051***	(0.005)
children aged 4 and over	0.024***	(0.004)	-0.047***	(0.005)
exp. as private employee	0.037***	(0.001)	0.025***	(0.001)
exp.^{2} as private employee	-0.001***	(0.000)	-0.000***	(0.000)
open-ended contract	0.034***	(0.005)		
part-time	-0.563***	(0.010)	-0.403***	(0.005)
partner working	0.024***	(0.003)	0.016***	(0.004)
constant	7.169***	(0.008)	7.145***	(0.008)
σ_u	0.375		0.369	
σ_e	0.152		0.144	
ρ	0.860		0.868	
R^2-within	0.142		0.192	
R^2-between	0.460		0.396	
R^2-overall	0.438		0.391	
no. of observations	88,966		63,934	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the nationality is Italian, while for the educational level it is compulsory education. Exp. refers to years of work. Children aged 0-3 or 4 and over is equal to one in presence of children in that age range. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E6: Monthly wages of public employees

	Male		Female	
	b	se	b	se
upper sec. degree	0.091***	(0.015)	0.073***	(0.012)
tertiary degree	0.431***	(0.018)	0.242***	(0.013)
exp. as public employee	0.026***	(0.002)	0.024***	(0.002)
exp.^{2} as public employee	-0.000***	(0.000)	-0.000***	(0.000)
children aged 0-3	0.048*	(0.025)	-0.067***	(0.014)
children aged 4 and over	0.060***	(0.014)	-0.033***	(0.009)
open-ended contract	0.254***	(0.029)	0.222***	(0.017)
part-time	-0.437***	(0.042)	-0.348***	(0.018)
constant	7.198***	(0.030)	7.159***	(0.022)
σ_u	0.473		0.366	
σ_e	0.341		0.266	
ρ	0.658		0.654	
R^2-within	0.034		0.019	
R^2-between	0.229		0.288	
R^2-overall	0.209		0.260	
no. of observations	16,637		24,175	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Exp. refers to years of work. Children aged 0-3 or 4 and over is equal to one in presence of children in that age range. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E7: Monthly wages of professionals

	b	se
female	-0.225***	(0.043)
exp. as professional	0.049***	(0.005)
exp.^{2} as professional	-0.001***	(0.000)
constant	7.221***	(0.045)
σ_u	0.856	
σ_e	0.457	
ρ	0.778	
R^2-within	0.014	
R^2-between	0.085	
R^2-overall	0.090	
no. of observations	8,311	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. Exp. refers to years of work. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E8: Monthly wages of self-employed workers

	b	se
female	-0.178***	(0.024)
upper sec. degree	0.092***	(0.015)
tertiary degree	0.161***	(0.027)
exp. as self-employed	0.022***	(0.003)
exp.^{2} as self-employed	-0.001***	(0.000)
constant	6.884***	(0.028)
σ_u	0.777	
σ_e	0.583	
ρ	0.640	
R^2-within	0.003	
R^2-between	0.044	
R^2-overall	0.044	
no. of observations	31,470	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Exp. refers to years of work. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E9: Monthly wages of atypical workers

	Male		Female	
	b	se	b	se
open-ended (lag)	0.396***	(0.046)	0.393***	(0.048)
fixed-term (lag)	0.316***	(0.066)	0.230***	(0.047)
experience	0.052***	(0.005)	0.022***	(0.004)
experience^2	-0.001***	(0.000)	-0.000	(0.000)
partner working	0.088***	(0.031)	-	-
children aged 0-3	0.076*	(0.041)	-0.039	(0.041)
constant	7.516***	(0.053)	7.593***	(0.055)
σ_u	0.592		0.430	
σ_e	0.406		0.462	
ρ	0.680		0.464	
R^2-within	0.053		0.034	
R^2-between	0.160		0.079	
R^2-overall	0.157		0.072	
no. of observations	3,747		3,440	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. Exp. refers to years of work. Children aged 0-3 is equal to one in presence of children in that age range. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E10: Monthly wages of working pensioners

	b	se
female	-0.130***	(0.040)
upper sec. degree	0.171***	(0.031)
tertiary degree	0.253***	(0.050)
experience	0.006***	(0.002)
fixed-term private	0.104*	(0.055)
professional	0.599***	(0.070)
self-employed	-0.141***	(0.045)
atypical	0.608***	(0.052)
constant	6.616***	(0.083)
σ_u	0.893	
σ_e	0.485	
ρ	0.772	
R^2-within	0.008	
R^2-between	0.139	
R^2-overall	0.126	
no. of observations	7,632	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Experience refers to years of work. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Table E11: Probability of working all year

	Male		Female	
	b	se	b	se
working all year (lag)	1.257***	(0.065)	1.197***	(0.063)
partner working	0.133***	(0.047)		
no. of months worked last year	0.191***	(0.008)	0.171***	(0.007)
part-time	0.851***	(0.068)	1.118***	(0.053)
experience	0.004**	(0.002)	0.009***	(0.003)
upper sec. degree	0.414***	(0.046)	0.333***	(0.054)
tertiary degree	0.694***	(0.070)	0.669***	(0.067)
fixed-term public	0.629***	(0.095)	0.619***	(0.064)
atypical	0.964***	(0.052)	0.520***	(0.060)
children aged 0-3			-0.323***	(0.074)
children aged 4 and over			-0.163***	(0.048)
constant	-3.638***	(0.108)	-3.971***	(0.119)
pseudo-R^2	0.327		0.267	
no. of observations	19,829		19,702	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Experience refers to years of work. Children aged 0-3 or 4 and over is equal to one in presence of children in that age range. All other variables are dummies except for no. of months worked. Source: authors' elaboration on AD-SILC data.

Table E12: Number of months in work if working less than 12 months

	Male		Female	
	b	se	b	se
experience	0.071***	(0.009)	0.063***	(0.010)
experience^2	-0.001***	(0.000)	-0.002***	(0.000)
studying	-1.088***	(0.093)	-0.806***	(0.084)
foreign	0.142	(0.095)		
upper sec. degree	0.573***	(0.062)	0.593***	(0.065)
tertiary degree	1.007***	(0.101)	1.299***	(0.088)
children aged 0-6	0.326***	(0.090)		
fixed-term private employee	-1.276***	(0.150)	-0.968***	(0.091)
atypical	-2.436***	(0.162)	-2.461***	(0.103)
retired	-0.542***	(0.160)	-0.520**	(0.215)
working (lag)	1.293***	(0.058)	1.400***	(0.056)
constant	4.819***	(0.188)	4.416***	(0.152)
σ_u	2.246		2.087	
σ_e	1.900		2.006	
ρ	0.583		0.520	
R^2-within	0.064		0.051	
R^2-between	0.136		0.172	
R^2-overall	0.113		0.143	
no. of observations	13,932		14,793	

Note: significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The omitted category for the educational level is compulsory education. Experience refers to years of work. Children aged 0-6 is equal to one in presence of children in that age range. All other variables are dummies. Source: authors' elaboration on AD-SILC data.

Appendix F Wealth module and Private Pensions sub-module regression estimates

Table F1: List of regressions adopted in the wealth module and in the private pensions sub-module

Process	Regression dependent variable	Data source
Financial investment decision	Ownership of government bonds	SHIW 2010-16
Financial investment decision	Ownership of corporate bonds	SHIW 2010-16
Financial investment decision	Ownership of stocks	SHIW 2010-16
Financial investment decision	Ratio of liquidity over total fin. wealth	SHIW 2010-16
Financial investment decision	Ratio of gov. bonds over total fin. wealth	SHIW 2010-16
Financial investment decision	Ratio of corp. bonds over total fin. wealth	SHIW 2010-16
Financial investment decision	Ratio of stocks over total fin. wealth	SHIW 2010-16
Inter vivos transfers	Probability of making transfers	SHIW 2014
Inter vivos transfers	Amount transferred (absolute value)	SHIW 2014
Inter vivos transfers	Probability of receiving transfers	SHIW 2014
Inter vivos transfers	Amount received (absolute value)	SHIW 2014
Inheritance	Probability of receiving inheritance	SHIW 2014
Inheritance	Amount received (absolute value)	SHIW 2014
House investment decision	Probability of buying house	SHIW 2010-16
House investment decision	Log-value of purchased house	SHIW 2010-16
Rent	Probability of paying rent	SHIW 2010-16
Rent	Ratio of rent paid over household income	SHIW 2010-16
Rent	Probability of received rent	AD-SILC 2015
Rent	Ratio of rent received over household income	AD-SILC 2015
Consumption	Log-level of household consumption	SHIW 2002-16
Private pensions	Probability of investing in II pillar	SHIW 2010-16
Private pensions	Probability of investing in III pillar	SHIW 2010-16
Private pensions	Amount of income invested in III pillar	SHIW 2010-16
Private pensions	Probability of investing zero in III pillar	SHIW 2010-16

Table F2: Probability of investing in financial activities

	$FA2_t$		$FA3_t$		$FA4_t$	
	b	se	b	se	b	se
$FA2_{t-1}$	1.662***	(0.150)				
$FA2_0$	0.886***	(0.150)				
$FA3_{t-1}$			1.296***	(0.185)		
$FA3_0$			0.899***	(0.187)		
$FA4_{t-1}$					1.350***	(0.143)
$FA4_0$					0.741***	(0.145)
age	-0.026	(0.019)	-0.099***	(0.021)	-0.038**	(0.017)
log fin. wealth	1.603***	(0.062)	2.230***	(0.088)	1.328***	(0.057)
avg age	0.056***	(0.019)	0.079***	(0.022)	-0.005	(0.017)
avg fin. wealth	-0.134**	(0.061)	-0.197**	(0.085)	0.014	(0.068)
female	0.337***	(0.117)				
tertiary degree	-0.776***	(0.134)			0.299**	(0.117)
constant	-19.730***	(0.727)	-22.684***	(0.855)	-13.969***	(0.522)
pseudo-R^2	0.654		0.774		0.639	
no. of obs.	6,019		6,019		6,019	

Note: significance levels are indicated by * < .1, ** < .05, *** < .01. FA2 corresponds to government bonds, FA3 corresponds to corporate bonds and FA4 corresponds to stocks. The subscript $t-1$ denotes the lagged variables. The subscript 0 denotes the initial conditions. The dependent variables are the probabilities of investing in one of the three forms of financial activities, named FA (households who own financial wealth always have a positive probability of detaining liquidity). To correctly estimate the dynamic relationship between ownership at time t and at time *t-1* we check for the initial conditions in the status (whether the household owned the financial activity in 2010) and we average time-varying variables, following the approach by Wooldridge (2005). Source: authors' elaborations on SHIW data (2010-2016).

Table F3: Panel estimates of log-consumption

	b	se
age	0.011***	(0.001)
2nd income decile	0.213***	(0.012)
3rd income decile	0.304***	(0.013)
4th income decile	0.375***	(0.014)
5th income decile	0.447***	(0.015)
6th income decile	0.516***	(0.016)
7h income decile	0.575***	(0.016)
8th income decile	0.653***	(0.018)
9th income decile	0.690***	(0.019)
10th income decile	0.777***	(0.022)
2nd fin. wealth quintile	0.037***	(0.008)
3rd fin. wealth quintile	0.056***	(0.008)
4th fin. wealth quintile	0.075***	(0.009)
5th fin. wealth quintile	0.118***	(0.011)
no. hh members	0.038***	(0.006)
retired	-0.067***	(0.011)
no. of income earners	0.035***	(0.007)
constant	8.176***	(0.045)
σ_u	0.382	
σ_e	0.354	
ρ	0.538	
R^2-within	0.146	
R^2-between	0.459	
R^2-overall	0.381	
no. of observations	39,559	

Note: significance levels are indicated by $^{*} < .1$, $^{**} < .05$, $^{***} < .01$. The dependent variable is the logarithm of consumption. We adopt a fixed effects estimator, where the correlation between error component and unobserved time-invariant household effect is introduced in the simulation. The main explanatory variables are the level of household income (in deciles) and of financial wealth (in quintiles); the regression coefficients illustrate a positive correlation between these two variables and consumption, as expected. The number of household members and of income earners increase the total consumption level, whereas the retired status of the head of the household reduces consumption and, as a result, increases savings. This result, typical of Italy, is known in the literature as the retirement consumption puzzle (Battistin et al., 2009). Source: authors' elaboration on SHIW data (2010-2016).

Table F4: Probability of investing in private pension plans

	II pillar		III pillar	
	b	se	b	se
age	0.247***	(0.036)	0.246***	(0.025)
age^2	-0.003***	(0.000)	-0.003***	(0.000)
log income	1.415***	(0.127)	0.811***	(0.071)
tertiary degree	0.473***	(0.087)	0.326***	(0.067)
high wealth	0.403***	(0.115)	0.414***	(0.091)
employee	-		0.425***	(0.080)
constant	-21.679***	(1.302)	-16.029***	(0.824)
pseudo-R^2	0.108		0.059	
no. of observations	14,288		25,198	

Note: significance levels are indicated by $^{*} < .1$, $^{**} < .05$, $^{***} < .01$. The regressions for private pensions are performed at the individual level. The estimate for the II pillar is made for private employees only. Among the explanatory variables are age, the level of individual income, tertiary education and the highest quintile of household net wealth. The estimate for the III pillar is made for all employed individuals. The high wealth variable is a dummy that denotes the highest quintile of wealth distribution. The employee dummy denotes the employment status. The coefficients of the two types of investment are similar, this means that the profile of the individuals who invest in collective schemes are similar to the ones who invest in individual plans. Source: authors' elaboration on SHIW data (2010-2016).

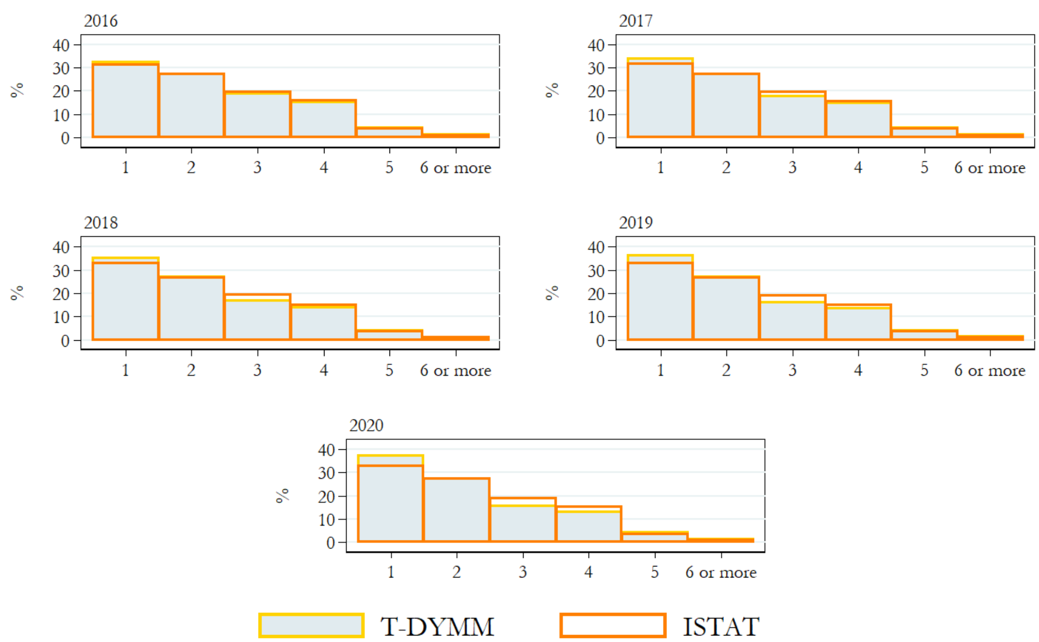
Appendix G Validation of T-DYMM's baseline results

Validating a dynamic microsimulation model involves a variety of activities that vary significantly between each other. We follow the categorisation put forward in Liégeois et al. (2021) which distinguishes between internal validation and external validation procedures. Internal validation refers to a series of processes in which modellers check whether the model operates in the way it is supposed to. These processes include the correct specification of model parameters of different nature (e.g. estimated parameters such as regression coefficients, tax-benefit policy parameters or model alignments) and model algorithms. On the other hand, with the aim of assessing the robustness of the model, external validation generally consists in the comparison of model results with administrative data (both historical and present data, in a practice that shows close similarities with the validation of static microsimulation models) or with projections from different models.

In what follows we compare non-aligned results with administrative statistics for the first simulation years. In order to perform our external validation, we have resorted to a number of institutional sources: the Italian National Institute of Statistics (ISTAT) for the Demographic and Labour Market module, the Italian National Institute of Social Security (INPS) for the Pension module, the Bank of Italy for the Wealth module, INPS and the Department of Finance of the Ministry of Economy and Finance for the Tax-Benefit module.

Figure G1 illustrates for the 2016-2020 period the evolution of households by number of family members. T-DYMM appears to closely follow the series by ISTAT, albeit the increasing trend for single-member families and the decreasing trend for three and four-member families appear stronger compared to what has been observed by ISTAT.

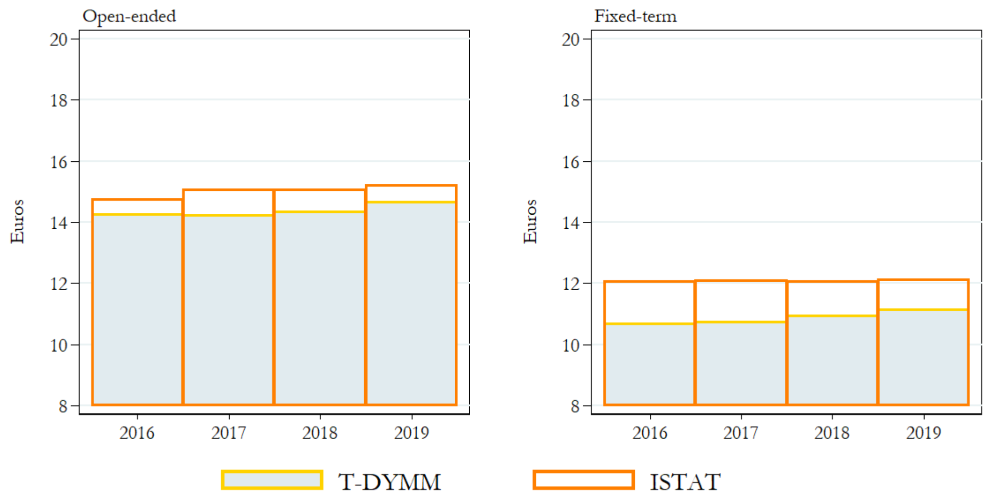
Figure G1: Quota of households by number of members



Note: In accordance with ISTAT, figures for period *t* are computed as average of *t* and *t-1*. Source: T-DYMM 3.0, authors' elaborations.

Concerning the Labour Market module, ISTAT data allow a comparison of gross hourly salaries for private sector employees by contract duration (Figure G2), educational attainment (Figure G3), area of birth (Figure G4), age group (Figure G5) and gender (Figure G6). T-DYMM performs fairly well in reproducing hourly wages, albeit some differences compared to official data are visible for less numerous categories (fixed-term, born-abroad and tertiary-educated employees). Within-group rankings are always maintained.

Figure G2: Average hourly salary by contract duration

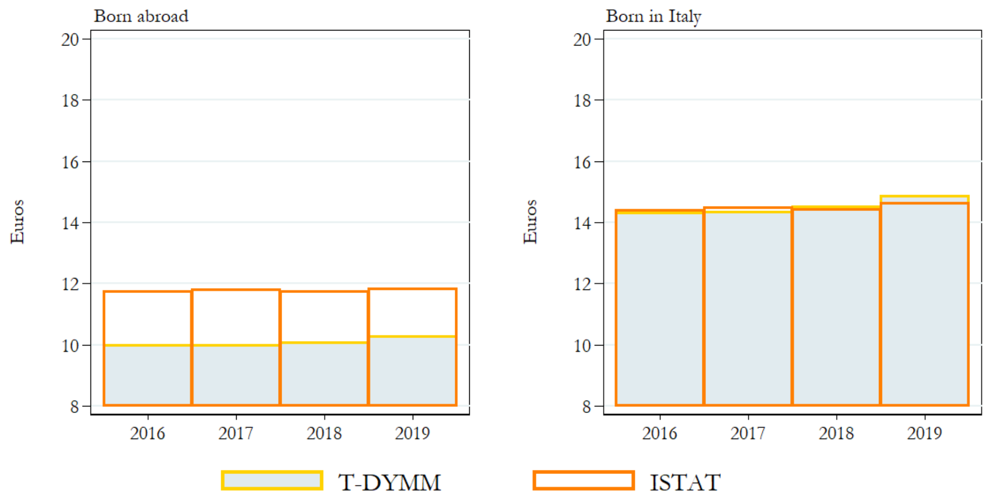


Source: T-DYMM 3.0, authors' elaborations.

Figure G3: Average hourly salary by educational attainment

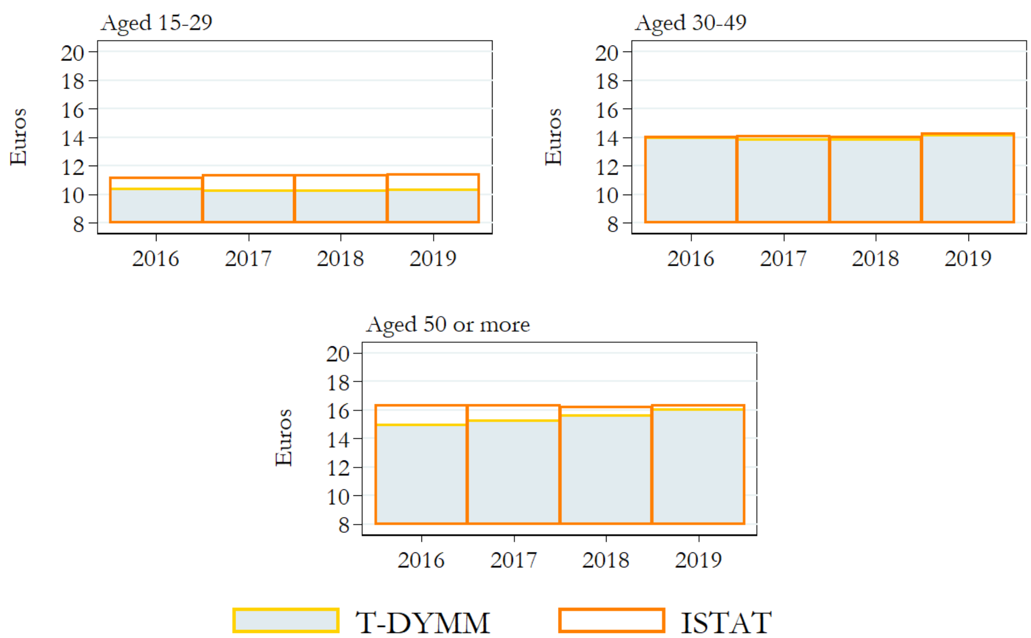
Source: T-DYMM 3.0, authors' elaborations.

Figure G4: Average hourly salary by area of birth



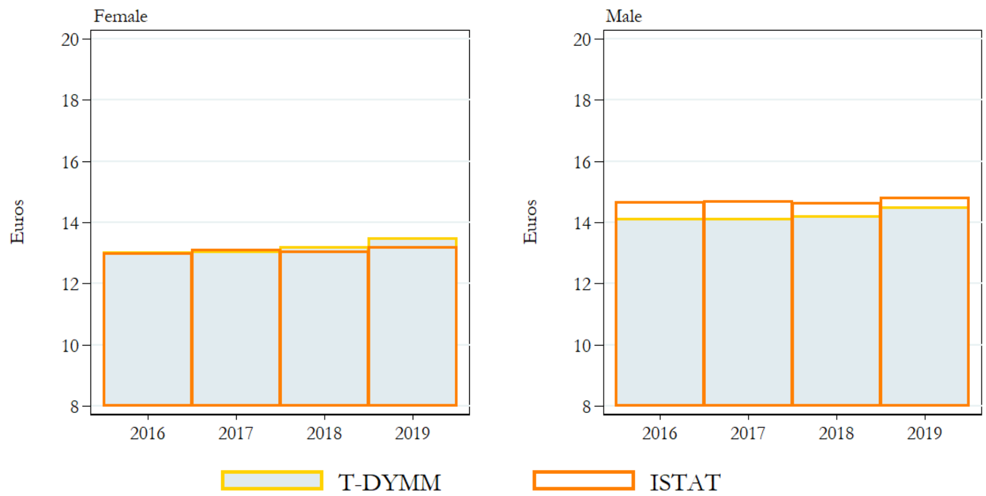
Source: T-DYMM 3.0, authors' elaborations.

Figure G5: Average hourly salary by age group



Source: T-DYMM 3.0, authors' elaborations.

Figure G6: Average hourly salary by gender



Source: T-DYMM 3.0, authors' elaborations.

Concerning the Pension module, Figure G7 compares the number of old-age/seniority, survivor and inability pensions for the 2017-2021 period[41], differentiating by gender. T-DYMM appears to correctly represent the differences across genders, as survivor pensions are much more common for women.

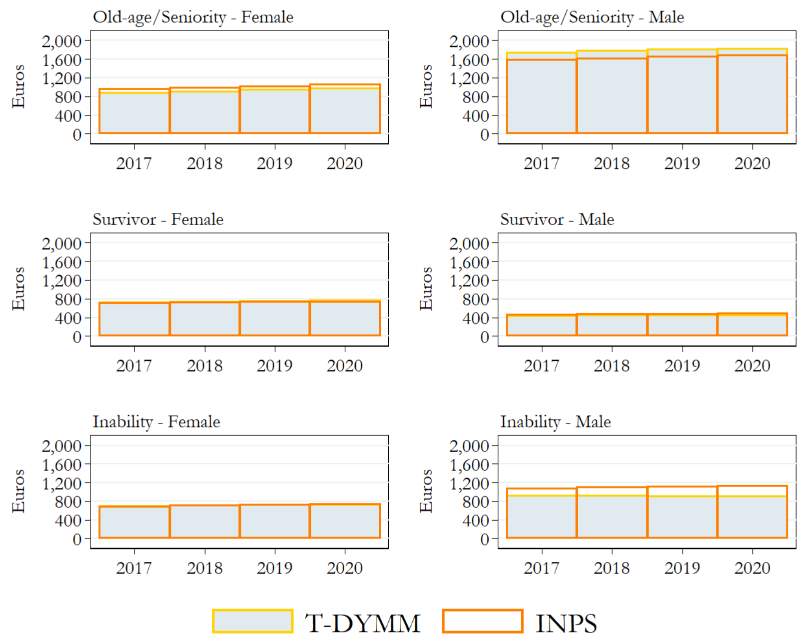
Figure G7: Number of pensions by typology and gender

Source: T-DYMM 3.0, authors' elaborations.

Figure G8 compares the average monthly amount by type of pension and gender. T-DYMM appears to slightly underestimate the average amount of inability pensions for male workers. As shown in Figure G9, average ages are also not in line with INPS data, as T-DYMM appears to have a younger pool of inability pensioners among males. Despite our work to ensure representativeness of the sample, it is to be expected that for smaller groups simulations may be less precise.

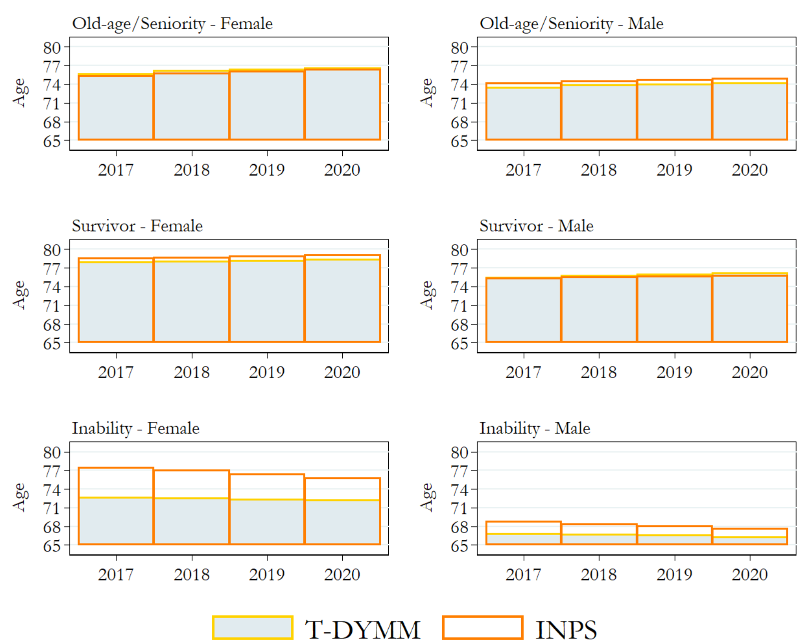
[41] Figures before 2017 are not publicly available for public sector workers from INPS.

Figure G8: Average monthly pension by typology and gender



Source: T-DYMM 3.0, authors' elaborations.

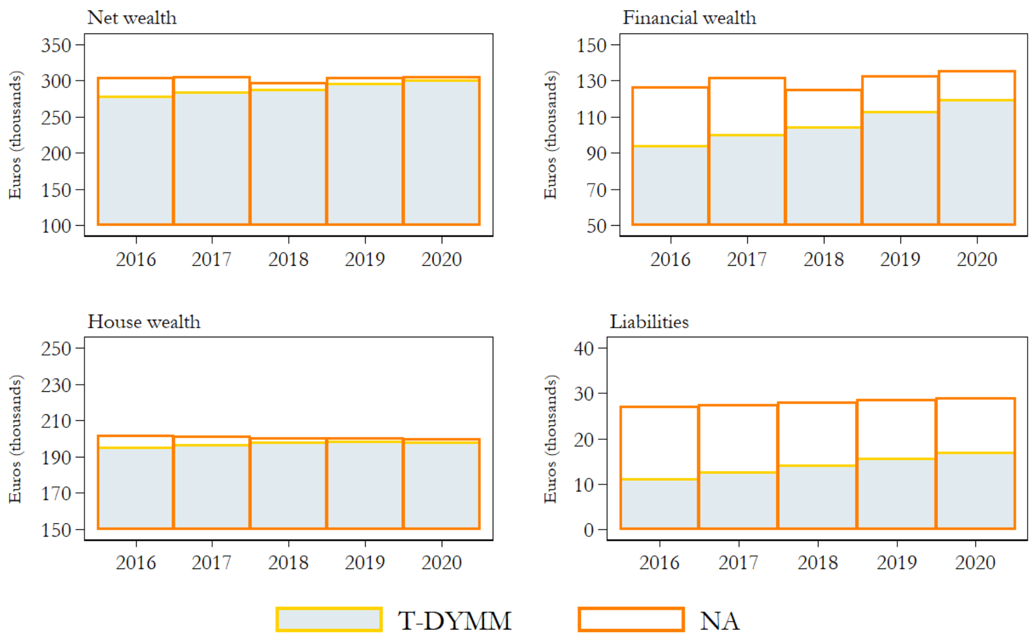
Figure G9: Average age of pensioners by typology and gender



Source: T-DYMM 3.0, authors' elaborations.

Regarding the validation of wealth outcomes, it is relevant to underline that there are not many sources to which we can compare results. We use the National Accounts (NA) totals provided by the Bank of Italy and ISTAT to compute average household levels. In Figure G10 we show the results of this comparison in the first 5 years of simulation. Net wealth averages get closer as we move forward in time. The house wealth averages from T-DYMM (whose initial values are taken from administrative sources) are almost equal to NA averages. For financial wealth, as explained in Section 3 and Appendixes A and B, we use an adjusted matched dataset derived from SHIW data that still suffers from under-reporting of financial wealth ownership and volume, visible above all for 2016. However, we see that this discrepancy reduces along the simulation period. The same reasoning holds for liabilities. This reduced discrepancy is then found in net wealth figures for the 2020 period (averages are almost identical).

Figure G10: Average household wealth

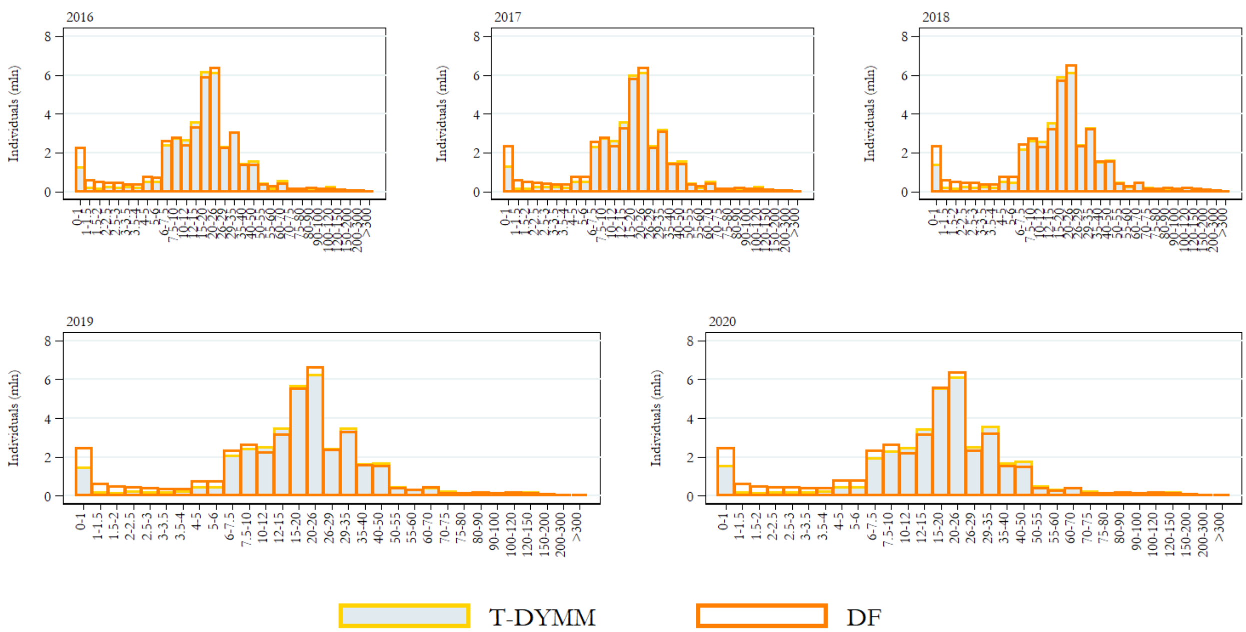


Source: T-DYMM 3.0, authors' elaborations.

The Tax-Benefit module comes at the lowest level of the model hierarchy, hence it is the one on which we can better evaluate the model's performance, given that the module's inputs are the outputs of all previous modules. Moreover, the validation of tax-benefit statistics is facilitated by the extensive availability of aggregate administrative data made publicly available by various institutions. We focus on recipients of income and most sizeable transfers (in terms of aggregate expenditure) instead of monetary amounts, because of the greater availability of external data on the former.

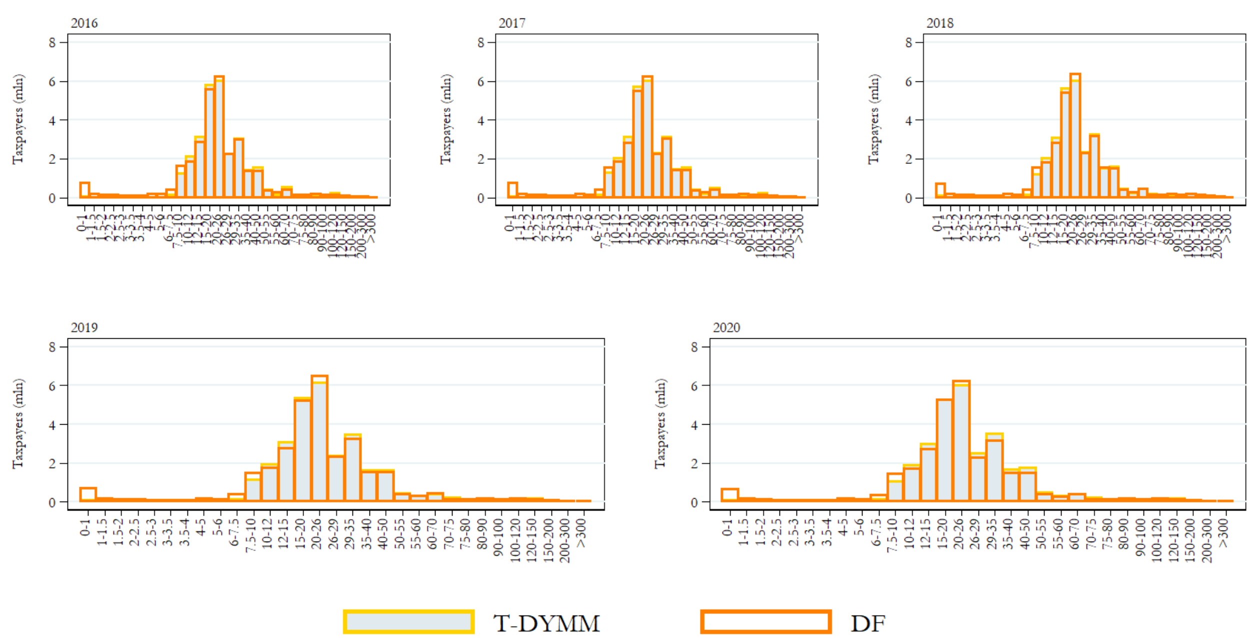
From the comparison we observe that model results adequately fit reference statistics in the 2016-2020 interval when it comes to gross income recipients (see Figure G11) and PIT taxpayers (see Figure G12) by income group.

Figure G11: Frequency density function for gross income (income groups in thousands of euros)



Note: gross income here is the sum of PIT gross income and rental income subject to the *cedolare secca*. Source: T-DYMM 3.0, authors' elaborations.

Figure G12: Distribution of PIT taxpayers by gross income group (in thousands of euros)

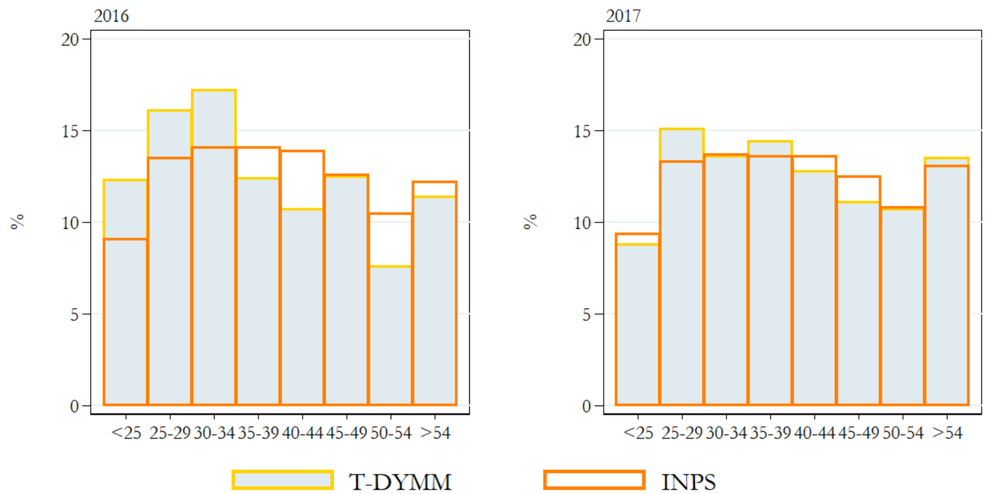


Note: gross income here is the sum of PIT gross income and rental income subject to the *cedolare secca*. Source: T-DYMM 3.0, authors' elaborations.

In both cases, the model performs rather poorly at the very end of the left tail of the gross income distribution, reflecting the difficulties we encounter in the exact prediction of labour income components – which make the greatest part of the gross income definition we are referencing to – for the most extreme segments of the distribution.

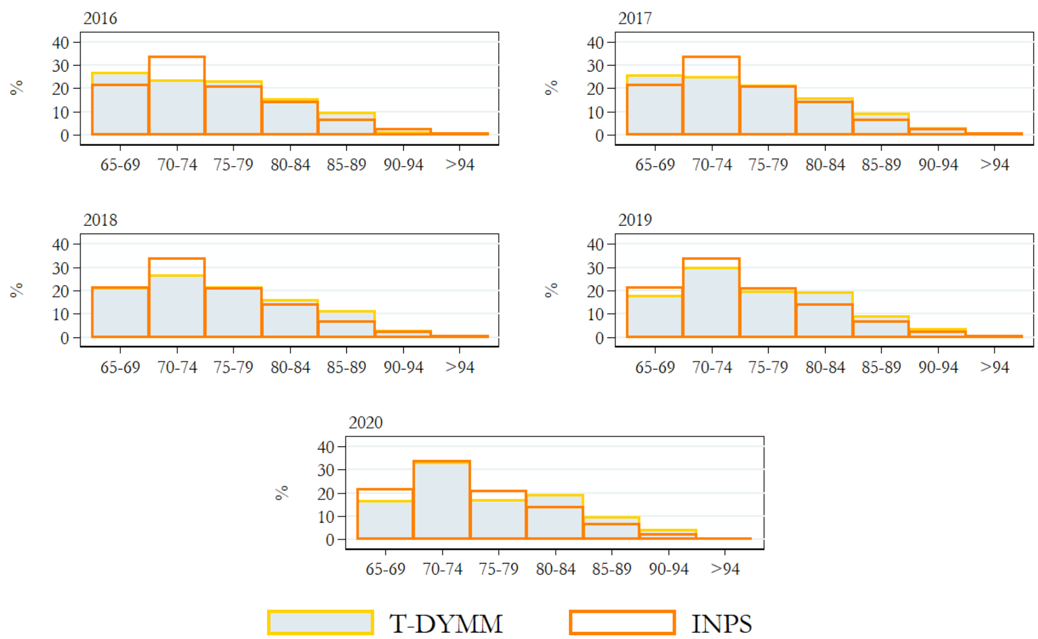
As for transfers, we register that the model reproduces the age distributions of NASpI recipients (see Figure G13) and social pension recipients (see Figure G14) quite accurately, and that goodness of fit increases as the simulation years go by.

Figure G13: Percentage distribution of NASpI recipients by age group



Source: T-DYMM 3.0, authors' elaborations.

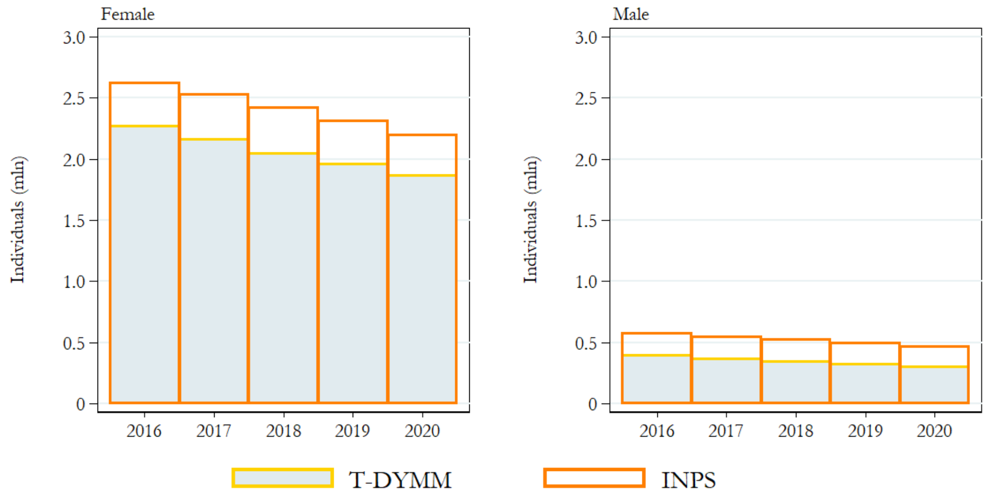
Figure G14: Percentage distribution of social pension recipients by age group



Source: T-DYMM 3.0, authors' elaborations.

Furthermore, the model is able to capture the gender differential in the group of recipients of the integration to old-age/seniority, survivor and inability pensions (*Integrazione al trattamento minimo*, see Figure G15).

Figure G15: Distribution of recipients of the integration to old-age/seniority survivor and inability pensions by sex



Source: T-DYMM 3.0, authors' elaborations.

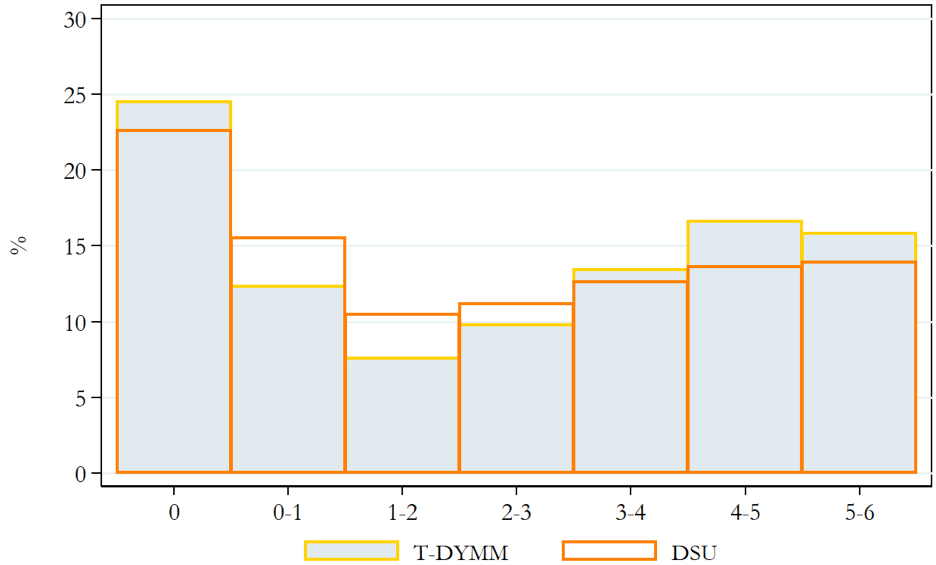
Unsurprisingly, we observe that the number of household beneficiaries of family allowances for employees' and pensioners' households is slightly higher than external statistics, as we assume that actual beneficiaries in T-DYMM equal potential beneficiaries. We find a rate that ranges from 96.6% to 97.5% in the 2016-2019 interval, keeping in mind that family allowances were likely to have a relatively high take-up rate for a number of reasons (e.g. they were paid in most cases by employers and included in the monthly pay check).

Figure G16 compares the distribution of households with an ISEE[42] value lower than 6,000 euros – which is the threshold below which households were eligible for the minimum income scheme in force in 2018 – with external statistics. We therefore focus on households who are very likely to be in severe poverty conditions and by doing so we validate the ISEE computation process for a population segment who is expected to be among the primary targets of social policies. Firstly, we must stress that the two distributions are derived on a similar basis, but they are not perfectly comparable. The DSU[43] distribution is derived from administrative records covering the whole population of households who filed the DSU declaration for claiming benefits of whatever nature in 2018, whose means-testing is based on the ISEE. The T-DYMM distribution is obtained by including all households who are beneficiaries of simulated benefits requiring ISEE (i.e. the 'mother bonus' and the 'newborn bonus', in addition to the minimum income scheme in 2018). We observe that the model distribution fits the administrative distribution with good precision, and in line with external statistics we find a substantial incidence of cases for households with ISEE equal to zero.

[42]The *Indicatore della Situazione Economica Equivalente* (ISEE) is one of the main pathway to social transfers in Italy, as it measures the overall economic level of households, considering all sources of wealth and income and a number of qualitative characteristics (e.g. presence of disabled individuals within the household). It is computed on the basis of the data indicated in the DSU (see below).

[43]The *Dichiarazione Sostitutiva Unica* (DSU) is a document containing household registry, income and estate data and is valid from the moment of presentation to December 31 of the same year.

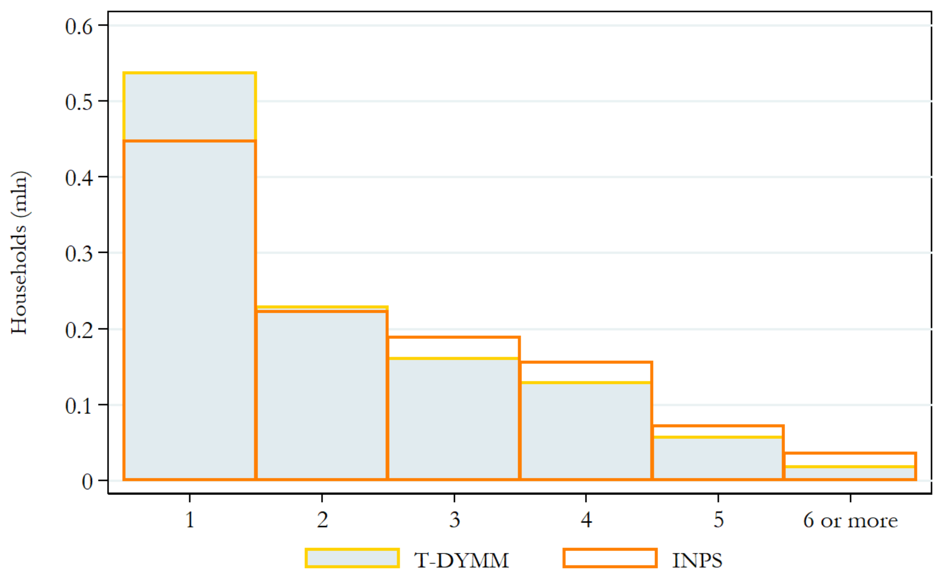
Figure G16: Percentage distribution of households with ISEE value lower than 6,000 euros in 2018 by class of ISEE value (in thousands of euros)



Source: T-DYMM 3.0, authors' elaborations.

When focusing on demographic characteristics of ISEE-related benefits' recipients (see Figure G17 on RdC beneficiaries in 2019), the model overestimates the number of single-member households and underestimates households with three or more members, while still reproducing the correct ranking.

Figure G17: Distribution of RdC recipients at the household level by number of household members in 2019



Source: T-DYMM 3.0, authors' elaborations.

Ministry of Economy and Finance

Department of the Treasury

Directorate I: Economic and Financial Analysis

Address:
Via XX Settembre, 97
00187 - Rome

Websites:
www.mef.gov.it
www.dt.mef.gov.it//it/

e-mail:
dt.segreteria.direzione1@tesoro.it

Telephone:
+39 06 47614202
+39 06 47614197

Fax:
+39 06 47821886