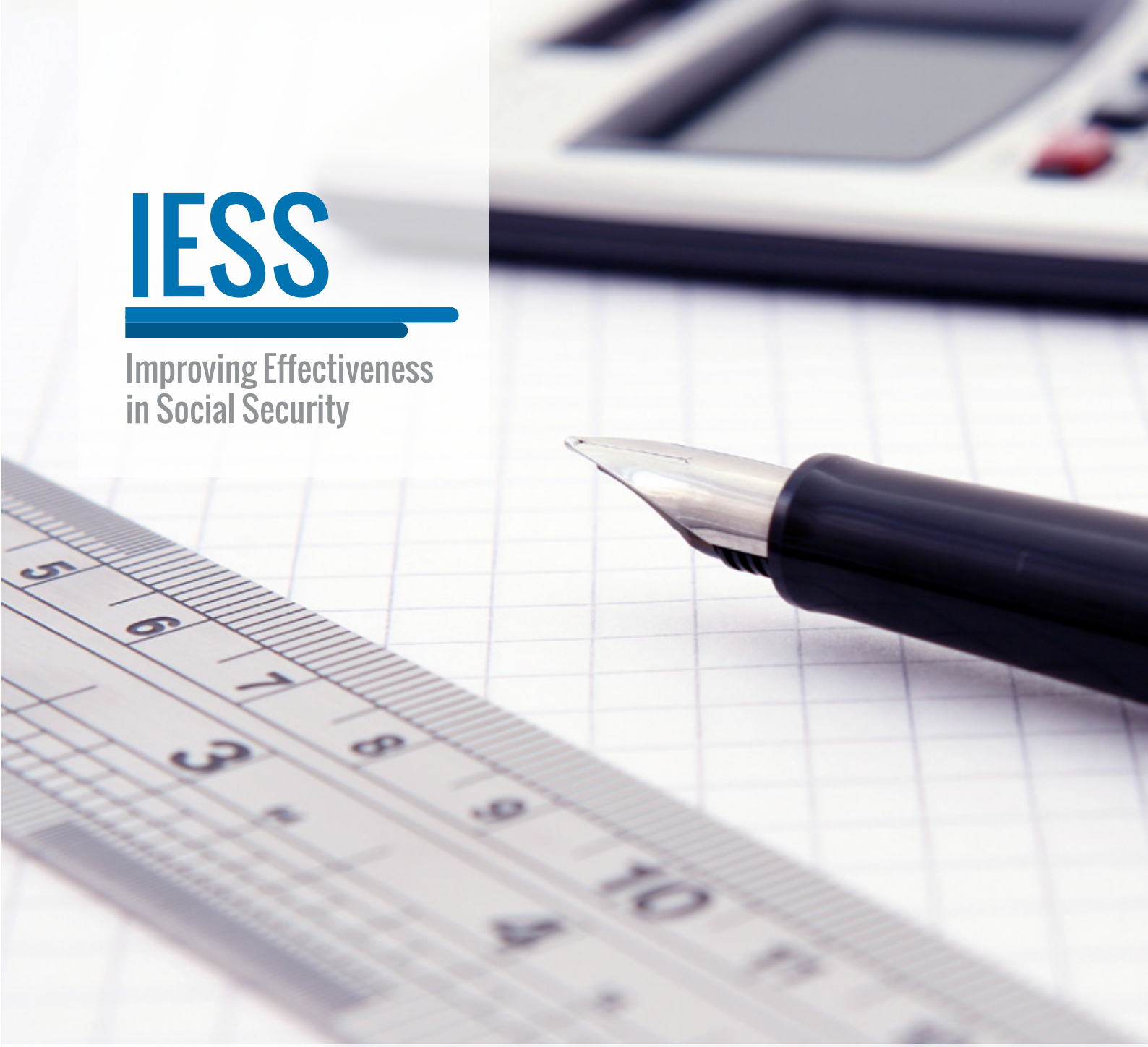


IESS

Improving Effectiveness in Social Security

The new versions of T-DYMM and AD-SILC

TECHNICAL NOTE

MEF Dipartimento del Tesoro

INPS

FGB Fondazione Giacomo Brodolini

centre for economic & social Inclusion

IESS project is funded by the European Union Programme for Employment and Social Solidarity – PROGRESS (2007-2013). The information contained in this report reflects only the author's view. The European Commission shall not be considered in any way responsible for any use that can be made of the information it contains.

TABLE OF CONTENTS

1. Introduction

The current project, "Improving effectiveness in social security" (henceforth IESS), has been launched to provide innovative analytical tools in order to improve the effectiveness of public policy evaluation in the fields of labour market analysis, labour income distribution, public and private social security programs and retirement behaviour and, consequently, in order to help policy makers in their decisional process.

In this respect, the project "Innovative Datasets and Models for Improving Welfare Policies" has already had as an aim to fill the severe deficiency coming from current policy toolkit by developing: i) a dynamic microsimulation model (henceforth, DMSM) – called T-DYMM (Tresaury Dynamic Microsimulation Model) – in order to simulate the evolution of cross-sectional sample representative of Italian population, with both individuals and households as units of analysis, and ii) a unique and innovative dataset – called "Administrative SILC", henceforth AD-SILC – by matching longitudinal information coming from several administrative archives gathered by INPS (National Institute of Social Security) with survey data collected by ISTAT (National Institute of Statistics).

T-DYMM has the national population as base. It simulates the evolution of cross-sectional sample representative of Italian population, with both individuals and households as units of analysis.

Following O'Donoghue's (2001) important taxonomy, T-DYMM presents the following features among others DMSMs:

i) It is a model with *dynamic ageing*;

ii) It is a *discrete time* model: transition in labour market and updating process are carried out year-by-year;

iii) The ageing process is *probabilistic*: simulation and transitional dynamics are achieved through probabilistic methodologies. In particular, discrete transition (in the labour market or in the others parts) are obtained by means of a Monte Carlo technique;

iv) It is a *closed* model: it simulates life-cycle evolution of the main demographic and economic population features within the sample, with new individuals that enter the population each year due to birth and others who exit due to death. At the state of the art, migration flows are not simulated.

Concerning the data source used for the microsimulations, a specific dataset has been built in order to respond to T-DYMM requirements. In particular, very detailed micro-data coming from the 2005 cross-sectional wave of the IT-SILC survey (e.g. the Italian database of the European Union Survey on Income and Living Conditions – EU-SILC, where thousands of variables are included over the ones collected in EU-SILC) have been merged with information collected in many administrative archives managed by INPS, which record, since the beginning of their working life, various characteristics of private and public employees, self-employed, recipients of unemployment benefits and retired. Furthermore, beside the addition of retrospective information to EU-SILC individual data (e.g. since the beginning of the working career up to 2005), a "forward-looking section" including information up to 2009-2010 has been added to the new panel dataset built merging survey and administrative data.

The AD-SILC dataset has been indeed the base used for analysing at the individual level Italian labour market performances in the last decades, focusing in particular on the dynamic of earnings distribution, on individual transitions among the various working statuses and on the adequacy of contributions accumulated by cohorts of workers belonging to the new notional defined contribution pension scheme, which have been some of the main subject matters of research project.

The current project IESS is a follow-up, aiming at extending and improving the tools built in the previous one. In particular, the project development consists of two main activities: i) extending the dynamic micro-simulation model T-DYMM, and ii) updating and extending the innovative longitudinal dataset AD-SILC.

Concerning the first activity, the new release of T-DYMM considers:

i) A new simulation platform LIAM2, that represents a natural evolution of the LIAM-based model;

ii) Some changes about the structures of the modules (demographic, market labour, pensions);

iii) The extension of the model with an extra sub-module which allows to analyse the dynamics of private pensions schemes.

Concerning the second activity, three mutually consistent directions can be distinguished:

i) updating micro-data collected in the administrative archives up to the end of 2013, in order to study in detail the impact of the current crisis on the individual dynamics on the labour market;

ii) adding new variables available in the administrative datasets to those already included in AD-SILC greatly increasing the sample size merging survey and administrative information also as regards individuals interviewed in all IT-SILC waves in the period 2004-2012, e.g. adding to the sample size also "new" individuals with respect to the old version in which only IT-SILC 2005 was included.

This technical note has the objective to provide a description of the structure and the characteristics of the new versions of the dynamic microsimulation model T-DYMM and the AD-SILC dataset.

The note starts with the presentation of the model. In particular, the next section summarises the recent history of T-DYMM: how it is born and the structures of the first DMSM. Section 3 gives an overview of new T-DYMM components, highlighting the most important progress of the model about demographic and pension modules. Section 4 considers the updating of the model, in particular the simulation of private pension scheme. The report proceeds with the part dedicated to the data description. Hence, section 5 presents the main advantages of AD-SILC compared to the currently available datasets, in particular regarding the IT-SILC survey data and the administrative archives collected by INPS. In addition, the first and the second versions of the dataset are compared in order to exemplify the improvements achieved in the new AD-SILC. Section 6 describes the contents of the two sources used to build AD-SILC also listing the main variables that will flow into the new dataset. Furthermore, section 6 introduces the possibility to use a further dataset, SHIW, as a source of information on wealth and private savings. Section 7 outlines the merging procedure and the structure of the new dataset, also providing some preliminary descriptive statistics about the data. Section 8 concludes.

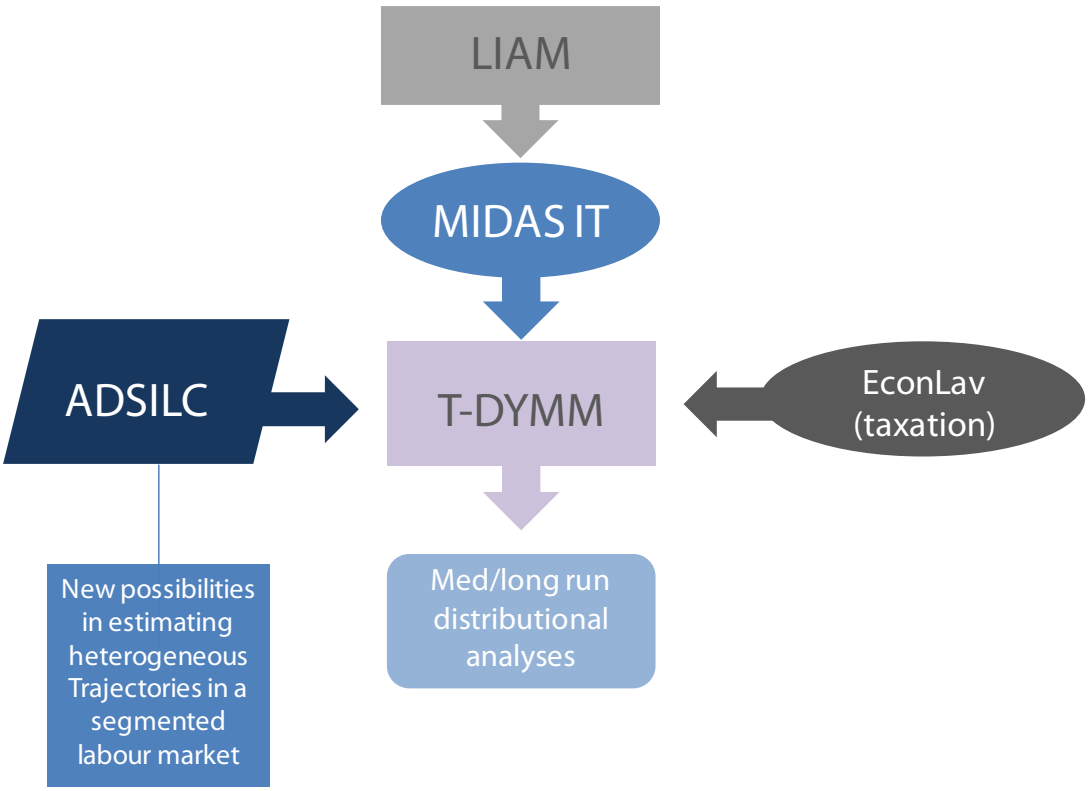
2. Recent history of T-DYMM project and the first release of the model

The first release of T-DYMM (henceforth and for to ease T-DYMM 1.0) is a dynamic microsimulation model, which significantly benefits and moves from the experience of MIDAS-IT, a DMSM developed by the ISAE (the Italian Institute for Studies and Economic Analyses).

In particular, T-DYMM 1.0 inherited from MIDAS-IT the general structure, the focus on pensions, the demographic module and the simulation platform, LIAM. Moreover, T-DYMM 1.0 integrates in the fiscal module the know-how coming from EconLav, a static micro-simulation model of the Italian tax-benefit system developed by ISFOL, with the support of the Ministry of Economy and Finance and the Ministry of Labour, for the analysis of the effects of tax and benefit system reforms (See Figure 1).

Starting from econometric estimates based on longitudinal datasets, T-DYMM 1.0 simulates individual transitions over the life cycle – such as births and deaths, marriages, educational and labour market decisions, retirement – and related outcomes such as earnings and pension benefits. By modeling with a high degree of detail the relevant pension and tax rules, the model simulates, in a life-cycle perspective, the future evolution (starting from 2006) of main demographic and economic events significant for carrying out medium-long run (both intra-and-inter-generational) distributional analyses. These with a particular focus on public pensions *adequacy,* also in connection with the radical pension reform process (mainly occurred in the nineties but still in the last years) and with recent labour market reforms resulting from demographic transition towards an ageing society. Moreover, the model aims at allowing *efficiency* evaluations of different pension schemes and it will be an aid tool for sustainability analyses, accounting for population ageing and public finance constraints.

FIGURE 1: THE MODEL AND ITS COMPONENTS



The stylised structure of the model consists of three main modules linked to each other by recursive feedbacks (i.e. in the same period the causal relationship is unidirectional), to be integrated with a fourth (so far external) one regarding the taxation system. In more detail, T-DYMM 1.0 comprises:

1. A **Demographic module**, inherited by MIDAS (Dekkers et al., 2009); it serves to estimate inter-generational persistence and highest education level achieved by the parents.
2. A **Labour market module** that probabilistically simulates individual labour market dynamics, employment transitions (full-time and part-time employed, unemployed, inactive people), transitions among sectors and contractual status (fixed term, permanent, parasubordinate).
3. A **Pension module** for the retirement benefit calculation and for the definition of eligibility requirements and retirement decisions.
4. A **Fiscal module**, running at the end of simulation process, that carries out the overall process relating gross to net labour and pension incomes and providing a high degree of details about the Italian tax-benefit system.

T-DYMM uses alignment procedures (i.e. calibration) – in particular in the demographic model – in order to hook some aggregated results (couples formation, fertility and mortality rates; employment rates; disability rates) on official projections. The main source of alignment is AWG (Ageing Working Group) baseline demographic and macroeconomic projections for the period 2006-2060.

3. The last release of T-DYMM

This section aims to provide an introduction to the new T-DYMM (henceforth, T-DYMM 2.0), highlighting the most important differences between the two versions and focusing on two main points:

i) The new platform of the simulation with new programming code (LIAM2);

ii) The structure of the model and the new characteristics of the modules.

The model has been written on the new simulation platform LIAM2, that represents a natural evolution of the LIAM-based model, ensuring additional assets in terms of speed and data capacity. LIAM2 is a microsimulation modelling toolbox which is built as generic as possible thus allowing to develop almost any microsimulation model as long as it uses cross-sectional ageing. By making it available for free , together with increased cooperation through meetings and code sharing could greatly reduce the development costs (in terms of both time and money) of microsimulation models and modules. In our view it will allow the use of huge dataset, such as AD-SILC, or expanding the survey data to the whole population in order to fulfil representativeness requirements. Due to its new programming code, LIAM2 – in the version 0.8.2 – is much more flexible and increases the simulation scopes, making a given model more precise.

The general structure of T-DYMM 2.0 (See Figure 2) has not been changed, therefore it consists of three main modules linked to each other by recursive feedback, plus a fourth external module as in the case of the previous version of DMSM. In particular:

1. A **Demographic module**, improved to account for intergenerational persistence and parental background (e.g. parents' highest education level). Compared to the old version, the main advantage is the application of a new more flexible programming code. The module simulates two types of demographic and educational events (or choices):

    a. events that mainly modify the population structure by sub-groups composition (age and education classes), such as mortality, fertility choices, referred to as external events and aligned to AWG 2015 projections (until 2060);

    b. events that affect the household structure, such as exit from the family of origin, cohabitation, marriage and separation. These are called internal events and are based on micro-econometric estimates.

2. A **Labour market module** that probabilistically simulates individual entry into the labour market, employment transitions (full time and part-time employed, unemployed, inactive people), transitions among sectors and contractual status (fixed term, permanent, parasubordinate). These transitions are based on Monte Carlo techniques mainly consisting in the comparison of a random number ($0<r<1$) drawn from a uniform distribution, and a conditional probability obtained from discrete choice estimates or transitional matrices. Both methods need longitudinal micro data and provide us with the parameters we use in the simulation program. This module heavily benefits from the AD-SILC dataset, presented and described in the following sections.

3. A **Pension module** that includes two sub-modules:

    a. Pension benefit calculation;

    b. Eligibility requirements and retirement decision.

The first sub-module computes old age and survivor pension benefits as well as minimum pensions, supplements and social assistance allowances, according to a number of schemes and different regimes which characterize the Italian social security system (i.e. from pre-1992-reform up to 2011, passing through the 1995 reform which introduced the NDC scheme for younger workers). Therefore, the amount of individual pension benefit depends on the combination of many parameters, among which the life-cycle profile of labour incomes, the contribution seniority at retirement, the contribution rate during working life, macroeconomic growth assumed during the period of pension contribution and the pension scheme.

In the T-DYMM 2.0 (like in the old version) the second sub-module is based on a deterministic transition conditional on achieving necessary requirements for old age pension claiming, and a probabilistic decision process based both on consumption smoothing and absolute living standard objectives for those who meet the requirements for early retirement (i.e. a combination of age and seniority or a seniority level above certain thresholds).

4. Whereas the first three modules are replicated in sequence year after year, the **Fiscal Module** is so far an additional block running at the end of the multi-period simulation process. In particular, the state-of-the-art T-DYMM 2.0's fiscal module (like in the old model version), starting from individual gross (of personal income tax) incomes provided by labour market and pension modules, computes the net figures. The module includes tax credits for labour incomes and pensions, fiscally dependent relatives, but it does not allow for other expenditures that can be partly subtracted from the gross personal income tax as these are not simulated by the dynamic model, such as health care expenditures of disabled relatives, supplementary pension premiums and alimonies.

4. Private pension module

T-DYMM 2.0 includes in the pension module a new sub-module to simulate and project the evolution of private pensions schemes (henceforth PPS) over time.

Complementary private retirement schemes constitute just one possible way for transferring purchasing power over time, possibly competing with other types of investment. Most of the analyses consider only the option whether an individual decides to transfer or not a part of his/her TFR to a complementary social security (hence, referring to employees only). However, considering both the share of self-employment and the recent diffusion of atypical jobs, just about 60 percent of Italian workers are employees and thus have a TFR to devolve to PPS (in the form of occupational funds, the so-called second pillar). The remaining 40 percent can participate in individual retirement accounts (IRA, third pillar) with voluntary contributions, in addition to those made to the first mandatory pension pillar.

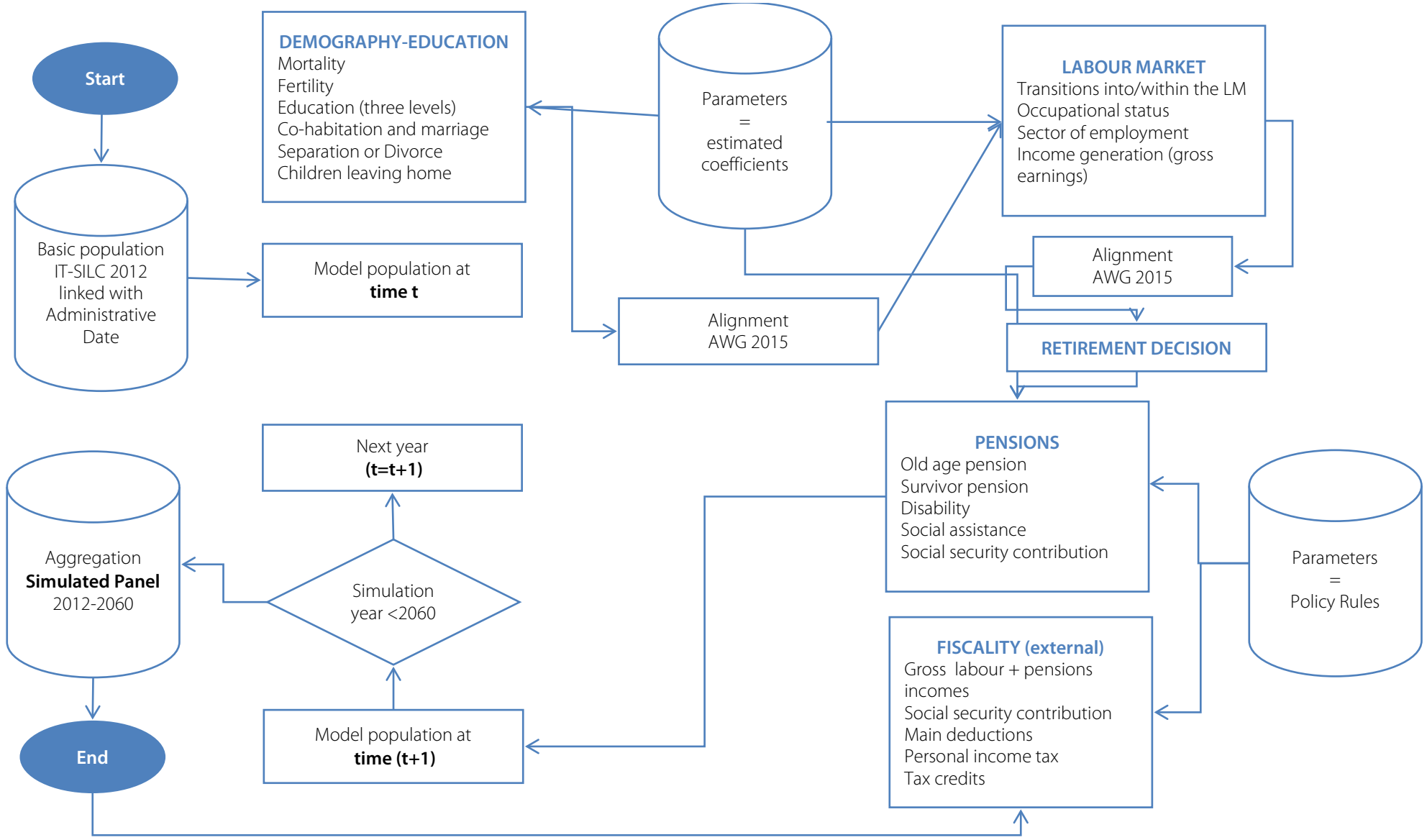
In this work, the sub-module for PPS focuses only on workers' investment decisions on private pension schemes, in a life cycle perspective, and does not consider the decision about the TFR. The sub-module has been achieved in two steps:

1. Analysing the individual choice about private pension investment;
2. Including the analysis results in the simulation platform LIAM2.

The individual choice of investment on PPS has been analysed using SHIW (Survey of Households' Income and Wealth). The probability to invest in PPS of a single agent has been estimated through a regression function.

The results of this econometric analysis have been implemented in the simulation platform LIAM2, using wealth's information contained in AD-SILC

FIGURE 2: THE STRUCTURE OF T-DYMM 2.0



5. The AD-SILC dataset: advantages, limits and improvements in the new version of the dataset

Assessing individual working histories, and in particular their transitions among different working statuses, requires the availability of a longitudinal micro dataset – e.g. the same individual has to be observed for many years – where detailed information about socio-economic characteristics of interviewed people has to be included. Currently, longitudinal information are provided by two different sources:

- Administrative datasets: e.g. INPS archives, Register of Active People and Register of Retired (*Casellario degli Attivi e Casellario dei Pensionati*).
- Survey datasets: for instance, panel data were provided by the SHIW biannual survey carried out by the Bank of Italy, by the PLUS survey carried out by ISFOL or by the longitudinal surveys covering EU countries ECHP and EU-SILC managed by EUROSTAT (ISTAT for the Italian version).

Both sources have some pros and cons. Administrative datasets main merit concerns their very wide coverage (sometimes they collect information about the whole reference universe) and the very long time span considered; usually administrative archives observe the individual since the moment he/she enrols into the group followed by the archive (e.g. he/she becomes a private employee or an unemployment benefit recipient). However, they usually collect only information needed for administrative modules, so they do not record crucial information needed to deeply analyse labour market issues. For instance, in INPS archives employees are followed for the whole working life, but individual educational attainments and marriage status are not recorded, so main determinants of workers' outcomes cannot be estimated.

On the contrary, panel surveys record dozens of individual characteristics, but they usually follow workers only for a short time span and/or have a very limited sample size. For instance, the EU-SILC dataset is based on a rotation scheme where individuals are followed for four years at most, while the panel quota of the SHIW dataset has a limited sample size, that makes very difficult deep analyses of specific groups of workers.

In any event, thanks to the previous project, "Innovative datasets and models for improving welfare policies", at present it is also available the innovative AD-SILC dataset –merging the SILC survey data collected by ISTAT with an administrative information provided by INPS – which has the value to overcome some of the drawbacks of the above mentioned sources. Therefore, micro-data from the survey IT-SILC 2005 have been merged with several variables collected in the INPS that provide retrospective information at the individual level since the beginning of their working life up to 2009. The improvement in AD-SILC, compared to the existing datasets, is put into effect in several respects:

- It contains both individual time variant variables regarding working conditions for the whole working career (collected in administrative Archives) and individual time invariant characteristics and time variant features (surveyed by IT-SILC in 2005; e.g. educational attainment, parental education and occupation, marital status, citizenship).
- Being based on INPS Archives and on Registers of Active People and Retired, it provides information concerning all typologies of jobs, including when the worker receives an unemployment benefit or a sickness/maternity allowance. Consequently, all kinds of workers' transitions can be observed: when the individual is not surveyed in a year it means that he/she has been inactive or has been an informal worker for the whole year.
- Collecting information also at the firm level (e.g. detailed sector and firm's size, both at the unit and the holding level) it offers the possibility to match employers and employees characteristics with much more precision than by individual datasets where employer's characteristics are inferred with a much lower detail through workers' responses.

AD-SILC has some limits as well, basically due to the drawbacks of the original sources. For instance, longitudinal information come from the administrative source of the dataset (e.g. INPS archives): on one hand, they refer only to individuals paying contributions to INPS or to the other pension funds (e.g. funds for self-employed managed by their professional order), hence excluding individuals who for some reasons do not belong to the labour force (at least formally), and on the other hand, it gives (longitudinal) information only about individual careers, hence excluding relevant individual characteristics (e.g. changes of family characteristics). Concerning the survey source (IT-SILC 2005), the available information refers only to 2005 which precludes the possibility to observe potential changes in the individuals' conditions, which albeit much less time variant, are not necessarily immutable (e.g. marital status, childbirths, etc.).

Therefore, a new version of AD-SILC has been built in order to update and extend the previous one. Besides, the new AD-SILC dataset represents an upgrade since some of the drawbacks deriving from the cross-sectional nature of IT-SILC are overcome by virtue of exploiting the longitudinal design of IT-SILC, that follows individuals and households for at most 4 years.

Therefore, additional advantages characterise the second version of AD-SILC. First of all, adding more waves significantly enlarges the sample size which will allow to analyse in detail some aspects that previously were impossible or unreliable; e.g. to study by means of robust techniques the dynamics followed by smaller and specific groups of the Italian labour force. Second, being updated up to 2013 (and for some employees also up to the second half of 2014) the extended AD-SILC will allow to study in a comprehensive way the weaknesses engendered by the current macroeconomic crisis and its impact on labour market dynamics and labour income distribution. Finally, the new longitudinal IT-SILC database allows to take into account changes in the individual statuses, different from those strictly linked to the labour market, that took place during the time interval present in IT-SILC (e.g. transitions from single to married, from married to divorced, childbirths, etc.). Thus, it allows to estimate divorce and marriage rates, as well as childbirth rates. Moreover, employment and contract transitions can be analysed in relation to changes occurred in the family structure and/or the education level.

In short, the new version of the dataset will allow, on the one side, to strengthen the results already obtained with the previous "shorter" version of AD-SILC and, on the other side, to greatly improve the possibility to carry out detailed researches on individual dynamics on the labour market. Furthermore, it will allow to strengthen and enlarge the scope of the dynamic microsimulations run by means of the T-DYMM model.

6. Structure and variables of IT-SILC and of Italian administrative datasets

The AD-SILC dataset has been built merging longitudinal data collected in several administrative archives and regarding all individuals belonging to the specific group recorded by an archive (e.g. private employees or professionals) and the survey micro-data IT-SILC, the Italian version of the EU country survey EU-SILC. In the first version of AD-SILC, however, only cross-section data of IT-SILC 2005 has been used, while eight more waves have been added in the current version of the dataset (in detail, all the cross-sectional waves of IT-SILC in the period 2004-2012 are now considered; notice that, according to the IT-SILC design, individuals can be re-interviewed for at most four consecutive waves of cross-sectional IT-SILC).

It has allowed to link rich information about individuals' background gathered in IT-SILC with information about their working histories collected in the administrative archives since the beginning of the individual working life (e.g. since the moment he/she starts to belong to a specific group) up until 2013/2014.

IT-SILC is the Italian database of the European Union Survey on Income and Living Conditions (EU-SILC), which has been developed as a flexible yet comparable instrument for the follow-up and monitoring of poverty and social exclusion at the EU and national levels. In general, EU-SILC covers data and data sources of various types: cross-sectional and longitudinal; household-level and person-level; economic and social; from registers and interview surveys; from new and existing national sources.

The sample design of IT-SILC dataset is based on two-stage sampling, where, for each region, municipalities are divided into auto-representative (with larger population size) and not auto-representative (smaller size) municipalities. The first belongs to the sample within which households are systematically drawn from the register office records. For the latter, instead, another two stages of sampling are designed, where a sample of municipalities is chosen and then households are randomly selected within each municipality.

Similarly to the European version, the scheme of IT-SILC comprises two components – cross-section and longitudinal. In particular, this integrated design involves a rotational panel in which a new sample of households and persons is introduced each year to replace a part (e.g. one quarter) of the existing sample[1]. Table 1 illustrates the rotation scheme of a four-year panel, as that adopted by IT-SILC[2]. For example, in the case of wave collected in time T+3 the sample group D is interviewed for the fourth time (e.g. D4), E for the third time (e.g. E3), F for the second time (e.g. F2), while G is introduced for the first time in the dataset.

TABLE 1. THE ROTATION SAMPLE SCHEME OF A FOUR-YEAR PANEL DATASET

time	year	A	B	C	D	E	F	G	H	I	J	K	L
T	2004	A4	B3	C2	D1								
T+1	2005		B4	C3	D2	E1							
T+2	2006			C4	D3	E2	F1						
T+3	2007				D4	E3	F2	G1					
T+4	2008					E4	F3	G2	H1				
T+5	2009						F4	G3	H2	I1			
T+6	2010							G4	H3	I2	J1		
T+7	2011								H4	I3	J2	K1	
T+8	2012									I4	J3	K2	L1

1 This scheme is equivalent to the rotation panel design (Duncan et al, 1989) since each panel time-limited duration and two or more panel samples are collected in the same reference period.

2 ISTAT has adopted a panel of four years for IT-SILC, which is the minimum duration requested by Eurostat. Some datasets has adopted a six-year panel.

Therefore, the cross-section sample is composed by the union of four longitudinal samples, each one assigned to its specific wave, where: a quarter of households participates in the survey for the first time, a quarter of the households participates for the second time, a quarter for the third time and a quarter for the fourth (and last) time of the survey. For instance, in the case of Italy, the first launched survey, IT-SILC 2004, is composed of 32.000 households on total and around 8.000 households in each longitudinal sample. For the second wave, relative to 2005, the sample size comprises ¾ of the households collected in IT-SILC 2004 (corresponding to B4, C3 and D2 presented in Table 1) and newly selected 8.000 households which belong to the new longitudinal sample (D1 in Table 1).

As specified, AD-SILC has been built by merging the administrative archives by INPS with all waves of IT-SILC (from 2004 to 2012). Consequently, as illustrated in Table 1, six sample groups (D, E, F, G, H, I) has been followed for four years (the maximum time span), two sample groups (C and J) for three years, two sample groups (B and K) for two years, and finally, two sample groups (A and L) appear as cross-section samples. Besides, only the (cross-section) waves between 2007 and 2009 contain (longitudinal) sample groups for which a full panel is present.

Another important feature of IT-SILC is the large number of collected variables. Four types of data are involved: (i) variables measured at the household level; (ii) information on household size and composition and basic characteristics of household members; (iii) income and other more complex variables measured at the personal level, but aggregated to construct household-level variables (which may then be ascribed to each member for analysis); and (iv) more complex non-income or 'social' variables collected and analysed at the personal level. The information collected in each wave refers to two distinct time periods: some variables refer to the time of selection (e.g. year t), and others, among which income, to the preceding year (e.g. year t-1).

For more clarity it is useful to look at the scheme provided by Graf et al (2011) which depicts the main features of the variables, as follows:

- Kind of data:

    a. If the variable is a household variable, it belongs to the following groups of variables:

1. Basic data (basic household data including degree of urbanisation);

2. Income (total household income and gross income components at household level);

3. Social exclusion (non-monetary household deprivation indicators, including problems in making ends meet, extent of debt and enforced lack of basic amenities);

4. Labour information (childcare);

5. Housing (Dwelling type, tenure status and housing conditions, amenities in dwelling and housing costs).

    b. If the variable is a person variable, it belongs to the following groups of variables:

1. Basic data (basic personal data and demographic data);
2. Education;
3. Labour information (basic labour information on current activity status and on current main job, including information on last main job for unemployed, basic information on activity status during income reference period, total number of hours worked on current second/third ... jobs, detailed labour information, activity history and calendar of activities);
4. Health (health, including health status and chronic illness or condition, and access to health care);
5. Income (net and gross personal income, total and components at personal level).

- Type of variable: cross-sectional or longitudinal;
- Reference period (constant, current, income reference period, last twelve months, since last year, working life and childcare reference period);

- Unit (household, household member, former household member, selected respondent and household members aged 16 and over);
- Mode of collection (constructed, frame, register, interviewer, household respondent, personal interview and proxy);
- Values (range);
- Flags.

Accordingly, IT-SILC provides a wide range of variables about individuals' background that can be linked to administrative datasets in order to enrich their representativeness of the Italian workers population.

In order to build the AD-SILC dataset we have made use of various administrative archives, provided by INPS (and collecting information also about individuals not paying contributions to INPS funds, e.g. public employees and self-employed enrolled to the private fund managed by their professional association). These datasets, and the variables included, do not substantially differ from those used for the previous version of AD-SILC, and are the following:

- **EC_INPS**: collecting up to 2014 information about all workers paying contributions to INPS funds, e.g. private employees, farmers, dealers, craftsmen, parasubordinate workers and professionals without a private fund managed by their association. The reference unit in EC_INPS is the individual's working relationship in the year; this means that individuals that have more than one work relationship in a year (or experiencing during the year also periods receiving unemployment benefits or maternity and sickness allowances) present more than one record at year.
- In EC_INPS the following variables are recorded for each year: age; year of birth and (eventually) of death; sex; the province of birth and the nation of birth for individuals birth outside Italy; the date of start and end of every work relationship or of every period receiving unemployment benefits of sickness or maternity allowances; the number of weekly contributions referred to each work relationship or the number of weeks during which the benefit is paid; the fiscal code of the firm (in case of private employees); the province where the individual works; the gross wage (e.g. including also workers' contributions) paid each year for every work relationship; the amount of welfare benefits; the specific INPS funds where the individual pays contributions (this variables allows to distinguish private employees, parasubordinate, farmers, craftsmen and dealers); the specific type of contributions paid to INPS, that allows to distinguish periods spent as workers from periods spent receiving unemployment benefits, sickness allowances and maternity allowances; the workers' qualification, that refers to employees and allows to distinguish managers, white-collars, blue-collars, apprentices and, since 1998, individuals working through part-time and through fixed-term employment arrangements.
- In the previous version of AD-SILC an additional archive (EC_NOINPS) was present for those workers not enrolled to INPS funds e.g. public employees (not enrolled to INPS until 2011) and professionals enrolled to the private fund managed by their association (e.g. lawyers, architects, dentists, psychologists). At present, EC_NOINPS is integrated in EC_INPS thus obtaining a unique archive for the active workers (henceforth, *Register of Active Workers*). Substantially, the same information is collected about these workers: age; sex; the eventual date of death; the specific fund (allowing to distinguishing public employees and the different typologies of professionals); the number of weeks of contribution; the gross wage; the date of start and end of the work relationship.
- **AZ_INPS**: collecting since 1990 up to 2013/2014 the variables recording the main firms' characteristics. This dataset can be merged to EC_INPS through the fiscal code of the firm. The variables recorded for each year in this archive are the following: the detailed sector of the firm (classified through 3 digits ATECO); the kind of firm (distinguishing single firms, "mother firm" and "daughter firm"; the number of workers of the holding and of the local unit of the firm.

- **PENSIONI**: collecting yearly for the period 2004-2013 information with respect to people receiving a pension benefit, thus it represents the *Register of Retired*. The following variables are recorded: sex; age; the eventual date of death; the starting date of the retirement; the monthly gross amount of the pension benefit; the region of residence; the professional status before retirement (e.g. employee or self-employed); the seniority at the retirement (computed in weeks); the typology of pension benefit, that allows to distinguish old age, survivor, invalidity and social pensions.

We already avowed that the intention of building AD-SILC has primarily arisen from the awareness that any effort to evaluate the Italian pension system and fiscal policy changes would be hardly attainable without a suitable dataset. As far as these are the main objectives of the IESS project, AD-SILC is a key tool to make operational the dynamic microsimulation model T-DYMM. Consequently, information on social insurance contributions and on supplementary pension plans is crucial for the afore-mentioned evaluations. In this regard, IT-SILC provides some variables by specifically asking respondents to indicate whether having paid voluntary contributions to banks, insurance and financial institutions for private pension plans and what the respective amounts were[3]. As reported in table 2[4], only a small part of the population aged 16-80 paid voluntary contribution for supplementary pension schemes, and the share has reduced over the years. On the other hand, focusing only on employees the share of contributors to private pension plans has increased over the years, especially in the last two years.

TABLE 2. SHARE OF POPULATION (EMPLOYEES) WHO PAID CONTRIBUTIONS TO SUPPLEMENTARY PENSION PLANS (% VALUES)

survey year	2004	2005	2006	2007	2008	2009	2010	2011	2012	average
population (aged 16-80)	7.4	6.9	6.1	5.5	5.3	5.1	4.8	5.2	3.6	**5.5**
only employees	6.2	6.1	6.2	7.4	9.2	9.2	9.7	14.5	15.9	**9.45**

Source: elaborations on IT-SILC data

However, as long as neither IT-SILC nor the administrative archives of INPS (which indeed collects data relative to the compulsory social contributions) give any more specific information about supplementary social insurance (in particular regarding the pension system) it could be necessary to appeal to external sources. The Survey on Household Income and Wealth (SHIW), developed by the Bank of Italy, perhaps is the most suitable source for the purpose since it is explicitly built with the aim of gathering data on the incomes and savings of Italian households. Furthermore, it has a specific section dedicated to the supplementary pension plans, where several variables are collected: whether a member of the household paid into a personal retirement plan or supplementary pension fund; the number and the type of pension funds or retirement plans the household subscribed in; the year the individual started to pay it; the amount paid in the year of the survey.

In any event, apart from some parameters, like birth/death and life expectancy rates, which are essentially taken from official statistical databases (e.g. ISTAT, AWG) and some other exceptions[5], the main database used for the series of estimations and simulations within the framework of the IESS project are taken from AD-SILC.

3 Contributions are referred to the year preceding the survey year.

4 Values reported in the first row are calculated on the basis of the following question of IT-SILC: "PENSIT: In year t-1, did you pay any voluntary contributions to banks, insurance or financial institutions for individual private pension plans?". Values reported in the second row refer to the following question: "CONVOL: In year t-1, were any voluntary contributions withheld from your paycheck and deposited to a company pension fund? ".

5 For example, the afore-mentioned variables collected in SHIW.

7. The procedures for building the AD-SILC dataset

As in the case of the first version of AD-SILC, the merging procedure between administrative and survey data has been made according to the rules regulating the SISTAN, e.g. the Italian National Statistical System. These rules state that Institutions members of SISTAN are allowed to exchange among themselves statistical information, including sensible data (as fiscal codes) for research purposes.

INPS and ISTAT are members of SISTAN, so they can exchange data including fiscal codes (that are blanked in the datasets provided to the public). Hence, the merging procedure started with the official request by INPS to receive the SISTAN version of IT-SILC, e.g. that including fiscal codes.

In particular, for the individuals sampled in IT-SILC, INPS took fiscal codes (e.g. a unique key characterizing all residents in Italy) and drew out from the three above mentioned archives all available records concerning those individuals. Once these records have been drawn out INPS again blanked fiscal codes for privacy reasons, replacing them with an individual identification key (common for INPS and IT-SILC). Hence, by means of the identification key the administrative archives have been merged in a single very large administrative dataset, and then information recorded by IT-SILC have been linked to the variables recorded by administrative archives for all individuals that have been registered in their lives in the Register of Active Workers or in the Register of Retired.

Once the merging procedure has been completed the very long retrospective panel AD-SILC, where individual data have been recorded since their entry in the labour market up to 2013/2014, has been built. Hence, AD-SILC is an unbalanced panel because, by definition, individuals are followed for a different number of years.

The retrospective unbalanced panel size amounts to 5,016,214 observations, referred to 147,777 individuals recorded at least once in an administrative archive (individuals sampled in IT-SILC and absent in administrative archives are children and people never active and not getting any pension). The number of observations recorded in each year is shown in table 3.

TABLE 3. NUMBER OF OBSERVATIONS IN AD-SILC BY YEAR

year	values	%	year	values	%	year	values	%
1927	1	0	**1958**	16,243	0.32	**1988**	82,574	1.65
1928	1	0	**1959**	18,690	0.37	**1989**	85,015	1.69
1930	1	0	**1960**	20,839	0.42	**1990**	101,189	2.02
1931	5	0	**1961**	23,362	0.47	**1991**	91,403	1.82
1932	8	0	**1962**	24,952	0.5	**1992**	92,759	1.85
1933	11	0	**1963**	27,298	0.54	**1993**	87,667	1.75
1934	19	0	**1964**	28,057	0.56	**1994**	87,329	1.74
1935	19	0	**1965**	30,617	0.61	**1995**	90,505	1.8
1936	30	0	**1966**	32,630	0.65	**1996**	93,696	1.87
1937	44	0	**1967**	34,850	0.69	**1997**	90,413	1.8
1938	68	0	**1968**	37,020	0.74	**1998**	91,348	1.82
1939	115	0	**1969**	38,906	0.78	**1999**	93,567	1.87
1940	150	0	**1970**	40,555	0.81	**2000**	97,867	1.95
1941	215	0	**1971**	42,272	0.84	**2001**	101,153	2.02
1942	280	0.01	**1972**	47,380	0.94	**2002**	103,956	2.07
1943	430	0.01	**1973**	50,929	1.02	**2003**	104,828	2.09
1944	444	0.01	**1974**	61,668	1.23	**2004**	107,879	2.15
1945	504	0.01	**1975**	58,933	1.17	**2005**	187,750	3.74
1946	579	0.01	**1976**	63,230	1.26	**2006**	191,693	3.82
1947	732	0.01	**1977**	63,094	1.26	**2007**	200,453	4
1948	937	0.02	**1978**	65,735	1.31	**2008**	201,853	4.02
1949	1214	0.02	**1979**	68,376	1.36	**2009**	203,223	4.05
1950	1,662	0.03	**1980**	71,120	1.42	**2010**	203,211	4.05
1951	2,200	0.04	**1981**	74,285	1.48	**2011**	207,814	4.14
1952	3,385	0.07	**1982**	75,184	1.5	**2012**	198,690	3.96
1953	3,850	0.08	**1983**	74,970	1.49	**2013**	176,506	3.52
1954	5,017	0.1	**1984**	75,292	1.5	**2014**	86,393	1.72
1955	6,525	0.13	**1985**	80,429	1.6	**2015**	29	0
1956	7,879	0.16	**1986**	78,457	1.56			
1957	14,165	0.28	**1987**	79,588	1.59	**Total**	**5,016,214**	**100**

Source: elaborations on AD-SILC data

It has to be remarked that the administrative archives concern each job relationship had by an individual during a year. For example, if the individual has changed job or the contribution typology during the year he/she would be recorded in the archive twice or more that year. Therefore, individuals often have more than one observation (e.g. a row in the dataset) every year, so that the total number of observations exceeds the number of individuals surveyed each year. Considering only one individual record for each year we observe the number of individuals sampled in IT-SILC for which administrative archives provide information every year. Indeed, the total number of individual observations reduces to 3,306,200 and their distribution up until 2015 is shown in table 4.

TABLE 4. NUMBER OF INDIVIDUALS IN AD-SILC BY YEAR

year	values	%	year	values	%	year	values	%
1927	1	0	**1958**	11,769	0.36	**1988**	58,508	1.77
1928	1	0	**1959**	13,511	0.41	**1989**	59,461	1.8
1930	1	0	**1960**	14,876	0.45	**1990**	60,303	1.82
1931	3	0	**1961**	16,503	0.5	**1991**	61,095	1.85
1932	5	0	**1962**	17,999	0.54	**1992**	61,341	1.86
1933	8	0	**1963**	19,267	0.58	**1993**	60,112	1.82
1934	12	0	**1964**	19,932	0.6	**1994**	59,904	1.81
1935	11	0	**1965**	21,858	0.66	**1995**	60,147	1.82
1936	19	0	**1966**	23,384	0.71	**1996**	62,139	1.88
1937	28	0	**1967**	24,791	0.75	**1997**	62,929	1.9
1938	39	0	**1968**	26,180	0.79	**1998**	63,844	1.93
1939	65	0	**1969**	27,646	0.84	**1999**	65,507	1.98
1940	97	0	**1970**	29,170	0.88	**2000**	67,202	2.03
1941	130	0	**1971**	31,214	0.94	**2001**	68,876	2.08
1942	163	0	**1972**	34,377	1.04	**2002**	70,792	2.14
1943	233	0.01	**1973**	37,076	1.12	**2003**	71,773	2.17
1944	277	0.01	**1974**	42,064	1.27	**2004**	72,273	2.19
1945	312	0.01	**1975**	43,278	1.31	**2005**	113,536	3.43
1946	385	0.01	**1976**	45,278	1.37	**2006**	115,789	3.5
1947	498	0.02	**1977**	46,260	1.4	**2007**	118,461	3.58
1948	634	0.02	**1978**	48,680	1.47	**2008**	119,899	3.63
1949	877	0.03	**1979**	50,194	1.52	**2009**	119,782	3.62
1950	1,196	0.04	**1980**	52,136	1.58	**2010**	119,847	3.62
1951	1,597	0.05	**1981**	53,090	1.61	**2011**	119,704	3.62
1952	2,113	0.06	**1982**	52,966	1.6	**2012**	117,086	3.54
1953	2,732	0.08	**1983**	53,438	1.62	**2013**	106,558	3.22
1954	3,539	0.11	**1984**	53,447	1.62	**2014**	56,985	1.72
1955	4,403	0.13	**1985**	54,392	1.65	**2015**	29	0
1956	5,350	0.16	**1986**	55,372	1.67			
1957	10,449	0.32	**1987**	56,992	1.72	**Total**	**3,306,200**	**100**

Source: elaborations on AD-SILC data

The data reported in table 3 and 4 refer to the overall sample (surveyed in IT-SILC) extracted from the administrative archives. Therefore, we have individuals present either in EC_INPS or in PENSIONI (e.g. active and retired persons).

Notice that the number of observations largely increases since 2005. This is due to the fact that information about pensioners provided by the Register of Retired are recorded on an annual basis since that year. However, information about pensions include also retrospective variables – e.g. the previous job (employee or self-employed), the seniority at the retirement and, mostly, the precise starting date of payment of pension benefits (e.g. a 2005 record reports that an individual retired on 15/11/1989); thus pension records can be easily reported on an annual basis also for the years prior to 2005.

The total number of observations in the Register of Retired archive amounts to 627,087 records corresponding to 61,165 individuals. Therefore, 4,389,136 observations, referred to 136,914 individuals are recorded in the Register of Active Workers.

It has to be remarked that the major part of the surveyed persons are present only in the EC_INPS archive as they result active, or in any event not retired, by their most recent record in the administrative archives. However, as far as AD-SILC is a retrospective panel database it provides the entire working history also for already retired workers; consequently, a significant number of individuals (50,311) are present in both registers. Finally, some individuals present in the Register of Retired have never paid any contributions in their lives (e.g. females receiving survivor pensions or disable individuals who receive disability benefits) In particular, 10,854 individuals do not appear in the Register of Active Workers due to above mentioned reasons, almost 70 percent of which are women.

8. Conclusions

We specified in the introduction that the IESS project is the follow-up of the studies and the analytical tools developed in the project "Innovative Datasets and Models for Improving Welfare Policies", which indeed has been realised by building the innovative longitudinal dataset AD-SILC useful for studying the dynamics of the Italian labour market in the last decades and the dynamic microsimulation model T-DYMM, based on the database AD-SILC, and useful for analyzing distributive issues about labour market dynamics and pension benefits adequacy in the long term. The current project, instead, has the objective to update and to extend T-DYMM model and AD-SILC dataset in order to produce more robust analyses about the individual dynamics on the labour market and the pension accumulation in the contributory scheme, thus enabling a more effective way to inform policy makers about the criticalities and opportunities emerging in the Italian labour market and pension system.

The new version of T-DYMM has been written using the new programming code (with new platform LIAM2), which is much more flexible and increases the simulation scopes, making a given model more precise. The new sub-module concerning private pensions schemes, which is now explicitly included in the simulation of public pension, increases the reliability of the model at analysing the impact of (public and private) pensions schemes on inequality.

Concerning the AD-SILC dataset is without doubts the best dataset currently available for Italy that allows one to carry out detailed analyses about the long term evolution of working conditions of a large sample of individuals and about the main socio-economic individual characteristics related to their trajectories on the labour market. Therefore, AD-SILC is a crucial database to be used for running the T-DYMM microsimulation model.

9. References

Dekkers, G., Buslei H., Cozzolino M., Desmet R., Geyer J., Hofmann D., Raitano M., Steiner V., Tanda P., Tedeschi S., Verschueren F. (2009). What Are the Consequences of AWG Projections for the Adequacy of Social Security Pensions? ENEPRI Research Report N. 65, AIM WP4.

European Commission (2007), *Employment in Europe 2007*, Bruxelles.

Graf, M., Wenger, A., & Nedyalkova, D. "Quality of EU-SILC data", 2011.

O'Donoghue, C. (2001), "Dynamic Microsimulation: A Methodological Survey", Brazilian Electronic Journal of Economics, Universidade Federal de Pernambuco, vol. 4, n.2.

Verma, V., Betti, G., & Ghellini, G. (2006). "Cross-sectional and longitudinal weighting in a rotational household panel: applications to EU-SILC", Università di Siena, Dipartimento di metodi quantitativi.