Università degli Studi di Modena e Reggio Emilia
Dipartimento di Economia Politica

Materiali di discussione

\\ 595 \\

CAPP_DYN: A Dynamic Microsimulation Model for the Italian Social Security System

by

Carlo Mazzaferro[1]
Marcello Morciano[2]

October 2008

[1] Department of Economics, University of Bologna and Capp
[2] University of East Anglia, and Capp

Viale Jacopo Berengario 51 – 41100 MODENA (Italy) tel. 39-059.2056711Centralino) 39-059.2056942/3 fax. 39-059.2056947

CAPP_DYN: A Dynamic Microsimulation Model for the Italian Social Security System

Carlo Mazzaferro*

Marcello Morciano**

October 2008

Abstract:

We present the technical structure of CAPP_DYN, a population based dynamic microsimulation model for the analysis of long term redistributive effects of social policies, developed at CAPP (Centro di Analisi delle Politiche Pubbliche) to study the intergenerational <u>and</u> the intragenerational redistributive effects of reforms in the social security system. The model simulates probabilistically the socio-demographic and economic evolution of a representative sample of the Italian population for the period 2005-2050. After a short review of the existing similar models for the Italian economy, a rather detailed analysis and discussion of the functioning of the model as well as a description of estimation procedures employed in each single module of the models is offered.

JEL Classification: C51, C52, H55

Keywords: Dynamic microsimulation, lifetime and intragenerational redistribution, social security systems

* Department of Economics, University of Bologna and Capp

** University of East Anglia, and Capp

1. Introduction[1]

Although dynamic micro simulation models (MSM) have a well-establish tradition in several developed countries as a tool for the evaluation of the long-run distributional effects of public policies (O'Donoghue, 2001; Zaidi and Rake, 2002; Klevmarken, 2005), in Italy their use is recent and not completely developed. All the same our country has a complex and wide-spread welfare state, with particular reference to pensions, and at the same time is facing an intense demographic aging process. Notwithstanding research on the long run redistributive effects of both reforms of the social security system and the ageing process are still little developed.

The first dynamic MSM for the Italian economy, DYNAMITE (Ando and Nicoletti Altimari, 2004) was developed at end of the 1990s within a Bank of Italy research project. It was employed mainly to analyze the effects of demographic transition and social security reforms on private savings. Following this work, Vagliasindi (2004) developed MINT, a dynamic population MSM which analyses medium-long run distributional effect of pension system and the medium term redistributive impact of changing in personal income taxation. Both these model are at present not in use and, according to our knowledge, other dynamic MSMs able to carry out a long-run distributive evaluation of public policies in Italy do not exist.

CAPP_DYN shares with the models mentioned above the aim of describing in details the socio-demographic structure of the Italian population as well as providing a micro analysis of the evolution in the supply side of the labour market, in the labour income structure and in the pension-related choices.

CAPP_DYN comes up within a research project carried out by the CAPP (*Centro di Analisi delle Politiche Pubbliche*) under the auspices of the Italian Department of Employment and Social Policies with the aim of assessing the distributional effects of social security reforms adopted in the previous decade (Ministero del Lavoro e delle Politiche Sociali, 2005). Afterwards, the model has been improved and further developed (Mazzaferro and Morciano, 2005; Morciano, 2007; Ministero della Solidarietà Sociale, 2008).

It allows the simulation of the socio-demographic and economic evolution of a representative sample of the Italian population for the period 2005-2050. The base year population (2005) is derived by the 2002 wave of the Bank of Italy's Survey of Households Income and Wealth (SHIW). The sample is reweighted in order to align socio-demographic distributions with the Italian population. The dynamic aging of micro-characteristics is probabilistic, in particular it is carried out

[1] We wish to thank Simone Tedeschi for technical assistance.

by means of finite and discrete markovian processes. Some behavioural functions have been introduced, the main being the one governing the retirement choice.

Once the population structure has been defined and labour incomes have been generated the model simulates the main social security benefits, with a high level of institutional detail and according to the pension scheme provision being in force. Then the model can estimate the distributional effects of important social security components as well as the impact of its reforms, allowing the implementation of both cross-sectional (at different point of time) and inter-temporal life cycle (on individuals living in different periods) analyses. Recently a module that estimates the number of disabled has been embedded in the model allowing the projection, over the whole period, of the number of not self-sufficient individuals and the related long term care expenditure.

CAPP_DYN is linked, through an alignment process, to the official demographic forecasts provided by ISTAT and is calibrated in order to follow the GDP and wage growth consistent with the evolution of the number of employed individuals.

CAPP_DYN shares, with other MSM models, advantages and drawbacks of this technique. In particular it allows a detailed redistributive analysis of the social security system, which can be carried out both in a cross sectional and in an intertemporal perspective. On the other hand, being based on a population derived from a survey, great attention must be paid to the initial selection to the sample representativeness and the difficulty to extract the effects of unobservables from data. Moreover it is important to remember that CAPP_DYN does not simulate the supply side of the economy. Therefore alignment is a tricky aspect that must always be considered with attention in order to guarantee consistency with external demographic as well as economic forecasts.

2. General Features

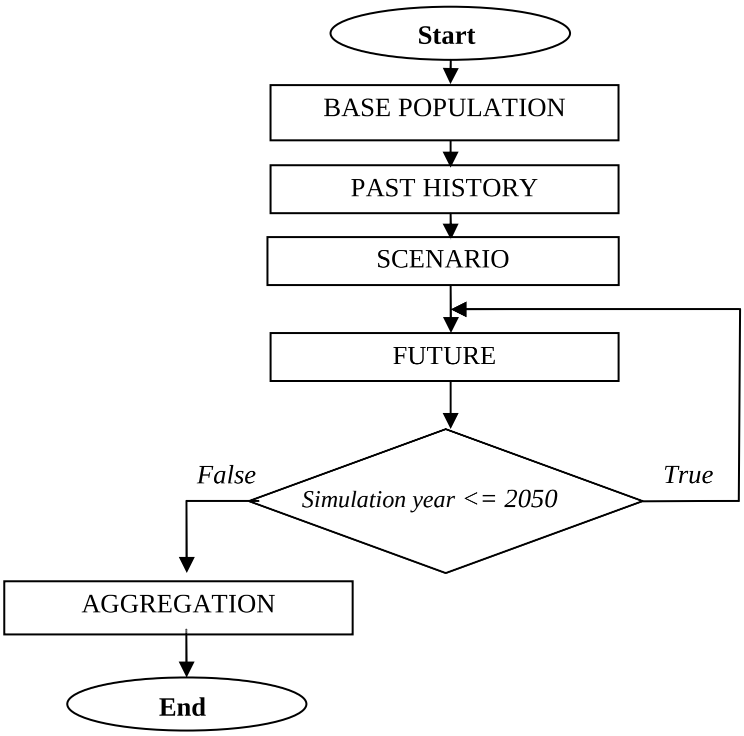
According to the taxonomy offered by O'Donoghue (2001), CAPP_DYN presents the following features:

- It is a closed model: it simulates life-cycle evolution of the main demographic and economic population features. New individuals enter in the population each year due to birth and net inflows migration, while others exit due to death.
- It is a dynamic ageing model: individual characteristics are periodically up to dated due to dynamic ageing processes based on discrete stochastic transitions among states.
- It is a discrete time model: transition and updating processes are carried out at end of each year.

- The ageing process is probabilistic: considering a particular event, partitioned into a number of mutually exclusive states at each point in time, transitions among states are achieved through probabilistic methodologies. In particular, transitions are obtained by means of a *Monte Carlo* technique.
- Units of analysis are both individuals and households.

More specifically, the model is structured in four blocks as shown in figure 1 below.

Figure 1 CAPP_DYN structure



- Base population: this block holds the procedures needed to generate the base year population. Socio-economic information for the basic units are drawn from Survey of Households' Income and Wealth (SHIW) 2002. A set of statistical methods is then employed in order to improve sample representativeness.
- Past history: it retrospectively reconstructs the working path and earnings for basic units already having a contributory history in the base year.
- Scenario: It defines the exogenous parameters values of the model. In particular, it depicts the dynamic path of macro-demographic (mortality, fertility and migration) and

macroeconomic (GDP and earnings growth) variables. Policy parameters and some behavioural rules – in particular pension-related decisions - are set within this model section too.

- Future: it is the main section of the model. This block contains a set of modules which simulate the socio-economic evolution of the micro-units according individual observed characteristics. More specifically, the model applies recursively the modules and sub-modules reported in table 1 below. Each module in the "Future" block produces a yearly cross-section of outputs from 2005 to 2050.
- Aggregation: this is the last step of the simulation. The set of annual outputs cross-sections is aggregated in order to produce a panel containing the socio-economic information for the population in the period 2005-2050.

In the following a detailed description of the contents of each block is offered.

Table1 "Future" block modules

Event	*Potential candidates*
	Demographic Module
1 Ageing	All individuals
2 Mortality	All individuals
3 Fertility	Married Women aged 16-49
4 Migration	Adds new individuals aged 16-65
5 Exit from household of origin	Children aged 18-34
6 Marriage	Singles divorced or widowed aged 16-60
7 Divorce	Married aged below 50
	Health Module
9 Disability	All individuals
	Education, Labour Market Module
10 Compulsory schooling	Individuals aged below 16
11 Choice of post-compulsory educational level	Individuals aged 16 which have completed compulsory education
12 Tertiary education	Individuals enrolled in tertiary education
13 Entry in the labour market	Individuals leaving or abandoning the school
14 Transitions between labour and non labour statuses	All individuals excepted pensioners and students
15 Transitions between contractual types	All active individuals in the labour market
16 Wages and Salaries	The same
	Social Security Module
17 Retirement	All non pensioners accruing retirement requirements
18 Survivors' pensions entitlement	Survivors (spouse and children) fulfilling law requirements
19 Social Pensions entitlement	Individuals aged above 65 entitled for assistance benefits
20 Pension benefits	All pensioners (old-age and seniority) in the three systems (defined benefit, defined contribution and mixed)
21 Supplements to minimum (*integrazioni al minimo*) and social assistance supplements (*maggiorazioni sociali*)	Pensioners fulfilling age and economic condition requirements

3. Blocks description

- **The base year population**

The SHIW_02 is the Italian most used data base for micro-econometric and distributional analysis. The survey unit is the household, i.e. “group of individuals linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes” (Brandolini, 1999); however, as information are gathered at individual level (interest, dividends and financial assets only being recorded at family level), analyses on personal income are allowed as well. SHIW income is net of taxes and social security contributions.

It represents the Italian population and the sampling scheme is organized in two-stages: firstly, municipalities are non-randomly selected according to 51 strata; in a second step households are randomly selected within the stratum. Hence, statistical inference must allow for sampling design: to this extent a bootstrapping method has been employed.

The wave 2002 contains information on 21,148 individuals within 8,011 households units.

As in other surveys, differential response rate among groups, under-reporting and mis-reporting (especially for capital income) are likely to bias estimation based on this source. In particular, under-reporting seem significantly widespread among self-employed (nearly 20% in 1987, according to Cannari and D’Alessio (1992) estimates) and inversely correlated to household income and wealth, causing an underestimation of mean income and inequality[2]. In addition, a comparison with National Accounts data (through a grossing-up procedure) shows a slight overestimation of wages while a severe underestimation for self-employment income and net interest on financial assets is recorded (respectively by 50% and 65-70%), resulting in an underestimation of total income of about 30% (32% when interest and dividends are included, Brandolini, 1999).

Finally, top and bottom coding problems have to be accounted for when the top or the bottom of the income distribution are analyzed.

Therefore, we tried, in building the base year population, to reduce at our best the biases due to the use of a not-fully-representative data set. To this extent we applied to original sample weights a post-stratification procedure, which uses information provided be the last ISTAT census on population and houses. This procedure, developed by Gomulka and employed at present by

[2] Response rate seems declining sharply from 26% of poorest to 14% of richest (Cannari, D’Alessio, 1992).

EUROMOD (Atkinson *et al*., 1988) allows increasing representativeness of the sample for the set of socio-economic characteristic which we control for[3].

For what concern the size of the initial population, a trade off exists between improving simulation heterogeneity with a larger sample, reducing this way the estimation variance (Orcutt *et al*., 1986) and the technological constraints in processing a set sample members which, in the final period of simulation, can reach a size of several millions unities.

According to the well-establish experience of important research groups in the micro simulation area, in the present setting the model simulates the evolution of a base year population composed of 107,000 household units and 270,000 individual observations.

- **The 'historical' block**

In order to obtain a complete contributory history for each micro unit present in the base year sample, the historical module builds up retrospectively the past working history of each active individual present in the base year[4].

The life-cycle profile of past earnings is built by means of econometric estimations implemented in the "income" module. Individual earnings are then discounted by an annual variable rate amounting to the growth of real earnings in the period 1952-2001[5].

- **The 'scenario' block**

This block allows to set the values of the exogenous parameters. Table 2 displays the list of exogenous variables and the official data sources the values used in the simulations are drawn from.

In particular, it is worth noting that demographic dynamics and macroeconomic variables are not independent. Therefore, at this stage, the model uses the central demographic forecasts provided by ISTAT, the same employed by Ragioneria Generale dello Stato (RGS) in its simulation of future GDP growth and earnings, in turn representing the macroeconomic benchmark scenario.

In addition, in this sub-blocks retirement decisions rules are set, accounting both for intertemporal choice optimizing framework and for elements linking retirement decision to the achievement of a certain level of the replacement rate (i.e. the last gross earnings to first pension benefit ratio).

---

[3] A detailed report of the procedure can be found in Morciano (2007).
[4] The re-construction of active individuals in 2002 employs information on contributory seniority professional attainments and sector (actual and previous) from SHIW_02.
[5] Values are taken from Golinelli 2002.

Table 2 Data sources and reference scenarios for the exogenous variables

EXOGENOUS VARIABLES	*SOURCE*	*REFERENCE SCENARIOS*
	Demographic Variables	
Age, gender and geographical area specific mortality rates	ISTAT 01/01/2007	High, median, low
Age, gender and geographical area specific fertility rates	ISTAT 01/01/2007	High, median, low
Net Migration	ISTAT 01/01/2007	High, median, low
	Macroeconomic Variables	
Real GDP growth	Ragioneria Generale dello Stato 2007	Country base and programmed
Productivity growth	Ragioneria Generale dello Stato 2007	Country base and programmed

- **The 'future' block**

This blocks contains the whole set of dynamic ageing procedures representing the core of the model. They can be grouped into four (five) main modules:

1. Demography
2. Health
3. Education, labour market and related incomes
4. Social security

Each module is in turn composed of sub-modules. The sequence of modules and sub-modules is presented in figure 3 where an illustration of the order of simulated events is also offered.

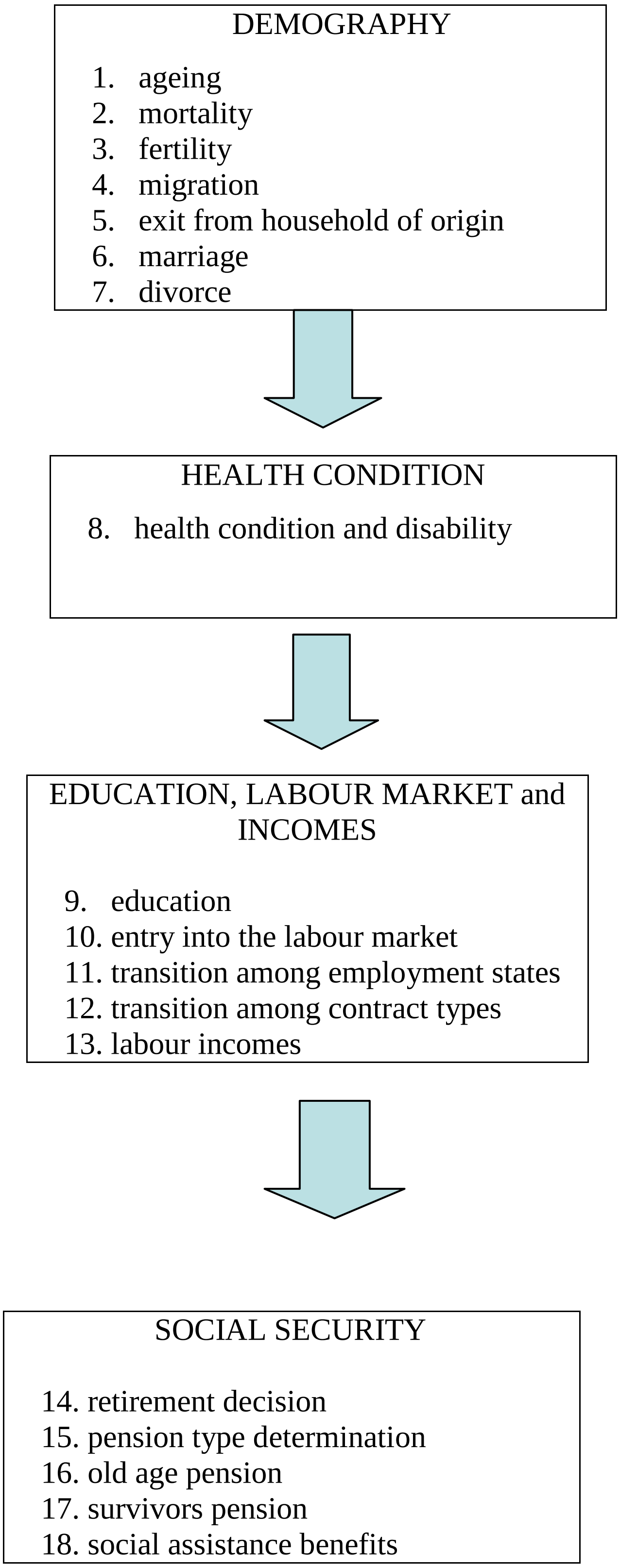
Two crucial issues of the model are worth to be mentioned:

i) the model is *sequential*
ii) the model is *recursive*

The first feature rules out from the analysis interactions between behaviors modeled within each single module. The second implies that, once all the modules have run, the model starts again the analysis of the same modules, in the same order, for the next year. These are two hypotheses usually employed in dynamic population micro simulation models. The introduction of reaction function in such a model would prove better in long term General Equilibrium Models (Auerbach and Kotlikoff, 1987) mainly aimed at studying aggregate supply, inter-temporal consumption saving and capital accumulation choices.

The general rule for the dynamic ageing of socio-economic variables - which are not exogenously defined in the “scenario” block – is *probabilistically* based. In practice, the model estimates the probability of transition among states by means of models estimated using different statistical sources. The predicted probability is then matched to a random numbers drawn from a uniform distribution with support [0;1] (i.e. *Monte Carlo* technique). The set of events simulated using this technique are reported in figure 3

Figure 2 Events simulated by CAPP_DYN



More formally, the dynamic ageing general rule of the socio-economic characteristics for units in the population is based on the discrete and finite Markovian processes (chain) theory. Given an event X, the probability of a transition from the state x_i in time t to the state x_j in time t+1 does *not* depend on the past history, but it is solely determined by the current characteristics in time t. So, the transition probabilities

$$P_{ij} = P(X_{t+1} = x_j | X_t = x_i)$$

can be represented by a strictly positive matrix, called transition or stochastic matrix:

$$P_{m*n} = \begin{vmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mn} \end{vmatrix}$$

where the *m* rows (*n* columns) identify the space of events in year t (t+1).

The i-th row of the transition matrix P : $| p_{i1} \; p_{i2} \ldots \; p_{ij} \ldots p_{in} |$, called *probability vector*, represents the probability of all possible transitions of state x_i into whatever else state in the space of the states, in period *t+1*.

Matrix *P* has the following properties:

- it is a square matrix, the number of states being the same in year t and t+1;
- $0 \le p_{ij} \le 1 \;\; \forall i, j$;
- $\sum_{j=1}^{n} p_{ij} = 1 \;\; i=1,2\ldots,m$;
- main diagonal elements represents the probabilities of inertia.

Transitions among states are yearly simulated through a Monte Carlo experiment: every year the simulator generates a random number (u_{ks}) for the k-th observation and the s-th event drawn from a uniform distribution with support [0,1]. The transition occurs if $p_{ks}-u_{ks}<0$.

4. The core of the model: the 'future' block

In this paragraph the set of modules composing the "Future" block are analyzed. Table 3 shows all the events yearly simulated by the model, the method employed for estimating the transition probabilities, the set of covariates and the data source.

Table 3 Estimation methods, covariates and data sources for the simulation of the model events

Event	***Estimation***	***covariates***	***Source***
Demography			
Mortality	Transition matrix	Age, gender, birth year	ISTAT forecast, 2005
Fertility	Transition matrix	Age, gender, birth year, area	ISTAT forecast, 2005
Migration	Transition matrix	Age, gender, birth year, area	ISTAT forecast, 2005
Exit from household of origin	Transition matrix	Age class, gender	"Famiglie e Soggetti Sociali" ISTAT, 2005
Marriage	Transition matrix	Age class, gender, area, education, marital status	"Famiglie e Soggetti Sociali" ISTAT, 2003
Divorce	Transition matrix	Wife's age class, area	"Famiglie, Soggetti Sociali e Condizioni dell'Infanzia" ISTAT, 2003
Health			
Disability	Transition matrix	Age, gender, area	Indagine sulle Condizioni di Salute, ISTAT 2003
Economy			
Education	Ordered Probit	Parents' education, gender, area	ISFol PLUS 2003
Entry in the labour market	Transition matrix	Education, age, gender, area	Rilevazione Trim. forze di Lavoro ISTAT
Transitions between labour and non labour statuses	Multinomial Logit	Education, polynomial of age , area, birth cohort, sector marital status	Rilevazione Trim. forze di Lavoro ISTAT
Transitions between contractual types	Logit	Education, age, gender, area	ISFOL Plus 2003
Work income	OLS	Age, contributory seniority, gender, area, citizenship, professional qualification, work time (part time/full time), contract, sector, education	ISFOL Plus 2003

4.1 The demographic module

The set of demographic events can be divided in two groups: external events, which modify the population structure by age, gender and geographical area and internal events, which affect the household structure only. Ageing, mortality, fertility and immigration are included in the former group, while exit from the family unit, marriage and divorce are part of the latter.

The general functioning of the demographic module is depicted in figures 4 and 5. First, external events are simulated. Each yearly simulation ages the population by one year. Then, simulation goes on in determining the number of observations that exit the model due to death. In the following step the model simulates new births. Finally, the stock of population varies every year also due to net migration.

Once the population size and composition have been defined for each period, the model starts the simulation of processes modifying the structure and the composition of household units (internal events). Children between 18 and 34 can leave their household unit of origin. Singles, living or not with their parents, can get married. The marriage event determines the creation of a new household unit. Widowed or divorced/separated individuals can get married, following the same rules applied to singles. Finally, the model simulates divorce for a share of married people, this event determining the split of the original household unit.

Figure 3 Demographic module - external events

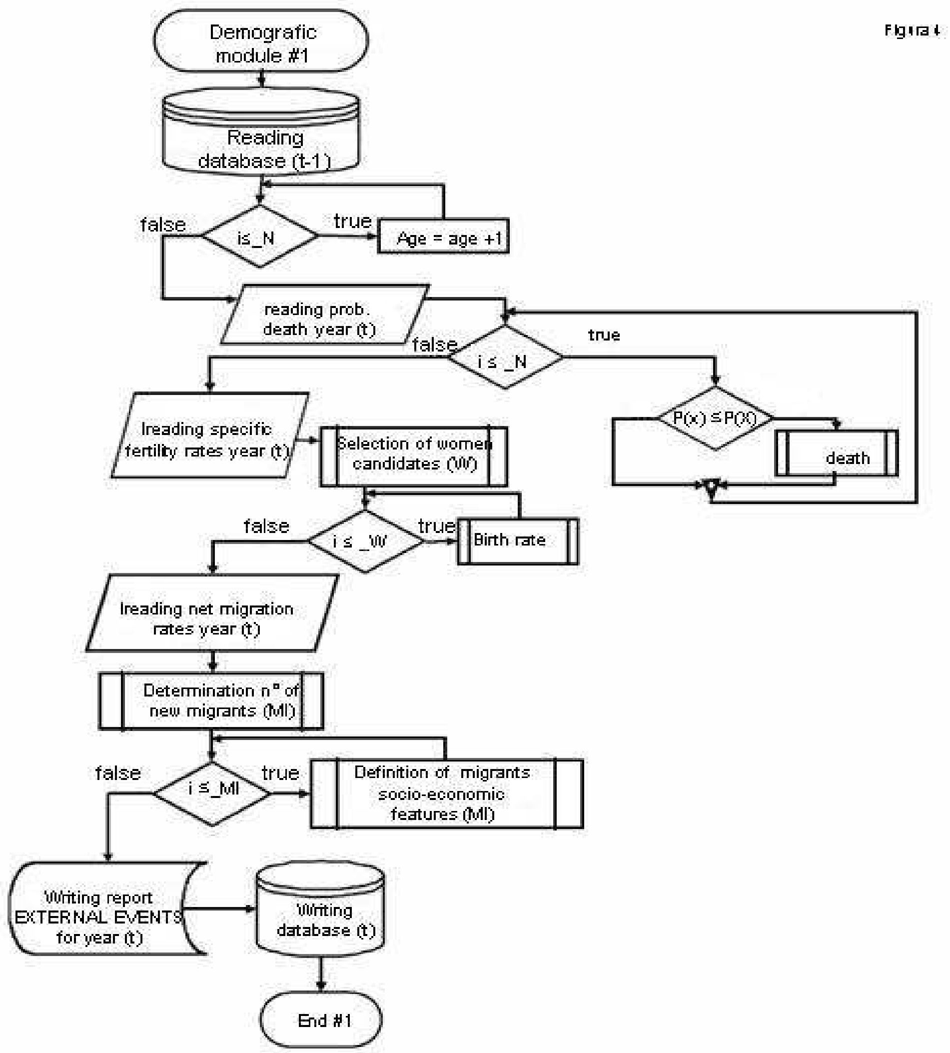
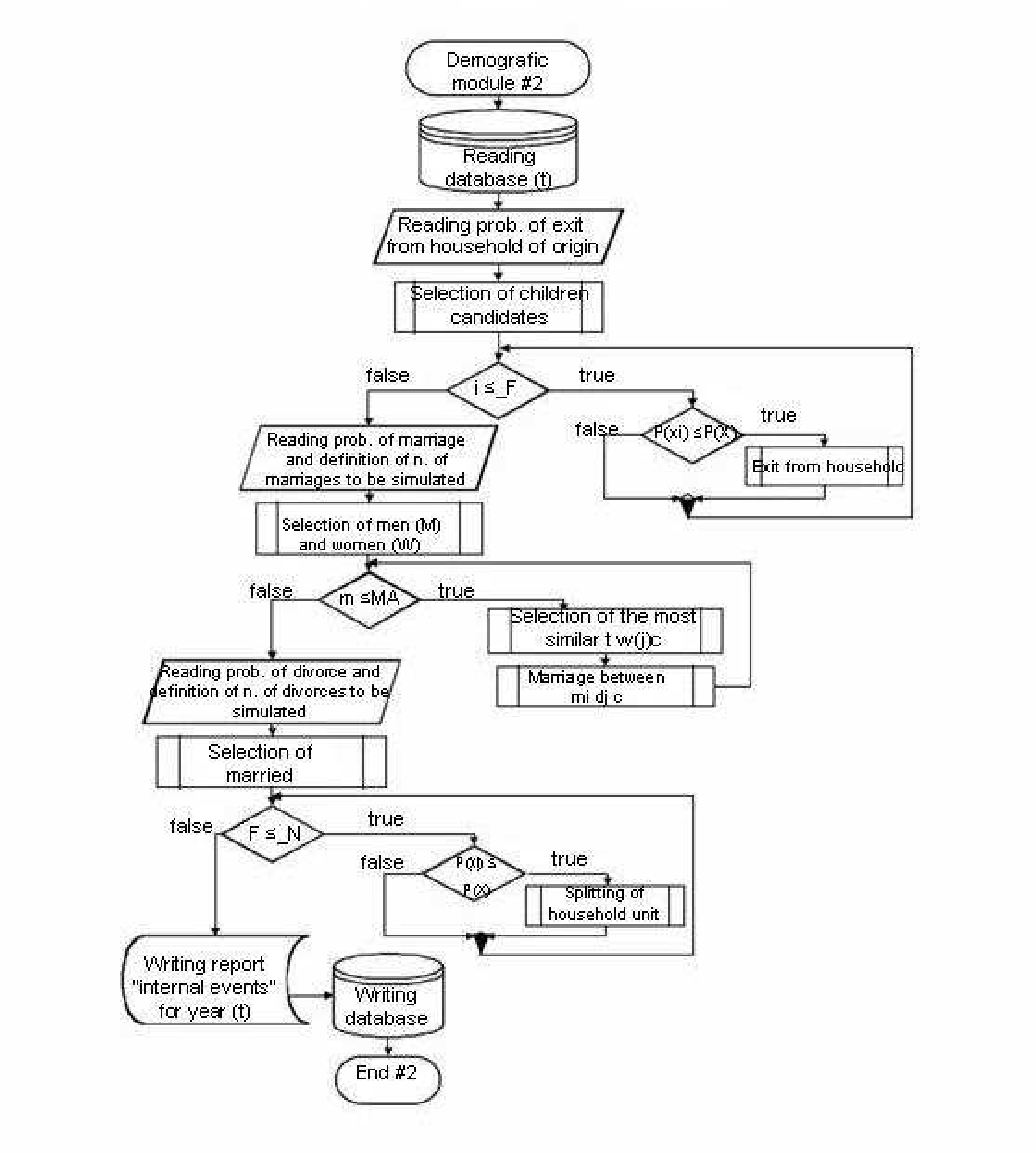
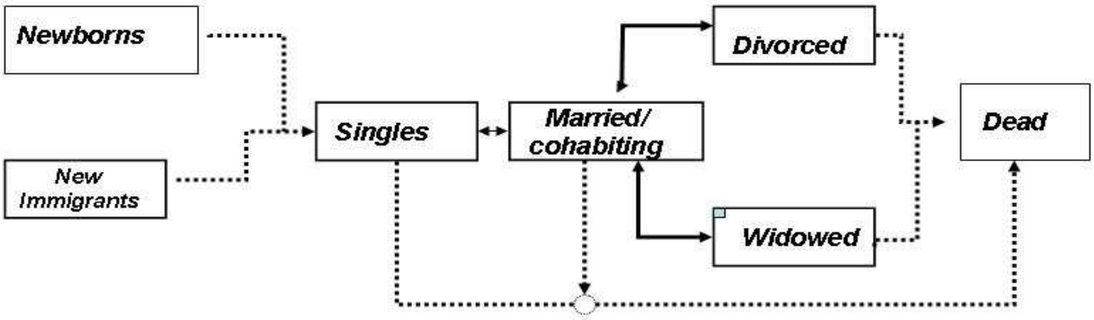


Figure 4 Demographic module - internal events



The model identifies four marital statuses (single, married/cohabiting, divorced, widower), allowing possible transitions across statuses according to the scheme showed in Figure 5.

Figure 5 Marital statuses transitions

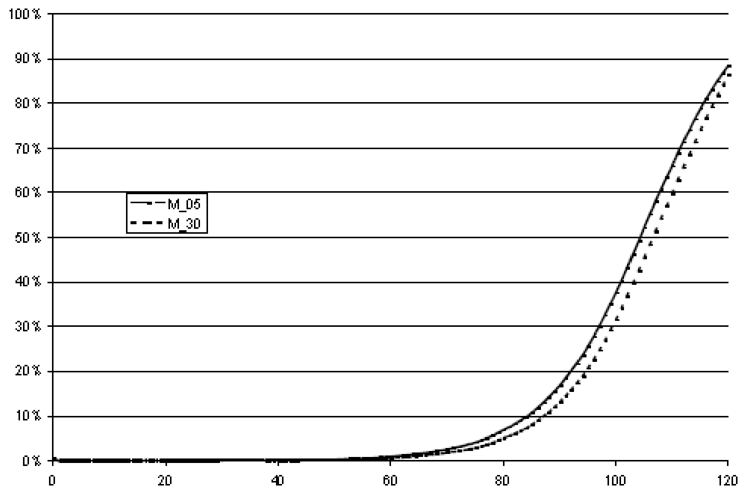


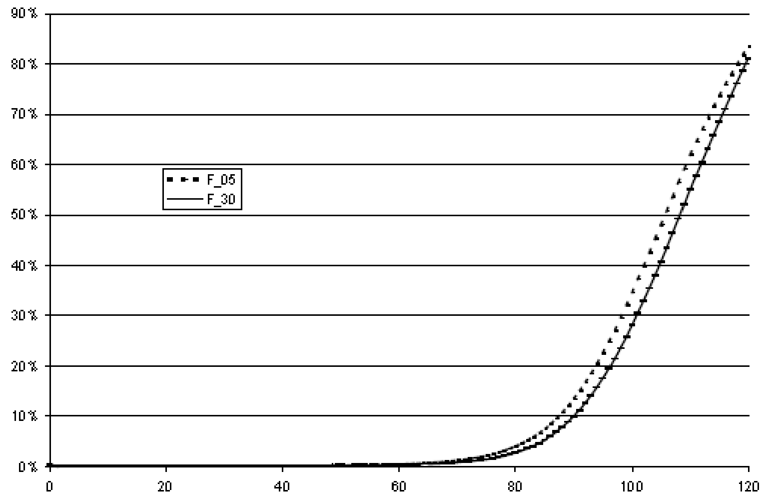
In the following, a detailed analysis of the main demographic sub-modules is presented.

4.2 The mortality module

The survival probabilities for the simulation are drawn from ISTAT official projections (1/2005). It is worth reminding that ISTAT employs an age-cohort approach for estimating death probabilities in order to allow for the recent procedure - widely used in all the developed countries - of a decreasing death probability across all ages and a substantial increase in old age survival probabilities, particularly for women.

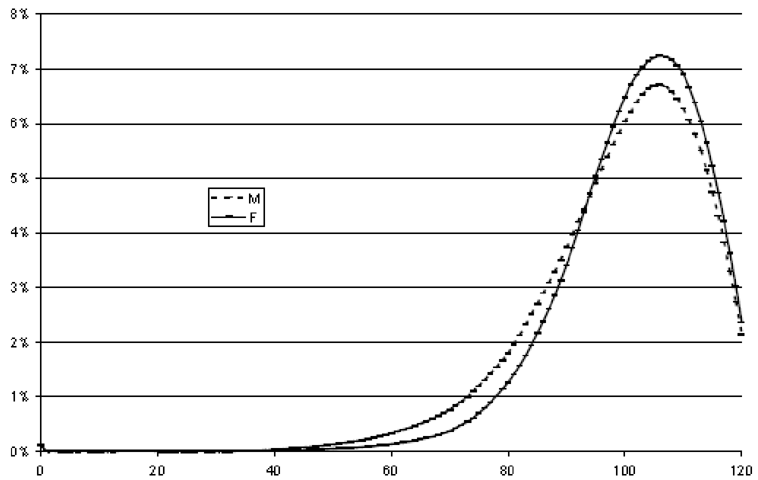
Figure 6 a and b Death probability by age and gender





Source: ISTAT, main projections at 1.1.2005. Central scenario.
Death probability on the right axis (national average).

Figure 7 Death probability variation by age and gender for year 2005 and 2030



Source: ISTAT, main projections at 1.1.2005. Central scenario.
Death probability on the right axis (national average).

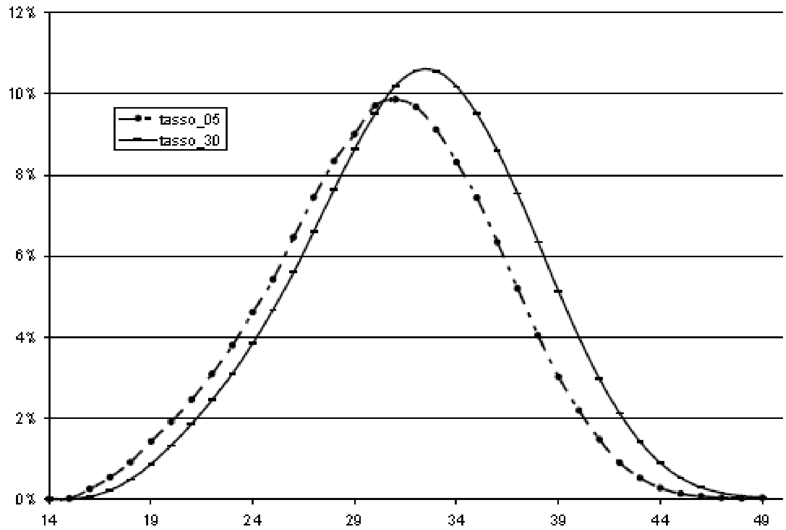
The technical working of the mortality module is the following: given the year of simulation, age and gender, a random number drawn from a uniform distribution [0,1] is attached to each observation. If the random value is smaller than the age-cohort specific ISTAT death probability, then the model simulates death and consequently modifies the cohabitant's marital status. Otherwise, in case the random number is greater, the model ages the observation by one year.

4.3 The fertility module

The annual flow of newborns is a function of the stock of women of child-bearing age (16-49) and of the ISTAT specific fertility rates. In this case too ISTAT adopts an age-cohort approach.

Figure 8 compares specific fertility rates in 2006 and 2030 according the ISTAT median scenario showing a mild increase of total fertility rates due to the higher specific rates for women over 31. On the opposite, fertility for women younger than 31 is expected to decrease probably due to the delay in marriages.

Figure 8 Specific fertility rates by mother age : 2005 and 2030 compared



Source: ISTAT forecasts, central scenario.

Once the flow of newborns by mother age class has been determined, the model selects women which are likely to have a child. Letting $f_a(c)$ be the probability distribution function for a married woman aged *a* with a number of previously born children equal to *c*, the probability of having a new child in year t+1 for a woman of child-bearing age is:

$$P(c_{t+1} = c_t + 1 | a_{t+1}, c_t) = (1 - F_{a(t+1)}(c_t))$$

where $F_{a(t+1)}(c_t)$ is the cumulative distribution function of $f_{a(t+1)}(c_t)$.

The procedure described above allows to allocate the flow of newborns by mother age, accounting for the number of children previously living in the household unit. Once the newborn is provided with the household *id*, the model determines her/his socio-demographic characteristics and updates the household unit size and composition. Gender is randomly assigned, probability to be male or female being the same.

4.4 The immigration module

The model simulates net migrations flows each year according to official forecasts provided by ISTAT[6]. Hence, the net expected migration flow for the next decades lies in the range of 145,000 to 150,000 individuals each year.

The entry age of immigrants is classified according to the legally registered immigrants distribution by age class as supplied by ISTAT. The model, however, excludes households' re-joining, assuming therefore, the immigrants being single at the entrance in the country. The flow of new immigrants is added to the stock of individuals (immigrants and natives) previously settled net of simulated deaths and births.

All the model modules are applied to the whole simulated population assuming immigrants' behavior being the same as natives'[7]. The imputation of socio-economics characteristics is carried out through the Monte Carlo method.

4.5 The exit from household unit

This sub-module allows the selection of children likely to leave the household of origin.

The increasing delay in household going-out by children is a well-established issue in Italy: according to ISTAT, 60.2% of children aged 18-34 lived in 2003 with at least one parent (table 5), both individuals' expectations and increasing troubles mainly related to economic conditions of the new generation being the main candidates to account for this phenomenon.

The recent estimations by ISTAT show an increase in the share of employed children living with parents, while a decrease is recorded for the share of ones looking for a fist job and 32.3% of children living with their parents is in education (ISTAT, 2004).

[6] In the official demographic forecasts international migrations are usually considered as less as important compared to fertility and mortality. In fact, forecasts on migrations are aleatory, mobility of populations being affected by (social, economic, psychological, political) factors which are hardly predictable(Blangiardo, 1997).

[7] This hypothesis could appear too strict for the simulation of some events, for instance fertility, while for others empirical evidences suggests less marked differences of behaviour. For instance, although working careers of immigrants are more mobile than the natives' ones, Anastasia Gambuzza Rasera 2005 finds on Giove 2004 archive behavioural pattern on labour market similar enough between immigrants and native workers after the first entrance into the labour market. The level of labour income is, on the opposite, *coeteris paribus* lower than natives' ones.

Table 4 Singles, aged 18-34 living with at least one parents

		1993			*1998*			*2003*	
Class	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
18-19	98,4	95,4	96,9	99,0	97,9	98,4	97,6	97,1	97,4
20-24	90,9	78,9	85,0	92,8	83,7	88,2	92,3	83,7	87,9
25-29	60,5	36,8	49,0	70,6	46,0	58,7	70,5	51,7	61,0
30-34	24,9	12,2	18,5	30,6	16,0	23,2	37,4	21,4	29,5
Total	64,0	48,9	56,5	66,2	51,1	58,7	66,8	53,6	60,2

Source: ISTAT (2004) "Indagine Multiscopo sulle famiglie:
Aspetti della Vita quotidiana; Famiglia, Soggetti Sociali 2003".
Mean value for 1993-1994, 1998 and 2003 for 100 young in the same age class.

In the lack of forecasts about future trends concerning this phenomenon, the model uses ex-post probabilities drawn from table 5 in order to establish a steady state exit rule: the selection of yearly flow of children likely to leave the household is carried out through a Monte Carlo process employing transition probabilities conditional on gender and age class equal to 1- the probabilities in column 8-9 in table 4.

4.6 Marriage module

The model allows singles to get married each year, and the simulation of this event consists of three steps: first, the flow of yearly marriages is defined as 4.3‰ of total population[8]. Once the number of marriages is known, potential candidates (aged 16-60) are selected through a Monte Carlo process relying on probabilities of marriage conditional on gender and age provided by ISTAT multiscope survey "Famiglie e soggetti sociali" (ISTAT, 2004)[9]. Candidates are then inserted in two distinct databases by gender in order to allow the birth of new household units.

Literature points out the presence in Italy of positive assortative mating in marriages (Becker, 1991), according to which spouses select themselves in a non random way, being similar in terms of

[8] http://www.istat.it/salastampa/comunicati/non_calendario/20060424_00/indicatori_demografici.pdf.
The steady state hypothesis does not appear in this framework particularly strict: in fact, the marriage rate has not substantially modified in the last year.
[9] ISTAT does not publish marriage probability by age and gender but reports the number of individuals getting married each year only. Starting from this information, cohort and periods effects apart, we obtain yearly marriage rates dividing the number of individual getting married by age and gender for the total number of marriages each year.

education (Rossetti, Tanda, 2000) and employment status (Del Boca et al., 2000). In addition, Borlini and Zajczyk (2001) find a high probability of matching between individuals coming from the same geographical area, attaining the same education level and the same professional condition.

Marriage age is generally lower for women than for man and the probability to join a man in a certain age class is assigned to all the select women in order to allow for this gap[10]. Therefore, the Monte Carlo technique conditionally on wife's age generates a variable containing the age class of a potential husband.

The simulation of a marriage is then carried out and the matching procedure allows the matching of similar spouses according to a vector of observable features including dummies on education, marital status (single, divorced and widowed), geographical area, and age class in line with a propensity score method of Rosembaum and Rabin (1983), Holland (1986), Rubin and Thomas (2000). Each new household unit (including children from previous relationship) is provided with a HID (Household IDentification number), which remains unvaried for the whole simulation period.

4.7 The divorce module

Married couples are allowed to divorce with the following splitting up of the household into two different units headed by the two divorced individuals. As for the marriage module, the divorce simulation is carried out through three steps. Firstly, the yearly flow of divorces is defined as 3‰ of the total number of married couple (ISTAT, 2003)[11]; secondly, couples which are likely to divorce are selected: as ISTAT finds a different incidence of divorce events both at geographical level and according to age, the selection is carried out through a Monte Carlo process relying on ISTAT probabilities conditional on geographical area and wife's age class. Within this group, a number of couples amounting to the yearly flow of divorces to be simulated is randomly the selected; the splitting up of the household in two different units and the updating of marital status and household composition variables are then carried out[12].

[10] Probabilities are computed on ISTAT data considering the distribution of women marriage age as a function of spouse's age class on the total number of marriages each year. The mean gap of age between men and women is about three years.
[11] The steady state hypothesis used in the divorce simulation appears stricter compared to the case of marriages as statistics on this topic suggest a growing propensity to divorce in the last years.
[12] Eventual children will belong to mother's household unit. According to ISTAT in 85% of cases underage children will be fostered to the mother.

4.8 The disability module

The simulation of the disability condition is based on external information taken from the ISTAT Survey on public health and the use of the national health services, which is carried out every five years on a sample of more than 100,000 individuals of all ages. The most recent survey, which is the one used for the purposes of this paper, was conducted in 2005. The survey collects information about individuals' ability to perform certain basic daily tasks such as washing, eating and dressing, without the need for the help of others. There are 19 questions of this type, and they may be grouped into four categories, each of which may indicate a different form of disability, namely: being unable to get out of the house; having serious difficulties with movements, everyday activities, or in communicating with others[13]. For each of the four categories, therefore, we end up with a dummy variable which is given the value 1 if the individual is unable to perform that set of activities. This classification has been used to distinguish three levels of disability, each of which depends on how many of these dummy variables takes the value of 1: the lowest disability condition (level 1) is that where the person is disabled in terms of only one of the four groups of variables; medium (level 2) disability corresponds to two dummy variables equal to 1; finally, a person is deemed severely disable (level 3) if three or four areas of disability take a value of 1. Table 5 provides some basic descriptive statistics regarding the survey.

Table 5 Descriptive statistics of the Survey on health conditions and use of health services

	Whole sample	Disabled, level 1	Disabled, level 2	Disabled, level 3
Age	41.8	68.1	75.5	78.8
Woman	51.4%	63.0%	68.4%	70.0%
Compulsory education	59.7%	88.6%	89.5%	93.5%
High-school diploma	26.4%	8.8%	7.7%	4.4%
University degree	8.2%	2.6%	2.7%	2.1%
North	45.2%	41.5%	38.8%	39.2%
Centre	19.2%	18.8%	20.1%	21.8%
South	35.6%	39.6%	41.1%	38.9%
Widow	7.9%	36.9%	45.6%	53.7%
Remaining life expectancy (in years)	40.9	19.2	13.5	11.3
Number of observations	128040	2797	1869	1324

[13] For example, a person is defined as unable to perform basic everyday activities if he/she indicates a serious difficulty as a reply to at least one of the questions that fall within this category.

Average age increases with the seriousness of the disability condition, as does the proportion of women. The level of education is negatively correlated with the level of disability as does the condition of widowhood.

In order to assign to each individual in the simulation database a disability status, we propose and compare two alternative approaches:

a) *Pure ageing*: the ISTAT Health Survey is used to compute the proportion of disabled people within classes defined by gender and age (Costello and Przywara, 2007). These relative frequencies by gender and age are used to select, following Monte Carlo methods, which sample members are attributed the disability status. Three levels of disability have been identified. Note that under this scenario no cohort effect is taken explicitly into account, nor it is assumed that any future gains in life expectancy will be spent in a state of bad health. As a consequence, these projections are rather mechanical and we incur in the risk of producing distorted estimations of the number of disabled.

b) *Compression of disability*: the probability of being disabled is not constant within groups of the same age and gender, but depends on a vector of socio-demographic determinants. If these variables change, the probability of suffering from a disability should change accordingly. In order to take account of this endogeneity, we have performed an ordered probit estimation on the 2005 Health survey, where the dependent variable may be classified at four different levels: no disability (95.7% of total sample), low disability (2.1%), average disability (1.3%), severe disability (0.9%). The explanatory variables must be restricted to those socio-demographic characteristics that are common to both the Health Survey and the microsimulation model database, namely: age, gender, educational level, geographic area, widowhood. In addition to these explanatory variables, we have also included the residual life expectancy (in years) of each person, such data (depending on age and gender) being taken from the latest ISTAT estimates. The introduction of residual life expectancy is important, since if overall life expectancy rises, one would not expect the probability of becoming disabled to remain constant for any given age. Indeed, it is now widely recognised that this probability increases rapidly during the last years of one’s life. In the presence of an ageing population, the omission of residual life expectancy from the regression would, at the simulation stage, result in an overestimation of the probability of becoming disabled, and therefore also of future LTC costs (Norton and Stearns, 2004). This second hypothesis may be considered to be a variant of the diverse theories asserting that the number of years spent in poor health should decrease as life expectancy increases (Manton et al. 2006). It is, nevertheless, more accurate and consistent with the data used to build the model than the mere application of a simple ad hoc rule whereby the probability of being disabled increases with life expectancy.

In order to check the results of the application of this rule, we also create a comparison scenario with a very simple rule whereby the probability of becoming disabled changes each year in proportion to the increase in life expectancy. Costello and Przywara (2007) refer to this rule as the “constant health scenario”.

The effect of observable socio-demographic characteristics on the probability to fall under a condition of disability is modelled in terms of an ordered probit model, which takes the following form. Define an ordinal variable y {i: 1 . . .,N} indicating the observed level of disability among the sample members and y_i^* is the associated latent variable. The model has the following general structure:

$$y_i^* = X_i \beta$$
$$y_i = j \quad \text{if} \quad c_{j-1} < y_i^* \leq c_j$$

where X_i denotes the vector of observable explanatory variables; β is a vector of coefficients, and ε is a random variable distributed as a normal. Given the nature of the data available, we ignore the possibility of unobservable personal characteristics which might influence both the level of disability and some of the explanatory variables. There are four different disability levels, denoted by j: 0 no disability, 1 low level of disability, 2 intermediate level of disability, 3 serious level of disability. The cut-off parameters c are estimated as part of the model. (A constant term is not identified in the model).

Table 6 shows the results of the ordered probit estimate on the 2005 Health Survey. The expalanatory variables relating to age are introduced using a spline function, and their coefficients show a marked increase in the probability of becoming disabled over the age of 70. Disability status is strongly dependent on the level of education (the omitted variable is the graduate level), and also on being resident in the southern part of Italy (the omitted geographic area). Residual life expectancy has a significant effect: if, in the future, life expectancy increases, this will lead to a reduction in the probability of becoming disabled for each year of age.

Table 6 Ordered probit estimates of the probability of being disabled

	Coef.	Robust Std. Err.
<=30 years	-0.0551	0.0100
31-50 years	-0.0400	0.0097
51-60 years	-0.0255	0.0104
61-70 years	-0.0012	0.0089
71-80 years	0.0463	0.0077
>=81 years	0.0604	0.0054
Female (D)	0.3137	0.0424
Compulsory education (D)	0.4726	0.0409
High-school diploma (D)	0.1472	0.0476
Northern Italy (D)	-0.2386	0.0188
Central Italy (D)	-0.1518	0.0234
Widow (D)	0.0997	0.0234
Residual life expectancy (in years)	-0.0541	0.0099
Cut-points:		
C_1	-1.5867	0.7821
C_2	-1.1616	0.7821
C_3	-0.6392	0.7819

Number of obs = 128040; LR $chi^2(13)$ = 15988.04; Prob > chi^2 = 0.0000; Log likelihood = -21515.021; Pseudo R^2 = 0.2709. (D) indicates dummy variables.

Since CAPP_DYN projects all the model predictors, we are able to use the estimated coefficients and the cut-off parameters of this regression for predicting, for each year, the probability for an individual with characteristics X of being in a condition of disability j as:[14]

$$pr(y_i^* = 0) = \int_{c_0}^{c_1} y_i^* dy = Norm[(c_1 - (X_i\beta + \varepsilon_i)]$$

$$pr(y_i^* = 1) = \int_{c_1}^{c_2} y_i^* dy = Norm[(c_2 - (X_i\beta + \varepsilon_i)] - pr(y_i^* = 0)$$

$$pr(y_i^* = 2) = \int_{c_2}^{c_3} y_i^* dy = Norm[(c_3 - (X_i\beta + \varepsilon_i)] - pr(y_i^* = 1)$$

$$pr(y_i^* = 3) = \int_{c_3}^{1} y_i^* dy = Norm[(X_i\beta + \varepsilon_i) - c_3]$$

[14] We assume that the gradient of the disability rates (by levels) with respect to the socio-economic characteristics observed in the cross-section (year 2005) will remain constant in the future.

In order to identify the disability level for the sample members in each year of the simulation, we use a Monte Carlo process. We assign no disability to sample members who receive a random number z, drawn from a uniform distribution between 0 and 1 below the conditional probability of having no disabilities $pr(y_i^* = 1)$; we assign low level of disability if the random number are between $pr(y_i^* = 1)$ and $[pr(y_i^* = 1) + pr(y_i^* = 2)]$; we assign intermediate level if $[pr(y_i^* = 0) + pr(y_i^* = 1)] \le z \le [pr(y_i^* = 0) + pr(y_i^* = 1) + pr(y_i^* = 2)]$; finally, if z is between $[pr(y_i^* = 0) + pr(y_i^* = 1) + pr(y_i^* = 2)]$ and 1 the individual is assigned the most serious level of disability.

We assume also that if a person is deemed to be disabled in year t, he/she cannot then return to being classified as non-disabled in future years of his/her life; however, if that person is deemed to be less than seriously disabled, then he/she may be attributed a worse degree of disability in any subsequent year, up until death. We then randomly select (for each of the two alternative imputation approaches described above), from among those classified as being severely disabled, and who had been disabled for more than three years, a subsample of individuals to be taken into nursing homes, this number corresponding to official estimates of the number of people recovered in such homes in Italy (ISTAT, 2007b).

Table 7 shows the association between socio-demographic characteristics of the population and the level of disability predicted by the probit model according to the method used in Ermish and Francesconi (2001). The predicted probabilities are computed at the sample values using estimated parameters (and cut point) from model presented in table 2. The results can be read as follow. The row “baseline probabilities” displays the predicted average probabilities for each of the four levels of disability when all the characteristics are set at their sample values for each person. It can be shown that the overall baseline predicted probability of being in one of the four states is equal to 1, while the sum of the predicted probabilities to have one of the three levels of disability is equal to the rate of disability presented in the text (sum of the baseline probability of having low level of disability=2.1%, intermediate level=1.3%, serious level=0.9%). The remaining rows of table 3 show predicted probabilities relative to particular values of the explanatory variables. In the case of age, for example, all the characteristics other than age are set at their sample values for each person, and predicted probability values for each person are averaged over the sample.

Women have a disability probability that is always higher than that for males. Individuals with a low level of education present a probability of having a low level of disability of 2.3%, compared with 1.4% for those with a high school diploma, while a higher level of education decreases this probability further. Living in the South of Italy increases the probability of disability compared with the probabilities computed for people who live in the Centre and the North.

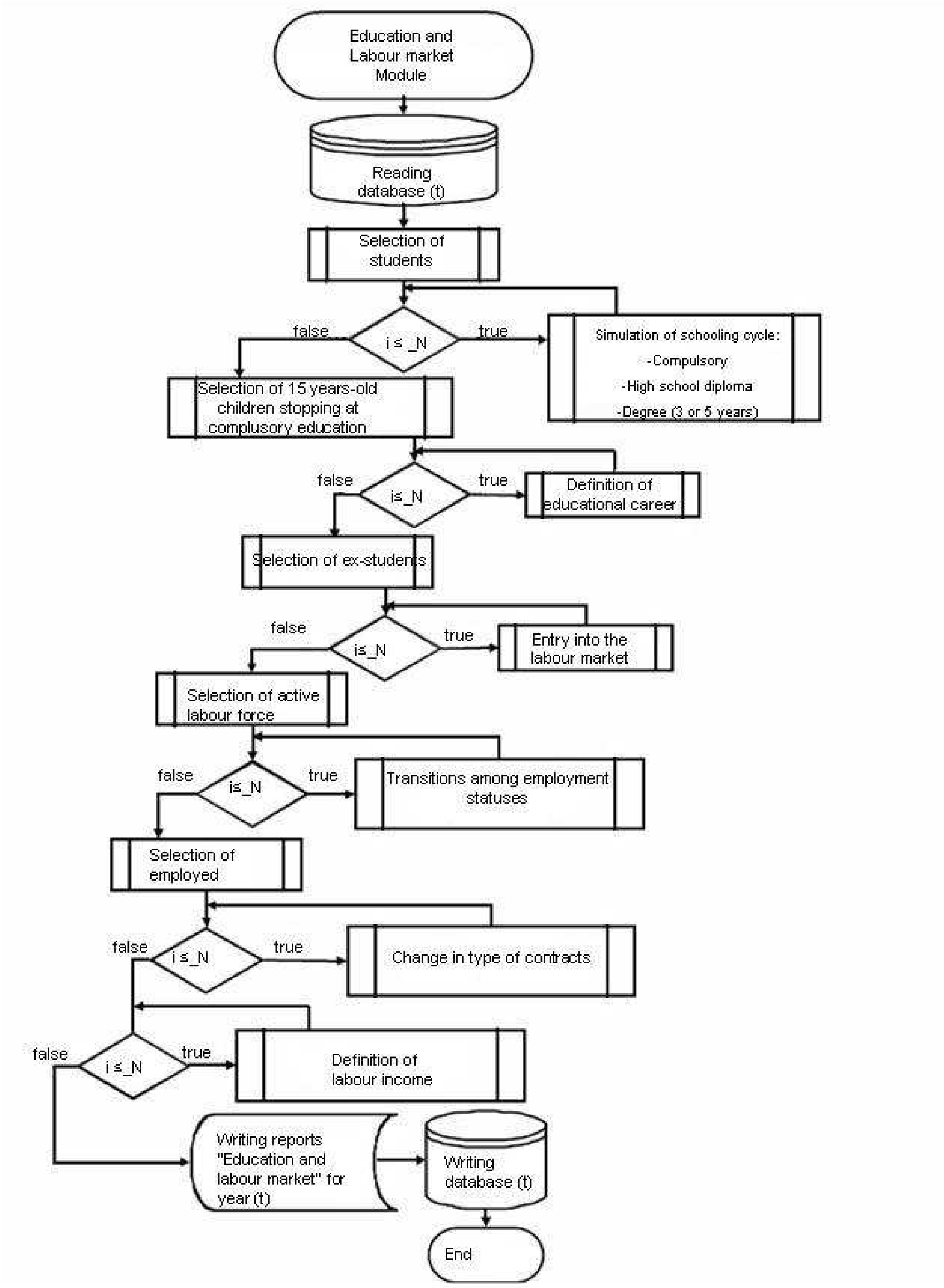
Table 7 Predicted probabilities of being disabled by level of gravity

	Not disabled	Low level	Intermediate level	Worst level
Baseline probabilities	0.957	0.021	0.013	0.009
55 years	0.970	0.016	0.009	0.004
65 years	0.976	0.013	0.007	0.003
75 years	0.966	0.018	0.011	0.006
85 years	0.925	0.034	0.024	0.017
Female	0.947	0.026	0.016	0.011
Male	0.967	0.017	0.010	0.006
Compulsory education	0.953	0.023	0.014	0.010
High school diploma	0.972	0.014	0.009	0.005
Degree	0.978	0.012	0.007	0.004
North	0.963	0.018	0.011	0.008
Center	0.958	0.020	0.013	0.009
South	0.948	0.025	0.016	0.012
Widow	0.952	0.023	0.015	0.010
Living with spouse	0.959	0.020	0.013	0.008

4.9 Education and labour market

After the demographic, a further module concerns education, entry and transitions in the labour market and the determination of earnings. The structure of this module is presented in figure 10. Firstly all individuals aged 16 are awarded the compulsory education level. A higher educational level delays the entry into the labour market up to the achievement of imputed educational level (high school certificate, three year degree and five year degree). The end of schooling is followed by the entry into the labour market. The in and out flows into/from the labour force and employment transitions are then simulated. Stock of active population is divided in two sub-groups: public and private employees and self employed (part time-full time). A share of population is employed with atypical and fixed term contracts. Finally, the model simulates labour income and updates the contributory history.

Figure 9 “Education and labour market” module structure



The steps illustrated in figure 9 are organized in the following sub-modules:

- education,
- entry into the labour market,
- employment transitions,

- transitions among contractual statuses,
- labour income

which will be analyzed in the following.

4.10 The education module

In this module three educational levels are accounted for:

1) compulsory
2) upper secondary school
3) tertiary school (three or five years degree)

All individuals aged 16 are awarded the compulsory educational level; then individuals can decide whether keep going with education or entry into the labour market.

Educational attainments are simulated by imputing coefficients obtained by an ordered probit estimation whose results are reported in table 6. The sample includes surveyed people aged above 16, which have completed their educational career or are enrolled at the university[15], included in the ISFol PLUS survey 2004[16]. The sample contains 34,324 individuals.

The empirical model is structured as follows: we define y_i the observed achieved educational level and y_i^* the corresponding latent variable. The alternatives have an ordinal form which implies the following general structure:

$$y_i^* = X_i\beta + \varepsilon_i$$
$$y_i = j \quad se \quad c_{j-1} < y_i^* \leq c_j$$

[15] Following Checchi, Flabbi (2005) students enrolled at university are supposed to end up their educational career getting the degree.

[16] A problem for the empirical analysis in the determinants educational careers in Italy comes from the lack of suitable statistical sources for dynamic estimation. Estimations have been conducted on different sample surveys. Several cross-section data does not allow the extrapolation of cohort and period effects, therefore which are less explored in the empirical analyses implemented in Italy so far. The polled cross-section of ISTAT work forces allows the analysis of cohort effects in the schooling rates dynamic (Leonbruni, Richiardi 2006) but it does not allow the conditioning of schooling choices to the household of origin characteristics. A pooling of SHIW surveys allows the joint analysis of the two effects. But for youngsters living with their parents only, implying possible estimation distortions due to “selection effect” (Heckman, 1979) the recent survey ISFOL PLUS on a sample of more than 40,300 individuals aged between 15 and 64 (ISFOL, 2006) points out through telephonic interviews detailed information of respondents’ schooling level and several interesting information on household social economic conditions. The information provided by this survey is therefore more suitable 2.2.compared to others to the estimation of educational choice determinants allowing for social and cultural, as well as normative, changes happened in Italy in the last years.

where X_i is the vector of individual (gender, geographical area and cohort dummies) and household (parents' presence and educational level[17]) characteristics; the parameters c_j, representing the thresholds, are estimated jointly to the column vector of β coefficients.

Regression results are reported in table 6: first column shows estimated coefficients, while columns 2-4 display marginal effects for every single value of the dependent variable. As it can be noticed from the regression results educational attainments appear to be strongly dependent from parents' educational level and geographical area. Women display higher probability to achieve a higher level of education: the chance to enroll to tertiary education is *coeteris paribus* 1.5% higher for women than for man and the probability to stop to compulsory schooling is 2.9% lower for women than for men. Cohort dummies suggest a positive trend in schooling for younger generations. *Coeteris paribus* an individual born after 1979 has a higher probability (+4.4%) to achieve a university degree than an individual born in the period 1971-75. On the opposite the same individual is 7.9% less likely to stop to lower secondary education.

In the dynamic simulation, coefficients in table 6 will be employed to impute educational choices. In practice, for each observation at the end of the compulsory school, the probability the dependent variable assumes value 1, 2 or 3 are calculated[18]. In other terms:

$$pr(y_i^* = 1) = \int_{c_0}^{c_1} y_i^* dy = Norm[(c_1 - (X_i\beta + \varepsilon_i)]$$

$$pr(y_i^* = 2) = \int_{c_1}^{c_2} y_i^* dy = Norm[(c_2 - (X_i\beta + \varepsilon_i)] - pr(y_i^* = 1)$$

$$pr(y_i^* = 3) = \int_{c_2}^{1} y_i^* dy = Norm[(X_i\beta + \varepsilon_i) - c_2]$$

[17] Presence of parents refers to the year interviewed was 15.

[18] The same procedure is applied to the students over 15 of the base year, in order to define the human capital level. The imputed value is constant all over the simulation.

Table 8 Ordered probit of educational level

Education(y)	coefficient	y=Pr(j==1) 0.4214	y=Pr(j==2) 0.4589	y=Pr(j==3) 0.1196
compulsory_mother***	-.8062	.2803	-.0655	-.2147
	(.0702)	(.0202)	(.0042)	(.0231)
High School_mother***	-.3054	.1209	-.0680	-.0529
	(.0689)	(.0273)	(.0171)	(.0102)
Compulsory_father***	-1.328	.4176	-.0312	-.3864
	(.0554)	(.0121)	(.0090)	(.0196)
High School father***	-.5229	.2062	-.1235	-.0826
	(.0550)	(.0212)	(.0146)	(.0068)
No_mother***	-.3204	.1270	-.0734	-.0536
	(.0581)	(.0230)	(.0151)	(.0079)
No_father***	-.1977	.0782	-.0425	-.0356
	(.0437)	(.0174)	(.0103)	(.0071)
women***	.0755	-.0295	.0144	.0151
	(.0211)	(.0082)	(.004)	(.0041)
Centre***	.1987	-.0765	.0339	.0425
	(.0287)	(.0108)	(.0043)	(.0066)
South***	.0961	-.0374	.0179	.0195
	(.0232)	(.0090)	(.0042)	(.004)
Co_min_1950***	-.4847	.1914	-.1123	-.0790
	(.0404)	(.0155)	(.0104)	(.0055)
Co_1951_1960***	-.2435	.0961	-.0515	-.0445
	(.0428)	(.0169)	(.0098)	(.0071)
Co_1961_1965**	-.1265	.0498	-.0259	-.0239
	(.0483)	(.0191)	(.0105)	(.0086)
Co_1966_1970**	-.1279	.0503	-.0260	-.0242
	(.0457)	(.0181)	(.0099)	(.0082)
Co_1976_1978***	.1895	-.0726	.0311	.0414
	(.0419)	(.0157)	(.0058)	(.0099)
Co_1979_plus***	.2047	-.0786	.0341	.0444
	(.0387)	(.0145)	(.0056)	(.0090)
_cut1	-21.282			
	(.0777)			
_cut2	-7530			
	(.0767)			
N	34323			
R:squared	0,935417			

In a second while, each observation aged at least 15 is assigned a random number (z) drawn from a uniform distribution with support [0,1] which will be compared with the probabilities estimated with the previous formula. If the number is lower than $pr(y_i^*=1)$, the i_{th} individual has the basic educational level; if z of the i_{th} is between $pr(y_i^*=1)$ and $[pr(y_i^*=1)+ pr(y_i^*=2)]$ the corresponding educational attainment is upper secondary school. If z is higher than $[pr(y_i^*=1)+ pr(y_i^*=2)]$ tertiary education is assigned[19].

4.11 Entry and transitions in the labour market

Missing information about employment status, occupational attainments, and type of activity from whom students in previous year are assigned using conditional probabilities drawn from transition matrices built on the sub-group of individuals moving from education to labour market (*Rilevazione trimestrale sulle forze di lavoro ISTAT 2001-2002 – RTFL* from now on).

Occupational attainments and sector are assumed to be time-invariant over the whole simulation period for each individual, while employment status and contractual arrangement are allowed to change along time.

Transitions among employment are simulated in the sub-modules "transitions" while contractual transitions (typical, atypical, permanent, fixed term) are determined in the sub-module "work contract".

Concerning the number of transition in the labour market, CAPP_DYN allows four employment statuses and, and in line with other dynamic MSMs, assumes employment decisions depend on individual characteristics only while being independent of demand side factors. Transitional probabilities are estimated on RTFL 1993-2003. Individuals aged 16 to 64, excluding pensioners and students, can be classified according to following statuses:

- full time worker (at least 31 working hours)
- part-time worker (less than 31 working hours)
- unemployed
- not in work force (unemployed individuals not looking for a job)

In table 7, ex-post transition probabilities are estimated on the selected sample which is composed of 68.8% of workers, 9.24% of unemployed and 22.70% of individuals out of the labour force. Each cell shows the share of individuals who starting from status *i* are in status *j* in the following year.

Results suggest a strong immobility among statuses as shown by the main diagonal figures. The higher stability is found for full-time workers, while the part-time workers appear more mobile:

[19] Since detailed information on university career is not available we assume 30% of enrolled students terminate at three-year degree, while the remaining 70% will achieve the five-year degree.

28.4% of them moves toward a full-time position in the following year, 5.55% becomes unemployed and 4.64% gets out from the labour force.

More than 23% of unemployed finds a job after one year (19.41% full-time) while 7.8% exits the labour market.

The 3.69% of inactive individuals enters the labour market among employed, while 2.70% searches for a job[20].

Table 9 Transition probabilities among employment statuses from year t to year t+1

Initial status	Employed_FT	Employed_PT	Unemployed	Non_active
Employed_FT	95,16	1,89	1,95	0,99
Employed_PT	28,41	61,4	5,55	4,64
Unemployed	19,41	3,92	69,18	7,48
Non_active	2,58	1,11	2,7	93,61
Mean	63,44	4,62	9,24	22,7

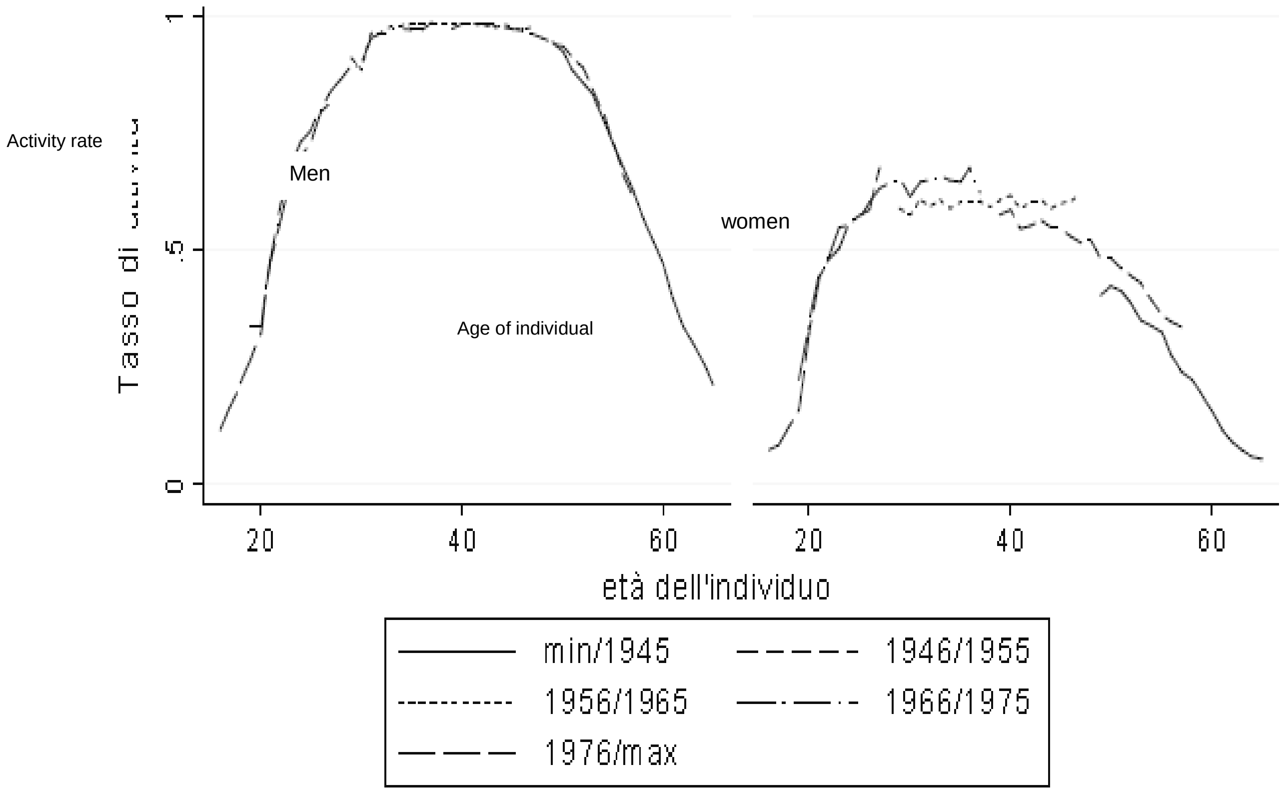
Figures in table 9 can be interpreted as mean trend indicators for the whole sample in the ten years period considered. It is however reasonable assuming transition probabilities depend on individual characteristics (gender, age, education, geographical area).

In addition, it has to be remarked that the considered period is characterized by a strong and prolonged increase of employment due to a greater female participation to the labour market and to a greater diffusion of part-time work with different features depending on birth cohorts.

Graphs in figure 10 show some results by cohort, gender and geographical area. The evidence is mainly in line with expectations and empirical evidences from other works[21]: while participation rates are approximately unvarying among men cohorts, among women a general increase in participation is recorded for observations aged 30 to 50, rates among the youngest women being almost constant suggesting a greater share of them choosing educational paths which delays the access to the labour market. Moreover, a significant share of part-time workers is recorded among youngest cohorts.

[20] Probability refers to mobility among employment statuses after 1 year. Of course, it is possible that during the year an individual experienced more than one transition. In these cases the transition between the initial and the final status is considered.
[21] See Trivellato et al. (2005), Leonbruni, Richiardi (2005), ISTAT (2004).

Figure 10 Prticipation rates to labour market 1993-2003 by age, gender and cohort.



Transition probabilities are estimated on a pseudo-panel RTFL 1993-2003 by using a multinomial logit. The dependent variable is the employment status in the final year[22]. Estimations have been carried out for each initial status and gender (four initial statuses *per* gender, therefore 8 models have been estimated).

Given the initial status of h_{th} individual, conditional probability of transition or immobility in the following year (J=j) can be represented as[23]:

$$P(J = j \mid I, X_h) = \frac{\exp(X_h \beta_j)}{1 + \sum_{j=1}^{4} \exp(X_h \beta_j)}, j = 1, \ldots, 4$$

[22] ISTAT data are not free from *caveat*. Trivellato et al. (2005) suggest not using too detailed classifications. Not with standing the well known limits, the transition matrices analysis on RTFL allows the use of wide data set which is representative of Italian population.

[23] The multinomial logit model is valid under some condition. The most important, known as independence of irrelevant alternatives (IIA) imposes errors ε_{ij} are independents of j, i.e. the odds-ratios are assumed to be constant between two alternatives, even if number of alternatives increases.

where *j* is one of the four feasible statuses, Xh is the covariates vector (education, second order polynomial in age, geographical area, marital status, activity and eight cohort dummies) and β_j is the vector of coefficients varying according to each state.

In table 10, 24 coefficients for both women and men are reported, while in column 12 transition typologies are presented[24]. A positive coefficient, *coeteris paribus*, increases the transition probability from the initial to the final status. For instance, the negative coefficient of the "public" variable for full-time women toward all the other statuses suggests a smaller chance of transition from a full time job to all the other conditions in the public sector; the negative coefficient for women in the Centre and in the South moving from unemployment or non-activity to employment suggests a higher probability to find a job in the North compared to the Centre and the South.

As Chies et al. (1998) pointed out, this table can be read by row and by column: by row it provides information whether and how a variable affects transition frequency, by column it displays the features of individuals which are more/less likely to move from a status to another.

The exit from unemployment or non-activity is more frequent in the North compared to Centre and South for both men and women. The education positively affects the probability of getting and keeping a job. The higher the age, the lower is the probability of a transition from employment to unemployment for both women and men, although the same variable has a negative effect for unemployed or inactive individuals willing to enter into the labour market. As expected, public employees have more stable careers compared to those in private sector. The share of younger which loses a job is declining with age cohort. An increasing trend in the transition toward a part time employment is recorded for young full time women compared to older cohorts, and this effect is stronger for married women, probably due to children[25].

The transition module uses coefficients in table 8 to compute the individual odds-ratios of employment transitions. Finally, a Monte Carlo process allows the simulation of mobility for each active individual[26].

[24] Coefficients are computed with reference to the inertia condition i.e. when the status is unchaged from period t to period t+1 (i = j).

[25] Presence of children in the household is not controlled by the model.

[26] Professional qualification (blue collar, white collar, self-employed, manager) never changes in life. This is a simplification widely employed.

Table 10 Multinomial logit for transitions in the labour market

	Employed_FT-> employed_PT	Employed_FT-> Unemployed	Employed_FT-> Inactive	Employed_PT-> Employed_FT	Employed_PT-> Unemployed	Employed_PT-> Inactive
Compulsory	0.273***	0.746***	1.109***	-0.482***	0,15	0.657**
High_school	-0.03	0.19	0.253*	-0.418***	-0.025	0,158
Age	0.069**	-0.04	-0.125***	-0.082**	-0.206***	-0.171***
Sq_age	-0.001**	0.00	0.001***	0.001	0.002	0.002**
centre	-0.125*	0.197**	0.13	0.066	0.264*	-0.151
South	-0.345***	0.965***	0.898***	0.441***	1.329***	0.810***
public	-0.200***	-0.556***	-0.652***	-0.016	0,103	0.024
Married	0.540***	-0.432***	1.035***	-0.170**	-0.540***	0.611***
C_48_52	-0.19	-0.346*	-0.321**	-0.363**	0,502	-0.496**
C_53_57	-0.16	-0.692***	-0.328*	-0.472***	0,162	-0.803***
C_58_62	0.08	-0.671**	-0.18	-0.733***	0,138	-1.140***
C_63_67	0.370*	-0.731**	-0.01	-0.898***	-0.227	-0.980***
C_68_72	0.397*	-0.949***	-0.18	-1.020***	-0.482	-0.879**
C_73_77	0.499*	-0.968**	-0.31	-1.133***	-0.923	-1.793***
c_78_max	0.526*	-1.038**	-0.15	-0.825**	-1.311*	-2.005***
_cons	-5.022***	-1.273*	-2.288***	2.368***	2.600*	1.292

	unemployed-> employed_FT	Unemployed-> Employed_PT	Inactive-> Inactive	Inactive-> Employed_FT	Inactive-> Employed_PT	Inactive -> Unemployed
Compulsory	-0.711***	-0.642***	0.669***	-1.074***	-0.884***	-1.006***
High_school	-0.386***	-0.533***	0.316**	-0.603***	-0.671***	-0.594***
Age	0.007	0.118**	0.009	0.102***	0.255***	0.059*
Sq_age	0	-0.001	0	-0.002***	-0.003***	-0.002***
centre	-0.692***	-0.750***	-0.295***	-0.067	-0.394***	-0.115
South	-1.378***	-1.534***	-0.228***	-0.355***	-1.206***	0.297***
Married	-0.003	0.299***	1.280***	-0.794***	-0.610***	-1.171***
C_48_52	-0.268	0.005	-0.362*	-0.239*	0.087	0.077
C_53_57	-0.38	0,233	-0.254	-0.371**	0.028	0.011
C_58_62	-0.533	0,491	-0.188	-0.513***	0,316	-0.063
C_63_67	-0.574	0.806*	-0.232	-0.502**	0.618*	0.029
C_68_72	-0.644	0,739	-0.383	-0.533**	0.789**	0.019
C_73_77	-0.543	1.127*	-0.364	-0.424	0.934**	0,152
c_78_max	-0.538	1.222**	-0.319	-0.641*	0,307	-0.007
_cons	0,383	-4.647***	-2.556***	-2.035**	-7.230***	-1.479**

4.12 The income module

The "income" module simulates yearly labour income for active population. The estimation of life-cycle labour income is based the ISFol PLUS cross-section survey 2004. This survey is specifically focused on active population and includes useful information for the purposes of dynamic simulation, as for instance nationality and work contract typology.

The econometric model specification is the following[27]:

$$\ln y_i = \alpha + \beta X_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0,\sigma^2_\varepsilon)$$

where lny_i is the log of individual labour income gross of personal taxation and the X vector includes the set of observables which are usually employed in a human capital model *a là* Mincer. The income level is separately determined for employees and self-employed workers. The group of employees is in turn divided by educational level and gender[28]. Table 9 shows the estimation results based on some ISFol sample subsets.

The signs of coefficients are in line with expectations. In particular, income is “bell” shaped with respect to age, and growing with educational level and seniority, while it appears, *coeteris paribus,* lower for women, Southern individuals, immigrants, blue-collars and public employees.

On average, blue-collars’ wage is, *coeteris paribus,* significantly lower than white-collars’, managers with secondary or tertiary education receiving in turn 6% to 10.5% higher income compared to white collars.

A public employee receives a lower wage compared to an employee in the private sector. The negative value of the gender dummy suggests a lower income for self employed women, while a self employed with tertiary education has an income 45.4% higher than those with just compulsory education. Finally, atypical workers earn a wage 45.8% lower than a self employed.

Once the coefficients are estimated, the level of gross income is computed for each individual considering the evolution of observable characteristics only (age, work experience, education, etc).

However individual income differs from the mean income for two reasons: first, due to an individual component (constant all over the period) which can be interpreted as a proxy of ability and effort at work; secondly, due to a yearly component which can be thought as the increase in productivity distributed to all workers in each simulation period.

[27] Errors are assumed to be normally distributed with zero mean and σ^2 variance.

[28] The limited availability of observations for graduated independent workers leads us not to disaggregate data by gender. For the same reason we decided not to decompose the subsample of self-employed by gender and education.

Table 11 OLS estimation coefficients of log gross labour income

	Employees			Self-employed
	Men	Women	Graduated	
Age	0.0351***	0.0233***	0.0436***	0.0790***
Sq_age	-0.0003***	-0.0002***	-0.0004***	-0.0008***
Women	-0.2052***	-0.5418***	-0.2052***	-0.5418***
Noth	0.0423**	0.0460***	0.0303	0.0662
South	-0.0600***	-0.0733***	-0.0643**	-0.1807***
immigrants	-0.0577	-0.0662	-0.1942**	-0.0969
Atypical				-0.4584***
Partime	-0.4079***	-0.3460***	-0.3556***	0
Years_contrib	0.0029***	0.0069***	0.0032**	0.0013
High_school	0.1176***	0.1360***	0.3593***	
Degree				0.4536***
fixed_term	-0.0702***	-0.0614***	-0.1644***	
Blue_collar	-0.0780***	-0.0990***	-0.1832***	
Mager	0.0602***	0.0620***	0.1057***	
Public	-0.0708***	-0.0104	-0.0852***	
_cost	9.0353***	9.0260***	9.0727***	7.9228***
N	4772	4819	3144	3127
Sq_R	0.3291	0.395	0.38306	0.2045

The distinction between these components allows computing for each individual the level of gross labour income according to the following formula:

$$\hat{y}_{i,t} = e^{\frac{1}{2}\hat{\sigma}^2 + \log \hat{y}_{it}} e^{(u_i)} (1 + \tau_t)$$

The first term on the right side is an unbiased estimation of the mean gross wage for individuals with similar observable characteristics, where $\hat{\sigma}^2$ is an unbiased estimator of the error variance and log $\hat{y}_{i,t}$ the estimated log $y_{i,t}$ (Wooldridge, 2003)

The term e^{ui} represents interpersonal variability among workers with similar observables which display different level of ability, talent, etc. This component is typically non-observable, and in the praxis, deviations from the mean are recovered as residuals from the first stage regression. Of course, this procedure is feasible for individuals employed in the base year only. For individuals which enters the active population in a second while (future births, students, unemployed), this term is randomly generated - when they enters into the labour market – from a normal distribution with zero mean and variance equal to the estimated variance from the first stage regression (Root MSE). All the workers are provided through this method with an individual fixed effect, the distribution of it being constant in time, implying differences in wage due to unobservable characteristics remain constant all over the simulation period.

Finally, the $(1+\tau_t)$ factor allows the wage level to be linked to the medium-long run productivity growth which is calibrated through the "scenario" block. There is again one point which needs to be made clear: the demographic evolution and the increase in the stock of human capital in the next decades make the mean income level increasing, since age and education have a positive effect on labour income[29]. However, in this model, endogenous growth is lower than the growth forecasts according to RGS, since it does not allow for the expected increase in productivity. In order to avoid over/under-estimations of earnings growth rates for the next decades, the following procedure is adopted: every year, a *pro-quota* growth factor *e* - equal to the difference between the exogenous earning growth fixed in the "scenario" and the growth estimated by the model - is added to the endogenous growth due to the socio-demographic evolution.

The term τ_t is given by:

$$\tau_t = \prod_{i=1}^{t} (1+m_i) - \left(\frac{E(\bar{y}_t)}{E(\bar{y}_{t-1})} \right)$$

[29] Other factors could have a negative effect, for instance the increase of female participation to labour market, the increase of immigrants and the diffusion of part-time contracts.

where *m* is exogenously determined in the "scenario", while $E(\bar{y}_t) / E(\bar{y}_{t+1})$ describes the endogenous growth rate generated by the model.

5. The social security module

Individual retirement choices and the computation of old age, seniority and survivors pension benefits, as well as of social allowances, social assistance increases (maggiorazioni sociali) and supplements *(integrazioni al minimo*) are simulated in this module.

The individual pension transfer depends on the following variables:

- the life-cycle profile of labour incomes;
- the seniority of social security contribution at the moment of retirement;
- the contribution rate during working life;
- the – exogenous - macroeconomic growth during the period of pension contribution;
- the pension scheme;
- the retirement age.

The first four variables depend on the results provided by the previous "demographic" and "education and labour market" modules. In particular, the life-cycle profile of labour incomes depends on the evolution of characteristics controlled for in the regressions reported in table 2.2.9. The seniority at the moment of retirement depends on the total amount of years the individual received a positive labour incomes according to the model. Due to the chance of transitions among different employment statuses during the working life, a share of simulated individuals may show periods with no contributions.

The model simulates the following pension benefits:

1. Old age and seniority pensions,
2. Survivors and indirect pensions,
3. INPS disability pensions,
4. Civil disability pensions,
5. Social allowances,
6. Supplements to the minimum (*integrazioni al minimo)*
7. Social assistance supplements (*maggiorazioni sociali*).

The benefit amount is held constant in real terms for the whole retirement period, accordingly to the general pension indexation system introduced by the 1992 reform[30]. Minimum pension amount, contribution caps and minimum and maximum thresholds for determining the benefit according to different pension schemes increase with time following the real GDP growth forecasts selected in the “scenario”[31].

5.1 The decision to retire

The selection of the new retirees is implemented each year of the simulation following a two stage procedure. In the first stage all individuals fulfilling the necessary conditions (in terms of age and contributory seniority) for old age and/or seniority pension are identified. The identification is implemented according to the law in force, in particular by law n. 243/2004 and law n. 247/2007[32].

The first chance for retirement occurs when an individual achieves age requirements for the earlier retirement (compared with the statutory age), which is nowadays fixed at 60 years for women and 65 for men[33]. In this case the model checks whether the exit from the labour market is inter-temporally advantageous. In practice, it compares two options: keep working one more year or exit immediately. If the net social security wealth is greater under the second option, then the retirement choice is effectively simulated if the replacement rate exceeds a certain threshold, set at 60%.

Therefore, the model allows for two relevant aspects in the retirement choice: the first is an evaluation of inter-temporal convenience and the second relates to the adequacy of the pension benefit provided by the social security system[34].

The second way for retirement is the achievement of the statutory age conditioned to the minimum requirements of social security contribution seniority, amounting to 20 years for

[30] The model does not allow for inflation.

[31] The path of real GDP growth is exogenous to the model. The implemented procedure differs from the pension scheme provisions being in force, which allows the pensions and the social allowances minimum amount adjustments through legislative action. However, we consider the option of keeping the minimum pensions constant to the base year values as being unrealistic as well as forcing discrete adjustments in certain years is quite arbitrary.

[32] The increases in requirements for an early retirement are revised with respect to age 65 for men and 60 for women, according to the law n. 243/2004. In particular, according to the provisions of the law, an increase to 58 of the age requirements for seniority retirement is provided in 2008, being unchanged to 35 the seniority requirement. From 2009 a system of quotas is introduced: the early retirement will be allowed if the sum of worker’s age and contributory seniority will be higher than 95 (until the end 0f 2010), 96 (until the end of 2012) and 97 (after 2013), with the age requirement however increasing from 59 in 2009 t0 61 in 2013. Requirements for self employed are one year higher.

[33] This chance will remain also with the DC system for the workers with a seniority of 40 years, apart from age.

[34] The choice of the particular threshold value adopted is clearly sensitive with respect the determination of the actual mean age of retirement. The choice of 60% seems to be a “reasonable” option which in practice restricts the seniority retirements just in the first part of the simulation period.

individuals subjected to the defined benefit (DB) or the mixed system and 5 years for those in the defined contribution (DC) system. In the latter case, the exit rule and the provision of the (old age) pension benefit depends on the gender of the worker. It is worth reminding the regulation in force determines a different retirement age for males and females (65 for the men, 60 for ladies). Once they are 65 and the minimum contributory requirements have been fulfilled, then men are supposed to retire, independently of the evaluation in terms of inter-temporal convenience and adequacy of the pension benefit. For women, a different rule has been adopted in order to avoid a substantial flow of pensioners receiving very low replacement rates, especially in the period when the DC system will be at work. We feel the hypothesis of an exit at 60 unrealistic for individuals which, especially in the later part of the simulation, will be more educated on average – therefore entering the labour market later – and in the same time will define their pension according to the DC rule which tends to penalize an early exit from the labour market.

In order to avoid the formation of a stock of young pensioners receiving very low pensions we calibrate the model so that the actual retirement age progressively increases every ten year up to 65 in 2050.

5.2 The computation of old age and seniority pensions

In the following, the computation formulas used for the estimation of the amount of the first pension transfer according to the three pension schemes, (the defined benefit, the mixed one and the defined contribution) are shown. In general, computing the pension benefit is not an easy task due to the joint consideration of both individual and household income criteria in the provision of social transfers such as supplements to the minimum (*integrazioni al minimo*) and social assistance supplements (*maggiorazioni sociali).* When possible, in the imputation procedure concerning these types of social benefits, the socio-economic features of the individuals are allowed for as well as their temporal evolution.

- <u>The defined benefit regime</u>

Individuals who have achieved at least 18 years contribution seniority at the end 1995 are under the DB regime. The calculation rule of pension benefit is synthesized by the formula:

$$P_{DB} = r^*(N_1 W_1 + N_2 W_2)$$

where r is the pension rate of return, N_1 and N_2 represent the year of contribution before and after 1992 respectively, W_1 and W_2 represent the pensionable earnings useful for pension estimation, for contributions paid before and after 1992 respectively.

The terms in the DB formula are not constant for all workers as they depend on the pension scheme and on the pensionable wage level. In particular, W_1 is equal to the last wage for employees in the public sector and to the five and ten years average for employees in the private sector and self-employed workers respectively. W_2 is the mean over ten years wage for public and private sector employees and over 15 years for self-employed workers. The rate of return r is equal to 2% for the pensionable earnings bracket between 0 and 36,980 Euros in 2002 and it decrease with earnings level down to 1.1% for the pensionable earnings bracket over 49,156 Euros.

Pensioners who have paid at least 20 annuities of contribution, but did not reach the minimum pension amount, are supplemented up to the minimum level.

- The mixed regime

This regime is applied to workers with less than 18 years of social security contribution seniority in 1995. In this case the old age/seniority pension benefit is determined as the sum of two components; the first component is P_A and it is computed according to the DB formula on the contribution paid before 1995, while the second, P_B, is computed according to a DC rule on the after 1995 contributions. In formula:

$$P_{mixed}=P_A+P_B$$

where the general rule for determining P_A is similar to the formula used in the DB regime. Nevertheless, in the "mixed" regime the pensionable wage for the contributions paid between 1992 and 1995 is determined differently, being computed as the mean wage over the years after 1992 indexed to 1% yearly rate according to a simple compounding rule. The P_B term of the mixed pension is figured according to a DC rule.
Pensioners which have paid at least 20 annuities of contribution but did not reach the minimum pension amount are topped up to the minimum level.

- The defined contribution regime

This regime involves people entered into the labour market after 1995, for which the amount of pension is defined as:

$$P_{DC}=k*MC$$

where k is the transformation coefficient that varies with retirement age so as to guarantee a quasi-actuarial equity between the present value of paid contribution and the present value of expected pension benefits. In order to allow for the expected evolution of mortality accounted for in the model by ISTAT official forecasts, the coefficients computed by RGS in its forecast model adjusting pension amounts according to varying demographic conditions every ten years have been employed.

MC is the "*montante contributivo*" i.e. the total of contributions accrued during the whole working life capitalized at the rate of growth of nominal GDP, defined by the sum of the contribution paid compounded at the GDP growth rate. The yearly contribution is computed as a share of the gross wage for employees and gross income for self-employed. The contribution rate is set at 33% for employees and 20% for self-employed workers. A contributory cap is set at 82,404 Euros. In order to benefit of minimum pension at least five annuities of contribution have to be paid. Finally, the amount of the pension cannot exceed the amount of social allowance increased by 20%. In the opposite case, pension will not be paid if the requirement is achieved before the statutory age of retirement which is set at 60 for women and 65 for men. For the pensions provided under the DC scheme no supplement up to the minimum is allowed, while a supplement up to the level of social allowance is provided if the income requirements are fulfilled.

- Survivors and indirect pensions

The death of a pensioner or of an insured entitles the survivors to benefit of survivors and indirect pension respectively.

The model allows the payments of indirect pensions if the dead had achieved at least 5 years contributions. When the total amount due to the survivor has been determined, the shares are assigned to every single components of the household accounting for their economic conditions[35].

In particular, the model distributes the total amount among the survivors assigning:

- 60% to the spouse
- 20% to each child in case the spouse is alive

[35] Current regulations can be found by consulting the INPS website in the section *"la pensione ai superstiti"*.

- 40% to each child in case the spouse is not alive.

The sum of the share cannot exceed the 100% of the pension the retired would be entitled for.

If one child is entitled only, the share is set at 70%. The amount of survivor pension depends on economic conditions of the beneficiary. The allowance is reduced by 20%, 40%, and 50% in case the income earned by the beneficiary exceed by 5, 4, 3 times respectively the amount of the minimum pension benefit. The last rule is not applied in case underage students or disable children are co-entitled.

5.3 Civil and INPS disability pensions

The social security module selects beneficiaries of disability allowances, disability and civil inability pensions.

The disability allowance amount is determined according to the standard system of old-age pension computation depending on the pensionable earnings and on the contributory seniority of the insured.

If the allowance amount is lower than the minimum benefit, it can be supplemented up to this limit.

The disability pension amount consists of two components: one share is determined according to the pensionable earnings and to the contributive seniority as for the inability allowance, while the remaining part, named "*maggiorazione*", is determined by the difference between the inability allowance and the pension she/he could benefit should she/he had accrued a seniority increased by a period amounting to the difference between the year the inability allowance started to operate and the pensionable age (set at 60 years for men and 55 years for women with a dispensation to disabled persons at least for the 80%). No seniority above the 40 years will be considered.

In order to determine the amount of disability pension with the DC system, the sum of contributions accrued is added to a share of contributions for the gap period between the pension starting year and the 60th birthday. Again, seniority cannot exceed 40 years.

The amounts of civil disability pension are in line with the values which are provided by the current law and are updated according to the GDP growth from the second year of simulation.

The module providing the entitled individuals with the civil disability pension benefit checks whether the yearly income requirements are fulfilled or not.

5.4 The social allowances

Over 65 persons which have an individual/household income under the statutory limit are supplied with the social allowance. In fact, the Italian regulations provide the social allowance can be paid out when certain beneficiary or her/his spouse's economic conditions occur. In order to allow – at least partially - for these conditions, the model sums up each year labour and pension incomes of the spouses. The benefits are then paid out when the total income is lower than the statutory thresholds, which are yearly revalued according to the real GDP growth rate. In general, the allowance amount is designed as to raise the overall income up to the legal threshold.

So for instance, in 2007 the monthly allowance amount is equal to 389.36 Euros. It implies the yearly amount in the same year is 5,061.68 Euros (389.36 x 13 monthly installments) corresponding also to the income limit for an applicant which is not married, the yearly limit being 5,061.68 x 2 = 10,123.36 whether the applicant is married. In general, the yearly allowance is equal to the gap between the current yearly social allowance and the effective overall income earned by the applicant – in case summed to the income of the spouse - [36]. Unlike the social pension[37], the social allowance - or a share of it - is paid to the applicant even if he/she has a personal income above the individual legal threshold, as long as the overall income of the spouses is lower than the legal threshold for the couple.

[36] So for instance, if the applicant is married and the yearly total income is 9,000 Euros, the allownce amount is reduced to 1,123.36 Euros (the difference between 10,123.36, current annual income threshold for a married applicant, and 9,000 Euros).

[37] From January the 1st 1996 the social allowance replaced the social pension, which however keeps on being payed out to those individuals which, being enitled for, applied within 31st of December 1995.

5.5 Supplements to the minimum (integrazioni al minimo) and social assistance supplements (maggiorazioni)

As already mentioned, when the supplied pension benefit is below the minimum legal amount the model provides the pensioner with a supplement or a social assistance increase, allowing for personal and household income as well as the pensioner age[38].

[38] The computation of the minimum treatments is carried out according to the provisions reported into the INPS website in the section *"il trattamento minimo"*.

REFERENCES

Ando A. and Nicoletti Altimari S., (2004), A Micro Simulation Model of Demographic Development and Households' Economic Behavior in Italy, Temi di discussione Banca d'Italia n. 533.

Atkinson A B, Gomulka J., Sutherland H. (1988), Grossing-up FES data for Tax-Benefit Models, in *Tax Benefit Models*, (ed.) Atkinson A. B., Sutherland H., STICERD Occasional Paper no. 10, LSE.

Auerbach A. and L: Kotlikoff, (1987), Dynamic Fiscal Policy, Chicago University Press.

Borlini B., Zajczyk F., (2001), "Matrimonio e lavoro: l'influenza del coniuge", in Bianco M. L. (ed), L'Italia delle diseguaglianze, Carocci, Roma

Blangiardo, G.C., (1997), Elementi di demografia, Il Mulino, Bologna.

Brandolini A., (1999), The distribution of personal income in post-war Italy: source description, data quality, and the time pattern of income inequality, Temi di discussione, n. 350, Banca d'Italia, Roma.

Cannari L., D'Alessio G., (1992), Mancate interviste e distorsione degli stimatori, Temi di Discussione, No. 172, Servizio Studi, Banca d'Italia, Roma.

Checchi D., (2003), Scelte di scolarizzazione ed effetti sul mercato del lavoro, in Lucifora, C. (a cura di), *Mercato, occupazione e salari: la ricerca sul lavoro in Italia*, Mondatori, Milano.

Checchi D., and Flabbi L., (2005), Intergenerational mobility and schooling decisions in Italy and Germany, mimeo.

Chies L., Lucchetti R., Staffolani S., (1998) Occupazione, disoccupazione, inattività: determinanti della mobilità tra stati in Italia, Rivista Italiana degli Economisti, n.3.

Costello D, Przywara B (2007) Projecting future health care expenditure: drivers, trends, policy recommendations, forthcoming in *European Economy*, European Commission, Directorate-general for economic and financial affairs.

Del Boca D., Locatelli M., Pasqua S., (2000), Employment Decision of Married Women: Evidence and Explanations, Labour, vol.14.

Ermisch J, Francesconi M (2001) 'Family Matters: Impacts of Family Background on Educational Attainments', *Economica*, 68 (270), 137-156.

Heckman J.J., (1979), Sample selection bias as a specification error, Econometrica, 47:153-161.

ISFOL, (2006), Impianto dell'indagine PLUS, Roma.

ISTAT, (2000), Le strutture familiari, Collana Informazioni n. 17, Roma.

ISTAT, (2002), La mobilità nel mercato del lavoro: principali risultati aprile 1998-aprile 2002, Collana Approfondimenti, Roma.

ISTAT, (2004), Forze di lavoro Istat. Anni 1993-2003, Annuario n 13, Roma.

ISTAT, (2004), Matrimoni, separazioni e divorzi. Anno 2000, Annuario n 13, Roma.

ISTAT (2007), Condizioni di salute, fattori di rischio e ricorso ai servizi sanitari (anno 2005), May (http://www.istat.it/salastampa/comunicati/non_calendario/20070302_00/) [accessed 15 July 2007].

Leombruni R., Richialdi M., (2006), LABORsim: an agent-based microsimulation of labour supply. An application to Italy, mimeo.

Klevmarken, N. A (2005)Dynamic Microsimulation for Policy Analysis – Problems and Solutions, paper presented at The Economics Conference of Australia 2005, Melbourne Sept 26-28

Mandrone E., Radicchia D., (a cura di.) (2006), PLUS: participation, labour, employment, survey. Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro, Rubbettino Editore, Soveria Mannelli (CZ).

Manton K G, Gu X, Lamb V L (2006) 'Changes in the prevalence of chronic disability from 1982 to 2004/05 as measured by long term changes in function ad health in the US elderly population', *Proceedings of the National Academy of Sciences*, 103, 18374-79.

Mazzaferro C., Morciano M., (2005), Un modello di microsimulazione a popolazione dinamica per la stima degli effetti distributivi della riforma pensionistica, paper presentato alla Conferenza SIEP 2005, Università di Pavia.

Ministero del Lavoro e delle Politiche Sociali, (2005), Aspetti distributivi del sistema pensionistico nella prospettiva del passaggio al sistema contributivo, Roma disponibile sul sito http://www.welfare.gov.it e sul sito www.capp.unimo.it.

Ministero della Solidarietà Sociale, (2008), Studio prospettico della povertà in Italia, Rapporto finale, Roma.

Morciano, M. (2007), Un modello di microsimulazione a popolazione dinamica per l'analisi del sistema di protezione sociale italiano, Unpublished PhD Thesis, University of Bologna.

Norton E C, Stearns S C (2004) 'Time to include time to death? The future of health care expenditure projections', *Health economics*, 13(4), 315-327.

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

Orcutt G.H., Mertz J., Quinke H (Ed.), (1986), Microanalytic Simulation Models to Support Social and Financial Policy, North-Holland, Amsterdam.

Rosembaum P.R., Rubin D.B., (1983), The Central role of the Propensity Score in Observational Studies Causal Effects, Biometrika, n. 70, 1.

Rossetti S., Tanda P., (2000), Human capital, wages and family interactions, Labour, vol.14.

Trivellato U., Battistin E., Rettore E,. (2005), Choosing among alternative classification criteria to measure the labour force state, IFS Working Papers W05/18, Institute for Fiscal Studies.

Vagliasindi P., (2004), Effetti redistributivi dell'intervento pubblico: esperimenti di microsimulazione per l'Italia, Giappichelli, Torino.

Zaidi A., Rake K., (2002), Dynamic microsimulation models: a review and some lessons for SAGE, Sage discussion paper n. 2, London.