2.º
CICLO

FCUP
2015

An Agent-Based approach of the Portuguese population projection and the Social Security sustainability

Renato da Silva Fernandes

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

An Agent-Based approach of the Portuguese population projection and the Social Security sustainability

Renato da Silva Fernandes
Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Matemática
2015

An Agent-Based approach of the Portuguese population projection and the Social Security sustainability

Renato da Silva Fernandes
Mestrado em Engenharia Matemática
Departamento de Matemática
2015

Orientador
Prof. Doutor Pedro Campos, FEP

Coorientador
Prof. Doutora Ana Rita Gaio, FCUP

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/_________

MSc

Resumo

O impacto da estrutura por idades da população portuguesa na segurança social é estudado nesta tese, focando nos reformados, pois o pagamento de pensões de reforma é de longe a função mais importante e mais dispendiosa da segurança social [Sta15c]. Nos últimos anos, a idade mínima de reforma aumentou, a percentagem das contribuições para a segurança social aumentou, as fórmulas para o cálculo das pensões foram alteradas e a função do estado na assistência social diminuiu. A incerteza relativa à sustentabilidade do sistema é uma justificação possível para estas alterações.

Neste trabalho foram criados alguns modelos baseados em agentes com a especificação da idade e género dos agentes para uma população humana e para as suas atividades econmicas, através de mecanismos de simulação, regressão e decisão. Estes modelos são então aplicados à população portuguesa e ao seu sistema de segurança social numa tentativa de mostrar o crescimento da população Portuguesa e para tentar descobrir se a segurança social Portuguesa é sustentável.

Um bom conhecimento da evolução da população por faixas etárias permite a realização de estudos mais aprofundados. Assim é possível compreender a interação da população com o mercado de trabalho. Adicionalmente, são encontrados padrões migratórios e é fornecida uma descrição populacional mais detalhada. Também é possível obter melhores estimativas para o crescimento da população como um todo.

Este trabalho mostra que o tamanho popuação Portuguesa irá decrescer até o ano de 2041, este decréscimo é em grande parte referente a pessoas com menos de 65 anos de idade. Isto é principalmente justificado pelo aumento das emigrações, que por sua vez leva a uma redução nos nascimentos, uma vez que as mulheres emigram em maior volume na sua idade fértil. Adicionalmente, mostra-se que o sistema de Segurança Social Portuguesa precisaria de ter pelo menos 250 mil milhões de euros de capital e estar a receber juros nesse mesmo valor desde o ano de 2011, para que o atual sistema de pensões seja sustentável.

Abstract

The impact of the age structure on social security is studied in this thesis. The focus of the thesis is on the Portuguese population and the elderly pensioners, because it is by far considered to be the most important and is the most costly function of social security [Sta15c]. In the past years, the retirement age has increased, the contribution percentage to the Social Security system has increased as well, the pension formulas have changed and social assistance has been reduced. This is maybe because of the risk that the system was not sustainable.

In this work, some age and gender specific Agent-based models for a Human population and its economical activity are created, through a combination of simulation, regression and decision making mechanisms. These models are then applied to the Portuguese population and its social security system in an attempt to shed some light on how the Portuguese population will grow, and on the sustainability of the Portuguese Social Security.

A good knowledge on how population age groups evolve enables further studies to be performed. It is then possible to understand how population and labor market interact. Furthermore, patterns on migration are found and a more detailed population characterization is described. It is also possible to obtain better estimates for the entire population growth.

This work shows that the Portuguese population size will decrease until the year of 2041, mainly in the number of persons of age below 65 years-old. This decrease is mostly justified by an increase on the emigration which also leads to a decrease on births because women usually emigrate in their fertile age. Also, it is shown that the Portuguese Social Security system would need to have a base capital of at least 250 thousand million euros and receiving interest on it since 2011, so that the current pension system remain sustainable.

Acknowledgements

I would like to thank my supervisor Prof. Pedro Campos for suggesting me this thesis theme even though I am from a different faculty and for helping me making the bridge between all the sub-themes that take part on this thesis, as well as my co-supervisor Prof. Ana Rita Gaio for helping me maintaining the formalism and scientific rigor demanded for the Master degree in Mathematical Engineering.

Also I thank all my friends and colleagues with a special emphasis to Ricardo Cruz for keeping up my moral during this degree and helping me finishing this degree and this thesis.

Finally, I send my love to my mother Rosa, father José, grandmother Teresa and uncle Manuel.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Today there are over 7 thousand million people alive [CIA13], composed of children, adults and elderly people. The proportion of population in each age group plays an important role both in the whole population growth and in social-economic activities.

Demography has been studied for hundreds of years. While many developments have been made so far, in the last few years the development has been less than significant when considering full population growth projections (Coale and Trussell [CT96]). This is a dire problem which can show that either the current developments are not intended for full population growth or that they are not effective.

With the evolution of modeling techniques, more specific studies are being made, nowadays. Some models focus on specific demographic phenomena, like nuptiality or mortality, and some models focus on phenomena for different population characteristics, such as age and gender.

We have found that trying to model several closely dependent entities, namely fertility, mortality and migrations, while being specific on the age and gender of the population was a true challenge for the author of this work. This is why we had got motivation to study this theme and wanted to move forward as further as possible.

In addition, the Portuguese Social Security system sustainability is a very mediated event in Portugal and as the author is also on the verge to enter the system, this system's sustainability is a great matter of personal concern. One of the possible key aspects for this system' sustainability is the population proportion that in the future will be contributing for the system and the proportion that will be supported by it. So, having in mind the very close link between the Portuguese population growth and the Portuguese Social Security system, it brought us additional motivation to study the economic events around the problem.

However, the aspect that influenced our decision most on whether we would pursue the research was the modeling technique to be used: the agent-based modeling. Agent-based modeling is a recent simulation technique which instead of modeling directly the event, models the intervenor that influences the event. In other words, this is a bottom-up approach. For the study at hand the chosen intervenor to be modeled had to be a human being and with a sufficiently large set of intervenors it is possible to recreate the Portuguese population. With this modeling technique we saw many innovation possibilities, which quickly built our interest over this project.

Therefore, a new age and gender specific agent-based model was created to study the demographic and social-economic evolution of the Portuguese population. This model has got some interesting particularities. A base model is created to support multiple fertility, mortality, migration, employment and social security system models and this model can easily be expanded to support other entities or events. Therefore, if there is the will to try different growth models for each considered entity or event, there is no need to make a new connection function. New models can be easily added to the base model that is presented here. Another particularity is on the present models, in which some use an additional modeling technique called Mic-Mac modeling, where the actions of an intervenor can affect the whole country and where events in the whole country independent to the intervenor actions can also affect each intervenor.

In addition, to the presentation of the models, their results and projections are also disclosed. Offering several possible scenarios for the evolution of the Portuguese population size and for the sustainability of the Portuguese Social Security system, always presented in a summarized and concise way.

The following chapters are organized in the following manner:

Chapter 2: a historical tour through the more remarkable aspects of modeling and the development of demographic and social security models by time and perspective will be displayed.

Chapter 3: the main engine of the developed agent-based model will be shown. This will showcase the main population structure and some applicable models for fertility, mortality and migration.

Chapter 4: the population model presented in Chapter 3 is extended to support some economic behaviors. The main mechanism to model the economic status of an agent is detailed. In addition, some models which can be used to model employment and social security systems are presented.

Chapter 5: the starting data is presented alongside with its sources. The initialization mechanism of both population and social security is also shown. Finally, the main results are exposed.

Chapter 6: the results and respective models are discussed, focusing on the models' strengths and weaknesses.

Chapter 7: the main conclusions of the work are described and some insights for Portugal are drawn.

A summarized overview on the created model is observable in Figure 1.1 on the following page, showing all the developed models and the variables for which several values will be tested. The diagram sequence goes from the top to the bottom.

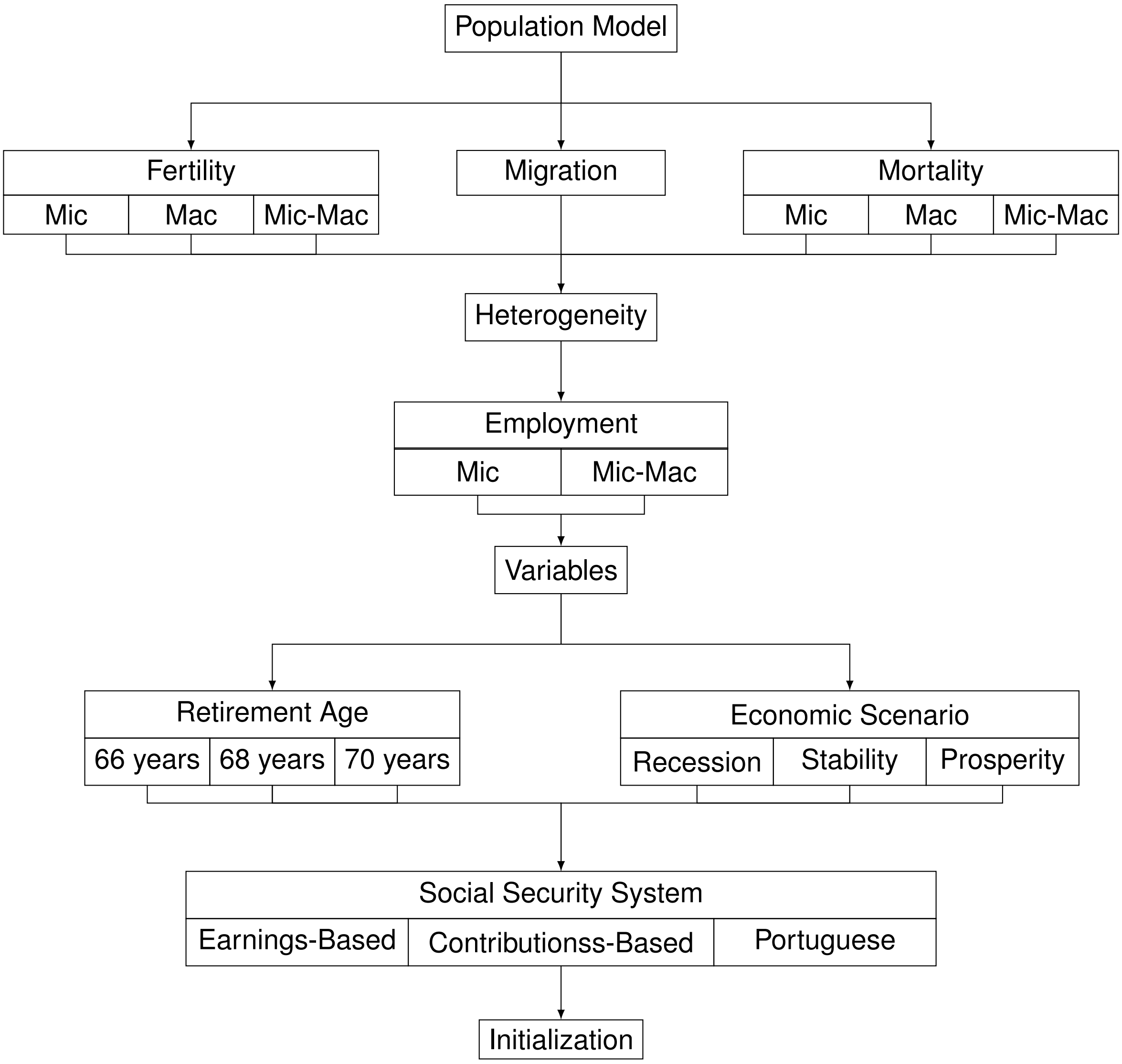

Figure 1.1: Model Overview Diagram.

Chapter 2

Literature Review

A model is an attempt to translate the real world into a mathematical formulation. As it is widely known, the real world is a very complex system and to translate it into one single model is impossible. Therefore, many simpler models are needed to translate the world, yet simplifying them implies a loss of accuracy. A model operates over some observable data. These observable data are the model variables.

This chapter presents basic ideas from demographic modeling. It starts with a general modeling section, where some modeling methods are shown and then progresses to a section of demographic modeling approaches, showing the demographic perspectives that have been studied. In addition, the origin of social security systems are presented along side some basic social security models for the elderly pension.

2.1 Modeling Approaches

A given phenomenon can be modeled by a multitude of methods. Though, there are special characteristics in common among them. These characteristics can be used to group them in big modeling approaches.

Here four groups are presented which are not necessarily incompatible:

The first two are the base models, the macro and the micro models. The first attends to model aggregated phenomena while the second tries to model disaggregated phenomena, hence the names macro for big entities and micro for small entities.

The remaining two groups have their own distinctions, but are combinations of the previous two. One is mic-mac group of models (micro-macro) where both micro and macro approaches are combined to establish connections and complementariness between them. Finally, agent-based models will be presented, which can be any combination of the previous models and the main distinction from them is that instead of modeling the phenomena directly, they model the intervenor.

2.1.1 Macro Models

Macro models attempt to govern the environment and usually resort to differential equations, regression methods, time series and others. A general rule is that macro models consist of solving a set of equations in order to obtain the general behavior of the considered phenomena.

The Lotka-Volterra equations are widely used to model the growth of populations (Lotka [Lot10]). They often resulted in exponential growths and some variants, such as the Malthusian model (Malthus [Mal98]) or the Verhulst model (Verhulst [Ver38]). Lotka-Voltera equations are defined as:

$$\frac{dx}{dt} = \alpha x - \beta xy \tag{2.1.1.1}$$

$$\frac{dy}{dt} = \delta xy - \sigma y \tag{2.1.1.2}$$

where x is the number of preys, y is the number of predators, t is the time and α, β, δ, σ are positive real parameters describing the interaction between the two species.

When large amounts of past data are available, a phenomenon may be modeled by regression methods, attempting to explain the phenomenon through a given set of explanatory variables; or may be modeled using time series, finding patterns on the past (Legendre [Leg05]).

If the set of used variables is sufficiently small, macro models may achieve results, analytically or numerically, in short time scales.

2.1.2 Micro Models

Micro models, in comparison to macro models, attempt to model disaggregated entities or events of the entire environment. Additionally, for models to be more realistic, when possible, researchers introduce stochasticity in the parameters. However, this addition has a severe drawback, the mathematical analysis becomes intractable. Instead, independently from its complexity, a model can always be simulated. The only limitations for the complexity of the model are the time that is acceptable to wait for the results or the computer power, such as ram, processor speed or graphics cards (Coale and Trussell [CT96]).

Micro models can be used to validate or evaluate the performance of many theoretical models for which testing data is lacking (Coale and Trussell [CT96]). They can also be used to study rare events and may lead to the discovery of unusual events not expected from the mere study of the theoretical model. Usually micro models are dealt with using the Monte Carlo method, from Neumann, Ulam, and Richtmmyer [NUR47], in juxtaposition to the theoretical analysis.

2.1.3 Mic-Mac Models

Mic-Mac models attempt to model a given phenomenon combining the two previously presented approaches. The mic (micro) part models smaller entities, such as small groups of individuals, individuals or particles. This part is usually modeled allowing some stochasticity and is solved through simulation. The mac (macro) part models large scale events, such as the weather, economic growth or net migration. This part is usually deterministic (Gaag, Beer, and Willekens [GBW05]).

In these models, the mic part is expected to influence the mac part and vice-versa. Also, this study can reveal the effect of each point of view on the other.

The great challenge within mic-mac models relies on the joint modeling of the mic and mac approaches. As one influences the other, the order by which they are executed is also an important choice.

2.1.4 Agent-Based Models

Agent-based modeling constitutes a recent approach to model phenomena. These have been used, for example, to model the life of the inhabitants of a town (Heiland [Hei03]) and the predator-prey dynamics (Wilensky [Wil99]). Each participant, which will be referred as agent in the future, in the given phenomena is modeled according to the following structure:

- definition of each participant permissions and constraints, like allowing individuals to move freely or forbidding two individuals to stand in the same place at the same time;
- definition of their rules and interaction patterns, like a buy-sell pattern or a simple rule where every individual must eat to remain alive;
- definition of environmental rules, such as climate influences on a person actions or the economic recession.

These models allow for a very large number of variables and/or constraints and admit stochasticity. To model and test a phenomena is not always needed past data; with simple and acceptable assumptions these models can be initialized. All this is achieved by the way these models are run. They do not try to solve a system of equations (possibly differential and usually stochastic), which with the allowed number of variables would most likely be impossible to solve. They also do not try to supply a model obtained through regression methods, nor do they try to discover a time series that governs the data, as they usually do not have enough data to achieve such results. These models run through simulation so all they require, besides a decent model and starting variables, is a sufficiently powerful computer to deal with all the interactions between the individuals and the environment. To detect the most usual outcomes, the simulation requires to be run several times (Billari, Ongaro, and Prskawetz [BOP03]).

The main aim of a model is to obtain the most probable outcome, however agent-based models and simulation models can also provide rarer events (Chattoe [Cha03]). Rare events are useful for the prediction of worst case scenarios and provide a broader horizon on the situation being modeled.

Agent-based models may also explain how a given phenomenon occurs, instead of only offering the final outcome. The interaction history can be stored and a theoretical analysis can be made, building the paths for the possible outcomes (Chattoe [Cha03]).

As in most modeling approaches, sensitivity analysis to a variable can be performed, as well as a study of the impact of adding or removing a variable. Due to the large flexibility of this modeling method, hybrid models can be created whereby other modeling methods are incorporated within this method. Incorporating models is possible when data is previously provided or is generated by the simulation itself (Wooldridge [Woo02]).

2.2 Demographic Models

In the Greek language, ”demo” means ”the people” and ”graphy” means ”writing, description”. And so, demography is the scientific and statistical study of populations, including of human beings. Demography encompasses the study of the size, structure, and distribution of these populations, and spatial and/or temporal changes in them in response to time, birth, migration, ageing, and death (Preston, Heuveline, and Guillot [PHG01]).

There are essentially two different approaches to the study of demography (Preston, Heuveline, and Guillot [PHG01], Caswell [Cas01]):

1. Cross-sectional, studying members of a population of different ages in a specified time interval;
2. Cohort, studying the evolution of a population with the same event (e.g. birthing year).

Events or entities, such as the birth rates or the number of deaths, are modeled using simple models with reduced number of variables. These simple models may, then, be used to construct more complex and complete models for the whole population structure. If the small models are complex then the whole population structure model would, most likely, become too complex to be dealt with.

Demographic models have often one of three objectives (Preston, Heuveline, and Guillot [PHG01], Caswell [Cas01]):

1. To model time series, which attempt to capture empirical regularities and are mostly used as theoretical frameworks, or to determine the quality of demographic data;
2. To attempt to estimate levels and trends in mortality, fertility, and other events or entities;
3. To smooth recorded age-specific time series for fertility, mortality and others.

Although most demographic models are macro models which try to model large events or the entire population; there are a few micro-models which try to model the individual (Coale and Trussell [CT96]). Also, a case was found where the modeling was done through the mic-mac approach (Gaag, Beer, and Willekens [GBW05]). These models can be either deterministic, where the outcome is certain, or stochastic, where outcome depends on chance. Usually, most phenomena can be treated both in a deterministic and in a stochastic way (Coale and Trussell [CT96]).

2.2.1 Conventional Life Table

Under a closed population model (no migrations), the size of a cohort for any given year can be computed. Let d_i be the number of persons in the cohort that die at age i and l_0 be the number of persons that were born in that cohort. Then

$$l_x = l_0 - \sum_{i=0}^{x-1} d_i \tag{2.2.1.1}$$

is the number of persons aged x years-old. In particular, $\frac{l_x}{l_0}$ is the proportion of individuals in the cohort surviving until age x. Defining T_x by

$$T_x = \int_x^{\infty} l_s ds \sim \sum_{i=x}^{\infty} i \; l_i, \tag{2.2.1.2}$$

which corresponds to the number of person-years to be lived by individuals surviving until age x, and $e_0 = \frac{T_0}{l_0}$ is the average number of years lived by the members of the cohort, i.e., the life expectancy at birth (Preston, Heuveline, and Guillot [PHG01]).

The study of a cohort corresponds to the study of a population subset that has the same year of birth finding all characteristics of this cohort implies the wait of a reasonable number of years, as every member must die. It is thus impossible to study a currently living cohort, prior to all members dying.

In order to solve this problem, the assumption that the death rates within a cohort do not differ from the current death rates in the population with age i for the individuals of the same age is made.

Therefore

$$l_x = \prod_{i=o}^{x} l_0 q_i \tag{2.2.1.3}$$

where q_x is the death rate of the individuals in the population that are aged $x - 1$ and that die before they reach x years old (Caswell [Cas01]).

2.2.2 Mathematical Models of Conception and Birth

Keyfitz [Key77], Bongaarts and Potter [BP83], and among others, created the first mathematical models for conception and birth, which have been vastly disseminated until now. Of these, birth interval models have been found to provide broader explanations for the birth cycles (Coale and Trussell [CT96]).

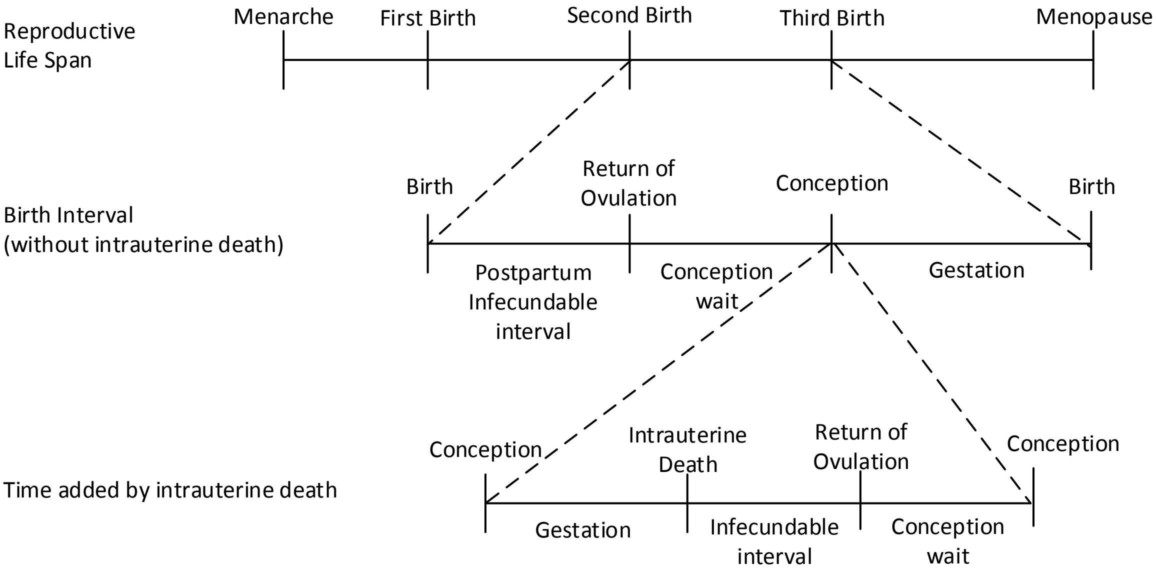

Figure 2.1: Birth Intervals.

From Figure 2.1, it is possible to observe that the fertile time of a woman can be split into multiple intervals. These intervals range from the date of menarche, which is the first menstruation, until the menopause, in which menstruation stops from occurring. Between these two moments a woman can have several births. The interval starting and ending in consecutive births is also divided into smaller intervals,

$$B_{int} = P_{int} + C_{int} + G_{int} \tag{2.2.2.1}$$

where

B_{int} is the between births interval;
P_{int} is the postpartum interval;
C_{int} is the conception interval;
G_{int} is the gestation interval.

If an intrauterine death occurs, then interval between the intrauterine death and the next birth is shorter than the usual interval between two births because the infecundable interval after an intrauterine death is shorter than a postpartum infecundable interval (Coale and Trussell [CT96]).

These models were essentially used as explanatory models for the effectiveness of anti-contraceptive methods, sterility, the effects of heterogeneity and sex-selection. This was done studying the difference between birth intervals when these variables were and were not present (Coale and Trussell [CT96]).

2.2.3 Age Schedules

The description of an event's value over the age of a individual is called an age schedule. This offers a very extensive report for the event at study and allows the creation of high quality life tables and it makes possible to test if already constructed life tables are coherent.

For an age schedule to be created it is needed all the accounts of a given event for age cohort, which, as already explained, proves to be impossible in some situations. So it is very important to use mechanisms that can work with missing data.

Age Schedules of Mortality

As in section 2.2.1 on page 8, the number of deaths at age x (d_x) is very important for the construction of a Life Table. So, it is very important to have very good models to predict age schedules of mortality.

In 1662, Graunt [Gra62] stated that $\frac{l_6}{l_0} \approx 64\%$ and $\frac{l_{76}}{l_0} \approx 1\%$, in London, where l_0, l_6, l_{76} were computed as in Equation 2.2.1.1 on page 8. He also found that the remainder values were almost proportional to those, which translate to an almost linear growth in mortality. From this, he made one of the first Life Tables in history. Later, in 1825, Gompertz [Gom25] changed the linear growth rate in mortality to an exponential growth, which was further improved in 1860 by Makeham [Mak60], by adding a constant term. Many more models appeared since then, but they will not be discussed here. It is worth noticing, however, that currently developed countries publish several variations of the models they use, varying their parameters to offer the biggest range of possible national outcomes.

Often, linear regressions on age schedules of mortality were used to correct life tables, on the years on which the information was incorrect or missing, so that future projections became more accurate, as stated by Coale and Trussell [CT96].

Age Schedules of Nuptiality

Nuptiality is usually modeled using empirical data, such as the the distribution of women's age at marriage, or by making simple and intuitive social assumptions.

The first model for nuptiality was developed by Coale and McNeil [CM72] who have found that the average year of the first marriage varied according to the geographic location. In most Asian and African countries, the women's mean age for the first marriage is bellow 15 years-old, while in Western Europe, it is over 25 years-old.

A second model was created by Hernes [Her72] and it is built over two assumptions;

1. The social pressure of a member of the cohort to marry is an increasing function of the percentage of the cohort members already married;
2. The odds to get married decreases as the age increases.

As stated by Hernes [Her72], this model appears to be in a high accordance with the United States marriage schedules.

Age Schedules of Fertility

The first age schedule of fertility was calculated using a polynomial function of the women's age, by Brass [Bra60] in 1960. About 20 years later, he attempted to create a relationship between a theoretical fertility schedule and an empirical fertility schedule. He showed that $\ln(-\ln(x))$, where x is the theoretical fertility schedule, has a linear correspondence with the same transformation on any empirical fertility schedule, as detailed in Brass [Bra81].

Henry [Hen61] has also created a model based on natural fertility, with the assumption that there was no deliberate conception control. Then, he introduced the conception control as a negative parameter in the birth rate. This parameter was related to the number of previous births.

Further improvements on the latter model were made incorporating marriage into the equation. Conception control started to increase with age, justified by the assumption that as age increased, the concept of "perfect" family size is reached.

Later, in 1977, Page [Pag77] improved the model that included marriage, adding the age of the woman and the time of marriage. This new model brought significant improvements to the fertility studies achieving results with correlations between this model and the empirical data over 0.99 in most societies.

Age Schedules of Migration

Little has been done by demographers in the field of migration, mostly because the processes involved are not governed by biological ones. However Rogers and Castro [RC81] created a model, relating migration with age. For this, eleven parameters were used. Seven parameters were used to model the individuals age and the other four were used to model the level of migration. This model is sometimes used to smooth data irregularities or to give confidence to the veracity of data or even in some cases to produce some interpolations for years with missing data.

Yet, the most widely used models are simpler models using only one parameter, by Hill [Hil81], or three parameters, by Zaba [Zab85].

2.2.4 Trends in life expectancy and survival

Population projection often involves the study of multiple scenarios and the results for all considered scenarios are published. This gives little certainty on what will happen, because we have too many possible outcomes to decide on.

In order to get a more concise response, researchers tried to model the future from the empirical relationship between life expectancy and the probability of survival up to a given age using period life tables (Mayhew and Smith [MS13]). In their model, a Gompertz-Makeham function is used in the fitting of the resulting distribution.

This method has a downside, it requires very accurate estimates of life expectancy which have been shifting very much. Also, it is expected that although the life expectancy of the Human kind might be increasing, it may be bounded by some age value. This fact presents a big problem to the estimation of the evolution of life expectancy.

2.2.5 Stable Population Equivalent

The stable population equivalent is used since World War II to show the implications of extreme conditions, such as the age composition and growth rate inherent in a combination of the highest fertility and the lowest mortality reliable record (Coale and Trussell [CT96]). Also, this population structure and growth is very close to the one observed in developed countries (Bras [Bra08]).

The long-term maintenance of constant demographic patterns may have important consequences in the population structure. A stable population emerges when some characteristics persist on the long term basis (Preston, Heuveline, and Guillot [PHG01], Caswell [Cas01], Keyfitz and Caswell [KC05]), namely:

- age-specific fertility rates are constant;
- age-specific death rates are constant;
- age-specific net migration rates are zero.

Setting the starting time at 0 for the study of the birth sequence of a population, then

$$B(t) = \int_0^t N(x,t)m(x)dx + G(t) \tag{2.2.5.1}$$

where

$B(t)$ is the number of births at time t;
$N(x,t)$ is the number of persons aged x at time t;
$m(x)$ is the rate of bearing female children for women aged x;
$G(t)$ are the number of births at time t from all the women alive at time 0.

Defining $p(x)$ as the probability of surviving until age x, then $N(x,t)$ can be expressed as

$$N(x,t) = B(t-x)p(x), \qquad t \geq 0 \tag{2.2.5.2}$$

therefore

$$B(t) = \int_0^t B(t-x)p(x)m(x)dx + G(t) \tag{2.2.5.3}$$

As in the Human population, the female fertile period ends at around the age of 50, then

$$B(t) = \int_0^t B(t-x)p(x)m(x)dx, \quad \text{for } t > 50 \tag{2.2.5.4}$$

As this is a homogeneous integral equation, it can be solved by a process of trial and error. Lotka showed that $B(t) = be^{rt}$ solves the equation. With this solution, the equation can be rewritten as

$$be^{rt} = \int_0^t be^{r(t-x)}p(x)m(x)dx, \quad \text{for } t > 50 \tag{2.2.5.5}$$

$$1 = \int_0^t e^{-rx}p(x)\phi(x)dx, \quad \text{for } t > 50 \tag{2.2.5.6}$$

Moreover, the fertile period of a woman is bounded below. Defining α and β as the starting and ending points of the fertile period, respectively, then the stable model, in its simplest continuous formulation, can be defined by

$$1 = \int_{\alpha}^{\beta} e^{-rx} p(x) m(x) dx. \tag{2.2.5.7}$$

Where r is the intrinsic growth rate of the population, which is the solution for the equation. An approximation for r can be given by $ln(R_0/K_1)$, where R_0 is the net reproduction rate and K_1 is the average fertility age (Stockwell, Shryock, and Siegel [SSS73], Pressat [Pre83], Tapinos [Tap85], Amegandjin [Ame89], Keyfitz and Caswell [KC05], Bras [Bra08], Preston, Heuveline, and Guillot [PHG01]). The intrinsic growth rate is a important value which reveals the growth of the entire population.

If the stability conditions previously mentioned are met, then it can be proven that the population will tend to the Stable Population Equivalent (Caswell [Cas01]). Some algebraic properties are needed to prove this.

Let A be a squared matrix of size n, where $n - 1$ is the maximum age achievable by a population. A will be the transition matrix of the population and it has the following shape

$$A = \left(\begin{array}{c|c|c|c} 0 & A_1 & 0 & 0 \\ \hline A_2 & 0 & 0 & 0 \\ \hline 0 & A_3 & 0 & 0 \\ \hline 0 & 0 & A_4 & 0 \end{array}\right) \tag{2.2.5.8}$$

where

A_1 is a positive line vector of size 35 corresponding to the the usually defined fertile period ranging from the age of 15 until the age of 49;

A_2 is a diagonal matrix with positive diagonal entries of size 15 with the survival rates $(1 - q_x)$ of the pre-fertile women;

A_3 is a diagonal matrix with positive diagonal entries of size 35 with the survival rates $(1 - q_x)$ of the fertile women;

A_4 is a diagonal matrix with positive diagonal values of size $n - 1 - 50$ with the survival rates $(1 - q_x)$ of the post-fertile women until the age of $n - 2$.

As it is specified by the stability conditions, all the values of A are constant. The women of post-fertile age do not influence the age structure of the population, therefore the study of the long term population should be directed to the pre-fertile and fertile women. Let A_{50} be the sub-matrix of A with the first 50 columns and rows

$$A_{50} = \left(\begin{array}{c|c} 0 & A_1 \\ \hline A_2 & 0 \\ \hline 0 & A_3 \end{array}\right) \tag{2.2.5.9}$$

From Cohen [Coh79], if A_{50} is primitive (Proposition 1), then it has a positive and real eigenvalue λ_1 that is greater than all norms of the remaining eigenvalues, that is $\lambda_1 > ||\lambda_i||, i \neq 1$. Furthermore, the author proves that the normalized eigenvector w_1 of the eigenvalue λ_1 is equal to the age proportions of the Stable Population Equivalent and also that $\lambda_1 = e^r$. As known from dynamical systems, in the long term, the eigenvector associated with the dominant eigenvalue will become the main direction of the system. It is said that the population will tend towards its Stable Population Equivalent.

So all that remains is to prove that A_{50} is primitive. The following propositions and definition follow Caswell [Cas01].

Proposition 1

A matrix is primitive if it is irreducible and if the greatest common divisor of the lengths of the cycles of its corresponding graph is 1.

Definition 1

A graph's cycle is a path starting in a node to and ending on itself.

Proposition 2

A matrix is irreducible if and only if its graph contains a path from every node to every other node.

Since A_1 is a positive vector, it connects every node with corresponding age between 15 and 49 to the age 0 corresponding node. Also since A_2 and A_3 are matrices with positive diagonals, then it connects every node to the next age corresponding node. So A_{50} is irreducible.

Since for the greatest common divisor of the length of the cycles of A_{50}'s graph, it is enough to notice that the graph has at least 35 cycles and from those there are at least 9 with length size equal to a prime number. As the greatest common divisor between two prime numbers is 1, then the greatest common divisor of the lengths of the graph's cycle is 1. And therefore A_{50} is primitive.

From the Stable Population Equivalent it is easy to define the Stationary Population Equivalent, which is the same as of the Stable Population Equivalent with an extra restriction, the intrinsic growth rate (r) must be equal to zero. This model is characterized by maintaining the entire population static, i. e., the number of persons of a given age and gender is the same every year.

2.2.6 Age Structure and Growth of Populations

The equation that generates the Stable Population value shows that the demographic parameters are highly correlated. The age-specific growth rates incorporate the relevant effects of past variations of fertility and mortality (Campos [Cam00], Coale and Trussell [CT96]).

Also they were used to improve the construction of life tables in certain situations, to show the importance of each demographic variable in population aging and even, to demonstrate the relation enclosed by actual and intrinsic growth rates (Coale and Trussell [CT96]).

2.3 Social Security History

Social Security is a system that tries to assure the basic rights of the citizens without discrimination. In addition it also aims at improving the well-being and social cohesion of all citizens or foreigners that exert professional activities or simply reside in the territory as stated in Diário da Républica [07b].

In principle, the social security systems also aim at redistributing resources towards low-income groups through social pensions and subsidies. This redistribution can also be done by defining floors or ceilings to the various social classes.

The principle of defining funds to support the elderly can be found since the Roman Empire, where individuals could make a single large payment to the annua and then receive an annual payment until death. [Inv15]

During the Medieval Age, several systems that supported the elderly could also be found. These systems were available for those who were members of a guild. A guild was an organization of people with the same occupation. There were guilds for merchants, craftsmen and religious men (Kieser [Kie89]).

Pension plans in private sector are old as the 16th century in France [Mou15] or in the 18th century in Canada [Nat07] and the USA [Uni15].

In 1881, a public pension plan was presented to the Emperor of Germany by Germany's Chancellor, Otto von Bismarck, but this plan was adopted only in 1889, being the first fully public sector pension system [Uni15]. Previously pension systems were already defined in England (Martins [Mar10]), France [Nat12] and Brasil (Alencar [Ale09]), although they were not governed by the state.

The USA set their public sector pension plan at 1935, with taxation beginning in 1937 along with the payment of pension in the same year [Uni15]. In Portugal, only by the end of 1963 under the administration of Oliveira Salazar, the public pension system was established [Por15].

2.3.1 Earnings-Based Model

As in most social security systems, in the earnings-based model, there is an attempt to redistribute the wealth in a more egalitarian way among the pensioners. Nevertheless, in the formulation of this model a usual work trend was forgotten. This model only takes into account the last years earnings of a person to calculate the retirement pension. Naturally a person tends to earn more money on the later years of work due to career progression. So this pension system becomes overgenerous to individuals with steeper earnings profiles, which are typically high earners (Borella [Bor01]).

The general formula for earnings-based model is

$$P_{\text{EB}} = \alpha\beta\frac{1}{\delta}\sum_{i=1}^{\delta}\mathbf{w}_i \tag{2.3.1.1}$$

where

P_{EB} is the value of the retirement pension;
α is the number of years on which the individual has been active in the labor market;
β is a scaling coefficient defined by the social security system ($\beta \in [0, 1]$);
δ is the number of working years used for the computation of the retirement pension;
$\mathbf{w}$ is a vector of size δ with the remunerations of the last δ years of work.

2.3.2 Contributions-Based Model

This social security model does not try to redistribute resource towards low-income groups but instead tries to return a fair amount of what a person invested in the social security through his contributions. The pension of a given person calculated by the contributions-based model uses all past contributions to compute the value of the retirement pension. The past contributions are corrected taking into account the yearly wealth increase. The total value is also adjusted for the life expectancy at the age of retirement (Belloni and Maccheroni [BM13]).

The general formula for contributions-based model is

$$P_{\text{CB}} = \left(\sum_{i=0}^{\text{age}} c_i (1+g)^{(\text{age}-i)} \right) \delta_{\text{age}} \tag{2.3.2.1}$$

where

P_{CB} is the value of the retirement pension;
c_i is the value of the contribution in the year i;
δ_{age} is a conversion factor; Mazzaferro and Morciano [MM12] defined it by the age of entry in the retirement system as

$$\delta_{\text{age}} = \left(\frac{\sum\limits_{j=m,f} dir_{\text{age},j} + ind_{\text{age},j}}{2} - k \right)^{-1}$$

with an actuarial adjustment factor k defined by the social security system and

$$dir_{\text{age},j} = \sum_{t=0}^{\Omega-\text{age}} \frac{l_{\text{age}+t,j}}{l_{\text{age}.j}} (1+g)^{-1}$$

$$ind_{\text{age},j} = \theta \sum_{t=o}^{\Omega-\text{age}} \frac{l_{\text{age}+t,j}}{l_{\text{age},j}} \left(1 - \frac{l_{\text{age}+t+1,j}}{l_{\text{age}+t,j}} \right) (1+g)^{-(t+1)} a^{W}_{\text{age}+t+1}$$

where

Ω is the maximum age that is expected a person to reach;
θ is the fraction of the retirement pension paid to the widow(er);
$a^{W}_{\text{age}+t+1}$ is the expected present value of a unitary annuity paid to the widow(er) at time $age+t+1$;
g is the expected growth of the gross domestic product.

2.3.3 PAYG Model

The PAYG (Pay As You Go) social security model is a model that requires no initial capital nor does it works based on the premise that previous contributions will pay for the retirement pension. This model also may or may not redistribute the wealth of the population. This system is an one year step balance between contributions and pensions (Ediev [Edi14]).

Let $P_{\text{PAYG}}(t)$ and $C_{\text{PAYG}}(t)$ be the the PAYG total pensions and total contributions defined as

$$P_{\text{PAYG}}(t) = \sum p_t \qquad C_{\text{PAYG}}(t) = \sum c_t.$$

As the balance is derived from

$$P_{\text{PAYG}}(t) = C_{\text{PAYG}}(t),$$

one of the three following choices has to be made every year:

- To fix c_t and vary p_t, keeping the yearly contributions constant and to recompute the pensions based on the available contributions;

- To fix p_t and vary c_t, keeping the yearly pensions constant and to recompute the contributions based on the vowed pensions;
- To vary both c_t and p_t, in order to keep $P_{\text{PAYG}}(t) = C_{\text{PAYG}}(t)$ in a fair manner for both tax payers and retired population.

Chapter 3

Population Model

An agent-based model for a Human population will be presented here. The basic functionalities are offered, i. e., an agent may give birth to another agent, may die or may leave/enter the country. An agent may get several other functionalities as well. In the case to be presented, the model was applied to the Portuguese population to test the sustainability of the Portuguese Social Security System. Therefore, this model requires the agents to have an activity, employment and retirement status. Additionally, they need to have an education and working proficiency. Finally, they must have a remuneration and pay their social security contributions, if they are working, or they must receive a retirement pension if they are retired.

In this chapter, the main population model will be presented, followed by a set of possible functions for the management of the births, deaths and migrations.

3.1 Closed Population Model

A closed population model is a population model in which it is assumed that there are no migrations, or that the net migration is 0. Therefore, only the fertility and mortality are considered.

The variables used in the closed model that will be presented can be computed using multiple methods. The simplest way to compute and understand them will be presented; however, during programming, other methods were used in order to enhance the performance speed and to reduce the RAM usage.

Firstly, the notation and the main components of the model are established. Variables starting with A, G and X correspond to agent variables, global variables and random variables, respectively. The indices a, s, k and y are used to denote age, sex, agent identification and year, respectively. Any variable indexed by a, s, k, y represents the realization of the variable in the agent k, aged a years-old and from sex s, in the year y. A similar interpretation applies to any subset of these indices. The following variables are then defined:

y_0, y_f: starting and finishing simulation years;

$A^{\text{Alive}}_{\text{a,s,k,y}}$: vital status, associated with a binary code where $A^{\text{Alive}}_{\text{a,s,k,y}} = 1$ means that the agent is alive, while $A^{\text{Alive}}_{\text{a,s,k,y}} = 0$ means that the agent is dead. Clearly if an agent is dead in a given year, then that agent will remain death in the following year, $A^{\text{Alive}}_{\text{a,s,k,y}} = 0 \Rightarrow A^{\text{Alive}}_{\text{a+1,s,k,y+1}} = 0$;

$G^{\text{Alive}}_{\text{a,s,y}}$: number of living agents and is given by

$$G^{\text{Alive}}_{\text{a,s,y}} = \sum_{\text{k}} A^{\text{Alive}}_{\text{a,s,k,y}}; \tag{3.1.0.1}$$

$G^{\text{MaleFreq}}_{\text{y}}$: relative frequency of male agents, which corresponds to total alive male agents divided by the total alive population,

$$G^{\text{MaleFreq}}_{\text{y}} = \frac{\sum_{\text{a}} G^{\text{Alive}}_{\text{a,M,y}}}{\sum_{\text{a,s}} G^{\text{Alive}}_{\text{a,s,y}}}; \tag{3.1.0.2}$$

$G^{\text{Births}}_{\text{a,s,y}}$: number of births of sex s, obtained from female agents aged a years-old;

$G^{\text{Deaths}}_{\text{a,s,y}}$: number of deaths. If an agent's age is greater than 0, then he/she must have been alive in the previous year and so, the value is the count of agents which were alive in the year y−1 and are death in the year y. If the age is equal to 0, then the number of deaths is the difference between the total number of agents born in the year y and the total number of agents of age 0 alive at the end of year y,

$$G^{\text{Deaths}}_{\text{a,s,y}} = \#\{A^{\text{Alive}}_{\text{a,s,k,y}} = 0 \wedge A^{\text{Alive}}_{\text{a}-1\text{,s,k,y}-1} = 1\}, a \neq 0, \tag{3.1.0.3}$$

$$G^{\text{Deaths}}_{\text{0,s,y}} = \sum_{\text{a}} G^{\text{Births}}_{\text{a,s,y}} - G^{\text{Alive}}_{\text{0,s,y}}. \tag{3.1.0.4}$$

Different update methods were used for the fertility and mortality in order to find the ones whose results closest to the reality. For the fertility case, there is a mic model in section 3.2.1, a mac model in section 3.2.2, a mic-mac model in section 3.2.3 on the following page and an agent heterogeneity model in section 3.2.4 on the following page. For the mortality case there is a mic model in section 3.3.1 on page 22, a mac model in section 3.3.2 on page 22, a mic-mac Model in section 3.3.3 on page 22 and an agent meterogeneity Model in section 3.3.4 on page 23.

3.2 Fertility

3.2.1 Mic Model

A simple micro model is defined here, where the next year fertility rate is computed using the data generated from the current simulated year.

Let

$G^{\text{FertR}}_{\text{a,y+1}}$ be the fertility rate indexed by the age a of the mother and the birth year y plus one; it is given by

$$G^{\text{FertR}}_{\text{a,y+1}} = \frac{\sum_{\text{s}} G^{\text{Births}}_{\text{a,s,y}}}{\sum_{\text{s}} \left(G^{Deaths}_{\text{a,s,y}} + G^{\text{Alive}}_{\text{a,s,y}}\right)}. \tag{3.2.1.1}$$

3.2.2 Mac Model

A simple macro model is defined here, where the next year fertility rate is computed using the current fertility rate and the expected fertility rate growth.

Let

G^{FertEvo} be a vector of size equal to the number of years to be simulated. This vector is composed by the expected mean fertility rate growth for those years and $G^{\mathrm{FertEvo}}_{\mathrm{y}}$ are G^{FertEvo} elements;

$G^{\mathrm{FertR}}_{\mathrm{a,y+1}}$ be the fertility rate indexed by the age a of the mother and the birth year plus one with

$$G^{\mathrm{FertR}}_{\mathrm{a,y+1}} = G^{\mathrm{FertR}}_{\mathrm{a,y}} \times G^{\mathrm{FertEvo}} \quad \text{and} \quad G^{\mathrm{FertR}}_{\mathrm{a,y_0}} = \frac{\sum\limits_{\mathrm{s}} G^{\mathrm{Births}}_{\mathrm{a,s,y_0}\,-\,1}}{\sum\limits_{\mathrm{s}} \left(G^{\mathit{Deaths}}_{\mathrm{a,s,y_0}\,-\,1} + G^{\mathrm{Alive}}_{\mathrm{a,s,y_0}\,-\,1} \right)}. \qquad (3.2.2.1)$$

3.2.3 Mic-Mac Model

A mic-mac model is defined here, where the next year fertility rate is computed using the current year generated data. The expected growth of the fertility rate is also used as a controlling factor, forcing the overall fertility rate to grow according to controlling factor.

Let

G^{FertEvo} be a vector with size equal to the number of years to be simulated. This vector is composed of the expected mean fertility rate growth for those years and $G^{\mathrm{FertEvo}}_{\mathrm{y}}$ are G^{FertEvo} elements;

$G^{\mathrm{FertR}}_{\mathrm{a,y+1}}$ be the fertility rate indexed by the age a of the mother and the birth year plus one, given by

$$G^{\mathrm{FertR}}_{\mathrm{a,y+1}} = \frac{\sum\limits_{\mathrm{s}} G^{\mathrm{Births}}_{\mathrm{a,s,y}}}{\sum\limits_{\mathrm{s}} \left(G^{\mathit{Deaths}}_{\mathrm{a,s,y}} + G^{\mathrm{Alive}}_{\mathrm{a,s,y}} \right)} \times G^{\mathrm{BirthEvo}}_{\mathrm{y}}. \qquad (3.2.3.1)$$

3.2.4 Agent Heterogeneity

One of the greatest advantages of agent-based models is the possibility of having heterogeneous individuals. That functionality is described and explored here.

The following variable is created in order to achieve heterogeneity among the agents:

$X^{\mathrm{FertR}}_{\mathrm{a,y+1}}$ a random variable with

$\sigma x^{\mathrm{FertR}}_{\mathrm{a,y+1}} = min\{0.02, \frac{G^{\mathrm{FertR}}_{\mathrm{a,y+1}}}{3}, \frac{1-G^{\mathrm{FertR}}_{\mathrm{a,y+1}}}{3}\}$ such that

$$X^{\mathrm{FertR}}_{\mathrm{a,y+1}} \sim N(G^{\mathrm{FertR}}_{\mathrm{a,y+1}}, \sigma x^{\mathrm{FertR}}_{\mathrm{a,y+1}}); \qquad (3.2.4.1)$$

This random variable must be between 0 and 1, so the standard deviation σ must be bounded. A factor of 3 is used in each fraction because a normal distributed sample has about 99.75% of its values between a range of three standard deviations from the mean. With this constraint, it is almost guaranteed that the variable will remain between 0 and 1, as desired.

With this distribution, each agent is assigned the parameter:

$A^{\mathrm{FertR}}_{\mathrm{a+1,k,y+1}}$ which is the probability for the female agent k, aged a+1 years-old, to give birth in the year y+1; it is given by

$$A^{\mathrm{FertR}}_{\mathrm{a+1,k,y+1}} = x^{\mathrm{FertR}}_{\mathrm{a+1,y+1}}, \quad x^{\mathrm{FertR}}_{\mathrm{a+1,y+1}} \in X^{\mathrm{FertR}}_{\mathrm{a+1,y+1}}. \qquad (3.2.4.2)$$

3.3 Mortality

3.3.1 Mic Model

A simple micro model is defined here, where the next year mortality rate is computed using the data generated on the current simulated year.

Let

$G^{\text{MortR}}_{\text{a,s,y+1}}$ be the mortality rate at year y+1 for agents aged a years-old and with sex s; it is given by

$$G^{\text{MortR}}_{\text{a,s,y+1}} = \frac{G^{\text{Deaths}}_{\text{a,s,y}}}{G^{\text{Deaths}}_{\text{a,s,y}} + G^{\text{Alive}}_{\text{a,s,y}}}. \tag{3.3.1.1}$$

3.3.2 Mac Model

A simple macro model is defined here, where next year mortality rate is computed using the current mortality rate and the expected mortality rate growth.

Let

G^{MortEvo} be a vector of size equal to the number of years to be simulated. It is composed of the expected mean mortality rate growth for those years and $G^{\text{MortEvo}}_{\text{y}}$ are G^{MortEvo} elements;

$G^{\text{MortR}}_{\text{a,s,y+1}}$ be the mortality rate at year y+1 for agents aged a years-old and with sex s; it is given by

$$G^{\text{MortR}}_{\text{a,s,y+1}} = G^{\text{MortR}}_{\text{a,s,y}} \times G^{\text{MortEvo}}_{\text{y}} \text{ and} \tag{3.3.2.1}$$

$$G^{\text{MortR}}_{\text{a,s,y}_0} = \frac{G^{\text{Deaths}}_{\text{a,s,y}_0 - 1}}{G^{Deaths}_{\text{a,s,y}_0 - 1} + G^{\text{Alive}}_{\text{a,s,y}_0 - 1}}. \tag{3.3.2.2}$$

3.3.3 Mic-Mac Model

A mic-mac model is defined here, where the next year mortality rate is computed using the current year generated data. The expected mortality rate growth is also used as a controlling factor, forcing the overall mortality rate to grow according to controlling factor.

Let

G^{MortEvo} be a vector of size equal to the number of years to be simulated. It is composed of the expected mean mortality rate growth for those years and $G^{\text{MortEvo}}_{\text{y}}$ are G^{MortEvo} elements;

$G^{\text{MortR}}_{\text{a,s,y+1}}$ be the mortality rate at year y+1 for agents of age a and sex s with

$$G^{\text{MortR}}_{\text{a,s,y+1}} = \frac{G^{\text{Deaths}}_{\text{a,s,y}}}{G^{\text{Deaths}}_{\text{a,s,y}} + G^{\text{Alive}}_{\text{a,s,y}}} \times G^{\text{MortEvo}}_{\text{y}} \tag{3.3.3.1}$$

If, for some characteristics, the population size is very small, then this formula is replaced by the correspondent mac model formula.

3.3.4 Agent Heterogeneity

Mortality agent heterogeneity is made in a completely analogous method to the fertility counterpart (3.2.4 on page 21). Then

$A^{\text{MortR}}_{\text{a+1,s,k,y+1}}$ which is the probability for the agent k of aged a+1 years-old and with sex s to die in the year y+1; it is given by

$$A^{\text{MortR}}_{\text{a+1,s,k,y+1}} = x^{\text{MortR}}_{\text{a+1,s,y+1}}, \quad x^{\text{MortR}}_{\text{a+1,s,y+1}} \in X^{\text{MortR}}_{\text{a+1,s,y+1}}; \tag{3.3.4.1}$$

3.4 Algorithm

The previous variables have to be updated at each time step. In real life any of the previously stated occurrences may happen in any given time. However, to maintain the control over the evolution, an update order needs to be forced.

The evolution process is done according to the following steps:

Step 1. Increase the simulation year by one;

Step 2. Age every living agents by one;

Step 3. Give birth to new agents according to the birth rates of the agents, i.e., two values u_1 and u_2 are randomly sampled from $U(0,1)$ and if $u_1 < A^{\text{FertR}}_{\text{a,s,k,y}}$ then a new agent is born. If $u_2 < G^{\text{MaleFreq}}_{\text{y}}$ then the agent is a set as male, otherwise it is set as female.

Step 4. Randomly ”kill” agents, i.e., u_3 is randomly sampled from $U(0,1)$ and if $u_3 < A^{\text{MortR}}_{\text{a,s,k,y}}$ then set $A^{\text{Alive}}_{\text{a,s,k,y}} = 0$.

Step 5. Compute the next year birth and death parameters and male proportion rates according to the chosen update model.

Step 6. Define each agent’s birth and death parameters for the following year.

3.5 Migration

3.5.1 Emigration Model

Since, in agent-based models, it is possible to map each intervenor individually, the modeling of the emigration is based on the will of each agent. Each agent determines his/her gain when moving to another country and the gain when staying in Portugal. This gain is computed using several variables as suggested by Baláž, Williams, and Fifeková [BWF14], namely health and safety indicators as well as wages and living costs. In addition the number of Portuguese people that were born in Portugal and that are living in each considered country, the distance of each country to Portugal and the spoken language in the foreign country is also used to compute the gains, as already studied by Anjos and Campos [AC10]. Moreover, observation of Portuguese data reveals that emigration is dependent on age and common sense dictates that work success is also an important factor in the emigration decision. So these two parameters are also considered, but they will not be used to directly model the gain in emigrating but to model the will to emigrate. The considered countries were the ones with more than 50 emigrants in 2011.

The following variables are considered. Henceforth, c is an index denoting a given country while c_0 denotes Portugal.

$G_{\mathrm{c}}^{\mathrm{Health}}$ is a health indicator with $A_{\mathrm{k}}^{\mathrm{HealthW}}$ as its corresponding weight;
$G_{\mathrm{c}}^{\mathrm{Safety}}$ is a safety indicator with $A_{\mathrm{k}}^{\mathrm{SafetyW}}$ as its corresponding weight;
$G_{\mathrm{c,l}}^{\mathrm{Wage}}$ is a wage indicator with $A_{\mathrm{k}}^{\mathrm{WageW}}$ as its corresponding weight;
$G_{\mathrm{c}}^{\mathrm{Pop}}$ is an indicator for the Portuguese population size, with $A_{\mathrm{k}}^{\mathrm{PopW}}$ as its corresponding weight;
$G_{\mathrm{c}}^{\mathrm{Dist}}$ is an indicator for the distance to Portugal, with $A_{\mathrm{k}}^{\mathrm{DistW}}$ as its corresponding weight;
$G_{\mathrm{c}}^{\mathrm{Lang}}$ is an indicator for the Portuguese language, with $A_{\mathrm{k}}^{\mathrm{LangW}}$ as its corresponding weight;
$G_{\mathrm{c}}^{\mathrm{Limit}}$ is the Portuguese emigration limit, defined by the destination country;
$G_{\mathrm{c}}^{\mathrm{ECounter}}$ is a counter for the number of emigrants.

The first five indicators range between 0 and 1. The used indicator must be the same for all countries and it is preferable that the data source is the same, because the same indicator may vary in different sources. $G_{\mathrm{c}}^{\mathrm{Lang}}$ equals 1 if Portuguese is the native language and 0 otherwise. The wage indicator changes every year according to the country expected mean wage growth.

For the parameter $G_{\mathrm{c}}^{\mathrm{limit}}$ there is no weight because this parameter cannot be influenced by any agent. For each weight a variable is created in order to achieve heterogeneity:

$X_{\mathrm{k}}^{\mathrm{HealthW}}$ for the health weight such that $X_{\mathrm{k}}^{\mathrm{HealthW}} \sim N(0, 0.75 \times A_{\mathrm{k}}^{\mathrm{HealthW}})$;
$X_{\mathrm{k}}^{\mathrm{SafetyW}}$ for the safety weight such that $X_{\mathrm{k}}^{\mathrm{SafetyW}} \sim N(0, 0.75 \times A_{\mathrm{k}}^{\mathrm{SafetyW}})$;
$X_{\mathrm{k,s}}^{\mathrm{WageW}}$ for the wage weight such that $X_{\mathrm{k,e}}^{\mathrm{WageW}} \sim N(0, 0.75 \times A_{\mathrm{k,e}}^{\mathrm{WageW}})$;
$X_{\mathrm{k}}^{\mathrm{PopW}}$ for the Portuguese population size weight such that $X_{\mathrm{k}}^{\mathrm{PopW}} \sim N(0, 0.75 \times A_{\mathrm{k}}^{\mathrm{PopW}})$;
$X_{\mathrm{k}}^{\mathrm{LangW}}$ for the spoken language weight such that $X_{\mathrm{k}}^{\mathrm{LangW}} \sim N(0, 0.75 \times A_{\mathrm{k}}^{\mathrm{LangW}})$;

and then the weights are recomputed as

$$A_{\mathrm{k}}^{\mathrm{HealthW}} = A_{\mathrm{k}}^{\mathrm{HealthW}} + x_{\mathrm{k}}^{\mathrm{HealthW}}, x_{\mathrm{k}}^{\mathrm{HealthW}} \sim X_{\mathrm{k}}^{\mathrm{HealthW}}; \tag{3.5.1.1}$$
$$A_{\mathrm{k}}^{\mathrm{SafetyW}} = A_{\mathrm{k}}^{\mathrm{SafetyW}} + x_{\mathrm{k}}^{\mathrm{SafetyW}}, x_{\mathrm{k}}^{\mathrm{SafetyW}} \sim X_{\mathrm{k}}^{\mathrm{SafetyW}}; \tag{3.5.1.2}$$
$$A_{\mathrm{k,e}}^{\mathrm{WageW}} = A_{\mathrm{k,e}}^{\mathrm{WageW}} + x_{\mathrm{k,e}}^{\mathrm{WageW}}, x_{\mathrm{k,e}}^{\mathrm{WageW}} \sim X_{\mathrm{k,e}}^{\mathrm{WageW}}; \tag{3.5.1.3}$$
$$A_{\mathrm{k}}^{\mathrm{PopW}} = A_{\mathrm{k}}^{\mathrm{PopW}} + x_{\mathrm{k}}^{\mathrm{PopW}}, x_{\mathrm{k}}^{\mathrm{PopW}} \sim X_{\mathrm{k}}^{\mathrm{PopW}}; \tag{3.5.1.4}$$
$$A_{\mathrm{k}}^{\mathrm{LangW}} = A_{\mathrm{k}}^{\mathrm{LangW}} + x_{\mathrm{k}}^{\mathrm{LangW}}, x_{\mathrm{k}}^{\mathrm{LangW}} \sim X_{\mathrm{k}}^{\mathrm{LangW}}. \tag{3.5.1.5}$$

One more parameter is used, in order to take into account age and employment history .

Also let

$A_{\mathrm{a,k,y}}^{\mathrm{Will}}$ be the will to emigrate from Portugal of agent k with age a in the year y.

This parameter is allowed to change every year, depending on the agent's current age and work success in the previous year.

From Portuguese data concerning the age distribution of emigrants, a resized Weibull probability density function was found to be the function that best describes the emigration behavior depending on age as in Section 5.2 on page 39. The update of the will of each agent needed a calculation of the derivative of the Weibull probability density function. Let

$$W(x;\lambda,k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(\frac{x}{\lambda})^k} \tag{3.5.1.6}$$

be the probability density function of the Weibull distribution, with parameters λ and k. The derivative of the probability density function is then given by

$$\begin{aligned}
w(x;\lambda,k) &= \frac{dW(x;\lambda,k)}{dx} && (3.5.1.7)\\
&= \frac{k}{\lambda}\left[(k-1)\frac{x^{k-2}}{\lambda^{k-1}}e^{-(\frac{x}{\lambda})^k} - \left(\frac{x}{\lambda}\right)^{k-1}\frac{ke^{-(\frac{x}{\lambda})^k}(\frac{x}{\lambda})^k}{x}\right] && (3.5.1.8)\\
&= \frac{k}{\lambda}e^{-(\frac{x}{\lambda})^k}\left[(k-1)\frac{x^{k-2}}{\lambda^{k-1}} - \frac{x^{k-2}}{\lambda^{k-1}}k\left(\frac{x}{\lambda}\right)^k\right] && (3.5.1.9)\\
&= \frac{k}{\lambda}e^{-(\frac{x}{\lambda})^k}\frac{x^{k-2}}{\lambda^{k-1}}\left[(k-1) - k\left(\frac{x}{\lambda}\right)^k\right] && (3.5.1.10)\\
&= \frac{kx^{k-2}}{\lambda^k}e^{-(\frac{x}{\lambda})^k}\left[k\left(1-\left(\frac{x}{\lambda}\right)^k\right) - 1\right] && (3.5.1.11)\\
& && (3.5.1.12)
\end{aligned}$$

The shape (λ) and scale (k) parameters for the Weibull distribution function are estimated by statistical fitting, as shown in Section 5.2 on page 39. Values are also jittered, and three variables are created: two variables to perform dilatation in the x and y-axis of the function and a third variable to define a base level for the will to emigrate.

Defining:
$A_{\mathrm{k}}^{\mathrm{Shape}}$ as the shape parameter;
$A_{\mathrm{k}}^{\mathrm{Scale}}$ as the scale parameter;
$A_{\mathrm{k}}^{\mathrm{x\text{-}axis}}$ as the x-axis dilatation parameter;
$A_{\mathrm{k}}^{\mathrm{y\text{-}axis}}$ as the y-axis dilatation parameter;
$A_{\mathrm{k}}^{\mathrm{Base}}$ as the base level parameter;
as well as the following random variables:
$X_{\mathrm{k}}^{\mathrm{Shape}}$ for the shape parameter such that $X_{\mathrm{k}}^{\mathrm{Shape}} \sim N(0, 0.1 \times A_{\mathrm{k}}^{\mathrm{Shape}})$;
$X_{\mathrm{k}}^{\mathrm{Scale}}$ for the scale parameter such that $X_{\mathrm{k}}^{\mathrm{Scale}} \sim N(0, 0.1 \times A_{\mathrm{k}}^{\mathrm{Scale}})$;
$X_{\mathrm{k}}^{\mathrm{x\text{-}axis}}$ for the x-axis dilatation parameter such that $X_{\mathrm{k}}^{\mathrm{x\text{-}axis}} \sim N(0, 0.1 \times A_{\mathrm{k}}^{\mathrm{x\text{-}axis}})$;
$X_{\mathrm{k}}^{\mathrm{y\text{-}axis}}$ for the y-axis dilatation parameter such that $X_{\mathrm{k}}^{\mathrm{y\text{-}axis}} \sim N(0, 0.1 \times A_{\mathrm{k}}^{\mathrm{y\text{-}axis}})$;
$X_{\mathrm{k}}^{\mathrm{Base}}$ for the base level parameter such that $X_{\mathrm{k}}^{\mathrm{Base}} \sim N(0, 0.1 \times A_{\mathrm{k}}^{\mathrm{Base}})$.

Then the main parameters are recomputed as

$$\begin{aligned}
A_{\mathrm{k}}^{\mathrm{Shape}} &= A_{\mathrm{k}}^{\mathrm{Shape}} + x_{\mathrm{k}}^{\mathrm{Shape}}, x_{\mathrm{k}}^{\mathrm{Shape}} \sim X_{\mathrm{k}}^{\mathrm{Shape}}; && (3.5.1.13)\\
A_{\mathrm{k}}^{\mathrm{Scale}} &= A_{\mathrm{k}}^{\mathrm{Scale}} + x_{\mathrm{k}}^{\mathrm{Scale}}, x_{\mathrm{k}}^{\mathrm{Scale}} \sim X_{\mathrm{k}}^{\mathrm{Scale}}; && (3.5.1.14)\\
A_{\mathrm{k}}^{\mathrm{x\text{-}axis}} &= A_{\mathrm{k}}^{\mathrm{x\text{-}axis}} + x_{\mathrm{k}}^{\mathrm{Shape}}, x_{\mathrm{k}}^{\mathrm{x\text{-}axis}} \sim X_{\mathrm{k}}^{\mathrm{x\text{-}axis}}; && (3.5.1.15)\\
A_{\mathrm{k}}^{\mathrm{y\text{-}axis}} &= A_{\mathrm{k}}^{\mathrm{y\text{-}axis}} + x_{\mathrm{k}}^{\mathrm{Scale}}, x_{\mathrm{k}}^{\mathrm{y\text{-}axis}} \sim X_{\mathrm{k}}^{\mathrm{y\text{-}axis}}; && (3.5.1.16)\\
A_{\mathrm{k}}^{\mathrm{Base}} &= A_{\mathrm{k}}^{\mathrm{Base}} + x_{\mathrm{k}}^{\mathrm{Base}}, x_{\mathrm{k}}^{\mathrm{Base}} \sim X_{\mathrm{k}}^{\mathrm{Base}}; && (3.5.1.17)\\
& && (3.5.1.18)
\end{aligned}$$

For these parameters, a standard deviation of 10% of the base value is used because they are very sensitive to small changes in comparison with the weight parameters.

Also defining

$A_{\text{k,l}}^{\text{Success}}$ as a work success parameter, with

$$A_{\text{k,y}}^{\text{Success}} = \begin{cases} -1 & \text{, if } A_{\text{a,s,k,y}-1}^{\text{Emp}} = 1 \\ 0 & \text{, if } A_{\text{a,s,k,y}-1}^{\text{Act}} = 0 \\ 1 & \text{, if } A_{\text{a,s,k,y}-1}^{\text{Emp}} = 0 \end{cases} \tag{3.5.1.19}$$

$A_{\text{k}}^{\text{SuccessW}}$ as the weight given to work success $A_{\text{k,y}}^{\text{Success}}$.

Like the previous weights, this will suffer a 75% random variation for each agent so, let

$X_{\text{k}}^{\text{SuccessW}}$ be a random variable for the work success weight such that

$$X_{\text{k}}^{\text{SuccessW}} \sim N(0, 0.75 \times A_{\text{k}}^{\text{SuccessW}}). \tag{3.5.1.20}$$

Then the work success weight is recomputed as

$$A_{\text{k}}^{\text{SuccessW}} = A_{\text{k}}^{\text{SuccessW}} + x_{\text{k}}^{\text{SuccessW}}, x_{\text{k}}^{\text{SuccessW}} \sim X_{\text{k}}^{\text{SuccessW}}. \tag{3.5.1.21}$$

Finally the will to emigrate can be computed as

$$A_{\text{a+1,k,y+1}}^{\text{Will}} = A_{\text{k}}^{\text{y-axis}} \times w(A_{\text{k}}^{\text{x-axis}} \times i; A_{\text{k}}^{\text{Shape}}, A_{\text{k}}^{\text{Scale}}) + A_{\text{k}}^{\text{SuccessW}} \times A_{\text{k,y}}^{\text{Success}} \tag{3.5.1.22}$$

and

$$A_{\text{a,k,y}_0}^{\text{Will}} = A_{\text{k}}^{\text{y-axis}} \times W(A_{\text{k}}^{\text{x-axis}} \times i; A_{\text{k}}^{\text{Shape}}, A_{\text{k}}^{\text{Scale}}) + A_{\text{k}}^{\text{Base}}. \tag{3.5.1.23}$$

Then, for each country, a gain is computed for either emigrating or staying in the home country. The emigration gain is given by

$$\begin{align} A_{\text{c,a,k,y}}^{\text{Gain}} = \Big(A_{\text{k}}^{\text{HealthW}} \times G_{\text{c}}^{\text{Health}} + A_{\text{k}}^{\text{SafetyW}} \times G_{\text{c}}^{\text{Safety}} + A_{\text{k,e}}^{\text{WageW}} \times G_{\text{c,e}}^{\text{Wage}} + \tag{3.5.1.24} \\ + A_{\text{k}}^{\text{PopW}} \times G_{\text{c}}^{\text{Pop}} + A_{\text{k}}^{\text{DistW}} \times G_{\text{c}}^{\text{Dist}} + A_{\text{k}}^{\text{LangW}} \times G_{\text{c}}^{\text{Lang}} \Big) \times A_{\text{a,k,y}}^{\text{Will}} \tag{3.5.1.25} \end{align}$$

while the gain to stay in Portugal is

$$\begin{align} A_{\text{c,a,k,y}}^{\text{Gain}} = \Big(A_{\text{k}}^{\text{HealthW}} \times G_{\text{c}}^{\text{Health}} + A_{\text{k}}^{\text{SafetyW}} \times G_{\text{c}}^{\text{Safety}} + A_{\text{k,e}}^{\text{WageW}} \times G_{\text{c,e}}^{\text{Wage}} + \tag{3.5.1.26} \\ + A_{\text{k}}^{\text{PopW}} \times G_{\text{c}}^{\text{Pop}} + A_{\text{k}}^{\text{LangW}} \times G_{\text{c}}^{\text{Lang}} \Big) \times \Big(1 - A_{\text{a,k,y}}^{\text{Will}} \Big) \tag{3.5.1.27} \end{align}$$

The $G_{\text{c,y,e}}^{\text{Wage}}$ parameter is updated assuming that the net migration on the foreign countries remains constant as well as the number of jobs.

The following variables are defined for each destination country of the Portuguese emigration:

$G_{\text{c,y}}^{\text{ForPop}}$ as the foreign population size;
$G_{\text{c,y}}^{\text{ForEmpR}}$ as the foreign population employment rate;
$G_{\text{c,y,e}}^{\text{ForEmpR}}$ as the foreign population employment rate indexed by education;
$G_{\text{c,y}}^{\text{ForVarR}}$ as the foreign population employment rate variation;

G_c^{ForEmp} as the existing jobs for foreign population, which is given by

$$G_c^{\mathrm{ForEmp}} = G_{c,y_0}^{\mathrm{ForPop}} \times G_{c,y_0}^{\mathrm{ForEmpR}} \tag{3.5.1.28}$$

G_c^{NetMig} as the net migration;
G_c^{MWage} as the mean wage for foreign population;
G_c^{PPP} as the Purchase Power Parity.

So the following values can be computed:

$$G_{c,l}^{\mathrm{ForPop}} = G_{c,l-1}^{\mathrm{ForPop}} + G_c^{\mathrm{NetMig}}; \tag{3.5.1.29}$$

$$G_{c,y}^{\mathrm{ForEmpR}} = \frac{G_c^{\mathrm{ForEmp}}}{G_{c,y}^{\mathrm{ForPop}}}; \tag{3.5.1.30}$$

$$G_{c,y}^{\mathrm{ForVarR}} = \frac{G_{c,y}^{\mathrm{ForEmpR}}}{G_{c,y-1}^{\mathrm{ForEmpR}}}; \tag{3.5.1.31}$$

$$G_{c,y,e}^{\mathrm{ForEmpR}} = G_{c,y-1,e}^{\mathrm{ForEmpR}} \times G_{c,y}^{\mathrm{ForVarR}} \tag{3.5.1.32}$$

$$G_{c,y,e}^{\mathrm{Wage}} = G_c^{\mathrm{MWage}} \times G_{c,y,e}^{\mathrm{ForEmpR}} \times G_c^{\mathrm{PPP}}. \tag{3.5.1.33}$$

Also lets define an emigration counter G_c^{ECounter}.

The emigration process is done according to the following sequence:

Step 1. Initialize the emigration counter G_c^{ECounter};
Step 2. Compute $G_{c,y,e}^{\mathrm{ForEmpR}}$ and $G_{c,y,e}^{\mathrm{Wage}}$;
Step 3. Update each agent's $A_{a,k,y}^{\mathrm{Will}}$ and $A_{c,a,k,y}^{\mathrm{Gain}}$;
Step 4. For each agent determine the desired emigration destination

$$A_k^{\mathrm{Dest}} = \arg\max_c \{A_{c,a,k,y}^{\mathrm{Gain}}\} \tag{3.5.1.34}$$

restricted to $G_{A_k^{\mathrm{Dest}}}^{\mathrm{ECounter}} < G_{A_k^{\mathrm{Dest}}}^{\mathrm{Limits}}$; add 1 to $G_{A_k^{\mathrm{Dest}}}^{\mathrm{ECounter}}$;
Step 5. Remove all agents with A_k^{Dest} different than Portugal from the simulation.

3.5.2 Immigration Model

The immigration model cannot rely on agents that are not initially considered by the model. These immigrant agents do not yet exist in the simulation, so the previous decision model does not apply.

In this model, only exogenous variables will be used. For each country with non-negative immigration in 2011, a linear regression model was fitted using as exogenous variable the year and as endogenous variable the number of immigrants. According to these estimated number of immigrants, new agents will be created.

It is important to notice that people coming from different countries have different cultures and habits. This means that agents created through immigration will not be equal to agent created through birth. Immigrant agents may be created as being of any age, from 0 to 99. The age of the immigrant agents is modeled using a Weibull distribution, as it is explained in Section 5.2 on page 39.

One further consideration is that the fertility rate of a woman is highly determined by her culture. As stated by Adsera and Ferrer [AF14], (for Canada), fertility-wise, Canada, European countries, and

United States of America, all share the same Western culture. So the findings from Canada can be easily adapted for European countries, including Portugal. Therefore, the fertility rate of the female agents created through immigration can be taken to be equal to the same of a Portuguese female agent multiplied by a factor determined by Adsera and Ferrer [AF14].

Defining

$G_{\mathrm{c,y}}^{\mathrm{Immi}}$ as the estimated number of immigrants to Portugal;
X^{ImmiAge} as the Weibull Distribution for the age of Portugal immigrants;
$G_{\mathrm{c}}^{\mathrm{ImmiProp}}$ as the male proportion of immigrants;
$G_{\mathrm{c}}^{\mathrm{FertF}}$ as the multiplying fertility factor of immigrants;
$A_{\mathrm{k}}^{\mathrm{ImmiC}}$ as the origin country of an immigrant agent.

The immigration process is now done according to the following sequence:

Step 1. Create the immigrant agents as determined by $G_{\mathrm{c,y}}^{\mathrm{Immi}}$ and define the origin country $A_{\mathrm{k}}^{\mathrm{ImmiC}}$ accordingly;

Step 2. For each newly immigrant agent randomly sample x from X^{ImmiAge} and assign x as its age;

Step 3. For each newly immigrant agent k, randomly sample u from $U(0,1)$ and if $u < G_{A_{\mathrm{k}}^{\mathrm{ImmiC}}}^{\mathrm{ImmiProp}}$ then set agent k as male;

Step 4. For each immigrant agent k, define their birth and death parameters for the following year;

Step 5. For each immigrant agent k, of age a and sex s, set $A_{a,s,k,y}^{\mathrm{BirthR}} = A_{a,s,k,y}^{\mathrm{BirthR}} \times G_{A_{\mathrm{k}}^{\mathrm{ImmiC}}}^{\mathrm{FertF}}$.

Chapter 4

Social Security Model

The social security model is an extension of the previously presented population model. The population model is used for an evolutive overview and social security parameters are included, such as activity, employment, retirement, schooling, job qualification, and remuneration. As in the previous chapter, the indexes a, s, k, y will be used to denote age, sex, agent identification and year. The parameters are defined as follows:

$A^{\text{Act}}_{\text{a,s,k,y}}$: economic activity status. $A^{\text{Act}}_{\text{a,s,k,y}} = 1$ means that the agent is active and $A^{\text{Act}}_{\text{a,s,k,y}} = 0$ means that the agent is inactive (retired, pensioner, student, etc.). For clarification, dead agents are inactive;

$G^{\text{ActR}}_{\text{a,s,y}}$: activity rate, given by

$$G^{\text{ActR}}_{\text{a,s,y}} = \frac{\sum_{\text{k}} A^{\text{Act}}_{\text{a,s,k,y}}}{\sum_{\text{k}} A^{\text{Alive}}_{\text{a,s,k,y}}} \tag{4.0.2.1}$$

$G^{\text{Act}}_{\text{a,s}}$: proportion of active individuals from the empirical data;

G^{ϵ}: a threshold parameter with $G^{\epsilon} \in [0, 1]$ fixed a priori;

$A^{\text{Emp}}_{\text{a,s,k,y}}$: employment status. $A^{\text{Emp}}_{\text{a,s,k,y}} = 1$ means that the agent is employed and $A^{\text{Emp}}_{\text{a,s,k,y}} = 0$ means that the agent is unemployed. For clarification, inactive agents are unemployed;

$G^{\text{EmpR}}_{\text{a,s,y}}$: employment rate, given by

$$G^{\text{EmpR}}_{\text{a,s,y}} = \frac{\sum_{\text{k}} A^{\text{Emp}}_{\text{a,s,k,y}}}{\sum_{\text{k}} A^{\text{Act}}_{\text{a,s,k,y}}} \tag{4.0.2.2}$$

$G^{\text{EmpProp}}_{\text{a,s}}$: proportion of employed individuals from the empirical data;

$A^{\text{Ret}}_{\text{a,s,k,y}}$: retirement status. $A^{\text{Ret}}_{\text{a,s,k,y}} = 1$ means that the agent is retired and $A^{\text{Ret}}_{\text{a,s,k,y}} = 0$ means that the agent is not retired. If G^{MinRet} is defined as the minimum allowed age for retirement, then $i < G^{\text{MinRet}} \Rightarrow A^{\text{Ret}}_{\text{a,s,k,y}} = 0$ and $i > G^{\text{MinRet}} + 10 \Rightarrow A^{\text{Ret}}_{\text{a,s,k,y}} = 1$. If the agent gets retired at some point, that status can never be reversed. For clarification, if an agent is retired then it is also inactive.

$G^{\text{RetR}}_{\text{a,s,y}}$: retirement rate, given by

$$G^{\text{RetR}}_{\text{a,s,y}} = \frac{\sum_{\text{k}} A^{\text{Ret}}_{\text{a,s,k,y}}}{\sum_{\text{k}} A^{\text{Alive}}_{\text{a,s,k,y}}} \tag{4.0.2.3}$$

$G^{\text{RetProp}}_{\text{a,s}}$: proportion of retired individuals from the empirical data.

Also, there is a need to separate agents based on their capabilities, namely education and work qualification. For these, the indeces e and j will be used to denote the educational attainment and the job qualification. Defining

$A^{\text{Edu}}_{\text{a,s,k,y}}$ as the education level of agent k. $A^{\text{Edu}}_{\text{a,s,k,y}} \in \{1,2,3,4,5,6\}$ where 1 refers to an education under the 4th grade, 2 refers to the 4th grade, 3 refers to the 6th grade, 4 refers to the 9th grade, 5 refers to high school and 6 refers to superior education;

$G^{\text{Edu}}_{\text{a,s}}$ as the empirical education rate;

$A^{\text{JQual}}_{\text{a,s,k,y}}$ as the job qualification of agent k. $A^{\text{JQual}}_{\text{a,s,k,y}} \in \{1,2,3,4,5,6,7,8\}$ where 1 refers to a qualification of apprentice or intern, 2 refers to a qualification of a not qualified professional, 3 refers to a semi-qualified professional, 4 refers to a qualified professional, 5 refers to a highly qualified professional, 6 refers to a team leader, 7 refers to a middle member of administration board and 8 refers to a high member of an administration board, as defined in the Portuguese staffing tables;

$G^{\text{JQual}}_{\text{a,s}}$ as the empirical job qualification rate;

$G^{\text{Wage}}_{\text{s,e,j}}$ as the mean wage of agents;

$A^{\text{Wage}}_{\text{a,s,k,y,e,j}}$ as the wage of agent k;

G^{GDP} as the yearly expected mean growth of the Portuguese gross domestic product until 2041

Both the activity status and the employment status are modeled because the national data on the employment rates is computed based on the employed population that it is also active.

4.1 Employment

4.1.1 Mic Model

A simple Micro model is defined here, where the current year employment rate is computed using the starting data or the data generated on the previous simulated year. Let

$G^{\text{EmplR}}_{\text{a,s,y}}$ be the employment rate of the agents in the year y, aged a years-old and with sex s, computed as

$$G^{\text{EmpR}}_{\text{a,s,y}} = \frac{\sum_{\text{k}} A^{\text{Emp}}_{\text{a,s,k,y}-1}}{\sum_{\text{k}} A^{\text{Act}}_{\text{a,s,k,y}-1}}. \qquad (4.1.1.1)$$

Then, for each agent k, a value u is randomly sampled from $U(0,1)$. If $u < G^{\text{EmpR}}_{\text{a,s,y}}$ then the agent k is set employed, else it is set unemployed.

4.1.2 Mic-Mac Model

A mic-mac model is defined here, where the current year employment rate is computed using the previous year generated data. The time between getting an employment contract and getting fired (work tenure) is set following a predetermined time distribution.

Defining

$G^{\text{FJobs}}_{\text{y}}$ as the number of available jobs;

$G^{\text{EmpR}}_{\text{y}}$ as the employment rate, which is equal to the number of available jobs at the end of the previous year divided by the number of unemployed agents:

$$G^{\text{EmpR}}_{\text{y}} = \frac{G^{\text{FJobs}}_{\text{y}-1}}{\sum\limits_{\text{k}} \left(A^{\text{Act}}_{\text{a,s,k,y}} - A^{\text{Emp}}_{\text{a,s,k,y}} \right)}. \tag{4.1.2.1}$$

$G^{\text{WTenure}}_{\text{a,s}}$ as a statistical distribution that represents the distribution of the work tenure;
$A^{\text{Contract}}_{\text{k,y}}$ as the remaining time until the end of a work contract;

In the beginning of the simulation:

For a given agent, say k, a value u is randomly sampled from $U(0, 1)$; if $u < G^{EmpR}_{\text{a,s,y}_0}$ then the agent k is set employed, else it is set unemployed; where $G^{EmpR}_{\text{a,s,y}_0}$ is computed as

$$G^{\text{EmpR}}_{\text{a,s,y}_0} = \frac{\sum\limits_{\text{k}} A^{\text{Emp}}_{\text{a,s,k,y}_0}}{\sum\limits_{\text{k}} A^{\text{Act}}_{\text{a,s,k,y}_0}}. \tag{4.1.2.2}$$

If the agent k is set employed, a value u is randomly sampled from $G^{\text{WTenure}}_{\text{a,s}}$. Then u is rounded to the considered time step size and assigned to $A^{\text{Contract}}_{\text{k,y}}$.

During the simulation:

First, the work contract of the agents will be reduced by one year, $A^{\text{Contract}}_{\text{k,y}} = A^{\text{Contract}}_{\text{k,y}-1} - 1$.
Then, if $A^{\text{Contract}}_{\text{k,y}} = 0$ and $A^{\text{Emp}}_{\text{a,s,k,y}-1} = 1$, the agent k is set unemployed ($A^{\text{Emp}}_{\text{a,s,k,y}} = 0$).
If $A^{\text{Emp}}_{\text{a,s,k,y}} = 0$, then a value u is randomly sampled from $U(0, 1)$ and if $u < G^{\text{EmpR}}_{\text{y}}$ then the agent k is set employed, $A^{\text{Emp}}_{\text{a,s,k,y}} = 1$. If the agent k is set employed, then a value u is randomly sampled from $G^{\text{WTenure}}_{\text{a,s}}$. Then u is rounded to the considered time step size and assigned to $A^{\text{Contract}}_{\text{k,s}}$.

4.2 Social Security System

In every considered social security model, each working agent contributes with a portion of its remuneration G^{EEContr} to the social security system. This is done with the expectation of an elderly pension for when the agent gets retired.

This contribution is not only made by the employee but also by the employer G^{ERContr}. Usually the latter is the one who contributes with the higher percentage, $G^{\text{EEContr}} < G^{\text{ERContr}}$. Both employee and employer contribution are based on the employee remuneration.

as the source of the contributions is not important for our purposes, for simplification, only the total contribution over a remuneration will be used: $G^{\text{Contr}} = G^{\text{EEContr}} + G^{\text{ERContr}}$.

The following variable is defined:

$A^{\text{Rem}}_{\text{k,y,e,j}}$: the remuneration of agent k. This remuneration is corrected for the last year values. In particular, for each passing year, the agent remuneration is multiplied by the expected GDP growth G^{GDP}, as a proxy for the real remuneration growth. The remunerations from before y_0 are corrected for the current values using the re-evaluation coefficients defined annually by the Portuguese Government [07a][13].

In order to simplify the next subsections, the following variables are also defined:

$A_{\mathrm{k}}^{\mathrm{FirstContr}}$: the year of the first contribution;
$A_{\mathrm{k}}^{\mathrm{RetAge}}$: the current retirement age;
$A_{\mathrm{k}}^{\mathrm{WYears}}$: the number of years that the agent k has worked until $A_{\mathrm{k}}^{\mathrm{RetAge}}$;
$G_{\mathrm{y}}^{\mathrm{TContr}}$: the total amount of contributions, $G_{\mathrm{y}}^{\mathrm{TContr}} = G^{\mathrm{Contr}} \times \sum_{\mathrm{k}} A_{\mathrm{k,y,e,j}}^{\mathrm{Rem}}$;
$A_{\mathrm{k,y}}^{\mathrm{Pension}}$: the pension value;
$G_{\mathrm{y}}^{\mathrm{TPensions}}$: the total amount of the pension values, $G_{\mathrm{y}}^{\mathrm{TPensions}} = \sum_{\mathrm{k}} A_{\mathrm{k,y}}^{\mathrm{Pension}}$;
$G_{\mathrm{y}}^{\mathrm{SS}}$: the total capital of the social security system, $G_{\mathrm{y}}^{\mathrm{SS}} = G_{\mathrm{y}-1}^{\mathrm{SS}} \times G_{\mathrm{y}}^{\mathrm{GDP}} + G_{\mathrm{y}}^{\mathrm{TContr}} - G_{\mathrm{y}}^{\mathrm{TPensions}}$.

4.2.1 Earnings-Based Model

The earnings-based model is now defined. Let

G^{EBScale} be a scaling coefficient defined by the social security system with $G^{\mathrm{EBScale}} \in [0, 1]$;
G^{EBYears} be the number of working years used for the computation of an agent's pension;
$A_{\mathrm{k}}^{\mathrm{EBRem}}$ be a vector of size G^{EBYears} with the remunerations of the last G^{EBYears} years of work re-evaluated for the current time;
$A_{\mathrm{k},i}^{\mathrm{EBRem}}$, $i \in [1, \cdots, G^{\mathrm{EBYears}}]$ be $A_{\mathrm{k}}^{\mathrm{EBRem}}$ elements.

Then $A_{\mathrm{k,y}}^{\mathrm{Pension}}$ can be computed by

$$A_{\mathrm{k,y}}^{\mathrm{Pension}} = A_{\mathrm{k}}^{\mathrm{WYears}} \times G^{\mathrm{EBScale}} \times \sum_{i=1}^{G^{\mathrm{EBYears}}} A_{\mathrm{k},i}^{\mathrm{EBRem}}. \tag{4.2.1.1}$$

4.2.2 Contributions-Based Model

The contributions-based model relies on the following definitions.

Let

$A_{\mathrm{a,k,y}}^{\mathrm{Contr}}$ be the contribution of the agent k to the social security system;
G^{ActAdj} be the actuarial adjustment factor defined by the social security system (Mazzaferro and Morciano [MM12]);
$G_{\mathrm{a,y}}^{\delta\mathrm{Age}}$ be a conversion factor (Mazzaferro and Morciano [MM12]) relating the retirement age and year in which the agent becomes retired.

$G_{\mathrm{a,y}}^{\delta\mathrm{Age}}$ is computed as

$$G_{\mathrm{a,y}}^{\delta\mathrm{Age}} = \left(\frac{\sum_{\mathrm{s}} G_{\mathrm{a,s,y}}^{\mathrm{Dir}} + G_{\mathrm{a,s,y}}^{\mathrm{Ind}}}{2} - G^{\mathrm{ActAdj}} \right)^{-1} \tag{4.2.2.1}$$

where s refers to the sex of the agents. $G_{\mathrm{a,s,y}}^{\mathrm{Dir}}$ and $G_{\mathrm{a,s,y}}^{\mathrm{Ind}}$ are direct and indirect terms for the computation of $G_{\mathrm{a,y}}^{\delta\mathrm{Age}}$. The direct term is related to the pension payed during the agent's life and the indirect term is related to a possible pension payment, as a widower/widow pension, for his/her spouse after the agent has died. These variables are computed as follows:

$$G^{\text{Dir}}_{\text{a,s,y}} = \sum_{t=0}^{G^{\text{MaxAge}}-\text{a}} \frac{G^{\text{StatProp}}_{\text{a}+t,\text{j,y}}}{G^{\text{StatProp}}_{\text{a,s,y}}}(1+G^{\text{GDP}})^{-1} \tag{4.2.2.2}$$

$$G^{\text{Ind}}_{\text{a,s,y}} = \sum_{t=0}^{G^{\text{MaxAge}}-\text{a}} \frac{G^{\text{StatProp}}_{\text{a}+t,\text{s,y}}}{G^{\text{StatProp}}_{\text{a,s,y}}}\left(1-\frac{G^{\text{StatProp}}_{\text{a}+t+1,\text{s,y}}}{G^{\text{StatProp}}_{\text{a}+t,\text{s,y}}}\right)(1+G^{\text{GDP}})^{-(t+1)}G^{\text{WidFrac}}(1+G^{\text{WidInc}})^{(t+1)}, \tag{4.2.2.3}$$

where

G^{MaxAge} is the maximum allowed age in the model;
$G^{\text{StatProp}}_{\text{a,j,y}}$ is the stationary population proportion as explained in Section 5.3.3 on page 52;
G^{WidFrac} is the percentage of the pension which is payed to the widow(er);
G^{WidInc} is the expected increase in the percentage of the pension which is payed to the widow(er) each year.

Then $A^{\text{Pension}}_{\text{k,y}}$ is computed by

$$A^{\text{Pension}}_{\text{k,y}} = \sum_{a=0}^{A^{\text{RetAge}}_{\text{k}}} A^{\text{Rem}}_{\text{k},a,e,j}(1+G^{\text{GDP}})^{(A^{\text{RetAge}}_{\text{k}}-a)}G^{\delta\text{Age}}_{A^{\text{RetAge}}_{\text{k}},\text{y}}. \tag{4.2.2.4}$$

4.2.3 Portuguese Model

In addition to the previous theoretical social security models, the model used by the Portuguese Social Security will also be considered. This model is a mixture of the two previous models.

In the early years, the Portuguese model followed an earnings-based model, which was then reformed into a contributions-based model, due to sustainability issues. So the current model has functionalities from both, in order not to penalize the Portuguese population which started working on the previous regime.

The following variables are defined:

$A^{\text{LastRem}}_{\text{k}}$: the vector of the last 40 work remunerations, corrected to present values.This vector might be smaller than 40 if the agent k has worked for less than 40 years;
$A^{\text{LastRem}}_{\text{k},t}$: $A^{\text{LastRem}}_{\text{k}}$ elements;
$A^{\text{SizeRem}}_{\text{k}}$: the size of $A^{\text{LastRem}}_{\text{k}}$;
$A^{\text{RefRem}}_{\text{k}}$: the reference remuneration used by the Portuguese Social Security system to compute the pensions value,

$$A^{\text{RefRem}}_{\text{k}} = \frac{\sum\limits_{t} A^{\text{LastRem}}_{\text{k},t}}{14A^{\text{SizeRem}}_{\text{k}}} \tag{4.2.3.1}$$

G^{IAS}: the reference value for the computation of aids and other expenses, and revenues of the Portuguese State administration [06];
$A^{\text{InvLastRem}}_{\text{k}}$: the sub-vector of $A^{\text{LastRem}}_{\text{k}}$ containing its last 15 elements at most. $A^{\text{InvLastRem}}_{\text{k}}$ is also ordered from the greatest to the smallest value;
$A^{\text{InvLastRem}}_{\text{k},t}$: $A^{\text{InvLastRem}}_{\text{k}}$ elements.

If $A^{\text{SizeRem}}_{\text{k}} \leq 10$ then

$$A^{\text{P1}}_{\text{k}} = A^{\text{RefRem}}_{\text{k}}. \tag{4.2.3.2}$$

else

$$A_{\mathrm{k}}^{\mathrm{P1}} = \frac{\sum\limits_{t=1}^{10} A_{\mathrm{k},t}^{\mathrm{InvLastRem}}}{14 * 10}. \tag{4.2.3.3}$$

If $A_{\mathrm{k}}^{\mathrm{SizeRem}} \leq 20$ then

$$A_{\mathrm{k}}^{\mathrm{P2}} = A_{\mathrm{k}}^{\mathrm{RefRem}} \times 0.02 \times A_{\mathrm{k}}^{\mathrm{SizeRem}} \tag{4.2.3.4}$$

else

$$A_{\mathrm{k}}^{\mathrm{P2}} = \begin{cases} 0.023 A_{\mathrm{k}}^{\mathrm{RefRem}} A_{\mathrm{k}}^{\mathrm{SizeRem}}, & \text{if } A_{\mathrm{k}}^{\mathrm{RefRem}} \leq 1.1 G^{\mathrm{IAS}} \\ [0.023 \times 1.1 G^{\mathrm{IAS}} & \\ \;+0.0225(A_{\mathrm{k}}^{\mathrm{RefRem}} - 1.1 G^{\mathrm{IAS}})] A_{\mathrm{k}}^{\mathrm{SizeRem}}, & \text{if } 1.1 G^{\mathrm{IAS}} < A_{\mathrm{k}}^{\mathrm{RefRem}} \leq 2 G^{\mathrm{IAS}} \\ [(0.023 \times 1.1 + 0.0225 \times 2) G^{\mathrm{IAS}} & \\ \;+0.022(A_{\mathrm{k}}^{\mathrm{RefRem}} - 2 G^{\mathrm{IAS}})] A_{\mathrm{k}}^{\mathrm{SizeRem}}, & \text{if } 2 G^{\mathrm{IAS}} < A_{\mathrm{k}}^{\mathrm{RefRem}} \leq 4 G^{\mathrm{IAS}} \\ [(0.023 \times 1.1 + 0.0225 \times 2 + 0.022 \times 4) G^{\mathrm{IAS}} & \\ \;+0.021(A_{\mathrm{k}}^{\mathrm{RefRem}} - 4 G^{\mathrm{IAS}})] A_{\mathrm{k}}^{\mathrm{SizeRem}}, & \text{if } 4 G^{\mathrm{IAS}} < A_{\mathrm{k}}^{\mathrm{RefRem}} \leq 8 G^{\mathrm{IAS}} \\ [(0.023 \times 1.1 + 0.0225 \times 2 + 0.022 \times 4 + 0.021 \times 8) G^{\mathrm{IAS}} & \\ \;+0.02(A_{\mathrm{k}}^{\mathrm{RefRem}} - 8 G^{\mathrm{IAS}})] A_{\mathrm{k}}^{\mathrm{SizeRem}}, & \text{if } 8 G^{\mathrm{IAS}} < A_{\mathrm{k}}^{\mathrm{RefRem}} \end{cases} \tag{4.2.3.5}$$

The model also defines

$A_{\mathrm{k}}^{\mathrm{C1}}$ as the working years of agent k until year 2006;
$A_{\mathrm{k}}^{\mathrm{C2}}$ as the working years of agent k from year 2007 until present;
$A_{\mathrm{k}}^{\mathrm{C3}}$ as the working years of agent k until year 2001;
$A_{\mathrm{k}}^{\mathrm{C4}}$ as the working years of agent k from year 2002 until present.

Finally, the pension value is set as

$$A_{\mathrm{k,y}}^{\mathrm{Pension}} = \begin{cases} A_{\mathrm{k}}^{\mathrm{P2}}, & \text{if } A^{\mathrm{FirstContr}} > 2002 \\ \frac{A_{\mathrm{k}}^{\mathrm{P1}} A_{\mathrm{k}}^{\mathrm{C1}} + A_{\mathrm{k}}^{\mathrm{P2}} A_{\mathrm{k}}^{\mathrm{C2}}}{A_{\mathrm{k}}^{\mathrm{WYears}}}, & \text{if } A^{\mathrm{FirstContr}} \leq 2002 \text{ and } A_{\mathrm{k}}^{\mathrm{RetAge}} \leq 2016 \\ \frac{A_{\mathrm{k}}^{\mathrm{P1}} A_{\mathrm{k}}^{\mathrm{C3}} + A_{\mathrm{k}}^{\mathrm{P2}} A_{\mathrm{k}}^{\mathrm{C4}}}{A_{\mathrm{k}}^{\mathrm{WYears}}}, & \text{if } A^{\mathrm{FirstContr}} \leq 2002 \text{ and } A_{\mathrm{k}}^{\mathrm{RetAge}} > 2016 \end{cases} \tag{4.2.3.6}$$

4.3 Algorithm

Most parameters related to the social security are updated every year, according to their values in the previously simulated year and the 2011 Portuguese data from Statistics Portugal and GEE databases [Sta15c],[GEE12].

Step 1. Update the activity of the agents in the current year $\mathrm{y_c}$. If $\left(G_{\mathrm{a,j,y_c}}^{\mathrm{ActR}} - G_{\mathrm{a,s}}^{\mathrm{ActP}}\right) > G^{\epsilon}$ then u is randomly sampled from $U(0,1)$ for each agent aged a years-old, with sex s and with $A_{\mathrm{a,s,k,y}}^{\mathrm{Act}} = 1$; if $u < G^{\epsilon}$, then, the agent k is set as inactive $A_{\mathrm{a,s,k,y}}^{\mathrm{Act}} = 0$ in order to reduce the excess of active population. If $\left(G_{\mathrm{a,s,y_c}}^{\mathrm{ActR}} - G_{\mathrm{a,s}}^{\mathrm{ActP}}\right) < -G^{\epsilon}$ then u is randomly sampled from $U(0,1)$ for each agent aged a years-old, with sex s and with $A_{\mathrm{a,s,k,y}}^{\mathrm{Act}} = 0$; if $u < G^{\epsilon}$, then then agent k is set as active $A_{\mathrm{a,s,k,y}}^{\mathrm{Act}} = 1$ in order to fix the lack of active population;

Step 2. Update the employment status of the agents;

Step 3. Update the retirement status of the agents with $A^{\mathrm{Ret}}_{\mathrm{a,s,k,y}} = 0$. If $G^{\mathrm{MinRet}} \leq a \leq G^{\mathrm{MinRet}} + 10$ and $G^{\mathrm{Act}}_{\mathrm{a,s,k,y}} = 0$ then u is randomly sample from $U(0,1)$ and if $u < G^{\mathrm{RetR}}_{\mathrm{a,s,y}}$, then the agent k is set as retired $A^{\mathrm{Ret}}_{\mathrm{a,s,k,y,}} = 1$. When an agent gets retired, its retirement pension is defined and remains the same until the agent's death.

Step 4. Update the agents education and job qualification. When an agent gets employed for the first time these parameters are defined. They are decided randomly applying the Inversion Method (Devroye [Dev86]) to randomly sampled u_1 and u_2 from $U(0,1)$ and using the rates of the education and job qualification for the age of the agent. On the following years the parameters are updated using the same mechanism but with the constraint that they have to be increasing with time.

Step 5. Update the wages. The wage of an employed agent is defined as $A^{\mathrm{Wage}}_{\mathrm{a,s,k,y,e,j}} = G^{\mathrm{Wage}}_{\mathrm{s,e,j}} \times \left(1 + G^{\mathrm{GDP}}\right)^{\mathrm{y}-\mathrm{y}_0}$;

Step 6. Impose that all agents with $A^{\mathrm{Emp}}_{\mathrm{a,s,k,y}} = 1$ pay their social security contributions;

Step 7. Pay the retirement pensions to all agents satisfying $A^{\mathrm{Alive}}_{\mathrm{a,s,k,y}} = 1$ and $A^{\mathrm{Ret}}_{\mathrm{a,s,k,y}} = 1$;

Step 8. Update the social security's capital G^{SS}.

Chapter 5

Model Initialization, Data and Results

The model presented sometimes requires information from before the start of the simulation. Here, the mechanism that was used to setup the simulation will also be shown.

Furthermore, a broad spectrum of variables is needed for the initialization of the previous model. The data used was obtained from several sources, national and international. Often, the data that was found was not exactly the one that was needed, so some arrangements were made in order to achieve the desired information. Preferably, the data source for a determined variable should always be the same, but unfortunately that was not always the case. The source data and the needed modifications will be presented in this chapter.

At the end, some results of the program will be presented, focusing in a set of sub-models that portrait the outcome of the other sub-models. These results showcase the Portuguese population structure and its social security system.

Figure 5.1 on the next page systematizes and summarizes the previous model application and study, showcasing the most important aspects. The diagram sequence goes from the top to the bottom and starts from the end of the diagram in Chapter 1 on page 1.

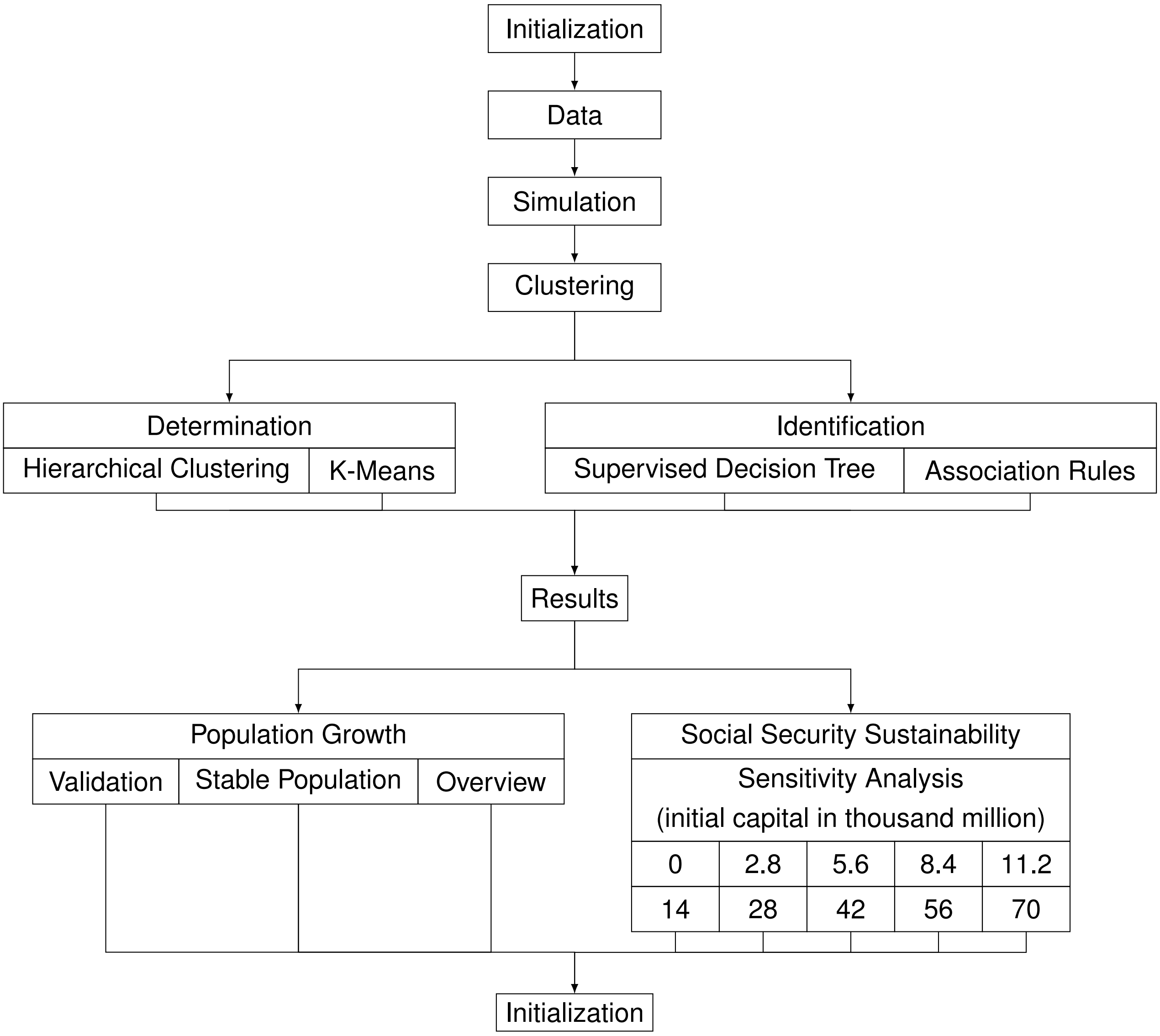

Figure 5.1: Diagram of how results will be presented.

5.1 Model Initialization

The model that is closed to migration does not require information prior to the starting year, but all other models do. Still, to initialize the migration model, the information needed from before the simulation starting year is only used to fit a few distributions, and so the supplied data concern only the distribution parameters.

However, the model for the social security system requires previous data to be inserted into the model. The computation of the retirement pension is highly dependent on the remuneration and work situation of an agent through all agent's working career. The data from the past Portuguese activity rates $P_{\mathrm{y}}^{\mathrm{ActR}}$, employment rates $P_{\mathrm{y}}^{\mathrm{EmpR}}$, mean wages $P_{\mathrm{y}}^{\mathrm{wage}}$ and GDP growth $P_{\mathrm{y}}^{\mathrm{GDP}}$ since 1950 were obtained from Statistics Portugal and Banco de Portugal [Sta15c; Ban15].

At the start of the simulation, the following procedure was performed for each living agent k:

Step 1. Find the year y_b for which the age of agent k was 15 years-old (minimum age for employment)

Step 2. For each year y, with $\mathrm{y}_b < \mathrm{y} < \mathrm{y}_0$, generate random numbers u_1, u_2 from $U(0, 1)$. If

$\left(u_1 < P_{\mathrm{y}}^{\mathrm{ActR}}\right) \wedge \left(u_2 < P_{\mathrm{y}}^{\mathrm{EmpR}}\right)$, then $A_{\mathrm{a,s,k,y}}^{\mathrm{Emp}} = 1$ and $A_{\mathrm{a,s,k,y,e,j}}^{\mathrm{Wage}} = P_{\mathrm{y}}^{\mathrm{wage}}$.

In section 4.2 on page 31, the social security capital was assumed to have an annual increase that is directly proportional to the increase of the GDP. The same will be done for previous contributions but, in the initialization, no pension will be paid. In addition, and for simplicity, each agent's contribution $A_{\mathrm{k,y}}^{\mathrm{Contr}}$ will be already increased in accordance to the GDP increase:

$$A_{\mathrm{k,y}}^{\mathrm{Contr}} = A_{\mathrm{k,y,e,j}}^{\mathrm{Rem}} \times G^{\mathrm{Contr}} \tag{5.1.0.1}$$

and

$$G_{\mathrm{y}}^{\mathrm{TContr}} = \sum_{\mathrm{k}} \left(A_{\mathrm{k,y}}^{\mathrm{Contr}} \times \prod_{j=y}^{\mathrm{y_0}} P_{\mathrm{j}}^{\mathrm{GDP}} \right), \ \mathrm{y} < \mathrm{y_0}. \tag{5.1.0.2}$$

Therefore, in the beginning of the simulation

$$G_{\mathrm{y_0}}^{\mathrm{SS}} = \sum_{\mathrm{y}<\mathrm{y_0}} G_{\mathrm{y}}^{\mathrm{TContr}}. \tag{5.1.0.3}$$

There is one more important information relating to the model initialization. This one is not related to the necessary past data, but it is related to the population size. For the computational simulation to work it was necessary to use a 2% sample of the Portuguese population. A higher sample would require many computer resources that were not available (processor speed and RAM) and a lower sample would make some age groups to be too small or inexistent.

5.2 Data

$G_{\mathrm{a,s,y_0}}^{\mathrm{Alive}}$	2011 Portuguese Census [Sta15c]
$G_{\mathrm{a,s,y_0}}^{\mathrm{Births}}$	Statistics Portugal [Sta15c]
$G_{\mathrm{a,s,y_0}}^{\mathrm{Deaths}}$	Statistics Portugal [Sta15c]
G^{FertEvo}	Resident Population Projections 2012-2060 [Sta15c]
G^{MortEvo}	Resident Population Projections 2012-2060 [Sta15c]

Table 5.1: Closed Population Model data source.

G^{FertEvo} and G^{MortEvo} are growth rates, but the existing data is about the expected fertility and mortality. Therefore, in order to obtain the growth rate for each year, the value of the following year from the data was divided by the value of the current year.

The chosen health $G_{\mathrm{c}}^{\mathrm{Health}}$ and safety $G_{\mathrm{c}}^{\mathrm{Safety}}$ indicators were the percentage of the public expenses spent on health and safety for each country. For $G_{\mathrm{c}}^{\mathrm{Health}}$ the needed data that was available, unlike $G_{\mathrm{c}}^{\mathrm{Safety}}$ which was available only as the nominal expenditure in safety and total GDP.

The population indicator $G_{\mathrm{c}}^{\mathrm{Pop}}$ was set equal to the percentage of Portuguese born residents in country c over the total resident population of the same country.

The distance indicator $G_{\mathrm{c}}^{\mathrm{Dist}}$ that was used was the distance between each country's capital to the Portuguese capital (Lisbon). The available data were the coordinates of each city according to its latitude and longitude. So defining lat_1, lon_1 as the coordinates in degrees of the first capital, lat_2, lon_2 as the coordinates in degrees of the second capital and setting $R = 6371$ as the radius of the planet

G_c^{Health}	The World Factbook [CIA13]
G_c^{Safety}	The World Factbook [CIA13] United Nations Statistics Division [UN15] Rea [Rea09] (New Zealand source) Statistics Austria [Sta15a] (Austria source) Comptroller General of the Union [Com15] (Brazil source) Statistics Canada [Sta15b] (Canada source) National Institute of Statistics, Geography and Informatics [NI15] (Mexico source)
G_c^{Pop}	United Nations Statistics Division [UN15]
G_c^{Dist}	CIA [CIA13]
G_c^{Limit}	OECD Statistics [OEC15] Emigration Observatory [Emi15] (through apparent limits on data)
$W(x;\lambda,k)$	Eurostat [Eur15]
$G_{c,y_0}^{\text{ForPop}}$	OECD Statistics [OEC15] United Nations Statistics Division [UN15] Japan Institute for Labour Policy and Training [Jap15] (Japan source) France's National Institute for Statistics and Economic Studies [Fra15] (France source)
$G_{c,y_0,s}^{\text{ForEmpR}}$	OECD Statistics [OEC15] Japan Institute for Labour Policy and Training [Jap15] (Japan source)
G_c^{NetMig}	The World Factbook [CIA13]
G_c^{MWage}	OECD Statistics [OEC15]
G_c^{PPP}	World Bank [Wor15]
$G_{c,y}^{\text{Immi}}$	OECD Statistics [OEC15]
X^{ImmiAge}	Immigration Annual Estimations [Sta15c]
G^{ImmiProp}	OECD Statistics [OEC15]

Table 5.2: Migration Model data source.

Earth in kilometers, then

$$dist = \frac{1-\cos\left((lat_2 - lat_1)\times\frac{\pi}{180}\right)}{2} \tag{5.2.0.4}$$

$$+\cos(lat_1\times\frac{\pi}{180})\times\cos(lat_2\times\frac{\pi}{180})\times\frac{(1-\cos\left((lon_2-lon_1)\times\frac{\pi}{180}\right)}{2} \tag{5.2.0.5}$$

$$dist = R\times 2\times\arcsin(\sqrt{dist}) \tag{5.2.0.6}$$

where *dist* is the distance in kilometers.

The parameters G_c^{Health}, G_c^{Safety}, G_c^{Pop}, $G_{c,y,e}^{\text{wage}}$, G_c^{Dist} were forced to be between 0 and 1, dividing them by their corresponding maximum, ranging between all values of c.

The parameter $W(x; \lambda, k)$ is an age distribution of Portuguese emigrants. The data containing the age of the Portuguese emigrants was fitted to a Weibull and Gamma distributions as they were the distributions that had the closest shape to the data density function. Figure 5.2 on the next page shows the data density function and the fitted distributions. The Weibull distribution was chosen as it presented the best fit.

$G^{\text{NetMig}}_{\text{c},\text{y}_0,\text{e}}$ values are on a 1 : 1000 ratio. Therefore, the data was scaled using the population size of each country.

X^{ImmiAge} is an age distribution of immigrants in Portugal. The data with the age of the Portuguese immigrants was fitted to a Weibull, Gamma and Log-Normal distributions as they were the distributions that had the closest shape to the data density function. Figure 5.3 on the next page shows the data density function and the fitted distributions. The Weibull distribution was again chosen as it exhibited the closest fit, even though it is indistinguishable from the gamma distribution in the figure.

$G^{\text{Act}}_{\text{a,s}}$	Active Population (2011 Series) [Sta15c]
$G^{\text{EmpProp}}_{\text{a,s}}$	Active Population (2011 Series) [Sta15c]
$G^{\text{RetProp}}_{\text{a,s}}$	Statistics Portugal [Sta15c]
$G^{\text{Edu}}_{\text{a,s}}$	Active Population (2011 Series) [Sta15c]
$G^{\text{JQual}}_{\text{a,s}}$	2011 Staffing Tables [GEE12]
$G^{\text{Wage}}_{\text{s,e,j}}$	2011 Staffing Tables [GEE12]
G^{GDP}	Trading Economics [Tra15] (projections)
$G^{\text{WTenure}}_{\text{a,s}}$	OECD Statistics [OEC15]
G^{IAS}	European Social Fund [IGF15]

Table 5.3: Social Security Model data source.

The data found for $G^{\text{WTenure}}_{\text{a,s}}$ was not as desired. The data that was needed was the mean work tenure by age of the worker at the beginning of the work contract, but the data that was found was for the mean work tenure by age of the worker at the end of the work contract. This problem was solved by shifting the data backwards considering the mean work tenure. For example, if the data showed that there were 200 people aged 60 years-old and their mean work tenure was 15 years, then the 200 value would be moved to people aged $60 - 15 = 45$ years-old having a mean work tenure of 15 years. If the work tenure was expressed with more precision than years, then the value was rounded to years. This was applied for all ages.

Inspired by Addison and Portugal [AP87], the work tenure was modeled using a statistical distribution. Due to unique characteristics of the data, namely a heavy right tail with a bump, the only distribution that was found to be capable of reproducing such behavior was the Log-Cauchy distribution. The graphic in Figure 5.4 on page 43, shows the closeness of the Cauchy distribution to the logarithm of the data, which is equivalent to show that the Log-Cauchy distribution is close to the data distribution itself.

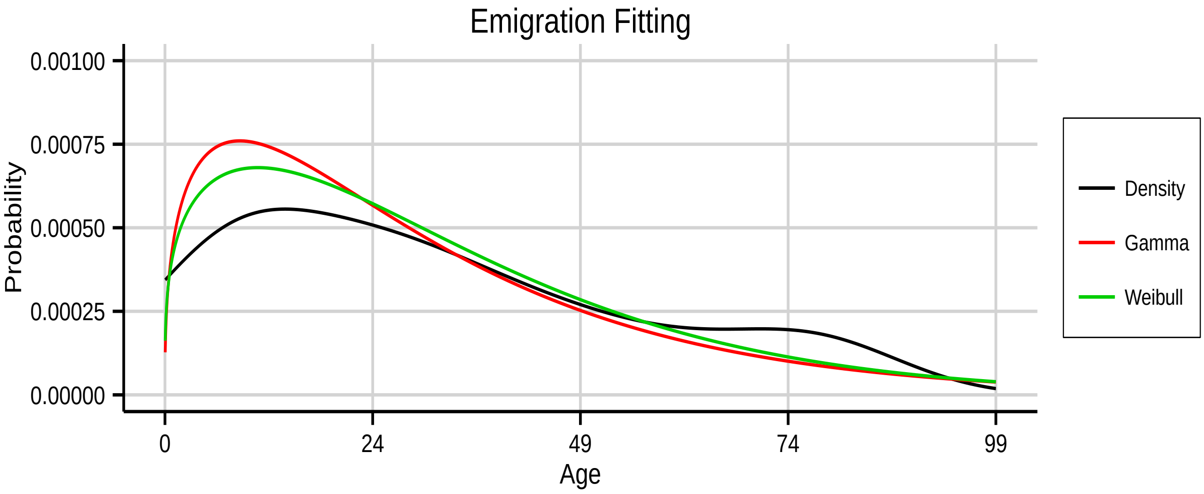

Figure 5.2: Portuguese emigrants fitted to Weibull and Gamma distributions.

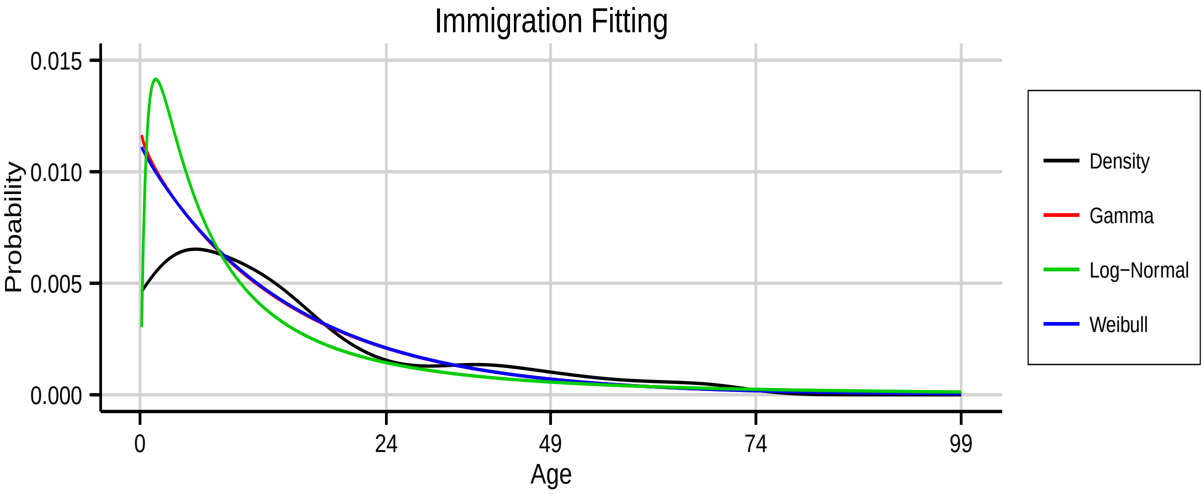

Figure 5.3: Immigrants to Portugal fitted to Weibull, Gamma and Log-Normal distributions.

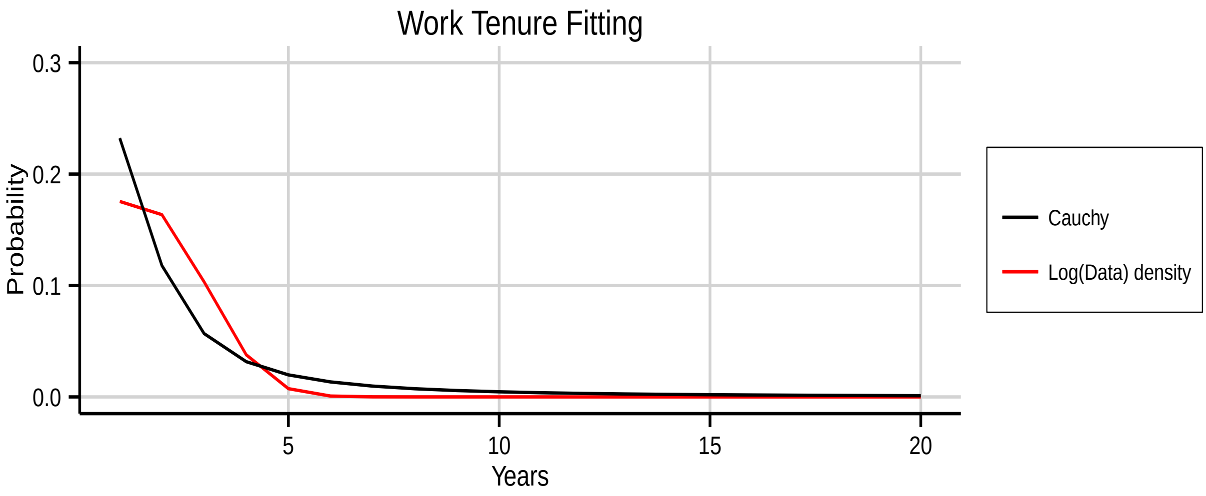

Figure 5.4: Logarithm of the work tenure of Portuguese with ages between 15 and 19 years fitted to a Cauchy distribution.

5.3 Results

Until now, three fertility models, three mortality models and two employment models were presented. The model must either have heterogeneous or homogeneous agents and may or may not have migrations. Also there are two exogenous variables, the retirement age, consisting in three possibilities (66 years, 68 years and 70 years), and the economic scenario, consisting in three possibilities (0.9 or recession, 1.0 or stability, and 1.1 or prosperity). The latter is used to change the number of jobs existing in Portugal, changing each year by the corresponding economic scenario percentage in relation to the previous year. This way, seven necessary dimensions can be defined for this work, namely fertility, mortality, employment, heterogeneity, migrations, retirement age and economic scenario. To perform a simulation it is necessary to select a model or a variable for each one of the seven dimensions.

As it is not evident which selection is the better suited for the Portuguese case, all possible combinations between the seven dimensions are considered for this study. Considering the 5 models dimensions there are $3 \times 3 \times 2 \times 2 \times 2 = 72$ sub-models to be considered. Also adding the 2 variables dimensions there are $3 \times 3 = 9$ possible variables combinations. So combining all this there is a total of $72 \times 9 = 648$ sub-models to be simulated.

Some experiments showed that both mac and mic-mac models for mortality provided similar results for the Portuguese population evolution. Therefore only the mic-mac model was used, thus reducing the number of sub-models to 432.

Furthermore, after some testing and some revision of the code, two models were found to be incompatible. These models were the mic employment model and the migration model, which resulted in instantaneous null migration. So at the end, 324 sub-models were left to be studied.

Subsequently, for each combination, 100 simulations were produced using the computer grid of Faculty of Engineering of the University of Porto (FEUP). The means and standard deviations for all output features were also computed.

The considered output features are the predictions of 2041. The outputs and their respective codes in the graphics are:

Total Population (total_pop);
Total Female Population (total_pop_F);
Total Male Population (total_pop_M);
Total Population with age until 14 years (total_pop_0_14);
Total Female Population with age until 14 years (total_pop_F_0_14);
Total Male Population with age until 14 years (total_pop_M_0_14);
Total Population within age of 15 years until 65 years (total_pop_15_65);
Total Female Population within age of 15 years until 65 years (total_pop_F_15_65);
Total Male Population within age of 15 years until 65 years (total_pop_M_15_65);
Total Population of age 66 years or over (total_pop_66p);
Total Female Population of age 66 years or over (total_pop_F_66p);
Total Male Population of age 66 years or over (total_pop_M_66p);
Total Births (total_births);
Total Female Deaths (total_deaths_F);
Total Male Deaths (total_deaths_M);
Earnings-Based Social Security's Capital (total_EB_SSV);
Contributions-Based Social Security's Capital (total_CB_SSV);
Portuguese Social Security's Capital (total_PT_SSV);
Total Emigration (total_emi);
Employment Rate (emp_rate).

5.3.1 Clustering

In order to filter the results to be presented, a clustering analysis was considering the mean from the output of all 324 model combinations.

In the first place, the minimum number of clusters that better illustrated the main differences between all models combinations was found. An unsupervised learning regression tree, considering all the output parameters, suggested an initial number of clusters equal to 7 (Figure 5.5 on the following page). In addition, according to the Elbow Rule from the K-means algorithm, the number of clusters to be used should be either 4 or 7 (Figure 5.6 on the following page). Yet, considering that using only 4 clusters the resulting groups would be unsatisfactory, the final choice for the number of clusters was 7.

After obtaining the 7 desired clusters through the K-means method, a supervised learning regression tree and association rules were applied to characterize and identify these clusters. For this, two partitions of the 324 sub-models were created: one containing 70% of the data, which rounded to 227 sub-models, and the other containing the remaining part of the data. While the first set was used as a training set, the second set was used to test the regression tree and the association rules.

As observable in Figure 5.7 on the following page, the clusters are well defined, considering only the output from the employment rates, total population, total population of age 66 years or over, and male population of age 66 years or over. This process of cross-validation provided an error of 2% in the identification of the clusters.

Alongside with an error of 0%, 11 association rules were identified for the input and the clusters (Figure 5.8 on the following page). Graphically, only 9 rules are identifiable, but in fact there are 2

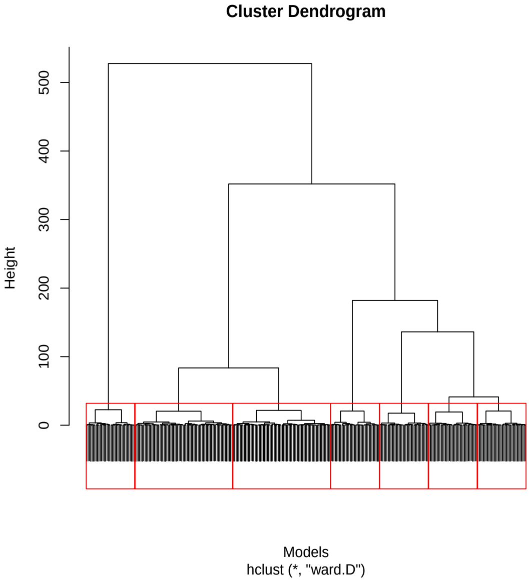

Figure 5.5: Unsupervised learning regression tree for the 324 model combinations.

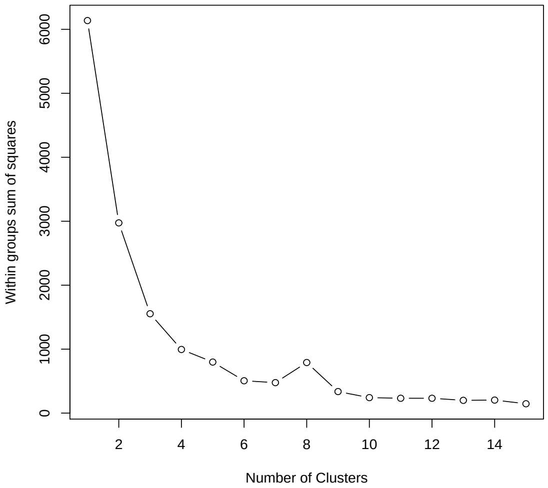

Figure 5.6: K-means Elbow Rule for the 324 model combinations.

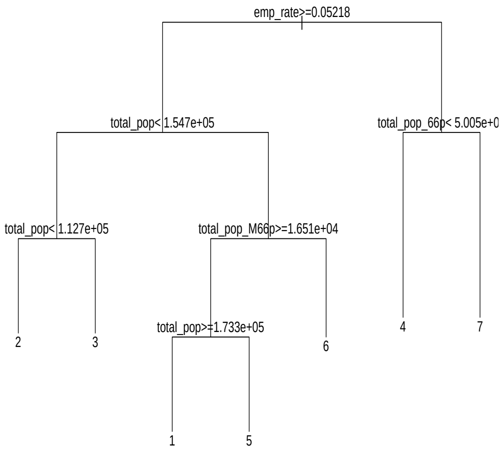

Figure 5.7: Supervised learning regression tree for the 7 clusters.

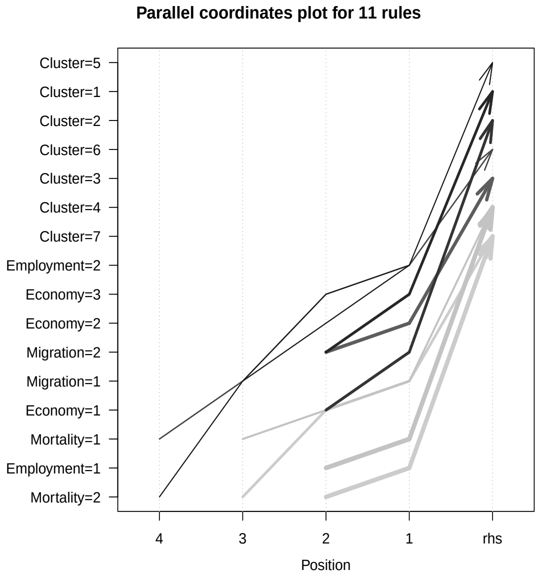

Figure 5.8: Association rules for the 7 clusters.

rules overlaying 2 others, namely the ones that end on clusters 5 and 6. Although an error of 0% is an odd value, when we directly observe the data, it is noticeably that the clusters are very well organized for the inputs, which suggests that the input is being mapped correctly into each cluster.

In Tables 5.4 on the next page and 5.5 on the next page, a summary of the supervised tree and association rules can be observed.

From Figure 5.9 on page 47, 20 bar plots can be observed, one for each output parameter, where each bar corresponds to the centroid value of each cluster and the small intervals on top of each bar correspond to the 95 confidence interval for the correspondent centroid. When the confidence intervals do not overlap, the bars correspond to significantly different means. In particular,

	Employment Rate	Total Population	Total Population Over 65	Male Population Over 65
Cluster 1	≥ 0.05218	≥ 173300		≥ 16510
Cluster 2	≥ 0.05218	< 112700		
Cluster 3	≥ 0.05218	$\geq 112700 \wedge < 154700$		
Cluster 4	< 0.05218		< 50050	
Cluster 5	≥ 0.05218	$\geq 154700 \wedge < 173300$		≥ 16510
Cluster 6	≥ 0.05218	≥ 154700		< 16510
Cluster 7	< 0.05218		≥ 50050	

Table 5.4: Supervised Tree for the 7 clusters.

	Mortality	Employment	Migration	Economy
Cluster 1		Mic-Mac	Yes	Prosperity
Cluster 2		Mic-Mac	Yes	Recession
Cluster 3		Mic-Mac	Yes	Stability
Cluster 4	Mic		No	Regression
	Mic	Mic		
Cluster 5	Mic-Mac	Mic-Mac	No	Prosperity
	Mic-Mac	Mic-Mac	No	Stability
Cluster 6	Mic	Mic-Mac	No	Prosperity
	Mic	Mic-Mac	No	Stability
Cluster 7	Mic	Mic-Mac		
	Mic-Mac		No	Recession

Table 5.5: Association Rules for the 7 clusters.

this suggests that the error for the supervised regression tree should be low, as it is.

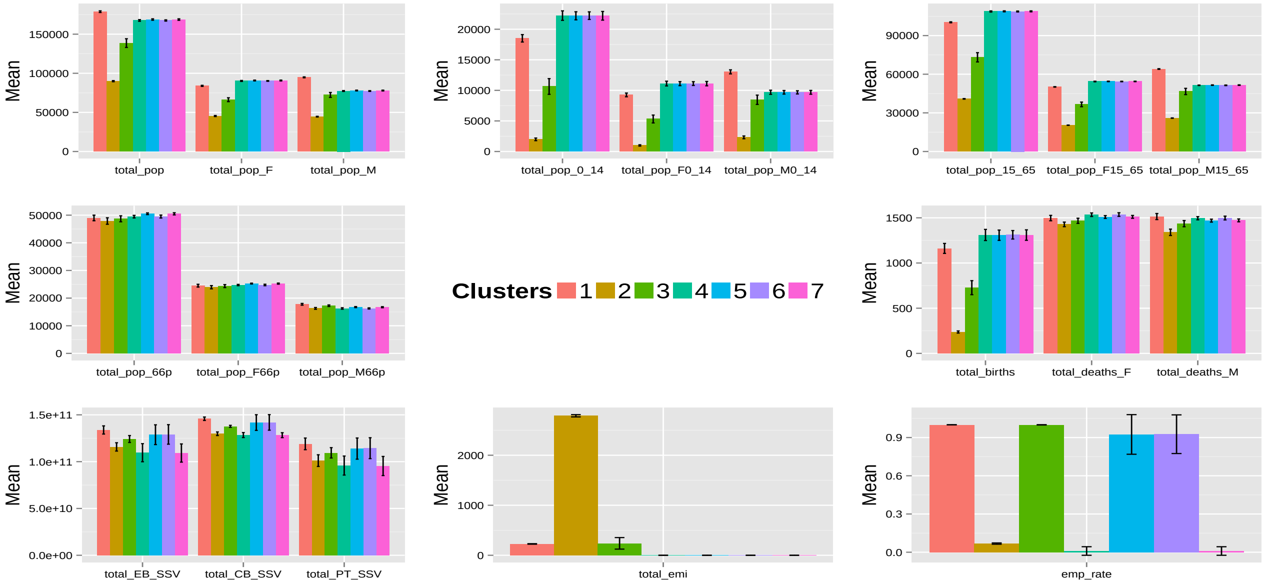

Figure 5.9: Graphical description of the clusters' centroids with 2 standard deviations error bars.

With all the information so far, some observations of the clusters can be made.

Cluster 1 This cluster has the highest mean estimation for the total Portuguese population, most likely due to the high fertility, low emigration and high immigration (as immigration evolves linearly and positively with time). The employment rate is close to 100% and the social security's capital is at its highest value. This is congruent with the precept of a prosperous country.

Cluster 2 This cluster, in opposition to the previous one, has the lowest mean estimation for the total Portuguese population, most likely due to the low fertility and high emigration. The employment rate is the lowest of them all and the social security's capital is one of the lowest. This is in line with the precept of a country in recession.

Cluster 3 This cluster has the centre mean estimation for the Portuguese population, most likely due to the average fertility and medium/low emigration rates that it is presenting. The employment rate is one of the highest and the social security's capital is intermediate in comparison to the others. This description fits a country that has a stable economy.

Cluster 4 This cluster corresponds to a set of sub-models closed to migration. Also, due to some sub-models that compose this set, namely the sub-models with the mic employment model, the employment rate bar plot is biased. Because in the mic employment model the employment rate is not computed. This cluster is one of the two clusters with the lowest social security capital (place shared with Cluster 7). It differs only from Cluster 7 on the total Portuguese population of age over 65, justified by the usage of the mic mortality model, which usually "kills" all the ages with low population density; cluster 7 uses the mic-mac mortality model.

Cluster 5 This cluster corresponds to a set of sub-models closed to migration on which the employment rate is relatively high, nearly 80%. This cluster is one of the two clusters with the highest social security capital (place shared with Cluster 6 and 1). It differs only from Cluster 6 on the total Portuguese population of age over 65, justified by the usage of the mic-mac mortality model, which for low density age population uses the mac mortality model; cluster 6 uses the mic mortality model.

Cluster 6 This cluster corresponds to a set of sub-models closed to migration on which the employment rate is relatively high, nearly 80%. This cluster is one of the two clusters with the highest social security capital (place shared with Cluster 5 and 1). It differs only from Cluster 5 on the total Portuguese population of age over 65, justified by the usage of the mic mortality model.

Cluster 7 This cluster corresponds to a set of sub-models closed to migration. Also, due to some sub-models that compose this set, namely the sub-models with the mic employment model, the employment rate bar plot is biased. Because in the mic employment model the employment rate is not computed. This cluster is one of the two clusters with the lowest social security capital (place shared with Cluster 4). It differs only from Cluster 4 on the total Portuguese population of age over 65, justified by the usage of the mic-mac mortality model.

Considering the 7 above defined clusters and the 11 association rules, a selection of 11 sub-models is made to refine the presentation in this work.

The 11 sub-models are the ones which satisfy the Association Rules and are closest to the centroid of their cluster. These 11 sub-models are:

Sub-Model 1 : mac fertility + mic-mac mortality + mic-mac employment + migration + age of retirement equal to 68 years + prosperous economy;
Sub-Model 2 : mac fertility + mic mortality + mic-mac employment + heterogeneity + migration + age of retirement equal to 68 years + recession economy;
Sub-Model 3 : mic-mac fertility + mic mortality + mic-mac employment + heterogeneity + migration + age of retirement equal to 68 years + stable economy;
Sub-Model 4 : mic fertility + mic mortality + mic employment + age of retirement equal to 70 years + prosperous economy;
Sub-Model 5 : mac fertility + mic mortality + mic employment + heterogeneous + age of retirement equal to 70 years + recession economy;
Sub-Model 6 : mic-mac fertility + mic-mac mortality + mic-mac Employment + age of retirement equal to 66 years + prosperous economy;
Sub-Model 7 : mic-mac fertility + mic-mac mortality + mic-mac employment + age of retirement equal to 70 years + stable economy;
Sub-Model 8 : mic fertility + mic mortality + mic-mac employment + age of retirement equal to 66 years + prosperous economy;
Sub-Model 9 : mic-mac fertility + mic mortality + mic-mac employment + heterogeneity + age of retirement equal to 70 years + stable economy;
Sub-Model 10 : mic-mac fertility + mic-mac mortality + mic employment + age of retirement equal to 70 years + prosperous economy;
Sub-Model 11 : mic-mac fertility + mic-mac mortality + mic employment + age of retirement equal to 70 years + recession economy.

5.3.2 Model Validation

Due to the lack of data for the years before 2011, it is impossible to test the validity of most sub-models. Nothing related to the employment or the social security's capital can be tested, nor it can be tested for the population open to migration. Nevertheless, model validation techniques can be applied to the population closed to migration with a sub-model consisting only of the models for the mic fertility and mic mortality.

A sliding window verification is performed using 100 simulations for the mic fertility and mic mortality sub-model, initiated with the data from the years of 1981, 1991 and 2001, until the years of 1991, 2001 and 2011. For each of these years there was a population census in Portugal and the data are therefore available.

Comparing the output from these simulations with the correct data from 1991, 2001 and 2011 census data, the error of the simulations can be computed as ($simulated - real$). These errors are stored in tables such as Table 5.6.

		Start-off instants		
		1981	1991	2001
Gaps	10 years	Error in 1991	Error in 2001	Error in 2011
	20 years	Error in 2001	Error in 2011	
	30 years	Error in 2011		

Table 5.6: Model Validation.

Error tables are made for the Total Population, Total Population by age, Total Population by gender and Total Population by gender and by age. With these, an overview of the simulation error can be presented.

Figures 5.10 on the next page, 5.11 on page 51 and 5.12 on page 51 show the error from the three starting years 1981, 1991 and 2001 respectively, alongside with their comparison with the years of 1991, 2001 and 2011 when possible. In all figures, the most dramatic errors are at the lowest and highest ages. In Figure 5.12 on page 51 additional errors are found at every ages, showing that the simulation under-predicted what would happen in the next 10 years. It should be noticed, however, that according to the database from the Statistics Portugal [Sta15c], the immigration in that time frame was very high, as well as the net migration.

		Start-off instants		
		1981	1991	2001
Gaps	10 years	4734	−3877	−16491
	20 years	9462	−8225	
	30 years	17979		

Table 5.7: Total Population Error.

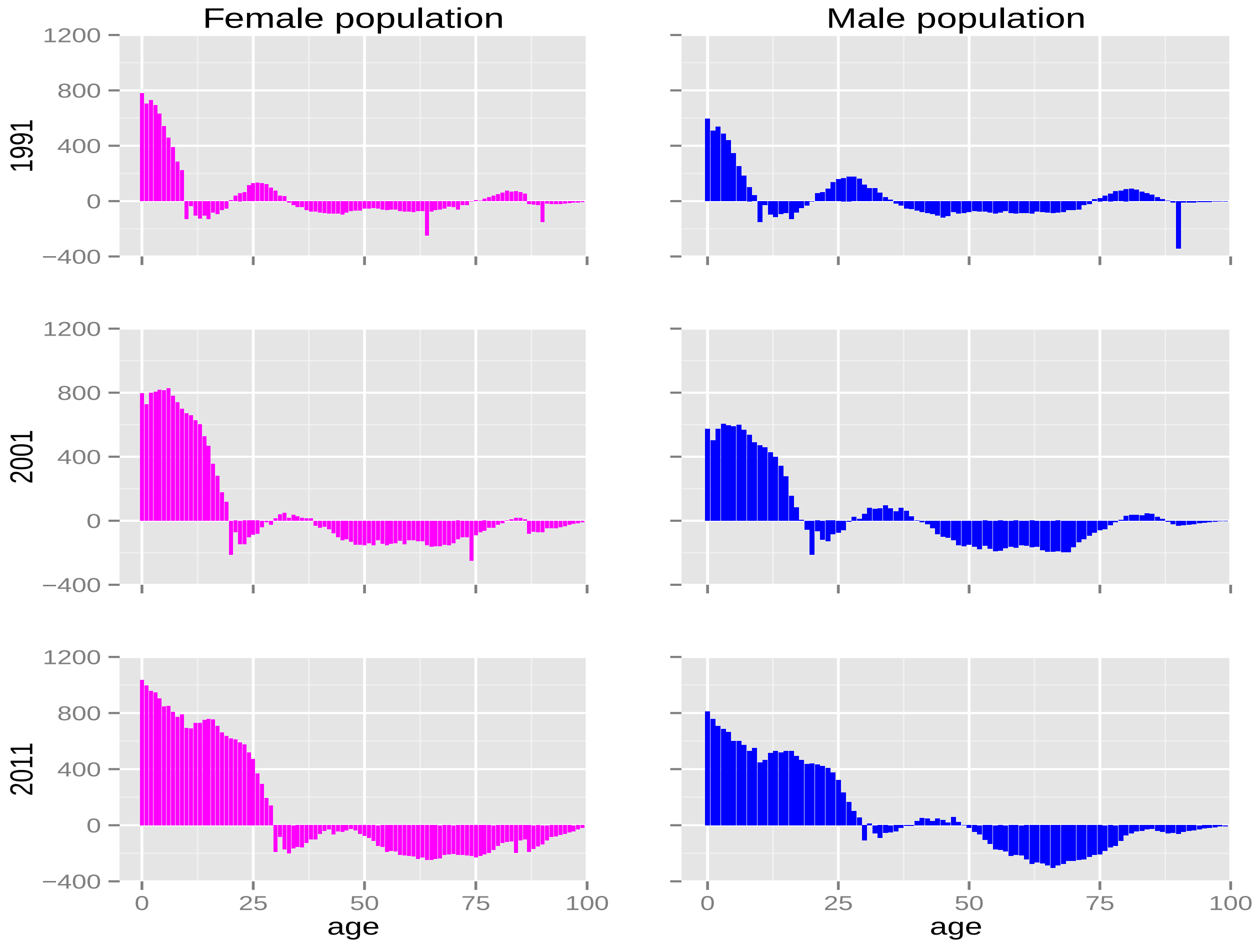

Figure 5.10: Error means bar plot starting at 1981.

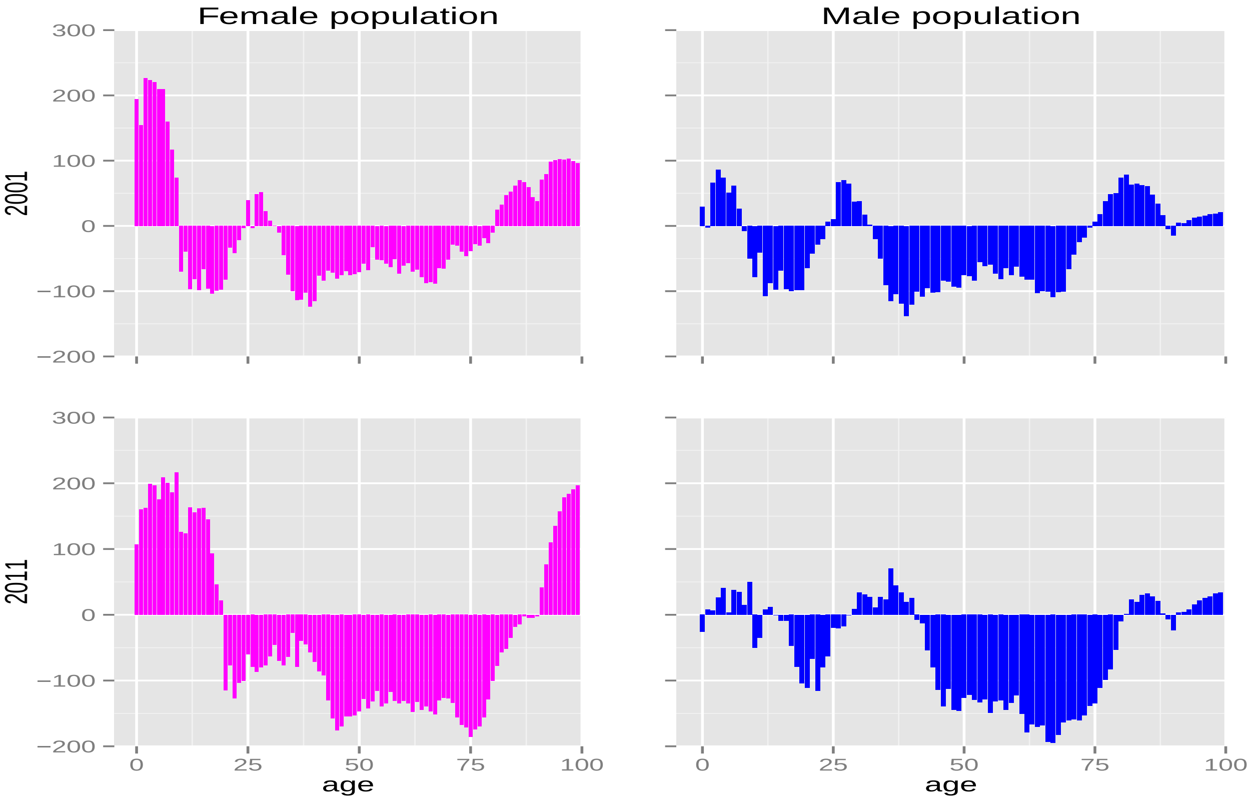

Figure 5.11: Error means bar plot starting at 1991.

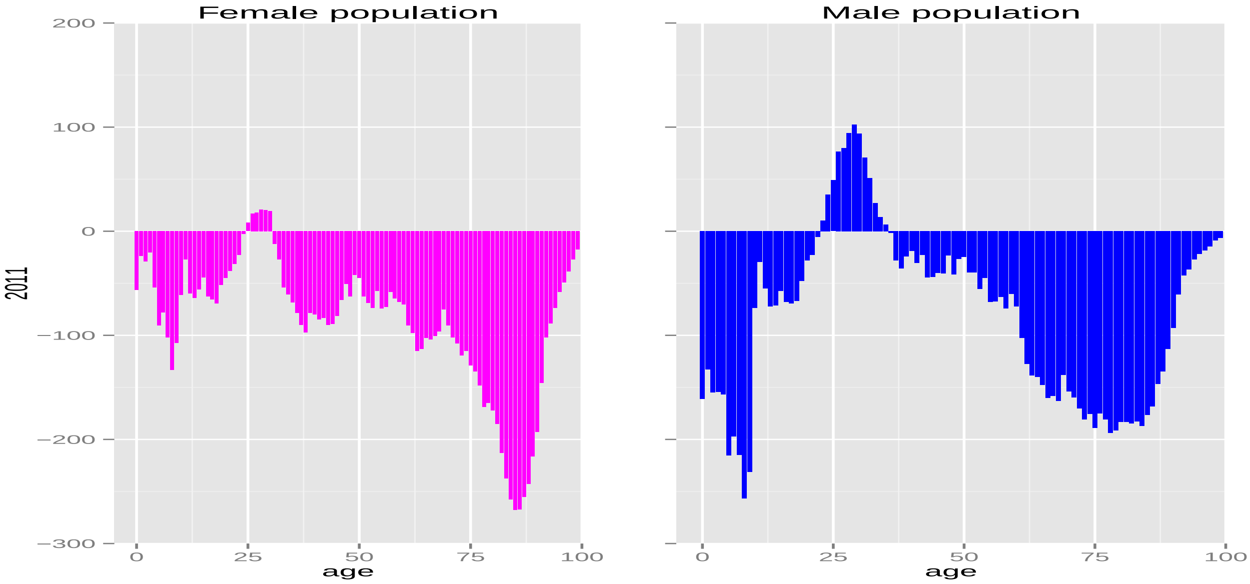

Figure 5.12: Error means bar plot starting at 2001.

		Start-off instants		
		1981	1991	2001
Gaps	10 years	2996	−849	−8589
	20 years	6201	−3177	
	30 years	10448		

Table 5.8: Total Female Population Error.

		Start-off instants		
		1981	1991	2001
Gaps	10 years	1738	−2029	−7902
	20 years	3261	−5048	
	30 years	7531		

Table 5.9: Total Male Population Error.

Tables 5.7 on page 49, 5.8 and 5.9 show total population, total female population and total male population errors organized as in table 5.6 on page 49. Not surprisingly, the error increases alongside the considered gap size, i. e., the longer the simulation lasts, the higher the errors will be produced. From year to year, the error is not independent; the error from the previous year influences the present year error. Therefore, there is an error propagation, so it is not interesting to increase the simulation time indefinitely. Also, except for the starting year of 1981, the total female population error is lower than the male counterpart.

5.3.3 Stationary and Stable Population

The stable population model, as presented in Section 2.2.5 on page 12, is used to study populations with constant age-specific fertility, mortality and net migration. Portugal, as a developed country, should tend towards a structure similar to those of other developed countries, because the population distribution of most developed countries is very close to their stable population equivalent (Bras [Bra08]).

1. Mortality-wise, there are variations between population distributions, but they are small;
2. Fertility-wise, studies say global fertility is expected to remain constant, but that is not necessarily the case for each age-specific fertility;
3. Migration-wise, the age-specific net migration is not yet null.

But even so, the stable population equivalent is still valid, because an approximation for the stable population equivalent might translate into a tendency towards the satisfaction of the three starting conditions for the stable population model.

In the following pages, the algorithm for the stable population will be presented. Even though stable population can be achieved immediately by complex analytic means, it is simpler if the stationary population is first computed and only then the stable population. The stationary population is a specific case of the stable population model, for whose intrinsic growth rate (r) equals to 0.

Usually this computation is not made for each age, but it is made for 5 year age groups (Preston, Heuveline, and Guillot [PHG01]). The data required is only the female population, female deaths, and births for each age group. The data on male population and deaths will only be used to show the outcome for the male population.

Defining:

$l_{i,s}^{\text{five}}$ as the number of persons in the age group $i \in \{1, \dots, 20\}$ with sex s;
$d_{i,s}^{\text{five}}$ as the number of deaths in the age group $i \in \{1, \dots, 20\}$ with sex s;
$sr_{i,s}^{\text{five}}$ as the survival rate in the age group $i \in \{1, \dots, 20\}$ with sex s, given by:

$$s_{i,s}^{\text{five}} = 1 - \frac{d_{i,s}^{\text{five}}}{d_{i,s}^{\text{five}} + l_{i,s}^{\text{five}}}; \tag{5.3.3.1}$$

b_i^{five} as the number of births in the age group $i \in \{1, \dots, 20\}$.

Defining $p_{i,s}^{\text{stati}}$ and $r_{i,s}^{\text{stati}}$ as the proportions of the stationary population and real stationary population respectively, then:

$$p_{i,s}^{\text{stati}} = 5 \times \prod_{x=1}^{i} sr_{i,s}^{\text{five}} \tag{5.3.3.2}$$

$$r_{i,s}^{\text{stati}} = p_{i,s}^{\text{stati}} \times \sum_{i} l_{i,s}^{\text{five}} \tag{5.3.3.3}$$

For the stable population, the following computations are also required:

$$r_0 = \left\langle p_{i,F}^{\text{stati}}, b_i^{\text{five}} \right\rangle; \tag{5.3.3.4}$$

$$r_1 = \left\langle p_{i,F}^{\text{stati}} \times (2 + 5(i-1)), b_i^{\text{five}} \right\rangle, \text{ where } (2 + 5(i-1)) \text{ is the mean age of each group;} \tag{5.3.3.5}$$

$$r_2 = \left\langle p_{i,F}^{\text{stati}} \times (2 + 5(i-1))^2, b_i^{\text{five}} \right\rangle; \tag{5.3.3.6}$$

and

$$m_1 = \frac{r_1}{r_0}; \tag{5.3.3.7}$$

$$m_2 = \frac{r_2}{r_0}; \tag{5.3.3.8}$$

$$k_2 = m_2 - m_1^2; \tag{5.3.3.9}$$

$$r = \frac{m_1 - \sqrt{m_1^2 - 2 \times k_2 \times \ln(r_0)}}{k_2} \tag{5.3.3.10}$$

The intrinsic growth rate r is already defined, so it will now be defined the proportions of the stable population and real stable population, $p_{i,s}^{\text{stable}}$ and $r_{i,s}^{\text{stable}}$ respectively, as

$$p_{i,s}^{\text{stable}} = e^{(2+5(i-1)) \times r} \times p_{i,s}^{\text{static}}; \tag{5.3.3.11}$$

$$r_{i,s}^{\text{stable}} = e^{(2+5(i-1)) \times r} \times r_{i,s}^{\text{static}}. \tag{5.3.3.12}$$

To determine if the Portuguese population is getting closer to its stable population equivalent, the difference between the simulated population and the stable population equivalent in age groups for each year is computed, and checked if the gap is getting bigger or smaller.

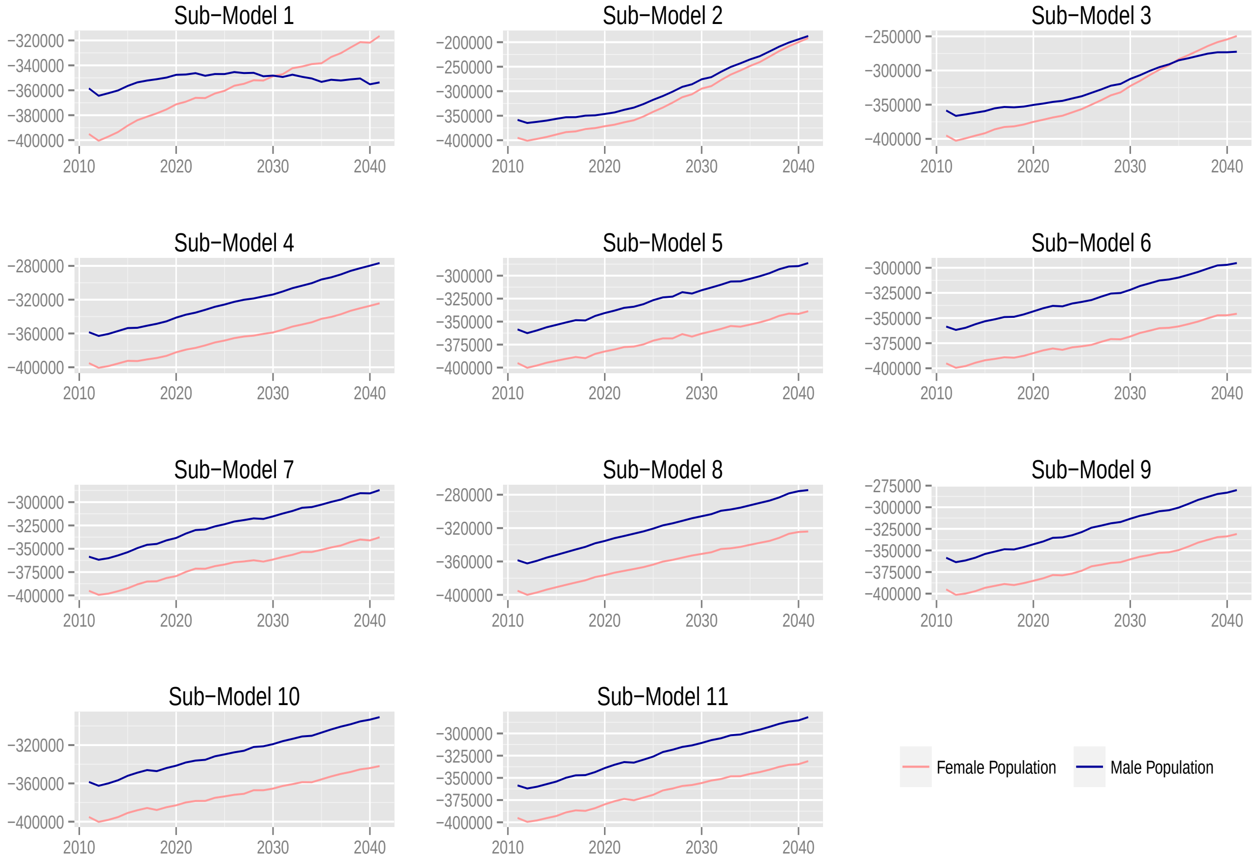

Figure 5.13: Distance between the 11 sub-models and their Stable Population Equivalent.

The evolution of the simulated population for the 11 sub-models, as previously considered, in relation to their stable population equivalent can be observed in Figure 5.13. The difference between the simulated population and its stable population equivalent is given by (*simulated* − *stable*). As can be observed, the distance is shortening every year, which may lead to conclude that we are approximating the stable population equivalent. However, without guarantees that, in the future, the Stable Population model's starting conditions will be met, there is no assurance that we will converge to it.

5.3.4 Population Evolution Overview

Although it is very important to know how the population will be in 2041, the knowledge over the intermediary years may be also very informative.

In this section, a yearly representation of the Portuguese population from 2011 until 2041 will be presented. The most important reasons for the 11 possible outcomes will be shown and explained.

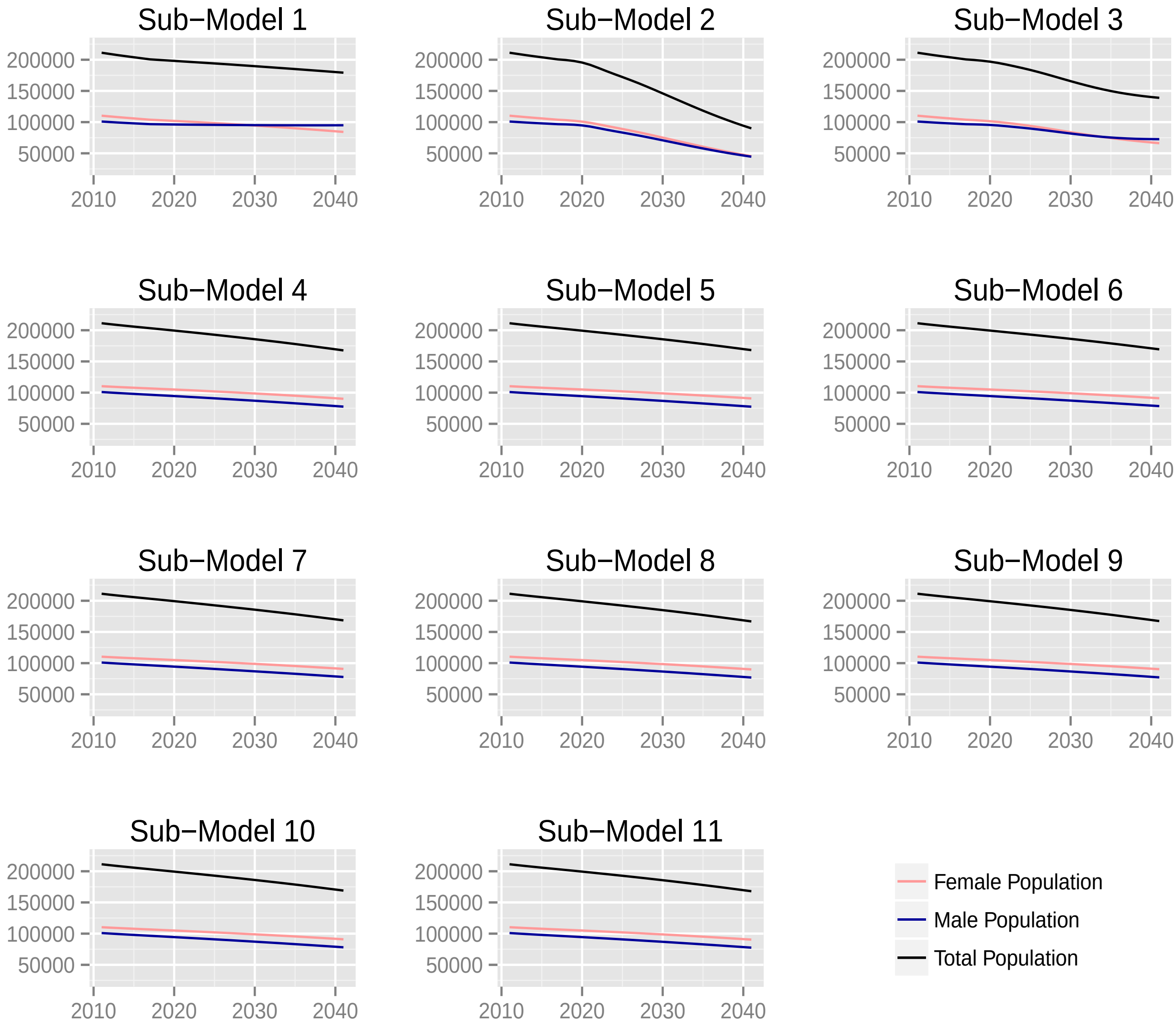

Figure 5.14: Total Population projection for the 11 sub-models.

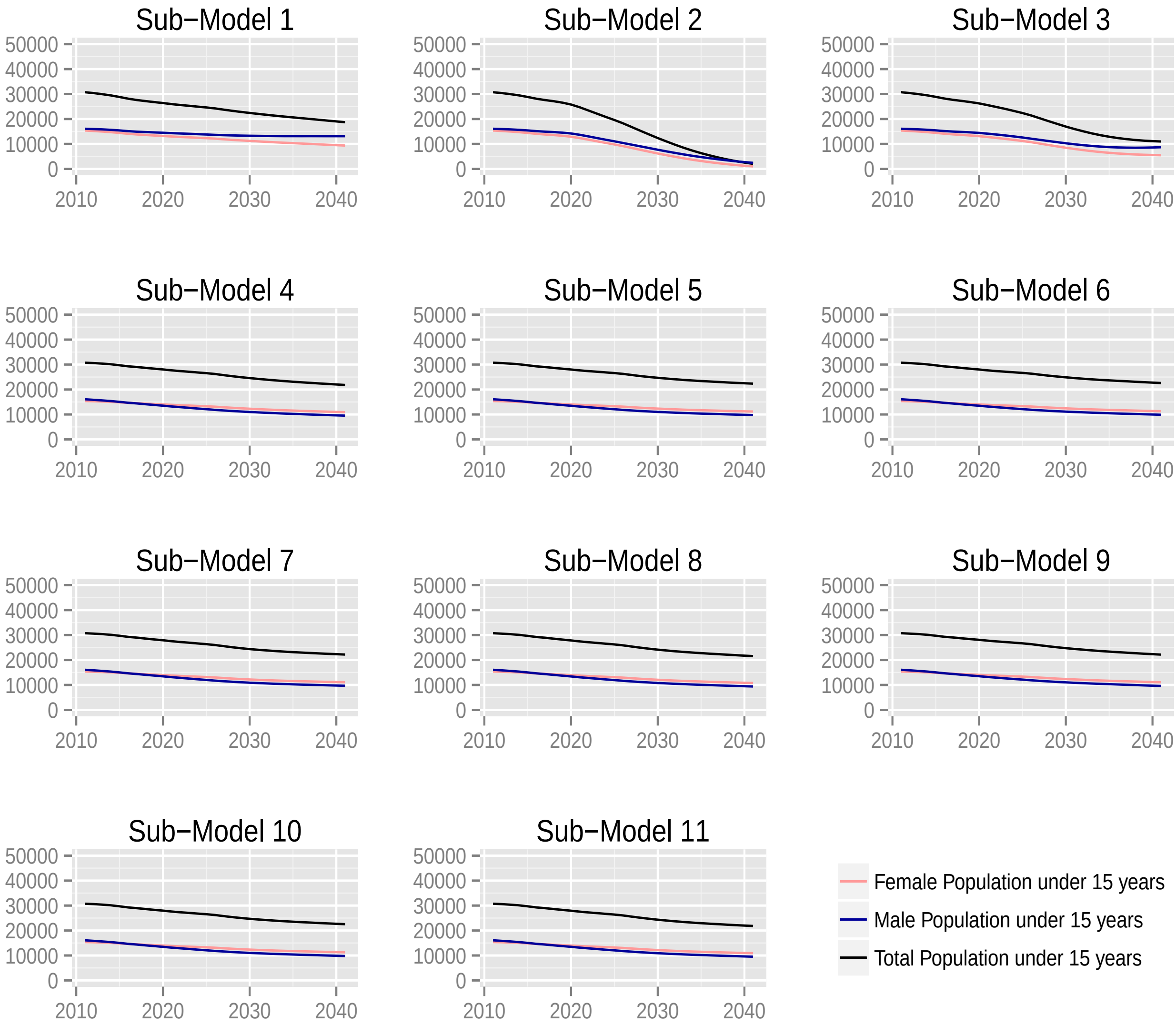

Figure 5.15: Total Population under 15 years projection for the 11 sub-models.

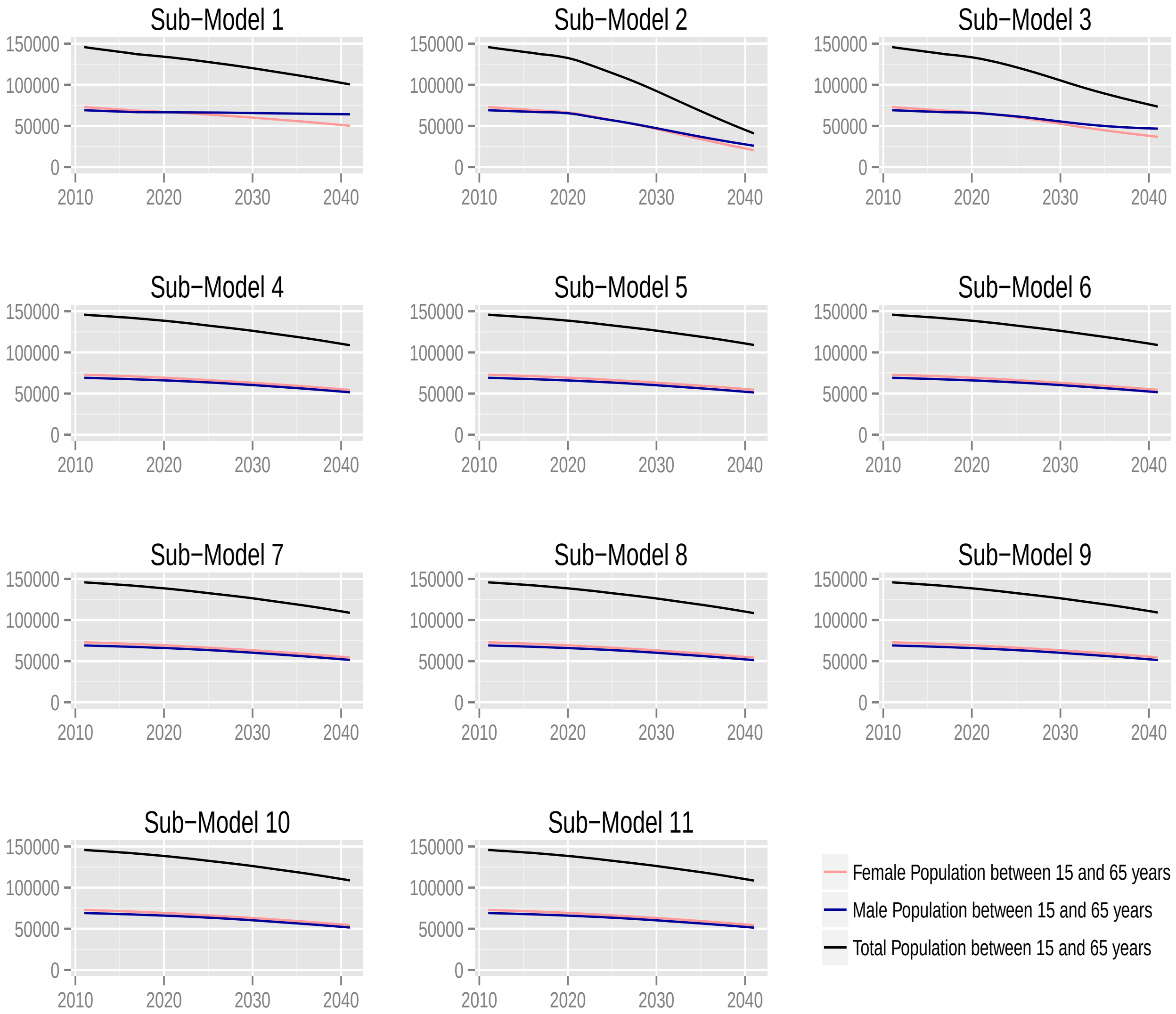

Figure 5.16: Total Population between 15 and 65 years projection for the 11 sub-models.

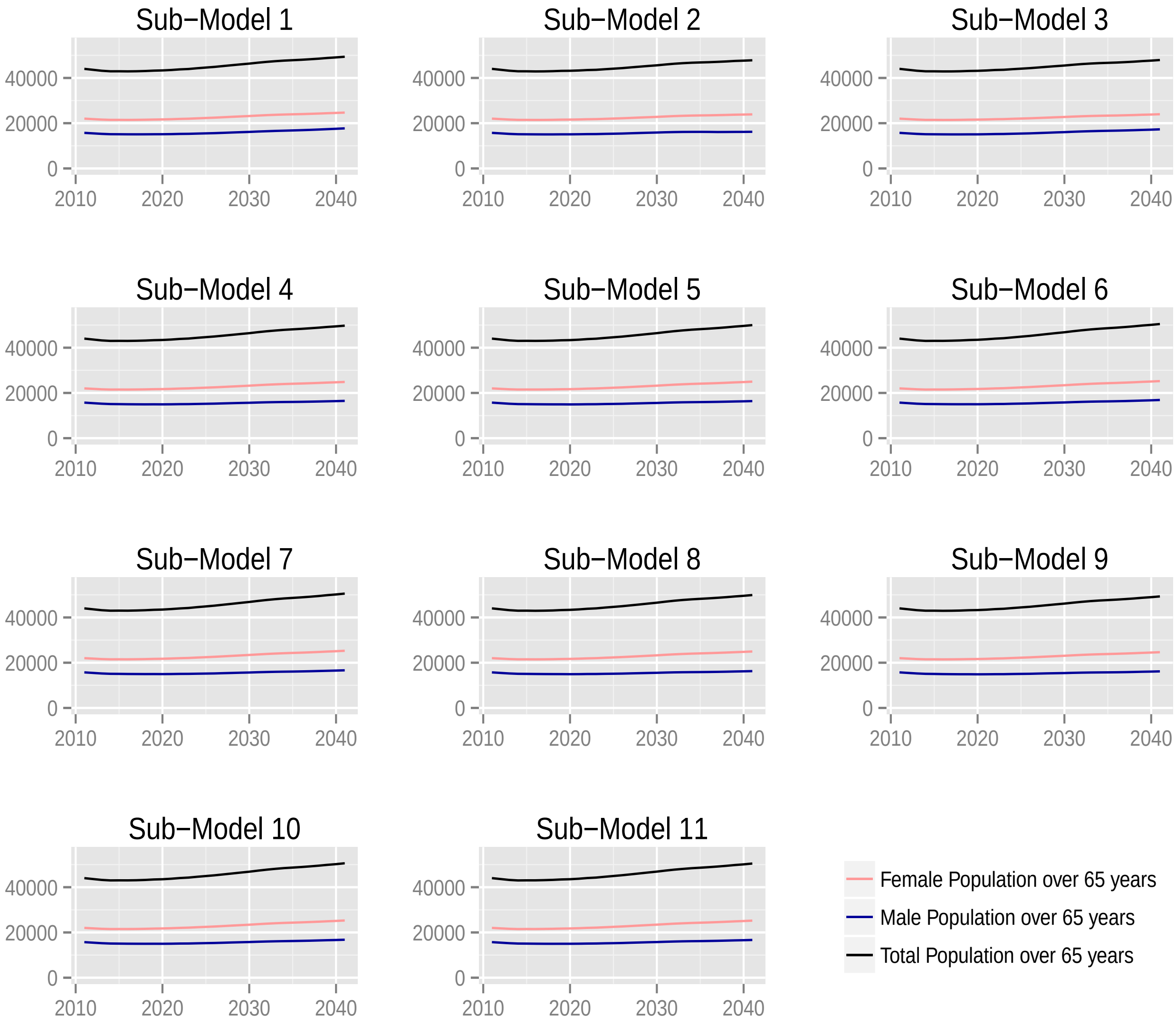

Figure 5.17: Total Population over 65 years projection for the 11 sub-models.

Figures from 5.14 on page 55 to 5.17 show an inevitable population reduction in all sub-models and in all age sub-groups except for the population over 65 years, in which case, after a small decrease, the tendency trend changes and starts to increase. This reveals that the Portuguese population is tending to an aged population structure. For the 8 sub-models with population closed to migration, the population decrease is almost linear, while, for the population open to migration, it shows a quadratic or cubic descent, stabilizing in the last years.

From Figure 5.15 on page 56, Sub-Model 2 shows a very drastic reduction of the population under 15 years of age, closing to 0 in 2041, though this is not due to a reduction in fertility, but it is due to a reduction of the number of females in the fertile period, as revealed in Figure 5.16 on the previous page. Recalling Subsection 5.3.1 on page 44, Sub-Model 2 refers to a sub-model with very high emigration, probably because of the simulated economy being in recession. This sub-model is one in which the active-inactive dependency ratio is very elevated. This is the most pessimist scenario present in this work.

In opposition, Sub-Model 1 and 3 show very positive developments. Although there is still a reduction in the population with less than 15 years-old and between 15 and 65 years-old, there are recovery signs near 2041. Additionally, the population reduction between 15 and 65 years-old is not very steep, so at the end, even with an increase of the dependency ratio, that increase will not be very significant.

It is important to note that the graphics presented for the population age groups only have in consideration the Portuguese-born population. So, immigrants are not taken into account in these graphics. In the total population graphic, however, this does not happen and the immigrant population is also considered.

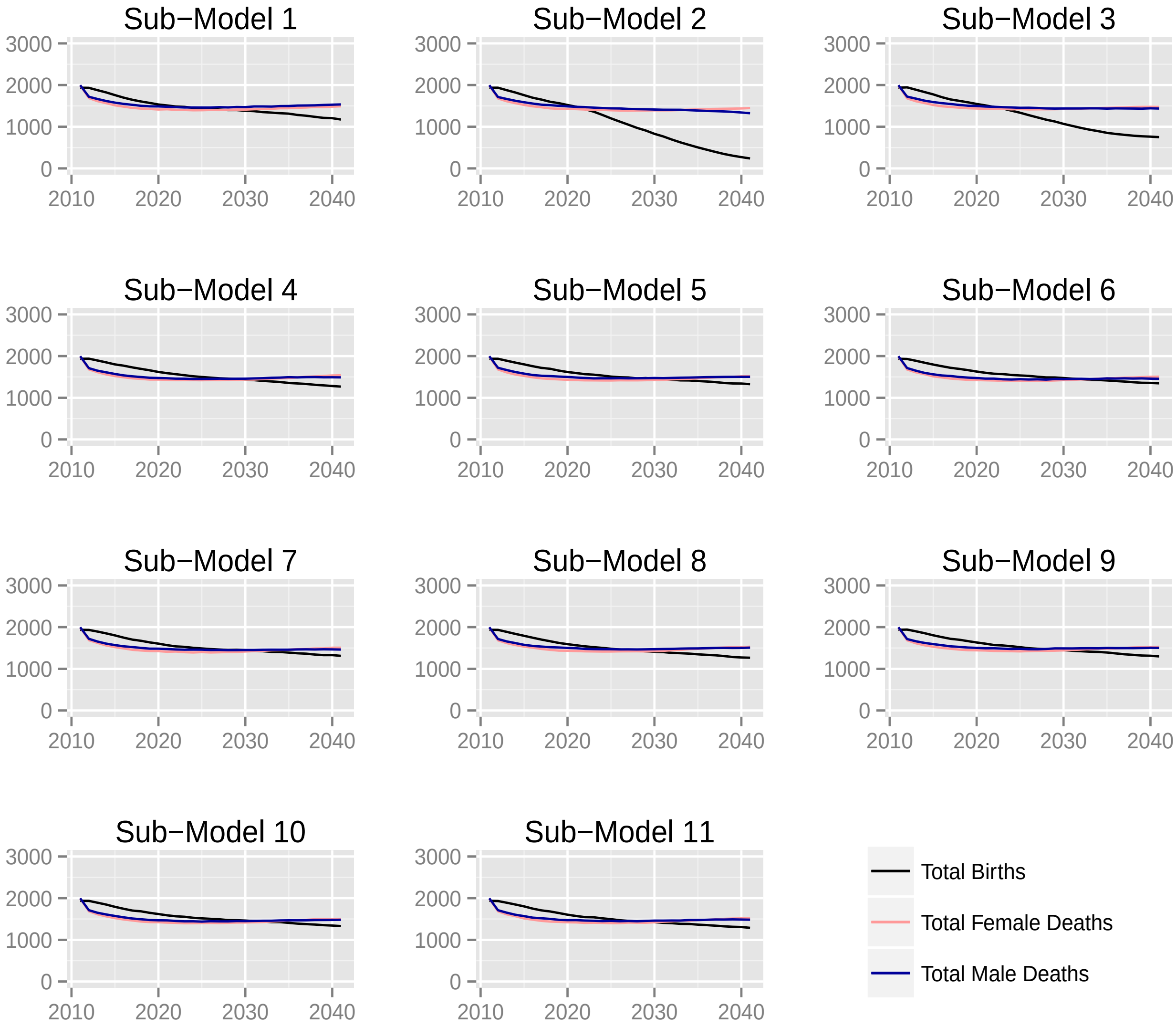

Figure 5.18: Total Births and Deaths for the 11 sub-models.

Figure 5.18 reveals one of the reasons for the population decrease. For almost all sub-models, the number of births in a given year is almost the same as the number of deaths of either the female or male population. This shows that our replacement level is about 50%. Only in Sub-Model 2 and 3, the number of births per year after 2025 is constantly and significantly lower than the number of deaths.

As for the migrations, which are only applied in Sub-Model 1 to 3, the influence in the population can be observed in Figure 5.19 on the next page. The immigration grows linearly by definition and as observable on the graphic. The emigration, however, is highly dependent of the economic scenario and the population structure. Sub-Model 1, which has the prosperous economic scenario, as well as Sub-Model 3, the net migration eventually gets positive and very close to immigration value, as the emigration value tends to 0. While for Sub-Model 2, the graphic shows a tendency for the emigration to decrease past 2041; most likely because most people are already gone, and so, eventually, the net migration may become positive.

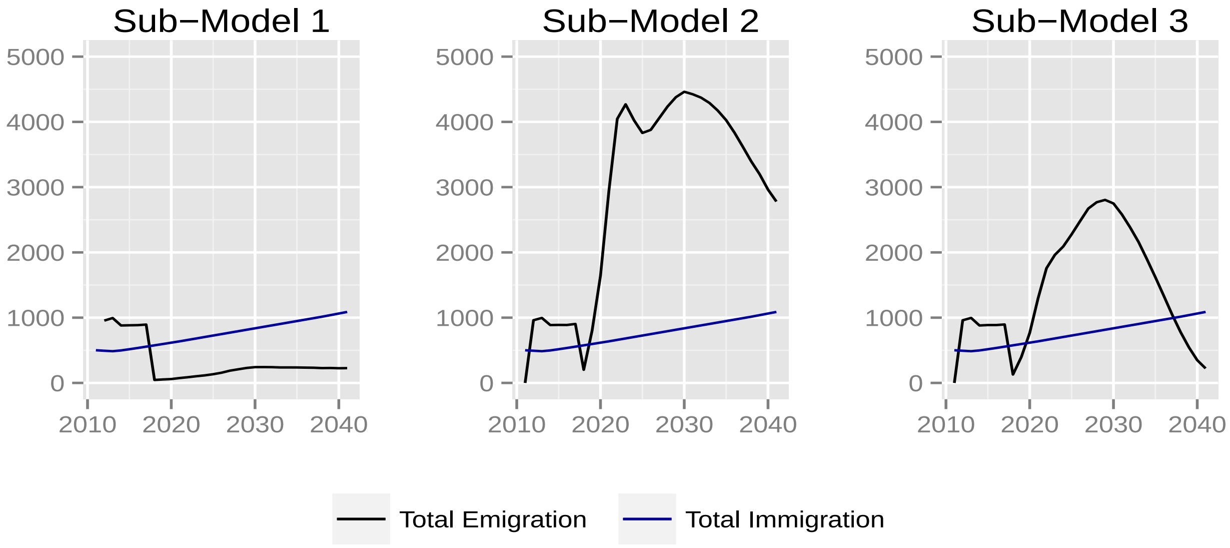

Figure 5.19: Total Emigration and Immigration for the 11 sub-models.

5.3.5 Social Security Sustainability

As explained in Section 5.1 on page 38, the starting value for the social security's capital is computed assuming that the Portuguese Social Security saved the surplus from previous years, and its financial application had profits. The GDP growth was used as a proxy for the financial application interest rates. However, since the data is not available, a sensitivity analysis was performed on this variable. The starting value for the social security's capital in the previous simulations was roughly 70 thousand million euros. The sensitivity analysis used the following set of values (in euros) {0, 2.8 thousand million, 5.6 thousand million, 8.4 thousand million, 11.2 thousand million, 14 thousand million, 28 thousand million, 42 thousand million, 56 thousand million}.

With this sensitivity analysis, an attempt is made to locate the position of the tipping point for the social security sustainability for the 32.7% Contributions, i.e., find the smallest interval for which the lower bound graphic with the social security's capital below 0, for at least 1 year, and the upper bound graphic with the social security's capital always above 0. Therefore, it is guaranteed that the tipping point for the sustainability of the Portuguese Social Security's capital is in that interval.

Figures 5.20 on the following page to 5.41 on page 64 show the tipping point interval for each sub-model considered. It is interesting to note that for some cases, even if the social security's capital reaches negative values, it can still recover and return to positive values before the end of the simulation. This is a result of inherent population structure of each model, but it must be noted that, if the initial capital for the social security gets significantly lower than the lower bound of the interval, then bankruptcy is inevitable. An increase on the contribution also helps to maintain the sustainability of the social Security, but as it is observable, one of the key aspects for the sustainability is the capital

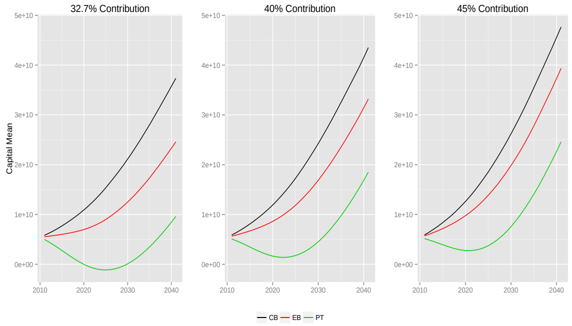

Figure 5.20: Social Security's capital lower bound (5.6 thousand million) for sub-model 1.

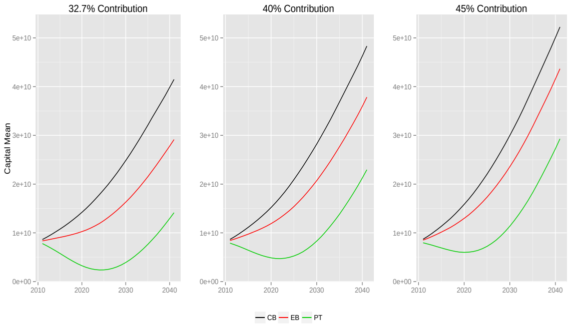

Figure 5.21: Social Security's capital upper bound (8.4 thousand million) for sub-model 1.

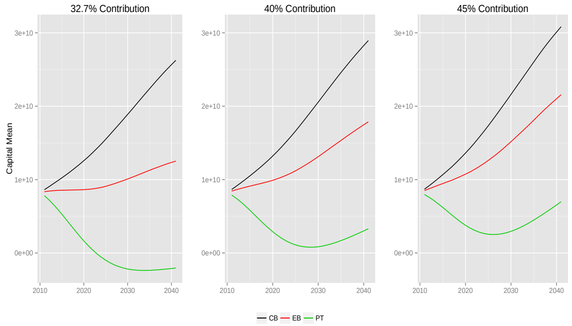

Figure 5.22: Social Security's capital lower bound (8.4 thousand million) for sub-model 2.

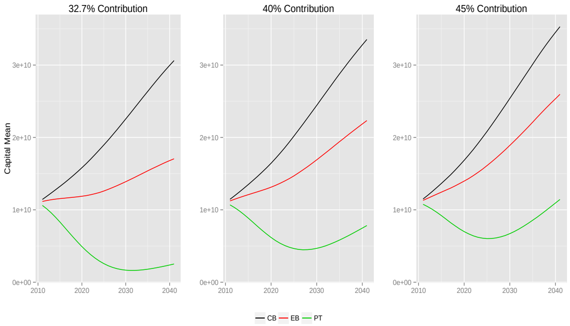

Figure 5.23: Social Security's capital upper bound (11.2 thousand million) for sub-model 2.

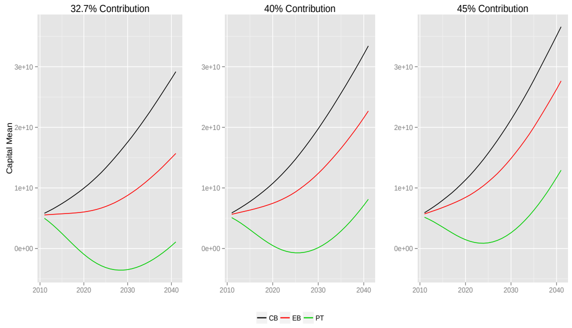

Figure 5.24: Social Security's capital lower bound (5.6 thousand million) for sub-model 3.

Figure 5.25: Social Security's capital upper bound (8.4 thousand million) for sub-model 3.

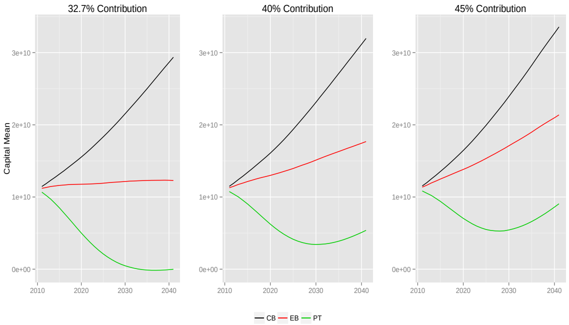

Figure 5.26: Social Security's capital lower bound (11.2 thousand million) for sub-model 4.

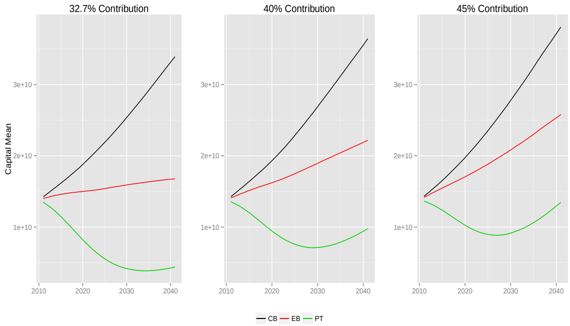

Figure 5.27: Social Security's capital upper bound (14 thousand million) for sub-model 4.

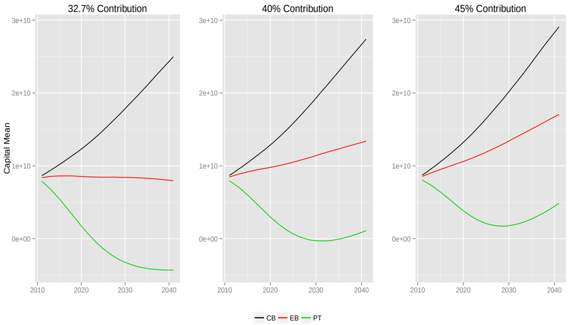

Figure 5.28: Social Security's capital lower bound (8.4 thousand million) for sub-model 5.

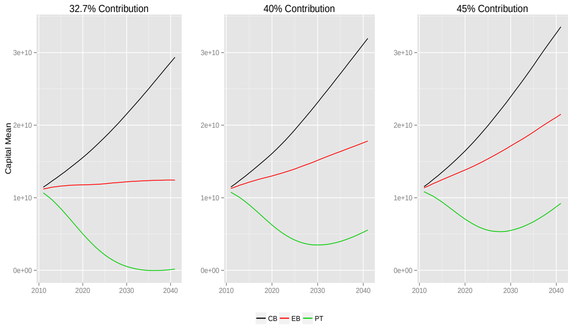

Figure 5.29: Social Security's capital upper bound (11.2 thousand million) for sub-model 5.

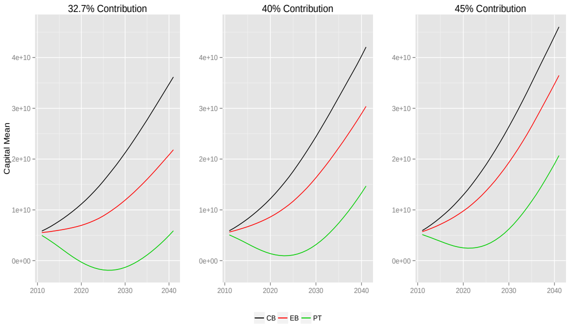

Figure 5.30: Social Security's capital lower bound (5.6 thousand million) for sub-model 6.

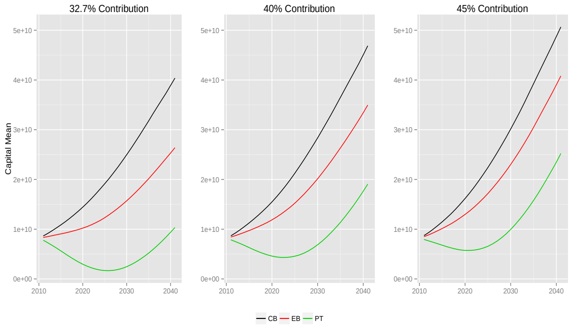

Figure 5.31: Social Security's capital upper bound (8.4 thousand million) for sub-model 6.

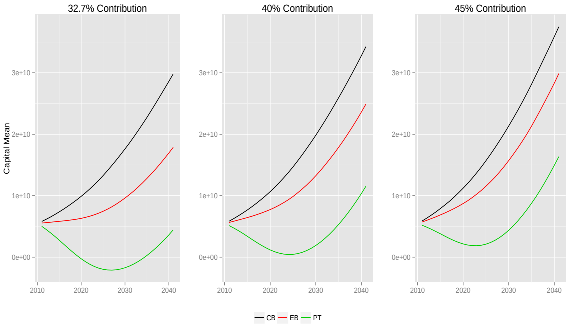

Figure 5.32: Social Security's capital lower bound (5.6 thousand million) for sub-model 7.

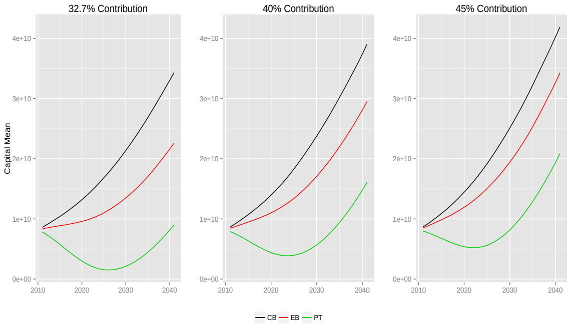

Figure 5.33: Social Security's capital upper bound (8.4 thousand million) for sub-model 7.

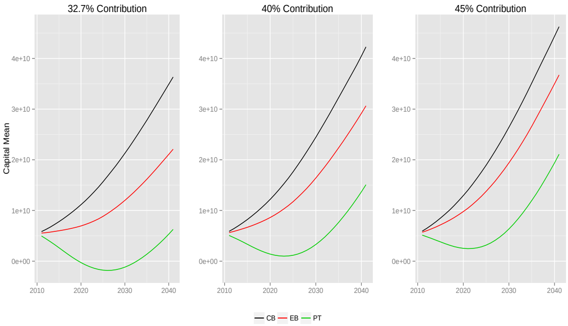

Figure 5.34: Social Security's capital lower bound (5.6 thousand million) for sub-model 8.

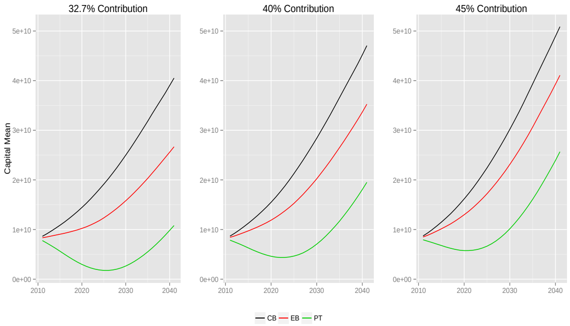

Figure 5.35: Social Security's capital upper bound (8.4 thousand million) for sub-model 8.

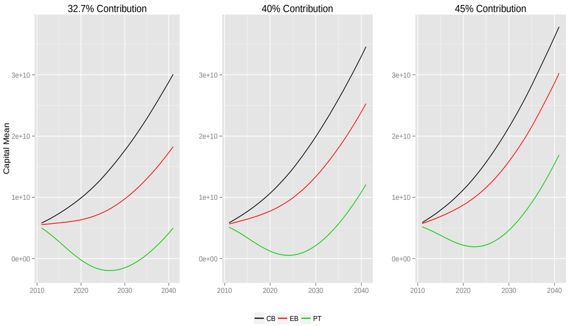

Figure 5.36: Social Security's capital lower bound (5.6 thousand million) for sub-model 9.

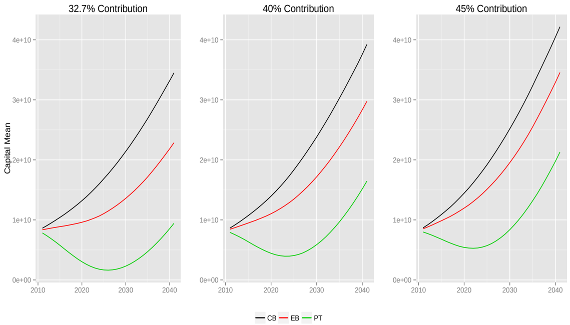

Figure 5.37: Social Security's capital upper bound (8.4 thousand million) for sub-model 9.

Figure 5.38: Social Security's capital lower bound (11.2 thousand million) for sub-model 10.

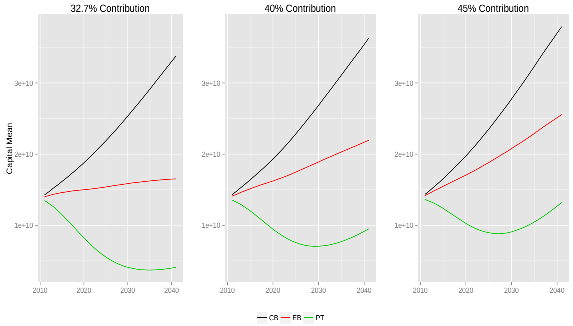

Figure 5.39: Social Security's capital upper bound (14 thousand million) for sub-model 10.

Figure 5.40: Social Security's capital lower bound (11.2 thousand million) for sub-model 11.

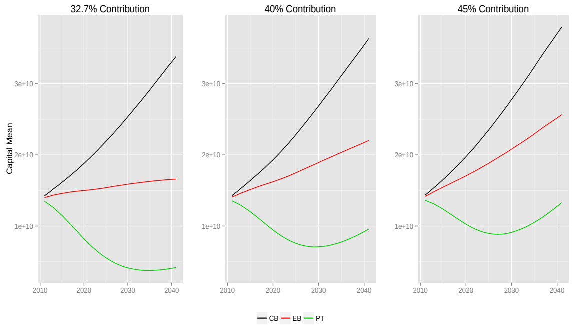

Figure 5.41: Social Security's capital upper bound (14 thousand million) for sub-model 11.

remaining for the social security, because this variable causes a big variability in the results.

In addition, the lower and upper bounds vary, depending on the model — this reveals a correlation between the sustainability of the social security and the inherent population age and employment structure, where the bounds are as low as the active-inactive dependency ratio. Through all submodels, a initial capital of 14 thousand million euros or higher is likely to guarantee the sustainability of the social security, while with a initial capital of 5.6 thousand million euros or lower it is likely to occur a bankruptcy of the same. So, while it is not possible to state if the Portuguese Social Security will be sustainable or not, it can be affirmed that if its capital is lower than 5.6 thousand million euros, it is unlikely to be sustainable.

Through all graphics it can also be noticed that using the earnings-based model or the contributions-base model, the Portuguese Social Security would be in a much safer position. However, it is not feasible to make a drastic change on the social security model. After all, besides having to maintain it sustainable, Portugal's government has to maintain its population happy and because of that, they try not to do severe changes on the pension system. This is clear from the Portuguese formulas; it can be deduced from the formulas that in earlier years, the Portuguese Social Security was using pension formulas similar to the ones of the earnings-based model, and, in later years, it is making use of formulae based on the contributions-based model.

Chapter 6

Discussion

In this chapter, a discussion over the main results will be presented, focusing on the detection of possible weaknesses of the proposed models and in suggestions on how to improve them. Also, some insights on the proposed models will be offered, showcasing the already known defects, characteristics and advantages.

6.1 Results

Population dynamics are a very complex system with high dependability between parameters. Any error committed predicting a year will be propagated to the following one. Finding better modeling approaches that minimize each year error is desired, and, for this, demographers and mathematicians must come together to discuss and improve each others approaches.

From the lack of data, all models presented here, except one, cannot be verified correctly for accuracy. Only insights and philosophical explanations can be shown to validate the models. Hopefully, within some years more data will be available, and the validation of a future model will be possible to be made. One of the best methods to verify if a model is correct is by initializing it in past years and check how far from the reality they are when predicting the present. In Section 5.3.2 on page 48 a mechanism for this evaluation is shown for which the data available is enough. As observed already from Table 5.7 on page 49, with the exception of the starting year of 2001, the error for the first 10 years is below 5 thousand persons and for the next 10 years the error is below 10 thousand persons. As for the starting year of 2001, the error for the first 10 years is near 15 thousand, which in comparison with the remainder ones is a very serious error. As previously stated, the period of 2001 until 2011 is marked with a positive net migration and that could produce a big influence on the results. On the remaining years, the greatest error is present on the ages below 10, 20 or 30, depending on the considered gap. This comes from the mic fertility model, which does not consider the fertility tendencies and, as visible on Figure 5.10 on page 50, the error is lower the closer the age is from the gap size. From these results, undoubtedly, a model incorporating the mic-mac fertility model and the migration model would make big changes to the predictions and possibly reduce the error.

With regard to all possible population's evolution that were evaluated, the general overview, which is supported by all the considered sub-models, is that the Portuguese Population will decrease in the years after 2011. This is also stated by Statistics Portugal [Sta15c] in a study called "Resident Population Projections" for the years 2012-2060. Figure 6.1 on page 69 is extracted from the aforementioned study, and shows that the Portuguese population projections will decrease until 2060.

However, this reduction is lighter than the ones presented in Subsection 5.3.4 on page 54 for the models with migration. This change is due to the migration predictions by Statistics Portugal that can be observed in Figure 6.2 on the following page, because Statistics Portugal predicts that the Portuguese net migration will increase starting on 2012 for either pessimistic and optimistic scenarios. Figure 5.19 on page 60 shows that, from the three sub-models with migration, only the sub-model with prosperous economy has a similar projection to Statistics Portugal's optimistic scenario. All the other migration projections showcased in this thesis are much more pessimistic than Statistics Portugal. Additionally, evaluating the tendency of the Portuguese net migration before 2011 in Figure 6.2 on the following page, the scenarios produced in this work seem to be the ones that better replicate that tendency.

The higher emigration explains why projections from this work are more pessimistic than the ones from Statistics Portugal. As already explained in Subsection 5.3.4 on page 54, the highest emigration volume is in fertile ages, so, along with an higher emigration, there is also a lower count of births in each year. Also, an high emigration in the first years deepens the effects on future years because newborns in the first years become mothers and fathers in the later ones.

Furthermore, Statistics Portugal's projections are in general optimistic on the evolution of fertility (Figure 6.3 on the following page) in which, considering the past fertility evolution trend, the pessimistic scenario is actually a very positive one. Although this predictions can be correct in the long term, closer to 2011 they seem very peculiar, where at least a slight decrease past 2011 should be expected before a possible recovery in subsequent years. In spite of this, the pessimistic hypothesis was used for the mac and mic-mac models for the fertility, as it was the most convincing scenario that was found in published data. Statistics Portugal's optimistic scenario is the result of 2013's Fecundity Survey conducted also by Statistics Portugal, where in average, women reported to be expecting 1.8 more children until 2060 besides their current children, which seems highly subjective and does not offer much confidence. In order to improve the sub-models without using previously predicted values for fertility, the following suggestion is presented.

Many studies were already conducted for Matching models; in this case, the matching would be between two persons (Kohler [Koh00], Billari and Kohler [BK04], Zinn [Zin12], White and Potter [WP13], Dribe, Hacker, and Scalone [DHS14]). Using social-economic values and past accounts on births from couples of Male-Female, Female-Female and Male-Male parents, a birth decision model could be designed. At the beginning of this project, this was one of the models that was to be created, but later it was found that adding this model was too complicated for the available time, mainly because a decent Matching model would require several months to be developed.

As for the Portuguese Social Security sustainability, the presented results show how much capital was needed in 2011 for the Portuguese Social Security to remain sustainable for the next 30 years for each of the considered sub-models. It is important to reinforce that the present values of Portuguese Social Security's capital are publicly unknown and, therefore, all that can be said is that, if the Portuguese Social Security's capital is lower than the lower bounds, then the odds are high that the system bankrupts. It is also important to notice that the Portuguese Social Security expenses are not just allocated to the retirement pension, since there are many other expenses for the system, such as unemployment subsidies or study scholarships. In the revenue's side, the Portuguese Social Security's capital is not only funded by the working contributions but, if needed, the State may move capital from other funds at its disposal.

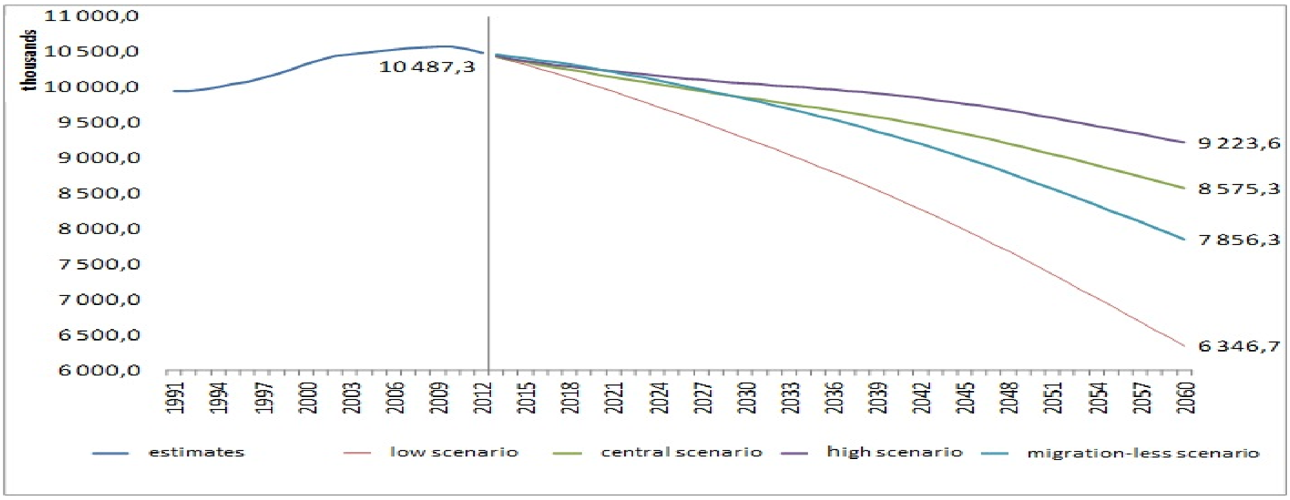

Figure 6.1: Statistics Portugal's Resident Population Projections from 2012 until 2060.

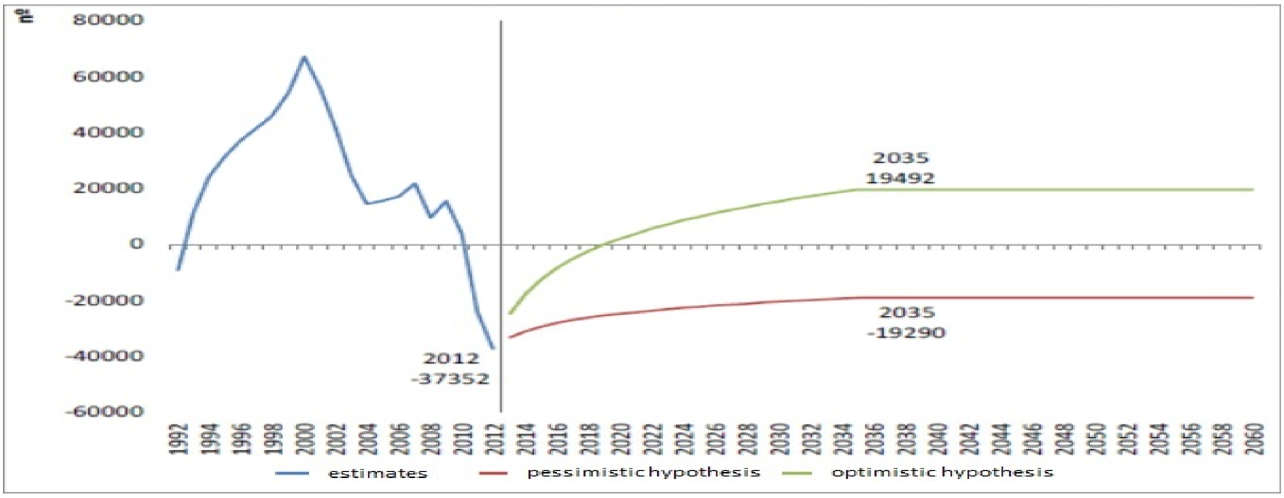

Figure 6.2: Statistics Portugal's Net Migrations Projections from 2012 until 2060.

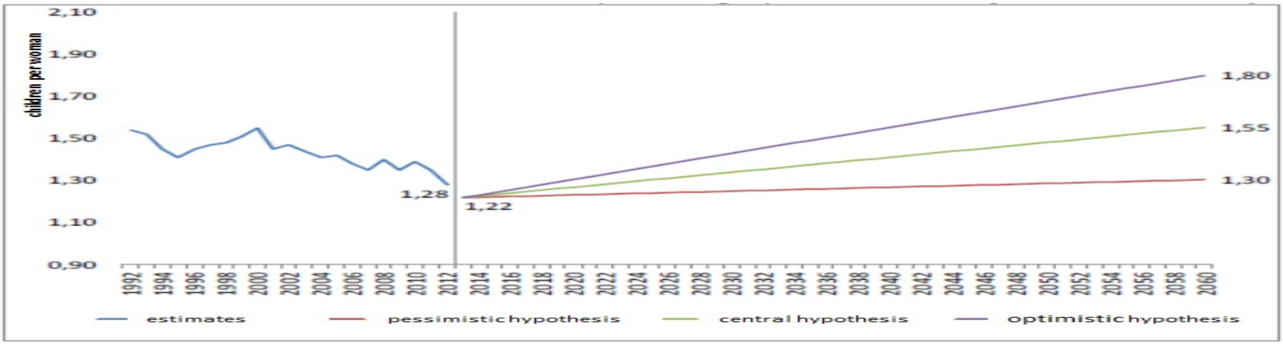

Figure 6.3: Statistics Portugal's Children per Woman Projections from 2012 until 2060.

6.2 Models

Many models for various situations were proposed. As already stated a model is an attempt to translate the reality and, obviously, a model is highly likely to not be 100% accurate. Some of their shortcomings and advantages were presented in the previous section. Now, the characteristics already expected to happen during their design and implementation will be presented.

Fertility (Section 3.2 on page 20) and mortality models (Section 3.3 on page 22) are very similar. Therefore, the discussion will not be focused on any of them but in the pair simultaneously. The mic model is a model with high variability for each year, since each year is very dependent on random effects — it is possible that two consecutive years present a very low and a very high fertility/mortality rate for any given age. But for a sufficiently big population per age, on average the sates will remain constant. This statement is due to the following line of thoughts. First, the process of having a child or dying is the same as a Bernoulli Process and for a sufficiently large population, this can also be seen as a binomial distribution. Second, for a sufficiently large number of trials n (or population in this case) and probability p, the binomial distribution $B(n, p)$ average occurrences equals to np and averaging this value through the population gives the yearly rate of p. So, for a sufficiently large population, the rates will remain constant. The mac model, however, has no variability and each simulation run produces the same global rate for each age. The mic-mac model maintains the same variability as the mic model, but on average tends to a previously predicted direction (either increases, stabilizes or decreases). Due to the already explained higher variability on the global rates for ages with small population density, in such cases, the mic-mac model should be forced to work exclusively with the mac model, as it is the case for the mortality model.

It is also important to discuss the closed population algorithm (Section 3.1 on page 19). In this algorithm, the most important detail is the order of the events. It was chosen to age the agents first, then to give birth to new agents and only after to do the random “killing”. The choice for the order was mainly to keep the algorithm simple while explaining the observable data. The existing data shows the number of deaths of persons of age 0 (newborns), which in the demographic field is a very important value, since it is used to evaluate the development indicators of a country; it also shows the number of births in the same year and persons alive at the end of the year. So, for simplicity, it was necessary to make the random “killing” to happen only after the births. And the aging comes first so that in the end of the simulation there are agents with age 0 (newborns that survived) and there are no agents of age 100 (which have all died at the “killing” step). The 99 age boundary was defined to control the population growth; if no boundary was defined, there would be a low (but positive) chance of existing agents of age 130 in the end of the simulation.

The emigration model attempts to translate the choice of a person to emigrate to a determined country through a gain function dependent on its own preferences (Subsection 3.5.1 on page 23). The preferences of a person are very diverse and it is hard to model them. One of the difficulties lays on the amount and type of the country’s aspects that influences a person. The chosen aspects were objective ones, which could be ordered by amount or distance, but a person decision is not only based on objective characteristics. For example, the climate was not taken into consideration for this model, but it is an important characteristic when choosing a destination country. The climate could not be used because it is a subjective characteristic and the country ordering through climate would be different for each person. Also, many objective characteristics were not used to maintain some simplicity. One other problem is the modeling of the person’s preferences. Even if all possible countries' characteristics were present, a person does not make its choice using all available possibilities. A characteristic could be unimportant to a person while for another it could be one of the most important characteristics. A good way to deal with that would be through a large scale survey to the Portuguese population, to obtain the personal interests of a significant number of persons.

For the immigration model (Subsection 3.5.2 on page 27), however, there is little to be said. It is not possible to make a decision model for agents that do not exist in the model, therefore the model has to be completely exogenous. Other modeling ways could be considered, like using Markov Chains or incomplete/censored data regressions; also more variables could be used, like Portuguese economic scenario and emigration volume. The model used here is not perfect but it is a very simple model and it is easy to compute.

For the employment model, there is both the mic model (Subsection 4.1.1 on page 30) and the mic-mac model (Subsection 4.1.2 on page 30). The mic model will not be discussed further, because the behavior is the same of the mic models for fertility and mortality. For the mic-mac model, some considerations must be made. This model makes the age-specific employment rate change; this is due to the fact that this model, instead of using the employment rates, uses the work tenure distributions. This makes the model much more dynamic than simpler models. This model can also be used together with legislative rules for work contract periods instead of just using work tenure. However, this model is highly dependent on the chosen distributions and a proper study of the data is very important before using it. A clear disadvantage of using this model is the necessary data. Ideally, the data should be age-specific, but that is hard to find; furthermore, the data must also be ordered on the date of contract. Therefore, this model is more delicate and its usage should be done with some care, so the produced results are the ones intended.

Another thing worth discussing, besides the models themselves, is the programming language that was used. Initially this model was programmed in NetLogo (Wilensky [Wil99]). As it is a prototyping programming language, the model development was quite fast and easy, but as the model became more and more complex, problems started to appear. The first problem was the memory usage — with the mic models alone, along with the Earnings-based model, the model, in NetLogo, required 2GB of available memory to function. This was a problem because NetLogo standard memory usage limit was 1GB, but with a simple manipulation on the batch files of NetLogo it was possible to increase this limit. With the same basic sub-model, after solving the memory issue, another problem arose. This time was the computing time — a single simulation run required about 1 hour and 30 minutes to be completed. This problem was then solved with the help of the Faculty of Engineering of the University of Porto's Grid, which allowed about 100 simulations to be run at the same time. This was however the most problematic issue. At the time, more sub-models were already planned to be modeled and running several simulations with them would become a major problem. To solve this problem, the whole model was reprogrammed in Java [AG98]. In this new programming language, the model's memory requirement fell bellow the 1GB limit and a simulation run time reduce from the 1 hour and 30 minutes to a stunning 30 seconds; and with further improvements to the code, the computing time was reduced even more, reaching the minimum time of 5 seconds per simulation run. Now, with the more complex sub-models, the memory usage remains a problem, requiring 4GB of memory, but the run time is no longer a problem since each simulation has been taking at most 1 minute.

Chapter 7

Conclusion

Population growth has been studied for centuries; many models and many approaches have been tried ever since. In the beginning, linear models were used but they were unable to capture the entire reality. Then, models became more and more complex, using more variables, which have brought a large improvement on population growth predictions. Nowadays, besides modeling population growth, the growth for each age and gender is also modeled, as is fertility and mortality rates and also migrations. Without a doubt, models are now able to capture more parts of the reality. However, as stated by Coale and Trussell [CT96], although more parts are explained and models became more sophisticated, Human population growth models have not improved significantly so far.

However, Human population growth is a very uncertain phenomenon, and is very different from general population models. Human population is dependent on their members and the actions of those members are almost random to an outside observer. Of course people's choices are not made by chance, however because local knowledge is distributed, the reasoning behind each person's decisions is not fully known to the statistician. Here, this work presented a model for the modeling of a person's actions. Far from being perfect, it is, nevertheless, a starting point.

A core algorithm was created for people's actions, accepting multiple models for each life components. For now, the fertility part is not based on the decisions of a person but is more of a random effect with some control. It has three different models to choose from, ranging from a random model to a fully controlled model. The mortality part is also not modeled as a personal decision, but neither should it fully be, as the time of death of a person is partly decision-based (how healthy a person lives, for example) but is also random (a healthy person may have a car accident and die, for example). This component also has three different models to choose from, ranging from a random model to a fully controlled model. The migration model, however, is based on a person's decision capacity. There are many approaches on how to model the decision of a person; here, a gain maximization algorithm was used. This algorithm uses the person's preferences (for health, safety, ...), its life history (employment status and remuneration) and exogenous factors (age, economic scenario, ...).

Independently of the combination of the chosen models, the predictions suggest that there will be a population decrease on the next 30 years, but the slope varies. A sample size of 2% of the population was simulated. The predicted population reduction ranged from 10% of the total population to 50%. The different outcomes were largely influenced by the usage of the migration model and different economic scenarios. If the migration model is not present, the reduction appears to be linear, but adding the migration model, the reduction appears to have a quadratic or cubic behavior. In addition, when the migration model is used, the population growth seems to stabilize towards the end of the simulation, which sheds light into the future of the Portuguese population. If nothing more, these results show the importance of using decision agent-based models in detriment to deterministic or purely stochastic models.

The Portuguese Social Security sustainability is an important issue for most Portuguese. It is a hot discussion topic in Portugal, largely observed in news reports. Studies from people of many different areas, like mathematics, economy, politics and so on..., were presented to the general public. However, there is a big problem; there are many studies which state that the Portuguese Social Security will bankrupt in the following years, while others assert it will maintain its sustainability. Joining in this hot topic, this work also sheds light on this issue.

There are plenty uncertainties related to the Portuguese Social Security sustainability and many facts are not made available to the general public. Therefore, in this work, there is no attempt at making a definitive statement on the topic. All that is supplied are possible scenarios; and with those scenarios some insights are provided. The big question to be answered is if the Portuguese Social Security is currently sustainable. The given answer here is: it depends. This work reveals two great variables which will affect the Portuguese Social Security outcome. One is the age-specific growth of the Portuguese population and its respective work structure; the other is the remaining capital on the Portuguese Social Security funds nowadays. Alongside these two great variables, there are two obvious conclusions to return. First, the higher the emigration, the worse the Portuguese Social Security sustainability will endure, as emigration is mainly focused on the active population. Second, the least the capital remaining nowadays, the higher the chances of bankruptcy.

Additionally, two important and more precise insights can be drawn from this work. An upper and lower bound were found for the social security's capital in 2011. The upper bound is 14 thousand million euros and if the value at 2011 was higher than that, then the sustainability of the social security is a probable event. The lower bound is 5 thousand million euros and if the value at 2011 was lower than that, then it is most likely that the Portuguese Social Security is unsustainable. It is also important to remember that the simulations were run for a sample of 2% of the population and so, the results provided are also for a 2% sample. Therefore, extrapolating the results to the total Portuguese population, the upper and lower bounds would be 700 thousand million euros and 250 thousand million euros respectively.

Bibliography

[06] *Decreto de Lei nº 53-B/2006 de 29 de Dezembro, Diário da República nº 11 - Série I Lisboa, pp. 8626(388)-8626(390)*. 2006 (cited on p. 33).

[07a] *Decreto de Lei nº 187/2007 de 10 de Maio, Diário da República nº 90 - Série I Lisboa, pp. 3100-3116*. 2007 (cited on p. 31).

[07b] *Decreto de Lei nº 4/2007 de 16 de Janeiro, Diário da República nº 249 - Série I Lisboa, pp. 345-356*. 2007 (cited on p. 14).

[13] *Portaria nº 281/2013 de 28 de Agosto, Diário da República nº 165 - Série I Lisboa, pp. 5196-5198*. 2013 (cited on p. 31).

[AC10] C. Anjos and P. Campos. "The role of social networks in the projection of international migration flows : an agent-based approach". In: *Joint Eurostat-UNECE Work Session on Demographic Projections* 28-30 April, Lisbon, Portugal (2010) (cited on p. 23).

[AF14] A. Adsera and A. Ferrer. "Factors influencing the fertility choices of child immigrants in Canada". In: *Population Studies* 68.1 (2014), pp. 65–79. DOI: 10.1080/00324728.2013.802007 (cited on pp. 27, 28).

[AG98] K. Arnold and J. Gosling. *The Java Programming Language*. Addison-Wesley, 1998 (cited on p. 71).

[Ale09] H. A. Alencar. *Benefícios previdenciários*. São Paulo: Universitária de Direito, 2009 (cited on p. 15).

[Ame89] J. Amegandjin. *Démographie Mathematique*. Paris: Economica, 1989 (cited on p. 13).

[AP87] J. Addison and P. Portugal. "On the Distributional Shape of Unemployment Duration". In: *Review of Economics and Statistics* 69 (1987) (cited on p. 41).

[Ban15] Banco de Portugal. *Banco de Portugal*. 2015. URL: https://www.bportugal.pt/en-US/Pages/inicio.aspx (visited on 01/01/2015) (cited on p. 38).

[BK04] F. C. Billari and H.-P. Kohler. "Patterns of low and lowest-low fertility in Europe." In: *Population studies* 58.2 (2004), pp. 161–176. DOI: 10.1080/00324720500099843 (cited on p. 68).

[BM13] M. Belloni and C. Maccheroni. "Actuarial Fairness When Longevity Increases: An Evaluation of the Italian Pension System". In: *Geneva Papers on Risk and Insurance-Issues and Practice* 38 (2013), pp. 638–674. DOI: 10.1057/gpp.2013.27 (cited on p. 15).

[BOP03] F. Billari, F. Ongaro, and A. Prskawetz. “Introduction: Agent-Based Computational Demography”. In: *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*. New York: Springer, 2003. Chap. 1 (cited on p. 7).

[Bor01] M. Borella. “Social Security Systems And The Distribution of Income : An Application To The Italian Case”. 2001 (cited on p. 15).

[BP83] J. Bongaarts and R. G. Potter. *Fertility, Biology, and Behavior: An analysis of the proximate determinants*. New York: Academic Press, 1983 (cited on p. 9).

[Bra08] H. L. Bras. *The Nature Of Demography*. Princeton: Princeton University Press, 2008 (cited on pp. 12, 13, 52).

[Bra60] W. Brass. “The graduation of fertility distributions by polynomial functions”. In: *Population Studies* 14.2 (1960), pp. 148–162 (cited on p. 11).

[Bra81] W. Brass. “The use of the relational Gompertz model to estimate fertility”. In: *International Union for the Scientific Study of Population* (1981), pp. 345–362 (cited on p. 11).

[BWF14] V. Baláž, A. M. Williams, and E. Fifeková. “Migration Decision Making as Complex Choice: Eliciting Decision Weights Under Conditions of Imperfect and Complex Information Through Experimental Methods”. In: *Population, Space and Place* (2014). DOI: `10.1002/psp.1858` (cited on p. 23).

[Cam00] P. Campos. “Os Modelos Determísticos e a População Estável”. In: *Actas do I Congresso Portugês de Demografia*. Tróia, 2000 (cited on p. 14).

[Cas01] H. Caswell. *Matrix population models: construction, analysis, and interpretation*. Vol. 2nd. 2001, p. 722 (cited on pp. 7–9, 12–14).

[Cha03] E. Chattoe. “The role of agent-based modelling in demographic explanation”. In: *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*. 2003, pp. 41–54 (cited on p. 7).

[CIA13] CIA. *The World Factbook 2013-14*. 2013. URL: `https://www.cia.gov/library/publications/the-world-factbook/index.html` (visited on 01/01/2015) (cited on pp. 1, 40).

[CM72] A. Coale and D. R. McNeil. “The distribution by age of the frequency of the first marriage in a female cohort”. In: *Journal of the American Statistical Association* 67.340 (1972), pp. 743–749 (cited on p. 10).

[Coh79] J. E. Cohen. *Ergodic theorems in demography*. 1979. DOI: `10.1090/S0273-0979-1979-14594-4` (cited on p. 13).

[Com15] Comptroller General of the Union. *Comptroller General of the Union*. 2015. URL: `http://www.cgu.gov.br/` (visited on 01/01/2015) (cited on p. 40).

[CT96] A. Coale and J. Trussell. “The development and use of demographic models.” In: *Population studies* 50.3 (1996), pp. 469–484. DOI: `10.1080/0032472031000149576` (cited on pp. 1, 6, 8–10, 12, 14, 73).

[Dev86] L. Devroye. *Non-Uniform Random Variate Generation*. 1st Editio. Springer, 1986 (cited on p. 35).

[DHS14] M. Dribe, J. D. Hacker, and F. Scalone. “The impact of socio-economic status on net fertility during the historical fertility decline: A comparative analysis of Canada, Iceland, Sweden, Norway, and the USA.” In: *Population studies* 68.2 (2014), pp. 135–149. DOI: `10.1080/00324728.2014.889741` (cited on p. 68).

[Edi14] D. M. Ediev. “Why increasing longevity may favour a PAYG pension system over a funded system.” In: *Population studies* 68.1 (2014), pp. 95–110. DOI: `10.1080/00324728.2013.780632` (cited on p. 16).

[Emi15] Emigration Observatory. *Emigration Observatory*. 2015. URL: `http://www.observatorioemigraca` `pt/np4/home.html` (visited on 01/01/2015) (cited on p. 40).

[Eur15] Eurostat. *Eurostat*. 2015. URL: `http://ec.europa.eu/eurostat/data/database` (visited on 01/01/2015) (cited on p. 40).

[Fra15] France’s National Institute for Statistics and Economic Studies. *France’s National Institute for Statistics and Economic Studies*. 2015. URL: `http://www.insee.fr/en/default.asp` (visited on 01/01/2015) (cited on p. 40).

[GBW05] N. Gaag, J. Beer, and F. Willekens. “MicMac Combining micro and macro approaches in demographic forecasting”. In: *Joint Eurostat-ECE Work Session on Demographic Projections*. Vienna, 2005 (cited on pp. 6, 8).

[GEE12] GEE. *Office for strategy and studies*. 2012. URL: `www.gee.min-ecoomia.pt/en/` (visited on 01/01/2014) (cited on pp. 34, 41).

[Gom25] B. Gompertz. “On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies”. In: *Philosophical Transactions of the Royal Society of London* 115 (1825), pp. 513–583 (cited on p. 10).

[Gra62] J. Graunt. *Natural and Political Observations Made Upon the Bills of Mortality*. 1662 (cited on p. 10).

[Hei03] F. Heiland. “The collapse of the Berlin Wall: Simulating state-level east to West German migration patterns”. In: *Agent-Based Computational Demography* (2003), pp. 73–96 (cited on p. 7).

[Hen61] L. Henry. “Some data on natural fertility”. In: *Eugenics Quarterly* 8.1 (1961), pp. 81–91 (cited on p. 11).

[Her72] G. Hernes. “The process of entry into first marriage”. In: *American Sociological Review* 37.2 (1972), pp. 173–182 (cited on p. 10).

[Hil81] K. Hill. “A proposal for the use of information on residence of siblings to estimate emigration by age”. In: *International Union for the Scientific Study of Population* 18 (1981), pp. 19–34 (cited on p. 11).

[IGF15] IGFSE. *Agency for Development and Cohesion*. 2015. URL: `http://www.igfse.pt/` (visited on 01/01/2015) (cited on p. 41).

[Inv15] Investopedia. *Investopedia*. 2015. URL: `http://www.investopedia.com/` (visited on 01/01/2015) (cited on p. 15).

[Jap15] Japan Institute for Labour Policy and Training. *Japan Institute for Labour Policy and Training*. 2015. URL: `http://www.jil.go.jp/english/index.html` (visited on 01/01/2015) (cited on p. 40).

[KC05] N. Keyfitz and H. Caswell. *Applied Mathematical Demography*. Statistics for Biology and Health. New York: Springer-Verlag, 2005. DOI: `10.1007/b139042` (cited on pp. 12, 13).

[Key77] N. Keyfitz. *Applied Mathematical Demography*. New York: John Wiley and Sons, 1977 (cited on p. 9).

[Kie89] A. Kieser. “Organisational, Institutional and Societal Evolution: Medieval Craft Guilds and the Genesis of Formal Organizations”. In: *Administrative Science Quarterly* 34 (1989), pp. 540–564. DOI: `10.2307/2393566` (cited on p. 15).

[Koh00] H.-P. Kohler. *Social interactions and fluctuations in birth rates*. 2000. DOI: `10.1080/713779084` (cited on p. 68).

[Leg05] A. M. Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: Huzard-Courcier, 1805, p. 80 (cited on p. 6).

[Lot10] A. J. Lotka. “Contribution to the Theory of Periodic Reaction”. In: *Journal of Physical Chemistry A* 14.3 (1910), pp. 271–274. DOI: `10.1021/j150111a004` (cited on p. 6).

[Mak60] W. M. Makeham. “On the Law of Mortality and the Construction of Annuity Tables”. In: *The Assurance Magazine, and Journal of the Institute of Actuaries* 8.6 (1860), pp. 301–310 (cited on p. 10).

[Mal98] T. R. Malthus. *An Essay on the Principle of Population*. 1st editio. London: J. Johnson, 1798 (cited on p. 6).

[Mar10] S. P. Martins. *Direito da seguridade social*. São Paulo: Atlas, 2010 (cited on p. 15).

[MM12] C. Mazzaferro and M. Morciano. “A Dynamic Micro-simulation Model for the Italian Social Security System”. In: *CAPPapers* 48 (2012) (cited on pp. 16, 32).

[Mou15] J.-L. Moussier. *French special retirement plan*. 2015. URL: `http://www.regimesspeciaux.org/spip.php?article1` (cited on p. 15).

[MS13] L. Mayhew and D. Smith. “A new method of projecting populations based on trends in life expectancy and survival”. In: *Population Studies* 67.2 (2013), pp. 157–170. DOI: `10.1080/00324728.2012.740500` (cited on p. 11).

[Nat07] National Union of Public and General Employees. *A Brief History of Pensions in Canada*. 2007 (cited on p. 15).

[Nat12] National Assembly. “The Declaration of the Rights of Man and of the Citizen (1789)”. In: *Sources of the making of the West, Volume II: Since 1500*. Vol. II. 2012, pp. 118–124 (cited on p. 15).

[NI15] G. National Institute of Statistics and Informatics. *National Institute of Statistics, Geography and Informatics*. 2015. URL: `http://www.inegi.org.mx/` (visited on 01/01/2015) (cited on p. 40).

[NUR47] J. von Neumann, S. Ulam, and R. D. Richtmmyer. *Statistical Methods on Neutron Diffusion*. Tech. rep. Los Alamos: Los Alamos Scientific Laboratory, 1947 (cited on p. 6).

[OEC15] OECD. *OECD Statistics*. 2015. URL: `stats.oecd.org` (visited on 01/01/2015) (cited on pp. 40, 41).

[Pag77] H. Page. “Patterns underlying fertility schedules: A decomposition by both age and marriage duration”. In: *Population Studies* 31.1 (1977), pp. 85–106 (cited on p. 11).

[PHG01] S. H. Preston, P. Heuveline, and M. Guillot. *Demography: Measuring and Modeling Population Processes*. Vol. 57. 3. 2001, xiii, 291 p. DOI: `10.2307/1535065` (cited on pp. 7, 8, 12, 13, 53).

[Por15] Portugal Government. *Governo de Portugal*. 2015. URL: `http://www.portugal.gov.pt/` (visited on 01/01/2015) (cited on p. 15).

[Pre83] R. Pressat. *L'analyse Démographique*. 4th Éditio. Paris: Presses Universitaires de France, 1983 (cited on p. 13).

[RC81] A. Rogers and L. Castro. "Model Migration Schedules". 1981 (cited on p. 11).

[Rea09] D. Rea. "Government Expenditure and Revenue in New Zealand: A Brief Overview". In: *Policy Quartely* 5.3 (2009), pp. 58–67 (cited on p. 40).

[SSS73] E. G. Stockwell, H. S. Shryock, and J. S. Siegel. *The Methods and Materials of Demography.* 1973. DOI: `10.2307/2060757` (cited on p. 13).

[Sta15a] Statistics Austria. *Statistics Austria*. 2015. URL: `http://www.statistik.at/web%5C_en/statistics/index.html` (visited on 01/01/2015) (cited on p. 40).

[Sta15b] Statistics Canada. *Statistics Canada*. 2015. URL: `http://www.statcan.gc.ca/start-debut-eng.html` (visited on 01/01/2015) (cited on p. 40).

[Sta15c] Statistics Portugal. *Statistics Portugal*. 2015. URL: `www.ine.pt/en/` (visited on 01/01/2015) (cited on pp. i, iii, 34, 38–41, 49, 67).

[Tap85] G. Tapinos. *Éléments De Démographie*. Paris: Armand Colin, 1985 (cited on p. 13).

[Tra15] Trading Economics. *Trading Economics*. 2015. URL: `http://www.tradingeconomics.com/` (visited on 01/01/2015) (cited on p. 41).

[UN15] UN. *United Nations Statistics Division*. 2015. URL: `unstat.un.org` (visited on 01/01/2015) (cited on p. 40).

[Uni15] United States of America Government. *The United States Social Security Administration*. 2015. URL: `http://www.ssa.gov/` (visited on 01/01/2015) (cited on p. 15).

[Ver38] P. F. Verhulst. "Notice sur la loi que la populations suit dans son accroissement". In: *Correspondence mathematique et physique* X (1838) (cited on p. 6).

[Wil99] U. Wilensky. "NetLogo. http://ccl.northwestern.edu/netlogo/." In: *Center for Connected Learning and ComputerBased Modeling Northwestern University Evanston IL* 2009.26.02.2009 (1999), Evanston, IL (cited on pp. 7, 71).

[Woo02] M. Wooldridge. *An Introduction to MultiAgent Systems*. Sussex: John Wiley and Sons, 2002 (cited on p. 7).

[Wor15] World Bank. *World Bank*. 2015. URL: `http://www.worldbank.org/` (visited on 01/01/2015) (cited on p. 40).

[WP13] K. White and J. E. Potter. *The impact of outmigration of men on fertility and marriage in the migrant-sending states of Mexico, 1995–2000*. 2013. DOI: `10.1080/00324728.2012.721518` (cited on p. 68).

[Zab85] B. Zaba. "A parameterized procedure for projecting population". In: *Union for the Scientfic Study of Population* (1985), pp. 137–159 (cited on p. 11).

[Zin12] S. Zinn. “A Mate-Matching Algorithm for Continuous-Time Microsimulation Models”. In: *International journal of microsimulation* 5.1 (2012), pp. 31–51 (cited on p. 68).