Long-term care insurance: a multi-state semi-Markov model to describe the dependency process for elderly people

Guillaume BIESSY[1]

EURo-Institut d'Actuariat[2]

SCOR Global Life[3]

Abstract

The models used today for the pricing of long-term care insurance products often consider dependency as a unique homogeneous state. As the risk carried by long-term care insurers is not yet well-known, we think that a multi-state model, taking into account several levels of dependency, will allow for a more accurate picture of this risk. A semi-Markov process is a multi-state process whose probabilities of transition depend not only on the current state but also on the time spent in this state. This process is core to numerous applications e.g. in epidemiology but its use in relation to long-term care insurance is still very limited.

The present article aims at providing the reader with the construction process of a 4 states semi-Markov model. This work is based on data from the French long-term care public aid: the "Allocation personnalisée d'autonomie" (APA). Firstly, we will introduce the various parameters used to model transitions between states. We will then proceed to the calibration of those parameters, using a likelihood maximization method, while taking into account the peculiarities of the APA data set. Finally, we will apply the model to the pricing of a new long-term care insurance product, using a Monte Carlo method.

Résumé

Les modèles utilisés aujourd'hui pour la tarification des contrats d'assurance dépendance considèrent souvent la dépendance comme un état unique et homogène. A l'heure où le risque de dépendance est encore méconnu, nous pensons qu'une modélisation multi-états, fondée sur la prise en compte de plusieurs niveaux de dépendance, permettrait une meilleure maîtrise de ce risque. Un processus multi-états est dit semi-markovien lorsque les probabilités de transition du processus dépendent à la fois de l'état actuel et du temps passé dans cet état. De tels processus ont trouvé de nombreuses applications par exemple en épidémiologie mais leur utilisation en assurance dépendance demeure aujourd'hui très limitée.

Cet article a pour but de présenter la construction d'un modèle semi-markovien considérant 4 niveaux de dépendance. Ce travail s'appuie sur des données recueillies dans le cadre de l'Allocation personnalisée d'autonomie (APA). Tout d'abord, nous introduisons les paramètres intervenant dans notre modélisation des transitions entre les différents états du modèle. Nous procédons alors à l'estimation de ces paramètres par la méthode du maximum de vraisemblance, en tenant compte des spécificités liées aux données APA. Enfin, nous proposons une application du modèle à la tarification d'un nouveau produit d'assurance dépendance, à l'aide d'une méthode de type Monte Carlo.

Keywords: APA data, heavily censored data, long-term care insurance, maximum likelihood, Monte Carlo, semi-Markov processes.

[1]Contact: gbiessy@scor.com.

[2]EURIA - 6 avenue le Gorgeu - 29238 Brest Cedex 3 - France

[3]SCOR Global Life SE - 5 avenue Kléber - 75795 Paris Cedex 16 - France

1 Introduction

1.1 High stakes

In developed countries, since the beginning of the 20$^{\text{th}}$ century, there has been a steady increase in life expectancy of around one quarter every year. This, along with the aging of the baby-boomer generations has resulted in the expectation that the number of people aged of 65 or more will double by 2060 (see Blanpain and Chardon, 2010). Among other consequences, this will drastically increase the number of dependent elderly people.

The first long-term care insurance products appeared in the 1980s, which is very late compared to other types of insurance, and presented a maximum age of subscription. As a result, the amount of data available to insurers is still very limited, especially for higher ages, and so is the knowledge insurers have of the underlying risk. Under the above circumstances, more complex models for the dependency process need to be developed.

1.2 About current models

Most insurers nowadays use discrete models with 3 states: autonomy, dependency and death. Figure 1 shows an example of such a model.

Three different laws are used in this model:

- $i(s)$ is the dependency incidence rate. It is a discrete annual law that only depends on age. Empirical data shows that incidence rate increase exponentially with respect to the age s of the subject.
- $q^a(s)$ is the mortality rate for autonomous people.
- $q^d(s,x)$ is the mortality rate for dependent people. It depends not only on the age s of entry in dependency, but also on the time x spent in the dependency state.

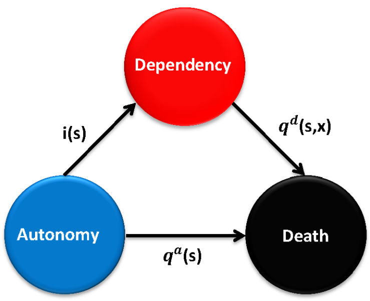


Figure 1: A simple model with 3 states: autonomy, dependency and death.

Models such as the one previously introduced are simple and reliable. However they only consider a single level of dependency. Nowadays, most long-term care insurance products offer two levels of guarantee with regard to the seriousness of the dependency state: partial or total dependency. To price such products, and because of the lack of available data, insurers often combine two separate models, one for partial dependency and the other for total dependency. In doing so, they make the approximation that the incidence rate toward total dependency is the same for both autonomous and partially dependent people. This leads to a simplistic view of the risk which results in less visible long-term trends for mortality and incidence rates.

1.3 Toward a new model

In this article, we will consider a model with 4 different states of dependency. To define these states we will rely on the "Autonomie, gérontologique, groupes iso-ressources" (AGGIR) grid. The AGGIR grid aims to categorize people by groups of similar needs based on dependency and is used in France for the attribution of the public aid for dependent elderly people, the APA, created in 2002. This grid describes 6 levels of dependency, from the more severe level GIR 1, which corresponds to people who are confined to bed or to a wheelchair and have lost their mental faculties, to GIR 6 which corresponds to nearly autonomous people. To determine to which group one belongs, the ability to perform activities of daily living is assessed, and these abilities are then weighted and factored into a complex algorithm (see Vetel et al., 1998). Only people in states GIR 1 to GIR 4 can benefit from the public aid, and we will only consider these levels as dependency in our model. The choice of an AGGIR-based definition is key, as it allows us to use data from the APA, that is gathered on the whole population and proves to be richer than most insurer's portfolio experience.

In our model, as for the mortality rate in dependency in the previous model, we will consider transition laws originating from dependency states as functions of both the age and the time spent in the current dependency state. As we will also work with a continuous time scale, such a model is called a multi-state semi-Markov model (see Cinlar, 1969; Janssen and Manca, 2002). Semi-Markov processes have been widely and successfully used in the field of epidemiology, to model for example the HIV disease, as in Mathieu, 2006. Although their use to model the dependency process may seem very natural under those circumstances, as far as we know, it has only be done in Lepez, 2006, the main reason for this being the lack of available data.

Finally, we will consider in our model that there is no return to a better health status, for several reasons:

- According to the French definition, dependency is a consolidated and irreversible state.
- From an insurer point of view, it is prudent to assume that an insured life with an improvement in his/her health status may not declare it. Also, performing regular checks for such improvements would prove far too expensive.
- From a model point of view, considering returns would result in doubling the number of transitions, and therefore raises computation time and parameter identification issues.

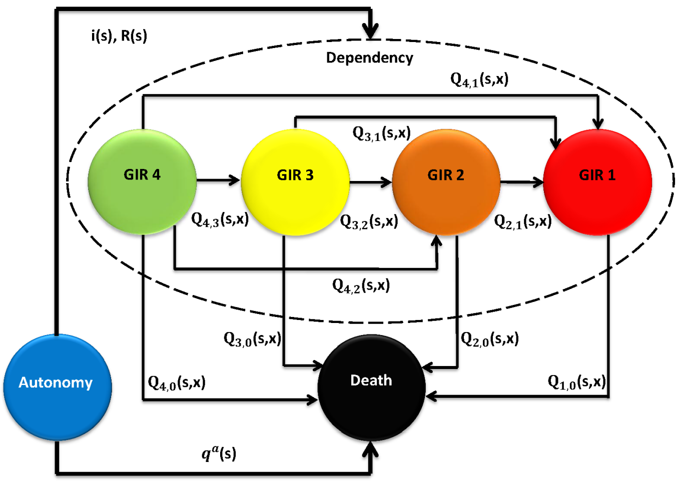


Figure 2: The new model with 6 states: autonomy, 4 levels of dependency, and death. A definition of the semi-Markov kernel $Q_{i,j}(s,x)$ where $i > j$, $0 \leq s, x$, is provided in the following section.

Figure 2 shows a representation of the new model, where some elements are the same as for the previous 3-states model. Indeed, we will continue to use the same mortality rate for autonomous people $q^a(s)$, and the same incidence rate $i(s)$ though we will need to complement the latter by adding a multidimensional function $R(s)$, which will give the probabilities, knowing the age s of entry, for dependency to begin in each state.

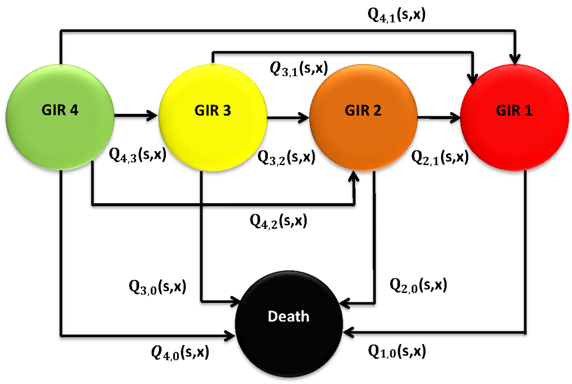


Figure 3: The semi-Markov sub-model with only transitions originating from dependency states.

In the following section, we will focus on the construction of the semi-Markov sub-model composed of transitions originating from dependency states. Figure 3 shows a representation of this sub-model, which has 5 states and 10 transitions.

2 Description of the new model

2.1 Elements of semi-Markov theory

Definition. *Let $X = (X_t)_{t\geq 0}$ be a continuous time process which takes its value in a finite set of states E and $(\mathcal{F}_t)_{t\geq 0}$ be the filtration associated to X. For each time $t \geq 0$, we denote by N_t the number of transitions made by X between 0 and t, and by τ_{N_t} the time of the N_t transition (with $\tau_0 = 0$). X is called a semi-Markov process if:*

$$\forall n \in \mathbb{N},\ \forall x > 0,\ \forall j \in E,\ P(\tau_{n+1} - \tau_n \leq x, X_{\tau_{n+1}} = j|\mathcal{F}_{\tau_n}) = P(\tau_{n+1} - \tau_n \leq x, X_{\tau_{n+1}} = j|X_{\tau_n}).$$

In other words, at times of transition, the future of the process is only determined by its current state. Information about the past of the process is irrelevant.

With the same notations, if X is a semi-Markov process, the following property is verified:

$$\forall j \in E,\ \forall\ 0 \leq s \leq t,\ P(X_t = j|\mathcal{F}_s) = P(X_t = j|X_s, s - \tau_{N_s}).$$

In other words, the future of the process is only determined by its current state and the time spent in that state. Markov processes are often described as processes with no memory. In comparison, semi-Markov processes do have a memory, although this memory is reset every time the process enters a new state.

Definition. *The semi-Markov kernel is defined by:*

$$\forall i, j \in E,\ \forall\ 0 \leq s,\ x, \quad Q_{i,j}(s,x) = P\left(X_{N_s+1} = j, \tau_{N_s+1} \leq s + x|X_{N_s} = i, \tau_{N_s} = s\right).$$

A semi-Markov process is entirely determined by its semi-Markov kernel, which makes the latter a key element in our modeling process.

Definition. *The jump process associated with a semi-Markov process is a Markov chain defined by:*

- *The same set of states as the original semi-Markov process.*
- *Probabilities of transition given by:*

$$\begin{aligned}\forall i, j \in E,\ \forall\ s \geq 0,\ p_{i,j}(s) &= \lim_{x \to +\infty} Q_{i,j}(s,x) \\ &= P\left(X_{N_s+1} = j | X_{N_s} = i, \tau_{N_s} = s\right).\end{aligned}$$

Definition. *The duration law of a semi-Markov process, given its semi-Markov kernel, is defined by:*

$$\begin{aligned}\forall i, j \in E, \forall\ s \geq 0,\ F_{i,j}(s,x) &= P\left(\tau_{N_s+1} \leq s + x | X_{N_s} = i, X_{N_s+1} = j, \tau_{N_s} = s\right) \\ &= \begin{cases} \dfrac{Q_{i,j}(s,x)}{p_{i,j}(s)} & \textit{if } p_{i,j}(s) > 0, \\ 0 & \textit{otherwise.} \end{cases}\end{aligned}$$

We have the fundamental relation: $Q_{i,j}(s,x) = p_{i,j}(s) \times F_{i,j}(s,x)$. It gives us a way to break down the modeling process into two subsequent simpler processes. We first pick a satisfying model for the jump probabilities $p_{i,j}(s)$. For each couple (i,j) with $i > j$ we then fit a model to the duration law for the transition from state i to state j.

The strength of this approach lies in the fact that both the jump probabilities and the duration laws can easily be interpreted. Furthermore this immediately translates into an algorithm to simulate trajectories. Firstly, we determine the next state using jump probabilities, then knowing this next state we determine the time until transition using the associated transition law. We continue to iterate these two steps until the current state becomes death.

2.2 Jump probabilities

Our first step is to find a parametric expression for the jump probabilities, which only depends on the age s. The right approach would be to first calibrate a non-parametric model for jump probabilities. Nevertheless, as our data is heavily censored, the results obtained this way cannot be trusted. Hence, we have decided to choose a linear model. This decision is based on Lepez, 2006 whose author fitted both a logistic and a linear model to the APA data and obtained results too close to justify using everything but the linear model.

$$\forall i \neq j \in E,\ \forall\ 0 \leq s,\ p_{i,j}(s) \ = \ a_{i,j} \ \times \ s \ + \ b_{i,j}.$$

As $p_{i,j}(s)$ are the terms of a stochastic matrix, we have the following constraints:

$$\begin{aligned}&\forall i \in E,\ \forall\ 0 \leq s, && \sum_{j \neq i} p_{i,j}(s) = 1 \iff \begin{cases} \displaystyle\sum_{j \neq i} a_{i,j} = 0. \\ \displaystyle\sum_{j \neq i} b_{i,j} = 1. \end{cases} \\ &\forall i \neq j \in E, && 0 \leq p_{i,j}(s) \leq 1.\end{aligned}$$

We consider 10 transitions so we should have 10 parameters for $a_{i,j}$ and 10 for $b_{i,j}$. However, as $(p_{i,j})_{0 \leq i,j \leq 4}$ is a stochastic matrix, the equality constraints associated force the value of one parameter in each line of the matrix. Thus we will only consider $10 - 4 = 6$ parameters of each type and express the probabilities in the last column of the matrix using those parameters.

2.3 Duration laws

We will use Weibull law as a base component for our model. Weibull law is a generalization of the exponential law. It uses the same scale parameter $\sigma > 0$ but also a shape parameter $\nu \geq 1$. Weibull laws play a key role in reliability theory (see Murthy et al., 2004; Zhang et al., 2006) and are often used to model human survival laws. Besides, dependency is a very heterogeneous state, as it can be caused by a wide range of illnesses. On one hand cancers and cardiovascular diseases are usually associated with short survival time. On the other hand, muscular and skeletal diseases as well as dementia, Alzeihmer and Parkinson diseases generally lead to longer survival time. To capture this effect, and still in accordance with the reliability theory (see Bucar et al., 2004; Jiang and Murthy, 1997), we decide to use a mixture of two Weibull laws for each transition, with a constant mixture parameter $\lambda \in \,]0\,;1[$. Lastly, we want the age s of entry to play a role in duration laws, as people offer less resistance to diseases the older they become, which translates into shorter survival times. To take this effect into account, we propose to use $xe^{\beta s}$ instead of x as the argument of the Weibull cumulative distribution function, where $\beta \in \mathbb{R}$ will be called the time dependency parameter.

Finally our duration law is defined by the following cumulative distribution function:

$$F(s,x) = (1-\lambda)W_1(s,x) + \lambda W_2(s,x)$$

where, for $i \in \{1;2\}$, $$W_i(s,x) = 1 - e^{-(\sigma_i x e^{\beta_i s})^{\nu_i}}.$$

Parameter summary

Finally, our model uses 82 parameters. A summary is provided in table 1.

parameter	count	definition interval	description
a	6	$a \in \mathbb{R}$	slope of jump process
b	6	$b \in \mathbb{R}$	intercept of jump process
ν	20	$\nu \geq 1$	shape parameter of Weibull law
σ	20	$\sigma > 0$	scale parameter of Weibull law
β	20	$\beta \in \mathbb{R}$	time dependency parameter
λ	10	$0 < \lambda < 1$	mixture parameter

Table 1: Summary of the different parameters used in the model.

3 APA data

3.1 Introducing the data

The data at our disposal has been gathered in four French administrative areas over the years 2002 to 2005 included. Every individual living in one of those areas that has been granted this aid at least once between 2002 and 2006 should therefore be included in this database.

Figure 4 shows an extract of the processed data base, after we harmonized variable format over the different administrative area's databases and removed the individuals with obvious errors in individual dates or chronology. The data contains the following information about individuals:

- The date of birth. As we only have the month and year of birth, the day of birth has been arbitrarily set to 15 for each individual.
- The date of up to 5 evaluations of the individual's dependency state using the AGGIR grid and the result of those evaluations.

- The date of death when death occurs during the observation period. As we can see in figure 5b, death is only observed starting from 2005.
- Other variables such as administrative area of residence, sex and matrimonial status that we will not be using here.

	date of birth	date eval1	date eval2	date eval3	date eval4	date evalf	date of death	gir eval1	gir eval2	gir eval3	gir eval4	gir evalf	area	sex	matri
1	15/06/1914	28/02/2005	NA	NA	NA	28/02/2005	NA	2	NA	NA	NA	2	4	2	2
2	15/04/1911	15/10/2003	NA	NA	NA	15/10/2003	26/12/2005	3	NA	NA	NA	3	4	2	2
3	15/03/1930	11/08/2004	NA	NA	NA	11/08/2004	NA	4	NA	NA	NA	4	4	1	3
4	15/04/1928	16/06/2004	NA	NA	NA	16/06/2004	30/01/2006	4	NA	NA	NA	4	4	2	2
5	15/04/1921	01/01/2005	NA	NA	NA	01/01/2005	NA	4	NA	NA	NA	4	4	1	2
6	15/11/1935	17/01/2003	03/02/2004	31/03/2005	NA	31/03/2005	NA	4	4	4	NA	4	4	2	2
7	15/03/1913	06/08/2004	23/01/2006	NA	NA	23/01/2006	NA	4	4	NA	NA	4	4	2	2
8	15/01/1937	09/02/2005	NA	NA	NA	09/02/2005	NA	3	NA	NA	NA	3	4	2	3
9	15/05/1917	02/10/2003	29/09/2004	NA	NA	29/09/2004	NA	4	4	NA	NA	4	4	1	1
10	15/12/1920	07/10/2005	NA	NA	NA	07/10/2005	06/03/2006	2	NA	NA	NA	2	4	1	1

Figure 4: Extract of the processed APA data.

We note that the first four dates of evaluation correspond to the first four dates of evaluations of the beneficiary in the period when the last date of evaluation is the last known date of evaluation. It means that if the beneficiary has undergone less than four evaluations, we will have his/her full trajectory, while if the beneficiary has passed 5 or more, we will only have the first four dates and the last date, but we will not even know if there is any evaluation in-between. Such trajectories are fortunately not very frequent (they represent around 1 % of the database) and can be safely discarded, without significantly altering the data. Once processed, the database contains information about 53,599 individuals.

3.2 Observations on the data

First of all, we note that in the place of dates of transitions, we only have dates of evaluation. However, as those evaluations are freely available and at any time, we will make the assumption that each change in the dependency status of an individual is immediately followed by an evaluation. Hence, we can consider that every transition that happened during the observation period is in our database.

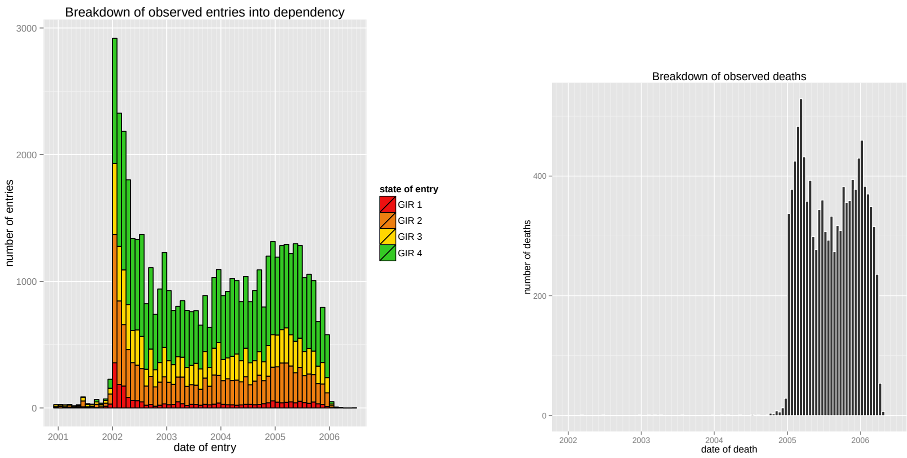
Figure 5a represents the distribution of observed first date of evaluation. We can observe a peak during the first year of observation. This is because 2002 is the year the APA was introduced. Thus we have a bias because of people who have been dependent for some time but whose dependency statuses were not assessed prior to the introduction of the APA. This phenomenon is called the "stock effect". To solve this problem we need to remove individuals whose first evaluation have taken place before 2003 from the base. They represent around one third of the database and we have 34,551 individuals left after this step.

As we follow individuals over a few years, many trajectories are incomplete at the end of the observation period. Those trajectories are right censored, and we need to account for this when we perform the calibration of parameters.

Figure 5b represents the distribution of observed death. We can see that there are hardly any observations during the first 3 years because, before 2005, no information was gathered on death of beneficiaries. Therefore, we can distinguish two situations for individuals whose death was not observed:

- The last evaluation happened after the 1^{st} of January 2005. In this case, if the individual was dead, we would have known it. Therefore there is no missing information.
- There has not been any evaluation since the 1^{st} of January 2005. In this case there may be some missing information about the death of the individual.

When death of the individual occurred before 2005, we do not know precisely when it happened between the last observed evaluation and the 1st of January 2005. This phenomenon is referred to as interval censoring. In our case however, things are even worse as we do not even know if death occurred at all. The next section will provide a way to take into account this peculiarity of the data.



(a) Distribution of observed entries (b) Distribution of observed deaths

Figure 5: Observed events on the data.

4 Calibration

To calibrate the parameters, we use a maximum likelihood method.

4.1 Likelihood function

The first step is to express the likelihood function. The global likelihood is the product of all individual likelihoods:

$$L = \prod_{p=0}^{N} L_p$$

where N is the number of individual in the observation base, $N = 34{,}551$ for our data.
The trajectory of each individual p is defined using:

- The number of observed transitions: n_p.
- The set of visited states: $X^p = (X_k^p)_{1 \le k \le n_p}$.
- The set of transition times: $t^p = (t_k^p)_{1 \le k \le n_p}$ and the times $\tau_k^p = t_k^p - t_{k-1}^p$ between transitions for $1 \le k \le N_p$ where t_0^p is the birth date of the individual.
- A couple (δ_1^p, δ_2^p) indicating if the trajectory is right censored or interval censored. If T_1^p is the age of the individual at the beginning of the death observation period (1st of January 2005) and T_2^p the end of the observation period (31st of December 2005), we can write:

  $$(\delta_1^p, \delta_2^p) = \left(\mathrm{I}[X_{n_p} \neq 0, T_1^p \le t_{n_p} < T_2^p], \mathrm{I}[X_{n_p} \neq 0, t_{n_p} < T_1^p]\right)$$

  where I(A) represents the indicator function associated with the subset A.

The global likelihood has the following expression:

$$L = \prod_{p=1}^{N} \left[\underbrace{\left(\prod_{k=1}^{n_p-1} c_{X_k^p, X_{k+1}^p}(t_k^p, \tau_{k+1}^p) \right)}_{\text{observed transitions}} \times \underbrace{c^1_{X_{n_p}^p}(t_{n_p}^p, T_2^p)^{\delta_1^p}}_{\text{right censoring}} \times \underbrace{c^2_{X_{n_p}^p}(t_{n_p}^p, T_1^p, T_2^p)^{\delta_2^p}}_{\text{interval censoring}} \right]$$

where $c_{X_k^p, X_{k+1}^p}(t_k^p, \tau_{k+1}^p)$ is the term associated with an observed transition, $c^1_{X_{n_p}^p}(t_{n_p}^p, T_2^p)$ and $c^2_{X_{n_p}^p}(t_{n_p}^p, T_1^p, T_2^p)$ are terms which take into account right censoring and interval censoring respectively. In practice, we use the log-likelihood to convert products into sums. The log-likelihood function can thus be expressed as the sum of 40,315 terms which takes into account the information we have on the remaining 34,551 individuals.

Observed transitions

The contribution associated with an observed transition has the following expression:

$$c_{X_k^p, X_{k+1}^p}(t_k^p, \tau_{k+1}^p) = p_{X_k^p, X_{k+1}^p}(t_k^p) \times f_{X_k^p, X_{k+1}^p}(t_k^p, \tau_{k+1}^p).$$

Right censoring

The trajectory of an individual is right censored when his/her last evaluation has taken place in 2005 and when death has not been observed. Under this hypothesis, the individual is alive at the end of 2005. Therefore, we need to take this information into account by adding another term into the likelihood function:

$$\begin{aligned} c^1_{X_{n_p}^p}(t_{n_p}^p, T_2^p) &= S_{X_{n_p}^p}(t_{n_p}^p, T_2^p - t_{n_p}^p) \\ &= \sum_{j < X_{n_p}^p} p_{X_k^p, j}(t_k^p) \times S_{X_{n_p}^p, j}(t_{n_p}^p, T_2^p - t_{n_p}^p) \\ &= 1 - \sum_{j < X_{n_p}^p} p_{X_k^p, j}(t_k^p) \times F_{X_k^p, j}(t_k^p, T_2^p - t_{n_p}^p). \end{aligned}$$

where, for $s,\ x \geq 0$ and $i \neq j$:

- $S_{i,j}(s,x) = 1 - F_{i,j}(s,x)$ is the survival function associated with the transition between i and j.
- $S_i(s,x) = \sum_{j<i} p_{i,j}(s) \times S_{i,j}(s,x)$ is the marginal survival function in state i.

Interval censoring

The trajectory of an individual is interval censored when his/her last evaluation has taken place before 2005 and when death has not been observed. Such individuals belong to one of the two following categories:

- Those who died between their last evaluation and the 1st of January 2005. In this case we also know that their status has not changed before death or it would have been observed.
- Those who are still alive at the 1st of January 2005. In this case we know that they did not die either during the year 2005 or it would have been observed, and that their status did not change at all during the remaining of the period.

To take this information into account we need to add another factor to the likelihood function, that precisely reflects the fact that the observed individual belongs to one of the two aforementioned groups:

$$\begin{aligned} c^2_{X^p_{n_p}}(t^p_{n_p}, T^p_1, T^p_2) &= p_{X^p_{n_p},0}(t^p_{n_p}) \times F_{X^p_{n_p},0}(t^p_{n_p}, T^p_1 - t^p_{n_p}) + S_{X^p_{n_p}}(t^p_{n_p}, T^p_2 - t^p_{n_p}) \\ &= p_{X^p_{n_p},0}(t^p_{n_p}) \times F_{X^p_{n_p},0}(t^p_{n_p}, T^p_1 - t^p_{n_p}) + \sum_{j < X^p_{n_p}} p_{X^p_k,j}(t^p_k) \times F_{X^p_k,j}(t^p_k, T^p_2 - t^p_{n_p}). \end{aligned}$$

where, for $s,\ x \geq 0$ and $i \neq j$: , $S_i(s,x)$ is the marginal survival function in state i as previously introduced.

4.2 Optimization

The optimization of a 40,315 term function with respect to 82 parameters should not be taken lightly.

Algorithm used

To solve this optimization problem, we need an algorithm that does not require an expression for the gradient of our likelihood function. Under those circumstances, the Nelder-Mead algorithm (see Nelder and Mead, 1965; Mathews and Fink, 2004) seems perfectly suited. This algorithm is conveniently implemented in the function *constrOptim()*, which allows for linearly constrained optimization in `R` . The Nelder-Mead algorithm is a heuristic which builds a simplex in the solution space and calculates the value of the optimization function at the vertexes. Then, depending on the results, the algorithm applies geometric transformations to the simplex in order to explore the most promising parts of the solution space. After many iterations, this heuristic should converge toward the global maximum of the function.

Log-likelihood subdivision

We observe that in the log-likelihood function, each term contains only parameters associated with the same departing state. Hence we can gather these terms according to their departing state, and write the log-likelihood function as a sum of 4 sub-functions, each one of them containing different parameters. It means that for solving the optimization problem, we can just solve those 4 sub-problems separately. Using this method, we are able to cut the computation time required by a factor of 4.

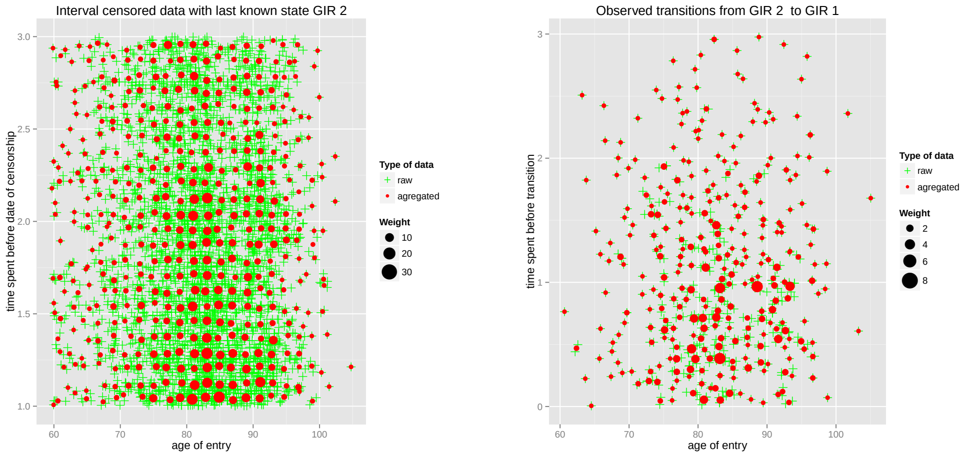
Data aggregation

Another way to reduce the algorithmic complexity of our problem is to aggregate data. The idea is that many observations are very similar, and instead of calculating the likelihood associated with all of them, we can aggregate them and calculate a likelihood function with weighted factors.

We decide that the observations meeting the following criteria should be put together:

- Same departing state and same arriving state (or type of censure for censoring terms).
- Age s of entry in the departing state in the same range of 2 years.
- Time x spent before transition (or before censoring) in the same range of 1 month.

For each group, we replace the observations by a single observation, whose age of entry s and time spent before transition x are the mean of the replaced observations', with a weight equal to the number of replaced observations it represents. Figure 6 illustrates the aggregation the effect of the aggregation for two different data sets. Using this method, the number of terms in the



(a) Effect of aggregation on a very dense set of data

(b) Effect of aggregation on a sparse set of data

Figure 6: Illustration of data aggregation.

log-likelihood function drops from 40,135 to 6,492, further cutting required computation time by a factor of 6.

Thanks to both methods, we are able to reduce computation time to the point where calibration of the model becomes possible on an office computer and in a reasonable amount of time.

4.3 Model selection

Now that we are able to calibrate models, we have to determine, using the APA data, which parameters are necessary and which ones are not.

Selection criterion

To compare two different models, one of them being a sub-model of the other, we will use the Bayesian Information Criterion (BIC) (see Lebarbier and Mary-Huard, 2006, for a review of the BIC.). This criterion gives a score to each model i, using the following formula:

$$BIC_i = -2\ l_i(\hat{\theta}_i) + K_i\ ln(N)$$

where:

- K_i is the number of parameters of the model.
- l_i is the log-likelihood function.
- $\hat{\theta}_i$ is the set of parameters that maximizes the log-likelihood function.
- N is the number of observations. It does not depend on the model used.

Then, according to this criterion, the model with the lower associated BIC value should be preferred over the others. When we take a closer look at the formula, we see that this criterion proposes a valuation of the trade-off between the accuracy of the fitting (represented by the log-likelihood function) and the explanatory power of each parameter (which is lower the more parameters there are in the model). These two factors have the associated weights - 2 and $ln(N)$ respectively. For our database, we have $N = 34{,}551$ and thus $ln(N) \simeq 10$.

Comparison of base models

In a first part we are going to compare models using survival laws of increasing complexity, as described previously. Results of calibration are summarized in table 2. Weibull laws with a time dependency parameter will be referred to as improved Weibull laws. According to the BIC, the last model, which uses a mixture of improved Weibull laws and includes 82 parameters, is the best model. In the next subsection, we will try to improve this model by removing non-essential parameters.

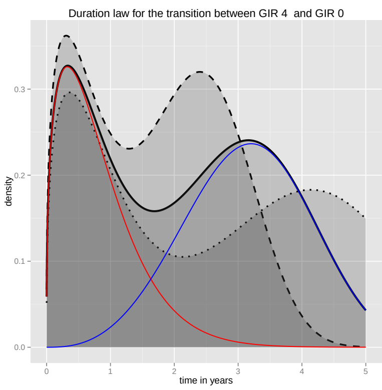
Duration laws model	Parameters	Log-likelihood	BIC	Calculation time
Simple Weibull laws	32	- 36,347	73,028	36 min
Mixture of simple Weibull laws	62	- 36,120	72 888	4.6 h
Improved Weibull laws	42	- 36,173	72,785	1.3 h
Mixture of improved Weibull laws	82	- 35,938	72,733	6.4 h

Table 2: Comparison of base models. Calculation times applies to an office computer with a 2.93 GHz dual core processor.

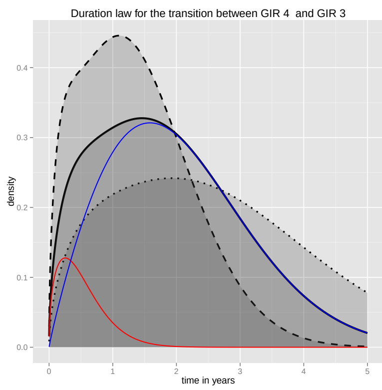
4.4 Suppression of non-significant parameters

Because of the high calculation time needed for calibration, we will not be able to use standard step-wise methods where at every step, we test every sub-model obtained by removing one of the parameters, and pick the model which gets the best BIC. Instead, we will have to remove several parameters at once, based on graphical observations or empirical criteria, and compare the BIC of the resulting model to the one of the original model. With this method, we improve our model at each step but with no guarantee of having the best model at the end, but this may also be said of step-wise methods. We will consider 3 successive waves of parameters removal.

Significance of mixture parameters



(a) Transition between GIR 4 and death



(b) Transition between GIR 4 and GIR 3

Figure 7: Mixture parameters. Bold line: duration law for an age of entry of 80, with its two components in red and blue. Dotted (resp. dashed) line: duration laws for an age of entry of 60 (resp. 100).

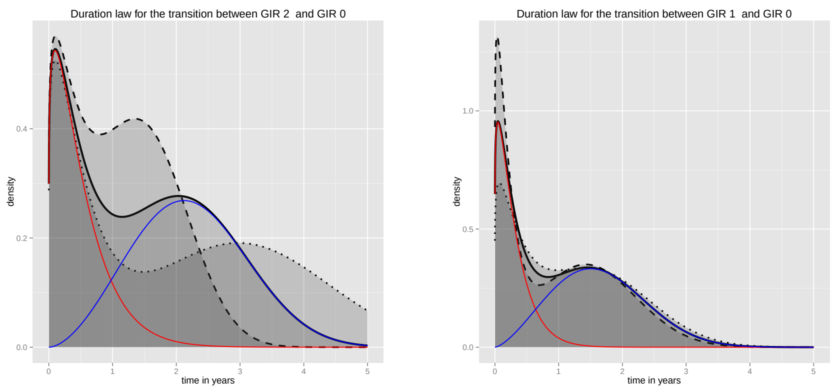
Firstly, we can observe that the use of mixtures is not always justified. Figure 7 represents 2 duration laws after calibration, for the model with mixtures of improved Weibull laws. In figure 7a, the duration law is clearly bi-modal, which justifies the use of mixtures, while in figure 7b the second mode can hardly be seen. The results of the calibration show that for transitions with death as the arrival state, a mixture is always needed. Transitions from GIR 4 to a heavy dependency state (GIR 2 or GIR 1) also require a mixture. For other transitions, an improved Weibull law is enough. In this first step, we will thus test the effect of the removal of 4 mixtures (for transitions GIR 4 to GIR 3, GIR 3 to GIR 2, GIR 2 to GIR 1 and GIR 2 to GIR 1), for a total of 16 parameters removed (4 mixture parameters and 3×4 parameters which were used in the second component of the mixture).

Significance of time dependency parameters

We also observe that improved Weibull laws are not always needed. To determine in which cases we can suppress the time dependency parameters, we will use the following empirical criterion:

$\beta_{i,j}$ is considered to be significant when $|e^{100\beta_{i,j}} - e^{60\beta_{i,j}}| < 0.25$.

This means that the transition takes on average at least 25 % less time (or 25 % more time) for people aged 100 than for people aged 60. If a parameter does not meet this criterion, we can regard the impact of the time dependency factor as not strong enough to appear in the model. Two parameters, $\beta^1_{3,0}$ and $\beta^2_{1,0}$, do not meet the criterion and are removed from the model. Figure 8 shows two duration laws: 8b is an example of significant time dependency parameters, while 8a has a very low time dependency on its second component.



(a) Transition between GIR 2 and death (b) Transition between GIR 1 and death

Figure 8: Time dependency parameters. Same legend as for figure 7.

Significance of slope parameter

In this last step, we will focus on the slope parameters of the models. We will also use an empirical criterion:

$a_{i,j}$ is considered to be significant when $|(100 - 60)a_{i,j}| < 0.01$.

It means that more than 1 % of the distribution of possible next states changes between 60 and 100 years. If a parameter does not meet this criterion, it does not have a significant impact on the probability distribution and can be suppressed. Two parameters $a_{4,1}$ and $a_{3,2}$ do not meet the criterion and are removed from the model. Figure 9 shows the jump probabilities from GIR

3. The slope parameter for the transition from GIR 3 to GIR 2 is not significant, while the two other parameters are significant.

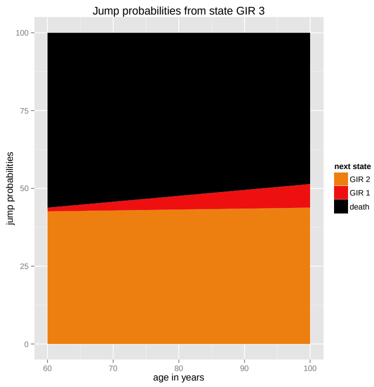


Figure 9: Slope parameters.

Results of previous steps

Table 3 gives the impact of the previously described steps on the model. We can see that each successive wave of removal contributes towards improving our model. From now on, we will only consider the model obtained after applying the three waves of removals, which contains 62 parameters. The expression of jump probabilities and duration laws used in this model and the value of these parameters obtained after calibration are given in Appendix B.

Duration laws model	Parameters	Log-likelihood	BIC
Complete model	82	- 35,938	72,733
including removal of mixtures	66	- 35,940	72,570
also including removal of time dependency parameters	64	- 35,941	72,551
also including removal of slope parameters	62	- 35,942	72,532

Table 3: Removal of non-significant parameters results.

5 Complete model

In the previous sections, we introduced and calibrated a model for the trajectory of individuals once they become dependent. In this section we will bring together the last elements required to generate complete trajectories.

5.1 Additional elements

To get a complete model as described in figure 2, we are still missing a few items:

- $i(s)$: the probability to become dependent.
- $q^a(s)$: the probability for autonomous people to die.
- $R(s)$: the distribution of probabilities to enter the different states of dependency, knowing the age of entry.

For the first two items, we are going to use experience laws graciously provided by SCOR. We will also introduce the following quantities:

- $p_i(s,x)$ the probability at age s to become dependent at age $s+x$.
- $p_q(s,x)$ the probability at age s to die at age $s+x$ without becoming dependent.
- $p_i(s)$ the probability at age s to become dependent before dying.
- $p_q(s)$ the probability at age s to die without becoming dependent.
- $f_i(s,x)$ the probability at age s knowing one will become dependent before dying that the entry in dependency occurs at age $s+x$.
- $f_q(s,x)$ the probability at age s knowing one will die without becoming dependent that the death occurs at age $s+x$.

These quantities can be expressed using i and q^a:

$$p_i(s,x) = \left(\prod_{k=s}^{s+x-1}(1-i(k))(1-q^a(k))\right) i(s+x), \quad p_i(s) = \sum_{x=0}^{\infty} p_i(s,x), \quad f_i(s,x) = \frac{p_i(s,x)}{p_i(s)}$$
$$p_q(s,x) = \left(\prod_{k=s}^{s+x-1}(1-i(k))(1-q^a(k))\right) q^a(s+x), \quad p_q(s) = \sum_{x=0}^{\infty} p_q(s,x), \quad f_q(s,x) = \frac{p_q(s,x)}{p_q(s)}.$$

While these quantities may seem redundant with the incidence and mortality rates, they present a real interest from an algorithmic point of view, as we will see in the following subsection.

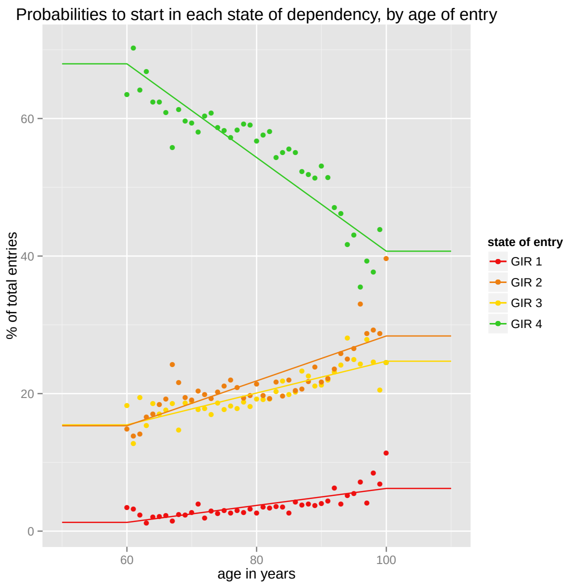


Figure 10: Distribution of entry states with regard to age of entry.

For the distribution $R(s)$, we will once again use the APA data. We have $N = 34{,}551$ observations of entry into dependency. For each age between 60 and 100 years, we count the number of individuals that became dependent and entered each state at that age. Then, we fit a parametric model to this distribution. Given the available data, a linear model seems quite appropriate. As the sum of fitted probabilities will not be exactly equal to 1, we have to normalize it so the laws are not exact linear functions, but are still extremely close. As little data is available for ages under 60 or over 100, we assume constant probabilities outside the interval. Figure 10 shows the empirical observation as well as the fitted distribution.

5.2 Simulation of trajectories

We denote by s_0 the age of subscription. This is the age at which the trajectories begin.

The trajectory of the individual p that we are trying to simulate is defined by:

- The number of visited states: n_p.
- The set of visited states: $X^p = (X_k^p)_{0\leq k\leq n_p}$.
- The set of transition times: $t^p = (t_k^p)_{0\leq k\leq n_p}$.

The possible states are $\{0, 1, 2, 3, 4, 5\}$ where 5 represents the autonomy state, 0 is death and states 1 to 4 correspond to states GIR 1 to GIR 4.

To simulate the trajectory of an individual p, we use the following algorithm:

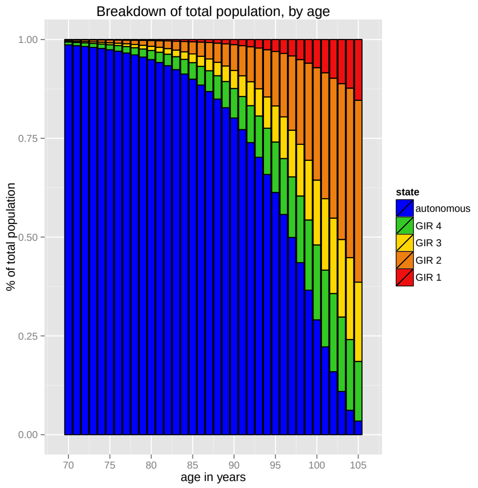
1. We set $X_1^p = 5$, and $t_1^p = s_0$. The underlying hypothesis is that all individuals are autonomous at subscription, which is the case for individual contracts, where medical underwriting is always performed.
2. We draw a variable r_1 uniformly distributed on $[0\,;1]$ and we test it against $p_i(s_0)$. If $r_1 > p_i(s_0)$, the individual will die without becoming dependent. We set $X_2^p = 0$ and go to step 3. Otherwise the individual will become dependent and we can go to step 4.
3. We draw a random discrete variable x distributed according to the probabilities $f_q(s_0, x)$, which allows us to determine the age of death $s + x$, to which we need to add a random variable r_2 uniformly distributed on $[0\,;1]$ as the fractional part of the year. We thus set $t_2^p = s_0 + x + r_2$. The trajectory ends with the death of the individual and the algorithm stops, with $n_p = 2$.
4. We draw a random discrete variable distributed according to the probabilities $f_i(s_0, x)$ to determine the age of entry in dependency $s+x$, to which we need to add a random variable r_2 uniformly distributed on $[0\,;1]$. We also set $t_2^p = s_0 + x + r_2$.
5. We draw a random discrete variable distributed according to $R(s_0 + x)$ to determine the state of entry in dependency X_2^p.
6. We set $k = 2$.
   While the individual is alive, i. e. $X_k^p \neq 0$ we do the following:
   - We draw a random discrete variable distributed according to $p_{X_k^p,j}(t_2^p)$ for all $j \in \{0, \ldots, X_k^p - 1\}$ to determine what the next state X_{k+1}^p will be.
     Then:
     - If $F_{X_k^p,X_{k+1}^p}$ is a simple duration law (i. e. $\lambda_{X_k^p,X_{k+1}^p} = 0$), the time X_k^p spent in the current state before transition follow is distributed according to the law:

       $$Y = \frac{e^{-\beta s}}{\sigma}\left[\ln\left(\frac{1}{1-U}\right)\right]^{\frac{1}{\nu}}$$

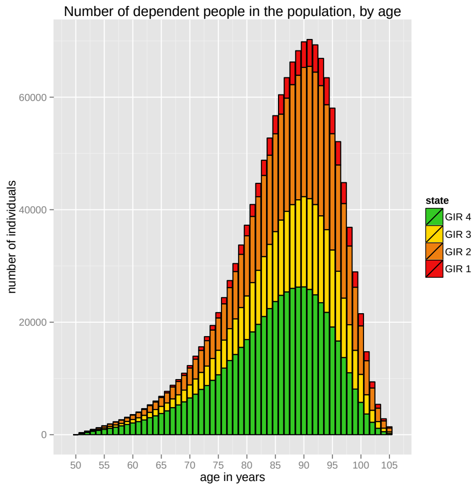
       where U is a random variable uniformly distributed on $[0\,;1]$. For clarity's sake, we omitted the index in this formula. They would be X_k^p and X_{k+1}^p. Hence we simulate u uniformly distributed on $[0\,;1]$ and we get y the time spent in state X_k^p, which gives us $t_{k+1}^p = t_k^p + y$.
     - If $F_{X_k^p,X_{k+1}^p}$ is a mixture, we draw a uniformly distributed variable r_3 on $[0\,;1]$ and we test it against $\lambda_{X_k^p,X_{k+1}^p}$: if $r_3 < \lambda_{X_k^p,X_{k+1}^p}$, we pick the first component of the mixture, otherwise the second. Then with this single component, we can proceed as previously described, and get t_{k+1}^p.
   - We increment k and go back to the beginning of step 6.
7. The trajectory ends with the death of the individual and the algorithm stops with $n_p = k$.

With discrete probabilities laws $f_i(s_0, x)$ and $f_q(s_0, x)$, we only need to draw 2 random variables to determine if the individual becomes dependent during his/her life or not and when he/she becomes dependent or dies. With incidence and mortality rates, for an age of subscription of 50 years, it takes on average more than 30 years before death or dependency occurs. We therefore need to perform 60 random variable draws just for this first step of the algorithm. This process allows us to cut the overall computation time by a factor of 10, and 10,000 trajectories can then be simulated by our algorithm in less than 3 seconds on an office computer.

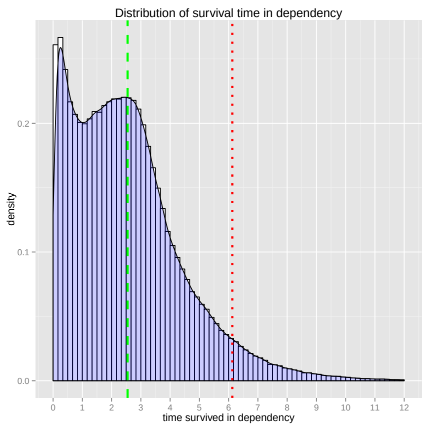
5.3 Statistical information on simulated trajectories



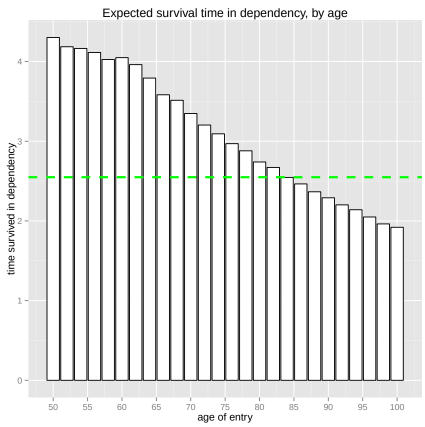
(a) Prevalence of dependency states in the general population, by age.



(b) Number of dependent people by age, for an initial population of one million 50 year old individuals.



(c) Distribution of survival time in dependency. Dashed green line: mean. Dotted red line: 95 % percentile.



(d) Average survival time in dependency by age of entry. Dashed green line: mean.

Figure 11: Statistical information on simulated trajectories.

Once we are able to generate trajectories, we can try to gather some statistical information about these trajectories:

- Life expectancy at 50 years is 34.8 years. It is slightly more than life expectancy of the general population, which is due to the fact that we used mortality rates from insurer portfolios experience, that are lower than those of the the general population.
- 57.1 % of people become dependent, and survive for an average time spent in this state of 2.55 years.
- The average age of entry in dependency is 85.9 years, which is close to what is observed on the APA data.

All these indicators are calculated based on one million simulations, with an age of subscription of 50 years. Figure 11a represents the structure of the population every year. Dependent people account for 5 % of the population only at 80 years, 20 % at 90 years and more than 70 % at 100 years. Figure 11b represents the number of dependent people every year for a base population of one million 50 year old subscribers. The maximum number of dependents is reached at age 91. Figure 11c represents the distribution of survival time in dependency. The green dashed line corresponds to the mean of the distribution at 2.55 years. The red dashed line corresponds to the 95 % percentile at 6.12 years. Finally, figure 11d shows the average survival time in dependency, given the age of entry. This survival time decreases almost linearly, from 4.2 years at age 50 to less than 1.9 years at age 100.

6 Pricing and results

In this part we are going to introduce a fictive long-term care insurance product and use our model to price it.

6.1 Product description

For this premium, the dependency is evaluated using the AGGIR grid. Only states GIR 1 to GIR 4 will be considered as dependency. A level premium will be paid by the insured life, at the beginning of every month, while he is alive and autonomous. If the insured life is alive and dependent at the end of the month, the premium no longer needs to be paid, and a monthly allowance is granted instead. The amount of this allowance is determined by the state of dependency:

- GIR 1: 1,300 €.
- GIR 2: 1,100 €.
- GIR 3: 800 €.
- GIR 4: 0 €.

An additional cash amount of 1 650 € is also granted alongside the first allowance, regardless of the dependency state. The contract also includes a deferral period of 3 months. No payment is made during the first three months spent in dependency. The additional cash amount is thus paid at the end of the fourth month in dependency, should the insured life still be alive at this time. An elimination period of 2 years is also added to the contract. It means that, should the insured life become dependent during the first two years after subscribing, the contract would be canceled. Finally, counter-insurance is also provided so that if the contract ends during the first two years because the insured life becomes dependent or dies, all premium paid are refunded to the insured life or his/her heirs.

Finally, a technical interest rate of 1.25 % will be set by the contract.

6.2 Pricing methodology

Premium determination

In most insurance models, the pricing is done directly using a closed formula, where the premium can be expressed according to different laws using simple or double sums. In the case of multi-state continuous models however, we would have to calculate multiple integrals, with absolutely no guarantee of finding an analytic function at the end. Therefore we will have to proceed otherwise.

The required premium is the amount that makes the expected payments for the insurer and the insured life equal. Both expected values are assessed using the Net Present Value (NPV). The required premium is the value p^* such that:

$$p^* \times E\left(\mathrm{NPV}(P)\right) = E\left(\mathrm{NPV}(B)\right),$$

where B represents the benefit cash flows and P represents the premium unit cash flows.

To find the value of the premium p^*, we use a Monte Carlo algorithm:

- We generate a large number n of trajectories with an age of subscription of s_0.
- For each trajectory, we compute the NPV of the benefit cash flows B and the NPV of unit premium cash flows P.
- We use the following formula to determine an estimate of the premium amount:

$$\widehat{p}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\mathrm{NPV}(B_i)}{\frac{1}{n}\sum_{i=1}^{n}\mathrm{NPV}(P_i)}$$

  where B_i (respectively P_i) represents the series of benefit cash flows (respectively unit premium cash flows) associated with the trajectory i. According to the law of large numbers, this is an unbiased estimator of p^*, i.e. $\widehat{p}_n \underset{n\to+\infty}{\to} p^*$ almost surely.

Uncertainty on premium estimation

The study of the law followed by the ratio of two normal variables has led to numerous results (Bastien, 1960; Marsaglia, 2006), with optimal bounds that depends on the correlation between those variables. However, in our case the empirical correlation coefficient between premiums and benefits is very low, only close to 5% and we resort to a much simpler approach.

Let us denote by $\widehat{\mu_B}(n)$, $\widehat{\mu_P}(n)$, $\widehat{\sigma_B}(n)$ and $\widehat{\sigma_P}(n)$ the estimators of empirical mean and empirical variance of $\mathrm{NPV}(B)$ and $\mathrm{NPV}(P)$ respectively, for $\alpha \in]0\,;1[$, according to the central limit theorem, we have:

$$\begin{aligned} |\mu_B - \widehat{\mu_B}(n)| &< \frac{\widehat{\sigma_B}(n)\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}} \\ |\mu_P - \widehat{\mu_P}(n)| &< \frac{\widehat{\sigma_P}(n)\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}} \end{aligned}$$

with a level of confidence of $1-\alpha$, where Φ is the cumulative distribution function of the standard normal law.

We want to upper-bound $|p^* - \widehat{p}_n| = |\frac{\mu_B}{\mu_P} - \frac{\widehat{\mu_B}(n)}{\widehat{\mu_P}(n)}|$.

Lemma. *Let $x \geq 0$ and $y > 0$. If $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$, $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ are four sequences such that:*

$$\begin{cases} x_n \underset{n\to+\infty}{\to} x \\ y_n \underset{n\to+\infty}{\to} y \end{cases} \qquad \begin{cases} a_n \underset{n\to+\infty}{\to} 0 \\ b_n \underset{n\to+\infty}{\to} 0 \end{cases} \qquad \begin{cases} |x_n - x| < a_n \\ \\ |y_n - y| < b_n \end{cases}$$

Then for all $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$:

$$|\frac{x}{y} - \frac{x_n}{y_n}| < \frac{1+\epsilon}{y_n}\left(a_n + \frac{x_n}{y_n} b_n\right)$$

We apply this lemma, whose demonstration is given in Appendix A, to:

$$x_n = \widehat{\sigma_B}(n) \qquad a_n = \frac{\widehat{\sigma_B}(n)\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}$$
$$y_n = \widehat{\sigma_P}(n) \qquad b_n = \frac{\widehat{\sigma_P}(n)\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}$$

With n large enough, we thus have the following formula for the uncertainty of the estimation:

$$|p^* - \widehat{p}_n| \leq (\widehat{\sigma_B}(n) + \widehat{p}_n \widehat{\sigma_P}(n)) \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\widehat{\mu_P}(n)\sqrt{n}}$$

with a level of confidence of $1-\alpha$.

For a portfolio of 50 year old insured lives, using the previous formula, we find that we need approximately $n = 10,000,000$ simulations to get an uncertainty inferior to 1 % of the premium, with a level of confidence of 95 %.

6.3 Results

Results of pricing

Age of subscription	40 years	50 years	60 years	70 years
Base monthly premium	15.83 €	21.96 €	32.79 €	54.63 €
Impact of deferral period	- 0.70 €	- 0.97 €	- 1.44 €	- 2.44 €
Impact of elimination period	- 0.02 €	- 0.05 €	- 0.25 €	- 1.43 €
Impact of counter-insurance	< 0.01 €	< 0.01 €	+ 0.02 €	+ 0.11 €
Final monthly premium	15.12 €	20.95 €	31.12 €	50.88 €

Table 4: Premium and impact of associated features on premium for several ages of subscription.

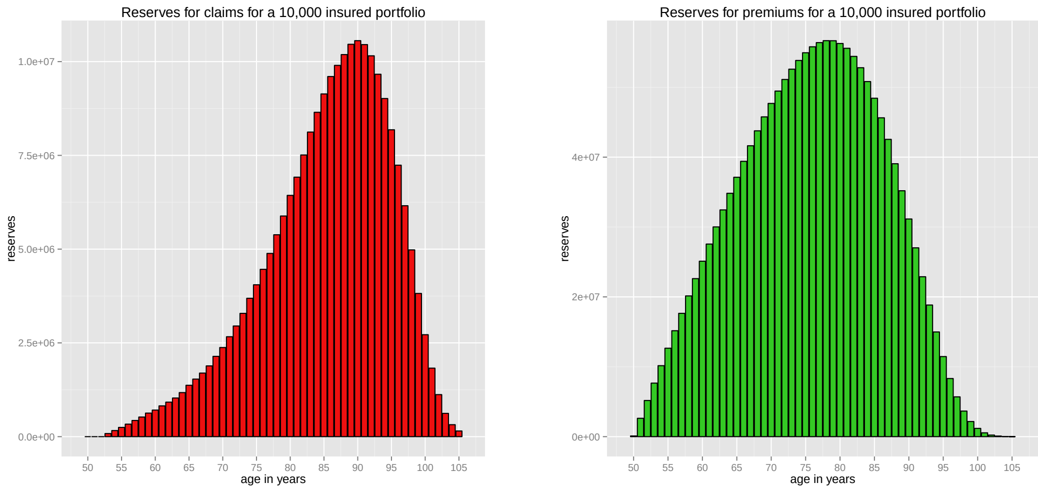
Reserves

In long-term care insurance, we are mostly interested in two kinds of technical reserves:

- **Reserves for premiums:** As the incidence rate increases with age, so does the long-term care risk. With level premiums, the insured lives pay more than the cost of their current risk at the beginning of the contract and less at the end. The difference thus needs to be added to reserves to cope with increasing future risk.

- **Reserves for claims:** These reserves aim at covering the cost of benefits for people who already are dependent. For each dependent insured life, annual reserves are created, where the amounts are equal to the expectancy of claims, given the situation of the individual. In our model, these reserves depend on age of entry s, current state of dependency i, and time x spent in this state of dependency. As for the premium, the amount of these reserves for a combination of the 3 variables can be evaluated by performing simulations. In practice this computation can be really difficult. We can only assess the amount of reserves for a finite number of variable combinations. Furthermore, we cannot generate trajectories where the insured life would already be dependent for some time at the start, so the combinations we can access are very limited.

Figure 12 displays the amount of reserves needed for a portfolio of 10,000 insured lives subscribing at the age of 50.



(a) Reserves for claims

(b) Reserves for premiums

Figure 12: Amount of technical reserves for a portfolio of 10,000 insured lives.

Sensitivity to different risk factors

Sampling risk associated with the insurer's portfolio

We first want to quantify the sampling risk associated with the insurer's portfolio. We consider a portfolio of 10,000 insured lives aged 50 at subscription. We run 20,000 simulations of the whole portfolio, for a total of 200 million individual trajectories simulated, in order to get the associated distribution of NPV. Figure 13 represents the fitted distribution. It looks very similar to a normal distribution, with a mean near 0, meaning that the estimated premium is pretty close to the true value. From a Solvency II point of view an insurer would need a capital of 3,032,000 € to cope with this specific risk on the portfolio.

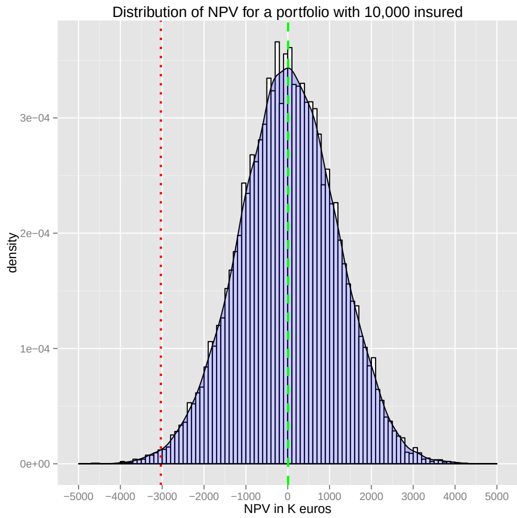


Figure 13: Distribution of NPV for a portfolio of 10,000 insured lives. Dashed green line: mean. Dotted red line: 0.5 % percentile.

Other factors of risk

The impact of other factors of risk: incidence rate, mortality rate for autonomous people, dependency duration laws and interest rates are given in table 5.

Shocks	Incidence rate	Autonomous mortality rate	Transition times in dependency	Interest rate
Value of shock	+ 10 %	- 10 %	+ 10 %	+ 100 bps
Resulting premium (20.95 € with no shock)	22.06 €	21.43 €	23.05 €	17.20 €
Relative variation	+ 5.3 %	+ 2.3 %	+ 10.0 %	- 17.9 %

Table 5: Sensitivity to different risk factors.

A shock on incidence rate naturally increases the long-term care risk, but in lesser proportions as it also reduces the size of the autonomous population and thus its exposure. A shock on mortality rate has a dual effect: an increase in the expectancy of premiums paid by autonomous insured lives but also an increase of the risk exposure of the population. This second effect actually outweighs the first, and the level of the premium needs to be increased to restore the balance. A shock on transition times in dependency has a straightforward effect on the level of premium required, as the amount of benefit is almost proportional to the time spent in dependency (except for the initial amount of cash and the payment deferral period). A shock on interest rate has a tremendous impact on the level of the required premium, as long-term care products have a very long duration. Hence, we can expect those products to become much more attractive in a situation of high interest rate.

7 Conclusion

In this paper, we have described the steps to construct a 4 state semi-Markov model for long-term care insurance, as well as the many problems encountered and the solutions we have chosen. This model can be used for the pricing of more complex products with several levels of guarantee, to match the level of dependency of the insured lives, and thus better meet their needs. It should also offer a more accurate quantification of the risk through reserves, as more information (state of dependency, time spent in dependency) about the dependent insured lives is taken into account. Besides, the use of a Monte Carlo method naturally allows us to get the distribution of the NPV of the portfolio. Hence we can compute the amount of capital needed from a Solvency II point of view. Finally, the various parameters or laws used in the model can be interpreted in regard to medical experience about the dependency process. Therefore, we could use the model to test the impact of various scénarii involving advances in the field of medicine, e.g. an increased prevalence of a particular disease among dependent people.

However, it should be noted that the model underestimates the survival times in dependency. Indeed, the life expectancy in dependency given by the model is only 2.55 where Debout 2010 computes an expectancy of 4.0 years based on APA data from 2002 to 2007 gathered over 300,000 beneficiaries. This spread can be linked to the limited length of the observation period. Observed trajectories are censored and therefore their length cannot exceed 3 years. Hence, the density of the duration law has to be extrapolated beyond the point where there is no data available. As a consequence, any error on observed trajectories is also reflected in the extrapolated part of the density. As the APA was only created in 2002, our data is strongly impacted by the stock effect we described before, even after deleting trajectories which started in 2002. Hence, observed transition times are underestimated and so is the whole density of duration laws. With longer series of data, whose observations would be further away from 2002, we should be able to eliminate this problem and get more realistic results.

The model we introduced could also be improved in several ways. The first thing that comes to mind is to calibrate a different model for each sex. This would be relevant as mortality rates are not the same for men and women. Furthermore, the incidence of diseases that cause dependency is not the same for men and women. Men are more vulnerable to cardiovascular diseases, while dementia is more frequent in women. Another way of improving the model would be to consider that the mixture parameters depend on the age (for example linearly). Indeed, the diseases responsible for entry in dependency change with age e.g. dependency following a cancer is very frequent at ages of 50 or 60, but become rare at ages of 80 or 90. However, we would have to be careful not to introduce too many parameters to model the effect of age, as they could eventually be substituted with one other, which would result in many local optima and make the calibration very difficult. In both cases, a better understanding of the diseases causing dependency, based on real data, is essential in order to be able to verify and interpret the results obtained after applying the improvements.

Acknowledgments

This paper is based on a master thesis (see Biessy, 2013) I wrote during a 6 month internship at SCOR Global Life, Paris, for my last year of actuarial studies at the EURIA. First of all, I would like to thank Vincent Lepez, deputy chief actuary at SCOR Global Life, for entrusting me with this position and for an exemplary overseeing. I would then like to acknowledge the help of Catherine Matias, CNRS research director at the "Statistique et Génome" laboratory in Evry, France, for several readings of the paper and a plethora of suggestions that led to improvements. Finally, my thanks go to all my colleagues at SCOR, especially to the people from the long-term care R&D team for their support and to Suhaila Binchy from the Dublin office for her gracious help.

References

Bastien, M. (1960). Loi du rapport de deux variables normales. *Revue de statistique appliquée 8*, 45–50.

Biessy, G. (2013). *Construction d'un modèle multi-états semi-markovien dans le contexte de l'assurance dépendance*. Mémoire d'actuariat, EURIA.

Blanpain, N. and O. Chardon (2010). Projections de population à l'horizon 2060. INSEE Première, 1320.

Bucar, T., M. Nagod, and M. Fajdiga (2004). Reliability approximation using finite Weibull mixture distributions. *Reliability Engineering and System Safety 84*, 241–251.

Cinlar, E. (1969). Markov renewal theory. *Advances in Applied Probability 1*, 123–187.

Debout, C. (2010). Durée de perception de l'Allocation personnalisée d'autonomie (APA). DREES, Document de travail.

Janssen, J. and R. Manca (2002). General actuarial models in a semi-Markov environment. In *Proceedings of ICA Cancun*.

Jiang, R. and D. Murthy (1997). Two sectionals models involving three Weibull distributions. *Quality and Reliability Engineering international 13*, 83–96.

Lebarbier, E. and T. Mary-Huard (2006). Une introduction au critère BIC : fondements théoriques et interprétation. *Journal de la SFdS 147*, 39–57.

Lepez, V. (2006). *Trajectoires en Dépendance des personnes âgées : modélisation, estimation et application en assurance vie*. Mémoire d'actuariat, Centre d'Etudes Actuarielles.

Marsaglia, G. (2006). Ratios of normal variables. *Journal of Statistical Software 16*, Issue 4.

Mathews, J. H. and K. D. Fink (2004). *Numerical Methods using Matlab, 4^{th} Edition*. Prentice-Hall Inc.

Mathieu, E. (2006). *Modélisations multi-états markoviennes et semi-markoviennes. Applications à l'état de sante des patients atteints par le virus du SIDA*. Thèse, Université de Montpellier.

Murthy, D., M. Xie, and R. Jiang (2004). *Weibull models*. Wiley series in Probability and Statistics. Wiley.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal 7*, 308–313.

Vetel, J., R. Leroux, and J. Ducoudray (1998). AGGIR, practical use, geriatric autonomy group resources needs. Soins Gerontol.

Zhang, L., M. Xie, and L. Tang (2006). Bias correction for the least squares estimator of Weibull shape parameter with complete and censored data. *Reliability Engineering and System Safety 91*, 930–939.

Appendix

Appendix A: Proof of the lemma

We first prove that:

If $x, x^{'} \geq 0$ and $y, y^{'} > 0$ and $a, b \in \mathbb{R}$ such that $|x - x^{'}| < a < x^{'}$ and $|y - y^{'}| < b < y^{'}$, then:

$$|\frac{x}{y} - \frac{x^{'}}{y^{'}}| < \frac{a}{y^{'}} + \frac{(x^{'} + a)b}{y^{'}(y^{'} - b)}$$

We have, on one hand:

$$\begin{aligned}
\frac{x}{y} - \frac{x^{'}}{y^{'}} &> \frac{x^{'} - a}{y^{'} + b} - \frac{x^{'}}{y^{'}} \\
&> \frac{x^{'} - a}{y^{'}} \frac{1}{1 + \frac{b}{y^{'}}} - \frac{x^{'}}{y^{'}} \\
&> \frac{x^{'} - a}{y^{'}} \left(1 - \frac{b}{y^{'}} + \frac{\frac{b^2}{y^{'2}}}{1 + \frac{b}{y^{'}}} \right) - \frac{x^{'}}{y^{'}} \\
&> \frac{x^{'} - a}{y^{'}} \left(1 - \frac{b}{y^{'}} \right) - \frac{x^{'}}{y^{'}} \\
&> -\frac{a}{y^{'}} - \frac{bx^{'}}{y^{'2}} + \frac{ab}{y^{'2}} \\
&> -\frac{a}{y^{'}} - \frac{bx^{'}}{y^{'2}}
\end{aligned}$$

On the other hand:

$$\begin{aligned}
\frac{x}{y} - \frac{x^{'}}{y^{'}} &< \frac{x^{'} + a}{y^{'} - b} - \frac{x^{'}}{y^{'}} \\
&< \frac{x^{'} + a}{y^{'}} \frac{1}{1 - \frac{b}{y^{'}}} - \frac{x^{'}}{y^{'}} \\
&< \frac{x^{'} + a}{y^{'}} \left(1 + \frac{b}{y^{'}} \frac{1}{1 - \frac{b}{y^{'}}} \right) - \frac{x^{'}}{y^{'}} \\
&< \frac{x^{'} + a}{y^{'}} \left(1 + \frac{b}{y^{'} - b} \right) - \frac{x^{'}}{y^{'}} \\
&< \frac{a}{y^{'}} + \frac{(x^{'} + a)b}{y^{'}(y^{'} - b)}
\end{aligned}$$

Therefore:

$$\begin{aligned}
|\frac{x}{y} - \frac{x^{'}}{y^{'}}| &< \max \left(\frac{a}{y^{'}} + \frac{bx^{'}}{y^{'2}}, \frac{a}{y^{'}} + \frac{(x^{'} + a)b}{y^{'}(y^{'} - b)} \right) \\
&< \frac{a}{y^{'}} + \frac{(x^{'} + a)b}{y^{'}(y^{'} - b)}
\end{aligned}$$

Let $x \geq 0$ and $y > 0$. If x_n, y_n, a_n, b_n are four series such that:

$$\left\{ \begin{array}{lcl} x_n & \underset{n \to +\infty}{\to} & x \\ y_n & \underset{n \to +\infty}{\to} & y \end{array} \right. \qquad \left\{ \begin{array}{lcl} a_n & \underset{n \to +\infty}{\to} & 0 \\ b_n & \underset{n \to +\infty}{\to} & 0 \end{array} \right. \qquad \left\{ \begin{array}{l} |x_n - x| < a_n \\ \\ |y_n - y| < b_n \end{array} \right.$$

Let $\epsilon > 0$. For n large enough, we have $|x_n - x| < a_n < x_n$ and $|y_n - y| < b_n < y_n$. Hence:

$$|\frac{x}{y} - \frac{x_n}{y_n}| \quad < \quad \frac{a_n}{y_n} + \frac{(x_n + a_n)b_n}{y_n(y_n - b_n)} \quad \underset{n\to+\infty}{\sim} \frac{1}{y_n}\left(a_n + \frac{x_n}{y_n}b_n\right)$$

which gives us the result.

Appendix B: Parameters used in the final model and their calibrated values

$p_{i,j}(s)$	GIR 3	GIR 2	GIR 1	death
GIR 4	$b_{4,3} + a_{4,3} \times s$	$b_{4,2} + a_{4,2} \times s$	$b_{4,1}$	$(1 - b_{4,3} - b_{4,2} - b_{4,1}) - (a_{4,3} + a_{4,2}) \times s$
GIR 3		$b_{3,2}$	$b_{3,1} + a_{3,1}$	$(1 - b_{3,2} - b_{3,1}) - a_{3,1} \times s$
GIR 2			$b_{2,1} + a_{2,1} \times s$	$(1 - b_{2,1}) - a_{2,1} \times s$
GIR 1				1

Table 6: Jump probabilities in the final model.

$F_{i,j}(s,x)$	GIR 3
GIR 4	$W_{\nu^1_{4,3},\sigma^1_{4,3}}(xe^{\beta^1_{4,3}s})$

$F_{i,j}(s,x)$	GIR 2
GIR 4	$(1-\lambda_{4,2})W_{\nu^1_{4,2},\sigma^1_{4,2}}(xe^{\beta^1_{4,2}s}) + \lambda_{4,2}W_{\nu^2_{4,2},\sigma^2_{4,2}}(xe^{\beta^2_{4,2}s})$
GIR 3	$W_{\nu^1_{3,2},\sigma^1_{3,2}}(xe^{\beta^1_{3,2}s})$

$F_{i,j}(s,x)$	GIR 1
GIR 4	$(1-\lambda_{4,1})W_{\nu^1_{4,1},\sigma^1_{4,1}}(xe^{\beta^1_{4,1}s}) + \lambda_{4,1}W_{\nu^2_{4,1},\sigma^2_{4,1}}(xe^{\beta^2_{4,1}s})$
GIR 3	$W_{\nu^1_{3,1},\sigma^1_{3,1}}(xe^{\beta^1_{3,1}s})$
GIR 2	$W_{\nu^1_{2,1},\sigma^1_{2,1}}(xe^{\beta^1_{2,1}s})$

$F_{i,j}(s,x)$	death
GIR 4	$(1-\lambda_{4,0})W_{(\nu^1_{4,0},\sigma^1_{4,0})}(xe^{\beta^1_{4,0}s}) + \lambda_{4,0}W_{\nu^2_{4,0},\sigma^2_{4,0}}(xe^{\beta^2_{4,0}s})$
GIR 3	$(1-\lambda_{3,0})W_{\nu^1_{3,0},\sigma^1_{3,0}}(xe^{\beta^1_{3,0}s}) + \lambda_{3,0}W_{\nu^2_{3,0},\sigma^2_{3,0}}(xe^{\beta^2_{3,0}s})$
GIR 2	$(1-\lambda_{2,0})W_{\nu^1_{2,0},\sigma^1_{2,0}}(x) + \lambda_{2,0}W_{\nu^2_{2,0},\sigma^2_{2,0}}(xe^{\beta^2_{2,0}s})$
GIR 1	$(1-\lambda_{1,0})W_{\nu^1_{1,0},\sigma^1_{1,0}}(xe^{\beta^1_{1,0}s}) + \lambda_{1,0}W_{\nu^2_{1,0},\sigma^2_{1,0}}(x)$

Table 7: Duration laws in the final model.

$a_{i,j}$	3	2	1	0	$b_{i,j}$	3	2	1	0
4		0.0016	0.0031	− 0.0048	4	0.166	− 0.0022	0.018	0.837
3			0.0012	− 0.0012	3		0.440	− 0.045	0.605
2			− 0.0037	0.0037	2			0.452	0.548
1				0	1				1

$\nu^1_{i,j}$	3	2	1	0	$\sigma^1_{i,j}$	3	2	1	0
4	1.46	1.67	1.95	7.76	4	0.07	0.032	0.58	0.10
3		1.47	1.62	1.28	3		0.15	0.12	0.67
2			1.51	1.17	2			0.12	1.96
1				1.14	1				1.20

$\nu^2_{i,j}$	3	2	1	0	$\sigma^2_{i,j}$	3	2	1	0
4		4.07	4.02	1.17	4		0.08	0.06	0.07
3				2.77	3				0.08
2				2.64	2				0.09
1				2.42	1				0.54

$\beta^1_{i,j}$	3	2	1	0	$\beta^2_{i,j}$	3	2	1	0
4	0.0216	0.0114	0.048	0.0135	4		0.0178	0.0233	0.0248
3		0.0154	0.0187	0.0106	3				0.0190
2			0.0164		2				0.0178
1				0.0129	1				

$\lambda_{i,j}$	3	2	1	0
4		0.52	0.51	0.72
3				0.57
2				0.64
1				0.63

Table 8: Values of parameters in the final model.