# A semi-Markov model to investigate the different transitions between states of dependency in elderly people

Vincent Lepez[1], Svetlana Roganova[2], Antoine Flahault[3]

The cost of the special government help dedicated to dependent elderly people has dramatically increased in France during the recent years. These costs are severe and will continue to grow steeply. It becomes questionable whether the existing social protection system in developed countries will be able to support them in the coming ageing society. In order to predict these future expenses, one needs to be able to estimate precisely the evolution of the number of elderly dependent people, the degree of their dependency, and time spent in various dependency states. As far as we know, rather few attempts have been made to take advantage of a semi-Markov type of approach to model both the inter-states transition probabilities and the laws of transition durations within dependency. Our model was used against a large amount of statistical data concerning the evolution of 52 000 dependent elderly people in France during the period from 2002 to 2005, evidencing more than 27 000 state transitions. Some public French data concerning total population by age was also used to directly compute the laws of entry into Long Term Care (LTC) states. We then built a statistical model describing the transitions included in the available data.

The semi-Markov model is a generalization of standard Markov chain models. The form we used allows to independently deriving the probabilities of transition and the duration probability laws of these transitions. Such a model is particularly appropriate, since all duration laws of passages to death as observed in the dataset clearly exhibit a non-conventional bimodal structure. Such a shape is typical for the description of mixed population groups. This confirms the assumption of at least two types of mortality in dependent people: one due to severe illness causing entry in dependency, with a very short life expectation, and a second one caused by the simple effect of age eroding capacities, such as dementia, with a much longer mean duration time. To our knowledge, this fine-tuned mixed effect is not well captured in dependency modeling so far, potentially leading to serious pitfalls in previous predictions in this field. This model may be used practically to construct annuities insurance products or evaluate financial needs for social coverage, both aligned with typical dependency multi-state process and examples can be found in [22].

We do believe that this area should deserve more attention and effort, by using stronger investments in data gathering, both from private and public sectors, and a deeper applied mathematical research than it did in the past.

---

[1] SCOR Global Life, 5 avenue Kleber, 75795 Paris Cedex 16 – (formerly working for CPRPSNCF, Marseille at the time this study was performed) – vlepez@scor.com

[2] URSSAF Marseille, 13009, Marseille – (formerly working for CPRPSNCF, Marseille at the time this study was performed) – roganova@gmail.com

[3] Université Descartes, Sorbonne Paris Cité, France – antoine.flahault@parisdescartes.fr

## Long Term Care as a new risk

During the last century life expectancy in France has almost doubled, increasing from 40 years in 1900 to 78 years in 2000. A similar trend can be observed in all developed countries. At the same time, the stabilization or even decrease in actual birth rates coupled with the considerable number of aging baby-boomers has led to a significant rise in the growth rate of the elderly population. In France for instance, till 2050, the number of people over 60 will double, the one for people over 75 will triple and the one for people over 85 will quadruple (see [1] for instance). However, very first insurance products only appeared in the early 80's. Since those products are usually bought starting at age 50 or more, they have a duration that could reach 50 years, meaning that we did not yet complete a full cohort statistical history of the underlying dependency risk. This is why this risk is so new to all of us and hence tough to manage and predict.

It is also known that, despite medical progress, many old people spend the last years of their lives in a state of dependency, total or partial, physical or moral. The costs of necessary long term care for dependent people are extremely high[4] and it is clear that the existing social security system is not and will not be able to support them fully. The development of new types of mandatory insurance, either state funded or individual, and related special funds could be the way forward. Some Asian and European countries, such as Singapore and Germany, already started trying to implement these strategies; others are lagging behind in terms of social policy developments.

In any case, it is essential that we find a way to estimate the increase in the number of elderly dependents, as well as the degree of their dependence and the time they will spend in the various states they will visit through their trajectories into dependency. It is the only way we will be able to precisely forecast future expenses.

The aim of this study is to build a strong mathematical model which can help us form a better understanding of these compelling issues.

First of all, let us be more specific about our purpose. We define "dependent elderly" as people who have suffered a deterioration of functional capacity due to advanced age and are forced to rely on mechanical or human assistance to carry out fundamental Acts of Daily Living (ADL). In the following study, we use the AGGIR grid defined by the French government [2], [3] to qualify dependency levels. It is one of several available grids measuring dependency and has been chosen as the legal reference in France. In this grid, each person is assigned one of a possible six GIR degrees (stands for *Groupe Iso-Ressources* – fr – or iso-resource group), based on the level of assistance they need to fulfill basic everyday activities. On the GIR scale, degree GIR 1 corresponds to most sever state of dependency and GIR 6 to self autonomy. GIR 2 to 5 are gradually decreasing levels of dependency. GIR 1 & 2

[4] Costs are over 2 500€ per months on average in France, while the average monthly pension is 1 200€ and the monthly average social "APA" allowance is 500€. There hence remains a significant gap to cover.

are often considered as “total” dependency states and 3 & 4 as “partial”. For the purposes of this study, and in line with French *Allocation Personnalisée d’Autonomie* (APA) [3] social allowances, we have considered GIR 5, in addition to GIR 6, as non-dependent levels, because it indicates very low level of dependency.

## Data

It is clear that a significant amount of data is needed to accurately estimate a statistically ambitious model. However, as stated above, dependency is a phenomenon which has only quite recently come under observation, meaning good statistical data is often hard to find, inadequate, insufficient, biased (in particular by underwriting and claims management discrepancies), etc. Most of the data provided by medical sources covers only the current state of dependency and not the changes in degrees of dependency. Most of the data available in insurers and reinsurers portfolios only cover one dependency level (a “total” one). Moreover, the use of different definitions and grids for degrees of dependency (ADLs, GIR, etc.) makes the task even more complex.

For this reason, the DREES[5] kindly agreed to share data they have collected from local French authorities on the evolution of people along the GIR scale, who received special government assistance allowance (APA) dedicated to dependent elderly people [4]. Note that this assistance is not heavily dependent on individual financial resources so, despite the unavoidable non self-declared bias, we consider the dependent population receiving APA to be a good representation of the total dependent population in France. The practical aspects of our research will rely on this data, which is comprehensive on the geographical areas of study[6], and covers 52 000 dependent elderly people in France, during the period 2002 to 2005. This dataset provides information on the initial degree of dependency and how this dependency evolves for each of these people: changes in degrees of dependency up to death and the time all these changes take. More than 27 000 GIR transitions were observed in the data sample.

We will focus specifically on people who died during the observation period. The related data defines a set of complete trajectories in dependency, with information from initial entry into dependency to death. For the numerous "still-alive-at-the-end-of-observation" others, the observation is said to be right-censored as observation is suspended at a level of dependency prior to death. The evolution of dependency for these people is not complete as we did not have enough time to observe them until the end of their lives and dependency trajectory.

[5] Direction de la recherche, des études, de l'évaluation et des statistiques, France

[6] Four French *Départements*: 04 – Alpes de Hautes Provence, 71 –Saône et Loire, 76 – Seine Maritime, 95 – Val d’Oise. Technically speaking, focusing on these geographical areas to infer global statistics can be interpreted as « grape sampling ».

Considering, as explained hereafter, that dependency is a degenerative process, a dependency trajectory end is inevitably death. The heavy censorship we face significantly complicates standard estimation procedures, but *ad-hoc* statistical methods do efficiently exist.

The structure of the data leads us to consider a stochastic model incorporating five different states: two partial dependency levels (GIR 4, GIR 3), one level of heavy dependency (GIR 2), one level of total dependency (GIR 1), and the final so-called absorbing death state. In the following analysis, we define $G_1$, …, $G_4$ the states of GIR 1 to GIR 4 and $G_0$ as the state of death.

Dependency process modeling is divided into two stages. First, we describe the process of becoming dependent or the "entry into dependency", *i.e.* transition from the non-dependent group of the elderly population to the dependent group. Then, we focus on dependent people and model their evolution through different states of dependency. By combining these two stages we get a complete model of dependency evolution from the initial state of dependency to death.

## Entry into dependency

To model the entry into dependency stage, we use dependent population data, but also data on the total French population of geographical area corresponding to our dataset, over the four years of observation [5]. Over this period we notice that, for a given age, both the ratio between the number of newly dependent people and the total population and the distribution of GIR degrees among these newly dependent people are stable. These elements can therefore be taken as reliable characteristics of the entry into dependency process. The rate of entry into dependency as a function of age is presented in Figure 1 below, where different GIR states are depicted by different colors.

We then model the entry process as follows: each year, and for each age group, we apply the aforementioned rate of non-dependent people becoming dependent and assign them to the appropriate level on the previously established GIR degree scale. Although this stage of modeling is relatively simple, it is important to note that the rate of "entry into dependency" plays a pivotal role in the model. Indeed, this rate is highly sensitive to the socio-economic structure of the population at stake. In this study, we base our research on APA data, which represents a sample of the entire French population. However we have also analyzed this phenomenon for retired SNCF[7] employees and their relatives (spouses and husbands). We noticed that the rate of entry into dependency for this particular population is almost half the rate indicated by APA data. This remarkable finding is probably a result of selection bias due to heavy pre-employment health screening and the provision of regular medical check-ups to

[7] Société Nationale des Chemins de Fer Français

these employees. As a matter of fact, the resulting total number of dependent people is here significantly lower than usual. It would therefore be more accurate to evaluate the entry into dependency rate of each specific population on a separate basis if possible. If not, confidence intervals should be set for the purpose of designing general insurance products.

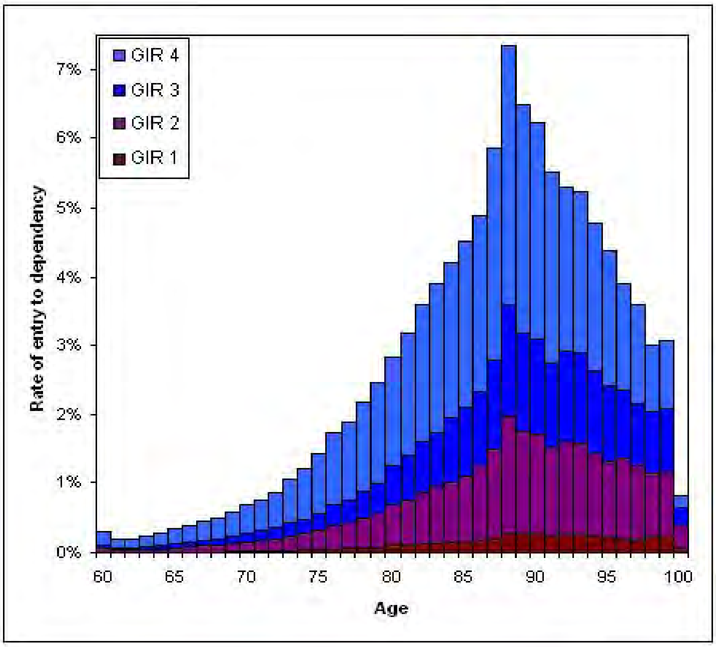


Figure 1. **Dependency entry rate as a function of age**.
Different colors correspond to entry into different GIR degrees.

## Modeling of transitions

As far as dependency modeling is concerned, we shall assume that it is not influenced by population structure or global characteristics, unlike the entry into dependency rate. The underlying assumption is that once people are dependent, their past health patterns do not influence much their future trajectories, which is in line with a Markov-type of hypothesis.

Our goal is to build a statistical model which describes the inter-state transitions included in the available data as accurately as possible. As some transitions are very unlikely or undesirable from an insurance standpoint, we first have to choose which of them are worth including in the model. We then need to find adequate probability laws for the chosen transitions and their different durations, and evaluate these laws before finally interpreting the results.

We begin by studying the observed frequencies of all possible transitions between different levels of dependency (see figure 2). Consequently, we take two assumptions.

First, the so-called “no miracle assumption”: the degree of dependency cannot decrease with time *i.e.* health improvements are very rare and for the purposes of this study will be considered non-existent. It is important to note that this would be seen as a reasonable statement to a cautious insurer, who may use this model to develop an annuity-type of insurance product.

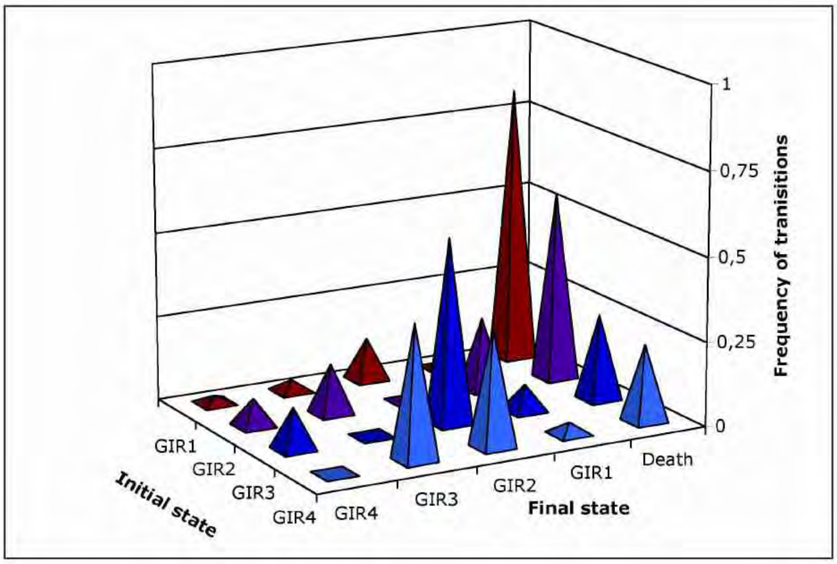


Figure 2. **Observed frequency of transitions between GIR states.**
Different colors correspond to different initial GIR degrees.

Second, “no hopping assumption”: a direct transition from light to total dependency is considered impossible. In other words, a person must pass through the state of heavy dependency (GIR 2) before becoming totally dependent (GIR 1). Nevertheless, a direct transition to death is still possible regardless of the initial state.

To sum up, eight transitions will be embedded in our "five states" model: four intermediate transitions to higher degrees of dependency and four transitions to death (see figure 3). Moreover, our assumptions clearly imply that once a person enters the GIR 1 state, the next state is inevitably death.

We have chosen to describe the transitions between different levels of dependency using a semi-Markov model [6], [7] (see the Methods section below for more details). Roughly speaking, this model is a generalization of the standard Markov chain model, where the next state only depends on the current visited state and each transition comes with its own probability.

In a process outlined by our version of a semi-Markov model, we first define the next state to be entered using probabilities of transition from the current state. We then define the duration of this transition using a random variable with a distribution that depends on its initial and final states. Once the transition is complete, and unless that transition leads to death, the whole procedure begins again.

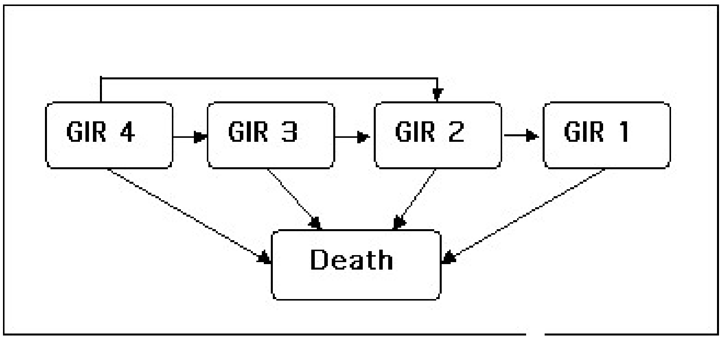


Figure 3. **diagram of the transitions between GIR states included into our semi-Markov model.**

There are fundamental advantages to using this model as opposed to the standard Markov model. With the type of semi-Markov model we intend to use, we independently define the probabilities of transition and the duration of that transition, whereas in a Markov model the duration is directly linked with probabilities of transitions[8]. However, the life expectancy in dependency is *the* crucial parameter for insurers when estimating costs so, being able to investigate duration exclusive of probability allows us to be much more flexible and accurate. Later in this study, we will present another argument which validates our reasons for choosing the semi-Markov model over the Markov model.

In this article, we use a special type of semi-Markov model, where all distributions considered are independent of age, *i.e.* the probability distribution function defined in the methods section is based purely on the duration of transition and not the age of transition. The same is valid for survival functions. This assumption may appear to be a strong one and maybe at odds with the argument for flexibility stated above. It will also be obvious to any reader familiar with mathematical modeling that such an assumption deeply simplifies computations and may have been used purely for that purpose. However, a naturalistic argument exists which supports this assumption. Indeed, it is generally accepted, but however difficult to prove, that increases in life expectancy over time are increases in non-dependent years (see [8] and [9] for instance). As a result, only the time of entry into given levels of dependency will be delayed by age effect (with possibly decreasing probabilities) while actual behavior within those dependent states will remain the same regardless of attained age. To sum up, in this version of semi-Markov model, only the age of a person determines the probability of transitions to the next level, but not the durations of these transitions.

[8] The transition duration probability law actually is an Exponential law with parameter the transition probability.

We then have a choice between parametric or nonparametric approaches. Parametric modeling requires the definition of functional forms of the probability and duration laws as a prerequisite to estimation. It allows us to forecast unobserved behavior under the assumption that the aforementioned choice of functional form adequately describes the phenomenon being studied. Nonparametric modeling allows wide variability of the functional form for the laws under consideration (piecewise constant, splines, etc.) and provides a very accurate fit for data over the given period of observation without restricting us to explicit parametric sets of laws such as standard probability densities. Unfortunately, it can only provide information over the four year range covered by our data and with usually poor prediction capabilities. We will therefore only use this estimation technique to guess the overall pattern of the probabilistic laws under consideration before calibrating the final parametric families to be used. Such a procedure ensures that we not only use an adequate functional form of laws for the modeling but also obtain a model adapted to prediction purposes.

Whether our laws for transition probabilities and duration are chosen from parametric or nonparametric families, we perform Maximum Likelihood estimation techniques (see [10] for instance). The Likelihood function is built on the basis of all available data, both censored and uncensored, using the procedure described in the methods section.

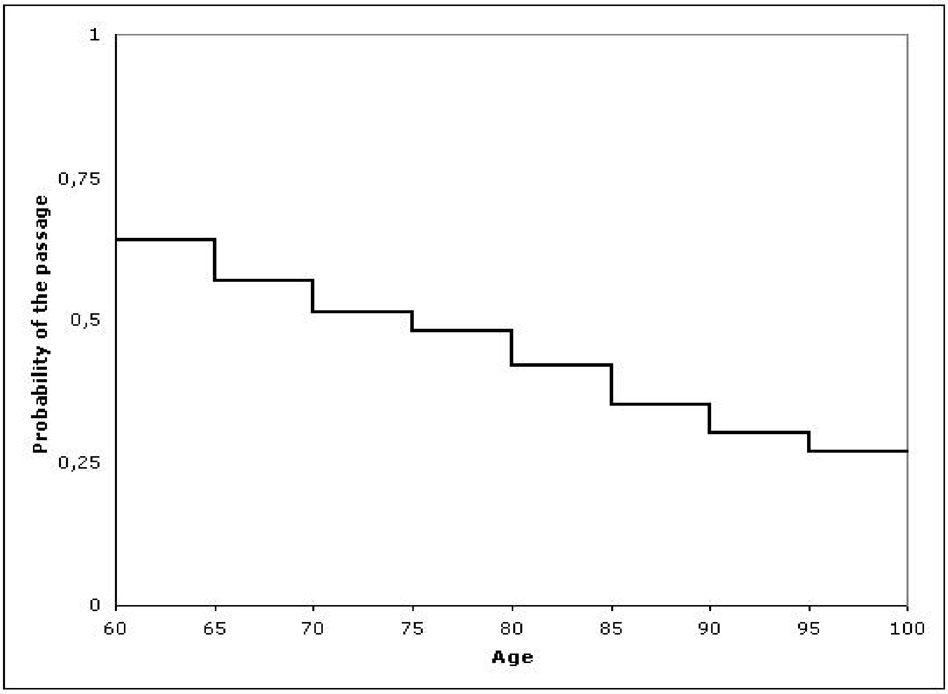


Figure 4. **Probability of passage from GIR 2 to death as function of age obtained by a nonparametric estimation**.

The nonparametric estimation of the probability of transitions was calculated using piecewise constant functions over the age interval [60, 100]. The results (see for example, the probability of the transition GIR 2 to death, figure 4) suggest that linear functions of age can be used in this estimation, even if they are rather unusual for describing probabilities.

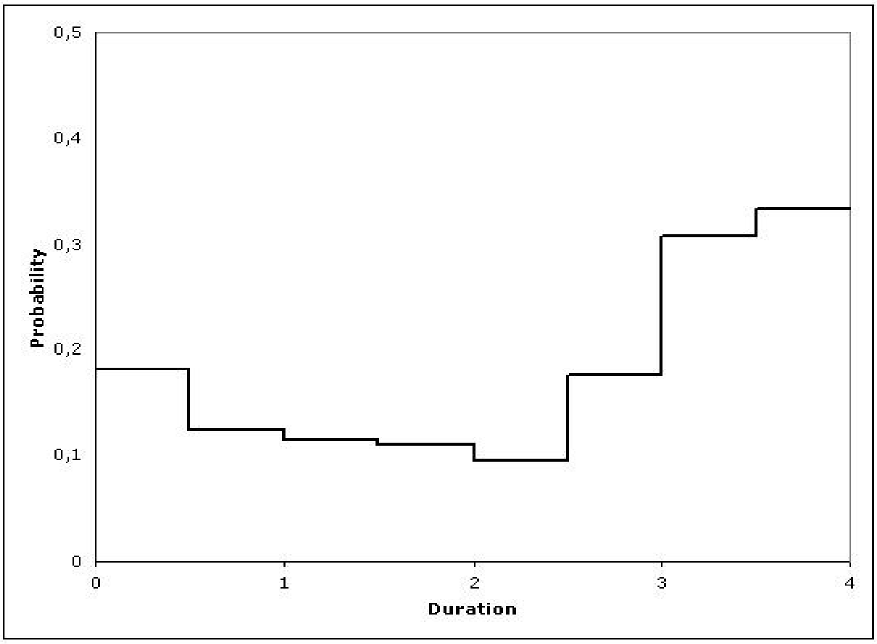


Figure 5. **Distribution of duration of passage from GIR 1 to GIR 0 (death) obtained by a nonparametric estimation**.

The same method is applied to the duration laws. It was observed that all transitions to death clearly present a bimodal structure (*i.e.* the duration law has two local maxima). An example of this is the duration distribution for the transition GIR 1 to death, shown in figure 5. On the other hand, transitions between non death states are unimodal and gently decrease over time (see figure 6).

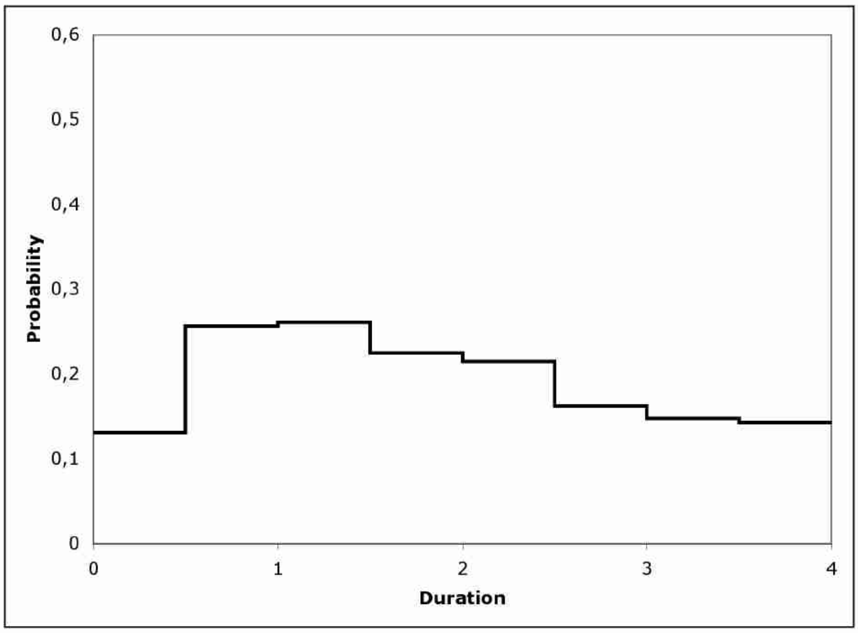


Figure 6. **Distribution of duration of passage from GIR 2 to GIR 1 obtained by a nonparametric estimation**.

The specific behavior for transitions to death is rather nonstandard. In fact, such bimodal laws are encountered when describing mixed samples of populations with different sets of characteristics. It leads us to suppose that there are indeed two different types of transition to death. The first maximum, corresponding to relatively quick death, could be the result of dependency caused by severe illness, whereas the second maximum represents elderly people whose dependency is a result of age alone through dementia, and whose life expectancy is actually much longer.

One should note that this bimodal pattern also reveals how a pure Markov model is inadequate for modeling multi-state trajectories into dependency. Indeed, implicitly embedding exponentially decreasing transition duration laws, it can neither guarantee a fair description of the data, nor provide reliable predictions. These arguments further validate our choice of a semi-Markov model for dependency trajectories modeling.

In light of the results of nonparametric estimation, we will assume that:

- Transitions between states of dependency, excluding death, mimic behavior illustrated by the Weibull distribution, a typical element of duration and reliability models [11] and [12]. Remember that the probability density of the Weibull distribution $W(\nu, \sigma)$, is given by

  $$f(x) = \nu\sigma^{\nu} x^{\nu-1} \exp\left(-(\sigma x)^{\nu}\right), \text{ with } \nu > 0, \sigma > 0;$$

- Taking into account the bimodal behavior mentioned above, transitions to death are well depicted by a convex combination of two Weibull distributions

  $$\alpha\, W(\nu_1, \sigma_1) + (1-\alpha)\, W(\nu_2, \sigma_2), \text{ with } 0 \leq \alpha \leq 1.$$

It is important to bear in mind that this choice, inherited from reliability theory for mixed samples with different behavior (see [13] and [14]), significantly increases the number of parameters we need to estimate:  from 2 in the case of a simple Weibull distribution, to 5 now, for each transition.

Summing up the ideas provided by nonparametric modeling, we define our semi-Markov model as follows.

Probabilities of transitions from state *i* to state *j*, denoted by $p_{i,j}$, will be described by linear functions of age $t$ (note that age 60 corresponds to $t = 0$) :

$$p_{i,j}(t) = a_{i,j} \times t + b_{i,j}.$$

All durations for transitions to death will be parameterized by convex combinations of two Weibull distributions. For other transitions, single Weibull distributions will be used.

Ultimately, our model includes 14 parameters for probabilities and 28 for durations. Taking into account standard constraints for probabilities however (see methods section for details), the number of independent parameters for transition probabilities is actually 8. Consequently, our semi-Markov model is based on 36 parameters.

## Results

We move on to estimate the model's parameters by applying the parametric maximum likelihood method to our statistical data. Results of the transition probabilities estimation are presented in table 1. Note in particular the increasing weight of transitions to death both with increasing level of dependency and age.

| | GIR 3 | GIR 2 | GIR 1 | GIR 0 |
|---|---|---|---|---|
| GIR 4 | $-0.008 \times t + 0.708$ | $0.006 \times t + 0.139$ | | $0.002 \times t + 0.153$ |
| GIR 3 | | $-0.001 \times t + 0.638$ | | $0.001 \times t + 0.362$ |
| GIR 2 | | | $-0.011 \times t + 0.652$ | $0.011 \times t + 0.348$ |
| GIR 1 | | | | $0.000 \times t + 1.000$ |

Table 1. **Functional forms of transition probabilities $p_{i,j}(t)$,**
initial state *i* on the *y* axis, final state *j* on the *x* axis.

Transition durations estimates are depicted in table 2. It is interesting to see that the differences between various transitions to death mainly vary by the weight associated with "fast" or "slow" Weibull component. Beside transition from GIR 4 to death, which probability is actually low, we may notice that the higher the level of dependency, the higher the weight of the slow Weibull component. This effect would need further socio-medical interpretation to be comfortably understood.

| Transitions GIR $i \rightarrow$ GIR $j$ | Transition GIR $i \rightarrow$ GIR 0 |
|---|---|
| $f_{4,3} = W(1.40, 0.22)$ | $f_{4,0} = 0{,}41 \times W(1.35, 0.69) + 0{,}59 \times W(5.08, 0.28)$ |
| $f_{4,2} = W(1.69, 0.40)$ | $f_{3,0} = 0{,}73 \times W(1.08, 0.31) + 0{,}27 \times W(5.90, 0.27)$ |
| $f_{3,2} = W(1.47, 0.30)$ | $f_{2,0} = 0{,}51 \times W(1.17, 0.51) + 0{,}49 \times W(5.98, 0.28)$ |
| $f_{2,1} = W(1.47, 0.20)$ | $f_{1,0} = 0{,}26 \times W(1.16, 0.95) + 0{,}74 \times W(4.14, 0.24)$ |

Table 2. **Functional forms of transition durations laws.**

It is revealed that some transition probabilities significantly depend on age. Observing the probability of a transition from GIR 2 or GIR 4 to death for example, we see that dependent 60 year olds have a much smaller chance of dying straight away than dependent 90 year olds. It seems natural that younger dependents, as opposed to older people, are more likely to increase their level of dependency before dying. In this way, the model seems based on common sense. Nevertheless, the probability of a transition from GIR 3 to death remains stable regardless of age, a phenomenon for which there doesn't seem to exist a straightforward explanatory argument.

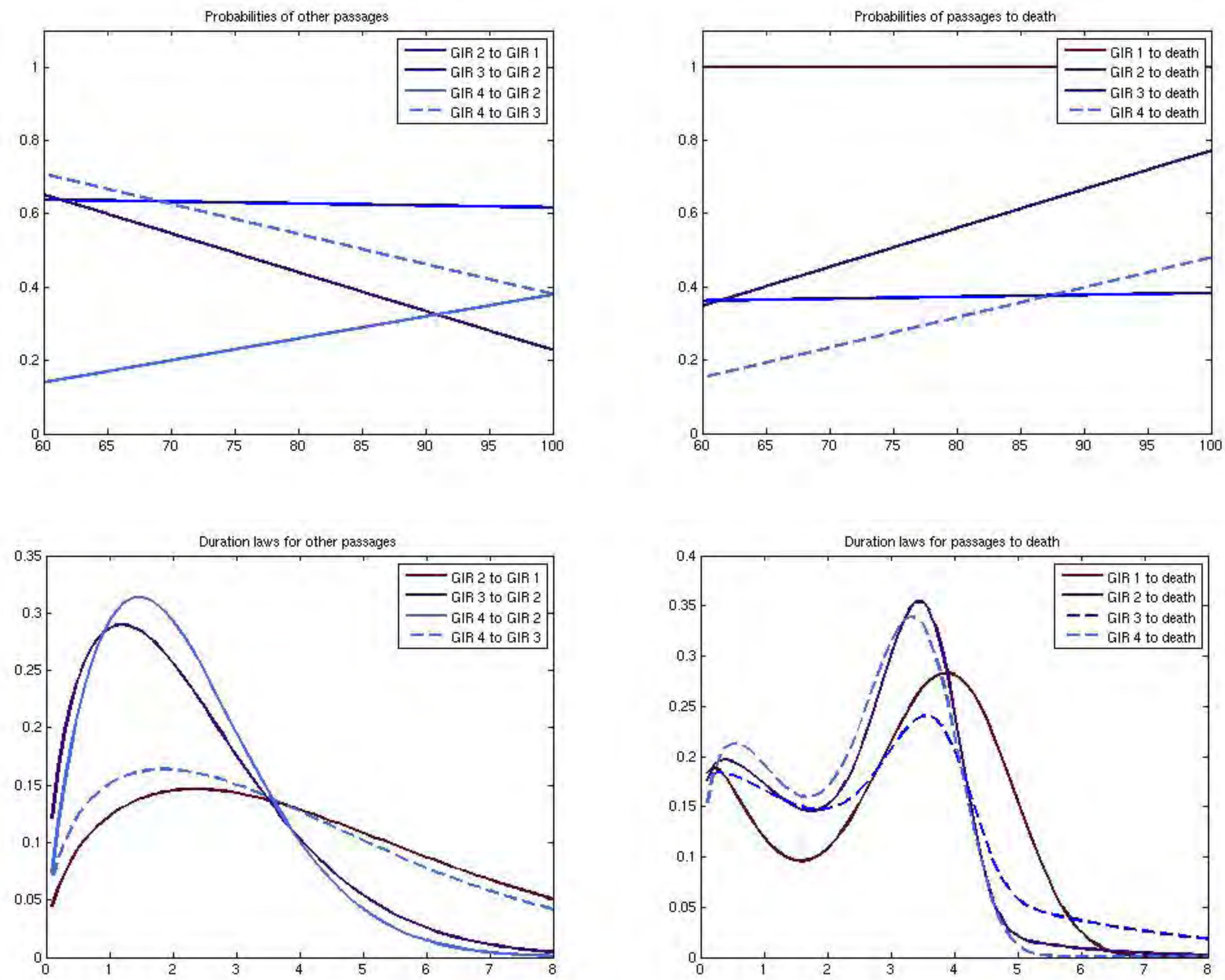


Figure 7. **Probabilities laws and duration laws of passages obtained by parametric estimation method.**

We also notice that the estimated laws for durations of transitions to death are effectively bimodal. This confirms our idea that there are two types of mortality for dependent people: first, as a result of severe illness, so a very short life expectancy, and second, as a result of old age, so a much longer life. To our knowledge, these characteristics, which seem obvious when highlighted here, are not well captured by standard insurers and socio-demographic dependency modeling. The only reference we were able to find regarding semi-Markov modeling of elderly dependency, focusing on socio-medical causes to dependency and their impacts on life expectancy is [16]. It accurately describes duration laws but requires underlying socio-medical data about individuals, which usually is a huge caveat.

As a fundamental and direct application of the parameter estimations, we compute the average time of transitions between various GIR states (see table 3). We rather discuss the average time of each transition included in our model, than the overall life expectancy in dependency. In fact, from the insurer's point of view, the detailed information about dependency evolution is much more important. A good insurance product inevitably adapts benefits to the level of dependency (*i.e.* to the long term care needs of a dependent person) therefore, information on the length of time spent in each state of dependency is crucial to any cost prediction.

| | GIR 3 | GIR 2 | GIR 1 | GIR 0 |
|---|---|---|---|---|
| GIR 4 | 4.1 | 2.2 | | 2.4 |
| GIR 3 | | 2.3 | | 3.2 |
| GIR 2 | | | 4.6 | 2.6 |
| GIR 1 | | | | 3.1 |

Table 3. **Mean duration (in years) of passages between GIR states.**
(initial state *i* on the *y* axis, final state *j* on the *x* axis).

It is important to note that the durations obtained through our study for transitions and, more generally, life in a dependency state are greater than we might usually expect considering the results of previous studies of elderly dependency (see [17], [18], [19] and [20]). There are several reasons for this irregularity. Firstly, different definitions of dependency, notably the inclusion or exclusion of the degree of light dependency in the model, can significantly alter results, as a lightly dependent person has a longer life expectancy. Secondly, our study uses an innovative semi-Markov model to directly model laws governing transition durations in dependency. We have also been untrusted enough by DREES to have access to a significant amount of statistical data related to the evolution of dependent people, while previous studies have been based on much smaller or less appropriate samples or cohorts. Ultimately, we can see that our results reflect the current situation observed in France, where the cost of APA funding, still in its early stage, is increasing rather faster than was initially expected (see [21] for instance).

Nevertheless, we must remind the reader that our estimations are uniquely based on data available over four years; with calculated average transition durations actually reflecting the same order of magnitude. Given the nature of the data and results, one could speculate that when data is observed over a longer period our model would return even longer average durations. Fortunately though, stress tests using the same data, censored to 3.5 and 3 years instead of 4, support our study by presenting us with very similar results.

Once the transition probabilities and duration laws are calibrated, it is rather simple to derive practical use of it in order to build up some insurance products based on the underlying model, where benefits are directly linked with the level of dependency (annuities or lump sum type of products can be envisaged). Those products can be constructed in several ways. The first natural one is an individual type of product, with a capitalized mathematical provision constituted and exhausted throughout the lifetime of the policy. Second, one can easily build up a group type of products, either financed with a risk adjusted premium, or financed as a "pay-as-you-go" type of guarantee, such as the APA system in France. Several illustrative examples of applications and related actuarial calculations can be found in [22].

## METHODS SUMMARY

### Semi-Markov model

Let us go back to some definitions for a semi-Markov model (see [6], [7]). Given a probability space ($\Omega$, $F$, $P$) and a state space $E = \{1, 2, \ldots, K\}$, we consider two stochastic processes: $G_n$: $\Omega \rightarrow E$ representing the current state at the $n$-th transition and $T_n$: $\Omega \rightarrow [0, +\infty)$ describing the chronological time of entry into this state. The variable $S_{n+1} = T_{n+1} - T_n$ is the duration of the $n$-th transition.

For both $i, j \in E$ and $t, s \in [0, +\infty)$, we define a semi-Markov kernel

$$Q_{i,j}(t, s) = P\,(G_{N(t)+1} = j,\ S_{N(t)+1} \leq s \mid G_{N(t)} = i,\ T_{N(t)} = t),$$

where $N(t)$ is the number of jumps between distinct states during the chronological time period $[0, t]$. The variable $s$ here determines the duration of the transition, *i.e.* the inner time scale starting at the entry into each new state.

Note that

$$p_{i,j}(t) = \lim_{s \rightarrow +\infty} Q_{i,j}(t, s) = P\,(G_{N(t)+1} = j \mid G_{N(t)} = i,\ T_{N(t)} = t)$$

are transition probabilities for an embedded (non-homogeneous while depending on global time $t$) standard Markov process.

We also define the probability distribution of the time spent in each state $i$, given the history of the process, as

$$F_{i,j}(t, s) = P\,(S_{N(t)+1} \leq s \mid G_{N(t)+1} = j,\ G_{N(t)} = i,\ T_{N(t)} = t),$$

and an attached survival function

$$S_{i,j}(t, s) = 1 - F_{i,j}(t, s).$$

For future purposes, it is also necessary to introduce a marginal survival function in state $i$

$$S_i(t, s) = P\,(S_{N(t)+1} > s \mid G_{N(t)} = i,\ T_{N(t)} = \mathrm{t}).$$

The following relation holds: whatever $t, s \in [0, +\infty)$,

$$S_i(t, s) = \Sigma_{j \neq i}\, S_{i,j}(s) \times p_{i,j}(t).$$

It is important to mention that contrary to a standard Markov process, our model only allows transitions between distinct states. In other words, a transition is defined as a change of state, with a corresponding duration. If the state is unchanged over a certain time period, it means that the transition is incomplete. We can rewrite this property as $p_{i,i} = 0$.

Of course, the transition probabilities have to satisfy the standard probability conditions whatever $j \geq i$, and $t \geq 0$:

$$p_{i,i}(t) = 0, \quad p_{i,j}(t) \in [0, 1], \quad \Sigma_{j \neq i}\, p_{i,j}(t) = 1.$$

## Construction of a likelihood function for the semi-Markov model

To build the likelihood function, we use standard techniques developed for instance in [10] with some modifications allowing us to take into account censored data. Suppose that $N$ people were observed. Assume moreover that the $p$-th observed person passes successively through $n_p$ distinct states $G^p{}_1$, $G^p{}_2$, … $G^p{}_{np}$. The chronological times (*i.e.* age of the person) of entry into those states are denoted by $t^p{}_1 < t^p{}_2 < \ldots < t^p{}_{np\text{-}1}$. At the end of the observation period, two different possibilities exist: either the final state is death and have been observed, or the transition is censored. For each observed transition from state $i$ to state $j$, starting at chronological time $t$, and completed during time $s$, we define the term

$$c_{i,j}(t, s) := p_{i,j}(t) \times f_{i,j}(s),$$

where $p_{i,j}$ and $f_{i,j}$ are the probability and duration laws for this transition.

If the transition starting in state $i$ was not completed at time $t$, *i.e.* the observation is censored, the corresponding term takes the form

$$c_i(t, s) := S_i(t, s) = \Sigma_{j \neq i}\, p_{i,j}(t) \times S_{i,j}(s),$$

where $S_{i,j}$ is the marginal survival function between states $i$ and $j$ defined above.

Terms $c_{i,j}$ and $c_i$ are referred as contribution to the partial likelihood function, which is an extension of Maximum Likelihood techniques to censored data. The partial likelihood function is then built as the product over all people and transitions, completed or censored, on the basis of the corresponding terms:

$$L = \prod_{p=1}^{N} \prod_{n=1}^{n_p} [c_{G^p_{n-1}, G^p_n}(t^p_{n-1}, x^p_n)]^{\xi^p_n} [S_{G^p_{n-1}}(t^p_{n-1}, x^p_n)]^{1-\xi^p_n};$$

where $\xi^{p}_{n} = 1$ if the $n$-th transition is completed and 0 if this transition is censored. Keep in mind the fact that $L$ actually depends on all the parameters of the laws $p_{i,\,j}\,(t)$ and $f_{\,i,\,j}\,(s)$, although we have not developed the formulas in detail here to avoid cumbersome notation.

Once the function is constructed, we need to maximize it (under some constraints such as sum of transition probabilities must equate 1) to find the maximum likelihood estimates of the parameters of all our laws.

The calculation was completed using the MATLAB function *fmincon*. As far as our data is concerned, with 52 000 individuals at stake and no less than 27 000 GIR transitions jumps through the 4 years of observation, the parameter estimation of our model made it necessary to maximize a function composed of the product of roughly one hundred thousand components based on 36 parameters.

## Bibliography


[1] Projections de population à l'horizon 2050, *INSEE Première*, **762** (2001).

[2] Journal Officiel de la République Française, 30 April (1997).

[3] Vetel, J.M., Leroux, R., Ducoudray, J.M., AGGIR, Practical use, Geriatric autonomy group resources needs, *Soins Gerontol* (1998).

[4] Données APA 2005, DREES (2005).

[5] Population totale par sexe et âge, INSEE, www.recensement.insee.fr (2005).

[6] Janssen, J., Manca, R., Non homogeneous semi-Markov reward processes for the management of health insurance models, *Proc. ASTIN Washington* (2001).

[7] Mathieu, E., Foucher, Y., Dellamonica, P., Daures, J.P., Parametric and non homogeneous semi-Markov process for HIV control, *Methodology and Computing in Applied Probability* Vol 9, **3**, 389-397 (2007).

[8] Manton KG, Stallard E, Changes in Health, Mortality, and Disability and Their Impact on Long-Term Care Needs, *Journal of Aging and Social Policy*, **7(3/4)**:25-52 (1996).

[9] Jitapunkul *et al*, Disability-free life expectancy of elderly people in a population undergoing demographic and epidemiologic transition, *Age and Ageing*, **32**, 401-405 (2003).

[10] Feller, W., *An Introduction to Probability Theory and its Applications*, Volume II, 2nd edition, John Wiley and Sons, New York, (2005).

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

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

[13] Bucar, T., Nagode, M., Fajdiga, M., Reliability approximation using finite Weibull mixture distributions. *Reliability Engineering and System Safety*, **84**, 241-251 (2004).

[14] Jiang, R., Murthy, D.N.P., Two sectional models involving three Weibull distributions, *Quality and Reliability Engineering international*, **13**, 83-96 (1997).

[15] Duée, M., Rebillard, C., La dépendance des personnes âgées : une projection à long terme, *Document de travail*, INSEE (2004).

[16] Robert, C., Denuit, M., Actuariat des assurances de personnes, modélisation, tarification et provisionnement, *Economica*, **80** (2007).

[17] Gisserot, H., Grass, E., Perspectives financières de la dépendance des personnes âgées à l'horizon 2025 : prévisions et marges de choix, Report, www.lesrapports.ladocumentationfrancaise.fr/BRP/074000235/0000.pdf

[18] *L'assurance dépendance*, Scor Tech (1995).

[19] Loones, A., David-Alberola, E., Eugene, S., Approche du coût de la dépendance des personnes âgées à domicile, *Cahiers de Recherche CREDOC*, **221** (2005).

[20] Eslous, L., *La dépendance des personnes âgées - éléments de travail*, Rapport RM 2007, 130P, août 2007.

[21] Espagnol, P., L'allocation personnalisée d'autonomie au 31 décembre 2006. *Études et Résultats*, **569** (2007).

[22] Lepez, V., *Trajectoires en Dépendance des Personnes Agées : Modélisation, Estimation et Applications en Assurance Vie*, Mémoire d’Actuariat, CEA (2009).

**Acknowledgements**: DREES