DISCUSSION PAPER PI-1306

Long Term Care and Longevity

Christian Gourieroux and Yang Lu

November 2013

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University London
106 Bunhill Row
London EC1Y 8TZ
UNITED KINGDOM

http://www.pensions-institute.org/

Long Term Care and Longevity

Christian Gouriéroux[(1)], and Yang Lu[(2)]

November 6, 2013

Abstract

The increase of the expected lifetime, that is the longevity phenomenon, is accompanied by an increase of the number of seniors with a severe disability. Because of the significant costs of long term care facilities, it is important to analyze the time spent in long term care, as well as the probability of entering into this state during its lifetime, and how they evolve with longevity. Our paper considers such questions, when lifetime data are available, but long term care data are either unavailable, or too aggregated, or unreliable, as it is usually the case.

We specify a joint structural model of long term care and mortality, and explain why parameters of such models are identifiable from only the lifetime data. The methodology is applied to the mortality data of French males, first with a deterministic trend and then with a dynamic factor process. Prediction formulas are then provided and illustrated using the same data. We show in particular that the expected cost of the long term care is increasing less fast than the residual life expectancy at age 50.

Keywords: Longevity, Long Term Care (LTC), Semi-Competing Risks, Unobserved Heterogeneity, Dynamic Frailty, Affine Process, Partial Observability, Identification, Markov Chain Monte-Carlo.

Acknowledgements: The first author gratefully acknowledges financial support of the Global Risk Institute and of the Chair: Regulation and Systemic Risk. The views are those of the authors. They do not necessarily reflect the views of the Autorité de Contrôle Prudentiel et de Résolution (ACPR) and of SCOR SE.

[1]CREST and University of Toronto. Email: Christian.Gourieroux@ensae.fr
[2]SCOR and CREST. Email: yang.lu@normalesup.org

1 Introduction

The general increase of human lifetime, that is the longevity phenomenon, has been largely illustrated in the demographic and insurance literatures [see e.g. Lee and Carter (1992)]. In average we observe each year an increase of 3 months of the life expectancy [see e.g. Oeppen and Vaupel (2002)]. This increase of the expected lifetime is accompanied by an increase of the number of old people who potentially need long term care (LTC henceforth), but also a decrease of the probability of entering into LTC at any given age [see e.g. Manton et al. (1998)], as well as a decrease of the mortality intensity for individuals in LTC *ceteris paribus*[1]. A person enters into LTC when he/she becomes unable to live independently, measured by the ability to do some special Activities of Daily Living (ADL). Because of the significant costs of LTC facilities, it is important to analyze this probability of entry, as well as the time spent in this state and how they evolve with longevity. Are they almost independent of the longevity feature or do they increase at a similar rate? Our paper will answer these questions, when the lifetime data are available, but the LTC data are either unavailable, or too aggregated[2], or unreliable.

Such questions are quite hard to answer, partly because of the competing risks nature of the two risks, that are, the risk of entering into LTC and the risk of dying directly. For instance, although the probability of entry is decreasing at any age, the total probability of entry during its lifetime could well be increasing, since the mortality intensity of dying directly is also decreasing; although the mortality intensities are decreasing in cohort for people under LTC all things equal, the average age of entry will probably increase; therefore the average time spent in LTC may not be necessarily increasing. On the other hand, since the mortality of disabled people is higher than that of the autonomous people, when the mortality is analyzed using only lifetime data, as is often the case in the literature, the health state at a given age[3] is an unobserved heterogeneity whose distribution changes over time. Therefore there is a spurious duration dependence in a similar way as in a population with static unobserved heterogeneity, or static frailty[4] [see e.g.

[1] That is, when all other parameters, for instance the current age, as well as the age of entry are equal.

[2] It is possible to use jointly longitudinal mortality data and aggregated cross-sectional LTC data, that is, data only disaggregated by age, but not by cohort. But this implies no evolution of the LTC data in cohort, which is unrealistic. The limits of this stationary approach is discussed at the end of the fourth paragraph.

[3] Either autonomous, or in LTC, the second group can be further decomposed by the age of entry.

[4] The difference is that, in our example, the heterogeneity of the health state is not time-invariant because of the transition of some people from the autonomous state into the LTC state.

Vaupel et al. (1979) and Elbers and Ridder (1982)]. This effect should be identified in order to study the true duration dependence, that is, the age dependence of the mortality evolution, and how this dynamics changes between different cohorts, that is, the longevity phenomenon. Therefore, it is essential to analyze the joint behavior of the two risks when it comes to the modeling and prediction of either the LTC, or the longevity.

We introduce in this paper a joint model of LTC and mortality, based on the intensity of entry and on the mortality intensities. The model disentangles the mortality intensities according to the time spent in LTC. Moreover we assume that these intensities depend on unobservable dynamic factors (or dynamic frailties) with nonstationary features, able to capture the longevity phenomenon and its potential impact on both the mortality and the long term care.

Such a joint model would be simple to estimate if individual data on both mortality and LTC were available. However there does not exist a universal definition of LTC, since the criteria of losing autonomy are quite vague and may differ both by country and insurance company. For instance, US insurers consider six Activities of Daily Livings, that are Eating, Dressing, Mobility, Bathing, Toileting, and Maintaining Continence, respectively, while their European peers, use only four of them called Instrumental Activities of Daily Living (IADL) [see e.g. Kessler (2008) for a review of the LTC insurance market]. This discrepancy is even larger between public LTC insurance plans in different countries (often Western European), where it is a pillar of the social security system. Concerning data quality, even in countries where public data exist, they often lack accuracy. Indeed, collecting the LTC data is a much more demanding task than, say, collecting the mortality data : besides the potential death age, it requires the knowledge of the entire history of each individual, especially the transition time(s) identified by accredited physicians. Most of the time, available public data of the national population only exist for a few years when there is a census. This is for instance the case for the "Enquête Handicaps-Incapacité-Dépendances" in France[5], as well as the NLTCS database in the US[6], which are also by far the two largest markets for private LTC insurance. On the other hand, as the sustainability of these Welfare States is more and more questioned, public's appetite for private LTC insurance

[5]Literally the Disability-Incapacity-Long Term Care Survey. This survey has been conducted in 1998/1999 and then in 2008/2009.

[6]National Long Term Care Survey. The data are based on surveys conducted in 1982,1984,1989,1994,1999, 2004 on a representative sample of the US population. Its official site is `http://www.nltcs.aas.duke.edu/`

is steadily increasing [see e.g. Rice (1989)], but insurers still lack experience, both in data management and modeling. Another problem is that most datasets are cross-sectional, either by nature or because the observation period is too short to deliver longitudinal information. So from the very beginning they are not suited for the understanding of the risk. Indeed, by using such a cross-sectional database one will in general ignore the evolution in cohort of different transition probabilities at given ages [see Keiding (1991) for a discussion on the limits of this stationary approach], and this nonstationary trend is exactly the biggest concern of all demographic studies.

This partly explains why data on LTC are often missing and not very reliable when they exist. For instance, for France, public databases based on different population samples show different trends of the LTC prevalence[7] [see e.g. the OECD report by Lafortune and Balestat (2007)]. This uncertainty is one of the biggest risks insurers will face [see e.g. Kessler (2008)] and is a serious obstacle to the further development of the private LTC insurance market in many countries. In the Insurance literature, as far as we know, Levantesi and Menzietti (2012) is the only attempt to model jointly the LTC and longevity behavior, and their analysis is based on observations in a ten-year interval of the Italian population but questions raised at the end of the first paragraph are not directly answered.

Our paper will develop a methodology to estimate this joint model when the mortality data are the only available ones. The parametric specification of the intensity functions characterizes explicitly the distribution of the heterogeneity of the health status which we explained in the second paragraph. The possibility to identify the characteristics of LTC from the mortality data is due to the jumps in mortality intensity arising when entering LTC. Our model allows us to predict jointly the future evolution of the LTC entry probabilities and the mortality rates.

The paper is organized as follows. In Section 2, we discuss the structural and reduced form approaches, which can be used for a joint modeling of LTC and mortality. This modeling is used in Section 3 to derive the joint distribution of the lifetime and of the date of entry. To derive this distribution we follow a progressive approach. We first consider the case of "observable" intensities, then we render them stochastic by introducing a static frailty. Section 4 extends the modeling to include nonstationary dynamic frailties able to capture the longevity phenomenon and its effect on LTC. In Section 5 we introduce parametric specifications for the intensities

[7] That is, the proportion of people in LTC.

and for the frailty dynamics, discuss the way of introducing a generation effect and derive the form of the log-likelihood functions when the lifetimes are observed with right censoring. The models are estimated for the population of French males in Section 6. We first consider a model without frailty in the spirit of the Lee-Carter model, but allowing for non degenerate intensities in a far future. We allow for either Markov or semi-Markov mortality intensity functions. Then the model is extended to include the uncertainty on the longevity factor by means of a dynamic frailty process. We also explain how to filter out this frailty process once the model is estimated. In Section 7 we implement the model for prediction purpose. Section 8 concludes. The proofs and the descriptions of the algorithms are gathered in appendices.

2 Structural versus reduced form approach

We consider a situation where an individual can either experience first a **non terminal event** and then fail, or he/she can fail directly. In both situations the failure is called the **terminal event**. In the second case, the terminal event censors the non terminal event. The corresponding model is called the semi-competing risks model in the literature [see e.g. Fine et al. (2001); Xu et al. (2010)]. In our framework, the non terminal event is the potential entering into LTC and the terminal event is the death. The migration from the autonomous state to the LTC is assumed irreversible.

We start by discussing the limits of the popular copula-based approach. Then we propose a structural approach with latent variables corresponding to the times elapsed up to the potential events and describe how the observable variables depend on the latent duration variables. We also introduce another existing methodology, the reduced form approach, in which the intensities of the different events are directly specified. This gives an alternative interpretation of the structural approach.

In the literature, most multivariate survival models are written in continuous time. The main reason is that in the continuous time intensity-based setting, the probability of observing tied events is naturally null[8]. In our example, we would like to avoid the simultaneous arrival of

[8]Under some mild regularity assumptions, for instance, the continuity of the joint distribution on the observable domain $0 < Y_1 \leq Y_2$. Note that this is still possible with a discrete time framework, but it will be more complicated to analyze.

both the non terminal and the terminal event. We follow this choice, at least for the theoretical model. The continuous time model is discretized when it comes to numerical estimation of the model with dynamic frailty.

We begin our analysis by considering only one cohort (generation). In this case and without left censoring (which we also assume for the time being), we can use either the terminology "age" or "time" to denote the elapsed duration. From Section 5 on, when the cohort effect (which we denote by t_0) is introduced, we will more frequently use the term "age", that is, the age of an individual since its birth. To describe the period effect, we use only the term "calendar time" and we have the following relationship between the three time measures:

$$\text{Cohort} + \text{Age} = \text{Calendar time}.$$

2.1 Structural approach

Semi-competing risks are traditionally written on the two duration variables Y_1 and Y_2, where Y_2 is the time of failure and Y_1 is the potential time of entering into LTC. Therefore, the variable Y_1 is latent in the sense that it is not observable when we observe first the variable Y_2, that is, when $Y_2 < Y_1$. Then the dependence between the two variables is modeled via a survivor copula C [see e.g. Fine et al. (2001) and Hsieh et al. (2008)], that is,

$$\mathbb{P}(Y_1 > y_1, Y_2 > y_2) = C(S_1(y_1), S_2(y_2)),$$

where C is assumed to belong to some specific parametric families, e.g. Archimedean copulas or other factor copulas. This bivariate copula approach is partly borrowed from the literature of competing risks models [see e.g. Zheng and Klein (1995)]. The model is often written with restrictions such as a continuous copula density, and a positive, symmetric dependence structure. Therefore it is not flexible enough to capture the peculiarities of semi-competing risks data. First, they are not adapted to characterize "regime switching" nature an individual may experience. Intuitively, if the individual enters into the non terminal event during its lifetime, then his residual lifetime distribution will be very different from the case when he never experiences the non terminal event. Therefore, using solely one variable Y_2 to model the lifetime is probably not

enough. Besides, the previous reasoning suggests also a possible discontinuity of the joint density distribution of (Y_1, Y_2). A more detailed discussion on the limits of this approach is provided in Gouriéroux and Lu (2013). We propose below another approach with an extra latent variable. More precisely, let us introduce:

- X_1 the potential time of the non terminal event,
- X_2 the (potential) time of death for an individual which has not experienced the non terminal event,
- X_3 the residual lifetime up to the death once the individual experienced the non terminal event.

Some of these variables are really latent even for an econometrician with the maximal available information. Indeed an individual dying before the potential time of the non terminal event will never experience spell X_1 or X_3. At most the observations include the indicator variable Z defined by: $Z = \mathbb{1}_{X_1 \leq X_2}$, that is, whether or not the individual experiences the non terminal event before the death, and the duration variable(s):

$$\begin{cases} Y_1^* = X_1 \text{ and } Y_2 = X_1 + X_3, & \text{if } Z = 1, \\ Y_2 = X_2, & \text{if } Z = 0. \end{cases} \tag{1}$$

In regime 1, we observe the time Y_1^* up to the entry into LTC and the lifetime Y_2. In regime 0, we observe the lifetime only.

The potentially observable model can be rewritten in another form, which avoids the explicit distinction between the regimes. For this purpose, we introduce a variable Y_1 defined by $Y_1 = Y_1^*$, if $Z = 1$, and $Y_1 = 0$, otherwise, which captures both the regime and the duration up to the non terminal event, if the latter is observed. We get:

$$\begin{cases} Y_1 &= X_1 Z, \\ Y_2 &= (X_1 + X_3)Z + X_2(1 - Z). \end{cases} \tag{2}$$

The first equation corresponds to a standard Tobit model [see e.g. Amemiya (1984)] and is completed by an equation providing the observed lifetime depending on the regime.

To our best knowledge, the idea of introducing explicitly a regime change dates back to Freund (1961), who considered only the case of constant hazards; it is later generalized to the previous general form by Tosch and Holmes (1980). Recently this latent model has been generalized to include frailty and applied to the pricing of joint insurance contracts for couples [see Gouriéroux and Lu (2013)].

In general, latent variables X_1, X_2, X_3 are specified by means of their hazard functions as well as some assumptions on the dependence between them. The next subsection gives a natural interpretation of these hazard functions in terms of transition intensities of an individual between different health states.

2.2 Reduced form approach

The model can also be defined by a chain with the three following states:

- state A: the individual is in good health,
- state B: the individual is under LTC,
- state C: the individual is dead. State C is the unique absorbing state.

The transitions are possible only from state A to state B, from state B to state C and from state A to state C. The history of the individual is represented by the discrete process $S = (S_t)$ which takes value in the state space $\{A, B, C\}$. The plot below gives an illustration of the path of an individual's lifetime.

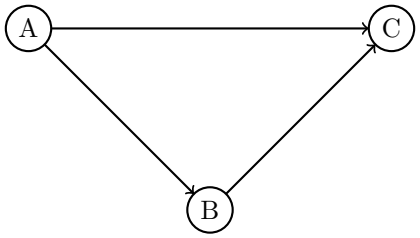


Figure 1: The potential transitions of an individual during its lifetime.

Let us denote by $\underline{S_t}$ the information on past individual history up to time t: $\underline{S_t} = \{S_u, 0 \leq u \leq t\}$, then we define the following transition intensities:

$$\text{If } S_t = A, \mu_1(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = B | \underline{S_t}) \Big\},$$
$$\text{If } S_t = A, \mu_2(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | \underline{S_t}) \Big\},$$
$$\text{If } S_s = S_t = B, S_{s-} = A, \mu_3(t|s) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | \underline{S_t}) \Big\}, \qquad \forall t > s.$$

Due to the discrete nature of process (S_t), the knowledge of $\underline{S_t}$ is equivalent to the knowledge of its current state, of its previous state (if it exists) and of the corresponding transition time. Therefore we can rewrite the transition intensities as follows:

$$\mu_1(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = B | S_t = A) \Big\},$$
$$\mu_2(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | S_t = A) \Big\},$$
$$\mu_3(t|s) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | S_{s-} = A, S_s = S_t = B) \Big\}.$$

The conditions on intensities μ_1 and μ_2 are Markov conditions. The condition on μ_3 is a semi-Markov condition since the transition also depends on the time of entry. This reduced form approach is more commonly called the illness-death model. Its usefulness in modeling semi-competing risks has only been rediscovered recently by Xu et al. (2010).

For expository purpose, the previous formulas are written without conditioning on any covariates, and the definitions can be easily extended to include time-varying exogenous covariates. It is also easily checked that (see the next section) this reduced form specification is essentially equivalent[9] to the structural model we defined if we carefully specify the intensity functions of the latent variables and the dependence structures between them. This should diminish the considerable confusion in the literature that the reduced form approach is different [see e.g. Imai and Soneji (2007)] from the structural approach and should be preferred, partly linked to the decades-long debate on the physical meaning of the latent variables in (semi)-competing risks models [see Prentice et al. (1978) and Andersen and Keiding (2012)]. However, in some cases

[9]Indeed, the only difference is that in the latent variable approach, for technical reasons, the variable X_3 is even defined even if $X_1 > X_3$. However in such cases the value of X_3 is not important.

one approach may appear to be more convenient than the other. To quote a summary from Han and Hausman (1990): while econometricians have emphasized the presence of unobserved heterogeneity (and therefore prefer the structural approach), (bio)-statisticians have instead emphasized the use of semiparametric models which do not require parametric specification of the baseline hazard (hence the choice of reduced form approach, often written without unobserved heterogeneity). Both approaches can be extended to include more LTC states. This is called a multi-state model and can be written either in reduced form [see e.g. Stallard (2011) for an application to LTC and Hougaard (1999) for a general review] or in structural form [see e.g. Abbring and Van den Berg (2003)]. For simplicity, our paper will focus on the case of three states (A, B, C).

3 The distribution of the potentially observable variables

Let us now derive the explicit expressions of the joint distribution of variables (Y_1, Y_2), and also of the marginal distribution of Y_2. We first consider the case in which the two latent variables X_1, X_2 are independent. Then we introduce the dependence by means of (static) frailties in the latent model.

3.1 The basic model

3.1.1 Joint distribution of the latent variables

Let us first assume that the latent variables X_1, X_2 are independent. Their joint distribution is then characterized by their marginal intensities:

$$\lambda_1(x_1) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(X_1 \leq x_1 + du | X_1 \geq x_1) \Big\},$$

$$\lambda_2(x_2) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(X_2 \leq x_2 + du | X_2 \geq x_2) \Big\}.$$

The variable X_3 is in general defined conditional on the values of X_1 and X_2, but it is often assumed independent of X_2. Therefore we denote by $\lambda_{2|1}(x_3|x_1)$ its intensity given the value of $X_1 = x_1$, which depends both on the non terminal event time x_1 and the time elapsed since the

non terminal event x_3.:

$$\lambda_{2|1}(x_3|x_1) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(x_3 \leq X_3 + du | X_3 > x_3, X_1 = x_1) \Big\}.$$

When this function depends on x_1, x_3 only via $x_1 + x_3$, the model is Markov; otherwise, it is semi-Markov.

The joint density function of the latent variables (X_1, X_2, X_3) is:

$$g(x_1, x_2, x_3) = e^{-\Lambda_1(x_1) - \Lambda_2(x_2) - \Lambda_{2|1}(x_3|x_1)} \lambda_1(x_1) \lambda_2(x_2) \lambda_{2|1}(x_3|x_1),$$

where $\Lambda_1, \Lambda_2, \Lambda_{2|1}$ are the cumulated intensities associated with $\lambda_1, \lambda_2, \lambda_{2|1}$, respectively. Therefore the joint survival function of the latent variables (X_1, X_2, X_3) is:

$$\begin{aligned} S(x_1, x_2, x_3) &= \int_{x_1}^{\infty} \int_{x_2}^{\infty} \int_{x_3}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_2(t_2) - \Lambda_{2|1}(t_3|t_1)} \lambda_1(t_1) \lambda_2(t_2) \lambda_{2|1}(t_3|t_1) dt_1 dt_2 dt_3 \\ &= e^{-\Lambda_2(x_2)} \int_{x_1}^{\infty} \int_{x_3}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_{2|1}(t_3|t_1)} \lambda_1(t_1) \lambda_{2|1}(t_3|t_1) dt_1 dt_3 \\ &= e^{-\Lambda_2(x_2)} \int_{x_1}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_{2|1}(x_3|t_1)} \lambda_1(t_1) dt_1. \end{aligned}$$

Under these independence assumptions, we get:

$$\begin{aligned} &\text{If } S_t = A, \qquad \mu_1(t) = -\frac{\partial}{\partial y_1} \log S_{12}(t,t) = \lambda_1(t), \\ &\text{If } S_t = A, \qquad \mu_2(t) = -\frac{\partial}{\partial y_2} \log S_{12}(t,t) = \lambda_2(t), \\ &\text{If } S_s = S_t = B, S_{s-} = A, \qquad \mu_3(t|s) = \lambda_{2|1}(t-s|s), \qquad \forall t > s. \end{aligned}$$

Therefore the latent variable approach is, under the assumptions of this section, equivalent to the reduced form approach. This equivalence is easily extended when (possibly unobserved and/or time-varying) covariates are introduced, if we assume that (X_1, X_3) and X_2 are independent given the whole history of the covariates and we define the transition intensities conditional on the whole history of the covariates. The rest of the paper will use the structural approach, but keeping in mind this equivalence can certainly help the reader better understand certain formulas.

3.1.2 Distribution of the (potentially) observable variables

Let us now derive the joint distribution of the potentially observable variables (Y_1, Y_2). The couple (Y_1, Y_2) has a bi-dimensional continuous component on domain $\mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\}$, and a one-dimensional continuous component on $\mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}$. The joint distribution of (Y_1, Y_2) admits a density with respect to the dominating measure $\lambda_{\mathcal{D}_1} + \lambda_{\mathcal{D}_0}$, where $\lambda_{\mathcal{D}}$ denotes the Lebesgue measure on domain $\mathcal{D}$. This density is:

$$f(y_1, y_2) = \lambda_1(y_1)\lambda_{2|1}(y_2 - y_1|y_1)e^{-\Lambda_1(y_1)-\Lambda_2(y_1)-\Lambda_{2|1}(y_2-y_1|y_1)}, \text{ on domain } \mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\}, \tag{3}$$

and

$$f(0, y_2) = \lambda_2(y_2)e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}, \text{ on domain } \mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}. \tag{4}$$

Many authors write instead the joint distribution of (X_1, Y_2) [see also Xu et al. (2010) for a discussion], in which case there will be no point mass, but instead a continuous component on the unobservable domain $\{X_1 > Y_2\}$ and the restriction of the density function adds up to $\mathbb{P}[X_1 > Y_2] = \mathbb{P}[Y_1 = 0]$ there. These two approaches are therefore equivalent, since in any applications the latent variable should be integrated out. Nevertheless, studying directly (Y_1, Y_2) is preferred because of our desire to distinguish explicitly the potentially observable information, that are (Y_1, Y_2), from the really latent one (X_1, X_2, X_3).

We deduce the marginal survival function and the p.d.f. of the lifetime Y_2, which is later on the only really observable duration variable:

Proposition 1. *The survival function of the lifetime Y_2 is:*

$$\begin{aligned} S_2(y_2) &= \mathbb{P}(Y_2 > y_2) \\ &= \int_0^{y_2} \lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}, \end{aligned} \tag{5}$$

and its p.d.f. is:

$$f_2(y_2) = \int_0^{y_2} \lambda_1(t)\lambda_{2|1}(y_2 - t|t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + \lambda_2(y_2)e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}. \tag{6}$$

Proof. See Appendix 2. □

3.2 Identification in a model with constant intensities

The aim of this subsection is to explain why it is possible to identify the parameters related to the intensity functions λ_1, λ_2, and $\lambda_{2|1}$ when none of them, but only the lifetime variable Y_2 is observed. For illustration purpose, we look at the special case when the intensities $\lambda_1, \lambda_2, \lambda_{2|1}$ are constant, that is, the latent variables X_1, X_2, X_3 are independent exponential variables. In this case the joint density becomes:

$$f(y_1, y_2) = \lambda_1 \lambda_{2|1} e^{-\lambda_1 y_1 - \lambda_2 y_1 - \lambda_{2|1}(y_2 - y_1)}, \text{ on domain } \mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\},$$

and

$$f(y_1, y_2) = \lambda_2 e^{-(\lambda_1 + \lambda_2) y_2}, \text{ on domain } \mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}.$$

The marginal survival function of lifetime Y_2 becomes:

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-\lambda_{2|1} y_2} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-(\lambda_1 + \lambda_2) y_2}\Big] + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2) y_2}, \text{ if } \lambda_1 + \lambda_2 \neq \lambda_{2|1}, \tag{7}$$

and

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[1 + (\lambda_1 + \lambda_2) y_2\Big] e^{-(\lambda_1 + \lambda_2) y_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2) y_2}, \text{ if } \lambda_1 + \lambda_2 = \lambda_{2|1}. \tag{8}$$

Both functions:

$$y \mapsto \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-\lambda_{2|1} y} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-(\lambda_1 + \lambda_2) y},$$

and

$$y \mapsto \Big[1 + (\lambda_1 + \lambda_2) y\Big] e^{-(\lambda_1 + \lambda_2) y},$$

are survival functions (see Appendix 1). Therefore, in both cases ($\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$, or $\neq 0$), the distribution of lifetime Y_2 is a mixture of two lifetime distributions. The first one is exponential with parameter $\lambda_1 + \lambda_2$, whereas the second one is a gamma distribution, $\gamma(2, \lambda_1 + \lambda_2)$ when

$\lambda_{2|1} = \lambda_1 + \lambda_2$. Let us now discuss if the observation of the lifetime only allows for identifying all the parameters including the parameter $\lambda_{2|1}$ providing the time spent in LTC. We have the following Proposition, which is a consequence of the equations (7) and (8):

Proposition 2. *Consider the model with constant intensities and assume that the lifetime Y_2 is the only observable variable.*

i) If $\lambda_1 + \lambda_2 - \lambda_{2|1} \neq 0$ and $\lambda_2 \neq \lambda_{2|1}$,
the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can be identified from the distribution of lifetime Y_2 given in equation (7).

ii) If $\lambda_2 = \lambda_{2|1}$,
the non terminal event has no effect on the mortality intensity. We get $S_2(y_2) = e^{-\lambda_{2|1} y}$. The parameter $\lambda_2 = \lambda_{2|1}$ is identifiable, but not the parameter λ_1.

iii) If $\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$,
the expression of $S_2(y_2)$ is given by equation (8), and the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can all be identified.

Therefore, the possibility of identifying the parameters is based on the jump in intensity upon entry into LTC, that is, the regime switch.

Note also that the following example is a special case of the phase-type distribution models [see Aalen (1995)]. If we allow for an arbitrary number of latent intermediate states, but keep the assumption of constant transition intensities, then the resulting lifetime follows a general phase-type distribution. Because of this homogeneous Markov property, the lifetime distribution has a tractable expression even with an arbitrarily large number of states[10]. The phase-type models are also strongly linked[11] to the static frailty which we will introduce in the next subsection.

3.3 Model with static frailty

Let us now include unobserved static characteristics $F = (F_1, F_2, F_3)$, called frailty, influencing the intensities. The conditional intensities of X_1, X_2 given F, and that of X_3 given X_1 and F,

[10] But the estimation is quite problematic when the number of states is too large.
[11] Yashin et al. (1994) proved that when the number of intermediate states goes to infinity, we get an equivalent model without intermediate state but with a static gamma frailty.

are denoted by $\lambda_1(x_1, F_1)$, $\lambda_2(x_2, F_2)$, $\lambda_3(x_3|x_1, F_3)$, respectively. It is often assumed that the duration variables X_1 and X_2 are independent given F. But in general the frailties F_1, F_2, F_3 are not independent, which implies a dependence between the latent variables X_1, X_2 once the static frailties are integrated out. Such frailties are called **correlated frailty** in the literature [see e.g.Yashin et al. (1995)]. In the special case $F_1 = F_2 = F_3$, we get the so-called **shared frailty** model. Let us now look at the distributions of observable quantities when the static frailties are integrated out. For expository purpose we consider a model with proportional frailties:

$$\begin{cases} \Lambda_1(x_1|F) & = A_1(x_1)F_1, \\ \Lambda_2(x_2|F) & = A_2(x_2)F_2, \\ \Lambda_{2|1}(x_3|x_1, F) & = A_{2|1}(x_3|x_1)F_3, \end{cases}$$

where A_1, A_2 and A_3 are deterministic functions. Let us denote by Ψ the (multivariate) Laplace transform of F:

$$\Psi(u_1, u_2, u_3) := \mathbb{E}[\exp(-u_1F_1 - u_2F_2 - u_3F_3)].$$

When the frailties are nonnegative, this multivariate Laplace transform is at least defined for nonnegative arguments u_1, u_2, u_3.

Then the joint density of X_1, X_2, X_3 becomes:

$$\begin{aligned} g(x_1, x_2, x_3) &= \mathbb{E}[g(x_1, x_2, x_3|F)] \\ &= \mathbb{E}\Big[e^{-A_1(x_1)F_1 - A_2(x_2)F_2 - A_{2|1}(x_3|x_1)F_3} a_1(x_1)F_1 a_2(x_2)F_2 a_{2|1}(x_3|x_1)F_3\Big], \end{aligned}$$

by applying the iterated expectation theorem. This joint density is easily written in terms of the joint Laplace transform of variables F_1, F_2, F_3. Let us denote

$$\tilde{\Psi}(x_1, x_2, x_3, x_1^*) := \Psi\Big(A_1(x_1), A_2(x_2), A_3(x_3|x_1^*)\Big).$$

Thus we have:

$$g(x_1, x_2, x_3) = -\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} \tilde{\Psi}\Big(x_1, x_2, x_3, x_1^*\Big) \mid_{x_1^* = x_1} .$$

Similarly we deduce the expression of the joint density of (Y_1, Y_2) in terms of Laplace trans-

form:

$$f(y_1,y_2) = \mathbb{E}\Big[a_1(y_1)F_1 a_3(y_2-y_1|y_1)F_3 e^{-A_1(y_1)F_1 - A_2(y_1)F_2 - A_{2|1}(y_2-y_1|y_1)F_3}\Big]$$
$$= \frac{\partial^2}{\partial x_1 \partial x_3}\tilde{\Psi}\Big(y_1, y_1, y_2-y_1, y_1\Big), \qquad \text{on domain } \mathcal{D}_1 = \{(y_1,y_2) : y_1 < y_2\},$$

and

$$f(y_1,y_2) = \mathbb{E}\Big[a_2(y_2)F_2 e^{-A_1(y_2)F_1 - A_2(y_2)F_2}\Big]$$
$$= -\frac{\partial}{\partial x_2}\tilde{\Psi}\Big(y_2, y_2, 0, 0\Big), \qquad \text{on domain } \mathcal{D}_0 = \{(y_1,y_2) : y_1 = 0, y_2 > 0\}.$$

Therefore the marginal density function of Y_2 is:

$$f_2(y_2) = \int_0^{y_2} \frac{\partial^2}{\partial x_1 \partial x_3}\tilde{\Psi}\Big(y_1, y_1, y_2-y_1, y_1\Big)dy_1 - \frac{\partial}{\partial x_2}\tilde{\Psi}\Big(y_2, y_2, 0, 0\Big).$$

4 Model with dynamic frailty

In this section we assume that the stochastic frailty F is dynamic, possibly multivariate and time-indexed. Passing from a static to a dynamic frailty is needed for several reasons. First the longevity phenomenon corresponds to a nonstationary underlying dynamic factor. Second, this enlarged version of the model provides much more flexibility to analyze the term structure of mortality and LTC rates.

4.1 The model

Let us consider a generation of individuals indexed by the birth date t_0, that is, the (stochastic) date of death of an individual of this generation is t_0+Y_2. Let us assume that the three intensities given the whole history $\underline{F}$ of the frailty are of the following form:

$$\lambda_1(x_1|\underline{F}, t_0) = \lambda_1(x_1, F_{t_0+x_1}),$$
$$\lambda_1(x_2|\underline{F}, t_0) = \lambda_2(x_2, F_{t_0+x_2}),$$
$$\lambda_{2|1}(x_3|\underline{F}, x_1, t_0) = \lambda_{2|1}(x_3|x_1, F_{t_0+x_1+x_3}).$$

The joint density function of (X_1, X_2, X_3) is (formally):

$$\begin{aligned} g(x_1,x_2,x_3) &= \mathbb{E}[g(x_1,x_2,x_3|F)] \\ &= \mathbb{E}\Big[\lambda_1(x_1,F_{t_0+x_1})\lambda_2(x_2,F_{t_0+x_2})\lambda_{2|1}(x_3|x_1,F_{t_0+x_1+x_3}) \\ &\quad e^{-\int_0^{x_1}\lambda_1(u,F_{t_0+u})du-\int_0^{x_2}\lambda_2(u,F_{t_0+u})du-\int_0^{x_3}\lambda_{2|1}(u|x_1,F_{t_0+u+x_1})du}\Big], \end{aligned}$$

where we assume that the conditional density $g(x_1, x_2, x_3|F)$ is well defined, summable, and the expectation is with respect to the stochastic path of the dynamic frailty. With the dynamic frailty, the density cannot be written in terms of cumulated intensities Λ_1, Λ_2, $\Lambda_{2|1}$, since an intensity such as $\lambda_1(u, F_{t_0+u})$ depends also on age u via the values of the dynamic frailty.

The density of observable variables (Y_1, Y_2) is:

$$\begin{aligned} f(y_1,y_2) &= \mathbb{E}\Big[\lambda_1(y_1,F_{t_0+y_1})\lambda_{2|1}(y_2-y_1,y_1,F_{t_0+y_2})e^{-\int_0^{y_1}\lambda_1(u,F_{t_0+u})du-\int_0^{y_1}\lambda_2(u,F_{t_0+u})du} \\ &\quad e^{-\int_0^{y_2-y_1}\lambda_{2|1}(u,y_1,F_{t_0+u+y_1})du}\Big], \text{ on domain } \mathcal{D}_1=\{(y_1,y_2): y_1<y_2\}, \end{aligned}$$

and

$$f(0,y_2) = \mathbb{E}\Big[\lambda_2(y_2,F_{t_0+y_2})e^{-\int_0^{y_2}\lambda_1(u,F_{t_0+u})du-\int_0^{y_2}\lambda_2(u,F_{t_0+u})du}\Big], \text{ on domain } \mathcal{D}_0=\{(y_1,y_2): y_1=0, y_2>0\}.$$

4.2 Special case: constant baseline intensity functions.

When the baseline intensities are constant, we get:

$$\begin{aligned} \lambda_1(x_1|\underline{F},t_0) &= \lambda_1' F_{t_0+x_1}, \\ \lambda_2(x_2|\underline{F},t_0) &= \lambda_2' F_{t_0+x_2}, \\ \lambda_{2|1}(x_3|\underline{F},x_1,t_0) &= \lambda_{2|1}' F_{t_0+x_1+x_3}, \end{aligned}$$

where λ_1, λ_2 and $\lambda_{2|1}$ are constant vectors. In this case, we can express the joint distribution of the observable variables in terms of the Laplace transform of the cumulated frailty.

Let us define the function $\tilde{\Psi}$ by:

$$\tilde{\Psi}(\lambda,x_1,x_2,x_3,x_1^*,t_0) := \mathbb{E}[\exp\Big(-\int_0^{x_1}\lambda_1' F_{t_0+u}du-\int_0^{x_2}\lambda_2' F_{t_0+u}du-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du\Big)],$$

where $\lambda = (\lambda_1, \lambda_2, \lambda_{2|1})$. Then the expression of the density of (X_1, X_2, X_3) becomes:

$$g(x_1, x_2, x_3) = -\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} \tilde{\Psi}\Big(\lambda, x_1, x_2, x_3, x_1^*, t_0\Big) \mid_{x_1^* = x_1},$$

and the density function for (Y_1, Y_2) becomes:

$$f(y_1, y_2) = \frac{\partial^2}{\partial x_1 \partial x_3} \tilde{\Psi}(\lambda, y_1, y_1, y_2 - y_1, y_1, t_0), \text{ on domain } \mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\},$$

and

$$f(0, y_2) = -\frac{\partial}{\partial x_2} \tilde{\Psi}(\lambda, y_2, y_2, 0, 0, t_0), \text{ on domain } \mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\},$$

When the frailty is static $F_{i,t_0+x_i} = F_i$, the formulas above are exactly the formulas derived in Subsection 3.3.

The expressions of the densities are simplified, when the joint process $F_t = (F_{1,t}, F_{2,t}, F_{3,t})'$ is an affine process [see e.g. Duffie et al. (2003); Gouriéroux et al. (2006)]. For an affine process, the Laplace transform of the factor process conditional on the information on the factors available at date t_0 is such that:

$$\begin{aligned}
&\mathbb{E}[\exp\Big(-\int_0^{x_1} \lambda_1' F_{t_0+u} du - \int_0^{x_2} \lambda_2' F_{t_0+u} du - \int_0^{x_3} \lambda_{2|1}' F_{t_0+x_1^*+u} du\Big) \mid \underline{F_{t_0}}] \\
&= \exp\Big(-a_1(\lambda, x_1, x_2, x_3, x_1^*, t_0) F_{1,t_0} - a_2(\lambda, x_1, x_2, x_3, x_1^*, t_0) F_{2,t_0} \\
&\qquad - a_3(\lambda, x_1, x_2, x_3, x_1^*, t_0) F_{3,t_0} - b(\lambda, x_1, x_2, x_3, x_1^*, t_0)\Big).
\end{aligned}$$

where a_1, a_2, a_3, b are defined in Appendix 3).

5 Statistical inference

We have seen that the parameters of the joint model for longevity and LTC are in general identifiable from lifetime data only. However the lifetimes are also partially observed due to censoring phenomena. In this section we introduce different parametric specifications for models with and without frailties, and derive the likelihood functions in closed form expressions, when the entry into LTC is unobserved and the lifetime is right censored.

In our model, the intensity function of the observed variable Y_2 depends in a non Markovian way on all the past of factor F. Indeed, for a given cohort t_0 and a given age y_2, the intensity at age y_2 depends on the proportion of people in LTC, and therefore depends on the past value of the factor via the mortality intensities $\lambda_2(y|F), \lambda_{2|1}(y|F)$ for all $y < y_2$, as well as the incidence intensity $\lambda_1(y|F)$ for $y < y_2$.

There are several ways to estimate a model with an unobserved factor F. The most straightforward one is to use a nonlinear semiparametric regression as in the Lee-Carter model [see also Mammen et al. (2011) for a more general framework], that is, when the distribution of the unobserved factor F is let unspecified. This methodology is not suitable due to the strong nonlinearity and non Markov features described above. Below we propose two specifications for the dynamic of the latent factor F. In the first case the evolution of F_t is completely deterministic and parametric; in the second one it is stochastic and its distributional dynamics is specified in a parametric way (more precisely as a CIR process). All models will be estimated by maximizing the likelihood estimation in the next section.

5.1 The likelihood functions

5.1.1 Specification of a model without frailty

Let us first consider the basic model introduced in Section 3.1. We denote by $i, i = 1, ..., n$, the individuals and assume that the latent variables $X_{1,i}, X_{2,i}, X_{3,i}$, $i = 1, ..., n$ are independent, with a joint distribution, which depends on the generation only. We denote by $\lambda_1(x_1|t_0)$, $\lambda_2(x_2|t_0)$, $\lambda_{2|1}(x_3|x_1, t_0)$ the intensities for the individuals with the same birth date t_0. Then the individual lifetimes $Y_{2,i}, i = 1, ..., n$ are also independent with a distribution depending on t_0 only. The associated p.d.f. and survival functions are denoted $f_2(y_2; t_0)$ and $S_2(y_2; t_0)$, respectively. Taking into account the right censoring of the lifetimes, the log-likelihood function is:

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \eta^u_{t_0}} \log f_2(y_{2,i}, t_0, \theta) + \sum_{i \in \eta^c_{t_0}} \log S_2(y_{2,i}, t_0, \theta) \Big\}, \tag{9}$$

where $\eta^u_{t_0}$ (respectively $\eta^c_{t_0}$) is the set of uncensored (resp. censored) individuals in generation t_0, $y_{2,i}$ denotes either the observed failure time if the individual is not censored, or the censoring

time, otherwise, and θ denotes the parameter.

Let us now specify the intensities by generation in a parametric way. We first disentangle the effects of the duration and of the current date in the intensities. More precisely we assume:

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = a_1(x_1) + b_1(x_1)F_{t_0+x_1}, \\ \lambda_2(x_2|\underline{F}, t_0) & = a_2(x_2) + b_2(x_2)F_{t_0+x_2}, \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = a_3(x_3|x_1) + b_3(x_3|x_1)F_{t_0+x_1+x_3}, \end{cases} \tag{10}$$

where $a_1(\cdot), a_2(\cdot)$,$a_3(\cdot|\cdot)$, $b_1(\cdot), b_2(\cdot)$,$b_3(\cdot|\cdot)$ are positive (hazard) functions which we will specify in the next section and (F_t) is a deterministic function of time, such as:

$$F_t = \exp(-mt), \tag{11}$$

where $m > 0$. The factor is deterministic and known up to the value of parameter m. The baseline intensities are affine functions of the factor F, this structure is well known in Finance [see e.g. Duffie et al. (2003)] and Insurance [see e.g. Gouriéroux and Monfort (2008)], although most of the time the affine structure is used to model observed processes such as the short interest rate whereas in our case it is written on latent intensity functions. It provides tractable expressions for many quantities when the process F is rendered stochastic and affine (see next subsection).

When t_0 goes to infinity, the intensities converge to $a_1(x_1)$, $a_2(x_2)$ and $a_3(x_3|x_1)$, respectively. Thus these functions can be interpreted as the long term intensities, that are the intensities in a far future. If the functions b_1, b_2, b_3 are nonnegative, the intensities are decreasing functions of t_0. The intensities change with the generation in a deterministic way as in the Lee and Carter (1992) model. They include a trend effect. The main difference with the basic Lee-Carter model is that the intensities in a far future are not equal to zero, that is, the individual does not necessarily become eternal.

To discuss the parametric specification of the intensities in terms of durations, it is convenient

to rewrite the model as an age-cohort model. We get:

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = a_1(x_1) + \Big[b_1(x_1)\exp(-mx_1)\Big]\exp(-mt_0), \\ \lambda_2(x_2|\underline{F}, t_0) & = a_2(x_2) + \Big[b_2(x_2)\exp(-mx_2)\Big]\exp(-mt_0), \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = a_3(x_3|x_1) + \Big[b_3(x_3|x_1)\exp(-mx_1 - mx_3)\Big]\exp(-mt_0), \end{cases}$$

or

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = a_1(x_1) + \tilde{b}_1(x_1)\exp(-mt_0), \\ \lambda_2(x_2|\underline{F}, t_0) & = a_2(x_2) + \tilde{b}_2(x_2)\exp(-mt_0), \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = a_3(x_3|x_1) + \tilde{b}_3(x_3|x_1)\exp(-mt_0), \qquad \text{say.} \end{cases} \tag{12}$$

At this stage, we get a semi-parametric modeling of the intensities where functions $a_j, \tilde{b}_j, j = 1, 2, 3$, are let unspecified. This semi-parametric approach is still similar to the Lee-Carter modeling, except that the trend effect has to be estimated (it depends on m) jointly with other parameters and is not a priori fixed[12].

To transform this semi-parametric model into a pure parametric one, the natural idea is to approximate the functions $a_j, \tilde{b}_j, j = 1, 2, 3$, by means of a well-chosen parametric family, such as splines, polynomials or exponentials. We consider spline approximations in the application of Section 6, since it provides a convenient closed form expression of the log-likelihood function, and facilitates the numerical computation of the maximum likelihood estimates. The explicit expression of the log-likelihood function is given in Appendix 4. In particular, we do not impose, say, collinearity assumptions for the baseline hazards, this allows for (non trivially) different decrease speeds of $\lambda_1, \lambda_2, \lambda_{2|1}$ in t_0.

Note that the functions a_3, $\tilde{b_3}$ are a priori bivariate. In next section's application, besides this general semi-Markov case, we will also consider the following special (Markov) case, where

[12]Indeed, the Lee-Carter model is estimated with a two-stage procedure. In the first stage, the factor process is regarded as observed; the factor value and the other parameters except the trend parameters are estimated. In the second stage, the trend parameter is estimated from the filtered factor process obtained from the previous step.

the intensity of X_3 given X_1 depends only on the current age $x_1 + x_3$:

$$\begin{cases} a_3(x_3|x_1) & = a_3(x_3 + x_1), \\ \tilde{b_3}(x_3|x_1) & = \tilde{b_3}(x_3 + x_1). \end{cases} \tag{13}$$

5.1.2 Specification of a model with dynamic frailty

The above deterministic model is easily extended to introduce a stochastic dynamic frailty. The aim of introducing a stochastic factor is to quantify the uncertainty of both the model fit and the future evolution, which should be taken into account when it comes to pricing insurance contracts written on the LTC risk, and/or the longevity risk.

Instead of using a deterministic trend [equation (11)] for the log-factor $\log F_t$, we assume now a special Cox, Ingeroll, Ross (CIR) dynamics without the mean reversion coefficient:

$$dF_t = -mF_t dt + \sigma\sqrt{F_t} dW_t, \tag{14}$$

where $\sigma > 0$, $m > 0$, and W is a standard Brownian motion. The initial condition is $F_{\min t_0} = 1$, where $\min t_0$ is the birth date of the first cohort. Appendix 5 summarizes some of the basic properties of this CIR process, in particular its existence, its potential hitting time at 0 and its behavior afterwards, as well as its discrete counterpart, which is an autogressive gamma process (ARG).

This CIR model includes the deterministic model of Section 5.1.1 as a limiting case. Indeed, if $\sigma = 0$, then the solution of the differential equation is $F_t = \exp(-mt)$, up to a multiple constant. Thus the CIR model is just introducing some uncertainty around the deterministic exponential model. Therefore, this CIR process still has a nonstationary feature, which reflects the longevity phenomenon.

The choice of a CIR process has several advantages. Firstly, it guarantees the positivity of the intensity functions $\lambda_1, \lambda_2, \lambda_3$ when functions $a_j, \tilde{b}_j, j = 1, 2, 3$ are nonnegative. Secondly, it allows for closed form expressions of the log-likelihood function under an appropriate approximation scheme. More precisely, when we integrate out the unobservable factor in the equations (25) and (26), in general we do not have a closed form for the likelihood function. Nevertheless, in

our example we can use the piecewise constant approximation for the baseline hazard functions and replace the continuous time CIR process by its discrete time counterpart (see Appendix 5). This discrete time process is an autoregressive gamma process, and in particular is still affine, an essential property which enables us to obtain closed form formulas for the Laplace transforms involved in the log-likelihood function.

Another possible candidate process which satisfies the same good properties is a geometric Brownian motion with a negative drift. For the rest of the paper we will use the CIR specification.

Let us now show in detail the second point mentioned above, that is, with an affine frailty process, it is still relatively easy to calculate the log-likelihood function once the continuous time model is discretized. Assume that the intensity functions are piecewise constant: for all x and the integer part of x, $n = \lfloor x \rfloor$, say, we have:

$$\lambda_1(x) = \lambda_1(n), \qquad \lambda_2(x) = \lambda_2(n), \qquad \lambda_{2|1}(x) = \lambda_{2|1}(n).$$

Then we get the link between the intensities in continuous and discrete time:

$$\mathbb{P}[X_1 > n+1 \mid X_1 > n] = 1 - \exp(-\lambda_1(n)),$$

and similarly for other duration variables. We keep all the other assumptions of the semi-Markov model without frailty. Then we get the following expression for the log-likelihood function:

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \eta^u_{t_0}} \log f_2^{\text{disc}}(y_{2,i}, t_0, \theta) + \sum_{i \in \eta^c_{t_0}} \log S_2^{\text{disc}}(y_{2,i}, t_0, \theta) \Big\}, \tag{15}$$

where f_2^{disc} and S_2^{disc} are discrete time approximations of the p.d.f. and the survival function respectively. They are calculated by first writing the corresponding p.d.f. and survival function $f_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)$ and $S_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)$ conditional on factor path F. Then the dynamic frailty F is integrated out:

$$f_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{E}[f_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)], \qquad S_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{E}[S_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)].$$

We give in Appendix 4.2 their exact expressions. In general, these expectations involve a high-

dimensional integral (whose dimension equals the number of observation years) and should be calculated by simulation. In this case, the estimation of the parameter involves simulation-based methods and is numerically cumbersome. However, in our example, we can use the affine property of the CIR process. In particular, we have,

$$\mathbb{E}[e^{-uF_{t+1}}|\underline{F_t}] = \exp(-\frac{e^{-m}u}{1+cu}F_t), \tag{16}$$

where $c = \frac{1-e^{-m}}{2m}\sigma^2$ and u is a nonnegative argument. Then all the expectations can be computed recursively, by using formulas (16) and conditioning with respect to the sigma-algebra $\underline{F_{t_0+y_{2,i}-1}}, \underline{F_{t_0+y_{2,i}-2}}, ..., \underline{F_{t_0}}, ...\underline{F_{\min t_0}}$, where $\min t_0$ is the birth date of the oldest cohort.

6 Application

6.1 The data

The methodology of Section 5 is now applied to a set of observations from the Human Mortality Database (HMD). The HMD was created to provide detailed mortality and population data to researchers, students, policy makers, and others, interested in the history of human longevity. It is maintained by the University of California, Berkeley, and the Max Planck Institute for Demographic Research in Rostock, Germany (see the official website `http://www.mortality.org`).

For instance, for France, the database gives, for each sex and each (birth) cohort t_0 since 1737, the size of the Population-at-Risk and the number of deaths[13] at each integer age, from 0 to $\min(2009 - t_0, 110)$. We use data from age 50 until age 110, and for cohorts starting from 1900. For the oldest cohort (1900), our period of observation begins in 1950 to avoid the period of World War II, and finishes in 2010; for the youngest cohort (1958), the observation begins in 2009 and finishes in 2010, which creates the right censoring effect.

Let us now provide some basic statistics of the French male population. Because of the longevity phenomenon, the distribution of lifetime is shifting to higher ages. This can be il-

[13] As a consequence, the corresponding estimates of the intensity function are available as well.

lustrated by the increase of cross-sectional life expectancy[14]. Because of the right censoring, the computation of the real, cohort-based longitudinal life expectancy involves the choice of a predictive model (and will be calculated in Section 7), while the cross-sectional quantities are model-free. As a consequence, cross-sectional life expectancies do not measure the real expected duration for any cohort. Nevertheless they are still widely used for simplicity. We plot in Figure 2 the mean age at death observed in a same calendar year.

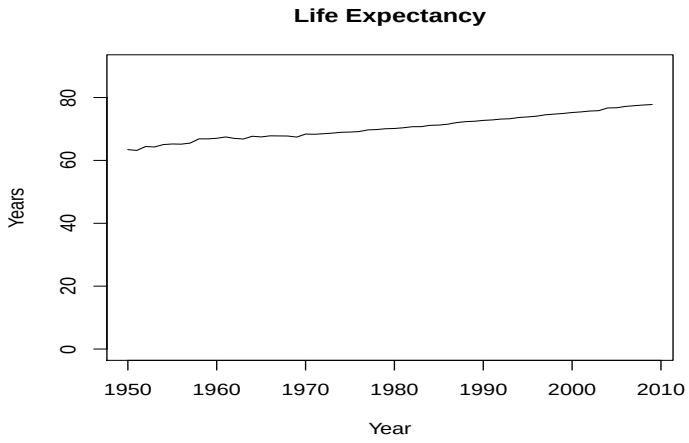


Figure 2: Evolution of the life expectancy at birth for deaths occurring in the same year.

For instance, during the past 40 years, the cross-sectional life expectancy for French males has been steadily rising at a rate of approximately 0.25 years per year. For year 2011, the cross-sectional life expectancy is around 78 years for male, which is about 6 years lower than that of French females', and the latter is also rising at a similar pace.

The longevity phenomenon results in a significant increase of the proportion of higher age population, which will potentially need LTC. Figure 3 shows, for each year, the dependency ratio, that is, the ratio between the size of the old people population (aged 65 or above) and that of the productive population (aged between 15 and 64). This statistics is widely used to measure the pressure on the productive population.

[14]This is also called period life expectancy in demography.

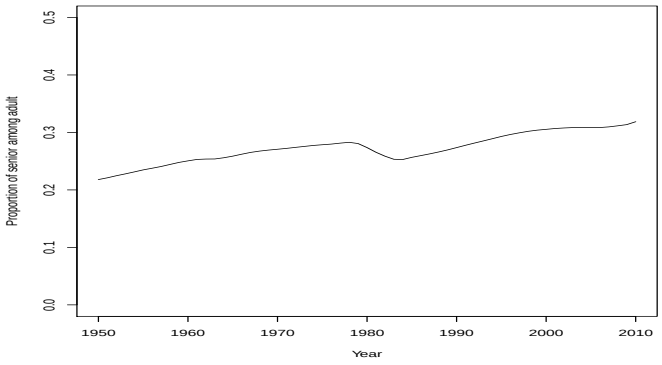


Figure 3: The dependency ratio by year.

The dependency ratio has consistently increased during the last three decades. This is expected to continue as the Baby Boomers reach their retirement ages. This phenomenon spells a huge threat to the sustainability of the social security system, since it may lead to a soaring cost for both the pension plans and the public LTC insurance scheme.

6.2 Estimation of a Markov model without frailty

We estimate the model introduced in Subsection 5.1.1 on the data of French male. We consider the population of males who survive up to age 50. As we suppose an homogeneous population[15], the left censoring is easily taken into account in the log-likelihood function by changing the date origin, which is now 50 instead of age 0.

The model is completed by specifying the functions $a_j, b_j, j = 1, 2, 3$ in the following way:

Assumption 1. [Markov model]

i) The function $a_1(x_1)$ is a linear spline for $x_1 \in]50, 110[$ with two knots at 60 and 70 and is null on the interval $]50, 60]$.

ii) The function $b_1(x_1)$ is such that $b_1(x_1)\exp(-mx_1)$ is a linear spline on $]50, 110]$ with two knots at 60 and 70 and is null on the interval $]50, 60]$.

iii) The function $a_2(x_2)$ is a linear spline for $x_2 \in]50, 110[$ with two knots at 80 and 90.

[15]Here by homogeneous population we mean a population without static frailty. Since we assume that at the beginning of the observation ($y = 50$) nobody is in LTC, there is no heterogeneity linked to the initial health status neither.

iv) The function $b_2(x_2)$ is such that $b_2(x_2)\exp(-mx_2)$ is a linear spline on $]50, 110[$ with two knots at 80 and 90.

v) The function $a_3(x_3|x_1) = a_3(x_3 + x_1)$ is a linear function of the current age $x_3 + x_1$, for $x_3 + x_1 \in]60, 110[$.

vi) The function b_3 is such that $b_3(x_3|x_1)e^{-m(x_3+x_1)}$ is a linear function of $x_3 + x_1$ function for $x_3 + x_1 \in]60, 110[$.

Let us now comment on these assumptions. We specify the baseline hazards under the age-cohort decomposition [see equation (12)]. The linear spline specification is a nonparametric method to approximate the baseline functions. It would be possible to choose more knots, but numerical experiments show that this offers little benefit and may induce the problem of over-parameterization. Empirically we find that other parametric specifications, such as exponential splines, can also fit the model relatively well. We show in Appendix 4 that the linear spline specification in the Assumption 1 provides closed form expressions of the log-likelihood function in some special cases. Assumptions *v*) and *vi*) written on the transition intensity function λ_3 are Markov conditions.

Finally let us now discuss the choice of the age range used in our estimation. We only look at people who survive age 50, since the mortality pattern at younger ages is significantly different from that of higher ages. In general, there are very few people in LTC before age 60; therefore we assume that functions a_1 and b_1 are null between 50 and 60. Our model is written up to age 110, which is approximately the limit age of the human being. It would equally be possible to restrict the observation window to, say, ages 50-90: this would (very slightly) improve the fit of the model, but will prevent us from predicting, say, the residual life expectancy.

The following Lexis diagram illustrates the relationship between the cohort, age and calendar years. The observed part of the history of each cohort is represented by a full 45° line whose left and right boundaries are respectively the age of the beginning and end of the observation (due to either right censoring). As for the censored parts, they are plotted in thick dashed lines. Of all the cohorts, we distinguish two cases:

- Cohorts born before 1900 (for instance cohort 1870 in the plot) are not taken into account

in the estimation. Indeed, their post age 50 history is impacted by the second world war, the aftermath of which marks a strong regime switch in terms of mortality improvement.

- Cohorts after 1900 are right censored, and the censoring age equals $\min(110, 2010 - t_0)$ for a cohort born in t_0. For instance, for cohort 1930, only the data from age 50 to 80 are used.

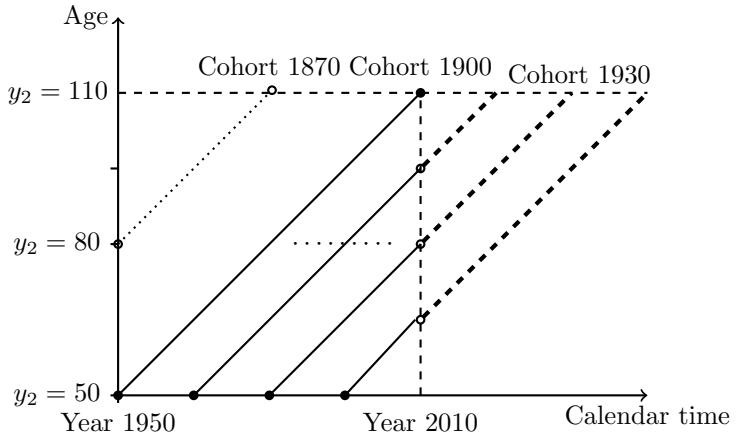


Figure 4: Lexis diagram of cohorts and their observability. The study period ranges from year 1950 to 2010.

The following table gives a summary of the linear splines $a_1, \tilde{b}_1, a_2, \tilde{b}_2, a_3, b_3$ in terms of their value at origin as well as their slopes between different knots.

Table 1: Parameters of the linear spline functions

	value at 50	slope between 50, 60	slope between 60, 70	slope between 70, 80	slope between 80, 90	slope between 90, 110
$a_1(x)$	0	0	w_1	w_2	w_2	w_2
$a_2(x)$	w_3	w_4	w_4	w_4	w_5	w_6
$b_1(x)$	0	0	w_7	w_8	w_8	w_8
$\tilde{b}_2(x)$	w_9	w_{10}	w_{10}	w_{10}	w_{11}	w_{12}
$a_3(x)$	w_{13}	w_{14}	w_{14}	w_{14}	w_{14}	w_{14}
$\tilde{b}_3(x)$	w_{15}	w_{16}	w_{16}	w_{16}	w_{16}	w_{16}

Under Assumption 1, the set of all parameters is $\theta = (w_1, w_2, ..., w_{16}, m)$. The model is estimated by maximum likelihood using the R package *DEoptim*. We report below the value

of the maximum likelihood estimator, and derive the standard deviation of its components by calculating numerically the inverse of the Fisher Information matrix.

Table 2: Estimation of the Markov model without frailty. All parameters are significant at 1% level.

variable	estimator	standard deviation	t-statistics
w_1	0.000398	0.0000158	25.1 ***
w_2	0.001441	0.0000338	42.7 ***
w_3	0.006955	0.0000256	271.3 ***
w_4	0.00024	0.0000051	47.2 ***
w_5	0.005047	0.0001091	46.3 ***
w_6	0.004713	0.0010629	4.4 ***
w_7	0.000285	0.0000225	12.7 ***
w_8	0.002342	0.0000385	60.8 ***
w_9	0.002037	0.0000408	50 ***
w_{10}	0.000784	0.0000071	110.7 ***
w_{11}	0.00259	0.0001255	20.6 ***
w_{12}	0.015769	0.0010415	15.1 ***
w_{13}	0.228108	0.0166392	13.7 ***
w_{14}	0.242871	0.0192654	12.6 ***
w_{15}	0.005123	0.0007004	7.3 ***
w_{16}	0.004978	0.0006665	7.5 ***
m	0.036432	0.0003179	114.5 ***

With the estimated value of parameter θ, we can derive the estimated intensity function for the lifetime variable Y_2 for a given cohort t_0 and a given age y_2 by using the following formula:

$$\lambda(y_2, t_0, \theta) = f_2(y_2, t_0, \theta)/S_2(y_2, t_0, \theta).$$

This estimated intensity is the mortality, when the unobserved heterogeneity of health status is integrated out. Therefore, it is a weighted average of the intensity functions of the two subgroups: autonomous and disabled. Indeed, using the expression of the p.d.f. f_2 and of the

survivor function S_2, we have:

$$\begin{aligned}\lambda(y_2) &= \lambda_{2|1}(y_2)\frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &\quad + \lambda_2(y_2)\frac{e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \lambda_{2|1}(y_2)p(y_2) + \lambda_2(y_2)\Big(1-p(y_2)\Big), \end{aligned} \tag{17}$$

where we have omitted the cohort index t_0, as well as the parameter θ to simplify the notations. The weight $p(y_2)$ is the proportion of people in LTC among the whole Population-at-Risk who survive up to a given age y_2 and is given by:

$$\begin{aligned} p(y_2) &= \frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \frac{\mathbb{P}[0<Y_1<y_2, y_2<Y_2]}{\mathbb{P}[y_2<Y_2]} \\ &= \mathbb{P}[0<Y_1<y_2|Y_2>y_2]. \end{aligned} \tag{18}$$

This probability is called the **prevalence** (at age y_2) and depends also on the cohort t_0.

Then we can compare the values of this intensity function of Y_2 at each integer age, to the historical values of the dataset for the corresponding cohort and age, to look at the goodness of fit of the model in terms of the observed intensity, first by cohort (see Figure 5), then by age (see Figure 6). These figures show a rather good fit for the mortality intensities. Then we plot the latent baseline hazard functions λ_1, λ_2, and $\lambda_{2|1}$ (see Figure 7). The model predicts that the mortality intensity of dependent people is larger than that of autonomous people ($\lambda_{2|1} > \lambda_2$), which is often the case in reality.

We can also plot the evolution of the prevalence function $p(y_2, t_0)$ for different cohorts (see Figure 8).

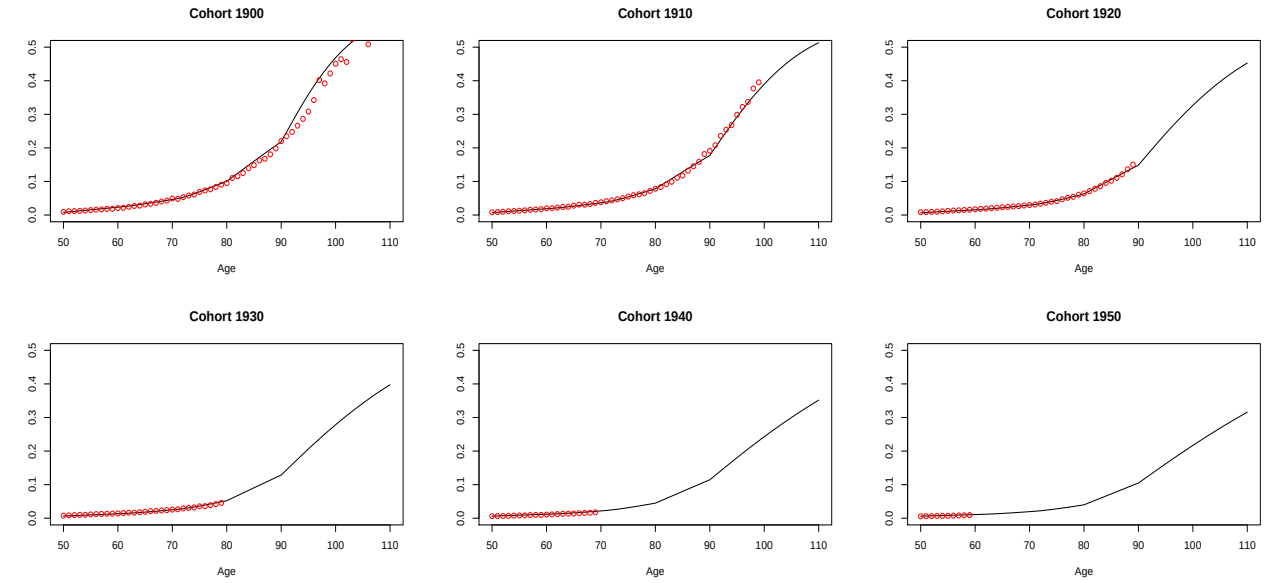


Figure 5: Fit of the observable mortality rates, for six different **cohorts**. Dotted line: historical data. Full line: the model (for both the past and future years). The x coordinate represents the age.

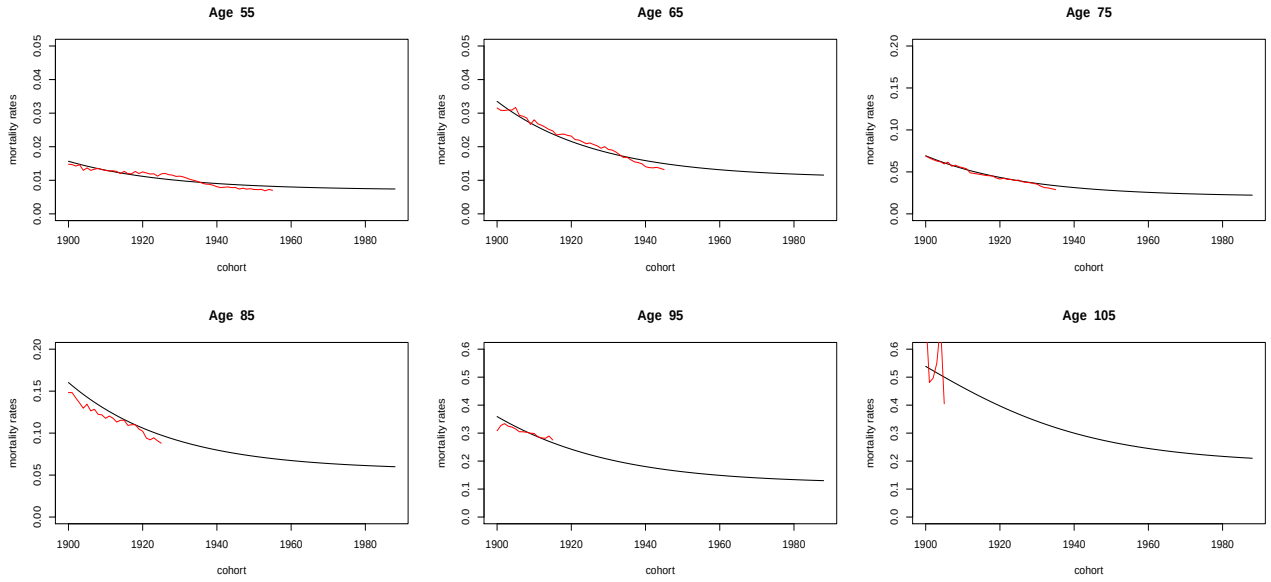


Figure 6: Fit of the observable mortality rates, for nine different **ages**. Dotted line: historical data. Full line: the model (for both the past and future years). The x coordinate represents the cohort.

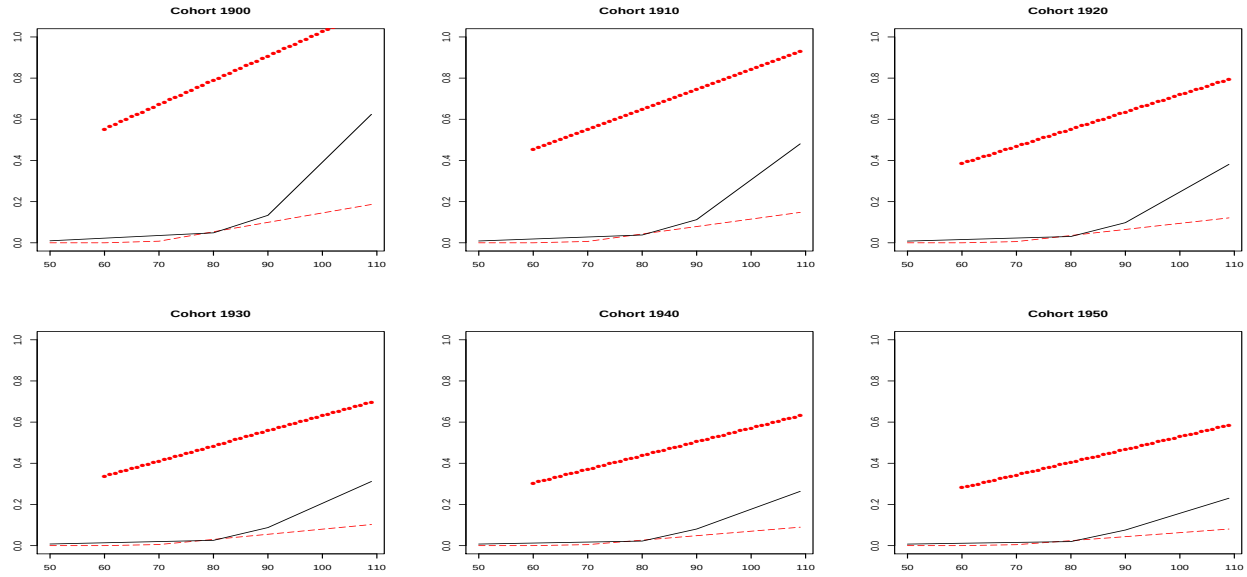


Figure 7: Evolution of the model based baseline hazard functions, respectively $\lambda_1(x)$ (for the intensity of entry, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\lambda_3(x)$ (mortality of person in LTC, dotted line).

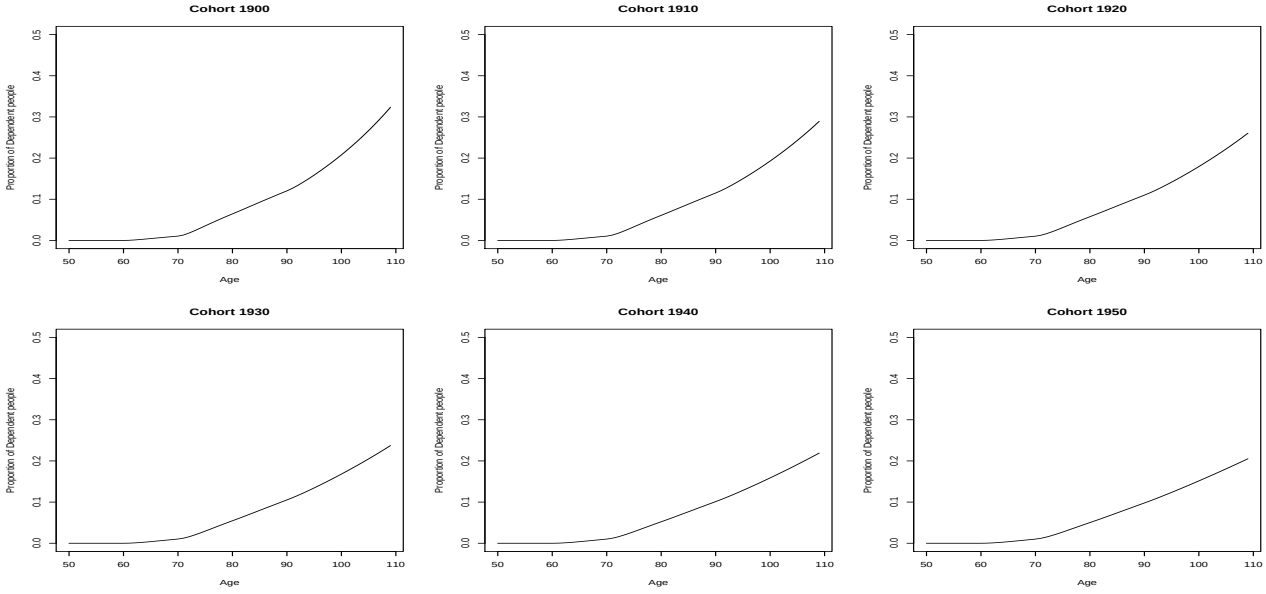


Figure 8: Evolution of the model based proportion of dependent people at a given age for each cohort.

The model predicts that the prevalence begins from 0 at young ages to around 40 percent at

age 110 for the cohort 1900, which corresponds roughly to the observed cross-sectional statistics. Moreover, this prevalence decreases in t_0 for each given age. For instance, the age at which this proportion reaches 10% is equal to respectively 82, 85 and 88 for the following cohorts: 1900, 1920, 1940. This corresponds approximately to an increase of 1.8 months per year for the age of entry into LTC to be compared with the 3-month increase for the cross-sectional expected lifetime. This does not mean a decrease of the cost of LTC for the whole society though, since the probability of survival until high age is increasing at the same time. Section 7 provides more prediction statistics to understand the future evolution of LTC.

6.3 Estimation of a semi-Markov model without frailty

In the previous Markov model, we have assumed that the mortality intensity for a person in LTC depends only on its current age. A more realistic and intuitive is that it depends also on, the age of entry z, or equivalently, on the time elapsed since entry $x - z$. Therefore, in this section, we consider the following semi-Markov assumption:

Assumption 2. [Semi-Markov model]

i) Functions $a_1(x)$, $b_1(x)$, $a_2(x)$ and $b_2(x)$ are specified in the same way as in Assumption 1.

ii) Function $a_3(x-z|z)$ and $b_3(x-z|z)\exp(-mx)$ are linear both in x and z:

$$\begin{cases} a_3(x-z|z) & = c_{0,a} + c_{1,a}(x-z) + \beta_1(z-60), \\ b_3(x-z|z)\exp(-mx) & = c_{0,b} + c_{1,b}(x-z) + \beta_2(z-60). \end{cases}$$

The additional parameters β_1, β_2 characterize the non Markovian feature. For this semi-Markov model, the set of parameters is:

$$\theta = (w_1, w_2, ..., w_{12}, c_{0,a}, c_{1,a}, c_{0,b}, c_{1,b}, \beta_1, \beta_2, m).$$

We report in Table 3 the value of the maximum likelihood estimator.

Table 3: Estimation of the semi-Markov model without frailty, all parameters are significant at 1% level.

w_1	0.000647 (***)
w_2	0.001983 (***)
w_3	0.005249 (***)
w_4	0.000234 (***)
w_5	0.003322 (***)
w_6	0.014902 (***)
w_7	0.000354 (***)
w_8	0.003278 (***)
w_9	0.002738 (***)
w_{10}	0.001389 (***)
w_{11}	0.003532 (***)
w_{12}	0.020574 (***)
$c_{0,a}$	0.234175 (***)
$c_{0,b}$	0.010442 (***)
$c_{1,a}$	0.0037 (***)
$c_{1,b}$	0.006254 (***)
β_1	0.014494 (***)
β_2	0.020769 (***)
m	0.034201 (***)

To illustrate the fit of the model, we compare for different cohorts the value of the estimated intensity $\lambda(y_2, t_0, \theta)$ with the historical mortality intensity function given by the data (Figure 9).

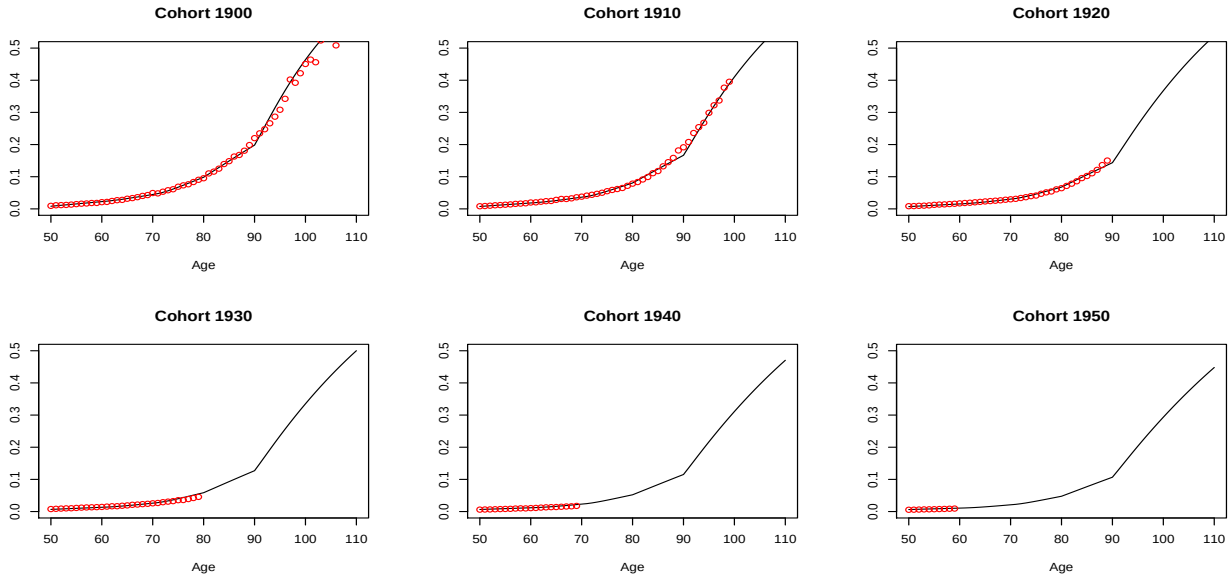


Figure 9: Fit of the observable mortality rates, for six different **cohorts**. Dotted line: historical data. Full line: the model (for both the past and the future years). The x coordinate represents the age.

The semi-Markov model provides also a very good fit. Then we plot (see Figure 10), for different cohorts, the baseline hazard functions λ_1 and λ_2, since they depend only on the age y_2. For the mortality intensity of people in LTC, we plot, for each cohort, the averaged mortality intensity of all the people aged y_2 in LTC $\ddot{\lambda}_{2|1}$, say. It is defined for each cohort by:

$$\ddot{\lambda}_{2|1}(y_2) = \frac{\displaystyle\int_0^{y_2} \lambda_1(z)\lambda_{2|1}(y_2 - z|z)e^{-\Lambda_1(z)-\Lambda_2(z)-\Lambda_{2|1}(y_2-z|z)}dz}{\displaystyle\int_0^{y_2} \lambda_1(z)e^{-\Lambda_1(z)-\Lambda_2(z)-\Lambda_{2|1}(y_2-z|z)}dz}.$$

Then we can check that equations (17) and (18) still hold when we replace $\lambda_{2|1}(y_2)$ by $\ddot{\lambda}_{2|1}(y_2)$. Figure 11 plots, for several cohorts, the evolution of the proportion of people in LTC.

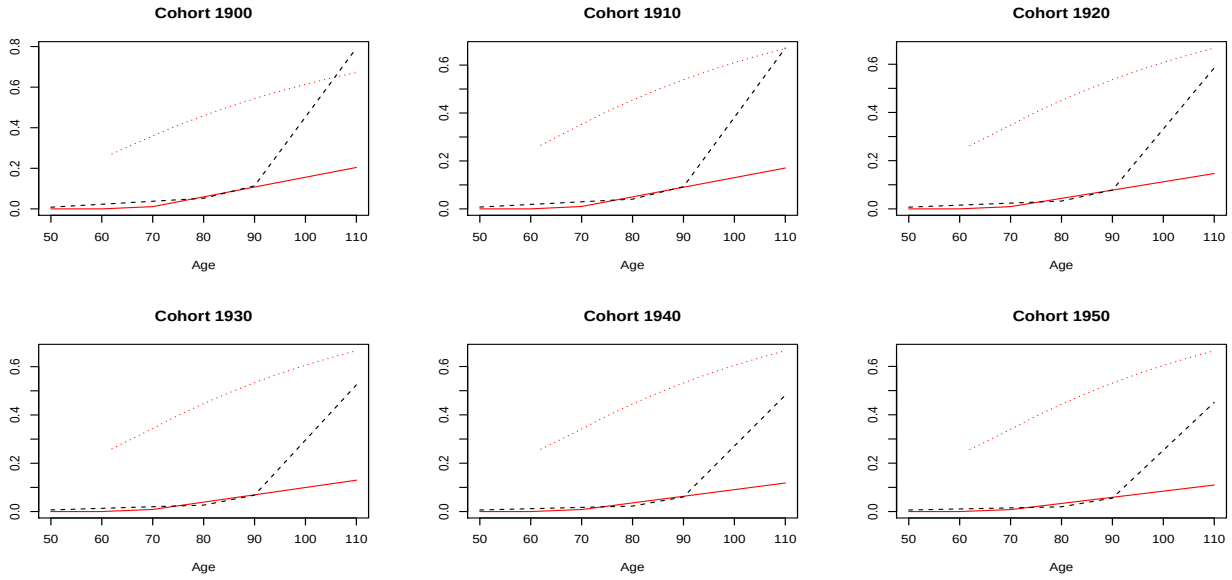


Figure 10: Evolution of the baseline hazard functions, respectively $\lambda_1(x)$ (for the probability of entering into LTC, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\ddot{\lambda}_3$ (mortality of people in LTC, dotted line).

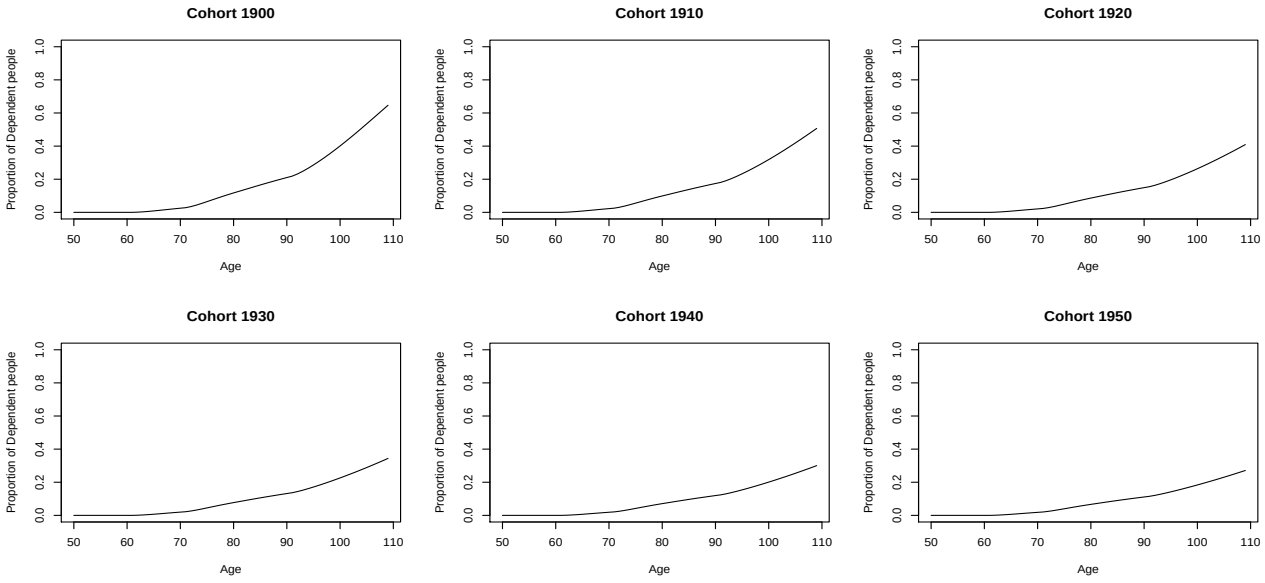


Figure 11: Evolution of model based proportion of people in LTC, for each cohort.

6.4 Estimation of the model with dynamic frailty

Let us now replace, in the previous semi-Markov model, the deterministic dynamic factor by a (common) dynamic frailty, as explained in Subsection 5.1.2. The parameters of the model, including those of the CIR process [equation (14)], m, σ, and those of the baseline hazard functions $a_j,\ b_j,\ j = 1, 2, 3$, are estimated jointly by maximizing the log-likelihood function given by equation (15). Since the model without frailty is the limiting case of the model with dynamic frailty, we can choose the initial value of the numerical algorithm used to optimize the likelihood function as $w = (w^*, 0)$, where w^* is the value of the maximum likelihood estimator of the semi-Markov model without frailty. We report in Table 4 the value of the estimator w.

Table 4: Estimator of the model with dynamic frailty, all parameters are significant at 1% level.

variable	estimator
w_1	0.000693 (***)
w_2	0.002568 (***)
w_3	0.005693 (***)
w_4	0.000168 (***)
w_5	0.003672 (***)
w_6	0.018114 (***)
w_7	0.000425 (***)
w_8	0.002639 (***)
w_9	0.002827 (***)
w_{10}	0.001485 (***)
w_{11}	0.002958 (***)
w_{12}	0.023078 (***)
$c_{0,a}$	0.177399 (***)
$c_{0,b}$	0.009781 (***)
$c_{1,a}$	0.003288 (***)
$c_{1,b}$	0.005822 (***)
β_1	0.004991 (***)
β_2	0.004737 (***)
σ	0.020561 (***)
m	0.034579 (***)

To look at the goodness of fit, we compute, as in Figure 5, the intensity function of the lifetime variable Y_2 for each cohort, when the dynamic frailty is integrated out. More precisely, we first compute the survivor function of the lifetime at different time by integrating out the whole history of the dynamic frailty, and then we calculate the hazard function by computing its

minus log-derivative:

$$h(y_2) = \lim_{h \to 0} \frac{\mathbb{P}[y_2 \leq Y_2 < y_2 + h]}{h} = -\frac{\partial}{\partial y_2} \log \mathbb{E}\Big[S_2(y_2|\theta, F)\Big] = \frac{\mathbb{E}\Big[f_2(y_2|\theta, F)\Big]}{\mathbb{E}\Big[S_2(y_2|\theta, F)\Big]}. \tag{19}$$

It is easily shown that:

$$h_2(y_2) = \mathbb{E}\Big[\frac{f_2(y_2|\theta, F)}{S_2(y_2|\theta, F)} \mid Y_2 > y_2\Big] \neq \mathbb{E}\Big[\frac{f_2(y_2|\theta, F)}{S_2(y_2|\theta, F)}\Big], \tag{20}$$

that is, the conditioning set $Y_2 > y_2$ is not independent from the factor path F.

We display in Figure 12 the hazard function of Y_2 and compare its values to the observed values from the data.

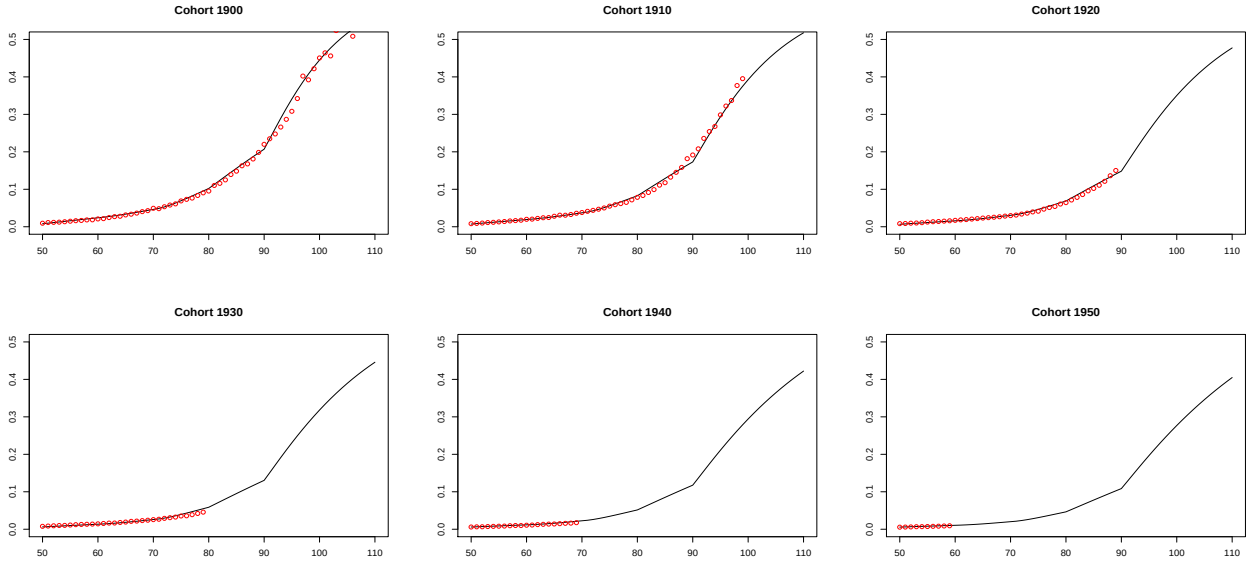


Figure 12: Hazard function of the lifetime variable. Dotted line: historical data. Full line: the model (for both the past and future years).

Once the parameters are estimated, we can infer the path of the unobserved frailty process F. This is useful for several reasons. First, after filtering out the unobserved frailty process, we can check the specification of its dynamics (CIR process), as well as the goodness of fit of the model in terms of observable mortality rates. Second, its values can be used for the predictions of the future mortality and of the LTC transition probability which depend on the frailty process.

There are at least two ways to filter out this unobserved process. First, the observed mortality rates can be written as (nonlinear) functions of the values of the unknown frailty and of parameters. We may invert these equations to obtain the values of the frailty process after replacing the parameter by its maximum likelihood estimate. This methodology is widely used in Finance, [see e.g. Chen and Scott (1993)]. However, since functions $f_2(y_{2,i}, t_0, \theta), S_2(y_{2,i}, t_0, \theta)$ depend on the frailty path in a non Markovian and nonlinear way, and the number of unknown frailty values is quite large when the process covers the period 1951-2009, this approach is numerically cumbersome. For the same reason, nonlinear filtering methods [see e.g. Gagliardini et al. (2012)] are equally forbidden.

The second method is based on simulations of the factor path after substituting the estimated parameters to their true values. More precisely, we can simulate a certain number of paths of the frailty process conditionally on both the estimated value of the parameter and on the observations $Y_{2,i}, i \in \eta^u \cup \eta^c$, that are either the dates of death or the right censoring ages of all individuals. This can be done by Gibbs sampling, similarly as in Duffie et al. (2009). Appendix 6 gives the details of this methodology. In Figure 13, we plot, for each year, the simulated factor mean $\mathbb{E}[F_t|\theta, Y_2]$ conditional on the observed $Y_{2,i}, i \in \eta^u \cup \eta^c$. As a comparison, we also plot the deterministic path $\mathbb{E}[F_t|\theta] = e^{-m(t-1950)}$, where m is the trend parameter of the CIR process (which is itself different from its counterpart in the model without frailty).

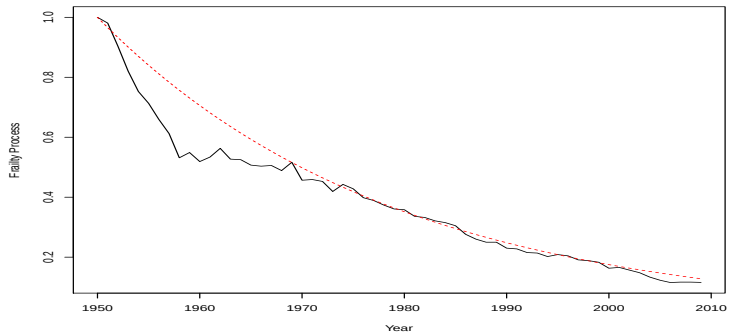


Figure 13: Simulated factor mean (full line) and the deterministic path (dotted line).

As expected, the path features a nonstationary (decreasing) trend, which corresponds to the longevity phenomenon. The filtered factor mean is different from the deterministic path, that

is $\mathbb{E}[F_t|\theta, Y_2] \neq \mathbb{E}[F_t|\theta]$, because of the conditioning on the information Y_2 (see also Equation (20)). Indeed for most t, we observe empirically that $\mathbb{E}[F_t|\theta, Y_2] < \mathbb{E}[F_t|\theta]$; this is quite natural, since the longevity phenomenon favors paths of the CIR process that feature a more pronounced decrease. The filtered paths of the factor can also be used to calculate the conditional intensity of Y_2, that is $\lambda_2(y_2|\theta, F)$, where the factor F is replaced by its filtered values. Not surprisingly, for each of its simulated paths, we get very satisfactory fit to the observed lifetime intensity similarly as in Figure 12. These figures are omitted due to lack of space.

This factor does not have the same influence on the different latent intensities $\lambda_1(x_1, t_0)$, $\lambda_2(x_2, t_0)$, $\lambda_{2|1}(x_3, t_0|x_1)$; indeed these effects depend on the ratios $a_1(x_1)/b_1(x_1)$, $a_2(x_2)/b_2(x_2)$, $a_3(x_3|x_1)/b_3(x_3|x_1)$, who depend themselves on the values of x_1, x_2, x_3. These values can be used to compare the improvement speed of different intensity functions. This was also true for the two previous models without frailty. For instance, for the Markov model without frailty, we see from Figure 7 that the reduction of $\lambda_{2|1}$ at age $x_3 + x_1 = 100$ is less important (about 50 %) than that of λ_2 (about 67 %).

6.5 Comparison of the models with and without frailty

The three models, that are the Markov and semi-Markov model without frailty as well as the semi-Markov model with frailty all provide satisfying fits. The maximized log-likelihoods are respectively: -38700240, -38709452, -38704065, and the corresponding values of the BIC are: 77400764, 77419205 and 77408448. It was expected that the semi-Markov model with frailty has a (very slightly) higher likelihood than the (nested) model without frailty. The comparison between the models with and without dynamic frailty requires more care. Indeed the standard BIC criterion is not necessarily the appropriate measure to compare the performance of the two models in terms of risk prediction and risk management. For instance we have already mentioned that a model with deterministic common factor will likely underestimate the risk. The next section offers a further comparison of them in terms of prediction.

7 Prediction of individual risk

Once the model is estimated from the lifetime data, we can infer for each individual the value of the unobserved variables given the observed ones. We consider below an individual of cohort t_0 at calendar date $t_0 + y_2$. For a model without frailty, it is rather easy to deduce the expressions of the predictive distributions; for a model with dynamic frailty, some expectations, such as the hazard function of the lifetime variable (see Equation (19)), admit explicit forms after integrating out the frailty process, but confidence intervals have in general to be computed by simulation. More precisely, for each simulated past history of process F obtained from the Gibbs sampler (see Subsection 6.4), we simulate its future path and obtain the predictional distributions conditional on the whole factor path, whose formulas are similar as for the model without frailty. This procedure is repeated to obtain the prediction intervals. The prediction problem depends on the observed variables. We have the following situations:

i) If the individual is already dead, we know the value of Y_2, but have to predict the potential date of entry into LTC Y_1 as well as the latent variables X_1, X_2, X_3.

ii) If the individual is still alive and we have no information on his/her health state, except that $Y_2 > y_2$, we have to predict Y_1, X_1, X_2, X_3 and Y_2.

iii) If the individual is in good health, that is, $X_1 > y_2, X_2 > y_2$, we have to predict Y_1, Y_2, X_1, X_2, X_3,

an so on. We first derive explicit prediction formulas for a model without frailty. Then we consider the prediction of future risks in Case *iii*) for the French males, by both the Markov model without frailty and the semi-Markov model with dynamic frailty. These quantities should be calculated for different cohorts, but for expository purpose we omit the cohort index t_0. Besides, since the individual observations are independent, we can perform the computation independently for each individual. For expository purpose we omit the individual index i.

7.1 Case *i*)

Let us first consider the case of predicting unobserved variables, which include the variable Y_1, and the latent variables (X_1, X_2, X_3), conditional on the complete observation of Y_2. The expressions of the predictive distributions are derived below.

Conditional distribution of Y_1 given Y_2. This distribution has a density with respect to the measure $\delta_0 + \lambda_{]0,y_2[}$, where δ_0 is the point mass at 0. This density is:

$$f(Y_1 = 0|Y_2 = y_2) = \frac{f(0, y_2)}{f(0, y_2) + \int_0^{y_2} f(y_1, y_2) dy_1} = \mathbb{P}(Y_1 = 0|Y_2 = y_2), \qquad \text{if } Y_1 = 0,$$

and

$$f(Y_1 = y_1|Y_2 = y_2) = \frac{f(y_1, y_2)}{f(0, y_2) + \int_0^{y_2} f(y_1, y_2) dy_1}, \qquad \text{if } Y_1 \neq 0,$$

where $f(\cdot, \cdot)$ is the joint density function [see equations (3) and (4)].

Conditional distribution of (X_1, X_2, X_3) given Y_2. This conditional distribution has two components, respectively on domain $\mathcal{D}_3 = \{(x_1, x_2, x_3) \in \mathbb{R}_{\geq 0}, x_1 + x_3 = y_2, x_2 \geq y_2\}$, and $\mathcal{D}_4 = \{(x_1, x_2, x_3) \in \mathbb{R}_{\geq 0}, x_2 = y_2, x_1 \geq y_2\}$. Both domains are subsets of a hyperplane. The joint distribution admits a density with respect to the measure $\lambda_{\mathcal{D}_3} + \lambda_{\mathcal{D}_4}$. This density is:

$$g(x_1, x_2, x_3|Y_2 = y_2) = \frac{g(x_1, x_2, y_2 - x_1)}{f_2(y_2)}, \qquad \text{on domain } \mathcal{D}_3,$$

and

$$g(x_1, x_2, x_3|Y_2 = y_2) = \frac{g(x_1, y_2, x_3)}{f_2(y_2)}, \qquad \text{on domain } \mathcal{D}_4.$$

7.2 Case *ii)*

Let us now consider the case when only the information $Y_2 > y_2$ is available.

Conditional distribution of Y_1 given $Y_2 > y_2$. This conditional distribution has three components corresponding to three different cases: $Y_1 = 0$, $Y_1 < y_2$ and $Y_1 > y_2$. Therefore it has a density with respect to the measure $\delta_0 + \lambda_{]0,y_2[}$, and this density is:

$$f(y_1|Y_2 > y_2) = \frac{\lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)}}{\int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}}, \qquad \text{on domain } \{y_1 \in]0, y_2]\},$$

$$f(y_1|Y_2 > y_2) = \frac{\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)}}{\displaystyle\int_0^{y_2} \lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}}, \qquad \text{on domain } \{y_1 \in]y_2, \infty[\},$$

and

$$f(0|Y_2 > y_2) = \frac{\displaystyle\int_{y_2}^{\infty} \lambda_2(t)e^{-\Lambda_1(t)-\Lambda_2(t)}dt}{\displaystyle\int_0^{y_2} \lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}}, \qquad \text{if } Y_1 = 0.$$

It is easily checked that this function $f(\cdot|Y_2 = y_2)$ sums up to 1 and we have:

$$\int_0^{y_2} f(y_1|Y_2 > y_2)dy_1 = p(y_2),$$

that is the prevalence at age y_2 [see Equation (18)].

Conditional distribution of Y_2 given $Y_2 > y_2$. This is already characterized by the hazard function of Y_2 (see e.g. Equation (17) for the Markov model).

The conditional distribution of (X_1, X_2, X_3) given $Y_2 > y_2$ can be obtained similarly and is omitted.

7.3 Case *iii)*

We assume now that the available information set is $X_1 > y, X_2 > y$. A special case is when $y = 50$, any individual enrolled in the study at this age is in good health at the beginning[16], and we are interested in the prediction of Y_1 and Y_2. First, let us compute the probability that a person will enter the LTC during his or her lifetime, given survival in good health up to age y. For each cohort, this probability is given by:

$$\mathbb{P}(Y_1 > 0|X_1 > y, X_2 > y) = \frac{\displaystyle\int_y^{\infty} \lambda_1(x)e^{-\Lambda_1(x)-\Lambda_2(x)}dx}{e^{-\Lambda_1(y)-\Lambda_2(y)}}, \tag{21}$$

This probability is also called the cumulative incidence (at age $Y_2 = \infty$) [see e.g. Kalbfleisch and Prentice (2002)].

[16] Since the transition intensity into LTC is null before age 60.

Other interesting quantities include the residual life expectancy with (potential) LTC.

$$
\begin{aligned}
e_1(y) &= \mathbb{E}[Y_2 - y | X_1 > y, X_2 > y] \\
&= \frac{\displaystyle\int_y^\infty (x_2 - y)\lambda_2(x_2) e^{-\Lambda_1(x_2)-\Lambda_2(x_2)} dx_2}{e^{-\Lambda_1(y)-\Lambda_2(y)}} \\
&\qquad + \frac{\displaystyle\int_y^\infty \Big(x_1 + \int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3 - y\Big) \lambda_1(x_1) e^{-\Lambda_1(x_1)-\Lambda_2(x_1)} dx_1}{e^{-\Lambda_1(y)-\Lambda_2(y)}},
\end{aligned}
$$

as well as the residual life expectancy without LTC (or Healthy Life Years[17]) defined by:

$$
e_2(y) = \mathbb{E}[\min(X_1, X_2) - y | X_1 > y, X_2 > y] = \frac{\displaystyle\int_y^\infty (x-y)\Big(\lambda_1(x) + \lambda_2(x)\Big) e^{-\Lambda_1(x)-\Lambda_2(x)} dx}{e^{-\Lambda_1(y)-\Lambda_2(y)}}.
$$

This term is very popular among sociologists. Indeed, the issue of increasing life expectancy in good health has become a huge concern for policy makers in recent years in developed countries.

Then we can compute the difference of these two terms, which is the expected duration spent in the potential LTC state[18]. It is of particular interest to insurance companies or public social security plan since it impacts the expected cost of an LTC insurance policy. We have:

$$
\begin{aligned}
e_1(y) - e_2(y) &= \mathbb{E}[X_3 \mathbb{1}_{Y_1>0} | X_1 > y, X_2 > y] \\
&= \frac{\displaystyle\int_y^\infty \Big(\int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3 \Big) \lambda_1(x_1) e^{-\Lambda_1(x_1)-\Lambda_2(x_1)} dx_1}{e^{-\Lambda_1(y)-\Lambda_2(y)}}. \qquad (22)
\end{aligned}
$$

In general, the term $\displaystyle\int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3$, that is, the expected residual lifetime upon entry at age x_1, depends on x_1 and cannot be factored out.

Let us now calculate the three quantities above for different values of age y and cohort t_0. For expository purpose, we use the Markov model without frailty and the semi-Markov model with dynamic frailty. For the latter one, 90% confidence bounds are also provided, that are, the 5% and 95% quantiles of the variable $\mathbb{P}[X_1 < X_2 | X_1 > y, X_2 > y, F]$, which is calculated for

[17] This term is introduced by Eurostat of the European Commission. It is calculated in a cross-sectional way while our $e_1(y), e_2(y)$ are longitudinal measures. Another similar terminology is the Disability-Free Life Expectancy (DFLE) [see e.g. Imai and Soneji (2007).

[18] For instance, for a person who never entered LTC during its lifetime, this duration is zero.

each simulated factor path F. Figure 14 displays the evolution of the probability of entering into LTC during its lifetime given survival up to age 50 as a function of the cohort t_0. The value of y is set to 50 years.

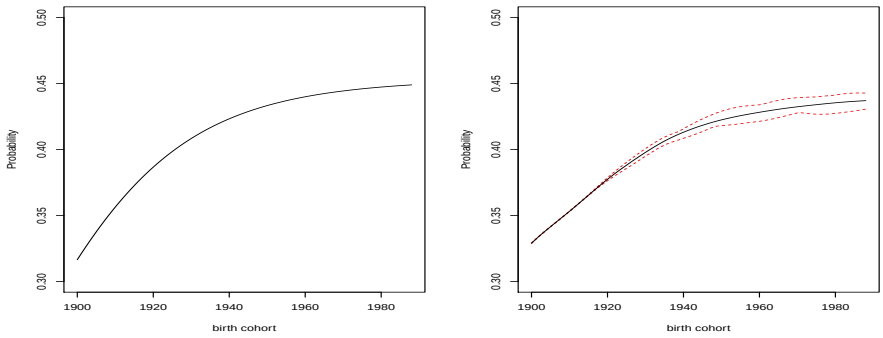


Figure 14: Evolution of the probability of entering into LTC during its lifetime as a function of the cohort. Left panel: the Markov model without frailty, right panel: the semi-Markov model with dynamic frailty; full line: the expected value, that is when frailty is integrated out, dashed lines: the 90% confidence bounds.

The Markov model predicts a slightly bigger probability than the semi-Markov model with dynamic frailty, but in both cases, this probability is increasing in cohort. For instance, the latter predicts that this probability is around 0.33 for the oldest cohort (born in 1900) and will be around 0.43 for the cohort 1980. This is the consequence of two opposite effects: for a given age, on the one hand, the probability of entering into LTC, $\lambda_1(x)$, is decreasing in t_0, and on the other hand, the decrease of both intensities, $\lambda_1(x)$ and $\lambda_2(x)$, means the Population-at-Risk of the LTC is increasing in t_0. The result is also to be compared to Figure 11, where we plot the proportion of people in LTC at any ages, which is decreasing in cohort[19]. For the semi-Markov model with dynamic frailty, the uncertainty, measured by the bandwidth of the confidence interval, is increasing in cohort: for the cohort 1900, the bandwidth is very close to (but not strictly equal to) zero, and becomes quite large for, say, cohort 1980. Indeed, the variation of the past path is considerably smaller than its future path (which follows a standard CIR dynamic given the value of the past history) because of the conditioning with respect to the information of Y_2. For cohort 1900, its history depends only on the filtered past history of the factor F, whereas for cohort 1980 it depends also on the future evolution of the path.

[19]Similarly, the probability of surviving until a given age, either with or without disability, is increasing.

Let us now plot in the same figures the evolution of the residual life expectancies (with and without LTC) for an individual in good health at age 50, for cohorts born from 1900 to 1988.

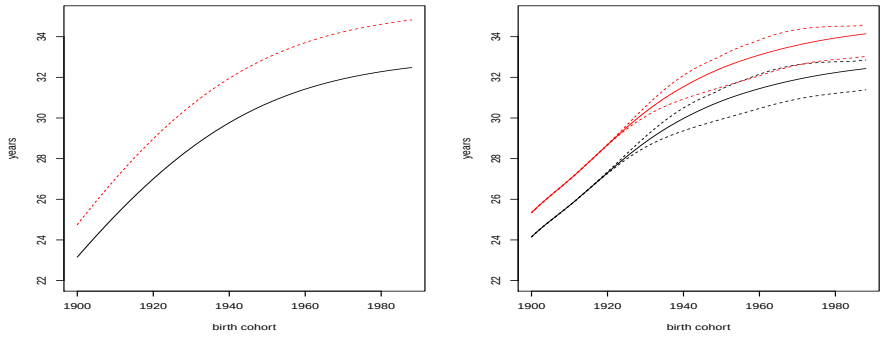


Figure 15: Evolution in t_0 of the residual life expectancy, with potential LTC (dashed line) and without (full line) LTC, at age 50. Left panel: the Markov model, right panel: the semi-Markov model with dynamic frailty; full lines: the expected values, dashed lines: the 90% confidence bounds; the three upper curves are for the life expectancy with potential LTC.

For a French male aged 50 in 2010, his residual life expectancy with potential LTC is around 33 years (with the semi-Markov model). The curve of the residual life expectancy with potential LTC is slightly concave, and increasing with an average improvement rate of around 0.1 year per annum. The difference between the two curves, which directly impacts the expected cost of an LTC insurance contract, is (slowly) increasing.

Finally, let us calculate the uncertainty of the quantities for a finite population:

$$\frac{1}{n}\sum_{i=1}^{n} Y_{2,i,t_0}, \qquad \frac{1}{n}\sum_{i=1}^{n} \min(X_{1,i,t_0}, X_{2,i,t_0}), \tag{23}$$

where Y_{2,i,t_0} [resp. $\min(X_{1,i,t_0}, X_{2,i,t_0})$] is the future death age (resp. age of either losing autonomy or death directly) for the individual i aged 50 in, say, year $t_0 = 2010$. In other terms, these two sums correspond to the average residual lifetime with (resp. without) LTC for a homogeneous portfolio of n individuals. We are interested in calculating their Value-at-Risk $VaR(\alpha)$, where $\alpha \in]0,1[$.

The computation of these VaR can be done by simulation, but this is very time consuming when the size of the portfolio is big. Nevertheless, it can be approximated by using the granularity theory [see e.g. Gagliardini and Gouriéroux (2013)]. For the model without frailty factor, the

distribution of the term (23) is approximately Gaussian by the Central Limit Theorem. For the model with dynamic frailty, conditional on each simulated factor path, (23) is still approximately Gaussian; therefore the unconditional distribution is approximately a mixture of, say, M Gaussian distributions, where M is the number of simulated factor paths. When the size of the portfolio goes to infinity, the asymptotic VaR is the so-called cross-sectional asymptotic (CSA), which corresponds to the un-diversifiable risk linked to the uncertainty of the model. This CSA VaR is easily calculated: for the Markov model without frailty, it is null; for the semi-Markov model with dynamic frailty, it equals the 95% quantile of the conditional expectation $e_1(y|F) = \mathbb{E}[Y_2|X_1 > y, X_2 > y, F]$ (resp. $e_2(y|F) = \mathbb{E}[\min(X_1, X_2)|X_1 > y, X_2 > y, F]$), which are already calculated (see Figure 15).

To illustrate this approach, let us take $n = 10, 100$, and $\alpha = 0.05, 0.95$. The confidence bounds are displayed in Table 5.

Table 5: 90% confidence bounds for the average residual lifetime for a portfolio of n individuals who are 50 years old in 2010.

Mean of Y_2	$n = 10$	$n = 100$	$n = \infty$
The Markov model without frailty	33.12, 33.60	33.29, 33.44	33.36 ± 0
The semi-Markov model with frailty	$31.95, 33.86$	$32.03, 33.85$	$32.18, 33.78$
Mean of $\min(X_1, X_2)$	$n = 10$	$n = 100$	$n = \infty$
The Markov model without frailty	30.98, 31.47	31.15,31.30	31.22 ± 0
The semi-Markov model with frailty	30.45, 32.10	30.47, 32.16	30.59, 32.08

As expected, for both means, the confidence interval is larger for the model with (common) frailty, since it incorporates the uncertainty of the frailty process (both its future and past). In other words, this model takes into account the common risk, whereas the Markov model without frailty assumes it equal to zero. The model with frailty is therefore more reliable from the insurer point of view.

8 Conclusion

In the paper we proposed a new methodology to predict the probabilities of entering in long term care along with the mortality intensities with or without LTC using solely the lifetime data. In

this modeling, the entry into LTC is characterized by a jump in the mortality intensity. In some sense we get a model based implied LTC state. It would be interesting to compare this implied LTC with the different dates of losing Eating, Dressing, ... abilities, when data will become available. This may lead to change the parameter of the Instrumental Activities of Daily Living, as well as the design of long term care products.

We have used a simplified model with a single state of LTC. However, long term care products usually have several levels of coverage, depending on the severity of the disability and a person can move from a lower level to a higher one. For an insurer who has a reliable database of long term care insurance policies, the methodology can be adjusted by taking into account other available information besides the lifetime data. The introduction of such extra states is theoretically possible, but new databases are needed for real data application since for such a multi-state model with several intermediate states, the estimation is computationally cumbersome, when only Y_2 is observed[20].

Appendices

Appendix 1 : Technical lemmas

Lemma 1. *Given $a, b, \alpha, \beta > 0$, let us consider the function g defined by:*

$$g(y) = a\exp(-\alpha y) - b\exp(-\beta y), \qquad y \in]0, \infty[;$$

then g is a survivor function if and only if $a = b + 1$ and $\frac{b}{b+1}\beta < \alpha < \beta$.

Proof. The necessary and sufficient condition for g to be a survivor function is $g(0) = 1$ and g is decreasing. The first condition gives $a = b + 1$. Let us now focus on the second condition. The

[20]Indeed, in a model with an extra state of disability, we should replace the latent variable X_3 by three new latent variables: X_4, the potential duration until moving to the more severe disability state, X_5, the potential duration until dying directly from the current disability state, as well as X_5, the duration spent in the next disability state. Therefore, the log-likelihood function based only on the lifetime data will involve triple integrals, or even integrals in higher dimensions if there are more disability states.

derivative of g is:

$$\frac{d}{dy}g(y) = -\alpha a \exp(-\alpha y) + b\beta \exp(-\beta y).$$

Therefore g is a survivor function if and only if:

$$a = b + 1 \text{ and } \frac{a\alpha}{b\beta} \geq \exp((\alpha - \beta)y), \qquad \forall y > 0,$$

or equivalently $a = b + 1$ and $\frac{b}{b+1}\beta < \alpha < \beta$. □

Lemma 2. *Given $a, b > 0$, let us consider the function g defined by:*

$$g(y) = (1 + by)e^{-ay}, \qquad y \in]0, \infty[;$$

then g is a survivor function if and only if $a \geq b$.

Proof. The condition $g(0) = 1$ is satisfied. Therefore g is a survivor function if and only if:

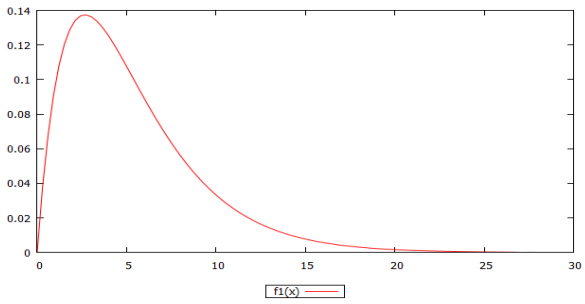
$$\frac{dg}{dy} = -e^{-ay}(aby + a - b) \geq 0, \qquad \forall y > 0,$$

or equivalently $a \geq b$. □

As an illustration, we plot below the corresponding p.d.f. of the survivor function:

$$S(y) := \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-\lambda_{2|1} y} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-(\lambda_1 + \lambda_2) y},$$

where we set the parameters as following: $\lambda_1 = 0.1, \lambda_2 = 0.3, \lambda_{2|1} = 0.35$.



Appendix 2 : Expressions of the survivor function and the p.d.f. of the lifetime variable Y_2

The expression of the p.d.f. of Y_2 is obtained by integrating out the joint density with respect to y_1. We get:

$$\begin{aligned} f_2(y_2) &= \int f_2(y_1, y_2) dy_1 \mathbb{1}_{0<y_1<y_2} + f(0, y_2) \\ &= \int_0^{y_2} \lambda_1(t)\lambda_{2|1}(y_2 - t|t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)} dt + \lambda_2(y_2) e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}. \end{aligned}$$

Let us now check the expression of the survivor function by computing its derivative. We get:

$$\begin{aligned} -\frac{dS_2(y_2)}{dy_2} &= -\lambda_1(y_2) e^{-\Lambda_1(y_2)-\Lambda_2(y_2)} \\ &\quad + \int_0^{y_2} \lambda_1(t)\lambda(y_2 - t|t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda(y_2-t|t)} dt \\ &\quad + \Big[\lambda_1(y_2) + \lambda_2(y_2)\Big] e^{-\Lambda_1(y_2)-\Lambda_2(y_2)} \\ &= f_2(y_2). \end{aligned}$$

Appendix 3 : Laplace transform of the affine factor process

Let us now explain that if the (possibly multivariate) factor process is affine, that is, there exists (possibly multivariate) functions A and B such that for all vector λ, ω and $t, h > 0$, we have:

$$\mathbb{E}\Big[e^{-\int_t^{t+h} \lambda' F_u du} \exp(-\omega' F_{t+h})\Big] = e^{-A'(\lambda,\omega,h)F_t - B'(\lambda,\omega,h)}, \tag{24}$$

where A' (respectively B') denotes the transpose of the vector A (resp. B), then the following Laplace transform has a closed form expression:

$$\begin{aligned}
&\mathbb{E}[\exp\Big(-\int_0^{x_1}\lambda_1' F_{t_0+u}du-\int_0^{x_2}\lambda_2' F_{t_0+u}du-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du\Big)\mid \underline{F_{t_0}}]\\
&=\exp\Big(-l_1(\lambda,x_1,x_2,x_3,x_1^*)F_{1,t_0}-l_2(\lambda,x_1,x_2,x_3,x_1^*)F_{2,t_0}\\
&\qquad -l_3(\lambda,x_1,x_2,x_3,x_1^*)F_{3,t_0}-l_4(\lambda,x_1,x_2,x_3,x_1^*)\Big),
\end{aligned}$$

Without loss of generality, let us consider the case $x_1 \le x_1^* \le x_1^*+x_3 \le x_2$. The following plot gives the illustration of order of these event times on the time axe.

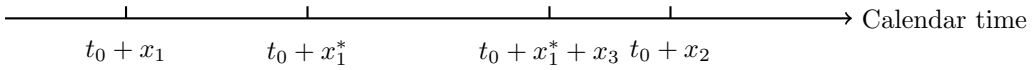


Figure 16: Illustration of the time axe

By conditioning recursively on the sigma-algebras $\underline{F_{t_0+x_1^*+x_3}}$, $\underline{F_{t_0+x_1^*}}$, $\underline{F_{t_0+x_1}}$, we get,

$$\begin{aligned}
&\mathbb{E}[\exp\Big(-\int_0^{x_1}\lambda_1' F_{t_0+u}du-\int_0^{x_2}\lambda_2' F_{t_0+u}du-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du\Big)\mid \underline{F_{t_0}}]\\
&=\mathbb{E}\Big[\mathbb{E}[\exp\Big(-\int_0^{x_1}\lambda_1' F_{t_0+u}du-\int_0^{x_2}\lambda_2' F_{t_0+u}du-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du\Big)\mid \underline{F_{t_0+x_1^*+x_3}}]\mid \underline{F_{t_0}}\Big]\\
&=\mathbb{E}\Big[e^{-\int_0^{x_1}\lambda_1' F_{t_0+u}du-A_1' F_{t_0+x_1^*+x_3}-B_1-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du}\mid \underline{F_{t_0}}\Big]\\
&=\mathbb{E}\Big[\mathbb{E}[e^{-\int_0^{x_1}\lambda_1' F_{t_0+u}du-A_1' F_{t_0+x_1^*+x_3}-B_1-\int_0^{x_3}\lambda_{2|1}' F_{t_0+x_1^*+u}du}\mid \underline{F_{t_0+x_1^*}}]\mid \underline{F_{t_0}}\Big]\\
&=\mathbb{E}\Big[e^{-\int_0^{x_1}\lambda_1' F_{t_0+u}du-A_2' F_{t_0+x_1^*}-B_2}\mid \underline{F_{t_0}}\Big]\\
&=\mathbb{E}\Big[\mathbb{E}[e^{-\int_0^{x_1}\lambda_1' F_{t_0+u}du-A_2' F_{t_0+x_1^*}-B_2}\mid \underline{F_{t_0+x_1}}]\mid \underline{F_{t_0}}\Big]\\
&=\mathbb{E}\Big[e^{-\int_0^{x_1}\lambda_1' F_{t_0+u}du-A_3' F_{t_0+x_1}-B_3}\mid \underline{F_{t_0}}\Big]\\
&=e^{-A_4 F_{t_0}-B_4},
\end{aligned}$$

where coefficients A_j, B_j are defined recursively:

$$
\begin{aligned}
&A_1 = A(\lambda_2, 0, x_2 - x_1^* - x_3), B_1 = B(\lambda_2, 0, x_2 - x_1^* - x_3)\\
&A_2 = A(\lambda_{2|1}, A_1, x_3), B_2 = B(\lambda_{2|1}, A_1, x_3) + B_1\\
&A_3 = A(0, A_2, x_1^* - x_1), B_3 = B(0, A_2, x_1^* - x_1) + B_2\\
&A_4 = A(\lambda_1, A_3, x_1), B_4 = B(\lambda_1, A_3, x_1) + B_3.
\end{aligned}
$$

Appendix 4 : Expressions of the log-likelihood functions

4.1 The model without frailty

In this section we give the detailed expression of the log-likelihood function (9) in the model without frailty. For expository purpose let us start by considering the Markov model. The semi-Markov case is slightly more complicated but is based on the same principle. By using the age-cohort decomposition (12), we have,

$$
\begin{aligned}
&f_2(y_{2,i}, t_0, \theta)\\
&= \Big(a_3(y_{2,i}) + \tilde{b}_3(y_{2,i})F_{t_0}\Big) \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds\\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx\\
&\quad + \Big(a_2(y_{2,i}) + \tilde{b}_2(y_{2,i})F_{t_0}\Big) \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned}
\tag{25}
$$

and

$$
\begin{aligned}
S_2(y_{2,i}, t_0, \theta) = &\int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds\\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx\\
&\quad + \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned}
\tag{26}
$$

where we have changed the time origin ($t = 0$ corresponds to age 50) to account for the left censoring.

Let us now derive the closed form expression of these functions under the linear spline Assumption 1. For any integer value of $y_{2,i}$, consider the interval $[y_{2,i} - 1, y_{2,i}]$. On this interval, functions $a_j, \tilde{b}_j, j = 1, 2, 3$ are all linear in x and the factor $F_{t_0} = e^{-mt_0}$ does not depend on x, we can write $a_1(x)+\tilde{b}_1(x)F_{t_0} = s_1x+i_1$, $a_2(x)+\tilde{b}_2(x)F_{t_0} = s_2x+i_2$, and $a_3(x)+\tilde{b}_3(x)F_{t_0} = s_3x+i_3$, where $s_1, s_2, s_1, i_1, i_2, i_3$ are constants and can be expressed by the coefficients of the linear splines and of $F_{t_0} = \exp(-mt_0)$. Let us now write:

$$\begin{aligned} S_2(y_{2,i}, t_0, \theta) = e^{-\int_0^{y_{2,i}}[a_3(s)+\tilde{b}_3(s)F_{t_0}]ds} \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^t [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\ - \int_0^t [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds + \int_0^t a_3(s) + \tilde{b}_3(s)F_{t_0}ds\Big)dx \\ + \exp(-s_3y_{2,i}^2/2 - i_3y_{2,i}), \end{aligned} \tag{27}$$

where we factored the term $e^{-\int_0^{y_{2,i}}[a_3(s)+\tilde{b}_3(s)F_{t_0}]ds}$ out of the first integral so that the integrand of the remaining integral does not depend on the upper bound $y_{2,i}$. This new integral can be calculated recursively by using the relationship: $\int_0^{y_{2,i}} = \int_0^{y_{2,i}-1} + \int_{y_{2,i}-1}^{y_{2,i}}$. We get:

$$\begin{aligned} &\int_{y_{2,i}-1}^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^t [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds - \int_0^t [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds + \int_0^t a_3(s) + \tilde{b}_3(s)F_{t_0}ds\Big)dx \\ &= e^{-s_3y_{2,i}^2/2-i_3y_{2,i}} \int_{y_{2,i}-1}^{y_{2,i}} (s_1x + i_1) \exp\Big(- (s_1 + s_2 - s_3)(x - y_{2,i} + 1)^2/2 - (i_1 + i_2 - i_3)(x - y_{2,i} + 1)\Big)dx \\ &\quad + \exp(-s_3y_{2,i}^2/2 - i_3y_{2,i}). \end{aligned}$$

The first term is of the form $\int A(x)e^{-B(x)}dx$ with A (respectively B) linear (respectively quadratic). If $s_1 + s_2 - s_3 > 0$, which is often the case, then this term can be expressed in terms of the cumulative distribution function of the normal distribution, therefore $S_2(y_{2,i}, t_0, \theta)$ and $f_2(y_{2,i}, t_0, \theta)$ can be expressed in (quasi) explicit form[21]. For the semi-Markov model, we cannot factor out the term $e^{-\int_0^{y_{2,i}}[a_3(s)+\tilde{b}_3(s|x)F_{t_0}]ds}$ because of the dependence on x. As a consequence the recursive formula is not valid but for fixed $y_{2,i}$, the integrand of the integral in

[21] Indeed, the cumulative distribution function has no closed form but its calculation is rather quick using standard softwares

(25) and (26) is still of the form $\int A(x)e^{-B(x)}dx$, where A and B are piecewise linear (resp. quadratic) therefore the integral can be calculated in explicit form by dividing the integration interval into several subintervals where A and B are linear (resp. quadratic).

4.2 The model with dynamic frailty

The expressions of the p.d.f. and of the survivor function are respectively:

$$
\begin{aligned}
&f_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} = y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} = y_{2,i} \mid F]\Big] \\
&= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}-1} \Big[1 - e^{-a_1(i)-b_1(i)F_{t_0+i}}\Big]\Big[1 - e^{-a_3(y_{2,i}|i)-b_3(y_{2,i}|i)F_{t_0+y_{2,i}}}\Big] \\
&\qquad \exp\Big(-\sum_{j=0}^{i-1}[a_1(j) + b_1(j)F_{t_0+j}] - \sum_{j=0}^{i-1}[a_2(j) + b_2(j)F_{t_0+j}] - \sum_{j=i+1}^{y_{2,i}-1}[a_3(j|i) + b_3(j|i)F_{t_0+j}]\Big)\Bigg] \\
&\quad + \mathbb{E}\Bigg[\Big(1 - e^{-a_2(y_{2,i})-b_2(y_{2,i})F_{t_0+y_{2,i}}}\Big) \exp\Big(- \sum_{i=0}^{y_{2,i}-1}[a_1(i) + b_1(i)F_{t_0+i}] - \sum_{i=0}^{y_{2,i}-1}[a_2(i) + b_2(i)F_{t_0+i}]\Big)\Bigg],
\end{aligned}
\tag{28}
$$

and

$$
\begin{aligned}
&S_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} > y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} > y_{2,i} \mid F]\Big] \\
&= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}} \Big[1 - e^{-a_1(i)-b_1(i)F_{t_0+i}}\Big] \exp\Big(-\sum_{j=0}^{i-1}[a_1(j) + b_1(j)F_{t_0+j}] - \sum_{j=0}^{i-1}[a_2(j) + b_2(j)F_{t_0+j}] \\
&\qquad - \sum_{j=i+1}^{y_{2,i}}[a_3(j|i) + b_3(j|i)F_{t_0+j}]\Big)\Bigg] \\
&\quad + \mathbb{E}\Bigg[\exp\Big(-\sum_{i=0}^{y_{2,i}}[a_1(i) + b_1(i)F_{t_0+i}] - \sum_{i=0}^{y_{2,i}}[a_2(i) + b_2(i)F_{t_0+i}]\Big)\Bigg].
\end{aligned}
\tag{29}
$$

Therefore these terms can be calculated in explicit form by means of the Laplace transform function (as well as its iterated compositions) of (F_t). Again, as for the model without frailty, the computation is faster for the special Markov model than for the general semi-Markov model, since in the case, we can factor out the term $\exp(-\sum_{j=0}^{y_{2,i}}[a_3(j|i) + b_3(j|i)F_{t_0+j}])$ since it does not depend on i and both $f_2(y_2)$ and $S_2(y_2)$ can be calculated recursively.

Appendix 5 : Properties of the latent CIR process

This section is a brief summary of the properties of the CIR process introduced in Subsection 5.1.2.

Lemma 3. *The stochastic differential equation (SDE)* (14) *defines a unique strong solution. With probability 1, this solution attains 0 in a stochastic finite time, and remains at 0 once it reaches it.*

Proof. The SDE verifies the condition that both the drift function and the diffusion function are Liptschitz with at most linear growth, therefore the SDE has a unique strong solution. Let us denote by τ the potential hitting time at 0.
The proof that $\tau < 0$ almost surely involves the knowledge that a CIR process is a time-changed squared Bessel process and can be found in Revuz and Yor (1999).
Once the solution hits 0, it remains at 0 thereafter, as a consequence of the uniqueness of the solution from that date on. □

It is also useful to recall the link between the continuous time CIR process and the discrete time autoregressive gamma process [ARG, see e.g. Gouriéroux and Jasiak (2006)], both of which are affine processes. Let us first give the definition of an ARG process.

Definition 1. We say that a random variable F follows a noncenterred gamma distribution $\tilde{\gamma}(\delta, \beta, c)$ if and only if there exists a Poisson variable with parameter β, $Z \sim \mathcal{P}(\beta)$ such that:

$$F \sim c\gamma(\delta + Z),$$

where γ is the standard gamma distribution.

A process $(F_t, t = 1, 2, ...)$ is an autoregressive gamma process (of order 1, with constant coefficients δ, β and c) if the conditional distribution of F_t given F_{t-1} is $\tilde{\gamma}(\delta, \beta F_{t-1}, c)$.

Lemma 4. *The CIR process defined by* (14) *is such that the discrete time process* $(F_t, t = 1, 2...T)$ *is an autoregressive gamma (ARG) process with coefficients* $\delta = 0$, $c = \sigma^2 \frac{1-e^{-m}}{2m}$, $\beta = e^{-m}/c$. *The ARG process is positive before the hitting time* τ *of the CIR process, and remains null afterwards.*

Proof. See Gouriéroux and Jasiak (2006). □

Intuitively, since $\delta = 0$, and Z is a Poisson variable, there is a nonnull probability that this ARG process hits zero at each date t. But this probability is negligible when the value of the process is large, or when σ is small.

Appendix 6 : Simulating the unobserved paths

The methodology used in this section is similar to that by Duffie et al. (2009). We also refer to Robert and Casella (2004) for a detailed documentation of the MCMC methods. Let us first rewrite the unobserved frailty process $(F_{\underline{t}+1}, ..., F_{\overline{t}=2009})$ by $F = (F_1, F_2, ..., F_T)$, where $T = \overline{t} - \underline{t}$ (by definition, the first value of F is constant: $F_{\underline{t}=1950} = 1$).

6.1 The Gibbs sampler

In order to generate samples of the path $(F_1, ..., F_T)$ conditional both on the value of parameter θ and all the observations Y_2, we can define a Markov chain $M = (M_k) = \Big((F_{1,k}, F_{2,k}, ..., F_{T,k})\Big)$ with values on the T-dimensional domain $(\mathbb{R}_{>0})^T$, where T is the length of the number of values of the dynamic factor process F. If this multivariate chain is stationary with stationary distribution $F \mid \theta, Y_2$, then for large k, M_k will correspond to a drawing from this distribution. Such a chain can be constructed by the multi-step Gibbs sampler. The following theorem explains its principle:

Theorem 1 (Hammersley and Clifford (1968))**.** *Let $(X_1, X_2, ..., X_p)$ be a distribution with joint density function $f(x_1, x_2, ..., x_p)$ then for all $(\xi_1, \xi_2, ..., \xi_p) \in supp(f)$, we have:*

$$f(x_1, ..., x_p) = \prod_{i=1}^{p} \frac{f_{(-j)(x_j | x_1, ...x_{j-1}, \xi_{j+1}, ..., \xi_p)}}{f_{(-j)(\xi_j | x_1, ...x_{j-1}, \xi_{j+1}, ..., \xi_p)}},$$

where $f_{(-j)}(\cdot \mid x_1, ...x_{j-1}, x_{j+1}, ..., x_p)$ is the conditional distribution function of X_j given all other X_i for $i \neq j$. These conditional distributions are called full conditional and the theorem states that they fully determine the joint distribution.

Now we explain how to define the multivariate Markov chain (M_k):

i) Initialize the value $M_1 = (F_{1,1}, F_{2,1}, ..., F_{T,1})$. For instance we set $F_{t,1} = \exp(-m(t-1))$ for all $t = 1, ..., T$, which corresponds to a deterministic factor as in the model without frailty.

ii) Given the k th value of the chain $M_k = (F_{1,k}, F_{2,k}, ..., F_{T,k})$, draw recursively the values $F_{1,k+1}, F_{2,k+1}, ..., F_{T,k+1}$ in the following conditional univariate distributions (often called **full conditionals**):

$$
\begin{aligned}
& F_{1,k+1} \mid F_{2,k}, ..., F_{T,k}, Y_2, \theta \\
& F_{2,k+1} \mid F_{1,k+1}, F_{3,k} ..., F_{T,k}, Y_2, \theta \\
& F_{3,k+1} \mid F_{1,k+1}, F_{2,k+1}, F_{4,k}, ..., F_{T,k}, Y_2, \theta \\
& \dots \\
& F_{T,k+1} \mid F_{1,k+1}, F_{2,k+1}, ..., F_{T-1,k+1}, Y_2, \theta
\end{aligned} \tag{30}
$$

In other words, the chain is updated step by step from its first component to the last one, by drawing at each iteration each time a univariate distribution of the $F_{t,k+1}$ conditional on the parameter θ, the current values of other components of F, as well as the observation Y_2. This approach above cannot be used directly since the conditional distributions do not have forms appropriate for such a drawing[22]. Indeed, only the p.d.f. is easily calculable, up to a multiple constant (see below). But samples from these distributions can be approximated by means of the Metropolis-Hasting algorithm. This is explained in the next subsection.

iii) Store the new value of the chain $M_{k+1} = (F_{1,k+1}, F_{2,k+1}, ..., F_{T,k+1})$ and return to step *ii*).

To generate each of the T distributions given by (30), we employ a Metropolis-Hasting algorithm. Thus to generate the first K values of the Markov chain (M_k), we need to use KT times the Metropolis-Hasting algorithm.

[22]More precisely, the corresponding cumulative distribution function, which should be used when simulating from a given distribution, cannot be calculated.

6.2 The Metropolis-Hasting algorithm

Now we explain the Metropolis-Hasting algorithm we employ in the previous step *ii*). For each t, we should draw from the distribution

$$F_{t,k+1} \mid F_{1,k+1}, ..., F_{t-1,k+1}, F_{t+1,k}, ..., F_{T,k}, Y_2, \theta,$$

or $F_t \mid F_{(-t)}, Y_2, \theta$ for simplicity, where $F_{(-t)}$ denotes the vector $(F_1, F_2, ..., F_{t-1}, F_{t+1}, ..., F_T)$. Let us first explain how to calculate the p.d.f. of this conditional distribution.

Using the same proof as in Duffie et al. (2009), especially the Markov property of F, we have:

$$p(F_t \mid F_{(-t)}, Y_2, \theta) \propto \mathcal{L}(\theta \mid Y_2, F) p(F_t \mid F_{t-1}, \theta) p(F_t \mid F_{t+1}, \theta). \tag{31}$$

The right hand side is the product of two terms. The first is $\mathcal{L}(\theta \mid Y_2, F)$, which is the likelihood of the lifetime data with given values F of the frailty process, that is,

$$\mathcal{L}(\theta \mid Y_2, F) = \exp \sum_{t_0} \Big\{ \sum_{i \in \eta^u_{t_0}} \log f_2(y_{2,i}, t_0, F) + \sum_{i \in \eta^c_{t_0}} \log S_2(y_{2,i}, t_0, F) \Big\},$$

where the expressions of $f_2(y_{2,i}, t_0, F)$ and $S_2(y_{2,i}, t_0, F)$ are respectively the integrand in the right hand side of equations (28) and (29). This can be calculated for given values of θ and F. The second term is $p(F_t \mid F_{t-1}, \theta) p(F_t \mid F_{t+1}, \theta)$, which involves only the one-step transition density of the process (F_t) (given θ). Since it is an autoregressive gamma process, this transition density can be calculated in an exact way[23]. Therefore the second term is equally easy to calculate. Thus the density function given by (31) can be evaluated at each point up to a multiple constant. Instead of drawing directly from this distribution, we can define an auxiliary univariate Markov chain denoted by $(F^{(n)}_{t,k}, n = 1, 2, ...)$, or $F^{(n)}_t$ for simplicity. This chain is also stationary and its stationary distribution is given by (31) Thus we can approximate $F_{t,k+1}$ by $F^{(n)}_t$ for a large value of n. The transition rule of this Markov chain $F^{(n)}_t$ is described as follows:

[23] However, this p.d.f. is numerically rather complicated [see Gouriéroux and Jasiak (2006)] and we can compute instead the transition p.d.f. for the Euler-approximation of the continuous time CIR process $(F_t, t \in \mathbb{R}_{\geq 0})$. More precisely, we have:

$$F_{t+1} \approx (1 - m) F_t + \sigma \sqrt{F_t} \mathcal{N},$$

where $\mathcal{N}$ is a standard normal variable.

1. Initialize the chain by setting $F_t^{(1)} = 1$.

2. For $n = 2, 3, ...$, draw a candidate from a proposal distribution, for instance, we can choose the log-normal distribution[24]:

$$f \sim F_t^{(n-1)} \mathcal{N}(0, \sigma),$$

where the standard deviation of the proposal density is chosen arbitrarily, say, $\sigma_p = 0.01$.

3. Compute

$$\alpha = \frac{p(F_t = f \mid F_{(-t)}, Y_2, \theta)}{p(F_t = F_t^{(n-1)} \mid F_{(-t)}, Y_2, \theta)} \tag{32}$$

both the numerator and the denominator can be calculated by equation (31).

4. Draw a uniform variable $u \sim U([0, 1])$ and set the $n-$th value F_t^n by the following rule: [25]

$$F_t^{(n)} = \begin{cases} f, & \text{if } u < \alpha \\ F_t^{(n-1)}, & \text{otherwise} \end{cases}$$

To ensure convergence of this univariate Markov chain to its stationary distribution (31), we take, say, the 300 th value of the chain as a sample from this distribution, which is used in step *ii*) of the Gibbs sampling algorithm.

References

Aalen, O. O. (1995). Phase Type Distributions in Survival Analysis. *Scandinavian Journal of Statistics*, 22(4):447–463.

Abbring, J. H. and Van den Berg, G. J. (2003). The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica*, 71(5):1491–1517.

Amemiya, T. (1984). Tobit Models: a Survey. *Journal of Econometrics*, 24(1):3–61.

[24] This choice is mainly motivated by simplicity reasons. Indeed it allows for a symmetric conditional density since $p(f|F_t^{(n-1)}) = p(F_t^{(n-1)}|f)$ so that there is no need to calculate the ratio $\frac{p(f|F_t^{(n-1)})}{p(F_t^{(n-1)}|f)}$. Besides, we should use a positive distribution, (since the factor F is nonnegative), which is the case for the log-normal distribution.

[25] The equation B4 in Duffie et al. (2009)[Appendix C] is not correct since their α does not depend on the factor $p(F_t \mid F_{t-1}, \theta)p(F_t \mid F_{t+1}, \theta)$.

Andersen, P. K. and Keiding, N. (2012). Interpretability and Importance of Functionals in Competing Risks and Multistate Models. *Statistics in Medicine*, 31(11-12):1074–1088.

Chen, R. and Scott, L. (1993). Maximum Likelihood Estimation of a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income*, (3):14–31.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty Correlated Default. *The Journal of Finance*, 64(5):2089–2123.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine Processes and Applications in Finance. *Annals of Applied Probability*, 13(3):984–1053.

Elbers, C. and Ridder, G. (1982). True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *The Review of Economic Studies*, 49(3):403–409.

Fine, J., Jiang, H., and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, 88(4):907–919.

Freund, J. E. (1961). A Bivariate Extension of the Exponential Distribution. *Journal of the American Statistical Association*, 56(296):971–977.

Gagliardini, P. and Gouriéroux, C. (2013). *Granularity Adjustment*. forthcoming, Cambridge University Press.

Gagliardini, P., Gouriéroux, C., and Monfort, A. (2012). Microinformation, Nonlinear Filtering, and Granularity. *Journal of Financial Econometrics*, 10(1):1–53.

Gouriéroux, C. and Jasiak, J. (2006). Autoregressive Gamma Processes. *Journal of Forecasting*, 25(2):129–152.

Gouriéroux, C. and Lu, Y. (2013). Love and Death: A Freund Model with Frailty. *CREST DP*.

Gouriéroux, C. and Monfort, A. (2008). Quadratic Stochastic Intensity and Prospective Mortality Tables. *Insurance: Mathematics and Economics*, 43(1):174–184.

Gouriéroux, C., Monfort, A., and Polimenis, V. (2006). Affine Models for Credit Risk Analysis. *Journal of Financial Econometrics*, 4(3):494–530.

Hammersley, J. M. and Clifford, P. (1968). *Markov Fields on Finite Graphs and Lattices*. Oxford University DP.

Han, A. and Hausman, J. A. (1990). Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics*, 5(1):1–28.

Hougaard, P. (1999). Multi-State Models: a Review. *Lifetime Data Analysis*, 5(3):239–264.

Hsieh, J.-J., Wang, W., and Adam Ding, A. (2008). Regression Analysis Based on Semicompeting Risks Data. *Journal of the Royal Statistical Society: Series B*, 70(1):3–20.

Imai, K. and Soneji, S. (2007). On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and its Extension. *Journal of the American Statistical Association*, 102(480):1199–1211.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data (2nd Edition)*. New York: Wiley.

Keiding, N. (1991). Age-Specific Incidence and Prevalence: a Statistical Perspective. *Journal of the Royal Statistical Society: Series A*, pages 371–412.

Kessler, D. (2008). The Long-Term Care Insurance Market. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 33(1):33–40.

Lafortune, G. and Balestat, G. (2007). Trends in Severe Disability Among Elderly People: Assessing the Evidence in 12 OECD Countries and the Future Implications. Technical report, OECD Health Working Paper.

Lee, R. D. and Carter, L. R. (1992). Modeling and Forecasting US Mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Levantesi, S. and Menzietti, M. (2012). Managing Longevity and Disability Risks in Life Annuities with Long Term Care. *Insurance: Mathematics and Economics*, 50(3):391–401.

Mammen, E., Nielsen, J., and Fitzenberger, B. (2011). Generalized Linear Time Series Regression. *Biometrika*, 98(4):1007–1014.

Manton, K. G., Stallard, E., and Corder, L. (1998). Economic Effects of Reducing Disability. *The American Economic Review*, 88(2):101–105.

Oeppen, J. and Vaupel, J. W. (2002). Broken Limits to Life Expectancy. *Science*, 296(5570):1029–1031.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., and Breslow, N. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*, pages 541–554.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, volume 293. Springer.

Rice, D. P. (1989). Health and Long-Term Care for the Aged. *The American Economic Review*, 79(2):343–348.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.

Stallard, E. (2011). Estimates of the Incidence, Prevalence, Duration, Intensity, and Cost of Chronic Disability Among the US Elderly. *North American Actuarial Journal*, 15(1):32–58.

Tosch, T. J. and Holmes, P. T. (1980). A Bivariate Failure Model. *Journal of the American Statistical Association*, 75(370):415–417.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16(3):439–454.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical Analysis of Illness–Death Processes and Semicompeting Risks Data. *Biometrics*, 66(3):716–725.

Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1994). A Duality in Aging: the Equivalence of Mortality Models Based on Radically Different Concepts. *Mechanisms of Ageing and Development*, 74(1):1–14.

Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995). Correlated Individual Frailty: an Advantageous Approach to Survival Analysis of Bivariate Data. *Mathematical Population Studies*, 5(2):145–159.

Zheng, M. and Klein, J. (1995). Estimates of Marginal Survival for Dependent Competing Risks Based on an Assumed Copula. *Biometrika*, 82(1):127–138.