Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2013-16

Long Term Care and Longevity

C. GOURIÉROUX[1]
Y. LU[2]

October 2013

1 CREST and University of Toronto. Email : christian.gourieroux@ensae.fr
2 SCOR and CREST. Email : yang.lu@normalesup.org

Long-Term Care and Longevity

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

(First version January 2014, Revised, July 2014)

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 loss of autonomy. Because of the significant costs of long-term care (LTC) facilities, it is important to analyze the time spent in LTC state, as well as the probability of entering into this state during its lifetime, and how they evolve jointly with longevity across the different cohorts. Our paper considers such questions, when lifetime data are available, but LTC data are either unavailable, or available on too short periods, or too aggregated, or unreliable, as it is frequently the case.

We specify joint structural models of LTC, mortality, and longevity, and explain why parameters of these models are identifiable from only the lifetime data under reasonable assumptions. More precisely, we model the potential entry into LTC as a latent state, which creates a dynamic unobserved heterogeneity in the population when only the lifetime is observed. The methodology is applied to the cohort mortality data of French males, first with a deterministic trend and then with a dynamic and stochastic common latent factor. Prediction formulas for the hypothetical date of entry into LTC or the time spent in this state are then provided and illustrated using the same dataset.

Keywords: Longevity, Long-Term Care (LTC), Semi-Competing Risks, Treatment effect, Unobserved Heterogeneity, Dynamic Frailty, Partial Observability, Identification.

Acknowledgements: The first author gratefully acknowledges financial support of the Global Risk Institute and of the Chair ACPR/Risk Foundation: Regulation and Systemic Risk. We thank G. Horny, J. Kalbfleisch, N. Keiding, B. Wouterse, as well as two anonymous referees for useful comments. The views are those of the authors. They do not necessarily reflect the views of the Prudential Supervision and Resolution Authority (ACPR) or of SCOR SE.

[1]CREST and University of Toronto. Email: gouriero@ensae.fr
[2]CREST and SCOR. Corresponding Author. 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 an increase of 3 months per annum of the life expectancy [see e.g. Oeppen and Vaupel (2002)]. This increase is accompanied by an increase of the number of old people who potentially need long-term care (LTC henceforth)[1], 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*[2]. 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). This entry into LTC state is in general irreversible and is accompanied by a huge increase of mortality intensity. Because of the significant costs of LTC facilities, it is important to analyze this probability of entry, the time spent in this state as well as how they evolve with longevity. Are they almost independent of the longevity feature or do they increase at a similar rate? Our paper answers these questions, when the lifetime data are available, but the LTC data are either unavailable, or available on too short periods, or weakly reliable.

We introduce in this paper joint models of LTC and mortality, based on the intensity of entry into LTC state and on the mortality intensities. The model disentangles the mortality intensities according to the time spent in LTC state. Moreover we assume that these intensities depend on an unobservable dynamic factor (or dynamic frailty) with nonstationary features, able to capture the longevity phenomenon and its potential impact on both mortality and LTC. This longevity factor can be assumed deterministic, or stochastic.

Such a joint model would be simple to estimate if individual data on both mortality and LTC were available [see e.g. Levantesi and Menzietti (2012), Majer et al. (2013)]. However data on LTC are often missing or not very reliable when they exist. Indeed, there does not even exist a universal definition of the LTC state. In the literature, the very terminology is often confounded[3] with "losing autonomy", "disability", "morbidity" or "nursing/home care" and differs by both country and insurance company; further more, it is subject to changes across time. In the US, insurers consider six limitations of Activities of Daily Livings, that are Eating, Dressing, Walking, Bathing, Toileting, and Maintaining Continence, respectively, while their

[1] Also called nursing care in the literature.

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

[3] For instance, Levantesi and Menzietti (2012) propose to price private LTC contracts using national disability benefit data.

European peers, use only four of them called Instrumental Activities of Daily Living (IADL) [see e.g. Rice (1989), 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. An OECD disability indicator even include extra criteria such as hearing and reading small letters [see McWhinnie (1980)]; French public databases based on different population samples show different trends of the LTC/disability prevalence[4] [see Lafortune and Balestat (2007)]. Finally, current data often measure the actual LTC use, instead of the need of LTC. There are various reasons for the two to differ in practice, such as administrative delay[5], the lack of self-diagnosis capacity of the disabled, or budget constraint, or even the incentive of false claim[6].

Moreover, even when data exist, they often lack accuracy. Indeed, collecting LTC data is a much more demanding task than collecting mortality data since it requires the knowledge of the entire history of each individual, especially the time(s) at which an IADL is lost, identified by accredited physicians. Most of the time, available public data of the national population only exist for a few years when there is either a census, or a sample population survey[7] with a large time spell between neighbouring surveys; their quality are quite limited because of the voluntary nature of the survey responses and the fact that surveys conducted in different years do not necessarily concern the same individuals. 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 evolution of the LTC risk. Indeed, by using such a cross-sectional database one will in general ignore the evolution in cohort of the different transition probabilities at given ages [see Keiding (1991) for a discussion on the limits of this stationary approach]; this is unrealistic and dangerous given the potentially large impact of the longevity on both LTC and mortality risks. This uncertainty on the future evolution and its poor understanding is a serious obstacle to the further development

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

[5]For instance, it is common practice for insurance companies to acknowledge the entry into LTC of a policy-holder (and begin periodic benefit payment) only six months after the effective entry, to make sure that the entry is really permanent.

[6]For instance, Dienst (1972) states that during past severe economic crisis, the number of people declaring disabled increased. This effect is produced mainly by people who have been medically disabled long time ago and in addition by people with relatively minor medical problems who would not consider themselves disabled in good times, but who in both instances are induced to claim insurance benefits only in case of a crisis.

[7]This is for instance the case for the survey "Handicaps-Incapacité-Dépendances" in France (literally the Disability-Incapacity-Long-Term Care Survey, this survey has been conducted in 1998/1999 and then in 2008/2009.), as well as the National Long-Term Care Survey (NLTCS) database in the US (which is based on surveys conducted in 1982,1984,1989,1994,1999, 2004 on a representative sample of the US population, see its official website `http://www.nltcs.aas.duke.edu/`). These two countries are also by far the two largest markets for private LTC insurance.

of the private LTC insurance market in many countries, in a period when the sustainability of the Welfare States is more and more questioned and the public's appetite for private LTC insurance is steadily increasing.

Our paper develops a methodology to estimate this joint model of risks using only the mortality data. Rather than relying on data with an *ad hoc* definition of the LTC state, we consider the autonomy state as a latent state variable and the entry into LTC is characterized by an unobservable mortality jump[8]. The assumption that we can capture an individual's aging history by such a model with two regimes, and interpret one of them as the entry into LTC is not just for identification convenience. Indeed, physiologically speaking, the entry into LTC is not an independent event, but is often caused by random events such as the onset of a disease or an accident[9]. Not all such events result in LTC, which becomes necessary only when there is a significant deterioration of the health, accompanied by a major rise of the mortality intensity. Thus it seems appropriate to regard the entry of LTC as the reaching of a critical stage of one's aging process, and the drastic change of the mortality rate spells a change of regime upon entry. This change of regime is by nature latent, and is only imperfectly captured by various existing data on LTC. Furthermore, given the definition and data quality issues discussed above, using an *ad hoc* definition of LTC state to define a regime split might not be optimal. Therefore, it is useful to introduce this latent, yet "canonical" LTC state which we will "filter" out of the lifetime data.

Due to the higher mortality for people in LTC, when the mortality is analyzed using only lifetime data, the autonomy state at a given age[10] is a time-dependent unobserved heterogeneity. Therefore there is a spurious duration dependence as in a population with static unobserved heterogeneity, or static frailty [see e.g. 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. Under reasonable assumptions, the possibility to identify the characteristics of LTC from the mortality data is due to the jumps in mortality intensity arising when entering into LTC and to the assumed effects of the unobserved longevity factor on both mortality and LTC across different cohorts. Thus, such a model allows us to predict jointly the

[8]The idea of introducing latent state variables is recently also proposed by Wouterse et al. (2013). With observations of a large number of health indicators including the LTC status, they construct a latent state variable as a synthetic measure of the individual's health status. However, in their framework, LTC is observable and their methodology does not allow for an analysis of the evolution of various risks across different cohorts.

[9]For instance, Kessler (2008) claims that more than 70 % of LTC entries is caused by chronic diseases such as cancer and dementia, others being triggered by events such as accidents or mental diseases.

[10]Either autonomous, or in LTC.

future evolution of the LTC entry probabilities and the mortality intensities.

The paper is organized as follows. In Section 2, we introduce a joint modeling of LTC and mortality risks. This modeling is used in Section 3 to derive the joint distribution of the lifetime and of the date of entry into LTC. 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. In Section 3 we consider a basic model with constant intensities and discuss its identification. Section 4 introduces semi-parametric specifications for the intensities and the frailty dynamics, discuss the way of introducing a nonstationary longevity generation effect, solve the identification issues, and derive the form of the log-likelihood function when the lifetimes are observed with right censoring. The models are estimated for the French male population in Section 5. We first consider a model with deterministic factor 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 6 we implement the model for prediction purpose. Section 7 concludes. Proofs and other technical details are gathered in Appendices.

2 Structural versus reduced form approach

Let us consider a situation where an individual can either experience first a **non terminal event** and then fail, or 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 semi-competing risks[11] 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. Thus there is an asymmetry between both types of events.

We first introduce a structural approach with latent variables corresponding to the times elapsed up to the potential events and describe how the ideally observable variables depend on the latent duration variables. Then we derive an alternative methodology in terms of intensities.

In the literature, most multivariate survival models are written in continuous time. The main

[11] In the microeconometric literature, the effect of the non terminal event on the terminal event is also called "treatment effect" [see e.g. Abbring and Van den Berg (2003b)], even if the exogenous entry in LTC cannot really be interpreted as a treatment as in other types of economic applications.

reason is that in the continuous time intensity-based setting, the probability of observing tied events is naturally null. In our example, we would like to avoid the simultaneous arrival of both the non terminal and the terminal event. Thus we follow the continuous time approach, 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 4 on, when the cohort effect is introduced, we will more frequently use the term "age" for the elapsed duration, that is, the age of an individual since its birth. To describe the period effect, we use the term "calendar time" and we have the following relationship between the three time measures:

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

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 failure time and Y_1^* is the potential time of entering into LTC. Therefore, the variable Y_1^* is latent since 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 often modeled via a survivor copula C [see e.g. Fine et al. (2001) and Hsieh et al. (2008)], that is,

$$\mathbb{P}(Y_1^* > y_1, Y_2 > y_2) = C(S_1(y_1), S_2(y_2)), \tag{1}$$

where C is assumed to belong to some specific parametric families, e.g. Archimedean copulas or other factor copulas and S_1, S_2 denote the marginal survivor functions of Y_1^* and Y_2, respectively. This bivariate copula approach is partly borrowed from the literature on 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. But such a direct modeling is not flexible enough to capture the peculiarities of semi-competing risks data. First, they are not adapted to characterize the "regime switching" nature that an individual may experience. Intuitively, if the individual enters into the LTC during its lifetime, then his residual lifetime distribution will be very different from the case when he never experiences the LTC. Therefore, using solely one variable Y_2 to model the lifetime is probably not enough. Besides, the idea

behind equation (1) is that instead of being latent, the variable Y_1^* is treated as observable (and is only censored when $Y_2 < Y_1^*$ instead of being nonexistant). This confusion explains also 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)]. We consider below an alternative approach with an extra latent variable. More precisely, let us introduce:

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

Some of these variables are really latent even for an econometrician with the maximal available information. Indeed an individual dying before the potential entry in LTC 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 LTC before the death, and the duration variable(s):

$$\begin{cases} Y_1^* = X_1 \text{ and } Y_2 = X_1 + X_3, & \text{if } Z = 1, \\ Y_2 = X_2, & \text{if } Z = 0. \end{cases} \tag{2}$$

In regime 1, we ideally 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 ideally 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:

$$\begin{cases} Y_1 &= X_1 Z, \\ Y_2 &= (X_1 + X_3)Z + X_2(1 - Z). \end{cases} \tag{3}$$

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 static frailty [see Abbring and Van den Berg (2003b)] and an extended version applied

to the pricing of joint insurance contracts for couples [see Gouriéroux and Lu (2013)]. The aim of our paper is to introduce dynamic (common) frailty featuring trends and able to capture the stochastic longevity phenomenon.

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 autonomous,
- 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 qualitative process $S = (S_t)$ which takes value in the state space $\{A, B, C\}$. The scheme below gives the possible paths of an individual's lifetime.

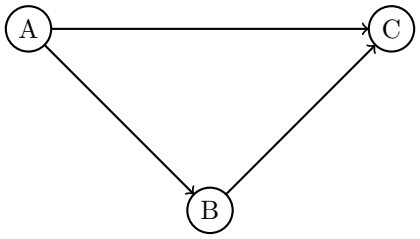


Figure 1: The potential transitions of an individual during its lifetime.

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 qualitative 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 into LTC. 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).

It is easily checked that (see Section 3.1) this reduced form specification is equivalent[12] to the structural model we defined in Section 2.1, 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 from the structural approach and that it should be preferred [see e.g. Imai and Soneji (2007)]. However, in some applications, one approach may be more convenient than the other one. To quote a summary from Han and Hausman (1990): "While econometricians have emphasized the presence of unobserved heterogeneity" (and therefore prefer the structural approach), "statisticians have instead emphasized the use of semi-parametric models which do not require parametric specification of the baseline hazard" (hence the choice of reduced form approach, often written without unobserved heterogeneity).

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 consider the case in which the latent variables X_1, X_2 are independent. Then we discuss the structural model with constant intensities to highlight the identification issues.

[12] The only difference is that in the latent variable approach, the variable X_3 is defined even if $X_1 > X_3$. But in such cases the value of X_3 is not important.

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 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, and 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:

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

where S_{12} is the joint survivor function $S_{12}(t_1,t_2) = \mathbb{P}[X_1 > t_1, X_2 > t_2]$. Therefore the structural approach with latent variables is equivalent to the reduced form approach. This equivalence is easily extended when (possibly unobserved and/or time-varying) stochastic factors are introduced, if we assume that (X_1, X_3) and X_2 are independent given the whole history of the factors and we define the transition intensities conditional on the whole history of the factors. 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 ideally observable variables

Let us now derive the joint distribution of the ideally 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:

$$f(y_1,y_2) = \lambda_1(y_1)\lambda_{2|1}(y_2-y_1|y_1)e^{-\Lambda_1(y_1)-\Lambda_2(y_1)-\Lambda_{2|1}(y_2-y_1|y_1)}, \text{ on domain } \mathcal{D}_1 = \{(y_1,y_2) : y_1 < y_2\}, \tag{4}$$

and

$$f(0,y_2) = \lambda_2(y_2)e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}, \text{ on domain } \mathcal{D}_0 = \{(y_1,y_2) : y_1 = 0, y_2 > 0\}. \tag{5}$$

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 equivalent, since in any application the latent variable should be integrated out. Nevertheless, as explained at the beginning of Section 2.1, studying directly (Y_1, Y_2) is preferred in order to distinguish explicitly the ideally observable information, that is (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:*

$$S_2(y_2) = \mathbb{P}(Y_2 > y_2) = \int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)} dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}, \tag{6}$$

and its p.d.f. is:

$$f_2(y_2) = \int_0^{y_2} \lambda_1(t)\lambda_{2|1}(y_2-t|t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)} dt + \lambda_2(y_2) e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}. \tag{7}$$

Proof. See Appendix 1. □

3.2 Identification in a model with constant intensities

For illustration purpose, let us assume a model with constant intensities λ_1, λ_2, and $\lambda_{2|1}$, that is with independent exponential latent variables. This simplified framework is useful to highlight the identification issue when only the lifetime variable Y_2 is observed.

For constant intensities 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 survivor function of lifetime Y_2 becomes:

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

and

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}\Big[1 + (\lambda_1+\lambda_2)y_2\Big] e^{-(\lambda_1+\lambda_2)y_2} + \frac{\lambda_2}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)y_2}, \text{ if } \lambda_1+\lambda_2 = \lambda_{2|1}. \tag{9}$$

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 survivor functions (see Appendix 2). In both cases ($\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$, or $\neq 0$), the distribution of lifetime Y_2 is a mixture of an exponential distribution with parameter $\lambda_1 + \lambda_2$, and a gamma distribution, $\gamma(2, \lambda_1 + \lambda_2)$, when $\lambda_{2|1} = \lambda_1 + \lambda_2$. This decomposition has the following interpretation:

$$\mathbb{P}(Y_2 > t) = \mathbb{P}(Z = 0)\mathbb{P}(Y_2 > t|Z = 0) + \mathbb{P}(Z = 1)\mathbb{P}(Y_2 > t|Z = 1),$$

with $\mathbb{P}(Z = 1) = \mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$.

Let us now discuss the identification of all parameters including the parameter $\lambda_{2|1}$ driving the time spent in LTC, when only the lifetime is observed. The following Proposition is a consequence of equations (8) and (9):

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 mixture representation has two distinct components and the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can be identified from the distribution of lifetime Y_2 given in equation (8).

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 (9), and the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can all be identified.

Therefore, under the assumption of constant intensities, the possibility of identifying the parameters is based on the jump in mortality intensity upon entry into LTC, that is, on the regime switch. Such a jump exists if and only if the point process associated with the LTC state causes the point process corresponding to mortality [see e.g. Abbring and Van den Berg (2003b)].

However, Proposition 2 *iii*) has to be interpreted carefully. The three parameters are identifiable, only if it is known *ex-ante* that the constraint $\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$ is satisfied.

4 Model with longevity effect

4.1 An identification issue

The model with constant intensity is not appropriate for modeling longevity effects in lifetime and LTC analysis. The longevity factor can be represented by introducing in the latent intensities a positive variable F indexed by calendar time. More precisely, let us consider a generation of individuals indexed by the birth date t_0, that is, the (stochastic) calendar date of death of an individual of this generation is $t_0 + Y_2$. The three intensities given the whole history $\underline{F}$ of the longevity factor are of the following form:

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

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.

The specification (10) disentangles the effect of age and of the current date in the intensities. The longevity factor is introduced as usual in a linear way. Since the factor is expected with a (deterministic or stochastic) trend, the linearity assumption implies cointegration between the different intensities with cointegrating vectors depending on age. This cointegration feature is introduced to capture the extension of lifespan going hand in hand with an extension or a diminution (according to the countries) of the amount of life spent in LTC. To get interpretable intensities for any generation, especially when t_0 tends to infinity, we consider a trend effect such that $\lim_{t\to\infty} F_t = 0$. Under this condition, 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 long term intensities, that are intensities in a far future. This is one difference with the basic Lee-Carter model [Lee and Carter (1992)] where in a far future the intensities are assumed equal to zero, that is, where the individual will necessarily become eternal.

Model (10) is semi-parametric with unknown functions $a_1(x_1)$, $a_2(x_2)$, $a_3(x_3|x_1)$, $b_1(x_1)$, $b_2(x_2)$, $b_3(x_3|x_1)$, and the dynamics of the longevity factor, which will be parameterized in the next subsection. This is a constrained structural model, but these constraints are not sufficient to identify all unknown parameters from just the observation of the lifetime Y_2, even if we have jump in the intensities and the generation can be viewed as a covariate. Indeed, in the limiting case when the generations have infinite sizes and all generations are observed, the observable

distribution summary is the survivor function indexed by the generation $S_2(y_2; t_0)$ [see Equation (13) for a typical expression of this function]. This is a function on $]0, \infty[^2$, but the set of functions to be estimated already includes two functions $a_3(x_3|x_1)$ and $b_3(x_3|x_1)$ defined on the same space. Then the order condition for identification is not satisfied. Such a lack of identification is standard in models with treatment effects [see e.g. Abbring and Van den Berg (2003b)]. It is here observed despite restrictions already introduced on the models and the effect of two exogenous variables, i.e., the observed indicator of the cohort and the unobserved longevity factor.

Thus to recover the identification of the joint distribution of the latent intensities (X_1, X_2, X_3), we need additional restrictions. We will assume that the conditional intensities $a_3(x_3|x_1)$ and $b_3(x_3|x_1)$ can be written in terms of univariate functions defined on $]0, \infty[$.

4.2 Constrained specifications

In the application we will consider two constrained specifications.

4.2.1 Specification of the baseline intensities

The first specification corresponds to the Markov case, where the intensity $\lambda_{2|1}(x_3|x_1, t_0)$ depends on x_3 and x_1 through the current age $x_3 + x_1$ only:

$$\begin{cases} a_3(x_3|x_1) & = a_3(x_3 + x_1), \\ \tilde{b_3}(x_3|x_1) & = \tilde{b_3}(x_3 + x_1). \end{cases} \tag{11}$$

We will also consider the following semi-Markov model,

$$\begin{cases} a_3(x_3|x_1) & = a_4(x_3) + a_5(x_1), \\ \tilde{b_3}(x_3|x_1) & = b_4(x_3) + b_5(x_1) \end{cases} \tag{12}$$

with additive decomposition of the conditional intensities. For instance, under the Markov model (11), the survivor function of the observed variable y_2 given the future factor path $\overline{F}_{t_0} = \{F_\tau, \tau \geq t_0\}$ is:

$$\begin{aligned} S_2(y_2, t_0) = \int_0^{y_2} [a_1(x) + b_1(x) F_{t_0+x}] \exp\Big(- \int_0^x [a_1(s) + b_1(s) F_{t_0+s}] ds \\ - \int_0^x [a_2(s) + b_2(s) F_{t_0+s}] ds - \int_x^{y_{2,i}} [a_3(s) + b_3(s) F_{t_0+s}] ds \Big) dx \\ + \exp\Big(- \int_0^{y_2} [a_1(x) + b_1(x) F_{t_0+x}] dx - \int_0^{y_2} [a_2(x) + b_2(x) F_{t_0+x}] dx \Big). \end{aligned} \tag{13}$$

4.2.2 Specification of the factor dynamics

i) **Deterministic factor.** Let us first assume a deterministic factor (F_t), with exponential pattern:

$$F_t = \exp(-mt), \tag{14}$$

where $m > 0$. The factor is known up to the value of the parameter m.

Under the exponential specification (14), the age-calendar time model (10) can be equivalently written as an affine age-cohort model[13]:

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = \lambda_1(x_1, F_{t_0}) = a_1(x_1) + \tilde{b}_1(x_1)F_{t_0}, \\ \lambda_2(x_2|\underline{F}, t_0) & = \lambda_2(x_2, F_{t_0}) = a_2(x_2) + \tilde{b}_2(x_2)F_{t_0}, \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = \lambda_{2|1}(x_3|x_1, F_{t_0}) = a_3(x_3|x_1) + \tilde{b}_3(x_3|x_1)F_{t_0}. \end{cases} \tag{15}$$

with, say, $\tilde{b}_1(x_j) = b_j(x_j)e^{-mx}, j = 1, 2, \tilde{b}_3(x_3|x_1) = b_3(x_3|x_1)e^{-mx_1 - mx_3}$.

In the age-calendar time model, the shocks on the factors depend on date t, whereas in the age-cohort model the factor has an impact at birth with consequences during the whole cohort lifetime. Thus, for exponential factor, it is not possible to distinguish between both interpretations of longevity, that is to say if longevity is associated with time, or with generation [see also Heckman and Robb (1985)].

The affine age-cohort specification is very similar to the popular proportional hazard models in survival analysis, in which the effect of the exogenous covariates, here the cohort t_0, appears often in a multiplicative way in the conditional intensity given the covariate. This model is mathematically easier to handle for nonparametric identification (see Appendix 7). The coefficient $\tilde{b}_j, j = 1, 2, 3$ measure the persistence of different intensities with respect to the generation effect F_{t_0}.

However, the age-calendar time specification is also widely used in demography and finance. It assumes that the longevity phenomenon is instead more influenced by calendar year fluctuations which incorporates, besides a general decrease of mortality (due to e.g. the progress in medicine), temporary effects such as pandemic, natural disasters, etc. The nonparametric identification of an age-calendar time model, with an unconstrained F, is more difficult to study. Indeed, for a same cohort t_0, the intensity of the observed variable y_2 depends on the age x via both the

[13] It is only in this exponential case that we have both an affine age-cohort model and an equivalent affine age-period model. Indeed, if we have both $\lambda_1(x_1|\underline{F}, t_0) = a_1(x_1) + b_1(x_1)F_{t_0+x_1}$ and $\lambda_1(x_1|\underline{F}, t_0) = \tilde{a}_1(x_1) + \tilde{b}_1(x_1)F_{t_0}$, it is easily shown that, given continuity assumptions on the function $t \mapsto F_t$, this function is necessarily an exponential function of time t.

baseline hazards a_j and $b_j, j = 1, 2, 3$ and the whole path of F between time t_0 and $t_0 + y_2$ (see the discussions in Section 4.3.2).

ii) **Stochastic factor.** Because of the stochastic nature of the longevity, we would also like to model the common factor (F_t) as an unobserved stochastic process, often called dynamic frailty since Duffie et al. (2009). For the comparison with the deterministic exponential specification above, we will assume in applications that the dynamics of the stochastic factor F is a Cox-Ingersoll-Ross (CIR) process [see Cox et al. (1985)]:

$$dF_t = -mF_t dt + \sigma\sqrt{F_t}dW_t, \tag{16}$$

where $\sigma > 0$, $m > 0$, W is a standard Brownian motion, and the initial condition is $F_{\min t_0} = 1$, where $\min t_0 := 0$, say, is the birth date of the first cohort.

This CIR model includes the deterministic model as a limiting case. If $\sigma = 0$, then the solution of the differential equation (16) is $F_t = \exp(-mt)$. Thus the CIR model is just introducing uncertainty around the deterministic exponential model. Therefore, this CIR process still has a nonstationary feature, which reflects the longevity phenomenon.

The advantage of introducing a stochastic specification of the factor over a deterministic, say, exponential specification, is that we can quantify the uncertainty of both the model fit and the future evolution. These uncertainties should be taken into account when pricing LTC insurance contracts, computing the regulatory required capitals and performing stress tests [see the discussion in Keilman et al. (2002) for macropolicy implications].

The choice of a CIR process has several other advantages. Firstly, it guarantees the positivity of the intensity functions $\lambda_1, \lambda_2, \lambda_{2|1}$ 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 by using the affine property of the process.

Appendix 5 summarizes the basic properties of this CIR process, including its existence, the potential hitting time at 0 and its behavior afterwards, as well as its discrete time counterpart, which is an autoregressive gamma process (ARG).

4.3 Nonparametric identification

Let us now discuss the identification issue. For expository purpose, we consider the Markov specification (11).

4.3.1 Deterministic exponential factor

Let us first consider the case where the factor F is deterministic and exponential, and the intensity of X_3 given X_1 is Markov. Assume that for each cohort, at the age origin $y_2 = 0$, the proportion of people already in LTC is null, and $F_{t_0} = 1$ for some pre-specified value of t_0.

Proposition 3. *Assume that we observe the lifetime of a continuum of cohorts of individuals indexed by t, where t varies in an open set $]t_0 - \epsilon, t_0 + \epsilon[$ for $\epsilon > 0$, that the six functions $a_j, b_j, j = 1, 2, 3$ are continuous and positive. Then, the parameter m is identified, and we have the following identification results for the six functions:*

1. *If $b_1 + b_2 = b_3$ for all y, then b_3 can be globally nonparametrically identified; the others cannot be identified.*

2. *If there exists constants c, c' such that $b_1 + b_2 - b_3 \geq c > 0$, and $|b_2 - b_3| > c'$ for each age y, then $b_1 + b_2$ is globally nonparametrically identified; the other functions are at least locally identified.*

3. *If there exists a constant d such that $b_1 + b_2 - b_3 \leq -d < 0$ for all y, then functions b_3, a_3 are globally identified; the other functions are at least locally identified.*

Proof. See Appendix 7. □

In other words, the repeated measurement across different cohorts of the nonlinear effect of the longevity factor on the aggregated lifetime behavior allows for identifying both the functional parameters and the longevity factor. The assumption that at origin, the proportion of people already in LTC is null is an implicit condition of our model and is already used in Equation (7). The assumption that all the functions are continuous means that, the entry into LTC is the only possible mortality jump during one's lifetime. The observation of a continuous-valued covariate t is also a standard assumption in the identification literature of survival models [see e.g. Abbring and Van den Berg (2003a)] and of treatment effects [see Abbring and Van den Berg (2003b) Proposition 2.3.4]. Indeed the proof of identification of m relies on the same "identification at zero" argument as in these papers. Nevertheless our identification result is not a consequence of theirs. Indeed this literature assumes that the time of treatment is observable and usually consider the mixed proportional hazard (MPH) specifications. For longevity models, the specification of the intensities cannot be multiplicative in the observable regressor, due to the need of a limiting model for the far future [see e.g. system (15)].

4.3.2 Stochastic factor

Let us now consider the identification of the Markov model with a stochastic factor. Loosely speaking a (functional) parameter is identifiable if it can be consistently estimated. Thus the notion of identification depends on the assumed asymptotics. For our problem, this is a double asymptotics, in which both the number T_0 of observed generations and the number of individuals observed in each generation tend to infinity. In the limiting case of this double asymptotics, the family of survivor functions $S_2(y, t_0)$ given in (13) is asymptotically known, that is, we can reconstitute the set of survivor functions given the existing factor path[14]. To summarize we have the following Proposition.

Proposition 4. *It is equivalent to consider the identification of the intensity components $a_1, b_1, ...$ in a model with stochastic factor, or to consider the identification problem for a model with (unconstrained) deterministic factor, where the factor path coincides with the realized path.*

Let us now consider system (10). This is a system of equations indexed by y_2 and t_0, which has to be solved w.r.t. functions $a_1, b_1, a_2, b_2, a_3, b_3, F_t$. This system is in general over-identified, except for some special factor paths such as deterministic exponential path. But since (F_t) is a diffusion process, the probability of reduced rank is zero. Thus we have the following Proposition:

Proposition 5. *Functions $a_1, b_1, a_2, b_2, a_3, b_3$ are locally identifiable, a.s., that is except for a negligible set of factor paths.*

The analysis of identification with unobserved stochastic dynamic frailty is completely different from the analysis in standard treatment effect models. Indeed, in models with treatment effects, the unobserved heterogeneity is individual and represented by a scalar or vector random variable. In our framework the longevity factor is a process, therefore much more complex. Nevertheless, the cross-sectional asymptotics allows for eliminating the uncertainty on this factor, that is for replacing the process by its underlying trajectory (Proposition 4). Then the observation of a large number of cohorts introduce the orthogonal dimensions leading to identification (Proposition 5).

Finally, wherever $a_j, b_j, j = 1, 2, 3$ are identifiable, from granularity theory [see e.g. Gagliardini and Gouriéroux (2014)], we can also identify the realized factor path, and then the parameters of the factor dynamics.

[14] For an asymptotics in T_0, with one observed individual in each cohort, say, it would only be possible to reconstitute the integrated survivor function $\overline{S}_2(y, t_0) := \mathbb{E}[S_2(y, t_0)]$, where the expectation is taken with respect to the stochastic future factor path.

5 Applications

Under the restrictions introduced in Section 4.2, the scalar and functional 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 consider the different specifications for models with deterministic or stochastic factors, and derive the likelihood functions, 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. But under the specifications of the factor that we consider, the likelihood function admits closed form formula when an appropriate discretization scheme is used. We also approximate the functionals $a_j, b_j, j = 1, 2, 3$ by parametric splines. We denote by θ the set of all parameters including both the splines parameters and the parameters characterizing the factor dynamics.

5.1 The likelihood function

5.1.1 Model with deterministic factor

Let us first consider the basic model with a deterministic factor $F_t = e^{-mt}$. We denote by $i, i = 1, ..., n$, the individuals and assume that the set of latent variables $(X_{1,i}, X_{2,i}, X_{3,i})$, $i = 1, ..., n$ are independent with identical joint distribution, which depends on the generation only. Then the individual lifetimes $Y_{2,i}, i = 1, ..., n$ are also independent with a distribution depending on t_0 only. Taking into account the right censoring of the lifetimes, the log-likelihood function is:

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log f_2(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log S_2(y_{2,i}, t_0, \theta) \Big\}, \tag{17}$$

where $\mathcal{I}^u_{t_0}$ (respectively $\mathcal{I}^c_{t_0}$) 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, the censoring time, otherwise, and θ denotes the parameter.

5.1.2 Model with dynamic frailty

The expression of the log-likelihood is similar as (17), except that the terms f_2, S_2 should be integrated with respect to the path of factor (F_t). More precisely, we define $\overline{S}_2(y_{2,i}, t_0, \theta) = \mathbb{E}[S_2(y_{2,i}, t_0, F)]$ the integrated survivor function, where $S_2(y_{2,i}, t_0, F)$ is the survivor function

conditional on the path of the factor (F_t) and with expression given by (13). Similarly we define $\overline{f}_2(y_{2,i}, t_0, \theta) = \mathbb{E}[f_2(y_{2,i}, t_0, F)]$. Then we get:

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log \overline{f}_2(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log \overline{S}_2(y_{2,i}, t_0, \theta) \Big\}. \tag{18}$$

This expression can be theoretically calculated in continuous time, but at the cost of numerically solving ordinary Riccati differential equations[15]. A simpler way is to approximate the continuous time model with its time-discretized version. This is useful when the available data are collected in discrete time, which is actually the case. More precisely, assume that the intensity functions are constant[16] between two neighboring integer dates: 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 the other duration variables. The log-likelihood function is therefore approximately:

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log f_2^{\text{disc}}(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log S_2^{\text{disc}}(y_{2,i}, t_0, \theta) \Big\}, \tag{19}$$

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:

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

[15] This treatment is standard in the literature of term structure of interest rates and credit spreads with affine underlying factors, see e.g. Duffie et al. (2000).

[16] This necessitates also to replace the continuous time CIR process with its time-discretized version, which is an ARG process. See Appendix 5.

We give in Appendix A.3.2 the expressions of these expectations. They can be written in terms of the Laplace transform of process F, and have closed form for affine processes such as the CIR process (otherwise, the calculation of the log-likelihood requires simulation of the factor paths and is numerically cumbersome).

5.2 The data

The methodology of the previous subsections 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 gender and each cohort t_0 since 1737, the size of the Population-at-Risk and the number of deaths[17] 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 summary statistics of the French male population. Because of the longevity phenomenon, the distribution of lifetime is shifting to higher ages. This can be illustrated by the increase of cross-sectional life expectancy[18]. 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 6), while the cross-sectional quantities are model-free, but they 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.

[17] As a consequence, the corresponding estimates of the mortality intensity function are available as well.
[18] Also called period life expectancy in demography.

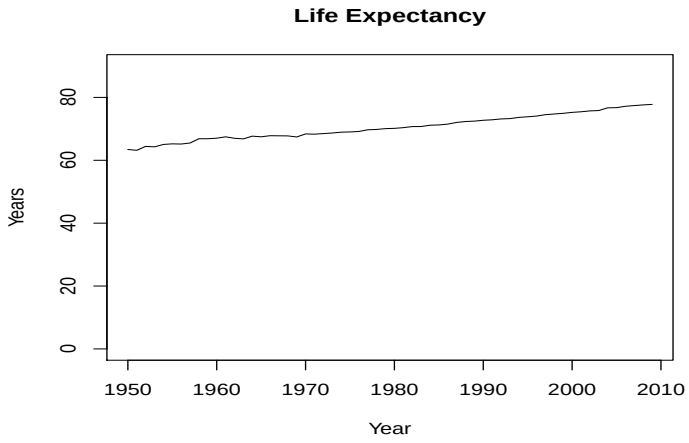


Figure 2: Evolution of the life expectancy at birth for deaths occurring in the same year.

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, that is 3 month 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 seniors in the 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.

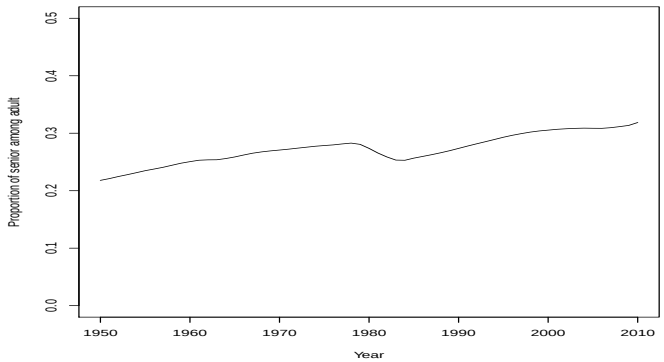


Figure 3: The dependency ratio by year.

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 and of the pension funds.

5.3 Markov model with deterministic exponential factor

We estimate the model introduced in Section 5.1.1 on the French male data. We consider the population of males who survive up to age 50. As we suppose an homogeneous population[19], 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 approximating the functions $a_j, b_j, j = 1, 2, 3$ by linear splines:

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 $\tilde{b_1}(x_1) = b_1(x_1)\exp(-mx_1)$ 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.

iv) The function $b_2(x_2)$ is such that $\tilde{b}_2(x_2) = b_2(x_2)\exp(-mx_2)$ 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 $\tilde{b}_3(x_3|x_1) = b_3(x_3|x_1)e^{-m(x_3+x_1)}$ 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-period decomposition [see equation (11)]. 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 over-parameterization and less robust results. Empirically we find that other parametric specifications, such as exponential splines, can also fit the model relatively well. We show in Appendix 3 that the linear

[19]By homogeneous population we mean a population without multiplicative unobserved heterogeneity as in Vaupel et al. (1979). Since we assume that at the beginning of the observation ($y = 50$) nobody is in LTC, there is no heterogeneity linked to the initial autonomy status neither.

spline specification 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.

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 current limit age of the human being[20]. 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 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.

[20]The oldest living human is currently a 116 years old man.

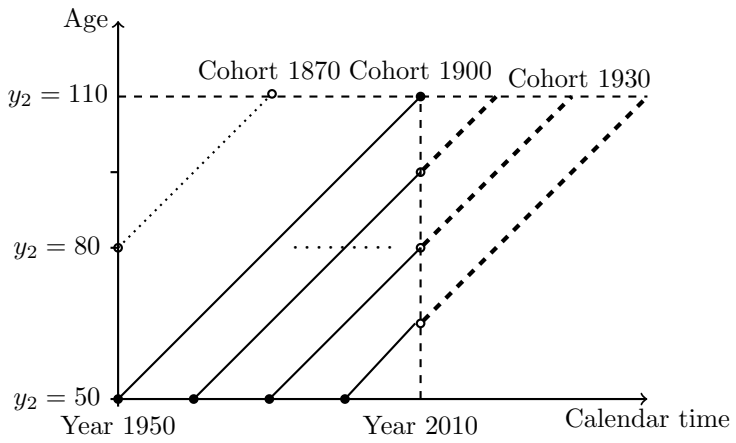


Figure 4: Lexis diagram of cohorts and their observability. The study period ranges from year 1950 to 2010.

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 $\theta = (w_1, w_2, ..., w_{16}, m)$. For brevity the value of the estimator, the goodness of fit, as well as the discussion of this model are given in Appendix A.4.1. We first compute the model implied intensity function of Y_2 and compare it to the historical data. Besides, we can also plot the evolution of the latent hazard functions, as well as the implied evolution of the proportion of people in long term care (i.e. prevalence), that is, the distribution of the unobserved heterogeneity.

5.4 Semi-Markov model with deterministic exponential factor

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 assumption is that it depends also on the age of entry into LTC z, or equivalently, on the time elapsed since this 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 becomes:

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

The estimation and discussion are gathered in Appendix A.4.2.

5.5 Model with dynamic frailty

Let us finally 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 (16)], 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 (19). Since the model with deterministic factor 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 with deterministic factor derived in Section 5.4. We report in Table 2 the value of the estimator w.

Table 2: 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 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 times by integrating out the whole history of the dynamic frailty, and then we calculate the hazard function by computing its minus log-derivative:

$$h(y_2) = \lim_{h\to 0} \frac{\mathbb{P}[y_2 \leq Y_2 < y_2 + h]}{h} = -\frac{\partial}{\partial y_2} \log \mathbb{E}\Big[S_2(y_2|\theta, F)\Big] = \frac{\mathbb{E}\Big[f_2(y_2|\theta, F)\Big]}{\mathbb{E}\Big[S_2(y_2|\theta, F)\Big]}. \tag{20}$$

We display in Figure 5 the intensity function of Y_2 and compare its values to the observed values from the data.

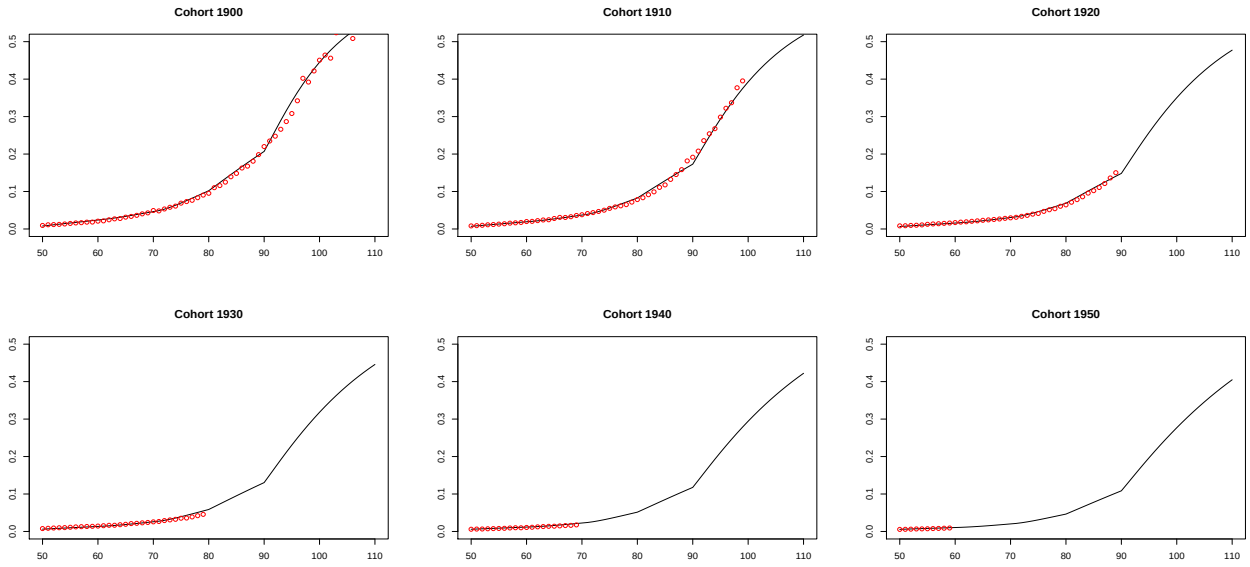


Figure 5: Hazard function of the lifetime variable. Dotted line: historical data. Full line: the model (for both the past and future years).

Once the parameters are estimated, we infer the path of unobserved frailty process (F_t). 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 predicting the future mortality and 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 $\overline{f}_2(y_2, t_0, \theta), \overline{S}_2(y_2, t_0, \theta)$ 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 simulate a certain number of paths of the frailty process conditionally on both the estimated value of the parameter and on the observations

$Y_{2,i}, i \in \mathcal{I}^u \cup \mathcal{I}^c$, that are either the dates of death or the right censoring ages of all individuals. This is done by Gibbs sampling, as in Duffie et al. (2009). Appendix 6 gives the details of this methodology. In Figure 6, we plot, for each year, the simulated factor mean $\mathbb{E}[F_t|\theta, Y_2]$ conditional on all the observed $Y_{2,i}, i \in \mathcal{I}^u \cup \mathcal{I}^c$. For 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.

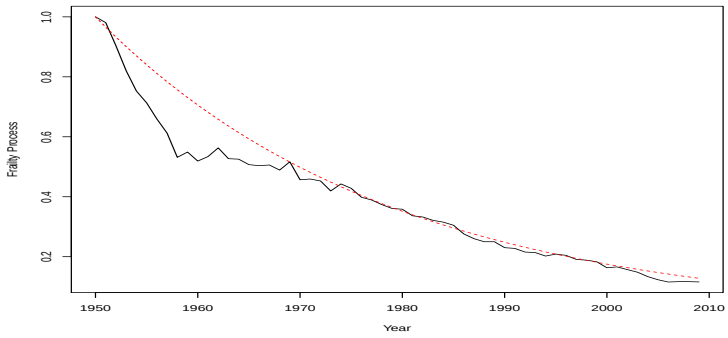


Figure 6: Simulated factor mean (full line) and the deterministic path (dotted line).

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. Indeed for most dates t, we observe empirically that $\mathbb{E}[F_t|\theta, Y_2] < \mathbb{E}[F_t|\theta]$. This result was expected, 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 values of factor F are replaced by their filtered values. Not surprisingly, for each of its simulated paths, we get very satisfactory fit to the observed lifetime intensity similarly as in Figure 5. 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 with deterministic factor. For instance, for the Markov model with deterministic factor, we see from Figure 14 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 %).

5.6 Comparison of the models with deterministic and stochastic factors

The three models, that are the Markov and semi-Markov model with deterministic factor as well as the semi-Markov model with stochastic factor all provide satisfactory fits. The maximized log-likelihoods are respectively: -38710240, -38709452, -38704065, and the corresponding values of the BIC are: 77420764, 77419205 and 77408448. It was expected that the semi-Markov model with deterministic factor (resp. the semi-Markov model with stochastic factor) has a higher likelihood than the nested semi-Markov model with deterministic factor (resp. Markov model with deterministic factor), but the difference is rather small. However, the comparison between the semi-Markov models with deterministic and stochastic factor 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 these models in terms of prediction.

6 Prediction of individual LTC and mortality risks

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 with deterministic factor, 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 (20)), admit explicit forms after integrating out the frailty process, but confidence intervals have to be computed by simulation. More precisely, for each simulated past history of process F obtained from the Gibbs sampler (see Subsection 5.5), we simulate its future path and obtain the predictive distributions conditional on the whole factor path, whose formulas are similar as for the model with deterministic factor. 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 autonomy 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 autonomous, that is, $X_1 > y_2, X_2 > y_2$, we have to predict Y_1, Y_2, X_1, X_2, X_3,

and so on. We first derive explicit prediction formulas for a model with deterministic factor. Then we consider the prediction of future risks in Case *iii*) for the French males, by both the Markov model with deterministic factor and the semi-Markov model with dynamic frailty. These quantities are calculated for different cohorts, but for expository purpose we omit the cohort index t_0. Since the individual observations are independent, we can perform the computation independently for each individual. For expository purpose we omit the individual index i.

6.1 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 (4) and (5)].

Conditional distribution of (X_1, X_2, X_3) given Y_2. This conditional distribution has two components 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 $\mathcal{D}_4 = \{(x_1, x_2, x_3) \in \mathbb{R}_{\geq 0}, x_2 = y_2, x_1 \geq y_2\}$, respectively. 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

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

6.2 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 $Y_1 > y_2$. It has a density with respect to the measure $\delta_0 + \lambda_{]0,y_2[}$, and this density is:

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

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

and

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

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 (30)].

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 (29) for the Markov model).

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

6.3 Case *iii)*

Let us now assume that the available information set is $X_1 > y, X_2 > y$. A special case is when $y = 50$, since any individual enrolled in the study at this age is autonomous[21], 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 autonomy 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 called the cumulative incidence (at age $Y_2 = \infty$).

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

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

as well as the residual life expectancy without LTC (or Healthy Life Years[22]) 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[23]. It is of particular interest to insurance companies or public social

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

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

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

security plans, since it impacts the expected cost of an LTC insurance policy in a direct way. We have:

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

In general, the term $\int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3$, 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 with deterministic factor 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 each simulated factor path F. Figure 7 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.

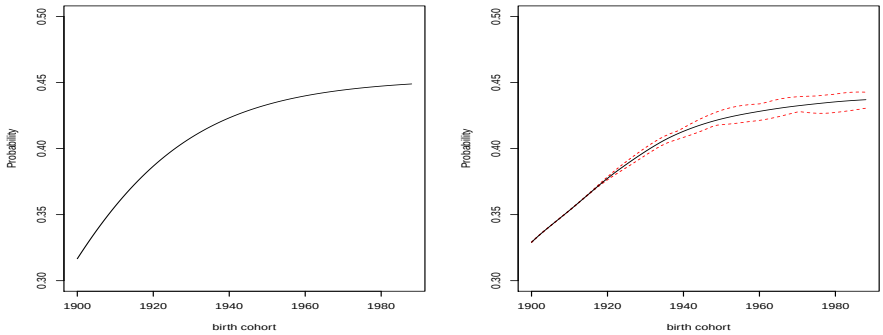


Figure 7: Evolution of the probability of entering into LTC during its lifetime as a function of the cohort. Left panel: the Markov model with deterministic factor, right panel: the semi-Markov model with dynamic frailty; full line: the expected value, that is when frailty is integrated out, dashed lines: the 90% confidence bounds.

The Markov model predicts a slightly higher probability of entering into LTC 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. Theses probabilities are in line with the

projection based on LTC use history of a sample of Americans by Spillman and Lubitz (2002), who predict that in 2020, the probability of a 65-year-old[24] ever entering a nursing home to will increase to 46 %. The result is also to be compared to Figure 18 in Appendix, where we plot the proportion of people in LTC at any ages, which is decreasing in cohort[25]. 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 filtered past path is considerably smaller than the variation of its predicted future path 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.

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.

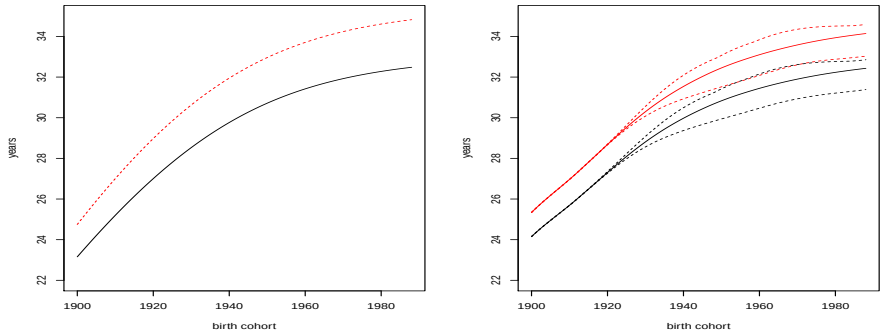


Figure 8: Evolution in t_0 of the residual life expectancy, with potential LTC (dashed line) and without (full line) LTC, at age 50. Left panel: the Markov model, right panel: the semi-Markov model with dynamic frailty; full lines: the expected values, dashed lines: the 90% confidence bounds; the three upper curves are for the life expectancy with potential LTC.

For a French male aged 50 in 2010, the 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 increases 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.

[24]Which is roughly of the same order than the probability for a 50-year-old given the relatively lower intensities between age 50 and 65.

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

Finally, let us calculate the uncertainty of the following 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 dying directly) for the individual i aged 50 in, say, year $\tau = 2010$. 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 large. Nevertheless, it can be approximated by using the granularity theory [see e.g. Gagliardini and Gouriéroux (2014)]. For the model with deterministic factor factor, the distribution of the quantities in (23) are approximately Gaussian by the Central Limit Theorem. For the model with dynamic frailty, conditional on each simulated factor path, these quantities are still approximately Gaussian; therefore their 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, i.e. cross-sectional asymptotic (CSA) VAR, provides the undiversifiable component of the risk. This CSA VaR is easily calculated: for the model with deterministic factor, it is equal to zero; for the 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. $e_2(y|F) = \mathbb{E}[\min(X_1, X_2)|X_1 > y, X_2 > y, F]$). These quantities have already been calculated (see Figure 8).

To illustrate this approach, let us take $n = 10, 100, \infty$, and $\alpha = 0.05, 0.95$. The confidence bounds are displayed in Table 3.

Table 3: 90% confidence bounds for the average residual lifetime for a portfolio of n individuals who are 50 years old in 2010.

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

For both empirical means, the confidence interval is larger for the model with (common) frailty, which incorporates the uncertainty of the frailty process (both its future and past), 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.

6.4 Comparison with real data on LTC

Let us finally compare the model-based prediction with data on LTC from a large insurance company. Such private proprietary database usually concern the customers and are not representative of the whole population. They are subject to selection biases due to both the behavior of the company and of the customers. Let us discuss the expected bias for the analysis of LTC.

- Since the LTC insurance market is young and small, products are not very differentiated. Thus the insurance company will try, for a given price of the contract, to select the least risky customers[26]. Thus we expect that in this database, the time spent in LTC is smaller than for the whole population.

- On the other hand, the standard economic literature insists on the role of adverse selection which tends to increase the average risk profile of the customers. However, this standard argument seems to be not valid in the LTC framework, a finding also confirmed by Finkelstein and McGarry (2006). They attribute this to the offsetting effect of selection into the market and find evidence that wealthier individuals and individuals who exhibit more cautious behavior are both more likely to have LTC insurance coverage and less likely to use LTC. Indeed, in insurance problems with irreplacable objects, individuals' utility function is in general state-dependent [see e.g. Dionne (1982), Karni (1983)], i.e. with a higher risk aversion in the LTC state. The preference to be better covered in this state will imply an increased demand. On the other hand, the weak effect of the adverse selection could also be partially explained by the long-term nature of the risk, which makes it difficult for individuals to exploit asymmetric information. In the same direction will be the bias coming from the income effect since the customers who can afford a private insurance are likely to have a higher income than the national average.

To summarize, we expect that the endogenous selection by both the insurance company and the customers are going in the same direction of overweighting of the best risks, i.e. smaller

[26] For instance, many insurance companies believe that living with one's partner, as well as being smoker, are indicators of small time spent in LTC.

probability of entering into in LTC in the database w.r.t. the whole population.

The database concerns a specific insurance product with only one LTC state; it has been launched in 1994 and sales continued up to[27] 2000, but the database is maintained even after that date. There are about 15000 male policyholders[28], the majority of whom were born between 1925 and 1940 (see Figure 9 for a histogram of the cohort of all policyholders) and bought the contract in their 60's. Thus they are quite young at the end of the observation period, that is 2014. As a consequence, observations are heavily right censored. Indeed, 20 % individuals died without LTC and only 5 % entered into LTC before the end of the observation period, with a potentially further censored final death date; the other observations are completely censored. No events are observed beyond age 90. The portfolio size is not sufficient to conduct a real cohort-specific analysis and the individuals from different cohorts are aggregated.

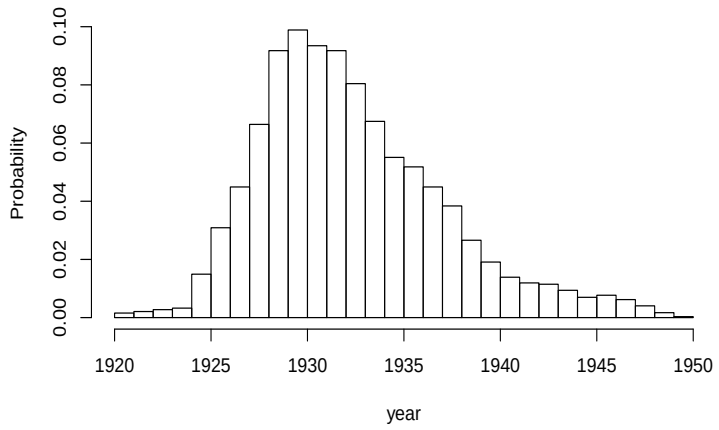


Figure 9: Histogram of birth cohort of all policyholders.

[27] After 2000, the company launched a new product with significant changes of policy terms; therefore the new product cannot be compared directly to the original one.

[28] The size of the portfolio is rather reduced with respect to the French population. Nevertheless, it is believed to be one of the largest and most reliable databases from one of the largest reinsurance companies in the world. This illustrates the difficulties of the insurance industry in providing comparable LTC products, and in maintaining quality databases.

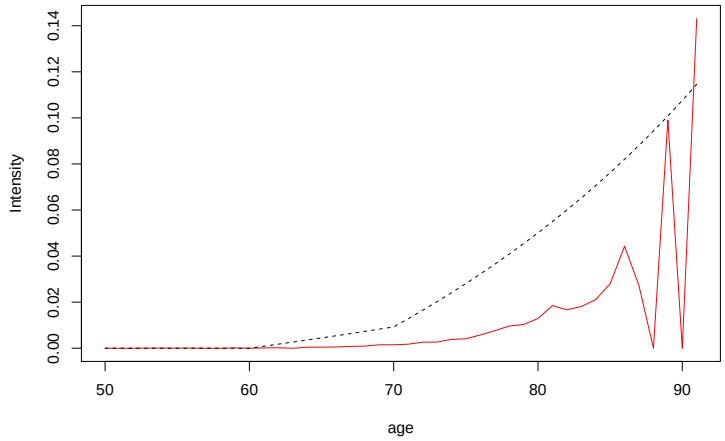


Figure 10: Comparison between the intensity of entry into LTC implied by the model (for general population) and that observed on the insurance data. Dashed line: the model; full line: the data.

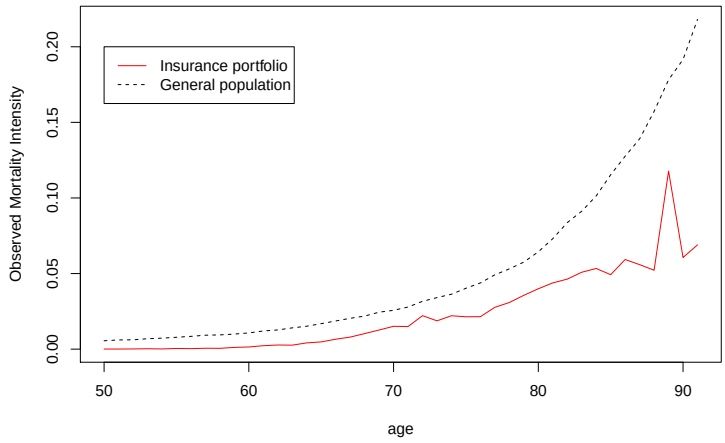


Figure 11: Comparison between the observed mortality intensity of the two populations.

Figure 10 compares the intensities of entering into LTC computed for the set of policyholders and deduced from the estimated model for the general population. For the insurance portfolio, the estimated intensity is $\hat{\lambda}_1(x) = -\frac{d}{dx}\log \hat{S}_1(x)$, where $\hat{S}_1(x)$ is the Kaplan-Meier estimator of the marginal survivor function of the entry into LTC. For the model based intensity at each age x, we took a weighted sum of $\Big(\lambda_1(x|t_0)\Big)$ for different cohorts t_0, where the weights are

determined by the share of each cohort among individuals that survive up to age x. This allows us to correct the longevity bias of the aggregated portfolio. Figure 10 shows that our model predicts a slightly higher intensity of entry into LTC for general population than that observed from the insurance data, especially for lower ages. This difference can be partly explained by the endogenous selection of policyholders by the insurance company and the choice of individuals to buy such a contract. Whereas the entry in LTC is exogenous, the enrollment in a private LTC coverage is endogenous (see the discussion at the beginning of the current subsection). To further confirm the selection effects, Figure 11 plots the aggregated mortality intensity (without distinguishing the autonomy state) for both the policyholders and the general population. The huge discrepancy between the two curves suggests that the insured population has a much better health than the general population, and, therefore are likely to have a lower intensity of entry into LTC[29]. This comparison shows the difficulty in taking into account the available LTC insurance data, when estimating the models, due to the poor data quality and the endogenous selection.

7 Conclusion

In this paper we proposed a new methodology to predict the probabilities of entering into LTC 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 which can be used as long as the data on LTC are either unavailable, or weakly reliable, or under endogenous selectivity. This implicit state may differ from that of a specific LTC database[30] and it would be interesting to compare the hypothetical date of entry in LTC with the different dates of losing Eating, Dressing, ... abilities, when longitudinal data will become available and reliable. This may lead to change the definition of the Instrumental Activities of Daily Living, as well as the design of LTC insurance products.

Our model is based on minimal[31] observability and thus assumes a single LTC state. In some cases it may be attempting to include other observed information, such as the regular measurement of various individual health indicators, or even direct observation of the LTC use. The inclusion of such information is theoretically possible, but since it often comes from a different database for a smaller population sample and/or a shorter period, its effective use requires

[29] In other words we assume a positive correlation between LTC and mortality risks. See the beginning of this subsection or Murtaugh et al. (2001) for a discussion of this assumption.

[30] Which is logical, especially given the lack of a universal definition and the poor quality of existing databases.

[31] Although repeated across different cohorts.

additional, case-dependent care. This can be an area of further research when appropriate database becomes available.

Finally, the joint statistical analysis of entry into LTC and mortality is a requested step, before checking if individual LTC risk is really insurable by insurance companies, or if it is profitable to combine mortality and LTC risks into a joint insurance product[32].

Appendices

Appendix 1 : 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 2 : Technical lemmas

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

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

[32]For instance, Murtaugh et al. (2001) argue that based on the assumption that the two risks are positively correlated, then combining the two risks would significantly lower the overall insurance premium, increase the attractiveness of the products, and thus also limit the adverse selection.

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 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:

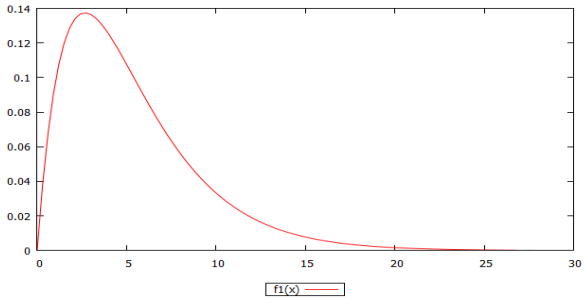
$$\frac{dg}{dy} = -e^{-ay}(aby + a - b) \geq 0, \qquad \forall y > 0,$$

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 3 : Expression of the log-likelihood function

A.3.1 Model with deterministic factor

In this section we give the detailed expression of the log-likelihood function (17) in the model with deterministic factor. 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, we have,

$$\begin{aligned}
&f_2(y_{2,i}, t_0, \theta) \\
&= \Big(a_3(y_{2,i}) + \tilde{b}_3(y_{2,i})F_{t_0}\Big) \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx \\
&\quad + \Big(a_2(y_{2,i}) + \tilde{b}_2(y_{2,i})F_{t_0}\Big) \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned} \tag{24}$$

and

$$\begin{aligned}
S_2(y_{2,i}, t_0, \theta) = &\int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx \\
&\quad + \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned} \tag{25}$$

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:

$$S_2(y_{2,i}, t_0, \theta) = e^{-\int_0^{y_{2,i}} [a_3(s)+\tilde{b}_3(s)F_{t_0}]ds} \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^t [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds$$
$$- \int_0^t [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds + \int_0^t a_3(s) + \tilde{b}_3(s)F_{t_0}ds\Big)dx$$
$$+ \exp(-s_3 y_{2,i}^2/2 - i_3 y_{2,i}), \tag{26}$$

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:

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

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[33]. 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 (24) and (25) 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).

A.3.2 The model with dynamic frailty

Let us adopt the discretization scheme described in Subsection 4.2.2, and replace the continuous time process (F_t) by its time discretized version $(F_{[t]})$. whose values at integer times is an ARG

[33]Indeed, the cumulative distribution function has no closed form, but its computation is rather fast using standard softwares.

process (see Appendix 5).

$$\overline{f}_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} = y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} = y_{2,i} \mid F]\Big]$$
$$= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}-1}\Big[1 - e^{-a_1(i)-b_1(i)F_{t_0+i}}\Big]\Big[1 - e^{-a_3(y_{2,i}|i)-b_3(y_{2,i}|i)F_{t_0+y_{2,i}}}\Big]$$
$$\exp\Big(-\sum_{j=0}^{i-1}[a_1(j)+b_1(j)F_{t_0+j}] - \sum_{j=0}^{i-1}[a_2(j)+b_2(j)F_{t_0+j}] - \sum_{j=i+1}^{y_{2,i}-1}[a_3(j|i)+b_3(j|i)F_{t_0+j}]\Big)\Bigg]$$
$$+ \mathbb{E}\Bigg[\Big(1 - e^{-a_2(y_{2,i})-b_2(y_{2,i})F_{t_0+y_{2,i}}}\Big)\exp\Big(-\sum_{i=0}^{y_{2,i}-1}[a_1(i)+b_1(i)F_{t_0+i}] - \sum_{i=0}^{y_{2,i}-1}[a_2(i)+b_2(i)F_{t_0+i}]\Big)\Bigg], \tag{27}$$

and

$$\overline{S}_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} > y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} > y_{2,i} \mid F]\Big]$$
$$= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}}\Big[1 - e^{-a_1(i)-b_1(i)F_{t_0+i}}\Big]\exp\Big(-\sum_{j=0}^{i-1}[a_1(j)+b_1(j)F_{t_0+j}] - \sum_{j=0}^{i-1}[a_2(j)+b_2(j)F_{t_0+j}]$$
$$- \sum_{j=i+1}^{y_{2,i}}[a_3(j|i)+b_3(j|i)F_{t_0+j}]\Big)\Bigg]$$
$$+ \mathbb{E}\Bigg[\exp\Big(-\sum_{i=0}^{y_{2,i}}[a_1(i)+b_1(i)F_{t_0+i}] - \sum_{i=0}^{y_{2,i}}[a_2(i)+b_2(i)F_{t_0+i}]\Big)\Bigg]. \tag{28}$$

These terms are lagged Laplace transform of the process (F_t) and can be calculated in explicit form by iterating the equation:

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

where $c = \frac{1-e^{-m}}{2m}\sigma^2$ and u is a nonnegative argument. Again, as for the model with deterministic factor, the computation is faster for the Markov model than for the semi-Markov model, since in the first case, we can factor out the term $\exp(-\sum_{j=0}^{y_{2,i}}[a_3(j|i)+b_3(j|i)F_{t_0+j}])$, which does not depend on i and both $f_2(y_2)$ and $S_2(y_2)$ can be calculated recursively.

Appendix 4 : Estimation results

A.4.1 Markov model with deterministic exponential factor

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 4: Estimation of the Markov model with deterministic exponential factor. 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 is the mortality intensity, when the unobserved heterogeneity of autonomy status is integrated out. Therefore, it is a weighted average of the intensity functions of the two subgroups: autonomous and non autonomous. Indeed, using the expression of the p.d.f. f_2 and of the

survivor function S_2, we have:

$$\begin{aligned}\lambda(y_2) &= \lambda_{2|1}(y_2)\frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &\quad + \lambda_2(y_2)\frac{e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \lambda_{2|1}(y_2)p(y_2) + \lambda_2(y_2)\Big(1-p(y_2)\Big), \end{aligned} \tag{29}$$

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:

$$\begin{aligned}p(y_2) &= \frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \frac{\mathbb{P}[0<Y_1<y_2, y_2<Y_2]}{\mathbb{P}[y_2<Y_2]} \\ &= \mathbb{P}[0<Y_1<y_2|Y_2>y_2]. \end{aligned} \tag{30}$$

This probability is 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 12), then by age (see Figure 13). 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 14). 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 plot also the evolution of the prevalence function $p(y_2, t_0)$ for different cohorts (see Figure 15).

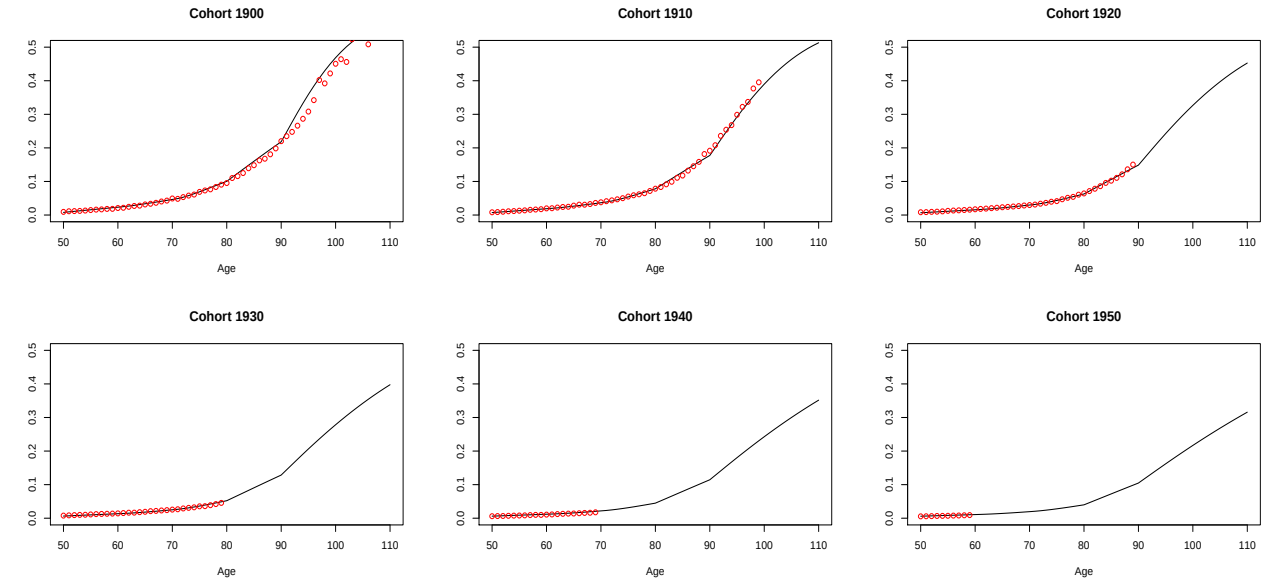


Figure 12: Fit of the observable mortality rates, for six different **cohorts**. Dotted line: historical data. Full line: the model (for both the past and future years). The x coordinate represents the age.

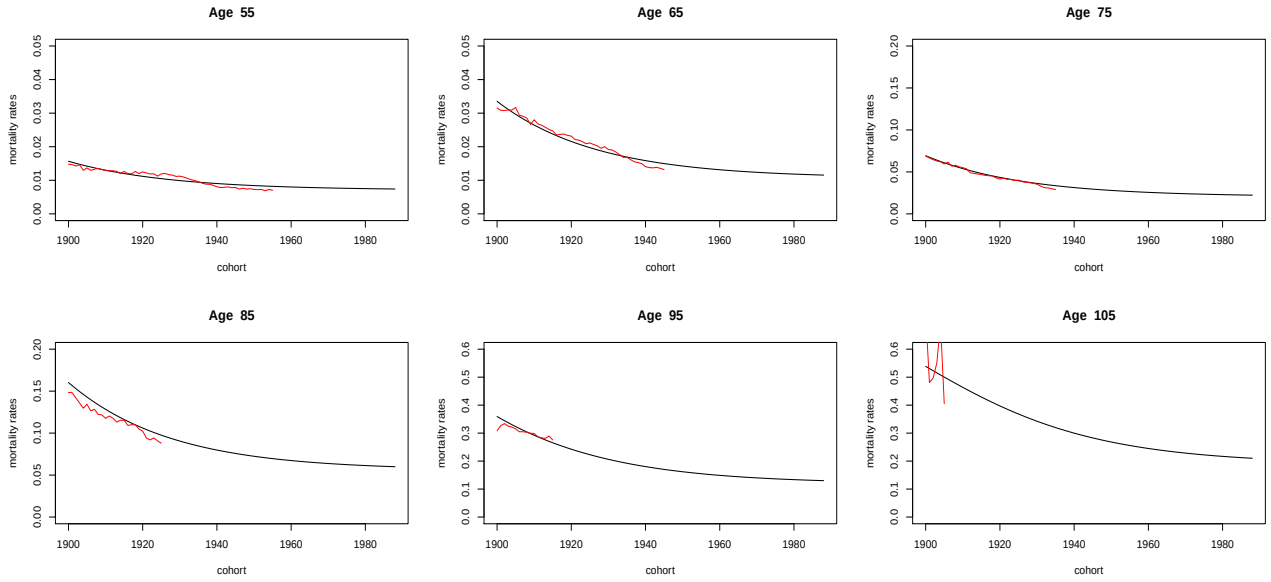


Figure 13: Fit of the observable mortality rates, for nine different **ages**. Dotted line: historical data. Full line: the model (for both the past and future years). The x coordinate represents the cohort.

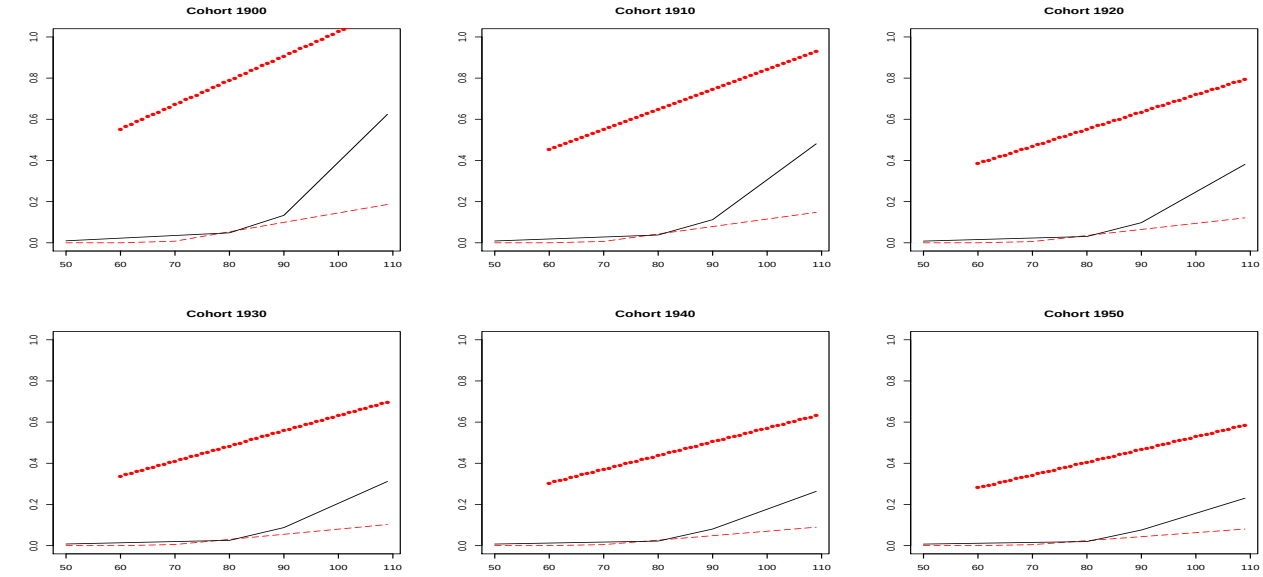


Figure 14: Evolution of the model based baseline hazard functions, respectively $\lambda_1(x)$ (for the intensity of entry, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\lambda_3(x)$ (mortality of person in LTC, dotted line).

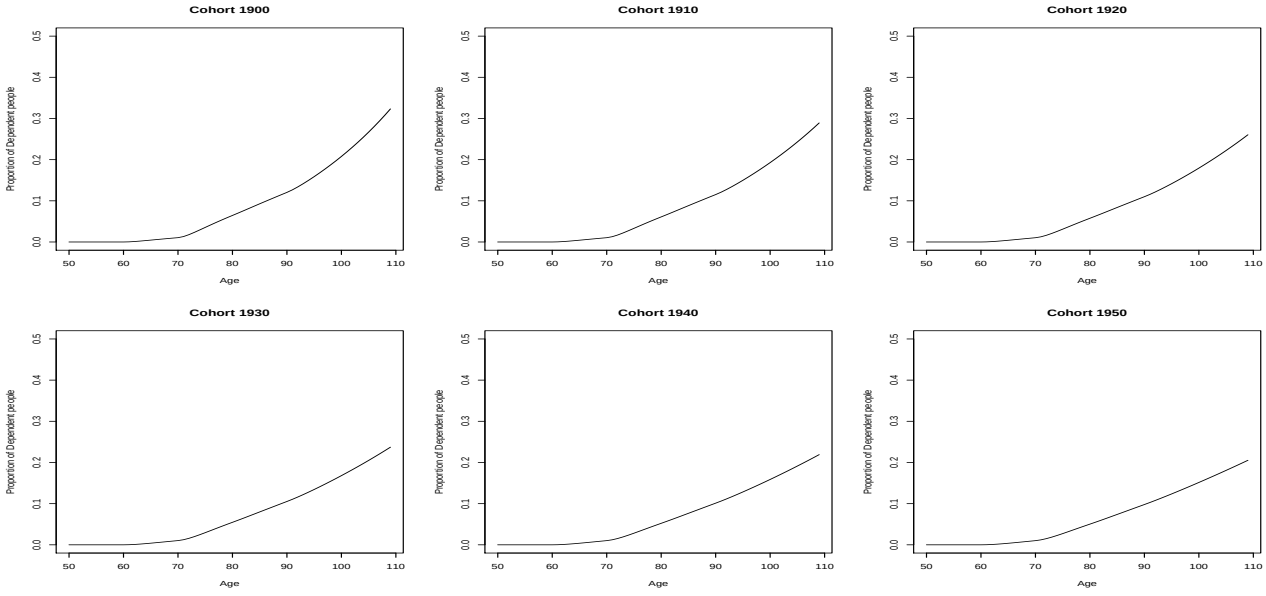


Figure 15: Evolution of the model based proportion of dependent people at a given age for each cohort.

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. This prevalence decreases in t_0 for each given age. This proportion reaches 10% at age 82, 85 and 88 for the following cohorts: 1900, 1920, 1940, respectively. This corresponds approximately to an increase of 1.8 months per annum for the age of entry into LTC to be compared with the 3-month increase for the cross-sectional life expectancy.

A.4.2 Semi-Markov model with deterministic exponential factor

As the previous Markov model, the parameter is estimated by maximizing the log-likelihood function. The estimated parameters are reported below:

Table 5: Estimation of the semi-Markov model with deterministic exponential factor; 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 16).

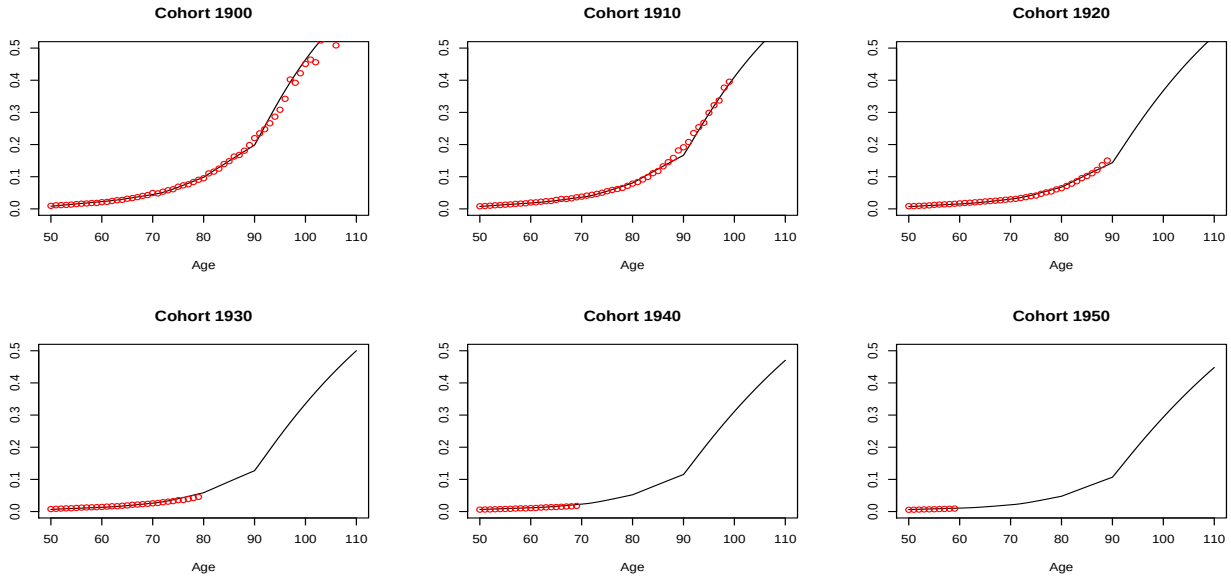


Figure 16: Fit of the observable mortality rates, for six different **cohorts**. Dotted line: historical data. Full line: the model (for both the past and the future years). The x coordinate represents the age.

The semi-Markov model provides also a very good fit. Then we plot (see Figure 17), 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:

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

Then we can check that equations (29) and (30) still hold when we replace $\lambda_{2|1}(y_2)$ by $\ddot{\lambda}_{2|1}(y_2)$. Figure 18 plots, for several cohorts, the evolution of the proportion of people in LTC.

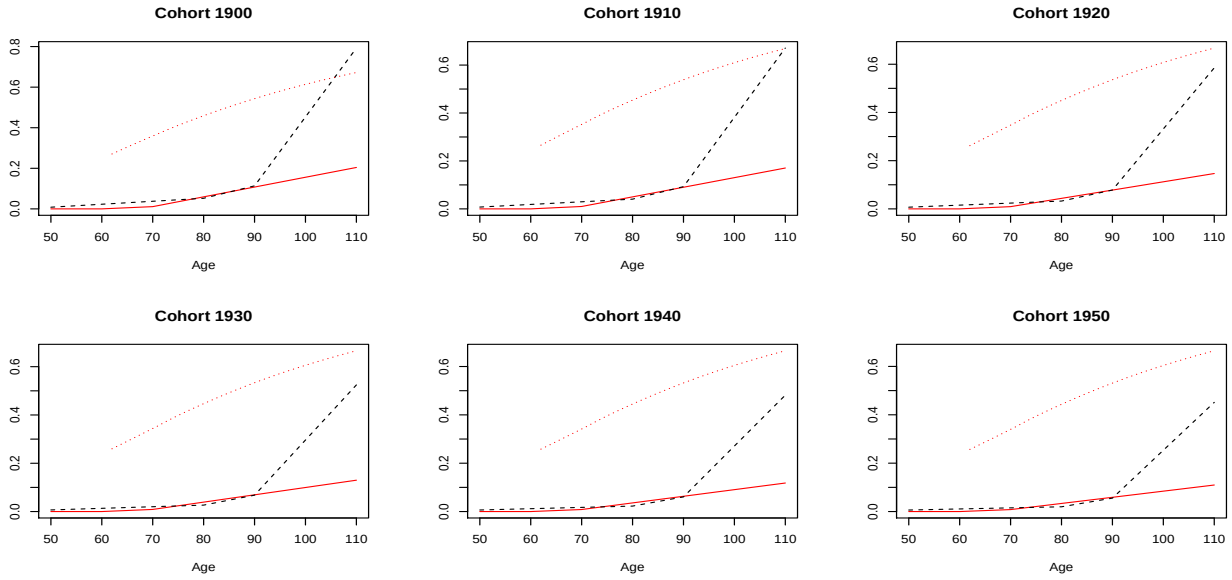


Figure 17: Evolution of the baseline hazard functions, respectively, $\lambda_1(x)$ (for the probability of entering into LTC, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\ddot{\lambda_3}$ (mortality of people in LTC, dotted line).

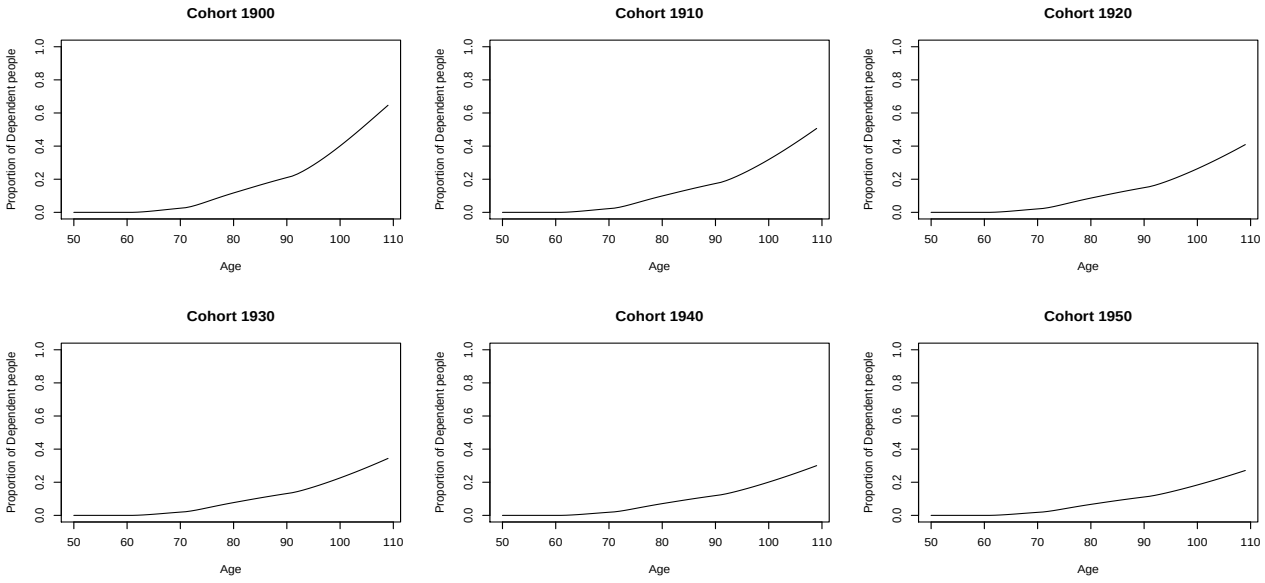


Figure 18: Evolution of model based proportion of people in LTC, for each cohort.

Appendix 5 : Properties of the latent CIR process

This section provides a brief summary of the properties of the CIR process satisfying:

$$dF_t = -mF_t + \sigma\sqrt{F_t}dW_t.$$

Lemma 3. *The stochastic differential equation (SDE) 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 [see e.g. 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. A random variable F follows a noncentered 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.

Definition 2. 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* (16) *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). □

Since $\delta = 0$, and Z is a Poisson variable, there is a non zero 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). For simplicity, let us denote the unobserved frailty process by $F = (F_1, F_2, ..., F_T)$ where T is the number of values of the dynamic factor process F.

A.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 $(\mathbb{R}^+)^T$. 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 let us 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 with deterministic factor.

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:

$$\begin{aligned}
&F_{1,k+1} \mid F_{2,k}, ..., F_{T,k}, Y_2, \theta \\
&F_{2,k+1} \mid F_{1,k+1}, F_{3,k}..., F_{T,k}, Y_2, \theta \\
&F_{3,k+1} \mid F_{1,k+1}, F_{2,k+1}, F_{4,k}, ..., F_{T,k}, Y_2, \theta \\
&\dots \\
&F_{T,k+1} \mid F_{1,k+1}, F_{2,k+1}, ..., F_{T-1,k+1}, Y_2, \theta
\end{aligned} \tag{31}$$

In other words, the chain is updated component by component, by drawing at each iteration in 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[34]. 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 (31), 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.

A.6.2 The Metropolis-Hasting algorithm

Now let us explain the Metropolis-Hasting algorithm we used 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:

$$p(F_t \mid F_{(-t)}, Y_2, \theta) \propto \mathcal{L}(\theta \mid Y_2, F) p(F_t \mid F_{t-1}, \theta) p(F_t \mid F_{t+1}, \theta). \tag{32}$$

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

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 the integrand in the right hand side of equations (27) and (28), respectively. 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. Therefore the second term is equally easy to calculate. Thus the density function given by (32) 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 (32). 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:

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

2. For $n = 2, 3, ...$, draw a candidate from a proposal distribution, for instance, we can choose the log-normal distribution[35]:

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

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

3. Compute

$$\alpha = \frac{p(F_t = f \mid F_{(-t)}, Y_2, \theta)}{p(F_t = F^{(n-1)}_t \mid F_{(-t)}, Y_2, \theta)}, \tag{33}$$

   where both the numerator and the denominator can be calculated by equation (32).

4. Draw a uniform variable $u \sim U([0, 1])$ and set the $n-$th value F^n_t by the following rule: [36]

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

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

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

To ensure the convergence of this univariate Markov chain to its stationary distribution (32), 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.

Appendix 7 : Identification proof of Proposition 3

Remind that with a deterministic exponential factor, the age-cohort and age-calendar time models are equivalent and that the survivor function for cohort t_0 is given by:

$$\begin{aligned} S_2(y_2,t_0) = \int_0^{y_2} [a_1(x)+\tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s)+\tilde{b}_1(s)F_{t_0}]ds \\ - \int_0^x [a_2(s)+\tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s)+\tilde{b}_3(s)F_{t_0}]ds\Big)dx \\ + \exp\Big(- \int_0^{y_2} [a_1(x)+\tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_2} [a_2(x)+\tilde{b}_2(x)F_{t_0}]dx\Big). \end{aligned} \tag{34}$$

A.7.1 Identification of m.

When $y_2 \to 0$, we have, for $t_1 \neq t_0 \neq t_2 \neq t_1$,

$$\lim_{y_2 \to 0} \frac{\lambda(y_2,t_2)-\lambda(y_2,t_0)}{\lambda(y_2,t_1)-\lambda(y_2,t_0)} = \frac{e^{-mt_2}-e^{-mt_0}}{e^{-mt_1}-e^{-mt_0}}.$$

Since the LHS in the equation above is observable, m is point identified. Note that the identification assertion remains valid even for a general functional parameter (F_t) without the exponential specification, under the limiting longevity assumption $\lim_{t\to\infty} F_t = 0$. Indeed, under this assumption the ratio $\frac{F_{t_2}-1}{F_{t_1}-1}$ is identified, where we remind that $F_{t_0} = 1$. If there is another path (F'_t) such that

$$\frac{F_{t_2}-1}{F_{t_1}-1} = \frac{F'_{t_2}-1}{F'_{t_1}-1},$$

then $\frac{F_{t_2}-1}{F'_{t_2}-1} = \frac{F_{t_1}-1}{F'_{t_1}-1}$ is equal to a constant that does not depend on t_2, t_1. Let t_2 go to infinity, by using the limiting condition $\lim_{t\to\infty} F_t = 0$, we deduce that this constant equals 1. Thus the path of the process (F_t) is nonparametrically identified. Moreover, the following identification of functional parameters remains valid for a general form of (F_t) and the age-cohort specification, but not the age-calendar time model, except with the exponential specification. See also the discussion in Section 4.2.2.

A.7.2 Identification of functional parameters $a_1, a_2, a_3, b_1, b_2, b_3$.

Under the assumption that all functions[37] $a_1, b_1, a_2, b_2, a_3, b_3$ are continuous, the conditional survivor function $S(y_2|t_0) = S(y_2|F)$ is an analytic function of F for a given y. Therefore it is equivalent to know this function or to know all its derivatives for any pre-specified t_0. These derivatives are simpler to deal with, especially if $t_0 = \infty$; equivalently we look at the derivative at $F = 0$. The case $t_0 < \infty$ is similar[38]. Thus we obtain, at order 0,

$$\int_0^{y_2} a_1(x) e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} dx + e^{-A_1(y_2)-A_2(y_2)} = S(y_2, F = 0), \tag{35}$$

and at each order $n \geq 1$,

$$\begin{aligned}
&\int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} \left(\frac{a_1(x)(-1)^n}{n!} \big[B_1(x) + B_2(x) + B_3(y_2) - B_3(x)\big]^n \right. \\
&\quad \left. + \frac{b_1(x)(-1)^{n-1}}{(n-1)!} \big[B_1(x) + B_2(x) + B_3(y_2) - B_3(x)\big]^{n-1} \right) dx \\
&\quad + e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} \Big[B_1(y_2) + B_2(y_2)\Big]^n = (-1)^n \frac{\partial S}{\partial F}(y_2, F = 0),
\end{aligned} \tag{36}$$

for all $y_2 \in [0, T]$, where the capital letters denote the cumulative integrals of the corresponding lower case functions.

Except in some special cases, one expects that (35) and (36) give a non degenerated infinite system of functional equations that $a_1, b_1, a_2, b_2, a_3, b_3$ should satisfy. This raises hopes that the solution to such a system is generically unique. Let us first look at Case 1 in Proposition 3.

Case 1 (global identification). If $b_1 + b_2 = b_3$, the $n-$th equation becomes:

$$\begin{aligned}
&B_3(y_2)^{n-1} \int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} \left(\frac{a_1(x)(-1)^n}{n!} B_3(y_2) - \frac{b_1(x)(-1)^{n-1}}{(n-1)!} \right) dx \\
&\quad + e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} B_3(y_2)^n = (-1)^n \frac{\partial S}{\partial F}(y_2, F = 0).
\end{aligned}$$

For $y > 0$, $B_3(y_2) > 0$, and large n, the LHS of the equation above is equivalent to:

$$B_3(y_2)^{n-1} \frac{(-1)^n}{n!} \int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} b_1(x) dx.$$

[37] Strictly speaking, the functional parameters are $a_1, a_2, a_3, \tilde{b_1}, \tilde{b_2}, \tilde{b_3}$. For ease of exposure, we omit the tilde symbol on b_1, b_2, b_3 for the rest of this section.

[38] If $t_0 < \infty$, we should look at the sequence of derivatives of the function $S_2(y_2|F_{t_0})$ for any given y_2 at the point $F_{t_0} \neq 0$. Their expressions are more complicated than at point $F_{t_0} = 0$.

Therefore $B_3(y_2)$ is globally identified[39], as well as the constant (in n) $\int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} b_1(x)dx$. Then by suppressing this dominating term, the LHS of the previous $n-$th equation reduces to the LHS in (35). Thus the infinite system reduces to only three independent equations and the model is not identified.

Case 2 (global identification). There exists constants $c, c' > 0$ such that $b_1 + b_2 - b_3 \geq c$ and $|b_2 - b_3| > c'$. For expository purpose let us introduce the following functions:

$$\begin{aligned}
C(y) &= e^{-A_1(y)-A_2(y)}, \\
D(y) &= B_1(y) + B_2(y), \\
f_n(y) &= \int_0^y e^{-A_1-A_2-A_3(y)+A_3} b_1 \Big[B_1 + B_2 + B_3(y) - B_3\Big]^n dx, \\
g_n(y) &= \int_0^y e^{-A_1-A_2-A_3(y)+A_3} a_1 \Big[B_1 + B_2 + B_3(y) - B_3\Big]^n dx.
\end{aligned}$$

Since $B_1(x)+B_2(x)+B_3(y)-B_3(x)$ is positive, increasing in x, and the term $e^{-A_1-A_2-A_3(y)+A_3} b_1$ is positive and bounded, we can prove that[40]:

$$(n+1) f_n(y) \sim \frac{e^{-A_1(y)-A_2(y)} b_1(y)}{b_1(y) + b_2(y) - b_3(y)} \Big[B_1(y) + B_2(y)\Big]^{n+1} \tag{37}$$

when n goes to infinity and

$$(n+1) g_n(y) \sim \frac{e^{-A_1(y)-A_2(y)} a_1(y)}{b_1(y) + b_2(y) - b_3(y)} \Big[B_1(y) + B_2(y)\Big]^{n+1}.$$

[39] By global identification, we refer to the standard definition of identification, that is, a function is identified if at any point y_2, the value of this function is uniquely determined. This notion has to be distinguished from the concept of local (nonparametric) identification, as in Chen et al. (2014), detailed later on in the proof.

[40] Intuitively, when n becomes large, the contribution of the integrand at a point x that is away from y_2 is negligible since $\Big[B_1(x) + B_2(x) + B_3(y) - B_3(x)\Big]^n$ is much smaller than $\Big[B_1(y) + B_2(y)\Big]^n$. Thus the asymptotic behavior of this integral depends only on the behavior of the integrand in a neighbourhood of point y. To get another informal explanation of this result, we can use the integration by parts:

$$\begin{aligned}
(n+1) f_n(y) = &\frac{e^{-A_1(y)-A_2(y)} b_1(y)}{b_1(y) + b_2(y) - b_3(y)} \Big[B_1(y) + B_2(y)\Big]^{n+1} - \frac{e^{-A_3(y)} b_1(0)}{b_1(0) + b_2(0) - b_3(0)} B_3(y)^{n+1} \\
&- \int_0^y \frac{\partial}{\partial x} \Big(\frac{e^{-A_1-A_2-A_3(y)+A_3} b_1}{b_1 + b_2 - b_3} \Big) \Big[B_1 + B_2 + B_3(y) - B_3\Big]^{n+1} dx
\end{aligned}$$

Since $\frac{B_1(y)+B_2(y)}{B_3(y)} > 1$, the second term is negligible with respect to the first one; if the partial derivative in the third term exists and is bounded, then the third term is $O(f_n(y))$ when n goes to infinity. By rearranging this equation, we get the desired asymptotic equivalent. The formal proof of this result uses solely real analysis techniques and does not requires the existence of the partial derivative which is needed the integration by parts.

Then we can study the behavior of the LHS of (36). We have:

$$(-1)^n \frac{\partial S}{\partial F}(y_2, F=0) = \frac{(-1)^n}{n!} g_n(y) + \frac{(-1)^{n-1}}{(n-1)!} f_{n-1}(y) + e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} \Big[B_1(y_2) + B_2(y_2)\Big]^n$$
$$\sim e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} \Big[B_1(y_2) + B_2(y_2)\Big]^n \Big(1 - \frac{b_1(y)}{b_1(y) + b_2(y) - b_3(y)}\Big)$$

provided that $b_2 - b_3$ is never null. Then $B_1(y) + B_2(y)$ is globally identified, as well as the function $\frac{e^{-A_1(y)-A_2(y)}\big(b_2(y)-b_3(y)\big)}{b_1(y)+b_2(y)-b_3(y)}$.

Case 3 (global identification). Similarly, if there exists $d > 0$ such that $b_1 + b_2 - b_3 \leq -d$, then $(n+1)f_n(y) \sim \frac{e^{-A_3(y)} b_1(0)}{b_1(0)+b_2(0)-b_3(0)} B_3(y)^n$. $B_3(y)$ is globally identified, as well as $e^{-A_3(y)}$, up to an additive constant. Since $A_1(0) = 0$, the constant is uniquely determined. Therefore A_3 is identified as well.

Cases 2,3 (local identification). Let us finally prove that the other functions are locally identified. We do this by following Chen et al. (2014), who give the definition of local identification on a functional space. Roughly speaking, a function h is locally identified at h_0, if h_0 is the unique solution to a certain system of equations when the unknown function is restricted to be in a certain neighbourhood of h_0 [see Definition 1, Chen et al. (2014)]. In our case the neighborhood has to be defined on an appropriate functional space and we have to find a functional operator, whose Gâteaux derivative is non degenerated. This is the infinite dimensional analogue of the standard full rank condition for local identification of parametric models. As explained in Chen et al. (2014), on the contrary to the finite dimensional case where the rank condition is also sufficient, in an infinite dimensional space, this condition alone implies only a rather weak notion of local identification [see Theorem 2, Chen et al. (2014)].

For expository purpose, let us focus on Case 3. For Case 2, the calculations are slightly more complicated, but the principle stays the same.

Let us denote by $\mathcal{B} = \mathcal{C}([0,T])$ the space of all continuous functions on the age domain $[0,T]$, where T is a fixed constant, that is, we assume that the observations are only available up to a maximum age, say, $T = 110$. We have deliberately chosen a fixed upper bound[41] so that the functional space $\mathcal{B}$, topologized by the uniform norm $||f|| = \max_{t\in[0,T]} |f(t)|$, is a Banach space. This fixed upper bound is not restrictive since, if we can prove local identification for any given T, then we will have local identification on the whole age domain $[0,\infty[$. Also remind that on the

[41] The assumption of a fixed upper bound for the observable attained age is compatible with the previous assumption $t_0 = \infty$, on the observed cohort.

space $\mathcal{B}$, all functions are bounded, and all positive functions are lower bounded by a positive constant. Under this framework, we have the following Lemma, which is a direct consequence of Theorem 2 in Chen et al. (2014):

Lemma 5. *The functions (a_1, b_1, a_2, b_2) are locally identified in the sense of Theorem 2 in Chen et al. (2014) if the following four conditions are satisfied:*

i) For each $n \geq 1$, the LHS of (36) is a continuous operator from $\mathcal{A} := \mathcal{B}^4$ to space $\mathcal{B}$, with the corresponding uniform topology for each space. These operators are denoted $m_n : \mathcal{A} \mapsto \mathcal{B}$.

ii) For each order $n \geq 1$, the operator m_n is Fréchet differentiable [see e.g. Chen et al. (2014) Equation 2.1]. For each element $\alpha \in \mathcal{A}$ we denote by $h \mapsto m'_n(h)$ the Fréchet derivative at point α, where h is the generic element of the space $\mathcal{A}$, $n \geq 0$. This derivative depends on the point $\alpha \in \mathcal{A}$, but we will omit the index α.

iii) The intersection of the null spaces $\cap_{n=1}^{\infty} Ker\, m'_n$ is reduced to $\{\mathbf{0}\}$.

These conditions are quite intuitive. Condition *i)* is a regularity condition at both infinity (since $y_2 \leq T < \infty$) and zero (since the integrands are all bounded at zero). It excludes in particular mixed proportional hazard (MPH) models with heavy-tailed unobserved heterogeneity distribution and an intensity function that is equal to infinity at time zero [see e.g. Ridder (1990)]. Condition *ii)*, that is the differentiability of these operators, is clearly satisfied, since each operator is a compounding of elementary (Gâteaux−) differentiable operators. Condition *iii)* is Assumption 1 in Chen et al. (2014), and is the infinite dimensional analogue of the full rank condition.

Let us give the proof of the lemma. For given (a_3, b_3) as well as path of (F_t), the survivor function $S(y|F)$ is a bivariate continuous function in arguments (y, F), that is $S(y|F) \in \mathcal{C}([0, T] \times [0, 1])$ which is a Banach space. Denote by M the operator from $\mathcal{A}$ to $\mathcal{C}([0, T] \times [0, 1])$, which maps the point (a_1, b_1, a_2, b_2) to the corresponding survivor function $S(y|F)$. Then by Chen et al. (2014), it suffices to prove that M', the Gâteaux derivative of M is nonsingular[42]. Because $S(y|F)$ (as well as its Gâteaux derivative) is analytical in F, $M'(y, F) = 0$ is equivalent to the derivatives of any order with respect to F being null functions. These derivatives are exactly[43] the sequence m'_n.

[42] $M'(\big(da_1, db_1, da_2, db_2)\big)$ is a bivariate function in arguments y and F.

[43] We have used the fact that it is equivalent to first take derivative with respect to F, then the Gâteaux derivative with respect to (a_1, b_1, a_2, b_2) or conversely.

Let us finally check that Condition *iii)* is satisfied in our framework. The expression of $m_n'(h)$ at point (a_1, b_1, a_2, b_2), for any $h = (da_1, db_1, da_2, db_2) \in \mathcal{A}$, is the following:

$$
\begin{aligned}
&(-1)^n(n-1)!m_n'(h)(y)\\
&= \int_0^y e^{-A_1-A_2-A_3(y)+A_3}\frac{a_1}{n}\Big[B_1+B_2+B_3(y)-B_3\Big]^n\Big[-dA_1-dA_2\Big]dx\\
&\quad+\int_0^y e^{-A_1-A_2-A_3(y)+A_3}\frac{da_1}{n}\Big[B_1+B_2+B_3(y)-B_3\Big]^n dx\\
&\quad+\int_0^y e^{-A_1-A_2-A_3(y)+A_3}a_1\Big[B_1+B_2+B_3(y)-B_3\Big]^{n-1}\Big[dB_1+dB_2\Big]dx\\
&\quad-\int_0^y e^{-A_1-A_2-A_3(y)+A_3}b_1\Big[B_1+B_2+B_3(y)-B_3\Big]^{n-1}\Big[-dA_1-dA_2\Big]dx\\
&\quad-\int_0^y e^{-A_1-A_2-A_3(y)+A_3}db_1\Big[B_1+B_2+B_3(y)-B_3\Big]^{n-1}dx\\
&\quad-\int_0^y e^{-A_1-A_2-A_3(y)+A_3}(n-1)b_1\Big[B_1+B_2+B_3(y)-B_3\Big]^{n-2}\Big[dB_1+dB_2\Big]dx\\
&\quad+e^{-A_1(y)-A_2(y)}\frac{\Big[B_1(y)+B_2(y)\Big]^{n-1}}{n}\Bigg(-\Big[dA_1(y)+dA_2(y)\Big]\Big[B_1(y)+B_2(y)\Big]+n\Big[dB_1(y)+dB_2(y)\Big]\Bigg),
\end{aligned}
\tag{38}
$$

where $dA_1, dA_2, ..$ are cumulative integral of the corresponding lower case functions. Let us explain this formula: lines 1-3 (resp. lines 4-6 and line 7) are the Gâteaux derivatives of the first (resp. second and third) term of the LHS of (36).

Assume now that $m_n'(h) = 0$ for a certain function $h = (da_1, da_2, db_1, db_2)$ and for all $n \geq 0$. Similarly as (37), when n goes to infinity, $m_n'(h)$ is equivalent to:

$$
-\frac{C(y)}{b_1(0)+b_2(0)-b_3(0)}\Big(b_1(y)dB_1(y)-(b_2(y)-b_3(y))dB_2(y)\Big)B_3^{n-1}(y)
$$

provided that this term is non null. Thus we should have:

$$
b_1(y)dB_1(y)-(b_2(y)-b_3(y))dB_2(y)=0 \tag{39}
$$

for all y. Then similarly, $m_n'(h)$ is equivalent to:

$$
-\frac{C(y)}{(n-1)\Big[b_1(0)+b_2(0)-b_3(0)\Big]}\Big(a_1(y)dB_1(y)+a_2(y)dB_2(y)+b_1(y)dA_1(y)-(b_2(y)-b_3(y))dA_2(y)\Big)B_3^n(y),
$$

provided that this term is non null. Therefore:

$$a_1(y)dB_1(y) + a_1(y)dB_2(y) + b_1(y)dA_1(y) - (b_2(y) - b_3(y))dA_2(y) = 0. \tag{40}$$

Similarly, we have

$$a_1(y)dA_1(y) - a_1(y)dA_2(y) = 0, \tag{41}$$

and finally

$$-\frac{C(y)D^n(y)}{n}dA_2(y) + C(y)D^{n-1}(y)dA_1(y) = 0. \tag{42}$$

Combining (39) to (42) we can get $dA_1 = dA_2 = 0$, and then we have $dB_1 = dB_2 = 0$, except when:

$$\frac{b_3 - b_2}{b_1} = \frac{a_1}{a_1} = 1,$$

which is not allowed since Case 3 assumes $b_1 + b_2 - b_3 < 0$. Thus Condition *iii*) above is satisfied.

References

Abbring, J. and Van den Berg, G. (2003a). The Identifiability of the Mixed Proportional Hazards Competing Risks Model. *Journal of the Royal Statistical Society: Series B*, 65(3):701–710.

Abbring, J. and Van den Berg, G. (2003b). 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.

Andersen, P. 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.

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014). Local Identification of Nonparametric and Semiparametric Models. *Econometrica*, 82(2):785–809.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407.

Dienst, H.-R. (1972). Gedanken zum Konjunkturrisiko in der Berufsunfähigkeits-Versicherung. *Blätter der DGVFM*, 10(4):521–531.

Dionne, G. (1982). Moral Hazard and State-Dependent Utility Function. *Journal of Risk and Insurance*, pages 405–422.

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

Duffie, D., Pan, J., and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6):1343–1376.

Elbers, C. and Ridder, G. (1982). True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *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.

Finkelstein, A. and McGarry, K. (2006). Multiple Dimensions of Private Information: Evidence From the Long-Term Care Insurance Market. *American Economic Review*, 96(4):938–958.

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

Gagliardini, P. and Gouriéroux, C. (2014). *Granularity Theory with Applications to Finance and Insurance*. 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*.

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.

Heckman, J. and Robb, R. (1985). Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations. In *Cohort Analysis in Social Research*, pages 137–150. Springer.

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.

Karni, E. (1983). Risk Aversion for State-Dependent Utility Functions: Measurement and Applications. *International Economic Review*, pages 637–647.

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

Keilman, N., Pham, D. Q., and Hetland, A. (2002). Why Population Forecasts Should be Probabilistic–Illustrated by the Case of Norway. *Demographic Research*, 6(15):409–454.

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. *OECD Health Working Paper*.

Lee, R. and Carter, L. (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.

Majer, I. M., Stevens, R., Nusselder, W. J., Mackenbach, J. P., and van Baal, P. H. (2013). Modeling and Forecasting Health Expectancy: Theoretical Framework and Application. *Demography*, 50(2):673–697.

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

McWhinnie, J. R. (1980). Disability Assessment in Population Surveys: Results of the OECD Common Development Effort. *Revue d'Épidémiologie et de Santé Publique*, 29(4):413–419.

Murtaugh, C. M., Spillman, B. C., and Warshawsky, M. J. (2001). In Sickness and in Health: An Annuity Approach to Financing Long-Term Care and Retirement Income. *Journal of Risk and Insurance*, 68(2):225–254.

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.

Ridder, G. (1990). The Non-Parametric Identification of Generalized Accelerated Failure-Time Models. *The Review of Economic Studies*, 57(2):167–181.

Spillman, B. C. and Lubitz, J. (2002). New Estimates of Lifetime Nursing Home Use: Have Patterns of Use Changed? *Medical Care*, 40(10):965–975.

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

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

Wouterse, B., Huisman, M., Meijboom, B. R., Deeg, D. J., and Polder, J. J. (2013). Modeling the Relationship Between Health and Health Care Expenditures Using a Latent Markov Model. *Journal of Health Economics*, 32(2):423–439.

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

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