Coherent extrapolation of mortality rates at old ages applied to Long Term Care

Léonie Le Bastard

Coherent extrapolation of mortality rates at old ages applied to Long Term Care

Léonie Le Bastard[*], SCOR[†] and SAF laboratory[‡]

2023/06/15

Abstract

In an insurance context, Long-Term Care (LTC) products cover the risk of permanent loss of autonomy, which is defined by the impossibility or difficulty of performing alone all or part of the activities of daily living (ADL). From an actuarial point of view, knowledge of risk depends on knowledge of the underlying biometric laws, including the mortality of autonomous insureds and the mortality of disabled insureds. Due to the relatively short history of LTC products and the age limit imposed at underwriting, insurers lack information at advanced ages. This represents a challenge for actuaries, making it difficult to estimate those biometric laws.

In this paper, we propose to complete the missing information at advanced ages on the mortality of autonomous and disabled insured populations using information on the global mortality of the portfolio. In fact, the three previous mortality laws are linked since the portfolio is composed only of autonomous and disabled policyholders. We model the two mortality laws (deaths in autonomy and deaths in LTC) in a Poisson Generalized Linear Model framework, additionally using the P-Splines smoothing method. A constraint is then included to link the mortality laws of the two groups and the global mortality of the portfolio. This new method allows for estimating and extrapolating both mortality laws simultaneously in a consistent manner.

Keywords: Long-Term Care Insurance; Actuarial Modelling; Generalized Linear Models; P-Splines; Extrapolation; Penalization

1 Introduction

Long-Term Care (LTC) is linked to the risk that an individual loses their autonomy, resulting in the impossibility or difficulty of performing Activities of Daily Living (ADL), such as washing, eating, moving and dressing. Many causes can lead to a loss of autonomy, but the need for LTC

[*]llebastard@scor.com

[†]SCOR, 5 avenue Kléber, 75795 Paris Cedex 16, France

[‡]Univ Lyon, Université Claude Bernard Lyon 1, Laboratoire de Sciences Actuarielle et Financière, Institut de Science Financière et d'Assurances (50 Avenue Tony Garnier, F-69007 Lyon, France)

is mostly due to illness occurring at old ages. With the persistent increase in life expectancy and the ageing of the Baby Boom generation, we are entering a period in which the number of people over 80 is likely to continue to grow. Eurostat (2022) estimates that the share of the population in Europe aged 80 years or above is likely to be multiplied by two and a half between 2021 and 2100. Therefore, it is expected that an increasing number of people will need financial support to cover the costs generated by the loss of autonomy. The average age at underwriting of an LTC product is approximately 60, while claims mainly occur after 85. This average underwriting age means for the insurer that only a few observations on the mortality of his portfolio at old ages will be available before the 25th anniversary of the product. This effect combined with the recency of LTC products makes it difficult to estimate the associated risk. Improving knowledge of the LTC risk is then a challenge for actuaries. In contrast, the mortality of the overall population was studied long before the emergence of LTC insurance products.

From an actuarial point of view, the insured's health condition is often represented by an illness-death model composed of three states, namely, "Autonomous", "Disabled" and "Dead". Some insurance contracts cover multiple levels of dependency with different levels of annuity. Actuaries may model these products with multi-state Markov models, with one state for each level of dependency. This choice multiplies the number of laws of transition from one state to another to estimate. The difficulty of calibrating multi-state Markov models comes from the scarcity of data. Most insurers do not observe enough transition from LTC states in their database due to the recency of their product. Therefore, papers modelling LTC products with multiple levels of dependency often make strong assumptions on the intensities of the model, or use big public data as in Biessy (2015) with data from the French LTC public aid called the "Allocation Personnalisée d'Autonomie" (APA). Fleischmann (2015) models an Austrian private health insurance product with 7 levels of severity and assumes that mortality is the same independently from the severity level and that the intensity to reach a severity level is independent of the state of origin and the time spent in that state. In addition to using public data, Biessy (2015) uses parametric laws. Another solution to model products covering multiple states of disability is to consider it as a set of illness-death models. This method is the solution mainly used by insurers. The method developed in this paper can therefore be used to model products covering multiple states of disability.

Since the state "Dead" is an absorbing state, the model is composed of 4 transitions, each one associated with a biometric law. The first one corresponds to the incidence rates in the disabled state, whereas the second one represents its reverse transition. The two remaining biometric laws correspond to distinct mortality rates for autonomous and disabled lives. In practice it is very hard to reverse loss of autonomy. Recovery probabilities are negligible, especially when the "Disabled" state is associated with a high level of dependency. The definition of LTC in France emphasizes the fact that the loss of autonomy must be permanent and irreversible. As in most of LTC product contracts, we consider in this paper that the loss of autonomy is final, which means that no return to autonomy is envisaged. With this hypothesis, which is representative of the real insurance market,

only 3 biometric laws need to be estimated. The impact of allowing recovery when taking into account a low level of disability is discussed in Section 6.

Since LTC risk is mostly due to ageing pathologies, the estimation of mortality at old ages is of importance for pricing and estimation of risk liabilities. In the context of mortality modelling, a common approach is to fit a parametric model on the crude death rates and assume that information available at younger ages would explain the behaviour at older ages, where we have no or not enough observations. Depending on the selected parametric model, a different underlying assumption on the shape of the mortality curve is made. Some of these models are compared in Hammond (2000) on 13 countries (European, Scandinavian and Japanese) using data from 1960 to 1990. A different approach is used in this paper to extrapolate mortality at old ages, relying on the P-Splines smoothing method introduced by Eilers and Marx (1996). This methodology was adapted to mortality for the first time by Currie et al. (2004).

The incidence rates and mortality rates of autonomous and disabled insureds are usually estimated and extrapolated independently. However, in this way, the consistency between the mortality laws is not guaranteed, and the predicted number of deaths in the whole portfolio might differ from the sum of the predicted numbers of deaths in autonomy and in LTC. Let D_x^G be the number of observed deaths between age x and $x+1$ in group $G \in \{A, D, gen\}$ and $\widehat{D}_x^G$ its predicted value, where A and D represent the groups of autonomous and disabled insureds, respectively, and gen represents the overall portfolio of insureds. Then, $D_x^{gen} = D_x^A + D_x^D$, and in the case of consistent mortality laws, the relation between the expected values must be given by the following equation

$$\widehat{D}_x^{gen} = \widehat{D}_x^A + \widehat{D}_x^D. \tag{1}$$

In the literature, the problem of consistency between mortality laws is mostly approached in the context of prospective modelling to ensure that the mortality laws do not diverge indefinitely over time between several groups. This idea of coherent mortality forecasting was first introduced by Li and Lee (2005). Li proposed a method based on the Lee-Carter model to forecast the mortality of a group of populations by allowing each population to have its own age pattern and level but have a common trend. Later, Zhou et al. (2019) and Li et al. (2017) approached this problem of coherent mortality forecasting with the concept of semicoherence. The idea is to fix a weaker assumption on the coherence between the mortality laws by allowing the mortality trajectories of two populations to diverge, as long as the difference between the two mortality laws does not exceed what they called a tolerance corridor. Noticing that the coherent assumption can be too strong, especially when it is imposed on a large number of populations, Guibert et al. (2020) proposed a new approach based on locally coherent mortality forecasts by assuming that the coherence principle is verified by subgroups of populations.

This paper aims to develop a method that improves the estimation and extrapolation of the mortality laws of autonomous and disabled groups, using knowledge on the mortality of the overall population

(union of the 2 groups) while keeping a smooth structure of the mortality laws. To this end, we use the P-Splines smoothing method proposed in Eilers and Marx (1996) and adapted to mortality estimation in Currie et al. (2004) and Macdonald et al. (2018). A constraint based on Equation (1) is then included to link the mortality laws of the two groups and the global mortality of the portfolio. The idea of adding constraints to the P-Splines method has already been used in Bollaerts et al. (2006) for research on the cognitive development of children and for mortality modelling in Camarda et al. (2016), Remund et al. (2018) and in Camarda (2019). The method proposed in this paper uses an algorithm converging under certain conditions discussed in Appendix A. The goal of this paper is to provide an algorithm for actuaries in charge of the pricing of LTC products. This paper is not intended to present limit theorems of convergence of estimators because they are difficult to obtain. The simulations support the interest of the method.

We show that this approach leads to a better estimate of the death rates for both autonomous and disabled insureds, providing an estimate of the predicted number of deaths of the overall population close to the sum of the predicted number of deaths in autonomy and LTC.

Section 2 of this paper introduces the dataset and the model. In Subsection 2.2, we present the continuous multi-state Markov model used in the context of LTC modelling and explain its link with the Poisson model and the Poisson generalized linear model (Poisson-GLM). The P-Splines smoothing method, used to maintain a smooth structure of the mortality laws, is proposed in Subsection 2.3. We add a constraint on the consistency between mortality laws based on Equation (1) in Subsection 2.4.

Section 3 focuses on the extrapolation of the mortality laws when no exposures are available at old ages. We propose an extrapolation method with reconstruction of the exposures using the model proposed in the first part of the paper. Section 4 addresses the problem of the choice of the hyper-parameter corresponding to the weight that we give to the consistency constraint. The larger this parameter is, the better the mortality laws estimated by the algorithm satisfy the coherence rule. An application on a real dataset is made in Section 5. Recovering from a high level of dependency can easily be assumed as impossible. However, one might wish to assume that recovery is possible when modelling low levels of LTC. Section 6 discusses how to use the loopback algorithm to model a product covering several levels of LTC, especially when the lower levels allow recovery. Concluding remarks are provided in Section 7.

2 Modelling

2.1 Data structure

The biometric laws are calibrated to an LTC portfolio observed in a given period. The trajectories of insured individuals, meaning their health status at each time of the period of observation, are observed. For each insured individual, the following information is available:

- date of birth,
- gender,
- underwriting date,
- date of loss of autonomy if it occurred,
- date of death if it occurred, and
- date of exit from the portfolio in case of a contract cancellation.

Since males and females do not have the same mortalities or same probabilities of loss of autonomy at each age, biometric laws are estimated separately by gender. The information of all the insureds is then aggregated to construct two databases per gender. The first one, denoted by DB^A, is used to study the autonomous experience, and the second one, denoted by DB^D, is used to study the experience in LTC. For each integer age x, DB^A and DB^D contain:

- the central exposure to risk between age x and $x+1$,
- the number of observed deaths between ages x and $x+1$, and
- the number of reported losses of autonomy (only for DB^A) between age x and $x+1$.

For a given age x, the central exposure to risk in autonomy (resp. LTC) corresponds to the sum of individual exposures of each insured in autonomous (resp. LTC) state between x and $x+1$. The individual exposure of an insured at age x in a given state (autonomous or LTC) is the fraction of time spent in this state between age x and $x+1$. For example, an insured in autonomous state at his 65th birthday, losing his autonomy 9 months after, and surviving at least until his 66th birthday in LTC has an exposure in the autonomous state equal to 0.75, which corresponds to 3/4 of a year, and an exposure in LTC equal to 1/4.

2.2 Modelling of a Long-Term Care product

We consider in this paper continuous multi-state Markov models, as shown in Figure 1, with three states: autonomy (A), LTC/Disability (D) and death. Such models are often used by insurers in practice. As no return to a better state of health is envisaged in this paper, e.g., in Nuttall et al. (1994) and Alegre et al. (2003), only three laws are needed in this model:

- i_x is the incidence intensity at age x,
- λ_x^A is the mortality intensity in state A at age x, and
- λ_x^D is the mortality intensity in state D at age x.

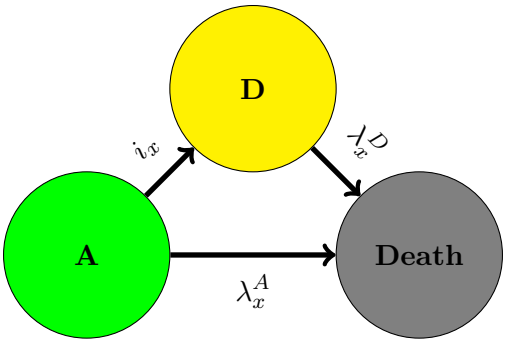


Figure 1: Modelling of an LTC product

Let X_x be the current state of an individual at age $x \in \mathbb{R}^+$. The transition intensities are defined as

$$i_x = \lim_{h \to 0^+} \frac{\mathbb{P}(X_{x+h} = D | X_x = A)}{h} ,$$

$$\lambda_x^A = \lim_{h \to 0^+} \frac{\mathbb{P}(X_{x+h} = \text{Death} | X_x = A)}{h} ,$$

$$\lambda_x^D = \lim_{h \to 0^+} \frac{\mathbb{P}(X_{x+h} = \text{Death} | X_x = D)}{h} .$$

Let x be an integer. The notation is as follows. Consider independent individuals $j = 1, \ldots, n$ observed at least one day in the autonomous state A between x and $x+1$. For each individual j, let $x + {}^j a$ and $x + {}^j b$ be the age at the beginning and end of observation in state A, respectively, for the age band $[x; x+1]$. Let $x + {}^j c$ be the age at the end of observation in the portfolio between x and $x+1$. Then, $0 \leqslant {}^j a \leqslant {}^j b \leqslant {}^j c \leqslant 1$. If j does not enter state D between x and $x+1$, then ${}^j b = {}^j c$. The end of observation in state A of an individual between x and $x+1$ can be due to three reasons:

1. right censoring,
2. loss of autonomy, or
3. death.

Let ${}^j d_x$ and ${}^j LTC_x$ indicate the cause of the end of observation in state A of j such that:

- ${}^j d_x = 1$ if the cause is death and 0 otherwise, and
- ${}^j LTC_x = 1$ if the cause is the loss of autonomy and 0 otherwise.

Let ${}^j d_x^{LTC} = 1$ if individual j dies in state D between x and $x+1$, 0 otherwise.

The main assumptions that we need in this paper are as follows:

1. the mortality rates remain constant throughout the age interval from x to $x+1$ (where x is an integer),
2. the logarithm of mortality rates may be decomposed on a P-Splines basis (see Section 2.3),

3. incidence and general mortality laws are assumed to be known and are not estimated.

Then, using classic tools of survival analysis and methods for modelling competing risks, as explained in Section 2.3 in Porta et al. (2007), and from the definitions of the intensities, the likelihood associated with individual j between x and $x+1$ is given by

$$ {}^{j}L_x = \underbrace{\exp\left(-\int_{{}^{j}a}^{{}^{j}b}(\lambda^A_{x+u}+i_{x+u})\,\mathrm{d}u\right)(i_{x+{}^{j}b})^{{}^{j}LTC_x}(\lambda^A_{x+{}^{j}b})^{{}^{j}d_x}}_{{}^{j}L^A_x}\underbrace{\exp\left(-\int_{{}^{j}b}^{{}^{j}c}(\lambda^D_{x+u})\,\mathrm{d}u\right)(\lambda^D_{x+{}^{j}c})^{{}^{j}d^{LTC}_x}}_{{}^{j}L^D_x}. \tag{2} $$

The likelihood of individual j in Equation (2) can be separated into 2 distinct partial likelihoods. ${}^{j}L^A_x$ corresponds to the experience of j in state A, whereas ${}^{j}L^D_x$ corresponds to its experience in state D. If j is not observed in state D between x and $x+1$, then ${}^{j}L^D_x = 1$. We can then study the experience in states A and D separately.

Under Assumption 1, ${}^{j}L^A_x$ becomes

$$ {}^{j}L^A_x = \exp\left(-(\lambda^A_x+i_x)\,{}^{j}e^A_x\right)(i_x)^{{}^{j}LTC_x}(\lambda^A_x)^{{}^{j}d_x}, \tag{3} $$

where ${}^{j}e^A_x = {}^{j}b - {}^{j}a$ is the time of exposure to the risk in the autonomous state of individual j between age x and $x+1$.

Therefore, the likelihood for the age band $[x;x+1]$ for the overall population of individuals $j = 1,\ldots,n$, being the product of all individuals likelihood is equal to

$$ L^A_x = \exp\left(-(\lambda^A_x+i_x)e^A_x\right)(i_x)^{LTC_x}(\lambda^A_x)^{d_x}, \tag{4} $$

where e^A_x, called the central exposure to risk is the sum of the time exposed to the risk in autonomy by all individuals in the age band, and LTC_x and d_x are the total number of observed losses of autonomy and deaths, respectively.

Maximizing the likelihood L^A_x with respect to λ^A_x and i_x is equivalent to maximizing separately

- $L^{A\to D}_x(i_x) = \exp\left(-i_x e^A_x\right)(i_x)^{LTC_x}$, and
- $L^{A\to\text{Death}}_x(\lambda^A_x) = \exp\left(-\lambda^A_x e^A_x\right)(\lambda^A_x)^{d_x}$.

We note that $L^{A\to D}_x(i_x)$ and $L^{A\to\text{Death}}_x(\lambda^A_x)$ are proportional to the likelihood of Poisson variables with expectancies equal to $i^A_x e^A_x$ and $\lambda^A_x e^A_x$, respectively. Therefore, the maximum likelihood estimators of i_x and λ^A_x obtained by maximizing Equation (4) are equal to the maximum likelihood estimators of the Poisson distributions. The incidence intensity i_x is considered as known in this paper. The likelihood has to be maximized with respect to λ^A_x only. One can therefore assume that

the number of observed deaths in autonomy (state A) at age x follows a Poisson distribution with parameter $\lambda_x^A e_x^A$. The same reasoning can be applied to deaths in the LTC state (state D). This result is interesting, allowing us to use the theory of Poisson-GLM to estimate λ_x^A. In the rest of the paper, we assume that the number of deaths at age x in both autonomous and LTC states follows a Poisson distribution with parameter $\lambda_x^G e_x^G$ where $G \in \{A, D, gen\}$.

Let x_{min} and x_{max} be integers corresponding to the minimum and maximum observed ages. Let d_x^A and d_x^D be the observed deaths at age x in states A and D, respectively. Since we may not have enough observations at certain ages, we introduce an indicator function w_x^G for each group indicating if the observations in the associated group at age x are reliable and can be included in the log-likelihood.

The total likelihood for all the observations from x_{min} to x_{max} in group G is given by

$$L^G\left(\lambda_{x_{min}}^G, \ldots, \lambda_{x_{max}}^G\right) = \prod_{x=x_{min}}^{x_{max}} \left[\frac{\left(\lambda_x^G e_x^G\right)^{d_x^G}}{d_x^G!} \exp\left(-\lambda_x^G e_x^G\right) \right]^{w_x^G}, \tag{5}$$

where $w_x^G = 1$ if the observations in the associated group at age x are reliable and 0 otherwise.

Equation (5) is the product of likelihoods of Poisson distributions for each age x between x_{min} and x_{max}.

The maximum likelihood estimators of the intensities $\hat{\lambda_x^G}$ for $x \in \{x_{min}; x_{min}+1; ...; x_{max}\}$ are given by the ratio d_x^G/e_x^G. Each intensity is fully determined by the deaths and exposure at this age, regardless of the observations at neighbouring ages. This can imply a very irregular curve of the mortality law that can be explained by the variance of the estimator, equal to d_x/e_x^2. Smoothing methods can be used to obtain a smoother mortality law, reducing volatility in the results. In this paper, we use the P-Splines smoothing method, which is widely used in the literature to smooth mortality intensities.

2.3 A model based on P-Splines

The method of P-Splines is a method of smoothing embedded in the GLM framework described in Eilers and Marx (1996), Marx and Eilers (1998), Eilers and Marx (2002), or Currie and Durban (2002).

Let $n = x_{max} - x_{min}$ be an integer. Let J be an integer representing the number of splines. In this method:

1. Let B be a basis matrix of cubic splines such that $B_{i,j}$ is the value of the cubic spline j at the i^{th} age. For a given group (A or D), the curve of the mortality intensities is considered a linear combination of J cubic splines. Let $\boldsymbol{\theta^G} = \{\theta_1^G, \ldots, \theta_J^G\}$ be the vector

of coefficients such that $\log(\lambda^G_{\boldsymbol{\theta^G},x_i}) = \sum_{j=1}^{J} B^G_{i,j}\theta^G_j$ for each $i \in \{1,\dots,n+1\}$. Let $\boldsymbol{\Lambda}^G_{\boldsymbol{\theta^G}} = \left(\lambda^G_{\boldsymbol{\theta^G},x_{min}}, \lambda^G_{\boldsymbol{\theta^G},x_{min}+1}, \dots, \lambda^G_{\boldsymbol{\theta^G},x_{max}}\right)^T$. We can write this linear combination in matrix form as

$$\log(\boldsymbol{\Lambda}^G_{\boldsymbol{\theta^G}}) = B^G\boldsymbol{\theta^G},$$

where $B^G \in M_{n+1,J}(\mathbb{R}^+)$ represents the matrix of the J splines at each point $\{x_{min},\dots,x_{max}\}$. B^G_{ij} is the value of the j^{th} spline at the i^{th} age of group G. This matrix is called the B-spline basis.

2. Let d be an integer. A penalty term $\rho(D^G_d\boldsymbol{\theta^G})^T(D^G_d\boldsymbol{\theta^G})$ depending on the penalty order d is added to the log-likelihood to avoid complex models with excessively large variability between coefficients of adjacent splines. $\rho \in \mathbb{R}$ is a smoothing parameter giving a weight to the penalty. Let $\Delta\theta^G_j = \theta^G_j - \theta^G_{j-1}$, $\Delta^2\theta^G_j = \Delta(\Delta\theta^G_j) = \theta^G_j - 2\theta^G_{j-1} + \theta^G_{j-2}$, …, $\Delta^d\theta^G_j = \Delta(\Delta^{d-1}\theta^G_j)$.

   D^G_d is defined as the matrix satisfying $D^G_d\boldsymbol{\theta^G} = \Delta^d\boldsymbol{\theta^G}$.

   Let $P^G_d = \rho(D^G_d)^T D^G_d$; a simpler way to write the penalty term that is used in the rest of the paper is $(\boldsymbol{\theta^G})^T P^G_d\boldsymbol{\theta^G}$.

Therefore, J coefficients θ^G_j must be estimated for each group G instead of one by age. In addition to having a smoother curve, the number of coefficients to estimate is then lower than if no smoothing method were used. The extrapolation mostly depends on the order of the penalty. The smoothing parameter ρ can be chosen to minimise the BIC as recommended in Currie and Durban (2002). The choice of other parameters, such as the number of nodes or the degree of the splines, may be less critical, as different choices often lead to similar smoothings. Ruppert (2000) and Eilers and Marx (1996) study the choice of the P-Splines parameters. The following rule of thumb is often sufficient:

- In the case of equidistant data, fix 1 node every 4 or 5 observations,
- Use cubic splines (order 3).

Let $(B^G\boldsymbol{\theta^G})_k$ be the k^{th} coefficient of the vector $\log(\boldsymbol{\Lambda}^G_{\boldsymbol{\theta^G}}) = B^G\boldsymbol{\theta^G}$. The penalized log-likelihood for group G is given by

$$
\begin{aligned}
l^G_{pen}(\boldsymbol{\theta^G}) &= \log(L^G(\boldsymbol{\theta^G})) - \frac{1}{2}(\boldsymbol{\theta^G})^T P^G_d \boldsymbol{\theta^G} \\
&= \sum_{x=x_{min}}^{x_{max}} w^G_x \left[d^G_x \log(\lambda^G_{\boldsymbol{\theta^G},x}) - \lambda^G_{\boldsymbol{\theta^G},x} e^G_x \right] - \frac{1}{2}(\boldsymbol{\theta^G})^T P^G_d \boldsymbol{\theta^G} \\
&= \underbrace{\sum_{k=0}^{x_{max}-x_{min}} w^G_{x_{min}+k} \left[d^G_{x_{min}+k} (B^G \boldsymbol{\theta^G})_{k+1} - \exp((B^G \boldsymbol{\theta^G})_{k+1}) e^G_{x_{min}+k} \right]}_{l^G(\boldsymbol{\theta^G})} - \frac{1}{2}(\boldsymbol{\theta^G})^T P^G_d \boldsymbol{\theta^G}.
\end{aligned}
\tag{6}
$$

In the following, the log-likelihood for autonomous and LTC groups is denoted by $l^A(\boldsymbol{\theta^A})$ and $l^D(\boldsymbol{\theta^D})$, respectively.

It is possible to smooth the intensities of the 2 groups simultaneously but independently (i.e. the observations of one group have no influence on the estimate of the mortality of the other). As the respective penalized log-likelihoods are independent, maximizing the sum of these log-likelihoods is equivalent to maximizing both of them. This is then equivalent to maximizing $l^{A/D}_{pen}\left(\boldsymbol{\theta^A}, \boldsymbol{\theta^D}\right)$ given by the following equation

$$
\begin{aligned}
l^{A/D}_{pen}\left(\boldsymbol{\theta^A}, \boldsymbol{\theta^D}\right) = &\sum_{k=0}^{x_{max}-x_{min}} w^A_{x_{min}+k} \left[d^A_{x_{min}+k} (B^A \boldsymbol{\theta^A})_{k+1} - \exp((B^A \boldsymbol{\theta^A})_{k+1}) e^A_{x_{min}+k} \right] - \frac{1}{2}(\boldsymbol{\theta^A})^T P^A_d \boldsymbol{\theta^A} \\
+ &\sum_{k=0}^{x_{max}-x_{min}} w^D_{x_{min}+k} \left[d^D_{x_{min}+k} (B^D \boldsymbol{\theta^D})_{k+1} - \exp((B^D \boldsymbol{\theta^D})_{k+1}) e^D_{x_{min}+k} \right] - \frac{1}{2}(\boldsymbol{\theta^D})^T P^D_d \boldsymbol{\theta^D}.
\end{aligned}
\tag{7}
$$

To this aim, we introduce the basis spline matrix $B = \begin{bmatrix} B^A & 0 \\ 0 & B^D \end{bmatrix} \in M_{2(n+1),2J}(\mathbb{R}^+)$ and the penalty matrix $P = \begin{bmatrix} P^A_d & 0 \\ 0 & P^D_d \end{bmatrix} \in M_{2J,2J}(\mathbb{R})$.

Let $\boldsymbol{\theta}$ be the vector of all the smoothing coefficients, i.e., $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta^A} \\ \boldsymbol{\theta^D} \end{pmatrix} \in \mathbb{R}^{2J}$, $\log(\boldsymbol{\Lambda_\theta}) = \begin{pmatrix} \log(\boldsymbol{\Lambda}^A_{\boldsymbol{\theta}^A}) \\ \log(\boldsymbol{\Lambda}^D_{\boldsymbol{\theta}^D}) \end{pmatrix} \in \mathbb{R}^{2(n+1)}$; then:

1. $\log(\boldsymbol{\Lambda_\theta}) = B\boldsymbol{\theta}$,
2. The sum of the penalties of the 2 groups can be written as $\frac{1}{2}\boldsymbol{\theta}^T P \boldsymbol{\theta}$.

We introduce the following vectors:

- $\mathbf{d} = \left(d^A_{x_{min}}, \ldots, d^A_{x_{min}+n}, d^D_{x_{min}}, \ldots, d^D_{x_{min}+n}\right)^T \in \mathbb{N}^{2(n+1)}$,
- $\mathbf{e} = \left(e^A_{x_{min}}, \ldots, e^A_{x_{min}+n}, e^D_{x_{min}}, \ldots, e^D_{x_{min}+n}\right)^T \in \mathbb{R}^{2(n+1)}$, and
- $\mathbf{w} = \left(w^A_{x_{min}}, \ldots, w^A_{x_{min}+n}, w^D_{x_{min}}, \ldots, w^D_{x_{min}+n}\right)^T \in \{0,1\}^{2(n+1)}$,

of length $2(n+1)$. Then, the penalized log-likelihood on all the observations is

$$\begin{aligned} l^{A/D}_{pen}(\boldsymbol{\theta}) &= l^A(\boldsymbol{\theta}) + l^D(\boldsymbol{\theta}) - \underbrace{\frac{1}{2}\boldsymbol{\theta}^T P \boldsymbol{\theta}}_{\text{P-Splines smoothing penalty } Pen^{smoothing}} \\ &= \sum_{k=0}^{n} \mathbf{w}_{k+1}\left[\mathbf{d}_{k+1}(B\boldsymbol{\theta})_{k+1} - \exp((B\boldsymbol{\theta})_{k+1})\mathbf{e}_{k+1}\right] \\ &+ \sum_{k=n+1}^{2n+1} \mathbf{w}_{k+1}\left[\mathbf{d}_{k+1}(B\boldsymbol{\theta})_{k+1} - \exp((B\boldsymbol{\theta})_{k+1})\mathbf{e}_{k+1}\right] - \frac{1}{2}\boldsymbol{\theta}^T P \boldsymbol{\theta}. \end{aligned}$$

2.4 Introduction of a second penalty on the log-likelihood

In this paper, the mortality intensities λ^{gen}_x of the general population are assumed to be known and to be piecewise constant as the mortality laws of groups A and D. The idea of our approach is to minimize the gap between the predicted number of deaths in the general population and the sum of the deaths in autonomous and LTC states (cf. Equation (1)). As the portfolio is composed of autonomous and dependent individuals, the total number of deaths at age x in the portfolio is equal to the sum of deaths in autonomy (state A) and LTC state (state D), as shown on Figure 2. This figure shows the possible transitions in an LTC product. One observed death is either a death in autonomy or in the LTC state. The exposures in these states are respectively denoted by e^A and e^D. Therefore, the exposure of the portfolio is $e^A + e^D$. In the context of a Poisson distribution for the number of observed deaths, Equation (1) can be written as follows

$$\lambda^{gen}_x(e^A_x + e^D_x) = \lambda^A_x e^A_x + \lambda^D_x e^D_x. \tag{8}$$

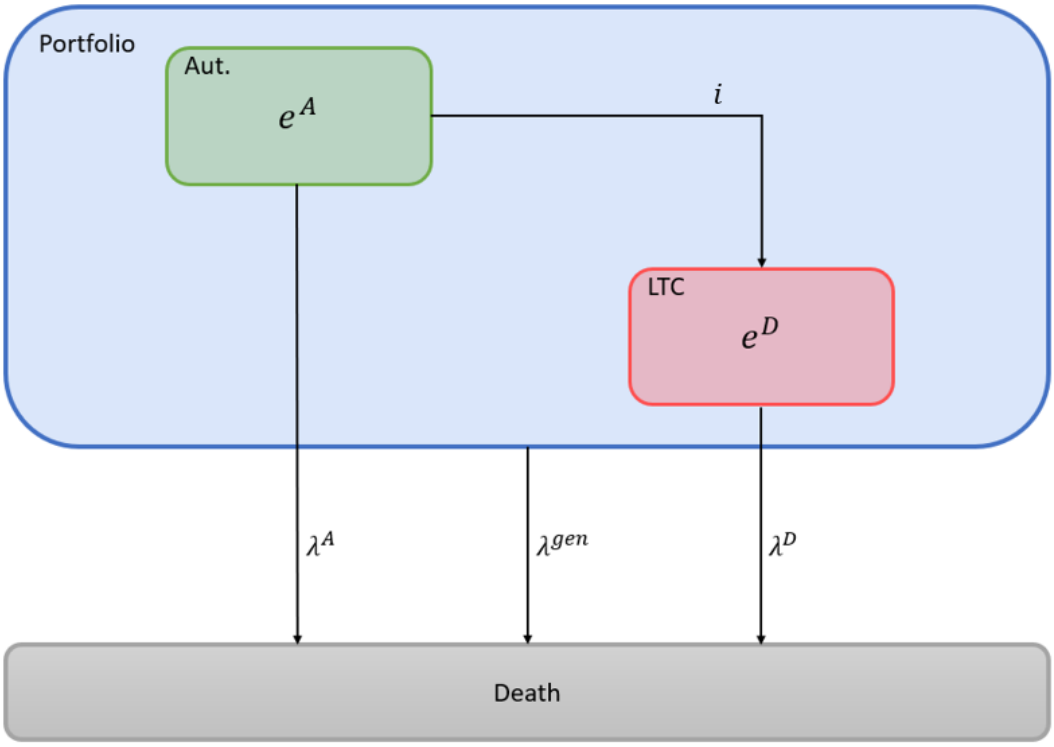


Figure 2: Modelling of an LTC product keeping consistency between mortality laws

To respect the coherence criteria between the 3 mortality laws (autonomous, LTC, and general), a second penalty term given by the following equation

$$
\begin{aligned}
Pen^{loopback}\left(\boldsymbol{\theta}\right) &= \frac{1}{2}K\sum_{x=x_{min}}^{x_{max}}\left(\frac{\lambda_x^{gen}\left(e_x^A+e_x^D\right)-\lambda_{\boldsymbol{\theta},x}^A e_x^A-\lambda_{\boldsymbol{\theta},x}^D e_x^D}{e_x^A+e_x^D}\right)^2 \\
&= \frac{1}{2}K\sum_{k=0}^{n}\left(\frac{\lambda_{x_{min}+k}^{gen}\left(\mathbf{e}_{k+1}+\mathbf{e}_{(n+1)+k+1}\right)-\Lambda_{\boldsymbol{\theta},k+1}\mathbf{e}_{k+1}-\Lambda_{\boldsymbol{\theta},n+1+k}\mathbf{e}_{(n+1)+k+1}}{\mathbf{e}_{k+1}+\mathbf{e}_{(n+1)+k+1}}\right)^2,
\end{aligned}
\tag{9}
$$

is added to the log-likelihood, where K is a parameter corresponding to the weight given to the consistency criteria in the estimation of the mortality laws.

The new penalized log-likelihood now becomes

$$
l_{pen}^{loopback}\left(\boldsymbol{\theta}\right) = l_{pen}^{A/D}\left(\boldsymbol{\theta}\right) - Pen^{loopback}\left(\boldsymbol{\theta}\right). \tag{10}
$$

This penalized log-likelihood is maximized by the Newton-Raphson algorithm. The first and second

derivatives with respect to the coefficients θ_i are needed. The matrix forms of the gradient and the Hessian are given by Equations (11) and (12), respectively.

$$\nabla_{\boldsymbol{\theta}}(l_{pen}) = B^T W(\mathbf{d} - \check{\mathbf{d}}_{\boldsymbol{\theta}}) - P\boldsymbol{\theta} - KB^T \left((\tilde{W}_3^{-1})^2 W_{\boldsymbol{\theta}}^Q \otimes I_2\right) \check{\mathbf{d}}_{\boldsymbol{\theta}}, \tag{11}$$

$$\begin{aligned} H_{\boldsymbol{\theta}}(l_{pen}) = -B^T W W_{\boldsymbol{\theta}} B - P - KB^T \left[W_{\boldsymbol{\theta}} \left([(\tilde{W}_3^{-1})^2 W_{\boldsymbol{\theta}}^Q] \otimes I_2\right)\right] B - \\ K \begin{bmatrix} \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A & \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \end{bmatrix}^T \begin{bmatrix} \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A & \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \end{bmatrix}, \end{aligned} \tag{12}$$

where:

- the basis matrices $B \in M_{2(n+1),2J}(\mathbb{R}^+)$, $B_A \in M_{(n+1),J}(\mathbb{R}^+)$, $B_D \in M_{(n+1),J}(\mathbb{R}^+)$, the penalty matrix $P \in M_{2J,2J}(\mathbb{R})$ and the vector of deaths $\mathbf{d} \in \mathbb{N}^{2(n+1)}$ are introduced in Section 2.3,
- $\check{\mathbf{d}}_{\boldsymbol{\theta}} = (\Lambda_{\boldsymbol{\theta},i}\mathbf{e}_i)_{i\in\{1,\dots,2(n+1)\}}$ is the expected number of observed deaths with the assumption of a Poisson distribution,
- $W = \text{diag}(\mathbf{w})$,
- I_2 is the 2×2 identity matrix
- $W_{\boldsymbol{\theta}} = \text{diag}(\check{\mathbf{d}}_{\boldsymbol{\theta}}) \in M_{2(n+1),2(n+1)}(\mathbb{R})$,
- $W_{\boldsymbol{\theta}}^G = \text{diag}(\check{\mathbf{d}}_{\boldsymbol{\theta}}^C) \in M_{(n+1),(n+1)}(\mathbb{R})$ where $G \in \{A, D\}$,
- $\tilde{W}_3 = \text{diag}\left((\mathbf{e}_{k+1} + \mathbf{e}_{(n+1)+k+1})_{k\in\{0,\dots,n\}}\right) \in M_{(n+1),(n+1)}(\mathbb{R})$ is the diagonal matrix of the total exposure at each age, and
- $W_{\boldsymbol{\theta}}^Q = \text{diag}\left((\Lambda_{\boldsymbol{\theta},k+1}\mathbf{e}_{k+1} + \Lambda_{\boldsymbol{\theta},(n+1)+k+1}\mathbf{e}_{(n+1)+k+1} - \lambda^{gen}_{x_{min}+k}[\mathbf{e}_{k+1} + \mathbf{e}_{(n+1)+k+1}])_{k\in\{0,\dots,n\}}\right)$ $\in M_{(n+1),(n+1)}(\mathbb{R})$.

The Newton-Raphson algorithm is used to find the optimal coefficients $\boldsymbol{\theta}$. The estimator at step $k+1$ denoted $\hat{\boldsymbol{\theta}}^{(k+1)}$ is given by

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} - H_{\widehat{\boldsymbol{\theta}}^{(k)}}(l_{pen})^{-1} \nabla_{\widehat{\boldsymbol{\theta}}^{(k)}}(l_{pen}). \tag{13}$$

The algorithm stops when the maximum relative difference between two coefficients from successive iterations is lower than a previously fixed tolerance ε (i.e $\max\limits_{i\in\{1,\dots,2M\}} \left|\dfrac{\widehat{\boldsymbol{\theta}}_i^{(k)} - \widehat{\boldsymbol{\theta}}_i^{(k-1)}}{\widehat{\boldsymbol{\theta}}_i^{(k)}}\right| < \varepsilon$). The final estimator of $\boldsymbol{\theta}$ is denoted by $\widehat{\boldsymbol{\theta}}$.

The convergence of the algorithm is discussed in Appendix A.

3 Extrapolation of mortality laws: calibration of theoritical exposures at old ages

Extrapolating the mortality laws is necessary when not having enough observations at certain old ages. In this case, the lack of observations does not enable the inclusion of this information at old ages in the likelihood in Equation (10). The weights of these ages are then fixed to 0 ($w_x = 0$). The corresponding intensities are exclusively determined in such a way as to minimize both the P-Splines and consistency penalties ($Pen^{smoothing}$ and $Pen^{loopback}$).

3.1 Research problem

The previous methodology is adapted when having non-zero exposures up to the maximum age at which we wish to estimate mortality law. However, the data often contain very few or no observations at old ages. In fact, as age increases, exposures in the portfolio of the insurer tend to decrease because of deaths. As a consequence, the loopback penalty, based on exposures, is of undetermined form. The extrapolation is therefore done entirely with the P-Splines penalty. However, the objective and the interest of the loopback is to extrapolate the mortality laws in a coherent way to better estimate the mortality at old ages, using the information on the general mortality at these ages.

We overcome this point by computing theoretical exposures at these old ages.

3.2 Estimation of theoritical exposures and extrapolation

To overcome this issue, a maximum age at which we have enough observations is fixed. All exposures above this age x_M are estimated to compute the distribution of autonomous and LTC people at each age in the general population. The theoretical and estimated exposures are only used in the loopback penalty term $Pen^{loopback}$ of Equation (10).

Suppose that the general mortality intensities of the overall portfolio and the incidence are known. Theoretical exposures at each age above x_M and in each group can be obtained by projecting the population observed at the maximum age previously fixed. To this aim, we start the projection with the exposures computed at age x_M. The mortality laws and the incidence intensities are then used to estimate the exposures at older ages.

Let $\mathbf{e}$ and $\mathbf{d}$ be the vectors of observed exposures and deaths in the portfolio, respectively, as defined in Section 2.3. Let $\mathbf{e^{est}}$ be the estimated vector of exposures. The k^{th} terms of $\mathbf{e}$ and $\mathbf{e^{est}}$ are equal ($\mathbf{e^{est}}_k = \mathbf{e}_k$) for each k corresponding to an age below x_M. Here, $\mathbf{e^{est}}$ is estimated by projection using mortality laws and incidence intensities, depending on the vector of coefficients of the splines $\boldsymbol{\theta}$. Therefore, $\mathbf{e^{est}}$ is denoted by $\mathbf{e^{est}}(\boldsymbol{\theta})$.

Let $l_{pen}^{loopback}\left(\boldsymbol{\theta}|\mathbf{e^{est}}, \mathbf{d}\right)$ be the penalized log-likelihood from Equation (10) given the estimated exposure vector $\mathbf{e^{est}}$ and the vector of death counts $\mathbf{d}$. We want to compute the mortality laws

by maximizing this penalized log-likelihood with respect to $\boldsymbol{\theta}$. However, the exposures are needed to compute the mortality laws, and the mortality laws are needed to estimate the exposures by projection.

The problem that we want to solve in this section is then:

$$\max_{\boldsymbol{\theta}} l_{pen}^{loopback}\left(\boldsymbol{\theta}|\mathbf{e}^{\mathbf{est}}(\boldsymbol{\theta}), \mathbf{d}\right).$$

The exposures at old ages are then estimated iteratively with Algorithm 1 by updating the mortality laws and exposures simultaneously. The Newton-Raphson algorithm in Section 2.4 is first used with $K = 0$ to compute mortality laws without the consistency constraint. From the resulting mortality laws and the known incidence intensities, exposures are first estimated by projection of the portfolio at age x_M. The following two steps are then repeated several times. First, mortality laws are computed using the Newton-Raphson algorithm from Section 2.4 with the estimated exposures and the chosen parameter K to link the estimation of the two mortality laws. The second stage consists of the re-computation of the exposures using the mortality laws from the previous step.

Let:

- $loopback(data, K, expo)$ be the loopback function applied to $data$ with the Newton-Raphson algorithm in Section 2.4, with $expo$ the exposures in both autonomous and LTC group (A and D respectively) at each age and K the loopback penalty chosen for the calibration of the model,
- $compute\_expo(expo_{x_M}, incidence, mortality_A, mortality_D)$ be the projection function that estimates the theoretical exposures given the incidence ($incidence$), the mortality laws ($mortality_A$ and $mortality_D$), and the exposures $expo_{x_M}$ at the age chosen for the projection x_M (the last age for which we consider the real exposures). This projection is made by considering the exposure in each group at age x_M as the number of insureds in each group. Under the assumption that all deaths and losses of autonomy occur at the end of the period (i.e., just before the birthday of the insured) and from the transition probabilities at each age, we are then able to estimate the number of insureds in states A and D at age $x \geqslant x_M$.
- $c(mortality_A, mortality_D)$ be the concatenation of the vectors of mortality intensities in states A and D,
- $expo_{data}$ be the real exposures observed at each age.

The algorithm stops when the maximum over all the ages between the exposures of two successive iterations is lower than a chosen tolerance ε_2. A maximum number of iterations is also fixed. Then, the algorithm returns the vector of the exposures in groups A and D at each age. The values are the real exposures for ages below x_M and theoretical exposures for ages above.

The mortality laws are then obtained by applying the Newton-Raphson algorithm in Section 2.4 by maximizing the penalized log-likelihood $l_{pen}^{loopback}$ (cf. Equation (10)). The weights w_x^G are fixed to 1

Algorithm 1 Exposures estimation algorithm

$K \leftarrow 0$
$expo \leftarrow expo_{data}$
$c(mortality_A, mortality_D) \leftarrow loopback(data, K, expo)$
$expo \leftarrow compute\_expo(expo_{x_M}, incidence, mortality_A, mortality_D)$
$K \leftarrow$ penalty parameter chosen for model calibration
for i=1,...,nb.iterations **do**
  $c(mortality_A, mortality_D) \leftarrow loopback(data, K, expo)$
  $expo \leftarrow compute\_expo(expo_{x_M}, incidence, mortality_A, mortality_D)$
end for
return $expo$

for the ages that are considered in the likelihood and 0 for the others.

The exposures used in the likelihood part of Equation (10) are the observed ones, even if $w_x^G = 1$ and the age is above x_M. The vector of theoretical exposure is only used in the loopback penalty.

We have introduced a hyper-parameter K on the penalized log-likelihood. Therefore, an important step for the user of this algorithm is to fix its value.

4 Choice of hyper-parameter K, an application on synthetic data

The choice of K is important for mortality law estimation. In fact, K can be considered as the weight given to the coherence criterion. The larger K is, the better the mortality laws estimated by the algorithm satisfy the coherence rule. Let us illustrate this aspect first on synthetic data.

4.1 Presentation of the synthetic data

Synthetic mortality laws have been constructed from age 50 to 120. Autonomous and LTC mortality laws have been independently estimated on a real French LTC portfolio covering severe LTC. The general mortality has then been constructed to satisfy the coherence criterion from Equation (8) in Section 2.4. The obtained laws are plotted from 50 to 120 years old in Figure 3. At age 50, the general mortality is equal to the autonomous mortality since we consider a population of 100% autonomous insured at age 50 to construct the general mortality law. As age increases, the proportion of disabled people in the general population changes, as does the general mortality law. We see that in this example of synthetic laws, the LTC mortality law converges to the general mortality since the population is composed of a majority of disabled individuals at old ages.

Let us assume then that the general mortality law is known but that no data are available above age 85 for both A and D groups. This means that exposures and number of deaths at these ages are null. The loopback algorithm should be able to find the mortality intensities for both autonomous and LTC groups that have been used to construct the general mortality law for ages above 85.

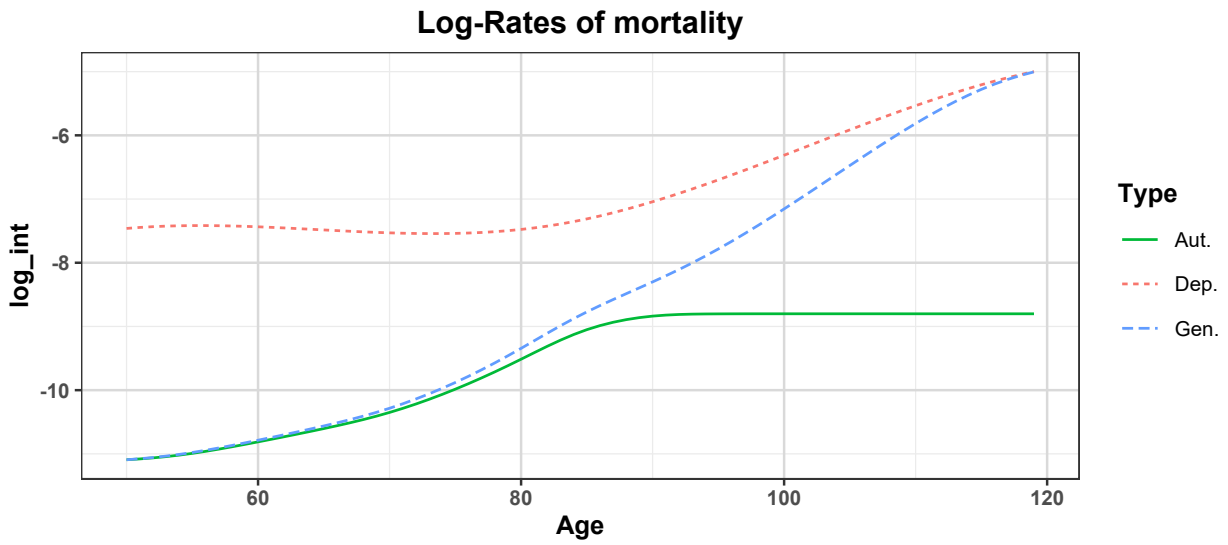


Figure 3: Synthetic mortality laws

4.2 Impact of the choice of K on the residual loopback error

The larger K is, the better the mortality laws estimated by the algorithm satisfy the coherence rule, and the lower the residual loopback error given by the formula

$$Error_{loopback} = \sum_{x=x_{min}}^{x_{max}} \left(\frac{\lambda_x^{gen} \left(e_x^A + e_x^D \right) - \lambda_x^A e_x^A - \lambda_x^D e_x^D}{e_x^A + e_x^D} \right)^2 . \tag{14}$$

The pattern of the residual loopback error as a function of K is illustrated in Figure 4.

4.3 Optimization of parameter K

A large value of K leads to a small value of the loopback error. Unfortunately, we cannot choose K as large as possible since it implies problems in the convergence of the algorithm. Indeed, the Hessian becomes non-invertible after a few iterations. We need to find a balance between minimizing the loopback error and having K small enough to converge the algorithm. Figure 4 shows that if we accept a residual error smaller than $2e-4$, we have to choose K such that the error is below the red line. This means here that we can choose all K larger than the one at the intersection between the red line and the error curve, which is approximatively equal to 2950.

The idea is to fix a tolerance criterion on the loopback error. We then choose the value of K that leads to a residual loopback error close to this tolerance. The smaller the tolerance is, the larger K. A function has been developed to optimize the choice of parameter K, leading to a tolerance close to the one previously fixed. In this example, we choose a tolerance named ε_3 in the R function equal to $2e-4$ (i.e., $\varepsilon_3 = 2e-4$).

The optimal value for K found by the algorithm is equal to 2741.65, and the residual loopback error

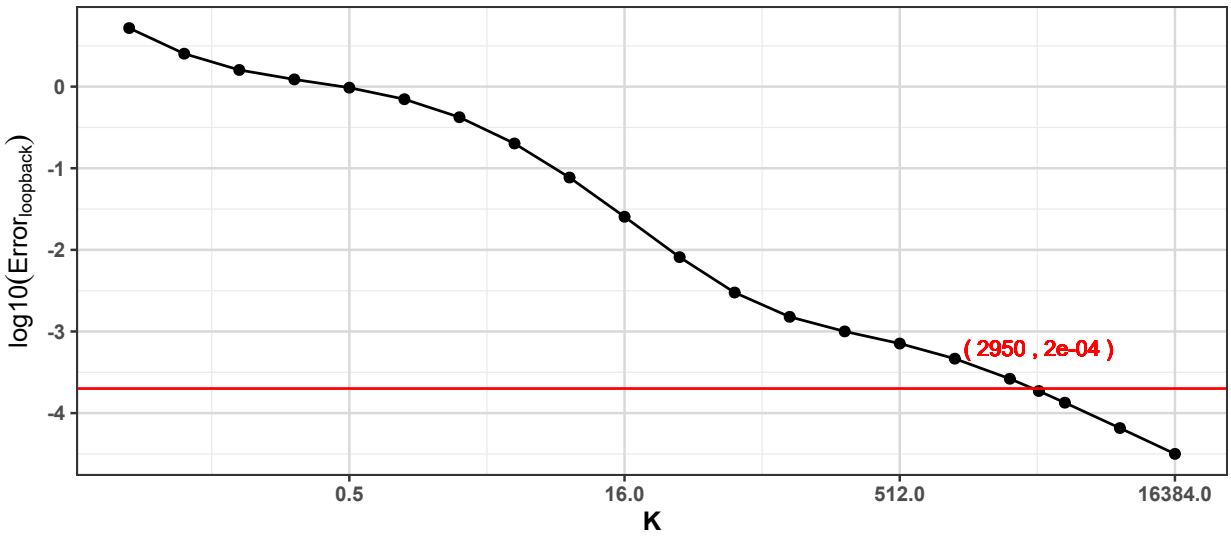


Figure 4: Choice of K given a tolerance on the residual loopback error

is equal to $2e-4$.

4.4 Application of the loopback with the optimal K

The loopback algorithm is then used with this optimized parameter K to estimate coherent autonomous and LTC mortality laws. Since no observations are available above 85, Algorithm 1 from Section 3.2 is used to estimate theoretical exposures at old ages appearing in the loopback penalty given by Formula 9. In this example, the maximum age x_M for which we used the real observed exposures is fixed to 80. All exposures used in the loopback penalty for ages above 80 are computed by projecting the population of age 80 using biometric laws. Proportions of autonomous and disabled individuals are sufficient to compute the penalty. Therefore, those values are computed from the theoretical exposures and plotted in Figure 5. Starting with almost only autonomous individuals (97.2%) at age 80, the proportion of disabled individuals increases and reaches 99.4% at age 119.

The resulting mortality laws, plotted in Figure 6, are very close to the laws we used to construct the general mortality law. In this figure, the triangle represents the observations used to fit the laws, and the dots represent the mortality intensities of the synthetic data that are not used in the loopback algorithm. The lines represent the estimated mortality laws with the loopback algorithm. Despite not using this information above age 85, the algorithm successfully manages to return mortality intensities close to those from the original synthetic data with the optimal K. Figure 6 shows the added value of the loopback algorithm. In fact, by not using any coherence penalty, the extrapolation of the mortality laws is driven only by the P-Splines order. Therefore, the extrapolation of the mortality in LTC is quadratic if the order is fixed to 3, as in our example. As a consequence, the mortality law in LTC (group D) diverges from the general mortality of the portfolio, whereas the

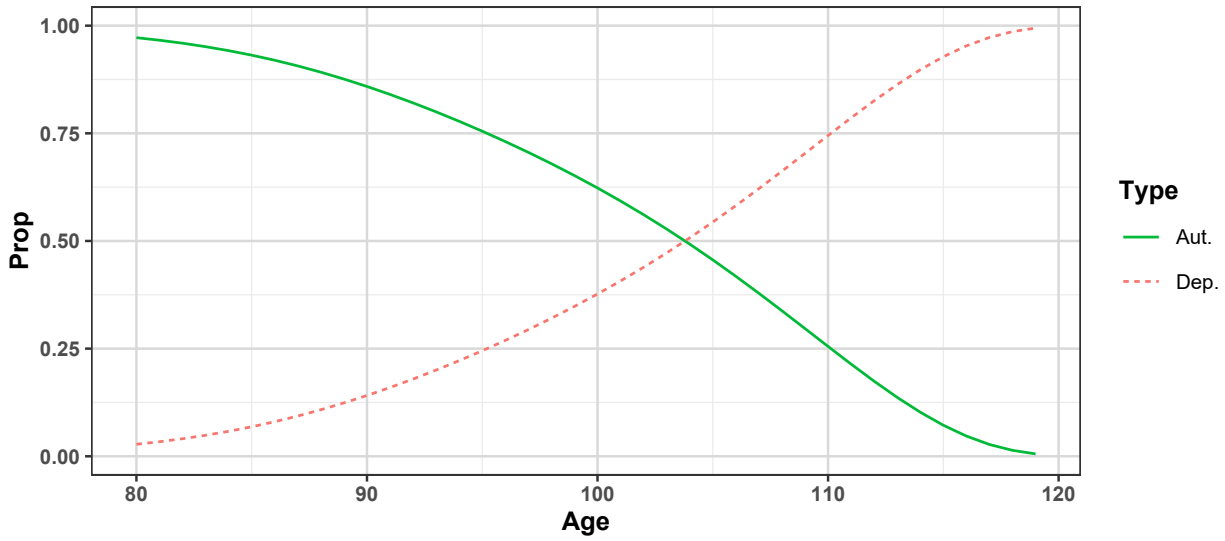


Figure 5: Proportions resulting from calibrated exposures with the optimal parameter K

population is mostly composed of disabled individuals at old ages.

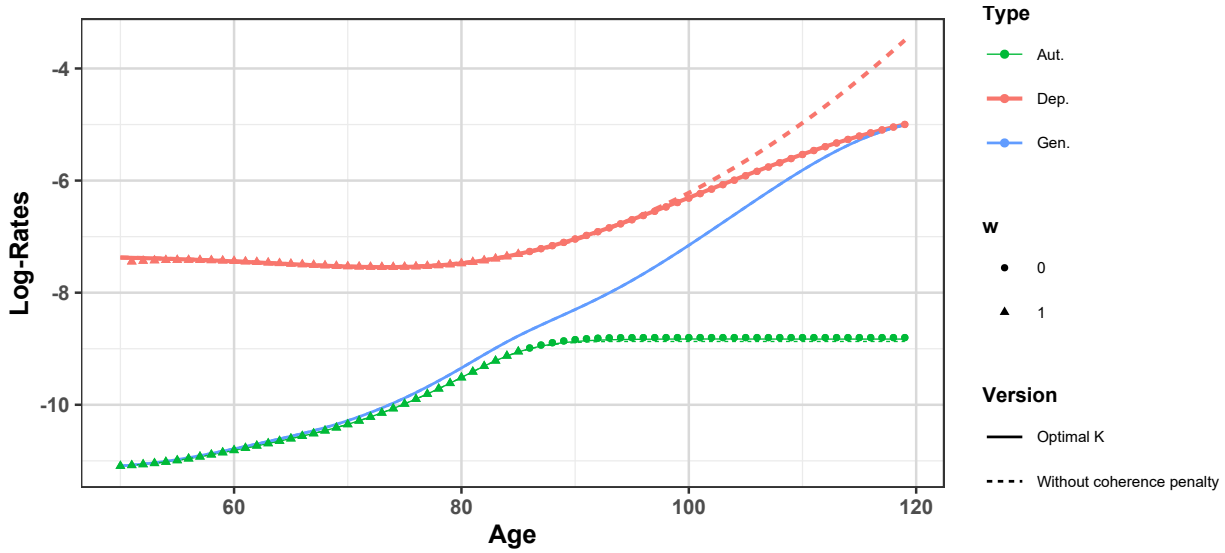


Figure 6: Estimated mortality laws with the optimal parameter K

5 A case study on real data

5.1 Data

We rely on data coming from 5 medium-to-large French LTC portfolios. The application focuses only on females. The level of the loss of autonomy varies from mild to severe. In this application, we consider only severe LTC, with the GIR12 definition from the AGGIR grid described in Dupourqué (2012), which is used by the French government for the attribution of public aid. From these

portfolios, 11 130 deaths are observed in the autonomous state (A), versus 3 681 in LTC (D). To calibrate the mortality laws, two datasets are constructed from the portfolios. The first one, called DB_A represents the dataset of the active contributors, and the second one, called DB_D, represents the dataset of the annuitants who are disabled. The first one is used to calibrate the autonomous mortality and incidence, while the second one is used to calibrate the mortality in LTC.

From these databases, only the observations between age 50 and 91 are used in the likelihood, as we decided to consider only ages with at least 10 observed deaths in our database. Observations at ages with fewer than 10 observed exits are too volatile. In Figures 8 and 10, representing estimated mortality laws, crude rates are represented as triangles or circles. Triangles represent the data points used in the likelihood, unlike the circles.

The general mortality law used in this section is calibrated on the same portfolios by aggregating the databases DB_A and DB_D and smoothing the crude rates by using the P-Splines smoothing methods. To extrapolate the general mortality law, we assume that the mortality law of the portfolio at old ages is close to the French mortality law, which is well known. The French mortality law used here comes from the « Human Mortality Database (HMD) » with an observation period from 2016 to 2018, available at www.mortality.org (data downloaded in May 2020) thanks to Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France) (2020). The BRASS model, explained in Brass (1971), is used to force the mortality law to converge to the HMD mortality.

The incidence law i used in this section is estimated on the same 5 French LTC portfolios with the P-Splines smoothing method with order $d = 2$. As many LTC products in France exclude recovery, we do not observe any transition from state D to state A.

5.2 Application

We first begin with the extrapolation of the mortality laws, excluding any coherence criterion. This is equivalent to using the loopback algorithm with a penalty K equal to 0. We then study the impact of K on the residual loopback error and choose the optimal hyper-parameter K. The mortality laws are then estimated and extrapolated using the loopback with this optimal penalization parameter. Life expectancy at age 50, an aggregate measure of mortality above age 50, is then computed.

5.2.1 Extrapolation without loopback penalization

When the penalization parameter K is fixed to 0, then the likelihood is equal to the sum of the likelihood of two P-Splines smoothing, one for each group (A and D). Maximizing the sum of these two likelihoods is equivalent to maximizing both of them independently. In this example, the smoothing penalty order is fixed to 1 for state A and 2 for state D. Therefore, as shown on Figure 8 with the dotted lines, the mortality law converges to a horizontal line at old ages for the first group and converges to a linear extrapolation for the second group.

5.2.2 Selection of parameter K and extrapolation of mortality laws

As seen in Section 4, parameter K has a large influence on the residual loopback error, with a larger K leading to a smaller residual loopback error.

For a tolerance ε_3 fixed to $1e-2$, the optimal K defined in Section 4 and chosen by the algorithm is equal to 411.56.

The maximum age x_M for which we use the real observed exposures in the loopback penalty is fixed to 90. All exposures used in the loopback penalty for ages above x_M are estimated using Algorithm 1 from Section 3.2. As in Section 4.4, the estimated proportions of autonomous (group A) and dependent individuals (group D) in the projected population needed for the loopback penalty are plotted in Figure 7. At age x_M, 80.6% of exposures are exposures in autonomy. At 99, estimated exposures in autonomy and disability are almost equal. The proportion of dependent individuals converges to 100% as age increases. Therefore, the population at old ages is composed almost only of dependent individuals, and we expect the estimated mortality law in LTC (group D) to converge to the global mortality of the portfolio denoted Gen.

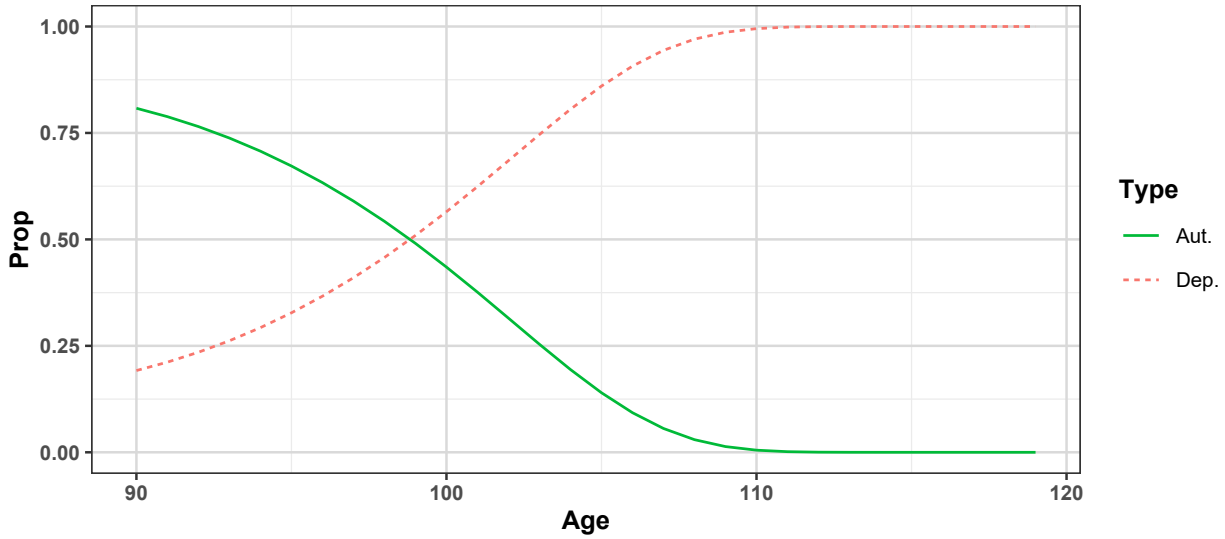


Figure 7: Proportions of autonomous and dependent individuals estimated with the optimal parameter K

As expected, the associated mortality laws obtained by the loopback algorithm, shown in solid lines in Figure 8, present convergence of LTC mortality to general mortality, while the log-intensity of autonomous mortality converges to a constant value. With the incidence law used in this example, the probability of remaining autonomous after 110 years is extremely low. Therefore, as shown in Figure 7, almost all the surviving insureds at 110 years are disabled, and the general mortality is equal to the mortality in LTC.

Figure 8 shows the impact of both smoothing and coherence penalties on the estimated mortality

laws. Mortality laws obtained by maximizing $l^G(\boldsymbol{\theta}^G)$ given in Equation (6) for each group G are represented in dotted lines. Without any smoothing penalty, the resulting mortality laws are very volatile and try to capture all the variance observed in the data. As shown in dashed lines, adding a smoothing penalty for each group reduces over-fitting and obtains better extrapolation in the sense that the mortality laws do not explode as age increases. Finally, solid lines represent the mortality laws obtained with optimal parameter K. Adding the coherence penalty allows consistency between the three mortality laws, having the mortality of group D (LTC) converging to the general mortality since the portfolio is composed almost only of dependent individuals at old ages.

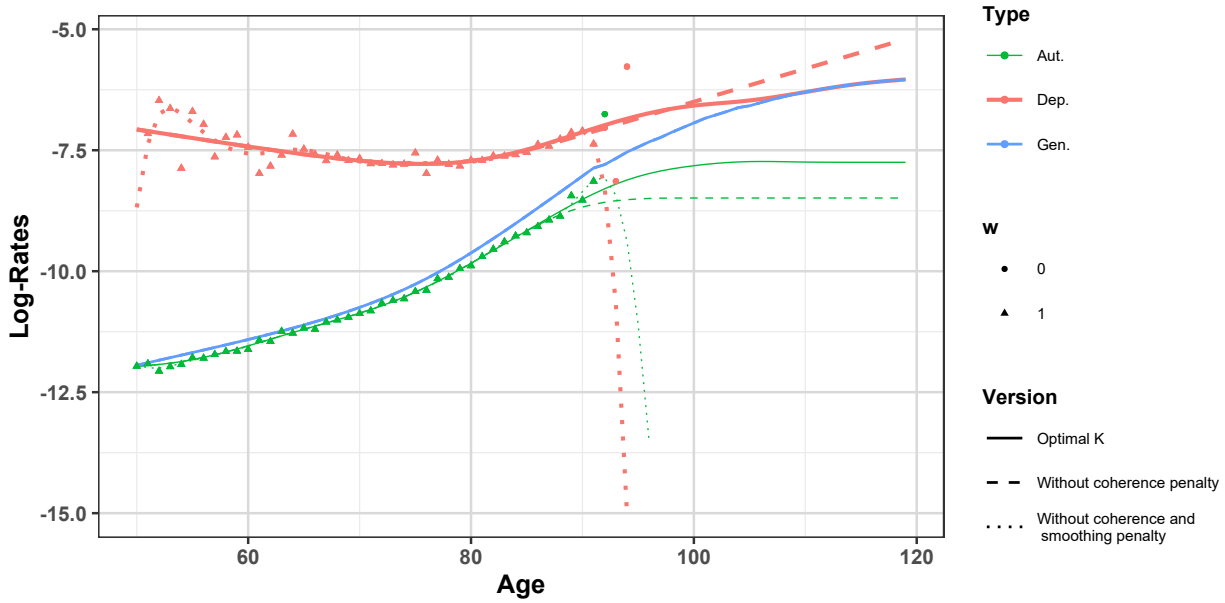


Figure 8: Comparison of the mortality laws for the GIR12 product obtained with the optimal K, without any coherence penalty, and without a smoothing penalty

Given the estimated proportions of autonomous and dependent individuals in Figure 7 and the associated calibrated mortality laws in groups A and D plotted in Figure 8, the implied mortality law of the porfolio is estimated with Equation 8. The proportions of individuals in states A and D are $\frac{e_x^A}{e_x^A + e_x^D}$ and $\frac{e_x^D}{e_x^A + e_x^D}$, respectively. Figure 9 shows how well the algorithm was able to replicate the general mortality law used as a reference. The two mortality laws are really close, except for ages below 70 where the implied mortality law is really close to the autonomous mortality law. This is explained by exposures in state D almost equal to zero at young ages. Therefore, the implied mortality law is almost equal to the autonomous mortality. Moreover, since exposures in state A are high at young ages, the weight of the likelihood of autonomous observations of Equation 10 is higher than the weight of the consistency penalty for these ages.

The confidence intervals of the two mortality laws are obtained with a simulation algorithm inspired by the bootstrap method. Using the fitted mortality laws and assuming that the number of deaths is Poisson distributed, new death counts per age and group (A and D) are simulated for each age

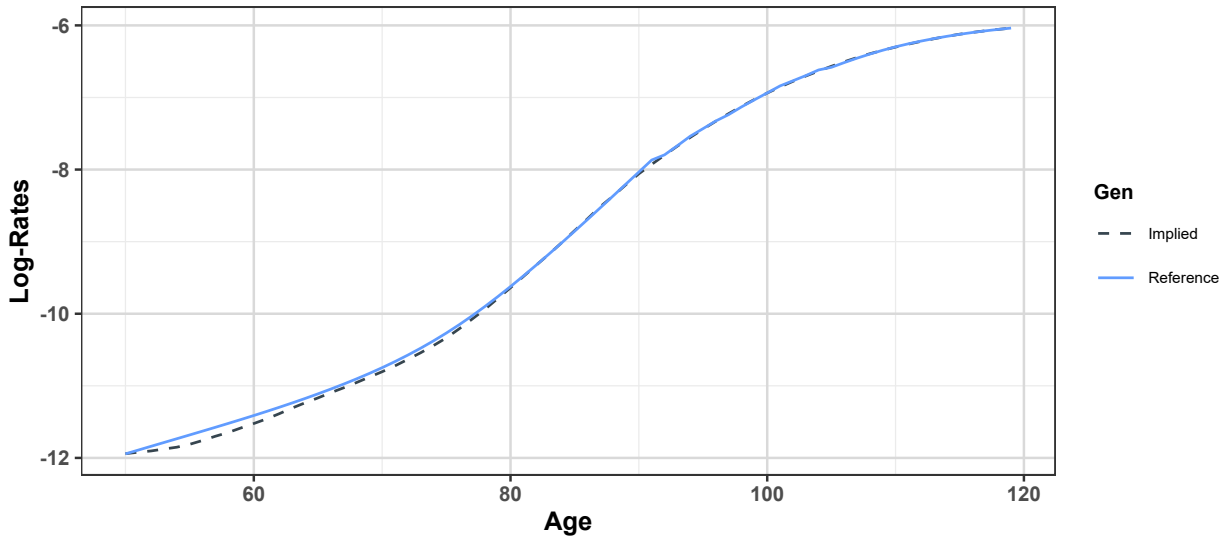


Figure 9: Comparison of the implied mortality of the portfolio to the general mortality law used as a reference

participating in the log-likelihood term of Equation (10). New mortality laws are then fitted using the loopback algorithm on these new simulated data. We must keep in mind that these confidence intervals are computed considering that the general mortality is known. This implies that the uncertainty on the general mortality is not taken into account when computing the confidence intervals of the mortality laws in autonomy and LTC.

The confidence intervals at 99% constructed with 800 simulations are shown in Figure 10.

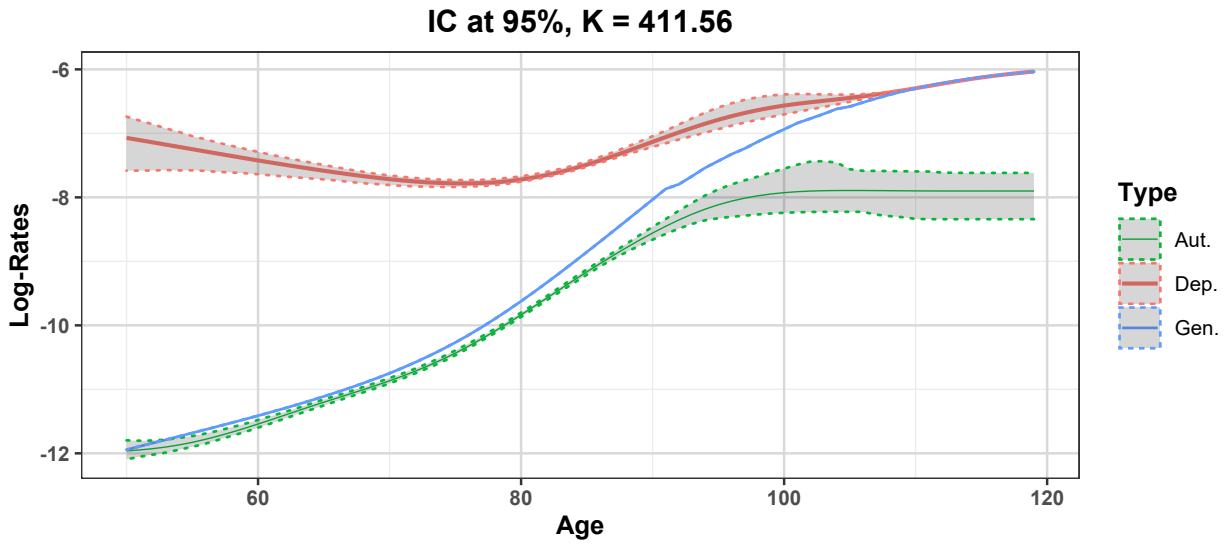


Figure 10: Confidence intervals for the GIR12 product obtained with the optimal K

Thanks to the loopback, the autonomous and LTC mortality and the incidence laws represented by

λ_x^A, λ_x^D and i_x, respectively, in Figure 1, are consistent with the general mortality of the portfolio that is known.

5.2.3 Actuarial application

Using the calibrated laws λ_x^A, λ_x^D and i_x, 50 000 trajectories of the future states of 50-year-old women are simulated. The aim is to estimate the probability for autonomous women to be in the autonomous, disabled or death state at each age above 50. The results are represented in Figure 11, where the obtained proportions in each group at each age are plotted. Starting with 100% of autonomous individuals at age 50, the proportion of autonomous individuals decreases with increasing age, since recovery is not considered in the model. Death is an absorbing state, and the proportion can only increase with age. The probability of being in LTC increases until age 93 before decreasing afterwards. Indeed, insureds can both enter and exit the LTC state. Under the calibrated biometric laws, up to age 93, the number of entries into LTC is larger than the number of deaths. This is reversed afterwards, with more deaths expected than loss of autonomy. The last survivor in this simulation dies in LTC at age 118. The last autonomous insured enters LTC at age 110 before dying.

Figure 12 represents the proportion of autonomous and disabled people among the survivors. Until age 99, there are still more autonomous than disabled insureds. For an insurer, this means that the proportion of insureds paying their premium is larger than the proportion of disabled insureds receiving an annuity.

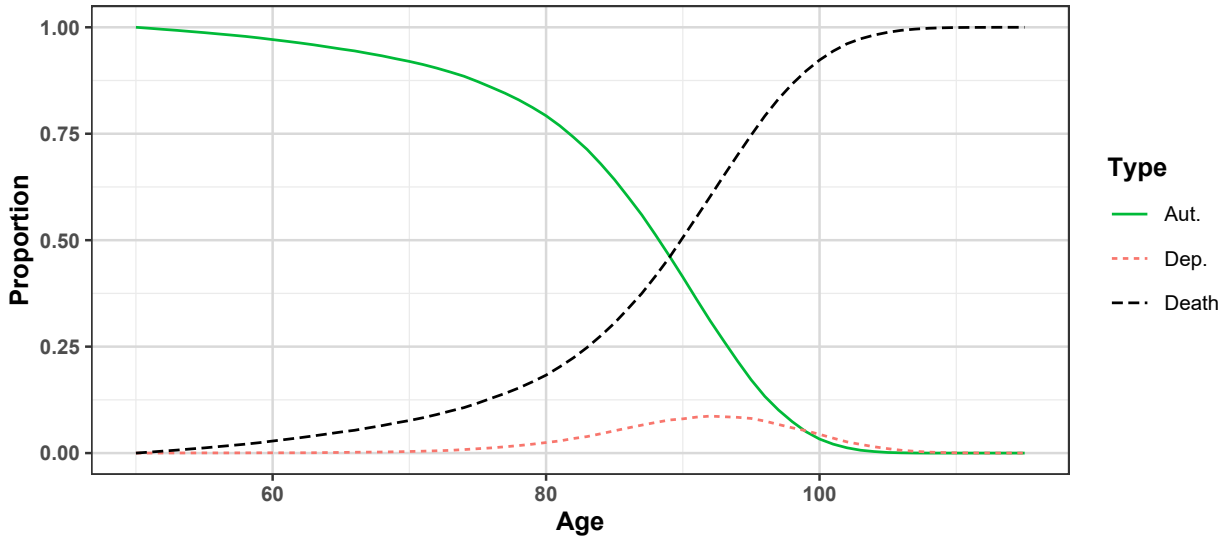


Figure 11: Proportion of insureds in each group considering a 100% autonomous population at age 50

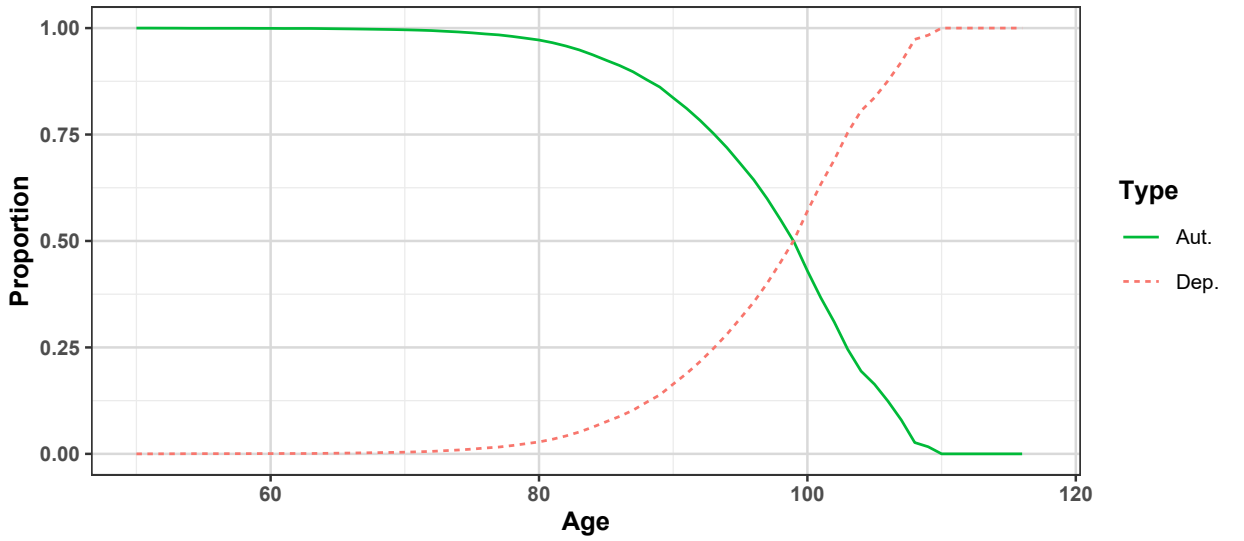


Figure 12: Proportion of autonomous insureds and annuitants considering a 100% autonomous population at age 50

6 Modelling products with several levels of dependency and allowing recovery

Let us consider in this section a product covering multiple levels of dependency, with different amounts of annuity depending on the degree of loss of autonomy.

Let us assume 3 levels of dependency:

- Total Dependency (**TD**)
- Partial Dependency (**PD**)
- Light Dependency (**LD**)

Recovering from severe dependency can be assumed to be impossible. However, one could want to allow recovery from the light level of dependency **LD**.

Section 6.1 presents two ways of modelling this product. Subsection 6.2 focuses on how to incorporate the recovery in the loopback algorithm.

6.1 Two ways of modelling a product covering multiple levels of dependency

In a first step, let consider a product without transition payments. We can model a product covering multiple levels of dependency with a 5-state Markov model (termed **Model 1** in this section), as shown in Figure 13. For clarity, the intensity notations are not mentioned in this figure, except for the incidence rates from the autonomous state **A** to the lower level of dependency **LD** denoted i_x and the recovery rates from **LD** to **A** denoted r_x.

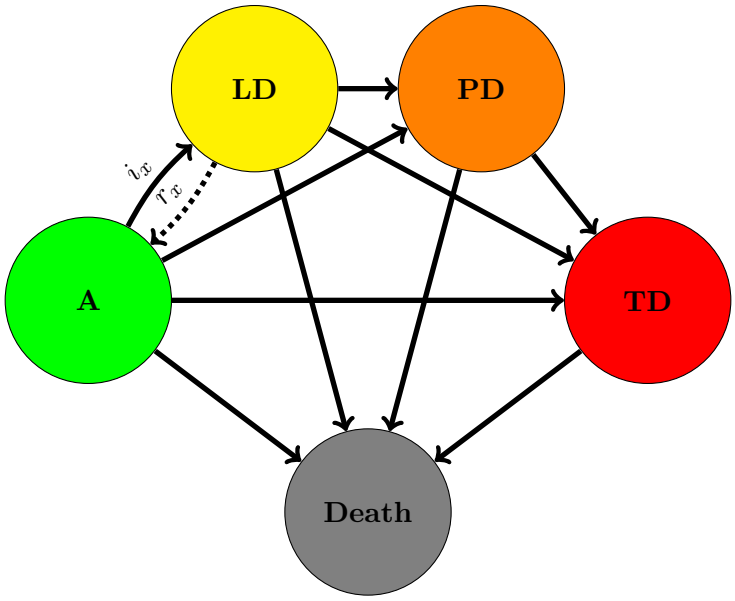
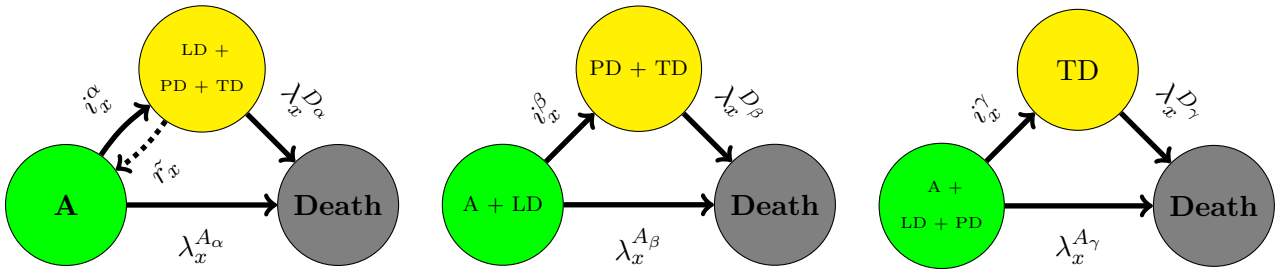


Figure 13: Modelling of an LTC product with multiple degrees of loss of autonomy (**Model 1**)

Most insurers suffer from scarcity of data due to the recency of LTC products covering multiple degrees of loss of autonomy. As a consequence, it is difficult for these insurers to calibrate this type of model without needing to introduce strong assumptions on the intensities, as in Fleischmann (2015), where, for example, the intensity to reach a specific level of dependency is assumed to be independent of the state of origin. Another way of modelling this product is to consider it as a set of 3 products. Underwriting to this product covering 3 levels of severity of loss of autonomy, with an annuity depending on this severity, is equivalent to underwriting to 3 LTC contracts denoted α, β and γ, as represented in Figure 14 and described as follows:

- Product α covers all degrees of LTC (light, partial and total dependency) with the same annuity amount. The conditions of the contract are as follows:
  - The insured pays the premium P_α as long as the insured is autonomous.
  - The insurer pays an annuity R_α as long as the insured is in light, partial or total dependency.
  - The insured can recover from dependency. In this case, the insurer stops paying the annuity, and the insured is considered autonomous.
- Product β covers partial and total dependency such that:
  - The insured pays the premium P_β as long as the insured is autonomous or in light dependency.
  - The insurer pays an annuity R_β as long as the insured is in partial or total dependency.
  - Recovery is not possible.

- Product γ covers only total dependency such that:
  - The insured pays the premium P_γ as long as the insured is alive and not in total dependency.
  - The insurer pays an annuity R_γ as long as the insured is in total dependency.
  - Recovery is not possible.



(a) Product α (b) Product β (c) Product γ

Figure 14: Modelling of an LTC product covering multiple degrees of loss of autonomy with a set of 3 LTC products (**Model 2**)

We note that $\tilde{r_x} < r_x$ because only insureds in light dependency (**LD**) can recover.

Let:

- P be the premium of the product covering the multiple levels of LTC.
- $\tilde{R}_{level}$ be the annuity paid to a dependent insured with the level of severity $level \in \{LD, PD, TD\}$.
- P_j and R_j be the premium and annuity amounts for Product $j \in \{\alpha, \beta, \gamma\}$, respectively.

Let us compare both models by analyzing the cash-flows of the insured depending on its health status. To model the multi-level product with a set of LTC products, the cash flows given in Table 1, of both models have to be equal. In this case, underwriting to a contract covering the 3 degrees of dependency is equivalent to underwriting to the 3 products described in Figure 14.

	Cash flows	
Health status	**Model 1**	**Model 2**
Autonomous	$-P$	$-P_\alpha - P_\beta - P_C$
LD	$\tilde{R}_{LD}$	$R_\alpha - P_\beta - P_\gamma$
PD	$\tilde{R}_{PD}$	$R_\alpha + R_\beta - P_\gamma$
TD	$\tilde{R}_{TD}$	$R_\alpha + R_\beta + R_\gamma$

Table 1: Comparison of cash flows of Model 1 and Model 2

A product offering transition payments, can also be considered as a set of 3 products as in Figure 14 if going through intermediate LTC states does not change the total amount received by the insured.

Here is an example of transition payments of such a product:

- In case of entry in LD from A, the insured receives a capital $\tilde{K_{LD}}$ at the time of entry in LD,
- In case of entry in PD from A, the insured receives a capital $\tilde{K_{PD}}$ at the time of entry in PD,
- In case of entry in TD from A, the insured receives a capital $\tilde{K_{TD}}$ at the time of entry in TD,
- If the insured enters state PD from state LD, then he receives $\tilde{K_{PD}} - \tilde{K_{LD}}$ at the time of entry in PD,
- If the insured enters state TD from state LD, then he receives $\tilde{K_{TD}} - \tilde{K_{LD}}$ at the time of entry in TD,
- If the insured enters state TD from state PD, then he receives $\tilde{K_{TD}} - \tilde{K_{PD}}$ at the time of entry in TD,
- An insured entering state $S \in \{LD, PD, TD\}$ from state A after recovering from LD receives $\tilde{K_S} - \tilde{K_{LD}}$ at the time of entry in S.

Model 2 is a good way to model a multi-level product when not having a large database without making strong assumptions on the shape of the rates. This model is often used by insurers.

6.2 Taking into account the possibility to recover

Product α represented in Figure 14a allows recovery. Allowing this transition has only a slight impact on the algorithm presented in this paper. In fact, allowing recovery has an impact only if one needs to estimate exposures, as in Section 3. In this case, Algorithm 1 becomes:

Algorithm 2 Exposures estimation algorithm in the case of recovery

$K \leftarrow 0$
$expo \leftarrow expo_{data}$
$c(mortality_A, mortality_D) \leftarrow loopback(data, K, expo)$
$expo \leftarrow compute\_expo(expo_{x_M}, incidence, \mathbf{recovery}, mortality_A, mortality_D)$
$K \leftarrow$ penalty parameter chosen for model calibration
for i=1,...,nb.iterations **do**
  $c(mortality_A, mortality_D) \leftarrow loopback(data, K, expo)$
  $expo \leftarrow compute\_expo(expo_{x_M}, incidence, \mathbf{recovery}, mortality_A, mortality_D)$
end for
return *expo*

where *recovery* represents the transition rates from the LTC to the autonomous state.

In this case, the function $compute\_expo(expo_{x_M}, incidence, \mathbf{recovery}, mortality_A, mortality_D)$ has to take into account the *recovery* law.

Let us assume that recoveries occur at the end of the period as the deaths and losses of autonomy in Section 3. An insured recovering at age x is in state A at age $x + 1$ and cannot enter state D a

second time or die before age $x+1$. At the i^{th} iteration,

$$expo^A_{x+1}(i) = expo^A_x(i-1)\exp\left(-(\lambda^A_x + i_x)\right) + \mathbf{expo^D_x(i-1)}\left[\mathbf{1} - \exp\left(-(\lambda^{\mathbf{D}}_{\mathbf{x}} + \tilde{\mathbf{r_x}})\right)\right]\frac{\tilde{\mathbf{r_x}}}{\lambda^{\mathbf{D}}_{\mathbf{x}} + \tilde{\mathbf{r_x}}},$$
$$expo^D_{x+1}(i) = expo^D_x(i-1)\exp\left(-(\lambda^D_x + \tilde{\mathbf{r_x}})\right) + expo^A_x(i-1)\left[1 - \exp\left(-(\lambda^A_x + i_x)\right)\right]\frac{i_x}{\lambda^A_x + i_x},$$

where $expo^G_{x+1}(i)$ denotes the exposure in group G at age x at the i_{th} iteration, and $\tilde{r_x}$ denotes the intensity rate of recovery. The terms added by allowing the recovery are highlighted in boldface.

7 Discussion

In this paper, we introduce an approach to simultaneously estimate the mortality laws of two subgroups A and D (where A and D represent the autonomous and disabled insured groups, respectively), knowing the mortality of the overall group ($A \cup D$). To do so, we rely on the P-Splines smoothing method combined with Poisson-GLM, to which we add a consistency constraint. The aim of this constraint is to link the mortality of the overall group, named general mortality in this paper, to both mortality laws in groups A and D. This constraint is based on the idea that each death in the overall group is a death in either subgroup A or subgroup D. Therefore, the sum of deaths in A and D is equal to the number of deaths in the overall group (gen). If D^G_x denotes the random variable of the death counts at age x in group $G \in \{A, D, gen\}$, then $D^{gen}_x = D^A_x + D^D_x$. As in the Poisson-GLM part of the model, we assume that the count of deaths in each group G at each age x exhibits a Poisson distribution of parameters proportional to the central exposure and the mortality intensities. This allows us to link the mortality rates of the three groups. This constraint is added in the form of a penalty in the likelihood. The mortality intensities are then estimated by maximizing the penalized log-likelihood.

We then address the problem of extrapolation of mortality laws in the case where no or not enough observations are available at old ages. This is often the case in an insurance context, particularly when estimating the risk associated with LTC products. In fact, the recency of these products combined with the fact that they are sold to individuals on average 60 years old are responsible for the data paucity beyond 85 years old. Extrapolation of mortality laws is therefore necessary for actuaries to assess the risk. We introduce an iterative approach to estimate the missing exposures at old ages. To do so, we successively estimate the mortality laws with the method described in Section 2 and then the exposures using the probability of transition from group A to D and the mortality laws from the previous step. We then re-estimate the mortality laws using the estimated exposures. These new mortality laws then lead to new estimations of the exposures. The algorithm stops when tolerance criteria are reached between successive estimations of exposures.

In the first step, the algorithm developed in this paper is tested on synthetic data. Mortality laws

are known until 119 years old, but we hide the observations above 85 to the algorithm and see how the algorithm is able to reproduce these mortality intensities between ages 86 and 119.

We introduce methods to fix the hyper-parameter and to construct confidence intervals and perform testing on the synthetic dataset. We show that our approach improves the extrapolation of the mortality laws. In fact, the extrapolation with a consistency penalty is much closer to the actual mortality intensities from 86 to 119 than the extrapolation without a penalty. In a second step, the approach presented in this paper is used on real data from five medium-to-large LTC portfolios. As with the synthetic data, we compare the results of estimations and extrapolations with and without a consistency penalty. Compared to the estimation without penalty, adding the consistency criteria results in lower estimated mortality rates in LTC (group D) at old ages and higher mortality rates for the autonomous group (group A). An insurer not using consistency criteria would overestimate the mortality of the annuitants, leading to underestimation of the provisions.

As the loopback algorithm is based on P-Splines, orders of splines penalties d introduced in Section 2.3 for each group (A and D) are considered as hyper-parameters. As seen in Section 2.3, the choice of d is crucial since it drives the age extrapolation results. In particular, without a loopback penalty, the age extrapolation is linear on the log-scale for $d = 2$ and constant for $d = 1$. Adding consistency penalty decreases the impact of this choice. The extrapolation is no longer driven only by this order but also by the consistency criteria. Nevertheless, a careful choice must be made whether one assumes that the mortality intensity continues to grow log-linearly with age even at old ages, as in Gavrilov and Gavrilova (2019), or if mortality stops growing at old ages, as in Barbi et al. (2018). In this paper, the order is fixed to 2 for disabled mortality. The order 2 allows a linear extrapolation and gives more degrees of freedom. At old ages, the probability of being autonomous is very low. Most insureds are either disabled or dead. Therefore, autonomous mortality at old ages has a relatively low impact on product pricing and reserving. In our application, fixing $d = 2$ for autonomous leads to intersecting mortality curves. In fact, in the first step of Algorithm 1, exposures are estimated with the mortality laws without a consistency constraint. With $d = 2$ for both autonomous and disabled groups, these extrapolated mortality laws at first step intersect, and the autonomous mortality is truly high at old ages (higher than the general mortality), leading to estimated exposures equal to 0 in autonomy. Therefore, for the second and next steps of the algorithm, the autonomous mortality in the constraint has a negligible or even zero weight. Hence, the constraint has only an impact on the extrapolation of the mortality in LTC, and the autonomous mortality remains higher than both the general and the disabled mortality.

In the context of modelling LTC products, many insurers use two-dimensional mortality rates for the LTC group, using semi-Markov models. In fact, mortality in LTC may depend on attained age but also on time spent in disability. Future research should implement this algorithm with a one-dimensional mortality law for group A depending on age and a two-dimensional mortality for group D depending on age and duration.

A Appendix A: Convergence of the Newton Raphson algorithm

To be the maximum penalized likelihood estimator of $\boldsymbol{\theta}$, the Hessian matrix $H_{\hat{\boldsymbol{\theta}}}$ at the final step of the algorithm has to be negative semi-definite.

Let us analyse the Hessian matrix.

- The first term $-B^T W W_{\boldsymbol{\theta}} B$ is negative semi-definite for all $\boldsymbol{\theta}$. Indeed, recalling that $W_{\boldsymbol{\theta}}$ is diagonal with only non-negative terms,

$$h^T B^T W_{\boldsymbol{\theta}} B h = (Bh)^T W_{\boldsymbol{\theta}} (Bh) \geqslant 0 \quad \forall h \in \mathbb{R}^{2M}.$$

- The second term $-P$, which does not depend on $\boldsymbol{\theta}$, is also negative semi-definite. Indeed, from 2.3, we know that $P_d = D_d^T D_d$. Therefore, $h^T P_d h \geqslant 0 \quad \forall h \in \mathbb{R}^{2M}$.

- The third term $-KB^T \left[W_{\boldsymbol{\theta}} \left([(\tilde{W}_3^{-1})^2 W_{\boldsymbol{\theta}}^Q] \otimes I_2 \right) \right] B$ is not necessarily negative semi-definite for all $\boldsymbol{\theta}$. In fact, the weight matrix $\left[W_{\boldsymbol{\theta}} \left([(\tilde{W}_3^{-1})^2 W_{\boldsymbol{\theta}}^Q] \otimes I_2 \right) \right]$ is diagonal, but not all coefficients are greater than 0 for some $\boldsymbol{\theta}$. The terms of the diagonal matrix are non-positive if some terms of $W_{\boldsymbol{\theta}}^Q$ are non-negative. This is the case when

$$\lambda_{\boldsymbol{\theta},x}^A e_x^A + \lambda_{\boldsymbol{\theta},x}^D e_x^D \leqslant \lambda_x^{gen} [e_x^A + e_x^D], \text{ for some } x_{min} \leqslant x \leqslant x_{max}.$$

- The fourth term $-K \left[\tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A \quad \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \right]^T \left[\tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A \quad \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \right]$ is negative semi-definite for all $\boldsymbol{\theta}$. In fact,

$$\begin{aligned} h^T \left[\tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A \quad \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \right]^T \left[\tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A \quad \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \right] h &= \left\| \left[\tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^A B_A \quad \tilde{W}_3^{-1} W_{\boldsymbol{\theta}}^D B_D \right] h \right\|_2^2 \\ &\geqslant 0. \end{aligned}$$

Then, a sufficient condition for $H_{\boldsymbol{\theta}}(l_{pen})$ to be negative semi-definite and therefore for $\hat{\boldsymbol{\theta}}$ to be the optimal parameter is that the third term is negative semi-definite. The condition is given by

$$\lambda_{\boldsymbol{\theta},x}^A e_x^A + \lambda_{\boldsymbol{\theta},x}^D e_x^D \leqslant \lambda_x^{gen} (e_x^A + e_x^D), \quad \forall x_{min} \leqslant x \leqslant x_{max}. \tag{15}$$

This means that the sum of the predicted number of deaths in states A and D has to be lower than or equal to the predicted number of deaths of the overall population.

References

Alegre, A., E. Pociello, A. Pons, J. Varea, and A. Vicente (2003). Actuarial valuation of long-term care annuities. *Insurance Mathematics and Economics Volume 32*.

Barbi, E., F. Lagona, M. Marsili, J. W. Vaupel, and K. W. Wachter (2018). The plateau of human mortality: Demography of longevity pioneers. *Science 360*(6396), 1459–1461.

Biessy, G. (2015). Long-Term Care insurance: A multi-state semi-Markov model to describe the dependency process in elderly people. *Bulletin Français d'Actuariat 15(29)*, 41–73.

Bollaerts, K., P. Eilers, and I. Mechelen (2006). Simple and multiple P-splines regression with shape constraints. *The British journal of mathematical and statistical psychology 59*, 451–69.

Brass, W. (1971). *Biological Aspects of Demography*. Taylor and Francis.

Camarda, C. (2019). Smooth constrained mortality forecasting. *Demographic Research 41*, 1091–1130.

Camarda, C. G., P. H. Eilers, and J. Gampe (2016). Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling 16*(4), 279–296.

Currie, I. D. and M. Durban (2002). Flexible smoothing with P-splines: A unified approach. *Statistical Modelling 2*(4), 333–349.

Currie, I. D., M. Durban, and P. H. Eilers (2004). Smoothing and forecasting mortality rates. *Statistical modelling 4*(4), 279–298.

Dupourqué, E. (2012). AGGIR, the work of grids. *Long-Term Care News 32*.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B -splines and penalties. *Statist. Sci. 11*(2), 89–121.

Eilers, P. H. C. and B. D. Marx (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics 11*(4), 758–783.

Eurostat (2022). Population structure and ageing. pp. 575–94.

Fleischmann, A. (2015). Calibrating intensities for long-term care multiple-state Markov insurance model. *European Actuarial Journal 5*, 327–354.

Gavrilov, L. and N. Gavrilova (2019). New trend in old-age mortality: Gompertzialization of mortality trajectory. *Gerontology 65*, 1–7.

Guibert, Q., S. Loisel, O. Lopez, and P. Piette (2020). Bridging the Lee-Carter's gap: a locally coherent mortality forecast approach. https://hal.archives-ouvertes.fr/hal-02472777.

Hammond, M. (2000). The forces of mortality at ages 80 to 120. *International Journal of Epidemiology 29*(2), 384–384.

Li, J. S.-H., W.-S. Chan, and R. Zhou (2017). Semicoherent multipopulation mortality modeling: The impact on longevity risk securitization. *Journal of Risk and Insurance 84*(3), 1025–1065.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of population: An extension of the Lee–Carter method. *Demography 42*, 575–94.

Macdonald, A. S., S. J. Richards, and I. D. Currie (2018). *Modelling Mortality With Actuarial Applications*. Cambridge University Press.

Marx, B. D. and P. H. Eilers (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis 28*(2), 193 – 209.

Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France) (2020). Human mortality database. www.mortality.org.

Nuttall, S. R., R. J. L. Blackwood, B. M. H. Bussell, J. P. Cliff, M. J. Cornall, A. Cowley, P. L. Gatenby, and J. M. Webber (1994). Financing long-term care in Great Britain. *Journal of the Institute of Actuaries 121*(1), 1–68.

Porta, N., G. Gomez, M. Calle, and N. r. Malats (2007). Competing risks methods. https://upcommons.upc.edu/bitstream/handle/2117/2201/TR_CR.pdf.

Remund, A., T. Riffe, and C. Camarda (2018). A cause-of-death decomposition of the young adult mortality hump. *Demography 55*, 957–978.

Ruppert, D. (2000). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics Volume 11*, 735–757.

Zhou, R., G. Xing, and M. Ji (2019). Changes of relation in multi-population mortality dependence: An application of threshold VECM. *Risks 7*(1).