Eur. Actuar. J. (2012) 2:105–127
DOI 10.1007/s13385-012-0047-3

ORIGINAL RESEARCH PAPER

# A subordinated Markov model for stochastic mortality

**Xiaoming Liu · X. Sheldon Lin**

Received: 16 August 2011 / Revised: 16 March 2012 / Accepted: 27 March 2012 /
Published online: 18 April 2012


**Abstract** In this paper we propose a subordinated Markov model for modeling stochastic mortality. The aging process of a life is assumed to follow a finite-state Markov process with a single absorbing state and the stochasticity of mortality is governed by a subordinating gamma process. We focus on the theoretical development of the model and have shown that the model exhibits many desirable properties of a mortality model and meets many model selection criteria laid out in Cairns et al. (ASTIN Bull 36:79–120, 2006; Scand Actuar J 2:79–113, 2008). The model is flexible and fits either historical mortality data or projected mortality data well. We also explore applications of the proposed model to the valuation of mortality-linked securities. A general valuation framework for valuing mortality-linked products is presented for this model. With a proposed risk loading mechanism, we can make an easy transition from the physical measure to a risk-neutral measure and hence is able to calibrate the model to market information. The phase-type structure of the model allows us to apply the matrix-analytic methods that have been extensively used in ruin theory in actuarial science and queuing theory in operations research (see Asmussen in Applied probability and queues. Wiley, New York, 1987; Asmussen and Albrecher in Ruin probabilities, 2nd edn. World Scientific Publishing, Singapore, 2010; Neuts in Matrix-geometrix solutions in stochastic models. Johns Hopkins University Press, Baltimore, 1981). As a result, many quantities of interest such as the distribution of future survival rates, prediction intervals, the term structure of mortality as well as the value of caps and floors on the survival index can be obtained analytically.

X. Liu
Department of Statistical and Actuarial Sciences, University of Western Ontario,
London, ON N6A 5B7, Canada
e-mail: xliu@stats.uwo.ca

X. S. Lin (✉)
Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, Canada
e-mail: sheldon@utstat.utoronto.ca

## 1 Introduction and motivation

Actuaries have been using extrapolative methods to project mortality rates for centuries, with the implication that the past represents the future. Consequently, traditional actuarial approaches for the pricing and risk management of life insurance and annuity products treat mortality rates deterministically. In other words, the calculation of premiums and risk reserves is based on pre-specified deterministic mortality schedule, with a belief that the differences between the projected rates and realized rates—the so-called mortality risk—can be diversified among individuals and/or over the time.

Over the last century, it has become evident that mortality risk is neither predictable nor diversifiable. In fact, the mortality projections in the last 50 years have systematically underestimated the overall mortality improvement. The consequent adverse financial impacts caused by mis-assessing mortality risk have been blamed as one of the reasons for the insolvency of the Equitable Life Assurance (UK) in 2000. For empirical studies on mortality trend changes and mortality projections, see [40, 41], GAD National Statistics Quality Review Report No. 8 [22], and a series of working papers by CMI [14–16]. Also see [4, 9, 37] for financial impacts of mortality risk on life insurances and annuities.

Many attempts have been made in the past decades to explore the use of stochastic approaches in modeling mortality risk and evaluating mortality-linked securities. The Lee-Carter time series model and its variations have been the most popular in modeling human mortality. See [6, 10, 27, 38, 39], and references therein. Based on the similar idea, Cairns et al. [12] propose a two-factor time series mortality model. The use of time series models by far have largely been limited to mortality prediction. The lack of mathematically desirable distributional properties and the difficulty in incorporating the martingale pricing principle often make the time series models challenging to use in the valuation of insurance products and mortality derivatives. In an effort to overcome these shortcomings, many researchers have recently turned to diffusion models and adopted affine interest rate models for mortality modeling, making use of many conceptual similarities between the mortality risk and the interest rate risk. See [5, 11, 13, 17, 18, 19, 34, 35], and references therein.

Despite many advantages of using affine-type diffusion models, there are some disadvantages. First, it can be difficult to establish a linkage between a diffusion model and the human aging process. We may not be able to interpret the diffusion model in a satisfactory manner. Second, external factors that influence interest rates and mortality rates are fundamentally different. Interest rates are influenced by economic factors and financial market conditions. Hence, interest rate dynamics often exhibit damped cyclical behavior. A mean-reversion diffusion model is suitable. On the other hand, mortality rates are influenced by medical and other life and environmental sciences' advances. As a result, mortality dynamics tend to have a downward trend, which might not be modeled by an affine model.

In this paper, we propose an alternative approach to model stochastic mortality. We start with a finite-state Markov process with one absorbing state to model the aging process of a life. The survival time or time until death of the life is then

defined as the time until absorption of the Markov process. The resulting distribution for the survival time is a so-called phase-type distribution. This approach is different from the traditional curve fitting mortality models in the sense that the survival probabilities are now linked with a Markov aging process. Stochasticity in mortality is introduced via a stochastic time change to (or subordination of) the underlying Markov process. The idea of changing time in a stochastic process in financial modeling has first been used by Madan and his collaborators to model the dynamics of the stock price (see [32, 33]). Recently, [24] use the same idea to model credit risk. But this method has not been considered in mortality modeling, to the best of our knowledge. As will be shown in this paper, the resulting stochastic mortality model has some desirable properties that reflect the stylised facts of the mortality risk. Furthermore, the term structure of mortality under the model has a phase-type representation. As a result, we are able to take advantage of the properties of phase-type distributions and matrix-analytic methods that have been extensively used in ruin theory in actuarial science and queuing theory in operations research (see [2, 3, 36]), in analyzing and evaluating the model, as well as in pricing mortality-linked products.

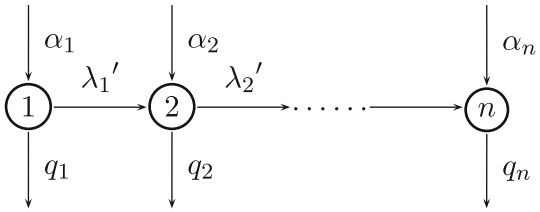
The paper is organized as follows. In Sect. 2, we describe the finite-state continuous-time Markov model for the aging process of a life. The model was first proposed in [30]. In Sect. 3, we propose a subordinated Markov aging process based on the Markov model in Sect. 2. The term structure of mortality and certain distributional properties are derived. The matrix-analytic methods are extensively employed. In Sect. 4, we fit the subordinated Markov model to the mortality projections made by the UK Government Actuary's Department (GAD). In Sect. 5, we first provide a general valuation framework for mortality-linked products under the subordinated Markov model and for illustration purposes, we then calibrate the model term structure of mortality to the EIB longevity bonds and compute the value of caplets written on a survival index. Section 6 concludes the paper with remarks on potential improvement of the model.

## 2 A deterministic phase-type mortality model

Traditional survival analysis has focused on modeling deterministic survival rates or the hazard rates (the force of mortality) of a population, as seen in Gompertz's model and its various extensions. In spite of apparent simplicity, these quantities are highly aggregating and affected by many factors. Therefore, mortality prediction models based on them are difficult to interpret. A process-based approach, viewing the survival time of a life as the occurrence time of certain events of a process, may be of certain advantages, as it may provide biological interpretations for a model and may utilize mathematical tools developed in the theory of stochastic processes.

In this section, we present a finite-state continuous-time Markov aging process $\{J_t;\ t \geq 0\}$ of Coxian type, as proposed in [30]. The state space of the Markov process is assumed to consist of a set of transient states $E = \{1, 2, \ldots, n\}$ that represent chronological health statuses before death (or physiological ages as termed in [30]) of a generic life aged $x$ at time 0 from a cohort and a single

absorbing state $\Delta$ representing the death. The process of transition from one state to the next state before death is referred to as the aging process, denoted by $J_t$ and described in the following diagram:




where $\lambda_i' > 0$ denotes the aging rate from status $i$ to status $i+1$, $q_i > 0$ denotes the death rate of the life given that the life is at status $i$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ with $\sum_{i=1}^{n} \alpha_i = 1$ is the initial probability distribution that may be interpreted as the distribution of health statuses among a particular cohort of age $x$ at time 0.

The intensity matrix for the transient states is thus given by

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1' & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2' & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_n \end{pmatrix}, \tag{1}$$

where

$$\lambda_i = \lambda_i' + q_i. \tag{2}$$

The aging process $J_t$ hence is a Markov process on a finite state-space $S = E \cup \Delta = \{1, 2, \ldots, n\} \cup \Delta$ with initial distribution being $(\boldsymbol{\alpha}, 0)$. Its intensity matrix is

$$\begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{q} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{3}$$

where $\boldsymbol{q} = -\boldsymbol{\Lambda}\boldsymbol{e}$, and $\boldsymbol{e}$ is the column vector of ones.

Let $\tau$ denote the time till absorption of the Markov process. Since $J_t$ represents the aging process, $\tau$ represents the time-till-death of life $(x)$, and it has a phase-type distribution with phase-type representation $(\boldsymbol{\alpha}, \boldsymbol{\Lambda})$ of order $n$. See [3, 36] for the definition of phase-type distributions and their properties. Applications of phase-type distributions as survival models can be found in [1, 26, 30, 31], and more recently, [21, 23].

One of the main advantages to use phase-type distributions is their mathematical tractability. With the help of matrix-analytic methods, one can compute various quantities of interest associated with a phase-type distribution such as its survival function and moments as well as its moment generating function. For example, the survival function of the future lifetime of $(x)$ based on the phase-type model (3) is

$$S_0(t) = \boldsymbol{\alpha} e^{\boldsymbol{\Lambda} t} \boldsymbol{e}, \quad t > 0. \tag{4}$$

Its moment generating function is

$$M_0(s) = \boldsymbol{\alpha}(-s\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}\boldsymbol{q}; \tag{5}$$

and non-central moments are

$$m_k = (-1)^k k! \boldsymbol{\alpha} \boldsymbol{\Lambda}^{-k} \boldsymbol{e}, \quad k = 1, 2, \ldots. \tag{6}$$

Furthermore, at any future time $s > 0$, the conditional survival function of the life, given survival at age $x + s$, is

$$\begin{aligned} \frac{S_0(t+s)}{S_0(s)} &= \frac{\boldsymbol{\alpha} e^{\boldsymbol{\Lambda}(s+t)} \boldsymbol{e}}{\boldsymbol{\alpha} e^{\boldsymbol{\Lambda} s} \boldsymbol{e}} \\ &= \boldsymbol{\alpha}_s e^{\boldsymbol{\Lambda} t} \boldsymbol{e}, \quad t > 0, \end{aligned} \tag{7}$$

where

$$\boldsymbol{\alpha}_s = \frac{\boldsymbol{\alpha} e^{\boldsymbol{\Lambda} s}}{\boldsymbol{\alpha} e^{\boldsymbol{\Lambda} s} \boldsymbol{e}}. \tag{8}$$

The future lifetime at time $s$ again has a phase-type distribution, with $\boldsymbol{\alpha}_s$ representing the probability distribution of health statuses of the cohort at time $s$.

Lin and Liu [30] have used this modeling framework to fit the entire schedule of human mortality. It has been shown that the resulting phase-type model is flexible and can fit well to different shapes of mortality pattern exhibited in historical mortality data. In order to reach high fitting accuracy for fitting a mortality schedule from age 0 up to age 100, the model adopted in their paper has a relative high dimension. Note that we can use formulas (7) and (8) to derive a model for lives aged $x$ from the model given in [30]. However, as shown in Figure 8 of [30], the distribution $\boldsymbol{\alpha}_s$ shifts its weights from early states to later states of the aging process. It indicates that we can adopt a lower dimensional Markov model for adult population and eliminate the developmental phases used in [30]. Moreover, we may relax the constraints on the transition rates that are specified in [30]. By doing so, we will be able to fit the model better to adult mortality data.

In this paper, we refit the model to the mortality rate at ages greater than or equal to $x$, using a lower dimensional phase-type distribution.[1] There are two other reasons for this. First, in many actuarial valuation problems and the problems regarding mortality-linked products in particular, we only need to focus on the population of age 65 or older. Thus, only part of the mortality schedule is relevant and a low-dimensional phase-type distribution is often sufficient. Second, a low-dimensional phase-type distribution allows us to use computational tools such as Matlab to obtain numerical results effectively and precisely. There is a trade off between computational efficiency and model fitting accuracy. Since there is a re-calibration of the model to the market information when the model is used for pricing purpose, the impact of using a lower-dimensional model should be minor.

[1] An example of the fitted model based on a 5-state Markov aging process is given in (27) and (28).

## 3 Stochastic mortality via time-change

The Markov aging process presented in Sect. 2 results in a deterministic survival function $S_0(t)$ given in formula (4) for the future lifetime random variable $\tau$. In this section, we incorporate stochasticity into this survival function by introducing a time-changed Markov process. For an individual aged $x$ from the cohort of interest, we assume that his/her underlying aging process follows the subordinated Markov process

$$Z_t = J_{\gamma_t}, \tag{9}$$

where $J_t$ is the Markov aging process as described in Sect. 2 that is now termed as the baseline process, and $\gamma_t$ is a nondecreasing continuous-time stochastic process. For notational convenience, age $x$ is suppressed throughout the rest of this paper. This idea of using a subordinated Markov process is similar to that in the works of Madan and his collaborators who used a subordinated Brownian motion to model the dynamics of the logarithm of stock prices (see [32, 33]). We may interpret that the baseline aging process $J_t$ models the uncertainty in a generic individual's survival time under the past living environment and the time changing process models the influences due to random effects that improve or worsen the living environment and attribute to systematic changes in the future. The realized survival probability at time $t$, $S(t)$, is now subject to the modification of the random time-change process $\gamma_t$ over period $[0, t]$. That is, under the model (9), we have

$$S(t) = S_0(\gamma_t). \tag{10}$$

Under this stochastic mortality modeling framework, the survival probability $S(t)$ is now a stochastic process and is referred to as the survival index for the cohort (see [11]).

### 3.1 Gamma time-change process

As in [32], in this paper the time-change process $\gamma_t$ is chosen to be a gamma process that is defined as follows:

1. $\gamma_0 = 0$;
2. it has independent increments, i.e., for any partition $0 \le t_0 < t_1 < \cdots < t_n$, the random variables $\gamma_{t_1} - \gamma_{t_0}, \ldots, \gamma_{t_n} - \gamma_{t_{n-1}}$ are mutually independent; and
3. the increment $\gamma_{t+s} - \gamma_t$ has a gamma distribution with mean $s$ and variance $vs$, for any $s, t \ge 0$. Thus the parameter $v$ may be interpreted as the volatility of the process. As will be seen in later sections, it represents the level of longevity risk under the proposed model.

It follows from (4) and (10) that

$$S(t) = S_0(\gamma_t) = \boldsymbol{\alpha}\, e^{\boldsymbol{\Lambda}\gamma_t}\, \boldsymbol{e}, \quad t > 0. \tag{11}$$

It will be shown that, with the gamma time-change process, the stochastic mortality model (9) not only becomes mathematically tractable but also exhibits many desirable properties of a mortality model.

### 3.2 Term structure of mortality

Let $\{\mathcal{F}_t;\, t \geq 0\}$ be the filtration generated by the survival index $S(t)$, $t \geq 0$. Further, for $0 \leq s \leq t$, let $P(s,t)$ be the survival function of a life aged $x$ at time 0 to be alive from time $s$ to time $t$ that is measured at time $s$. $\{P(s,t); 0 \leq s \leq t\}$ is commonly referred to as the term structure of stochastic mortality (see [11]). Thus, it follows from (11) that

$$P(s,t) = \frac{1}{S(s)} E[S(t)|\mathcal{F}_s]. \tag{12}$$

In the following, we present an explicit analytical expression for $P(s,t)$.

**Theorem 3.1** *Suppose that the eigenvalues $-\lambda_1, \ldots, -\lambda_n$ of the intensity matrix $\mathbf{\Lambda}$ are distinct. Let $\boldsymbol{h_1}, \ldots, \boldsymbol{h_n}$ and $\boldsymbol{v_1}, \ldots, \boldsymbol{v_n}$ be their corresponding right and left eigenvectors such that $\boldsymbol{v_i h_i} = 1$. It is known that $\boldsymbol{v_i h_j} = 0$, $i \neq j$, $i,j = 1, \ldots, n$. Then, $P(s,t)$ has the phase-type representation $(\boldsymbol{\alpha}_{\gamma_s}, \tilde{\mathbf{\Lambda}})$, where*

$$\boldsymbol{\alpha}_{\gamma_s} = \frac{\boldsymbol{\alpha}\, e^{\mathbf{\Lambda}\gamma_s}}{\boldsymbol{\alpha} e^{\mathbf{\Lambda}\gamma_s} \boldsymbol{e}}, \tag{13}$$

*and*

$$\tilde{\mathbf{\Lambda}} = -\sum_{i=1}^{n} \tilde{\lambda}_i \boldsymbol{h_i}\, \boldsymbol{v_i}, \tag{14}$$

*with $\tilde{\lambda}_i$ being given by*

$$\tilde{\lambda}_i = \frac{1}{v} \ln(1 + v\lambda_i).$$

*That is,*

$$P(s,t) = \boldsymbol{\alpha}_{\gamma_s} e^{\tilde{\mathbf{\Lambda}}(t-s)} \boldsymbol{e}, \quad 0 \leq s \leq t. \tag{15}$$

*As a special case, the term structure at time 0 is given by*

$$P(0,t) = \boldsymbol{\alpha}\, e^{\tilde{\mathbf{\Lambda}}t} \boldsymbol{e}, \quad t \geq 0. \tag{16}$$

*Proof* Since the eigenvalues of the intensity matrix $\mathbf{\Lambda}$ are distinct, $\mathbf{\Lambda}$ is diagonalizable and has the following eigendecomposition:

$$\mathbf{\Lambda} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^{-1}, \tag{17}$$

where $\boldsymbol{H} = (\boldsymbol{h_1}, \ldots, \boldsymbol{h_n})$, $\boldsymbol{H}^{-1} = (\boldsymbol{v_1}, \ldots, \boldsymbol{v_n})^t$ and $\boldsymbol{D} = diag(-\lambda_1, \ldots, -\lambda_n)$.

We can rewrite (17) as

$$\mathbf{\Lambda} = -\sum_{i=1}^{n} \lambda_i \boldsymbol{h_i}\, \boldsymbol{v_i} = -\sum_{i=1}^{n} \lambda_i \boldsymbol{h_i} \otimes \boldsymbol{v_i}. \tag{18}$$

where $\otimes$ is the symbol for the Kronecker product.

It follows from (A.16) and (A.17) of [3] that

$$e^{\mathbf{\Lambda} t} = \sum_{i=1}^{n} e^{-\lambda_i t} \boldsymbol{h_i} \otimes \boldsymbol{v_i} = \sum_{i=1}^{n} e^{-\lambda_i t} \boldsymbol{h_i}\, \boldsymbol{v_i}.$$

Write

$$\frac{S(t)}{S(s)} = \frac{\boldsymbol{\alpha}\, e^{\mathbf{\Lambda}\gamma_t} \boldsymbol{e}}{\boldsymbol{\alpha}\, e^{\mathbf{\Lambda}\gamma_s} \boldsymbol{e}} = \boldsymbol{\alpha}_{\gamma_s}\, e^{\mathbf{\Lambda}(\gamma_t - \gamma_s)}\, \boldsymbol{e}. \tag{19}$$

Then we have

$$\begin{aligned} P(s,t) &= E\left[\frac{S(t)}{S(s)}\middle| \mathcal{F}_s\right] \\ &= E[\boldsymbol{\alpha}_{\gamma_s}\, e^{\mathbf{\Lambda}(\gamma_t - \gamma_s)}\, \boldsymbol{e} | \mathcal{F}_s] \\ &= \boldsymbol{\alpha}_{\gamma_s} E[e^{\mathbf{\Lambda}(\gamma_t - \gamma_s)} | \mathcal{F}_s]\, \boldsymbol{e} \\ &= \boldsymbol{\alpha}_{\gamma_s} \left\{ \sum_{i=1}^{n} E[e^{-\lambda_i(\gamma_t - \gamma_s)} | \mathcal{F}_s]\, \boldsymbol{h_i}\, \boldsymbol{v_i} \right\} \boldsymbol{e}. \end{aligned}$$

Since $\gamma_t - \gamma_s$ is independent of $\mathcal{F}_s$ and has a gamma distribution with mean $t - s$ and variance $v(t-s)$, we have

$$\begin{aligned} P(s,t) &= \boldsymbol{\alpha}_{\gamma_s} \left\{ \sum_{i=1}^{n} E[e^{-\lambda_i(\gamma_t - \gamma_s)}]\, \boldsymbol{h_i}\, \boldsymbol{v_i} \right\} \boldsymbol{e} \\ &= \boldsymbol{\alpha}_{\gamma_s} \left\{ \sum_{i=1}^{n} (1 + v\lambda_i)^{-\frac{t-s}{v}}\, \boldsymbol{h_i}\, \boldsymbol{v_i} \right\} \boldsymbol{e} \\ &= \boldsymbol{\alpha}_{\gamma_s} \left\{ \sum_{i=1}^{n} e^{-\tilde{\lambda}_i (t-s)}\, \boldsymbol{h_i}\, \boldsymbol{v_i} \right\} \boldsymbol{e} = \boldsymbol{\alpha}_{\gamma_s}\, e^{\tilde{\mathbf{\Lambda}}(t-s)} \boldsymbol{e}. \end{aligned} \tag{20}$$

The proof is complete. □

Further, since the matrix $\tilde{\mathbf{\Lambda}}$ has expression (14), $-\tilde{\lambda}_1, \ldots, -\tilde{\lambda}_n$ are its eigenvalues and

$$\boldsymbol{H}^{-1} \tilde{\mathbf{\Lambda}} \boldsymbol{H} = \tilde{\boldsymbol{D}} = diag(-\tilde{\lambda}_1, \ldots, -\tilde{\lambda}_n). \tag{21}$$

Furthermore, formula (16) indicates that $P(0,t)$ has a phase-type representation; hence, the matrix analytic methodology developed for phase-type distributions is applicable.

In the following, we derive the variance of the survival index at time $t$ using the Kronecker operations. Let $\boldsymbol{D} \oplus \boldsymbol{D}$ be the Kronecker sum of $\boldsymbol{D}$ to itself. Then

$$\boldsymbol{D} \oplus \boldsymbol{D} = diag(\boldsymbol{D} - \lambda_1 \boldsymbol{I}, \boldsymbol{D} - \lambda_2 \boldsymbol{I}, \ldots, \boldsymbol{D} - \lambda_n \boldsymbol{I}).$$

It is obvious it is a diagonal matrix whose diagonal entries are $-\zeta_k, k = 1, \ldots, n^2$ with $\zeta_{i+j} = \lambda_i + \lambda_j, i, j = 1, \ldots, n$. Similar to the definition of $\tilde{\lambda}_i$, define

$$\tilde{\zeta}_k = \frac{1}{v}\ln(1+v\zeta_k),$$

and $\boldsymbol{D}\widetilde{\oplus}\boldsymbol{D} = diag(-\tilde{\zeta}_1,\ldots,-\tilde{\zeta}_{n^2})$.

**Theorem 3.2** *Denote*

$$\boldsymbol{\Lambda}\widetilde{\oplus}\boldsymbol{\Lambda} = (\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{D}\widetilde{\oplus}\boldsymbol{D})(\boldsymbol{H}\otimes\boldsymbol{H})^{-1}.$$

*Then, the variance of $S(t)$ is given by*

$$Var[S(t)] = (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})[e^{(\widetilde{\boldsymbol{\Lambda}\oplus\boldsymbol{\Lambda}})t} - e^{(\tilde{\boldsymbol{\Lambda}}\oplus\tilde{\boldsymbol{\Lambda}})t}](\boldsymbol{e}\otimes\boldsymbol{e}). \tag{22}$$

*Proof* By using the Kronecker operations on matrices, it follows from (12) and (16) that

$$\begin{aligned}(E[S(t)])^2 &= \boldsymbol{\alpha}e^{\tilde{\boldsymbol{\Lambda}}t}\boldsymbol{e}\boldsymbol{\alpha}e^{\tilde{\boldsymbol{\Lambda}}t}\boldsymbol{e}\\ &= (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})[e^{\tilde{\boldsymbol{\Lambda}}t}\otimes e^{\tilde{\boldsymbol{\Lambda}}t}](\boldsymbol{e}\otimes\boldsymbol{e})\\ &= (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})\,[e^{(\tilde{\boldsymbol{\Lambda}}\oplus\tilde{\boldsymbol{\Lambda}})t}](\boldsymbol{e}\otimes\boldsymbol{e}).\end{aligned}$$

Similarly, it follows from (11) that the second moment of the survival index random variable $S(t)$ can be written as

$$E[(S(t))^2] = (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})E[e^{(\boldsymbol{\Lambda}\oplus\boldsymbol{\Lambda})\gamma_t}](\boldsymbol{e}\otimes\boldsymbol{e}). \tag{23}$$

Since

$$\begin{aligned}\boldsymbol{\Lambda}\oplus\boldsymbol{\Lambda} &= \boldsymbol{\Lambda}\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{\Lambda}\\ &= \boldsymbol{HDH^{-1}}\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{HDH^{-1}}\\ &= \boldsymbol{HDH^{-1}}\otimes\mathbf{HIH^{-1}}+\mathbf{HIH^{-1}}\otimes\boldsymbol{HDH^{-1}}\\ &= (\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{D}\otimes\boldsymbol{I})(\boldsymbol{H^{-1}}\otimes\boldsymbol{H^{-1}})+(\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{I}\otimes\boldsymbol{D})(\boldsymbol{H^{-1}}\otimes\boldsymbol{H^{-1}})\\ &= (\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{D}\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{D})(\boldsymbol{H^{-1}}\otimes\boldsymbol{H^{-1}})\\ &= (\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{D}\oplus\boldsymbol{D})(\boldsymbol{H^{-1}}\otimes\boldsymbol{H^{-1}})\\ &= (\boldsymbol{H}\otimes\boldsymbol{H})(\boldsymbol{D}\oplus\boldsymbol{D})(\boldsymbol{H}\otimes\boldsymbol{H})^{-1},\end{aligned} \tag{24}$$

the proof of Theorem 3.1 can then be carried out exactly for the expectation (23) to show that the second moment has the phase-type representation $(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}, \boldsymbol{\Lambda}\widetilde{\oplus}\boldsymbol{\Lambda})$. Equation (22) follows immediately from

$$Var[S(t)] = E[(S(t))^2] - (E[S(t)])^2.$$

□

We would also like to present an alternative formula to compute the second moment $E[(S(t))^2]$ at time $t$ where $t/v$ is an integer. In this case, $\gamma_t$ has an Erlang distribution and its density function can be expressed in a phase-type form:

$$g_{\gamma_t}(x) = \boldsymbol{b}\, e^{\boldsymbol{\Gamma} x} \boldsymbol{I}_{\frac{1}{v}},$$

where $m = t/v, \boldsymbol{b} = (1\,0\ldots0)_{1\times m}$, $\boldsymbol{I}_{\frac{1}{v}} = (0,\ldots0,\frac{1}{v})_{m\times 1}$, and

$$\boldsymbol{\Gamma} = \begin{pmatrix} -\frac{1}{v} & \frac{1}{v} & 0 & \cdots & 0 \\ 0 & -\frac{1}{v} & \frac{1}{v} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{v} \end{pmatrix}_{m\times m}.$$

Note that the dimension of the vectors $\boldsymbol{b}$ and $\boldsymbol{I}_{\frac{1}{v}}$ as well as the matrix $\boldsymbol{\Gamma}$ are functions of time $t$.

Following (23), we have

$$\begin{aligned} E[(S(t))^2] &= \int 0^{\infty} (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})(e^{\boldsymbol{\Lambda} x}\otimes e^{\boldsymbol{\Lambda} x})(\boldsymbol{e}\otimes\boldsymbol{e})\,\boldsymbol{b}\, e^{\boldsymbol{\Gamma} x}\boldsymbol{I}_{\frac{1}{v}}\,dx \\ &= (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}\otimes\boldsymbol{b})\left(\int_0^{\infty} e^{(\boldsymbol{\Lambda}\oplus\boldsymbol{\Lambda}\oplus\boldsymbol{\Gamma})x}\,dx\right)(\boldsymbol{e}\otimes\boldsymbol{e}\otimes\boldsymbol{I}_{\frac{1}{v}}) \\ &= -(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}\otimes\boldsymbol{b})(\boldsymbol{\Lambda}\oplus\boldsymbol{\Lambda}\oplus\boldsymbol{\Gamma})^{-1}(\boldsymbol{e}\otimes\boldsymbol{e}\otimes\boldsymbol{I}_{\frac{1}{v}}). \end{aligned} \tag{25}$$

This will provide us the variance at points $t = mv$ where $m$ is a positive integer. The advantage of formula (25) is that the computation does not depend on the eigenvector matrices $\boldsymbol{H}$ and $\boldsymbol{H}^{-1}$. When the eigenvalues of the intensity matrix $\boldsymbol{\Lambda}$ are very close to each other or when the dimension of matrix $\boldsymbol{\Lambda}$ is very high, the eigenvector matrices might be numerically unstable and hence might lead to incorrect numerical results. Since formula (25) avoids the use of the eigenvectors, it can always be computed accurately. We thus may use it to check the accuracy of the variance computation.

## 4 Fitting model to mortality projections

In the previous section, we propose a subordinated Markov model to describe stochasticity of mortality. We further derive an explicit formula for its term structure $P(s,t)$ under this model framework, as given in Theorem 3.1. We remark that the values of $P(s,t)$ reflect not only the baseline mortality law of the cohort but also the uncertainty of the mortality risk. As a result, when $P(s,t)$ is calculated under the physical probability measure $P$, we may use the projected mortality data to calibrate the model to obtain the values of the model parameters.

In this section, we will fit the subordinated Markov model to data. The data we are using are those in [12] and listed in column 1 of Table 1. They are the mortality projections, made by the UK Government Actuary's Department (GAD), of the survival rates for UK males aged 65 in Year 2003. We fit a 5-state Markov model subordinated by the gamma process to the projection data using the least square method. The objective function is set to be

$$\min \sum_{t=1}^{25} [P(0,t) - \hat{P}(0,t)]^2, \tag{26}$$

where $P(0,t)$ is the value calculated by (16) and $\hat{P}(0,t), t = 1, \ldots, 25$, are the GAD projections given in column 1 of Table 1. The fitted model has parameter $v_0 = 0.36934357$ for the gamma process and the phase-type representation $(\boldsymbol{\alpha}, \boldsymbol{\Lambda})$ given as follows:

$$\begin{aligned} \boldsymbol{\alpha} = (&0.86290752, 1.00009920e-03, 1.00009920e \\ &-03, 0.035092392, 0.09999989), \end{aligned} \tag{27}$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} -0.2381937 & 0.2378846 & 0 & 0 & 0 \\ 0 & -0.2384834 & 0.237982 & 0 & 0 \\ 0 & 0 & -0.2390828 & 0.2380828 & 0 \\ 0 & 0 & 0 & -0.2423468 & 0.2413468 \\ 0 & 0 & 0 & 0 & -0.2433471 \end{pmatrix}. \tag{28}$$

The five eigenvalues of $\boldsymbol{\Lambda}$ are $-0.2381937$, $-0.2384834$, $-0.2390828$, $-0.2423468$, and $-0.2433471$, and the eigenvector matrices are

$$\boldsymbol{H} = \begin{pmatrix} 1 & -821.1412 & 106228.9990 & -257363.0801 & 3.042575423\ \mathrm{e}+07 \\ 0 & 1 & -397.0337 & 4493.1644 & -659126.6599 \\ 0 & 0 & 1 & -72.9420 & 13470.7429 \\ 0 & 0 & 0 & 1 & -241.2744 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\boldsymbol{H}^{-1} = \begin{pmatrix} 1 & 821.1412 & 219791.7224 & 1.259989615\ \mathrm{e}+07 & 5.90085112\ \mathrm{e}+08 \\ 0 & 1 & 397.0337 & 24467.2814 & 1214116.8407 \\ 0 & 0 & 1 & 72.9420 & 4128.3040 \\ 0 & 0 & 0 & 1 & 241.2744 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Applying Theorem 3.1, we have

$$\tilde{\boldsymbol{\Lambda}} = \begin{pmatrix} -0.22829274 & 0.21863814 & 0.00882993 & 0.000475265 & 0.0000293618 \\ 0 & -0.22855900 & 0.21869466 & 0.00882899 & 0.000481546 \\ 0 & 0 & -0.22910983 & 0.21864396 & 0.00894398 \\ 0 & 0 & 0 & -0.23210733 & 0.221481270 \\ 0 & 0 & 0 & 0 & -0.23302529 \end{pmatrix}.$$

The fitted survival rates based on the above estimated parameter and phase-type representation are shown in Table 1. The least square error of the fitted model is 0.0024369559.

**Table 1** GAD projected survival rates, and values of $P(0,t)$ and $S_0(t), t = 1, 2, \ldots, 25$

| Column# t | 1 GAD projection | 2 $P(0,t)$ | 3 $S_0(t)$ |
|---|---|---|---|
| 1 | 0.9836 | 0.9778 | 0.9772 |
| 2 | 0.9661 | 0.9584 | 0.9576 |
| 3 | 0.9475 | 0.9410 | 0.9405 |
| 4 | 0.9278 | 0.9245 | 0.9246 |
| 5 | 0.9068 | 0.9076 | 0.9088 |
| 6 | 0.8845 | 0.8894 | 0.8919 |
| 7 | 0.8610 | 0.8691 | 0.8728 |
| 8 | 0.830 | 0.8462 | 0.8508 |
| 9 | 0.8095 | 0.8202 | 0.8256 |
| 10 | 0.7816 | 0.7913 | 0.7970 |
| 11 | 0.7522 | 0.7594 | 0.7650 |
| 12 | 0.7213 | 0.7250 | 0.7301 |
| 13 | 0.6888 | 0.6885 | 0.6927 |
| 14 | 0.6548 | 0.6503 | 0.6533 |
| 15 | 0.6195 | 0.6110 | 0.6127 |
| 16 | 0.5828 | 0.5712 | 0.5713 |
| 17 | 0.5448 | 0.5312 | 0.5298 |
| 18 | 0.5059 | 0.4917 | 0.4888 |
| 19 | 0.4661 | 0.4531 | 0.4486 |
| 20 | 0.4258 | 0.4156 | 0.4098 |
| 21 | 0.3853 | 0.3796 | 0.3726 |
| 22 | 0.3450 | 0.3453 | 0.3373 |
| 23 | 0.3054 | 0.3129 | 0.3040 |
| 24 | 0.2667 | 0.2825 | 0.2730 |
| 25 | 0.2297 | 0.2541 | 0.2441 |

Since the fitted model is a parametric model, not only can it be used for fitting but also for extrapolation. In Fig. 1, the curve of the fitted model as well as its baseline model are displayed. The curves between 0 and 25 represent the fitting results and the curves between 25 and 60 are the extrapolation based on the fitted models. Based on Fig. 1, we remark that the calibration is not perfect but satisfying enough for illustrative purposes. Due to the non-uniqueness property of phase-type distributions, we can always increase the dimension of the Markov process to improve the accuracy of the model.

The rest of the section examines the impact of the parameter $v$ on the survival index $S(t)$. For this purpose, we choose three values 0.5, 1 and 2 for $v$ and compute the survival function or the term structure $P(0,t), t \geq 0$, and the variance function $Var[S(t)], t \geq 0$, using formulas (16) and (22). The results are shown in Figs. 2 and 3, respectively.

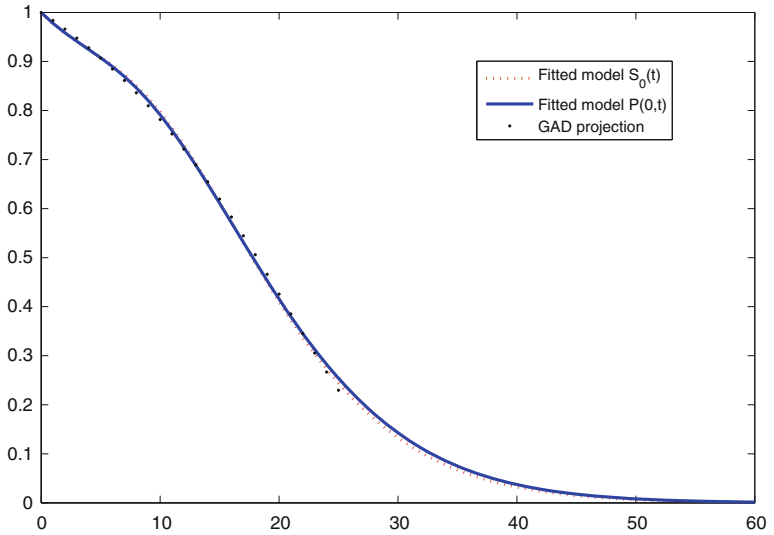



**Fig. 1** GAD projected survival rates, fitted survival function $P(0,t), 0 \leq t \leq 60$, and baseline survival function $S_0(t), 0 \leq t \leq 60$

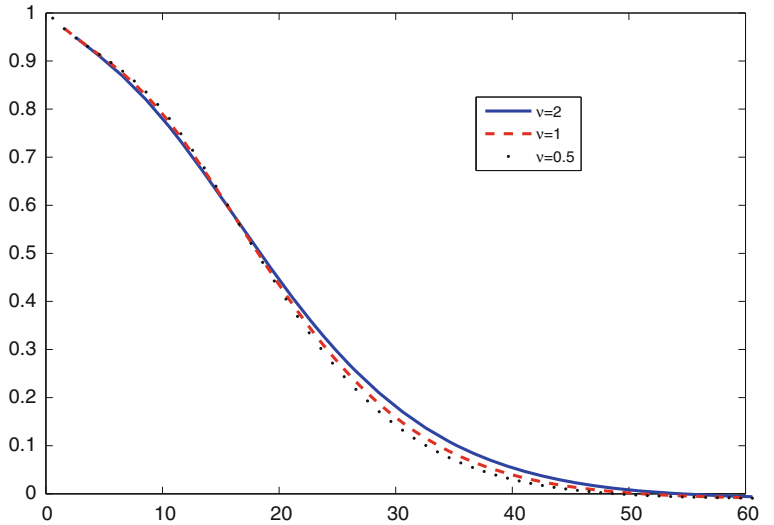



**Fig. 2** Term structure $P(0,t), t \geq 0$, for $\nu = 0.5$, 1 and 2

In Fig. 2, the curve of the term structure $P(0,t), t \geq 0$, exhibits a twisted upward shift as the value of $\nu$ increases. Moreover, in Fig. 3, the variance function $Var[S(t)], t \geq 0$, increases as $\nu$ gets larger. Thus, the parameter $\nu$ may be interpreted as the level of longevity risk and for this reason we call $\nu$ the longevity parameter of the model from now on. Combining the longevity parameter $\nu$ and the phase-type representation $(\boldsymbol{\alpha}, \boldsymbol{\Lambda})$ of the baseline aging process, $P(0,t)$ provides a flexible

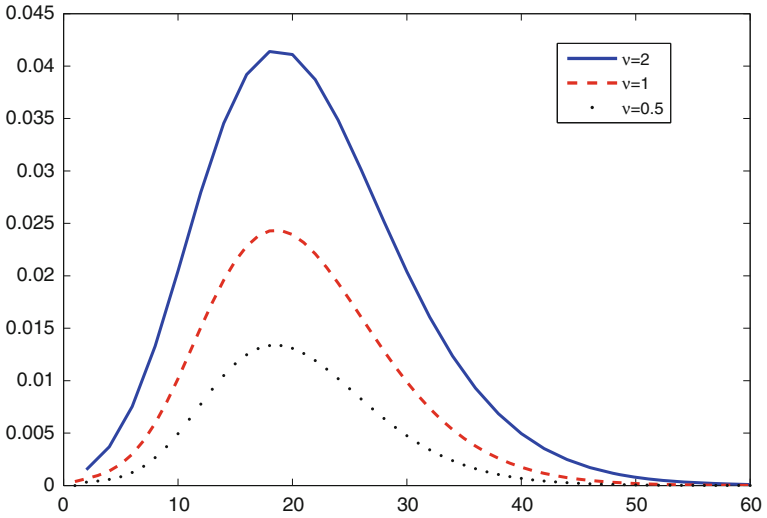



**Fig. 3** Variance function $Var[S(t)], t \geq 0$, for $v = 0.5,\ 1$ and 2

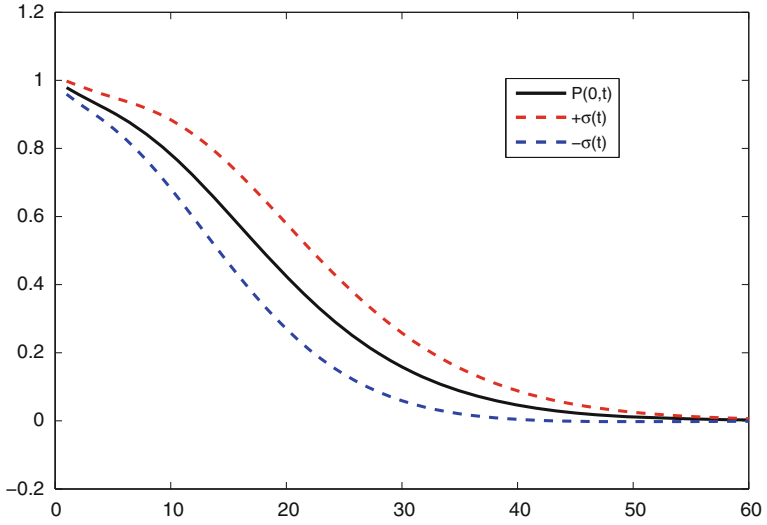



**Fig. 4** Term structure $P(0, t)$ with one-$\sigma$ confidence intervals, based on $v = 1$

pattern that is similar to the observed and predicted survival curves and should be able to calibrate to any term structure of mortality with high accuracy.

Another interesting observation from Fig. 3 is that the location of the mode of the variance function is relatively independent of the value of $v$ and the variance function peaks at around 20 years or so for UK males aged 65. This phenomenon can be interpreted as the result of two offsetting random effects. On the one hand, longer term survival index is generally subject to more uncertainties. On the other hand, the survival index itself is a decreasing function of time $t$, and eventually

diminishes to 0. Therefore, as $t$ increases, the variance function $Var[S(t)]$ presents a first increasing and then decreasing pattern. Similar behavior from the survival index has been found in [20] based on the Cairns–Dowd–Blake two-factor mortality model [12]. In Fig. 4 we draw the one-$\sigma$ confidence band for the survival rates for $v = 1$, based on the variance function in Fig. 3. It is similar to the fan charts for the survival function in [7]. This may be useful for mortality projection evaluation purposes.

## 5 Valuation of longevity bonds

Exposure to longevity risk has been a serious issue to annuity providers and pension plans due to the past and ongoing uncertainty surrounding mortality improvement. A recent innovation for managing the longevity risk exposure is to issue mortality-linked securities that transfer longevity risk from an insurer to capital markets (see [8], for an overview). As discussed in [6, 11, 19] and others, mortality-linked securities should be valued using an equivalent risk-neutral probability measure, say $Q$, assuming that there is a mortality-linked securities market. Further, [11] pointed out that the risk-neutral measure $Q$ might not be uniquely determined due to market incompleteness but the explicit use of a risk-neutral measure for fair valuation ensures that the market is free from riskless arbitrage. It is also true that in the current situation when the market is highly illiquid or has high transaction costs, the choice of $Q$ is in fact part of the modeling process.

In this section, we propose a measure change mechanism from the physical measure to a risk-neutral measure such that the stochastic mortality model under the risk-neutral measure remains to be a subordinated Markov model and discounted pure endowment value processes under the risk-neutral measure are martingales. We derive the term structure of mortality under the risk-neutral measure and present a general valuation framework for mortality derivatives. We then use longevity bonds (LBs) to illustrate how our proposed approach works.

### 5.1 A valuation framework

We begin with the subordinated Markov model described in Sect. 3. Let $(\boldsymbol{\alpha}, \boldsymbol{\Lambda})$ be the phase-type representation of the model and $v_0$ be its longevity parameter, both of which are estimated using projected survival rates as in Sect. 4 and hence the model is under the physical probability measure. As mentioned above, mortality-linked securities should be valued using an equivalent risk-neutral probability measure $Q$. Borrowing an idea from [25] on pricing defaultable bonds, we assume that under the risk-neutral measure $Q$, our mortality model remains to be a subordinated Markov model with the intensity matrix of the form

$$\boldsymbol{\Lambda}_{\boldsymbol{u}} = \boldsymbol{u}\boldsymbol{\Lambda}, \tag{29}$$

where the matrix $\boldsymbol{u}$ is a diagonal matrix with positive diagonal entries: $\boldsymbol{u} = diag(u_1, \ldots, u_n)$, with $u_i > 0$, $i = 1, \ldots, n$, which represents risk loadings with respect to the health statuses or frailty. Further, we let $v_1 \geq 0$ be a longevity risk loading such that the longevity parameter of the model is $v = v_0 + v_1$. As a result,

the term structure of mortality under the risk-neutral measure has the phase-type representation $(\boldsymbol{\alpha}, \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{u}})$, where $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{u}}$ is calculated using Theorem 3.1. We remark that the risk loading matrix $\boldsymbol{u}$ can have a more general form under an equivalent measure change (see [25, page 488] for details).

The survival index under the measure $Q$ is then

$$S_Q(t) = \boldsymbol{\alpha}\, e^{\boldsymbol{\Lambda}_{\boldsymbol{u}}\gamma_t}\boldsymbol{e}. \tag{30}$$

It can be shown that $S_Q(t)$ is a strictly decreasing function in $\boldsymbol{u}$, i.e. for $\boldsymbol{u}^* \neq \mathbf{u}$ and $u_i^* \leq u_i, i = 1, \ldots, n$,

$$S_Q^*(t) > S_Q(t), \quad t > 0. \tag{31}$$

The proof is lengthy and omitted. Thus, for the purpose of pricing longevity bonds and options, constraints $0 < u_i < 1, i = 1, \ldots, n$, are imposed. If one is to price mortality-linked insurances, then constraints $u_i > 1, i = 1, \ldots, n$, are imposed.

Consider now a pure endowment insurance or longevity insurance[2] of one monetary unit payable at time $t$. Let $E(s,t), s \leq t$, be the value of the pure endowment insurance at time $s$. Assuming independence between the term structure of mortality and the term structure of interest rates under the risk-neutral measure $Q$ as in [11], we have

$$\begin{aligned}\mathbf{I}_{\{\tau \geq s\}} E(s,t) &= E_Q[e^{-\int_s^t r(u)du}\mathbf{I}_{\{\tau \geq t\}}|\mathcal{F}_s] \\ &= D(s,t)E_Q[\mathbf{I}_{\{\tau \geq t\}}|\mathcal{F}_s] \\ &= \mathbf{I}_{\{\tau \geq s\}} D(s,t)\, P_Q(s,t)\end{aligned}$$

or

$$E(s,t) = D(s,t)\, P_Q(s,t), \tag{32}$$

where $\tau$ is the time of death or the absorption time of our subordinated Markov aging process, $D(s,t)$ is the term structure of zero-coupon bonds and $P_Q(s,t)$ is the term structure of mortality under the measure $Q$ given by

$$P_Q(s,t) = \boldsymbol{\alpha}_{\gamma_s} e^{\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{u}}(t-s)}\boldsymbol{e}. \tag{33}$$

Thus, the 'price' dynamics of a pure endowment insurance are identifiable under this model.

The calibration of the model under the risk-neutral measure is fairly straightforward. Suppose that the market values of $\hat{E}(0,t_i), i = 1, \ldots, n+1$, are available. We may numerically solve the following $n+1$ equations:

$$\hat{E}(0,t_i) = D(0,t_i)P_Q(0,t_i) = D(0,t_i)\boldsymbol{\alpha} e^{\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{u}t_i}}\boldsymbol{e}, \quad i = 1, \ldots, n+1, \tag{34}$$

for the loadings $u_i, i = 1, \ldots, n$, and $v_1$. In general, the model term structure of mortality may be calibrated to market by using approximation or the least square method as

[2] We use the same terminology as in [34, 35]. However, in [11], it is called the zero-coupon longevity bond or $(t,x)$-bond.

$$\min \sum_{i=1}^{k} \left[ \frac{\hat{E}(0,t_i)}{D(0,t_i)} - P_Q(0,t_i) \right]^2, \tag{35}$$

where $k$ is the number of longevity bonds available in the insurance market. We remark that we may also impose certain constraints on the risk loadings to simplify the calibration.

We next present a valuation formula for caplets written on the survival index. The value of a floorlet can be obtained using the put-call parity. Recall (19) and define the survival index at time $t$ under the measure $Q$ is

$$S(t) = S(s)\boldsymbol{\alpha}_{\gamma_s} e^{\boldsymbol{\Lambda_u}(\gamma_t - \gamma_s)}\boldsymbol{e}, \quad 0 \le s \le t, \tag{36}$$

where $\gamma_s$ satisfies the equation

$$S(s) = \boldsymbol{\alpha}\, e^{\boldsymbol{\Lambda}_u \gamma_s}\boldsymbol{e}.$$

Let $C(t) = [S(t) - K]_+$ be the payoff of a caplet on the survival index, where $K$ is a cap rate at exercise time $t$. Then the time-$s$ value of the option is

$$c(s,t,K) = D(s,t)E_Q\{C(t)|\mathcal{F}_s\}.$$

In the following we derive an explicit Black–Scholes type formula for $E_Q\{C(t)|\mathcal{F}_s\}$.

We have

$$\begin{aligned}
E_Q\{C(t)|\mathcal{F}_s\} &= E_Q\{[S(s)\boldsymbol{\alpha}_{\gamma_s} e^{\boldsymbol{\Lambda_u}(\gamma_t-\gamma_s)}\,\boldsymbol{e} - K]_+|\mathcal{F}_s\} \\
&= \int_0^\infty [S(s)\boldsymbol{\alpha}_{\gamma_s} e^{\boldsymbol{\Lambda_u} x}\boldsymbol{e} - K]_+ g(x)\,dx \\
&= S(s)\boldsymbol{\alpha}_{\gamma_s}\left[\int_0^{x_s} e^{\boldsymbol{\Lambda_u} x} g(x)\,dx\right]\boldsymbol{e} - K\int_0^{x_s} g(x)\,dx \\
&= S(s)\boldsymbol{\alpha}_{\gamma_s}\left[\int_0^{x_s} \sum_{i=1}^{n} e^{-u_i\lambda_i x} g(x)\,dx\, \boldsymbol{h}_i^{\boldsymbol{u}}\, \boldsymbol{v}_i^{\boldsymbol{u}}\right]\boldsymbol{e} - K\,\Gamma(x_s/v;(t-s)/v) \\
&= S(s)\boldsymbol{\alpha}_{\gamma_s}\left[\sum_{i=1}^{n}\int_0^{x_s} e^{-u_i\lambda_i x} g(x)\,dx\, \boldsymbol{h}_i^{\boldsymbol{u}}\, \boldsymbol{v}_i^{\boldsymbol{u}}\right]\boldsymbol{e} - K\,\Gamma(x_s/v;(t-s)/v) \\
&= S(s)\boldsymbol{\alpha}_{\gamma_s}\left[\sum_{i=1}^{n}(1+vu_i\lambda_i)^{-(t-s)/v}\Gamma((1+vu_i\lambda_i)x_s/v;(t-s)/v)\boldsymbol{h}_i^{\boldsymbol{u}}\, \boldsymbol{v}_i^{\boldsymbol{u}}\right]\boldsymbol{e} \\
&\quad - K\,\Gamma(x_s/v;(t-s)/v),
\end{aligned}$$

where $g(x)$ is a gamma density with mean $t-s$ and variance $v(t-s)$, the value $x_s$ satisfies

$$S(s)\boldsymbol{\alpha}_{\gamma_s} e^{\boldsymbol{\Lambda_u} x_s}\boldsymbol{e} = K,$$

$\Gamma(x;\alpha) = \frac{1}{\Gamma(\alpha)}\int_0^x y^{\alpha-1}e^{-y}\,dy$ is the incomplete gamma function, $\mathbf{H_u} = (\boldsymbol{h}_{\mathbf{1}}^{mathbfu}, \ldots, \boldsymbol{h}_{\mathbf{n}}^{\boldsymbol{u}})$ is the right eigenvector matrix of $\boldsymbol{\Lambda_u}$, and

$$\boldsymbol{H}_{\mathbf{u}}^{-1} = \begin{pmatrix} \boldsymbol{v}_{\mathbf{1}}^{\boldsymbol{u}} \\ \vdots \\ \boldsymbol{v}_{\mathbf{n}}^{\boldsymbol{u}} \end{pmatrix}.$$

Thus,

$$c(s,t,K) = D(s,t)[S(s)\boldsymbol{\alpha}_{\gamma_s} e^{\boldsymbol{\Lambda}_{\boldsymbol{u}}(s,t)}\boldsymbol{e} - K\,\Gamma(x_s/v;(t-s)/v)], \tag{37}$$

where

$$\boldsymbol{\Lambda}_{\boldsymbol{u}}(s,t) = \boldsymbol{H}_{\boldsymbol{u}}\boldsymbol{\Omega}(s,t)\boldsymbol{H}_{\boldsymbol{u}}^{-1}$$

and

$$\boldsymbol{\Omega}(s,t) = diag(-\omega_1(s,t),\ldots,-\omega_n(s,t))$$

with

$$\omega_i(s,t) = -\ln[(1+vu_i\lambda_i)^{-(t-s)/v}\Gamma((1+vu_i\lambda_i)x_s/v;(t-s)/v)].$$

In particular, when $s = 0$,

$$\boldsymbol{\alpha}e^{\boldsymbol{\Lambda}_{\boldsymbol{u}}x_0}\boldsymbol{e} = K,$$

and

$$c(0,t,K) = D(0,t)[\boldsymbol{\alpha}e^{\boldsymbol{\Lambda}_{\boldsymbol{u}}(0,t)}\boldsymbol{e} - K\,\Gamma(x_0/v;t/v)]. \tag{38}$$

We remark that to obtain (37), we utilize a decomposition technique on the intensity matrix $\boldsymbol{\Lambda}_{\boldsymbol{u}}$ similar to that in Theorem 3.1.

The option pricing formula (37) is analytical and hence allows us to compute quickly and efficiently the value of longevity caps, which we will demonstrate in the next subsection, and in turn the value of many annuity options or guarantees. Furthermore, option price (37) increases as $\boldsymbol{u}$ decreases due to (31) and it has upper limit $S(s) - K$ as $\boldsymbol{u}$ tends to $\boldsymbol{0}$. This relation clearly indicates the role of $\boldsymbol{u}$ as risk loadings on the health statuses and gives a one-to-one correspondence between the risk premium of the option and the loading to a health status. The monotonic property also ensures that the formula can be calibrated to the market price of the option.

### 5.2 EIB longevity bonds

In this section, we use an EIB longevity bond as an example to illustrate the implementation of the valuation framework proposed in Sect. 5.1. An EIB longevity bond is a financial contract in which annual coupon payments are proportionally linked to the realization of the survival index on a reference population over the next 25 years. The reference population for the EIB longevity bond is the English and Welsh males aged 65 in 2003. As described in [12], the price of the bond at issuance is the present value of the future cash flow based on the GAD projections, at discount rates that are 20 basis points above the

**Table 2** Market implied survival rates $\hat{P}_Q(0,t)$ and calibrated model survival rates $P_Q(0,t)$

| t | $\hat{P}_Q(0,t)$ | $P_Q(0,t)$ | t | $\hat{P}_Q(0,t)$ | $P_Q(0,t)$ |
|---|---|---|---|---|---|
| 1 | 0.9837 | 0.9853 | 14 | 0.6675 | 0.6663 |
| 2 | 0.9662 | 0.9696 | 15 | 0.635 | 0.6299 |
| 3 | 0.9477 | 0.9533 | 16 | 0.6015 | 0.593 |
| 4 | 0.9281 | 0.9361 | 17 | 0.5672 | 0.5558 |
| 5 | 0.9074 | 0.9179 | 18 | 0.5321 | 0.5189 |
| 6 | 0.8856 | 0.8982 | 19 | 0.4965 | 0.4825 |
| 7 | 0.8626 | 0.8766 | 20 | 0.4606 | 0.4469 |
| 8 | 0.8384 | 0.8529 | 21 | 0.4245 | 0.4125 |
| 9 | 0.8129 | 0.8269 | 22 | 0.3885 | 0.3793 |
| 10 | 0.7862 | 0.7682 | 23 | 0.353 | 0.3476 |
| 11 | 0.7583 | 0.7358 | 24 | 0.318 | 0.3174 |
| 12 | 0.7292 | 0.7358 | 25 | 0.2841 | 0.2890 |
| 13 | 0.6989 | 0.7017 | | | |

corresponding yield rates of the conventional fixed-interest EIB bonds. Suppose that the notional amount is one monetary unit (one pound). The cash flow of the longevity bond is then equal to the term structure $P(0,t)$ under the physical probability measure which is displayed in column 1 of Table 1. As shown in [12], the price of the bond is 11.442 and a two-factor time series model (the CBD model) was calibrated to the price with several choices of the two market prices of risk embedded in the model. In the same spirit of [12], we impose the following constraint on the risk loading matrix $\boldsymbol{u}$:

$$u_1 = u_2 = \cdots = u_n = u. \tag{39}$$

As a result, our model is now a two-factor model with $u$ being the frailty risk loading and $v_1$ being the longevity risk loading. We could calibrate the model to the longevity bond price directly but to better illustrate the calibration, we use the term structure of mortality given in Column 4, Table 1 of [12] (reproduced in the first column of Table 2) as the market implied term structure of mortality.

Solving for the optimal values of parameters $(u, v_1)$ in the minimization problem

$$\min \sum_{t=0}^{25} (P_Q(0,t) - \hat{P}_Q(0,t))^2, \tag{40}$$

we have

$$u = 0.9483659, \quad v_1 = 0.8887757. \tag{41}$$

where $P_Q(0,t)$ has the form (33) with $s = 0$, The calibrated model survival rates are given in column 2 of Table 2.

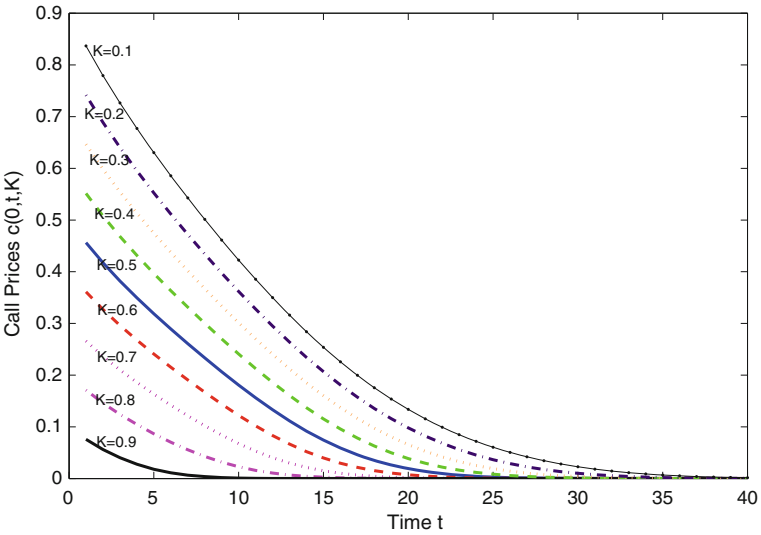



**Fig. 5** Caplet values based on formula (38), using $D(0,t) = e^{-0.05t}$

**Table 3** Caplet values with different $t$ and $K$

| $t$ | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| $c(0,t,0.5)$ | 0.3188 | 0.1808 | 0.0743 | 0.0191 | 0.0025 |
| $K$ | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| $c(0,20,K)$ | 0.0655 | 0.0387 | 0.0191 | 0.0071 | 0.0016 |

As we mentioned earlier, once the model is calibrated the value of a caplet can be computed easily by using the option pricing formula (38). With the calibrated parameters in (41) and assuming a force of interest 5 %, we compute the value of a wide range of caplets. They are plotted in Fig. 5. Selected values are given in Table 3.

We further compute the value of the caplets when their capped survival rates are the projected survival rates. The numerical results are given in Fig. 6. They show that the value of the caplets starts from a relative low level because the uncertainty in the near future is low. Their values increase quickly and present a plateau after 10 years. For comparison, we use $D(0,t) = e^{-\delta t}$ with two values of $\delta$: 0.05 or 0.0392 (which is used by Cairns et al. [12], in their illustration). The higher value of $\delta$ implies a larger discount effect from the term structure of interest rates. With $\delta = 0.0392$, the highest caplet value occurs at t = 18, 19, and 20 with $c(0,t,K) = 0.0408$. The total cost for setting a portfolio of caplets or a cap with the projected survival rates as strikes is 0.7546, which is obtained from adding up all the caplets' values in the solid line of Fig. 6 corresponding to $\delta = 0.0392$ and the nominal amount of 1 for next 25 years.

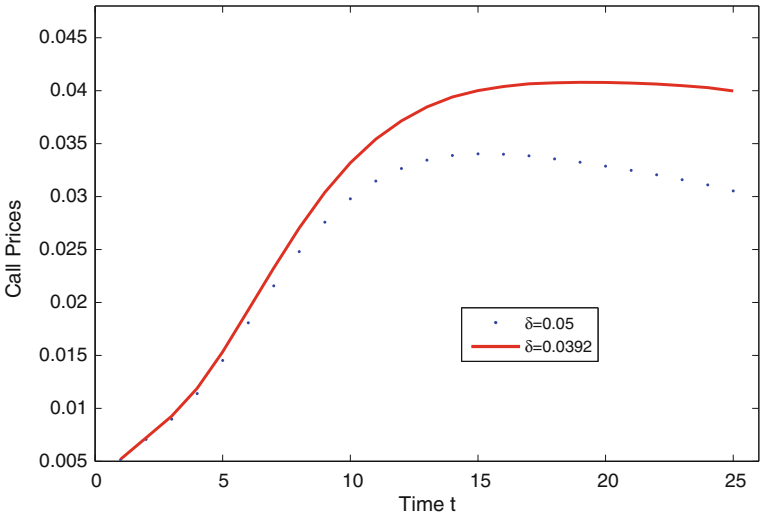



**Fig. 6** Caplet values based on formula (38) with predicted survival rates in column 1 of Table 1 as strike prices

## 6 Concluding remarks

In this paper, we have proposed a phase-type stochastic mortality model. The stochasticity of mortality is governed by a subordinating stochastic process. We focus on the theoretical development of the model and have shown that the model exhibits many desirable properties of a mortality model and meets many model selection criteria laid out in [11, 13]. Under our proposed model, the stochastic force of mortality is always positive. The underlying Markov process of the model provides certain reasonable biological interpretation of the aging process (see [30]). The model is flexible and fits either historical mortality data or projected mortality data well. Further, with a proposed risk loading mechanism, we can make an easy transition from the physical measure to a risk-neutral measure and hence is able to calibrate the model to market information.

The phase-type structure of the model allows us to apply the matrix-analytic methods developed for queuing theory to the model (see [2, 36]). As a result, many quantities of interest such as the distribution of future survival rates, prediction intervals, the term structure of mortality as well as the value of caps and floors can be obtained analytically. However, there are several issues that require further exploration and investigation. We intend to develop more effective fitting algorithms. One possibility is to re-examine the use of the EM algorithm by utilizing the special structure of the model, similar to that in [28, 29]. Incorporating a cohort effect into the model is another interesting but challenging problem. It seems that a time-dependent or non-homogeneous subordinated Markov model might be suitable for this purpose but the mathematical tractability will suffer if such a model is considered. Another approach might be the use of Bayesian modeling techniques as did in the Cox proportional hazard model.

**Acknowledgments** This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). We would like to thank two reviewers, Andrei Badescu and Soohan Ahn for their insightful comments.

## References

1. Aalen OO (1995) Phase type distributions in survival analysis. Scand J Stat 22:447–463
2. Asmussen S (1987) Applied probability and queues. Wiley, New York
3. Asmussen S, Albrecher H (2010) Ruin probabilities, 2nd edn. World Scientific Publishing, Singapore
4. Ballotta L, Haberman S (2006) The fair valuation problem of guaranteed annuity options: the stochastic mortality environment case. Insur Math Econ 38:195–214
5. Biffis E (2005) Affine processes for dynamic mortality and actuarial valuation. Insur Math Econ 37:443–468
6. Biffis E, Denuit M (2006) Lee-Carter goes risk-neutral: an application to the Italian annuity market. Giornale dell'Istituto Italiano degli Attuari LXIX:1–21
7. Blake D, Dowd K, Cairns AJG (2008a) Longevity risk and the grim reaper's toxic tail: the survivor fan charts. Insur Math Econ 42:1062–1066
8. Blake DP, Cairns AJ, Dowd K (2008b) The birth of the life market. Asia Pa J Risk Insur 3:6–36
9. Boyle P, Hardy M (2003) Guaranteed annuity options. ASTIN Bull 33:125–152
10. Brouhns N, Denuit M, Vermunt J (2002) A Poisson log-bilinear regression approach to the construction of projected lifetables. Insur Math Econ 31:373–393
11. Cairns AJG, Blake D, Dowd K (2006) Pricing death: frameworks for the valuation and securitization of mortality risk. ASTIN Bull 36:79–120
12. Cairns AJG, Blake D, Dowd K (2006) A two-factor model for stochastic mortality with parameter uncertainty. J Risk Insur 1:687–718
13. Cairns AJG, Blake D, Dowd K (2008) Modelling and management of mortality risk: a review. Scand Actuar J 2:79–113
14. CMI (2004) Projecting future mortality: a discussion paper. Continuous mortality investigation working paper 3
15. CMI (2005) Projecting future mortality: towards a proposal for a stochastic methodology. Continuous mortality investigation working paper 15
16. CMI (2007) Stochastic projection methodologies: Lee-Carter model features, example results and implications. Continuous mortality investigation working paper 15
17. Dahl M (2004) Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. Insur Math Econ 35:113–136
18. Dahl M, Melchior M, Moller T (2008) On systematic mortality risk and risk minimization with survivor swaps. Scandinavian Actuarial Journal 2:114–146
19. Dahl M, Møller T (2006) Valuation and hedging of life insurance liabilities with systematic mortality risk. Insur Math Econ 39:193–217
20. Dowd K, Cairns AJG, Blake D (2006) Mortality-dependent financial risk measures. Insur Math Econ 38:427–440
21. Foo J, Kevin Leder K, Michor F (2011) Stochastic dynamics of cancer initiation. Phys Biol. doi: 10.1088/1478-3975/8/1/015002
22. GAD (2001) National population projections: review of methodology for projecting mortality. National statistics quality review series, report no. 8. Government Actuary's Department
23. Hautphenne S, Latouche G (2011) The Markovian binary tree applied to demography. J Math Biol. doi:10.1007/s00285-011-0437-1
24. Hurd T, Kuznetsov A (2007) Affine Markov chain model of multifirm credit migration. J Credit Risk 3:3–29
25. Jarrow RA, Lando D, Turnbull SM (1997) A Markov model for the term structure of credit risk spreads. Rev Financial Stud 10:481–523
26. Kay R (1986) A Markov model for analysing cancer markers and disease states in survival studies. Biometrics 42:855–865
27. Lee RD, Carter LR (1992) Modeling and forecasting U.S. mortality. J Am Stat Assoc 87:659–675
28. Lee SCK, Lin XS (2011) Modeling dependent risks with multivariate Erlang mixtures. ASTIN Bulletin, to appear

29. Lee SCK, Lin XS (2010) Modeling and evaluating insurance losses via mixtures of Erlang distributions. N Am Actuar J 14(1):107–130
30. Lin XS, Liu X (2007) Markov aging process and phase-type law of mortality. N Am Actuar J 11:92–109
31. Longini IM, Clark WS, Gardner L, Brundage JF (1991) The dynamics of CD4+ T-lymphocyte decline in HIV-infected individuals: a Markov modelling approach. J Acquir Immune Defic Syndr 4:1141–1147
32. Madan DB, Carr P, Chang EC (1998) The Variance Gamma process and option pricing. Eur Finance Rev 2:79–105
33. Madan DB, Milne F (1991) Option pricing with V.G. martingale components. Math Finance 1:39–55
34. Milevsky MA, Promislow SD (2001) Mortality derivatives and the option to annuitise. Insur Math Econ 29:299–318
35. Milevsky MA, Promislow SD, Young J (2006) Killing the law of large numbers: mortality risk premiums and the Sharpe ratio. J Risk Insur 73:673–686
36. Neuts MF (1981) Matrix-geometrix solutions in stochastic models. Johns Hopkins University Press, Baltimore
37. Pitacco E (2004) Survival models in a dynamic context: a survey. Insur Math Econ 35:279–298
38. Renshaw A, Haberman S (2003) Lee-Carter mortality forecasting with age-specific enhancement. Insur Math Econ 33:255–272
39. Renshaw A, Haberman S (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insur Math Econ 38:556–570
40. Tuljapurkar S, Boe C (1998) Mortality changes and forecasting: how much and how little do we know? N Am Actuar J 2:13–47
41. Willets R (1999) Mortality in the next millennium. Presented at the Staple Inn Actuarial Society on 7 December