Asmussen, Søren; Laub, Patrick J.; Yang, Hailiang

**Article**

# Phase-type models in life insurance: Fitting and valuation of equity-linked benefits

Risks

**Provided in Cooperation with:**
MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Asmussen, Søren; Laub, Patrick J.; Yang, Hailiang (2019) : Phase-type models in life insurance: Fitting and valuation of equity-linked benefits, Risks, ISSN 2227-9091, MDPI, Basel, Vol. 7, Iss. 1, pp. 1-22, https://doi.org/10.3390/risks7010017

risks

MDPI

Article

# Phase-Type Models in Life Insurance: Fitting and Valuation of Equity-Linked Benefits

**Søren Asmussen [1,*], Patrick J. Laub [2] and Hailiang Yang [3]**

1 Department of Mathematics, Aarhus University, 8000 Aarhus, Denmark
2 Institut de Science Financière et d'Assurances, Université Lyon 1, 69007 Lyon, France; patrick.laub@univ-lyon1.fr
3 Department of Statistics & Actuarial Science, Hong Kong University, Hong Kong 999077, China; hlyang@hku.hk
* Correspondence: asmus@math.au.dk

Received: 11 December 2018; Accepted: 4 February 2019; Published: 11 February 2019


**Abstract:** Phase-type (PH) distributions are defined as distributions of lifetimes of finite continuous-time Markov processes. Their traditional applications are in queueing, insurance risk, and reliability, but more recently, also in finance and, though to a lesser extent, to life and health insurance. The advantage is that PH distributions form a dense class and that problems having explicit solutions for exponential distributions typically become computationally tractable under PH assumptions. In the first part of this paper, fitting of PH distributions to human lifetimes is considered. The class of generalized Coxian distributions is given special attention. In part, some new software is developed. In the second part, pricing of life insurance products such as guaranteed minimum death benefit and high-water benefit is treated for the case where the lifetime distribution is approximated by a PH distribution and the underlying asset price process is described by a jump diffusion with PH jumps. The expressions are typically explicit in terms of matrix-exponentials involving two matrices closely related to the Wiener-Hopf factorization, for which recently, a Lévy process version has been developed for a PH horizon. The computational power of the method of the approach is illustrated via a number of numerical examples.

**Keywords:** generalized Coxian distribution; EM algorithm, guaranteed minimum death benefit; high-water benefit; jump diffusion, matrix-analytic methods; reserves; Wiener-Hopf factorization

## 1. Introduction

A random variable (r.v.) $\tau$ is phase-type (PH) with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ if it is distributed as the lifetime $\inf\{t \geq 0 : J_t = \dagger\}$ of a killed (terminating) time-homogenous Markov process $J = \{J_t\}_{t \geq 0}$ with $p < \infty$ states, initial distribution $\boldsymbol{\alpha}$ (a row vector), and generator $\boldsymbol{T}$, where $\dagger$ is an additional absorbing state. PH distributions originate from queueing theory and the work of A.K. Erlang, A. Jensen and M.F. Neuts. However, in recent decades, they have found numerous applications to different areas. The main reason for this is that many calculations that are explicit for exponential distributions are often computationally tractable with PH assumptions and that, in addition, PH distributions are dense so that a given distribution $F$ on $[0, \infty)$ can be approximated arbitrarily well by a PH distribution provided one takes $p$ large enough. For surveys of the theory and applications, see (Asmussen 2003; Asmussen and Albrecher 2010; Bladt and Nielsen 2017). The present paper is concerned with some applications to life insurance, an area where PH distributions have been far less employed than in the related areas of non-life insurance. There are many papers in ruin theory that assume that claim sizes are PH distributed; see (Cai and Li 2005b) and the review paper (Bladt 2005). Some papers assume that the claim time is PH distributed; see (Stanford et al. 2000).

(Zadeh et al. 2014) considered disability insurance using a PH model. (Zadeh and Stanford 2016) studied credibility theory using PH distributions. In finance, PH distributions have also been used extensively. For example, (Asmussen et al. 2004) considered option pricing under an exponential phase-type Lévy model; (Cai and Li 2005a) dealt with applications of PH distributions to risk measures; and (Egami and Yamazaki 2014) considered PH approximations of Lévy models. On the contrary, the only references of applications of PH distributions to life insurance we know of are (Lin and Liu 2007) and (Zadeh et al. 2014).

The motivation for the present study is some valuation problems of equity-linked products in life insurance considered in Gerber et al. (2012, 2013, 2015) (GSY). Let $\tau = \tau_x$ be the remaining lifetime of an insured of age $x$ when signing the contract, then the payment of this class of benefits has the form $\Psi(\overline{S}_\tau, S_\tau)$ where $S_t$ is the price at time $t$ of some stock or stock index and $\overline{S}_t = \max_{v\le t} S_t$ the running maximum. An example of such a benefit is the guaranteed minimum death benefit (GMDB) $\max(S_\tau, K)$, but there are many others; see the list in Section 4. Defining $X_t$ as $X_t = \log(S_t/S_0)$, $M_t = \overline{X}_t$ as the running maximum of $X$, and $D_t = M_t - X_t$ as the displacement from the maximum, we then have:

$$\Psi(\overline{S}_\tau, S_\tau) \;=\; \Phi(M_\tau, D_\tau) \qquad \text{where } \Phi(m,d) = \Psi(S_0 \mathrm{e}^{m}, S_0 \mathrm{e}^{m-d)}) \tag{1}$$

(note that $X_t = M_t - D_t$), and the fair value of the benefit is the expected discounted payment.

A widely-used model (Cont and Tankov 2004) assumes that $X = \{X_t\}_{t\ge 0}$ is a Lévy process, and the key vehicle in GSY for the computations for such expectations is the classical Wiener-Hopf factorization:

**Lemma 1.** *For a Lévy process $X_t$ and an independent exponential time $\tau$, the r.v.'s $M_\tau = \max_{0\le s\le\tau} X_s$ and $D_\tau = M_\tau - X_\tau$ are independent. Further,*

$$D_\tau \;\stackrel{\mathrm{d}}{=}\; -\underline{X}_\tau \;=\; -\min_{0\le s\le\tau} X_s\,.$$

For the geometric Brownian motion model where $X$ is a Brownian motion (BM) with drift, $M_\tau$ and $D_\tau$ both have an exponential distribution, and for more general jump diffusions, a number of computational procedures have been developed (see Section 5 for references). In GSY, they approximate the density/tail of $\tau_x$ by a linear combination of exponential terms, which allows them to reduce the computations to simple integrations using Lemma 1.

The approximation of the distribution of $\tau_x$ by a linear combination of exponential terms in GSY uses an ad hoc optimization procedure. The main problem there is that it is very difficult to make the linear combination of exponential densities to be a density function; in particular, very often, in the tail part, the linear combination of exponential densities can be negative. Since, in the application to valuing equity-linked insurance products, the distribution at the tail part plays an important role, this motivated us to consider using the PH distribution to approximate the future lifetime distribution. For PH distributions, more classical statistical methods using maximum likelihood have, however, been developed (Asmussen et al. 1996), and in view of the denseness of PH distributions and the fact that they have proven a computationally-tractable generalization of the exponential distribution in other settings, a natural idea is therefore to use a PH approximation $\tau$ to $\tau_x$. The first step in implementing this for pricing equity-linked benefits is to develop a version of the Wiener-Hopf factorization for the case where $\tau$ is PH and independent of $X$. This has been done in a companion paper (Asmussen and Ivanovs 2019), and the contribution of the present paper is to provide the further steps such as PH fitting of human mortality and computations of prices of equity-linked benefits for BM and jump diffusions.

## 2. Preliminaries

The vector of killing rates of a PH$(\boldsymbol{\alpha}, \boldsymbol{T})$ r.v. $\tau$ is given by $\boldsymbol{t} = -\boldsymbol{T}\mathbf{1}$, where $\mathbf{1}$ is a column vector with all entries equal to one. The density is $\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}z}\boldsymbol{t}$, and the tail $\mathbb{P}(\tau > z)$ is $\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}z}\mathbf{1}$. Here, the exponential $\mathrm{e}^{\boldsymbol{R}}$ of a matrix is defined as $\sum_0^\infty \boldsymbol{R}^n/n!$. We shall often use formulas like:

$$\int_a^b \mathrm{e}^{\boldsymbol{R}x}\,\mathrm{d}x \;=\; \boldsymbol{R}^{-1}(\mathrm{e}^{\boldsymbol{R}b} - \mathrm{e}^{\boldsymbol{R}a}) \;=\; (\mathrm{e}^{\boldsymbol{R}b} - \mathrm{e}^{\boldsymbol{R}a})\boldsymbol{R}^{-1}\,, \quad \int_0^\infty \mathrm{e}^{\boldsymbol{R}x}\,\mathrm{d}x \;=\; -\boldsymbol{R}^{-1}\,,$$

that are immediate analogues of standard formulas for univariate exponentials. For example, taking $\boldsymbol{R} = -s\boldsymbol{I} + \boldsymbol{T}$ gives the Laplace transform $\mathbb{E}[\mathrm{e}^{-s\tau}]$ as $\boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t}$. Of course, the existence of $\boldsymbol{R}^{-1}$ has to be verified, which for a PH generator $\boldsymbol{R} = \boldsymbol{T}$ is a standard fact.

The subclass of generalized Coxian (GC) distributions will play a particular role. Here, the states are run through in lexicographical order, with exponential $(\lambda_i)$ holding time in state $i$, and it is possible to enter or exit from any state (see Figure 1), where $q_i$ is the probability of exit after a visit to $i$ and $r_i = 1 - q_i$ the probability of going to the next state. The structure of Coxian distributions is similar, except that only State 1 can be entered so that $\alpha_1 = 1$. Coxian distributions are in particular used in (Lin and Liu 2007) and (Zadeh et al. 2014).

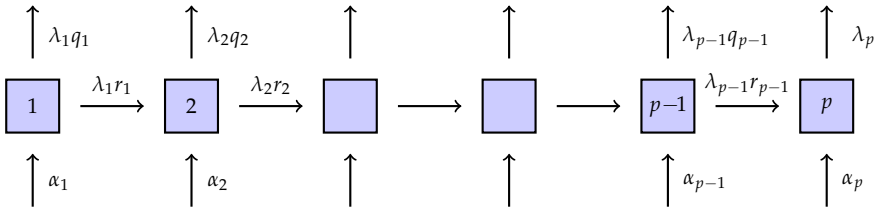



**Figure 1.** A GC distribution.

**Remark 1.** Though computational tractability and denseness are the main reasons for the popularity of PH distributions, other advantages have also sometimes been advocated. For example, whereas in the majority of examples, the PH modeling is purely descriptive, there are others where attempts have been made to give the Markov states a physical meaning like in compartment models and, in the life insurance context, in (Lin and Liu 2007) and (Zadeh et al. 2014). This makes the number $p$ of phases quite large there, of order 200, whereas we have taken the descriptive point of view. The fit obviously becomes better when increasing $p$, but how good it needs to be depends very much on what it will be used for. In this paper, we consider the valuation problem, that is to calculate the net premium (expected discounted payoff) for equity-linked insurance products, and our numerical examples will show that taking $p = 50$ will be sufficient for the precision one needs in this context.

## 3. PH Fits of Human Mortality Data

The distribution of human lifetimes of course varies somewhat according to the period where the data were collected, geographical and social conditions, etc., but the basic shapes are much alike. We have therefore concentrated on a single example, the illustrative life table in Appendix 2A of (Bowers et al. 1997). The form of the distribution, as illustrated via the number of deaths in consecutive years, is given in Figure 2.

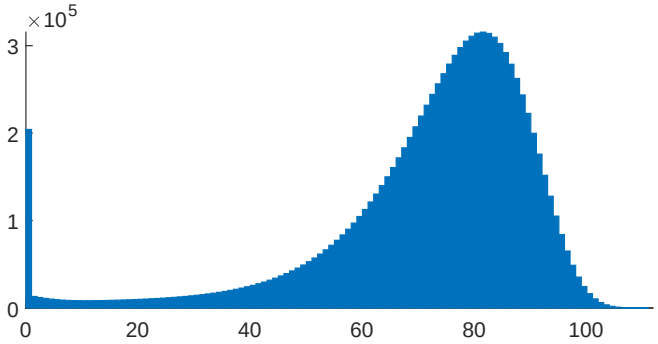



**Figure 2.** Life table data.

We use an EM algorithm to fit various types of phase-type distributions to this life table. This idea was originally developed in (Asmussen et al. 1996), with the case of censored data and other implementation details given in (Olsson 1996). Olsson (Olsson 1998) has written a C implementation of this algorithm called EMpht, which is available online. The overwhelming computational burden is in the E step, where many complicated integrals involving matrix exponentials must be evaluated. This is done approximately by converting the problem into a system of ODEs, which are approximately solved via a Runge-Kutta method. While the original C implementation is still very adequate for creating fits with a small number of phases $p$, it can be improved upon for larger $p$.

We wrote a new implementation using the Julia language. It allows for the E step to be evaluated in three different ways: (i) using an ODE solver (as in (Asmussen et al. 1996; Olsson 1996)), (ii) using quadrature methods, and (iii) using the uniformization method of Okamura et al. Okamura et al. (2011a, 2011b). We used this last method to generate the fits found later in this paper; the code we used to create the figures below will be made available online (Asmussen et al. 2019). Each fit was given one hour execution time restricted to one thread on the processor Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20 GHz. The EM algorithm could have been terminated much earlier without a major difference to the quality of the fits, but as we were comparing the various values of $p$, we wanted to ensure no individual fit was judged poorly for simply being stuck in a local maxima.

The PH representation in terms of $\boldsymbol{\alpha}, \boldsymbol{T}$ is heavily overparameterized. There is an abundance of sets $\boldsymbol{\alpha}, \boldsymbol{T}$ that leads to the same distribution. Consider, e.g., the PH with $\boldsymbol{\alpha} = \frac{1}{p}\mathbf{1}^\intercal$ (where $^\intercal$ stands for transposition) and where $\boldsymbol{T}$ is a diagonal matrix with $(-1, -2, \ldots, -p)$ on the diagonal (i.e., a hyperexponential distribution). If the diagonal elements in $\boldsymbol{T}$ are permuted in any of the $p!$ possibilities, then the distribution is unchanged. This example indicates that the maximum $\ell^*$ of the likelihood function $\ell$ is attained at several parameter values so that the choice of the initial parameter set will determine which one is found by the EM algorithm. The presence of local maxima with values $< \ell^*$ cannot be excluded, but we are not aware of serious cases where the EM algorithm has been stuck in one of these.

In life insurance calculations, the relevant lifetime is not the overall one $\tau_0$, but rather the remaining lifetime $\tau = \tau_x$ of the insured at the time when he/she signs the contract. That is, if the age of the insured at that time is $x$, the relevant distribution is that of $\tau_0 - x$ given $\tau_0 \geq x$. As a typical example, we took $x = 35$ and fitted PH distributions with various structure and various numbers $p$ of phases. Note that while we fit the distribution of $\tau_0 - x$, which starts at zero, in the following, we plot $\tau_0$ given $\tau_0 \geq x$ (i.e., the $x$ axes begin at $x = 35$ and not zero).

As a first illustration, we fit both a general structure PH and a generalized Coxian (GC) one with $p = 20$ phases. The results are in Figure 3, with the densities of the fitted PH distributions in the left panel together with the data and the corresponding hazard rates in the right panel.

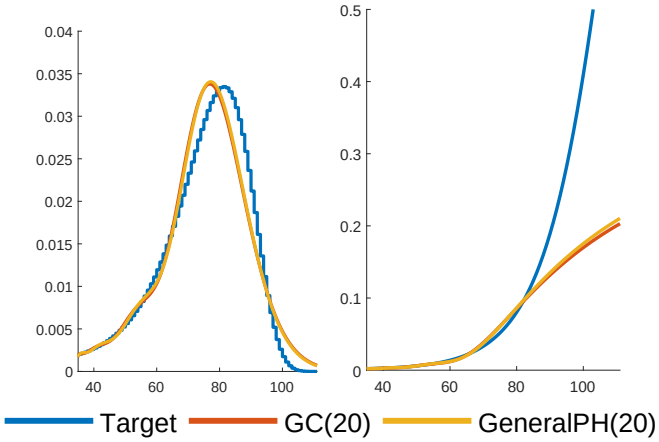



**Figure 3.** Probability density functions and hazard rate functions for the life table and phase-type (PH) fits using the general and the generalized Coxian (GC) structures with $p = 20$ phases.

The fits of the two structures are almost identical (but the shape of the density is somewhat off the data, which indicates that a larger valuer of $p$ could be appropriate). The general structure leads to a system of ODEs of dimension $q = p^2 + p$, whereas the corresponding number for GC is only $q = 2p - 1$. Due to this curse of dimensionality and the marginal differences of the fits, we have only considered GC structures for higher values of $p$.

Figure 4 shows GC fits for $p = 35$ and $p = 40$. As expected, $p = 40$ improves upon $p = 35$.

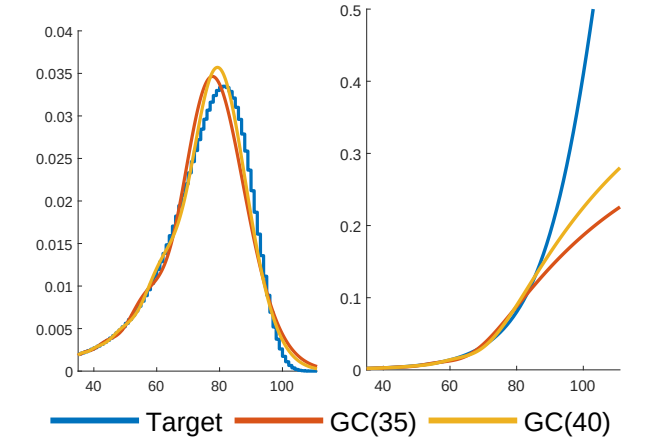



**Figure 4.** Probability density functions and hazard rate functions for the life table and phase-type fits using the generalized Coxian structures with $p = 35, 40$ phases.

In Figure 5, we complement Figure 4 by plots of the tails. Such plots are frequently used to illustrate the quality of fits, but in our minds, a good agreement is not an argument in itself: two tails may come out quite similar in a comparison even if the shapes of the densities are rather different. Finally, Figure 6 gives GC fits with $p = 50, 75, 100$.

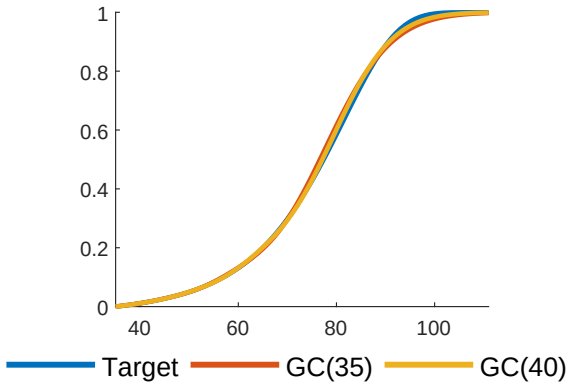



**Figure 5.** Tails of the life table and GC fits with $p = 35, 40$ phases.

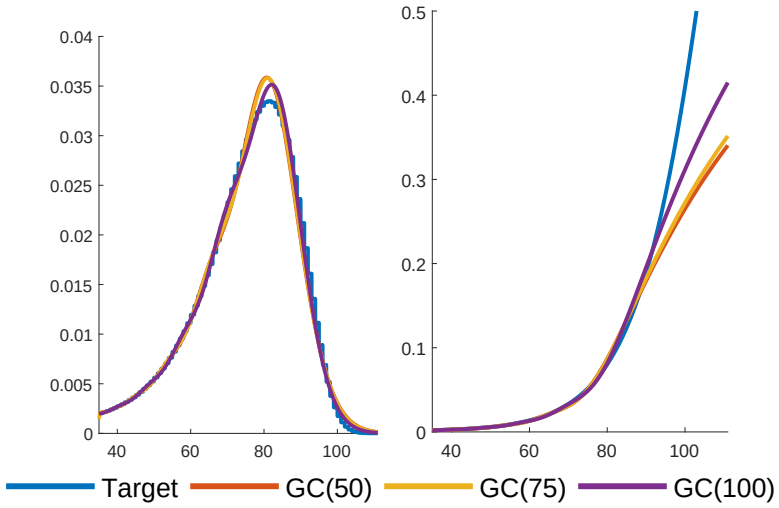



**Figure 6.** Probability density functions and hazard rate functions for the life table and GC fits with $p = 50, 75, 100$ phases.

Which value of $p$ should one then work with? There exists hardly a single criterion for answering such questions, i.e., assessing the quality of the fits to the data. Some concerns may be aesthetic, such as avoiding negative values of the density (cf. Appendix A below), support outside the obvious range of the data (cf. kernel estimates), or an excessive number of parameters (as is the case for most PH fits as compared to, e.g., the classical Gompertz-Makeham fit of life table data). Other concerns may have to do with the features of the distribution which are considered especially important, be it the overall shape of the density (caught well by maximum likelihood), the cumulative distribution function (c.d.f.), although c.d.f.'s may look close even for rather different distributions, the tail (where maximum likelihood fails and other tools such as the Hill estimator may be more appropriate), the hazard rate (as showing up in numerous life insurance contexts), or the behavior in a more narrow range (cf. a life insurance contract signed at the start of employment and running up to the time of retirement). For the present purposes and related ones like applications in queueing and insurance risk, the main point is, however, that PH fitting is not done for its own sake, but for the purpose of facilitating some calculation, here of the prices of equity-linked products, in queueing, say of waiting time characteristics, etc. In the rest of the paper, we therefore study the performance of the PH approach and the various fits and return to a discussion of the issue in the concluding Section 8.

## 4. Valuation of Benefits

The main application of the results in this paper is valuing equity-linked insurance products, in particular GMDBs in various deferred annuities. The time of payment of this kind of product is at

the time $\tau$ of death of the policyholder, often denoted $\tau_x$, where $x$ is his/her age at $t = 0$ when he/she signs the contract. The amount of the payment depends on the price $S_\tau$ of a stock (or stock index) at that time and possibly also on the history of the stock price as, e.g., in (1). There are many different products (benefit functions) in the market. One example is the GMDB with benefit function being:

$$\Psi(\overline{S}_\tau, S_\tau) \;=\; \max(S_\tau, K) = S_\tau + [K - S_\tau]_+, \tag{2}$$

where $K$ is the guaranteed amount, e.g., $K = 0.9S_0$. The second term on the right-hand side (r.h.s.) of (2) is the payoff of a life-contingent put option. Another popular product is the so-called high-water benefit (HWB) with the payoff function being:

$$\Psi(\overline{S}_\tau, S_\tau) \;=\; \overline{S}_\tau.$$

Another form of the HWB function is:

$$\Psi(\overline{S}_\tau, S_\tau) \;=\; \max(a\overline{S}_\tau, S_\tau)$$

where $0 < a < 1$. Since:

$$\max(a\overline{S}_\tau, S_\tau) = S_\tau + \max(a\overline{S}_\tau - S_\tau, 0), \tag{3}$$

the second term on the r.h.s. of (3) is the payoff of a life-contingent fractional lookback option.

In practice, the policyholders can withdraw. Note that the payment of the GMDB can be written as the payoff of a put option (plus a share of underlying stock). When the stock price rises, put options become less valuable; hence, policies may lapse. This motivates considering the benefit function:

$$\Psi(\overline{S}_\tau, S_\tau) \;=\; I_{\{\overline{S}_\tau < H\}}[K - S_\tau]_+,$$

where $I_{\{.\}}$ is the indicator function and $H$ is a "barrier". This is a barrier option payoff form.

Since we can approximate the distribution of $\tau_x$ by a PH distribution, the valuation problem becomes calculating:

$$\mathbb{E}[\mathrm{e}^{-\delta\tau}\Psi(\overline{S}_\tau, S_\tau)], \tag{4}$$

where $\delta$ denotes a discounting factor and $\tau$ is PH r.v. One example is to take $\delta$ as the force of interest, but the company may want to use a lower $\delta$ in its technical reserve to ensure a conservative pricing.

## 5. The Factorization: Implementation for Jump Diffusions

Our basic assumption is that $X$ is a jump diffusion with Brownian parameters $\mu, \sigma^2$, upward jumps at rate $\lambda_1$ and with a PH$(p_1, \boldsymbol{\alpha}_1, \boldsymbol{T}_1)$ distribution of the sizes, and similar parameters $\lambda_2, p_2, \boldsymbol{\alpha}_2, \boldsymbol{T}_2$ for the downward jumps.

We will first need to introduce the equivalent time-reversed PH representation $(\boldsymbol{\alpha}^*, \boldsymbol{T}^*)$ of $\tau$. It is obtained by considering:

$$J_t^* = \begin{cases} J_{(\tau-t)-}, & \text{if } t < \tau, \\ \dagger, & \text{otherwise.} \end{cases}$$

It turns out (Asmussen 2003, III.5) that $J^*$ is a terminating time-homogeneous Markov process with initial distribution $\boldsymbol{\alpha}^*$ and generator $\boldsymbol{T}^*$ given by:

$$\boldsymbol{\alpha}^* = \mathbb{E}\tau \cdot \boldsymbol{t}^\intercal \boldsymbol{\Delta}_\nu, \quad \boldsymbol{T}^* = \boldsymbol{\Delta}_\nu^{-1} \boldsymbol{T}^\intercal \boldsymbol{\Delta}_\nu, \tag{5}$$

where $\boldsymbol{\Delta}_\nu$ denotes the diagonal matrix with the positive vector $\boldsymbol{\nu} = -\boldsymbol{\alpha}\boldsymbol{T}^{-1}$ on the diagonal. To see this, consider a doubly-infinite stationary Markovian arrival process (Asmussen 2003, XI.1) with generator $\boldsymbol{T} + \boldsymbol{D}$, where $\boldsymbol{D} = \boldsymbol{t}\boldsymbol{\alpha}$ gives the rates of phase changes with arrivals, and note that the corresponding stationary distribution of the phase process is proportional to $\boldsymbol{\nu}$. Now, just use the standard time

reversion for non-killed Markov processes. Considering, for the ease of exposition, the case $\delta = 0$ of no discounting, the form of the factorization identity of (Asmussen and Ivanovs 2019) that is most convenient for our purposes is then:

$$\mathbb{P}(M_\tau \in \mathrm{d}x, D_\tau \in \mathrm{d}y) = \sum_{k=1}^{p} \frac{1}{c_k} \mathbb{P}^*_{\boldsymbol{\alpha}}(M_\tau \in \mathrm{d}x, J_{\overline{\sigma}_\tau} = k) \mathbb{P}^*_{\boldsymbol{\alpha}^*}(M_\tau \in \mathrm{d}y, J_{\overline{\sigma}_\tau} = k)\,, \tag{6}$$

where $\mathbb{P}^*$ is the probability measure where $\tau$ is generated by the time-reversed representation and $X$ evolves as $-X$ did under $\mathbb{P}$, $\overline{\sigma}_\tau$ is the time at which the maximum is attained, and $c_k = \mathbb{P}_{\boldsymbol{\alpha}}(J_{\overline{\sigma}_\tau} = k) = \mathbb{P}^*_{\boldsymbol{\alpha}^*}(J_{\overline{\sigma}_\tau} = k)$.

The distributions of $M_\tau, D_\tau$ are particularly simple if $X$ is BM$(\mu, \sigma^2)$ and $\tau$ exponential $(\lambda)$. In fact, they are then exponential with rates $\rho^+$, respectively $\rho^-$, where:

$$\rho^\pm \;=\; \mp\frac{\mu}{\sigma^2} + \sqrt{\frac{\mu^2}{\sigma^4} + \frac{2\lambda}{\sigma^2}}.$$

If more generally, $X(t)$ is a jump diffusion with both up- and down-ward PH jumps, there is an abundance of results and algorithms giving the distributions in various representations or their Laplace transforms. See, among others, (Bean et al. 2005; Breuer 2008; Dieker and Mandjes 2011; Horváth and Telek 2017; Jacobsen 2005; Jiang and Pistorius 2008; Lewis and Mordecki 2008; Mordecki 2002; Pistorius 2006). The most appropriate form for our purposes is that for (computable) matrices $\boldsymbol{U}, \boldsymbol{U}^*$ of dimension $p(1+p_1) \times p(1+p_1)$, respectively $p(1+p_2) \times p(1+p_2)$, one has:

$$\mathbb{P}_i\big(M_\tau \in \mathrm{d}x,\, J_{\overline{\sigma}} = k\big) \;=\; \boldsymbol{e}_i^\intercal \mathrm{e}^{\boldsymbol{U}x} \boldsymbol{e}_k \cdot u_k \tag{7}$$

$$\mathbb{P}^*_j(M_\tau \in \mathrm{d}y, J_{\overline{\sigma}_\tau} = k) \;=\; \boldsymbol{e}_j^\intercal \mathrm{e}^{\boldsymbol{U}^*y} \boldsymbol{e}_k \cdot u^*_k \tag{8}$$

where $u_k, u^*_k$ are the $k^{\text{th}}$ element of the vectors $-\boldsymbol{U}\boldsymbol{e}$, respectively $-\boldsymbol{U}^*\boldsymbol{e}$; this occurs already in (Asmussen 1995), as well as several later sources, but we have in fact used a new and (we feel) somewhat more straightforward approach; see Appendix B.

Multiplying (7), (8) by $\alpha_i$, respectively $\alpha^*_j$ and summing over $i, j$, the factorization (6) then takes the form:

$$\mathbb{P}(M_\tau \in \mathrm{d}x, D_\tau \in \mathrm{d}y) \;=\; \sum_{k=1}^{p} r_k \cdot \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}x}\boldsymbol{e}_k \cdot \boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}^*y}\boldsymbol{e}_k$$

where $r_k = u_k u^*_k / c_k$. The modification with discounting is explained in Section 4 of (Asmussen and Ivanovs 2019) and becomes:

$$\mathbb{E}[\mathrm{e}^{-\delta\tau};\, M_\tau \in \mathrm{d}x, D_\tau \in \mathrm{d}y] = \sum_{k=1}^{p} r_k \cdot \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{e}_k \cdot \boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}^*_\delta y}\boldsymbol{e}_k \tag{9}$$

where $\boldsymbol{U}_\delta$, $\boldsymbol{U}^*_\delta$ are calculated as $\boldsymbol{U}, \boldsymbol{U}^*$ with $\boldsymbol{T}$ replaced by $\boldsymbol{T} - \delta\boldsymbol{I}$ (here, the $r_k$ remain unchanged, i.e., $u_k$, $u^*_k$ and $c_k$ are computed for the case $\delta = 0$). This can be rewritten in matrix form as follows:

**Corollary 1.** (i) *For $x, y \geq 0$,*

$$\mathbb{E}[\mathrm{e}^{-\delta\tau};\, M_\tau \in \mathrm{d}x, D_\tau \in \mathrm{d}y] = \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}^{*\intercal}_\delta y}\boldsymbol{\alpha}^{*\intercal} = \boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{U}^*_\delta y}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}_\delta^\intercal x}\boldsymbol{\alpha}^\intercal;$$

(ii) *For functions $f$ and $g$,*

$$\mathbb{E}\big[\mathrm{e}^{-\delta\tau} f(M_\tau) g(D_\tau)\big] \;=\; \boldsymbol{\alpha}\boldsymbol{F}_{\boldsymbol{U}_\delta}\boldsymbol{\Delta}_r\boldsymbol{G}^\intercal_{\boldsymbol{U}^*_\delta}\boldsymbol{\alpha}^{*\intercal} \;=\; \boldsymbol{\alpha}^*\boldsymbol{G}_{\boldsymbol{U}^*_\delta}\boldsymbol{\Delta}_r\boldsymbol{G}^\intercal_{\boldsymbol{U}_\delta}\boldsymbol{\alpha}^\intercal$$

*where* $\boldsymbol{F}_{\boldsymbol{U}_\delta}$, $\boldsymbol{G}_{\boldsymbol{U}^*_\delta}$ *are the matrices* $\int_0^\infty f(x)\mathrm{e}^{\boldsymbol{U}_\delta x}\,\mathrm{d}x$, *respectively* $\int_0^\infty g(y)\mathrm{e}^{\boldsymbol{U}^*_\delta y}\,\mathrm{d}y$.

**Proof.** Since $\boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}^*_\delta y}\boldsymbol{e}_k \in \mathbb{R}^1$, it may be replaced by its transpose $\boldsymbol{e}_k^\intercal \mathrm{e}^{-\boldsymbol{U}_\delta^{*\intercal} y}\boldsymbol{\alpha}^{*\intercal}$ in (9), and this gives the first identity in (i); the second follows by transposition. Part (ii) then follows by integration. □

For computations, the matrix form rather than the sum form is convenient for keeping the code short and transparent, and in many software packages like MATLAB, it also is faster. In a number of our examples, the matrices $\boldsymbol{F}_{\boldsymbol{U}}$, $\boldsymbol{G}_{\boldsymbol{U}^*}$ come out in a simple form, avoiding integration. Here is a first example:

**Example 1.** For the HWB, we need to compute:

$$\mathbb{E}\left[\mathrm{e}^{-\delta\tau}\max(a\overline{S}_\tau, S_\tau)\right] = S_0\mathbb{E}\left[\mathrm{e}^{-\delta\overline{\tau}}\mathrm{e}^{M_\tau}\cdot\mathrm{e}^{-\delta\underline{\tau}}\max(a,\mathrm{e}^{-D_\tau})\right]. \tag{10}$$

Thus (assuming $S_0 = 1$ w.l.o.g.), we have $f(x) = \mathrm{e}^x$, $g(y) = \max(a,\mathrm{e}^{-y})$ in Part (ii) of Corollary 1, and we get:

$$\begin{aligned}
\boldsymbol{F}_{\boldsymbol{U}_\delta} &= \int_0^\infty \mathrm{e}^x\mathrm{e}^{\boldsymbol{U}_\delta x}\,\mathrm{d}x = -(\boldsymbol{I}+\boldsymbol{U}_\delta)^{-1},\\
\boldsymbol{G}_{\boldsymbol{U}^*_\delta} &= a\int_{-\log a}^\infty \mathrm{e}^{\boldsymbol{U}^*_\delta y}\,\mathrm{d}y + \int_0^{-\log a}\mathrm{e}^{-y}\mathrm{e}^{\boldsymbol{U}^*_\delta y}\,\mathrm{d}y\\
&= -a\mathrm{e}^{-\boldsymbol{U}^*_\delta\log a}\boldsymbol{U}_\delta^{*-1} + (\boldsymbol{I}-a\mathrm{e}^{-\boldsymbol{U}^*_\delta\log a})(\boldsymbol{I}-\boldsymbol{U}^*_\delta)^{-1}.
\end{aligned}$$

♢

**Corollary 2.** *Define* $\boldsymbol{S}_\delta = \int_0^\infty \mathrm{e}^{\boldsymbol{U}_\delta y}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}_\delta^{*\intercal} y}\,\mathrm{d}y$. *Then:*

$$\mathbb{E}[\mathrm{e}^{-\delta\tau};\, X_\tau\in\mathrm{d}x] = \begin{cases}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{S}_\delta\boldsymbol{\alpha}^{*\intercal} & \text{for } x>0\\ \boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}^*_\delta x}\boldsymbol{S}_\delta^\intercal\boldsymbol{\alpha}^\intercal & \text{for } x<0.\end{cases}$$

**Proof.** For $x > 0$, we have:

$$\mathbb{E}[\mathrm{e}^{-\delta\tau};\, X_\tau\in\mathrm{d}x] = \int_0^\infty \mathbb{E}[\mathrm{e}^{-\delta\tau};\, M_\tau\in\mathrm{d}x+y, D_\tau\in\mathrm{d}y],$$

and so, the result follows immediately by Part (ii) of Corollary 1. The proof for $x < 0$ is similar; one only has to note that if $\boldsymbol{S}^*_\delta$ is defined as $\boldsymbol{S}_\delta$ with $\boldsymbol{U}_\delta$ and $\boldsymbol{U}^*_\delta$ interchanged, then $\boldsymbol{S}^*_\delta = \boldsymbol{S}_\delta^\intercal$. □

**Remark 2.** The distribution of $X_\tau$ restricted to $(0,\infty)$ is in fact (defective) PH, but in general, with a different generator $\widetilde{\boldsymbol{U}}$ than $\boldsymbol{U} = \boldsymbol{U}_0$ (for this one would have needed that $\boldsymbol{S}_\delta\boldsymbol{\alpha}^{*\intercal}$ be proportional to $-\boldsymbol{U}\boldsymbol{e}$). This follows from the general fact that if $\boldsymbol{\alpha}, \boldsymbol{s} \geq 0$ and $\boldsymbol{T}$ is a phase generator, but possibly $\boldsymbol{s} \neq -\boldsymbol{T}\boldsymbol{e}$, then $\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}x}\boldsymbol{s} = \widetilde{\boldsymbol{\alpha}}\mathrm{e}^{\widetilde{\boldsymbol{T}}x}\widetilde{\boldsymbol{t}}$ where:

$$\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\boldsymbol{\Delta}_v,\ \ \widetilde{\boldsymbol{T}} = \boldsymbol{\Delta}_v^{-1}\boldsymbol{T}\boldsymbol{\Delta}_v,\ \ \widetilde{\boldsymbol{t}} = -\widetilde{\boldsymbol{T}}\boldsymbol{e} = \boldsymbol{\Delta}_v^{-1}\boldsymbol{s}\ \text{ with } \boldsymbol{v} = -\boldsymbol{T}^{-1}\boldsymbol{s},$$

and here, $\widetilde{\boldsymbol{T}}$ is a phase generator; all needed positivity properties follow from $\boldsymbol{S}_\delta$ and $-\boldsymbol{T}^{-1}$ being non-negative.

Similar remarks apply to the negative part of $X_\tau$. These observations may be seen as an extension to jump diffusions at PH times of the standard fact that Brownian motion at an independent exponential time has a two-sided exponential distribution (often called the Laplace distribution). A similar result for matrix-exponential distributions is on p. 478 of (Bladt and Nielsen 2017), but is in terms of Laplace transforms rather than densities.

**Example 2.** For the GMDB, the price is (taking $S_0 = 1$):

$$\mathbb{E}[\mathrm{e}^{-\delta\tau}\max(S_\tau, K)] \;=\; K\mathbb{E}[\mathrm{e}^{-\delta\tau}; X_\tau \le \log K] + \mathbb{E}[\mathrm{e}^{-\delta\tau + X_\tau}; X_\tau > \log K]\,.$$

In the most common case $K < S_0$, the first term equals:

$$K\int_{-\infty}^{\log K} \boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}_\delta^* x}\boldsymbol{S}_\delta^{\intercal}\boldsymbol{\alpha}^{\intercal}\,\mathrm{d}x \;=\; -K\boldsymbol{\alpha}^*\boldsymbol{U}_\delta^{*-1}\mathrm{e}^{-\boldsymbol{U}_\delta^*\log K}\boldsymbol{S}_\delta^{\intercal}\boldsymbol{\alpha}^{\intercal}$$

when $S_0 = 1$, whereas the second is:

$$\begin{aligned}&\int_{\log K}^{0} \mathrm{e}^{x}\boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}_\delta^* x}\boldsymbol{S}_\delta^{\intercal}\boldsymbol{\alpha}^{\intercal}\,\mathrm{d}x \;+\; \int_0^\infty \mathrm{e}^{x}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{S}_\delta\boldsymbol{\alpha}^{*\intercal}\,\mathrm{d}x\\ &\quad= \boldsymbol{\alpha}^*(\boldsymbol{I} - \boldsymbol{U}_\delta^*)^{-1}\big(\boldsymbol{I} - K\mathrm{e}^{-\boldsymbol{U}_\delta^*\log K}\big)\boldsymbol{S}_\delta^{\intercal}\boldsymbol{\alpha}^{\intercal} - \boldsymbol{\alpha}(\boldsymbol{I} + \boldsymbol{U}_\delta)^{-1}\boldsymbol{S}_\delta\boldsymbol{\alpha}^{*\intercal}\,.\end{aligned}$$

◊

**Example 3.** The calculation of reserves:

$$V_t \;=\; \mathbb{E}\big[\mathrm{e}^{-\delta(\tau-t)}\Phi(M_\tau, D_\tau)\,\big|\,X_s,\, 0 \le s \le t,\, \tau > t\big]$$

is closely related to price calculations. However, the details may be somewhat more cumbersome since for a fixed $t$, the expressions for $M_\tau, D_\tau$ have two components, the quantities $m_0 = \max_{s\le t} X_s$ and $d_0 = m_0 - X_t$, which are known at time $t$, and the unknown random ones:

$$m_1 \;=\; \max_{t\le s<\tau}(X_s - X_t)\,,\quad d_1 = m_1 - x_1 \;\text{ where } x_1 = X_\tau - X_t\,.$$

In terms of these and the relations:

$$\begin{aligned}M_\tau \;&= \max(m_0, x_0 + m_1),\\ D_\tau \;&=\; \max(m_0, x_0 + m_1) - x_0 - x_1 \;=\; \max(d_0, d_1) - x_1\\ &=\; [d_0 - m_1]^+ + d_1\,,\end{aligned}$$

we can write $\Phi(M_\tau, D_\tau) = \widetilde{\Phi}(m_1, d_1)$, where:

$$\widetilde{\Phi}(x, y) \;=\; \Phi\big(\max(m_0, x_0 + x), [d_0 - x]^+ + y\big).$$

Since the conditional distribution of $\tau - t$ given $\tau > t$ is $\mathrm{PH}(\boldsymbol{\alpha}_t, \boldsymbol{T})$ where $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}t}/\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}t}\boldsymbol{e}$, this gives:

$$V_t \;=\; \int_0^\infty\int_0^\infty \widetilde{\Phi}(x, y)\boldsymbol{\alpha}_t\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}_\delta^{\intercal} y}\boldsymbol{\alpha}^{*\intercal}\,\mathrm{d}y\,\mathrm{d}x \tag{11}$$

(note that for the reversal, we need not calculate $\boldsymbol{\alpha}^*$ by replacing $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}_t$ in (5), but can use the same $\boldsymbol{\alpha}^*$ for all $t$, cf. Section 2.2 of (Asmussen and Ivanovs 2019)).

**Example 4.** For the GMDB, we have:

$$\Phi(M_\tau, D_\tau) \;=\; \max(K, \mathrm{e}^{X_\tau}) \;=\; \max(K, \mathrm{e}^{x_0 + x_1}) \;=\; \mathrm{e}^{x_0}\max(\widetilde{K}, \mathrm{e}^{x_1})$$

where $\widetilde{K} = K\mathrm{e}^{-x_0}$. Combining with Remark 2, we therefore arrive at the same expression as in Example 2 when $K\mathrm{e}^{-x_0} \le 1$, only with $K$ replaced by $\widetilde{K}$ and $\boldsymbol{\alpha}$ by $\alpha_t$. The case $K\mathrm{e}^{-x_0} > 1$ may arise by randomness, but is just a minor modification.

**Example 5.** The details for the HWB reserves are somewhat more complicated than for the GMDB. Here, one needs to consider $\mathrm{e}^{M_\tau}\max(a,\mathrm{e}^{-D_\tau})$, corresponding to:

$$\widetilde{\Phi}(x,y) \;=\; \exp\bigl\{\max(m_0,x_0+m_1)\bigr\}\max(a,\exp\bigl\{-[d_0-m_1]^+-d_1\bigr\}.$$

One approach (potentially the easier one) may be simply to insert in (11) and do the two-dimensional numerical integration. However, in fact, one can reduce to a one-dimensional numerical integration similar to Corollary 2. To this end, note that $\mathrm{e}^{M_\tau}\max(a,\mathrm{e}^{-D_\tau})$ after simple derivations comes out as:

$$\begin{cases} \mathrm{e}^{x_0+m_1}a & \text{if } m_1 > d_0,\ d_1 > -\log a, \quad (12a)\\ \mathrm{e}^{x_0+m_1-d_1} & \text{if } m_1 > d_0,\ d_1 \le -\log a, \quad (12b)\\ \mathrm{e}^{m_0}a & \text{if } m_1 \le d_0,\ d_0-m_1+d_1 > -\log a, \quad (12c)\\ \mathrm{e}^{m_0}\mathrm{e}^{-d_0+m_1-d_1} & \text{if } m_1 \le d_0,\ d_0-m_1+d_1 \le -\log a. \quad (12d) \end{cases}$$

of these four possibilities, (12d) (and to a slightly lesser extent, (12c)) is the most complicated one with which to deal. The contribution to the reserve from (12c) is zero if $m_0 < d_0+\log a$, since then, the defining region is empty. Otherwise, it becomes:

$$\mathrm{e}^{-\delta t+m_0-d_0}\int_{m_0+\log a}^{d_0}\int_0^{x-d_0-\log a}\mathrm{e}^{x-y}\boldsymbol{\alpha}_t\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}_\delta^{*\intercal}y}\boldsymbol{\alpha}^{*\intercal}\,\mathrm{d}y\,\mathrm{d}x$$

which after two integrations reduces to an expression, which, though complicated, only involves known matrices and their exponentials, except that:

$$\boldsymbol{S}_\delta(b,c) \;=\; \int_b^c \mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{\Delta}_r\mathrm{e}^{\boldsymbol{U}_\delta^{*\intercal}x}\,\mathrm{d}x$$

is left to be determined by numerical integration. ◊

## 6. Numerical Examples

For the underlying stock price process $S_t = S_0\mathrm{e}^{X_t}$, the most popular model is the geometric Brownian motion model. Since this model cannot capture the asymmetric leptokurtic features and the volatility smile, many modified models have been proposed in the literature; one of these is the jump diffusion model, cf. (Kou 2002). Empirical studies show that the daily return distribution tends to have larger kurtosis than the distribution of monthly returns, and the jump diffusion model can capture this feature; see (Das and Foresi 1996). As an illustrative example, we suppose that the yearly interest rate is $r = 3\%$ and the yearly volatility $\sigma = 0.25$. The jumps follow a compound Poisson process with upward jump rate $\lambda_1 = 3$ with the jump sizes being exponential with rate $\nu_1 = 50$ (mean $1/50$) and downward jump rate $\lambda_2 = 2$ with jump size being exponential with rate $\nu_2 = 30$. When considering the valuation of an insurance product alone, it is not necessary to use a risk-neutral valuation approach. However, since in our setting, the underlying involves an equity, it is natural to use an equivalent martingale measure $\mathbb{Q}$, defined by the requirement $\log\mathbb{E}_\mathbb{Q}\mathrm{e}^{X(1)} = r$. This does not define $\mathbb{Q}$ uniquely since the market is incomplete due to the jumps and the mortality risk. Our choice was to let the parameters $\sigma^2,\lambda_1,\lambda_2,\nu_1,\nu_2$ remain unchanged, which gives the risk-neutral Brownian drift as:

$$\mu \;=\; r-\frac{\sigma^2}{2}-\frac{\lambda_1}{\nu_1-1}+\frac{\lambda_2}{\nu_2+1}.$$

For the illustrative example, we have:

$$\mu \;=\; 0.03 - \frac{0.25^2}{2} - \frac{3}{50-1} + \frac{2}{30+1} \;=\; 0.002042.$$

It is left to compute the $\boldsymbol{U}_\delta, \boldsymbol{U}_\delta^*$. Writing:

$$\boldsymbol{U} \;=\; \boldsymbol{U}(\boldsymbol{T}, \mu, \sigma^2, \lambda_1, \lambda_2, \nu_1, \nu_2)$$

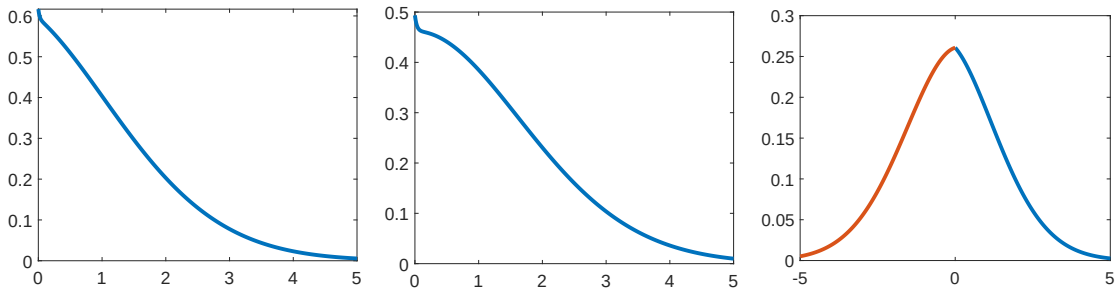
we obtain:

$$\boldsymbol{U}^* \;=\; \boldsymbol{U}(\boldsymbol{T}^*, -\mu, \sigma^2, \lambda_2, \lambda_1, \nu_2, \nu_1)\,,\;\; \boldsymbol{U}_\delta \;=\; \boldsymbol{U}(\boldsymbol{T} - \delta \boldsymbol{I}, \mu, \sigma^2, \lambda_1, \lambda_2, \nu_1, \nu_2)\,,$$

and thus, it suffices to consider $\boldsymbol{U}$. Essentially, two types of algorithms are available. One set is based on finding roots in the complex plane of equations of the form $\kappa(s) = a$ where $\kappa(s) = \log \mathbb{E}e^{sX(1)}$ is the Lévy exponent of $X$. The other uses fixed-point equations of the form $\boldsymbol{U} = \psi(\boldsymbol{U})$ and iteration, $\boldsymbol{U} = \lim_{n\to\infty} \boldsymbol{U}_n$, with $\boldsymbol{U}_n = \psi(\boldsymbol{U}_{n-1})$ and $\boldsymbol{U}_0$ suitably chosen.

We give a few examples of previous algorithms of both types in Appendix B. When implementing these with $\tau$ having one of the GC representations fitted in Section 3, our experience was that the root-finding method was only feasible for a rather small number $p$ of phases. More precisely, for $p = 50$ or $p = 75$, meaningless results were produced, with warnings of the matrix $\boldsymbol{H}$ being close to singular. This is understandable since many eigenvalues of the relevant matrices are almost identical. Since the iterative algorithms ran without problems also for large values of $p$, we have therefore concentrated on these. In fact, we have developed a new iterative algorithm, which may be simpler than previous ones. It is presented in Appendix B, but for now, we concentrate on some numerical examples.

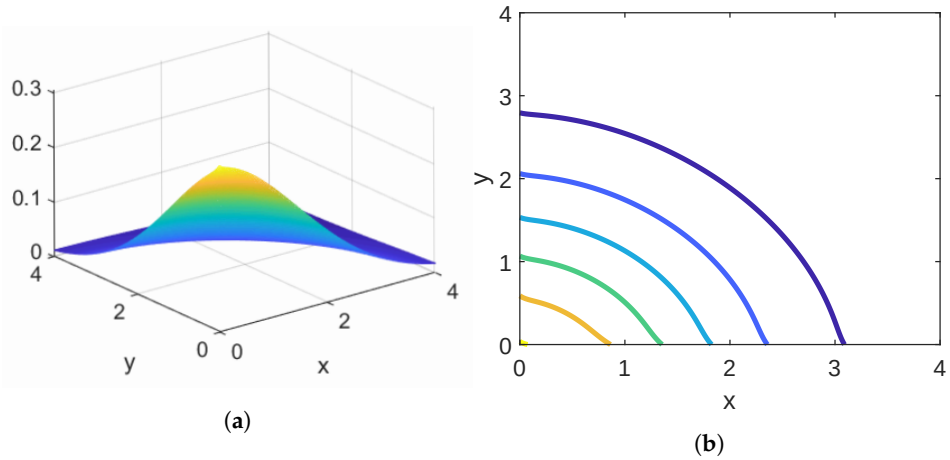
**Example 6.** In Figure 7, we have plotted the marginal densities of $M_\tau$, $D_\tau$, and $X_\tau$.




(**a**) (**b**) (**c**)

**Figure 7.** Densities of (**a**) $M_\tau$, (**b**) $D_\tau$, and (**c**) $X_\tau$.

**Example 7.** Figure 8 gives the joint distribution of $M_\tau$ and $D_\tau$ (computed by Corollary 1) for our jump diffusion with the parameters given above.




(**a**) (**b**)

**Figure 8.** Joint distribution of $M_\tau$ and $D_\tau$. The joint density is in the left panel and a contour plot in the right.

**Example 8.** To consider the sensitivity of the algorithms to the value $p$ of chosen phases for $\tau$, we took $\delta = r = 0.03$ or 0.00 and computed the price (10) for $p = 20$, 35, 40, 50, 75, 100 and the GC fits presented in Section 3 for two products, the HWB with $a = 0.85$ and the GMDB with $K = 0.85$; the relevant formulas are available from Examples 1 and 2. The results are in Table 1. We also complemented with a Monte Carlo simulation estimate with an associated 95% confidence interval based on $R = 10^6$ replications, where $\tau$ was generated from the discrete life table distribution with an additional randomization within the year.

**Table 1.** Price estimates for $\delta = r$ and varying $p$. HWB, high-water benefit; GMDB, guaranteed minimum death benefit.

| $p$ | 20 | 35 | 40 | 50 | 75 | 100 | Monte Carlo |
|---|---|---|---|---|---|---|---|
| HWB, $\delta = r = 0$ | 2.703 | 2.703 | 2.704 | 2.704 | 2.704 | 2.704 | $2.665 \pm 0.008$ |
| HWB, $\delta = r = 0.03$ | 1.698 | 1.699 | 1.699 | 1.699 | 1.699 | 1.699 | $1.680 \pm 0.008$ |
| GMDB, $\delta = r = 0$ | 1.468 | 1.468 | 1.468 | 1.468 | 1.468 | 1.468 | $1.467 \pm 0.008$ |
| GMDB, $\delta = r = 0.03$ | 1.080 | 1.080 | 1.079 | 1.079 | 1.079 | 1.079 | $1.079 \pm 0.010$ |

A related question is how varying different fits can be for a fixed $p$ due to issues such as non-uniqueness of PH representations and possibly local maxima of the likelihood. We illustrate this point in Table 2, which is a parallel of Table 1. We fix $p$ at 50, but use five different initial values (seeds) for the EM algorithm, all allowed the same running time. The fitted densities are in Figure 9.

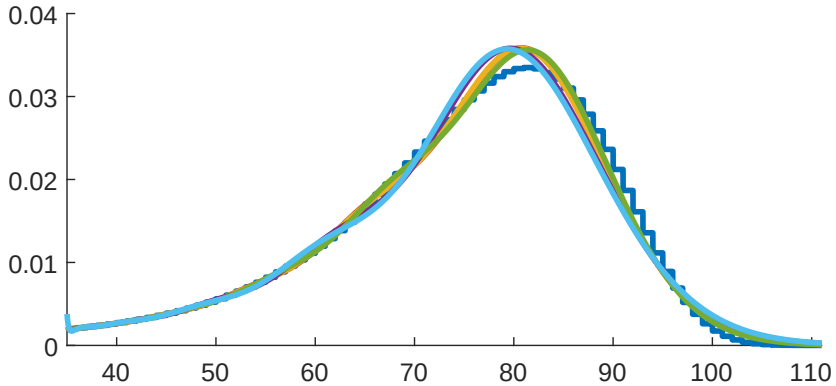



**Figure 9.** Fitted densities for five different seeds (against the life table data).

Our conclusion is that the differences are too minor to be of any concern, and certainly, using a single fit with $p \geq 50$ will produce acceptable results. Insisting on more than 2–3-digit precision in

price estimates does not make sense due to issues such as model uncertainty and that payments will be adjusted later via bonus and dividend arrangements.

**Table 2.** Price estimates for $\delta = r$ and varying seeds.

| **Seed** | **1** | **2** | **3** | **4** | **5** | **Monte Carlo** |
|---|---|---|---|---|---|---|
| HWB, $\delta = r = 0$ | 2.704 | 2.704 | 2.703 | 2.702 | 2.704 | $2.665 \pm 0.008$ |
| HWB, $\delta = r = 0.03$ | 1.699 | 1.699 | 1.699 | 1.698 | 1.699 | $1.680 \pm 0.008$ |
| GMDB, $\delta = r = 0$ | 1.468 | 1.468 | 1.468 | 1.467 | 1.468 | $1.467 \pm 0.008$ |
| GMDB, $\delta = r = 0.03$ | 1.079 | 1.079 | 1.079 | 1.078 | 1.080 | $1.079 \pm 0.010$ |

The programming was done in MATLAB on a 2014 MacBook Air with a 1.4 GHz Intel Core i5 processor. As an example of execution times, MATLAB's tic-toc command reported approximately two seconds for each HWB or GMDB value using the PH methodology. The figures for the Monte Carlo values were much higher, approximately 30,000 replications per minute (recall that $10^6$ replications were used). The programming is very simple, with the main effort lying in writing a routine for computing $\boldsymbol{U}$ for a given set of input parameters. This amounted to 50 lines of code in our implementation. Given this ease, we did some further price computations, even if our aim here is not to illustrate pricing aspects, but rather the computational ones.

**Example 9.** We first consider varying the discount rate $\delta$, but not the force of interest $r$, cf. the remarks following (4) (that in particular motivate taking $\delta \le r$). The results for the HWB are in Table 3 and show the expected picture, that the price is decreasing in the discount rate.

**Table 3.** HWB price estimates for $r = 0.03$ and varying $\delta$.

| $\delta$ | **0.00** | **0.01** | **0.02** | **0.03** |
|---|---|---|---|---|
| Price | 6.24 | 3.99 | 2.58 | 1.70 |

**Example 10.** The next example is discounting with the force of interest, i.e., taking $r = \delta$ and varying $r$. The numerical values are in Table 4, and the picture is again that the price is decreasing in $r = \delta$. The intuition is less clear-cut here, since the Brownian drift $\mu$ is increasing and thus pulls in the opposite direction.

**Table 4.** HWB price estimates for varying $r = \delta$.

| $r = \delta$ | **0.00** | **0.01** | **0.02** | **0.03** | **0.05** |
|---|---|---|---|---|---|
| Price | 2.70 | 2.23 | 1.92 | 1.70 | 1.44 |

**Example 11.** Next, we give an illustration of a feature of the factorization: if $\tau$ is exponential, then $M_\tau$ and $D_\tau$ are independent. cf. Lemma 1. In the PH case, there is dependence caused by the common value of $\overline{\tau} = k$. How does this influence the price? A numerical illustration is in Table 5, where the entries are the ratios between the correct price with dependence and the incorrect one computed assuming independence, i.e., as:

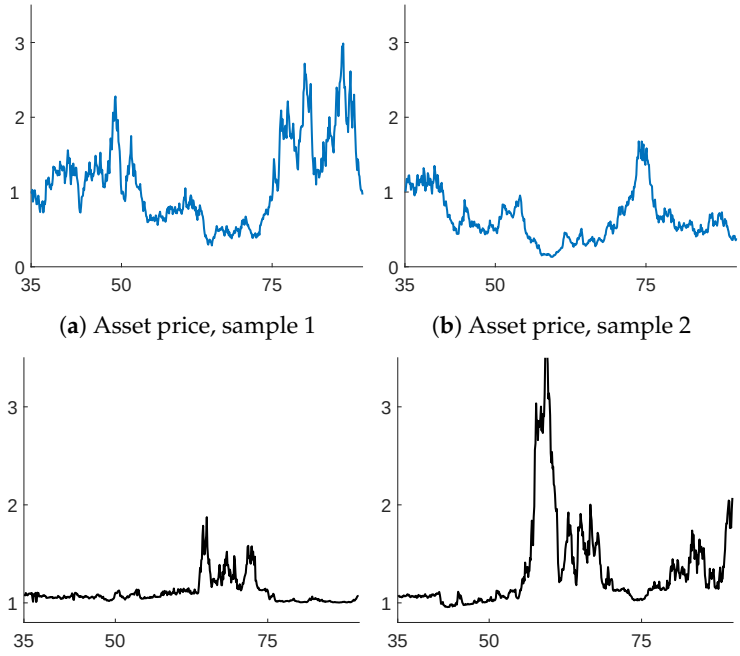
$$\begin{aligned}
&\mathbb{E}\mathrm{e}^{-\delta\overline{\tau}}\mathrm{e}^{M_\tau} \cdot \mathbb{E}\mathrm{e}^{-\delta\underline{\tau}} \max(a, \mathrm{e}^{-D_\tau}) \\
&\quad = \int_0^\infty \mathrm{e}^x \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{U}_\delta x}\boldsymbol{u}\,\mathrm{d}x \cdot \left(a\int_{-\log a}^\infty \boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{U}_\delta^* y}\boldsymbol{u}^*\,\mathrm{d}y + \int_0^{-\log a} \mathrm{e}^{-y}\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{U}_\delta^* y}\boldsymbol{u}^*\,\mathrm{d}y\right) \\
&\quad = -\boldsymbol{\alpha}(\boldsymbol{I}+\boldsymbol{U}_\delta)^{-1}\boldsymbol{u}\left(-a\,\boldsymbol{\alpha}^*\mathrm{e}^{-\boldsymbol{U}_\delta^*\log a}\boldsymbol{U}_\delta^{*-1}\boldsymbol{u}^* + \boldsymbol{\alpha}^*(\boldsymbol{I} - a\mathrm{e}^{-\boldsymbol{U}_\delta^*\log a})(\boldsymbol{I}-\boldsymbol{U}_\delta^*)^{-1}\boldsymbol{u}^*\right).
\end{aligned}$$

The other settings are as in Table 3. Table 5 shows that in some cases, dependence indeed influences the price quite a bit.

**Table 5.** Ratios between prices with and without dependence.

| $\delta$ | **0.00** | **0.01** | **0.02** | **0.03** | **0.05** |
|---|---|---|---|---|---|
| Ratio | 0.97 | 0.88 | 0.78 | 0.68 | 0.51 |

**Example 12.** As a follow-up of Example 3, we finally consider the evaluation of reserves. Two paths of the jump diffusion were simulated, and the reserve was calculated for each value of $t$ using the formulas of Examples 3, 4 and 5. The results are given in the graphs of Figure 10. The upper panel gives the stock price $S_{a-35} = \mathrm{e}^{X_{a-35}}$ (assuming $S_{35} = 1$) and the lower one the ratio $V_{a-35}/\mathrm{e}^{X_{a-35}}$ as function of the age $a$ of the insured.




(**a**) Asset price, sample 1 (**b**) Asset price, sample 2

(**c**) Reserve/asset price, sample 1 (**d**) Reserve/asset price, sample 2

**Figure 10.** Asset price paths and excess ratio of associated GMDB reserves.

When $\delta = r$,

$$\begin{aligned} V_t &= \mathbb{E}\big[\mathrm{e}^{-\delta(\tau-t)}\max(K,\mathrm{e}^{X_\tau}) \,\big|\, \tau > t, X_t\big] \\ &\geq \mathbb{E}\big[\mathrm{e}^{-\delta(\tau-t)}\mathrm{e}^{X_\tau}) \,\big|\, \tau > t, X_t\big] \;=\; \mathrm{e}^{X_t} \end{aligned}$$

because of the martingale property. Hence, the ratio in the lower panel should always be at least one. This is also confirmed from the figure, as well as the expectation that the ratio should be close to one when $X_t$ is large.

In the calculations, we used the fact that $\alpha^*$ and $\boldsymbol{U}^*$ need not be recalculated for each $t$, cf. the remarks following (11). This gave a substantial speed-up.

## 7. Erlangization and Extrapolation

A multitude of papers in finance, insurance, and queuing theory explore the idea of Erlangization or Canadization, that is approximation of a deterministic time $T$ by an Erlang$(q)$ time $E = E_{q,q/T}$ with the same mean. Here, $q$ is the number of stages and $q/T$ is the rate, so that $E_{q,q/T}$ is the sum of $q$ mean $T/q$ exponential r.v.'s. This idea has various applications: explicit identities (see, e.g., (Asmussen et al. 2002; Carr 1998; Stanford et al. 2005)); this is used in (Kuznetsov et al. 2011) together with the traditional Wiener-Hopf factorization to provide an algorithm simulating $(M_E, X_E)$

to approximate $(M_T, X_T)$ and as a numerical tool for calculating $f(T)$ as the limit of $\mathbb{E}f(E_{q,q/T})$ (or just $f(E_{q,q/T})$) under suitable continuity conditions on $f$; see, e.g., (Asmussen et al. 2002; Carr 1998; Leung et al. 2015). This last procedure uses the fact that $E_{q,q/T}$ converges in probability to $T$ as $q \to \infty$. In (Asmussen et al. 2002), it is combined with extrapolation, which is based on the asymptotics:

$$\mathbb{E}f(E_{q,q/T}) \;=\; f(T) + \frac{C}{q} + \mathrm{O}(q^{-2}) \tag{13}$$

valid for $f$ smooth and some $C = C(f)$. This $C$ is typically not available, but (13) suggests eliminating $C$ by estimating $f(T)$ using:

$$q\mathbb{E}f(E_{q,q/T}) - (q-1)\mathbb{E}f(E_{(q-1),q/T}) \tag{14}$$

which converges to $f(T)$ at the improved rate $\mathrm{O}(q^{-2})$. The numerical examples of (Asmussen et al. 2002) indicate that often, calculations with $q$ set as small as 2–4 will give very good precision.

The relevance for the present purposes is that the time horizon of many life insurance contracts is not the remaining lifetime $\tau$ of the insured, but rather $T \wedge \tau$, where $T$ is a fixed deterministic number. For example, when the age of the insured is 35 at the time of underwriting, $T = 45$ corresponds to contract expiry at age 80 if the insured is still alive. Some calculations can be found in Section 10 of (Gerber et al. 2012) for $\tau$ exponential and a call or put option, but rely heavily on this particular structure. For our life insurance applications, it is crucial that $E_{q,q/T} \wedge \tau$ is again PH when $E_{q,q/T}$ and $\tau$ are independent. The details involve the Kronecker product $\otimes$ and sum $\oplus$ so that:

$$\boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}, \quad (\boldsymbol{A} \oplus \boldsymbol{B})(\boldsymbol{v} \otimes \boldsymbol{h}) = \boldsymbol{A}\boldsymbol{v} \otimes \boldsymbol{h} + \boldsymbol{v} \otimes \boldsymbol{B}\,.$$

For two Markov process generators, the Kronecker sum is the generator of independent versions, and so, $E_{q,q/T} \wedge \tau$ is PH with generator $\boldsymbol{E} \oplus \boldsymbol{T}$ of dimension $nq \times nq$ where $\boldsymbol{E} = \boldsymbol{E}_{q,q/T}$ is the PH generator of $E_{q,q/T}$. Written out with lexicographical order of the states:

$$\boldsymbol{E} \oplus \boldsymbol{T} \;=\; \begin{pmatrix} \boldsymbol{T}-\beta\boldsymbol{I} & \beta\boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}-\beta\boldsymbol{I} & \beta\boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & & \ddots & & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{T}-\beta\boldsymbol{I} & \beta\boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{T}-\beta\boldsymbol{I} \end{pmatrix}.$$

**Example 13.** We consider a HWB or GMDB contract signed at age 35 of the insured and expiring at the time $\tau$ of death or at age 70, whatever comes first. We took $r = \delta = 3\%$ and used the GC fit with $p = 50$, leading to the numerical values given in Table 6; in comparison, Monte Carlo simulation with $R = 50{,}000$ replications gave the 95% confidence interval $1.636 \pm 0.034$ for the HWB and $1.097 \pm 0.029$ for the GMDB.

**Table 6.** Erlangized price estimates.

| $q$ | HWB Simple | HWB Extrapolated | GMDB Simple | GMDB Extrapolated |
|---|---|---|---|---|
| 1 | 1.523 | 1.523 | 1.092 | 1.092 |
| 2 | 1.583 | 1.642 | 1.097 | 1.102 |
| 3 | 1.606 | 1.652 | 1.097 | 1.099 |
| 4 | 1.618 | 1.655 | 1.097 | 1.096 |
| 5 | 1.626 | 1.657 | 1.097 | 1.095 |
| 6 | 1.631 | 1.658 | 1.096 | 1.095 |
| 7 | 1.635 | 1.658 | 1.096 | 1.094 |
| 8 | 1.638 | 1.659 | 1.096 | 1.094 |
| 9 | 1.640 | 1.659 | 1.096 | 1.094 |
| 10 | 1.642 | 1.659 | 1.095 | 1.094 |

The extrapolation is done via (14), with the required smoothness being obvious for the GMDB. For the HWB, $d \mapsto \max(a, \mathrm{e}^{-d})$ is not smooth at $d = -\log a$, but some rough heuristics (that we omit) suggest that (13) remains in force for this special $f$. Certainly, extrapolation appears to produce sensible results also for the GMDB. In both cases, it seems that $q = 4$–$6$ is enough to obtain sufficiently precise estimates, whereas without extrapolation, one would need at least $q = 10$. This is a considerable improvement upon simple Erlangization, since the dimension of the matrices in the computations (where inversion is the heavy part) is $2pq$. This means that for $p = 50$, one would need to invert matrices of dimension 400 when $q = 4$, but 1000 when $q = 10$, and since the inversion of an $n \times n$ matrix has complexity $n^3$ (or slightly less with sophisticated methods), this corresponds to reducing computation time by a factor of about $2.5^3 = 15.6$.

## 8. Conclusions

In this paper, we have considered pricing calculations for life insurance products like the guaranteed minimum death benefit and the high-water benefit. The payout occurs at the death time $\tau$ of the insured, and its size is linked to an asset price $S$ either via its current value $S_\tau$ or to features of its past evolution during the time of the contract such as the maximum $\overline{S}_\tau$.

The key observation is that for the most popular model of $S$, an exponential jump diffusion, the distributions of $S_\tau$ and/or $\overline{S}_\tau$ are available when the jumps and $\tau$ are phase-type. This has long been known in the geometric Brownian motion setting without jumps when $\tau$ is exponential and is essentially just the Wiener-Hopf factorization of Brownian motion. There is also an extensive literature for PH jumps, and for $\tau$ itself being PH, the relevant Wiener-Hopf factorization has recently been developed in (Asmussen and Ivanovs 2019).

The program thus involves two steps, first fitting a PH distribution of $\tau$ to lifetime data and next implementing the pricing calculations.

The shape of lifetime data with a steep decline after the mode makes the fitting of PH distributions non-trivial, necessitating the number $p$ of phases to be relatively high. However, we already made the points that a good fit is one that produces reliable estimates of the prices of the life insurance products, and that giving more than 2–3-digit precision in price estimates does not make sense due to issues such as uncertainty about the model and its parameters (think, e.g., of longevity risk) and bonus/dividend arrangements. We found $p = 50$ to be more than sufficient in the sense of providing appropriate accuracy of the calculated price of life insurance products and could well have worked with smaller values. An alternative is developed in GSY, to use matrix-exponentials fitted at a selected number of points, but we show in Appendix A that this approach has its pitfalls.

Fitting of so high-dimensional PH distributions is slow, though we developed in part some new software giving substantial speed-up. However, given this step has been passed, pricing calculations are both very easy to program and run very fast, also compared to Monte Carlo simulation, and this may be the main advantage of our approach. A limitation is of course that we need the pay-out to be of the form (1), and there are many products for which this is not the case. Alternatives such as Monte Carlo simulation and traditional methods based on Thiele-type differential equations may then be more suitable. Therefore, we do not claim the superiority of our own approach for all purposes, but have rather tried to clarify its advantages and disadvantages.

**Author Contributions:** Conceptualization and methodology, S.A. and H.Y.; software and validation, S.A. and P.J.L.; writing–original draft preparation, all authors; writing–review and editing, P.J.L.

**Funding:** Hailiang Yang was funded by Research Grants Council of the Hong Kong Special Administrative Region (Project No. HKU 17305018). Patrick J. Laub conducted this research with the support of the ANR under the "Investments for the Future" program and the support of DAMI—Data Analytics and Models for Insurance—Chair under the aegis of the Fondation du Risque, a joint initiative by UCBL (Université Lyon 1, SAF EA2429) and BNP Paribas Cardif.

**Acknowledgments:** We are grateful to Jevgenijs Ivanovs for useful discussions and comments.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

## Appendix A. Fits Using Sum-of-Exponentials

We here briefly reinspect the approach of (Siu et al. 2015) for fitting a distribution with tail of the form:

$$\overline{F}_{\boldsymbol{a},\boldsymbol{b}}(t) \;=\; \sum_{i=1}^{\ell} a_i \mathrm{e}^{-b_i t}$$

to lifetime data where $\boldsymbol{a}$ is the vector $(a_1, \ldots, a_\ell)$ (some $a_i$ may be negative) and similarly for $\boldsymbol{b}$. Such a tail is of matrix-exponential form, meaning $\overline{F}(t) = \boldsymbol{a}\mathrm{e}^{\boldsymbol{\Delta} t}\boldsymbol{s}$ for suitable $\boldsymbol{a}, \boldsymbol{\Delta}, \boldsymbol{s}$ (not restricted to the Markov process interpretation in PH tails). In fact, the only restriction is that $\boldsymbol{\Delta}$ must be diagonalizable. We refer to (Bladt and Nielsen 2017) for further background on matrix-exponential distributions.

In (Siu et al. 2015), the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are found by minimizing:

$$\sum_{t=1}^{n} \left({}_t p_x - \overline{F}_{\boldsymbol{a},\boldsymbol{b}}(t)\right)^2 \tag{A1}$$

for some fixed $\ell$ and $n$, where ${}_t p_x$ is the survival probability (derived from the life table) of a person aged $x$ surviving until age $x + t$. In other words, they minimize the mean squared error between the empirical and theoretical survival probabilities at ages $x, x + 1, \ldots, x + n$.

In (Siu et al. 2015), the values $x = 30$ and $n = 25$ are used as examples, and so, the age span considered begins at 30 and goes until 55. We fit this type of approximation to our life table with similar parameters and plotted the resulting densities; we used $\ell = 10$ exponential terms, the same as in (Siu et al. 2015), but we set the starting age as $x = 35$ rather than $x = 30$.

The relevant ages for life insurance are often greater than 60 years old, and so, we create a sequence of fits with $n = 25, 50, 75$ (shown in Figures A1 and A2a) to try to approximate accurately an increasing portion of the life table. For each fit, we plot time across the whole age range of 35–110, even though one cannot expect a sum-of-exponentials fit to be accurate outside of the time period $t \in (x, x + n)$.

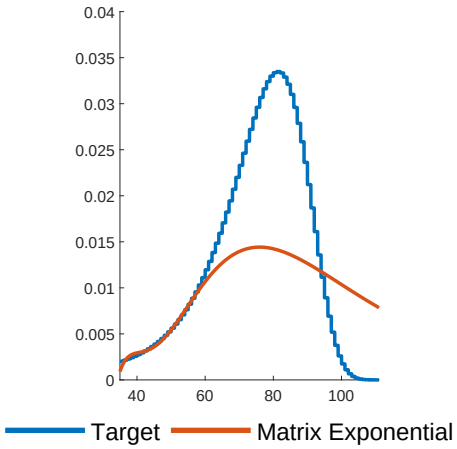
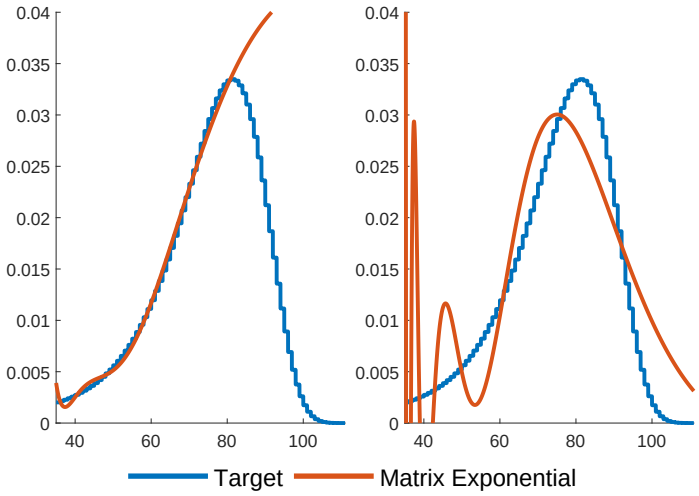



**Figure A1.** Sum-of-exponentials fits in age range 35–60.




(**a**) (**b**)

**Figure A2.** Sum-of-exponentials fits in age ranges 35–85 (left) and 35–110 (right).

In Figures A1 and A2a, the fits are fine in the selected age range $(x, x+n)$, but fail miserably beyond. This is not necessarily a drawback if the fit has been made for the pricing of a product where $x+n$ is the upper limit of the relevant age range. The picture in Figure A2b is more rough, but presumably repairable by increasing $\ell$ beyond $\ell = 10$. Another phenomenon that is observed is the possibility of densities that are negative and exhibit weird oscillations. This is seen most markedly in Figure A2b.

This behavior is a result of the fitting criterion (A1), which only pays attention to the selected grid $x, x+1, \ldots, x+n$, and not to what goes on in between the grid points.

## Appendix B. Algorithms for Computing $U$

As an example of the root-finding algorithms, we state the following particularly simple example, which is a special case of Corollary 5.2 of (Asmussen 1995):

**Proposition A1.** *Assume $X$ is $BM(\mu, \sigma^2)$ and that $\tau$ is GC, as in Figure 1. Then, $\mathbb{P}_i(M_\tau \in \mathrm{d}y;\, J_{\overline{\sigma}} = k) = \boldsymbol{e}_i^\top \mathrm{e}^{\boldsymbol{U}y} \boldsymbol{e}_k \,\mathrm{d}y$ where:*

*$\boldsymbol{U} = \boldsymbol{H}\boldsymbol{\Delta}_{-\boldsymbol{s}}\boldsymbol{H}^{-1}$;*
*$\boldsymbol{s}$ is the vector with elements $s_j = -\mu/\sigma^2 + \sqrt{\mu^2/\sigma^4 + 2\lambda_j/\sigma^2}$;*
*$\boldsymbol{H}$ is the matrix with columns $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n$;*
*$\boldsymbol{h}_j$ is the eigenvector of $\boldsymbol{T} + \lambda_j \boldsymbol{I}$ corresponding to eigenvalue zero, with elements $h_{ij}$ given by $h_{ij} = 0$ for $i > j$, $h_{1j} = 1$ and:*

$$h_{ij} = \prod_{k=1}^{i-1} \frac{\lambda_k - \lambda_j}{\lambda_k q_k} \quad \text{for } 1 < i \le j.$$

*In particular, $\mathbb{P}(J_{\overline{\sigma}} = k) = -\boldsymbol{\alpha}\boldsymbol{U}^{-1}\boldsymbol{e}_k \cdot u_k$ where $u_k$ is the $k^{th}$ element of $-\boldsymbol{U}\boldsymbol{e}$.*

**Remark A1.** For a GC distribution, the characteristic polynomial, its roots, and the eigenvalues are explicit because of the upper triangular form of the generator matrix. This may of course be a curiosity in the computer age, and it is probably faster to code the program of Proposition A1 via matrix software than to use the explicit formulas. However, continuing in the same vein, note that not only is the

determinant of an upper triangular matrix explicit, but also the inverse. Indeed, let $\boldsymbol{k}$ be the vector of diagonal elements of $\boldsymbol{H}$ and write $\boldsymbol{H} = \boldsymbol{\Delta}_{\boldsymbol{k}}(\boldsymbol{I} + \boldsymbol{N})$, i.e., $\boldsymbol{N} = \boldsymbol{\Delta}_{1/\boldsymbol{k}}(\boldsymbol{H} - \boldsymbol{I})$. Then, $\boldsymbol{N}$ is nilpotent and:

$$\boldsymbol{H}^{-1} \;=\; (\boldsymbol{I} - \boldsymbol{N})^{-1}\boldsymbol{\Delta}_{1/\boldsymbol{k}} \;=\; \sum_{k=1}^{n-1}(-1)^k \boldsymbol{N}^k \boldsymbol{\Delta}_{1/\boldsymbol{k}}\,.$$

The iterative algorithms are hardly simpler for the GC case than for a general PH structure. An example (Theorem 4.1 of (Asmussen 1995)) follows.

**Proposition A2.** *Assume $X$ is $BM(\mu, \sigma^2)$ and that $\tau$ is $PH(\boldsymbol{\alpha}, \boldsymbol{T})$. Define $\lambda_i = -T_{ii}$,*

$$\omega_i = -\mu/\sigma^2 + \sqrt{\mu^2/\sigma^4 + 2\lambda_j/\sigma^2}\,, \qquad \eta_i = \mu/\sigma^2 + \sqrt{\mu^2/\sigma^4 + 2\lambda_j/\sigma^2}\,.$$

*Then, $\boldsymbol{U}$ is the solution of the fixed-point problem $\boldsymbol{U} = \psi(\boldsymbol{U})$ where $\psi(\boldsymbol{V})$ is the matrix whose $i^{th}$ row is the $i^{th}$ row of the matrix:*

$$\frac{2}{\sigma^2}(\boldsymbol{\Delta}_{\boldsymbol{\lambda}} + \boldsymbol{T})(\eta_i \boldsymbol{I} - \boldsymbol{V})^{-1} - \boldsymbol{\Delta}_{\boldsymbol{\omega}}\,.$$

*Further, $\boldsymbol{U} = \lim_{n\to\infty} \boldsymbol{U}_n$ with $\boldsymbol{U}_n = \psi(\boldsymbol{U}_{n-1})$ and $\boldsymbol{U}_0 = -\boldsymbol{\Delta}_{\boldsymbol{\omega}}$.*

In our numerical examples, we have used an algorithm that is somewhat related, but also contains some genuinely new ideas. It is applicable to more general models than jump diffusions and thereby also in other application areas than life insurance. For this reason, the details will be provided elsewhere, and we give here only a very brief outline. The basic idea is illustrated in Figure A3, which contains a realization of the jump diffusion $X$ up to the PH time $\tau$. The underlying Markov process has two states, blue B and red R for $\tau$, two cyan C and magenta M for the upward jumps, and one green G for the downward ones. The phases when $X$ is at a maximum are shown in the left-hand vertical bar and have four possible states B, R, CB, MB, CR, MR with legends given in the upper right corner.

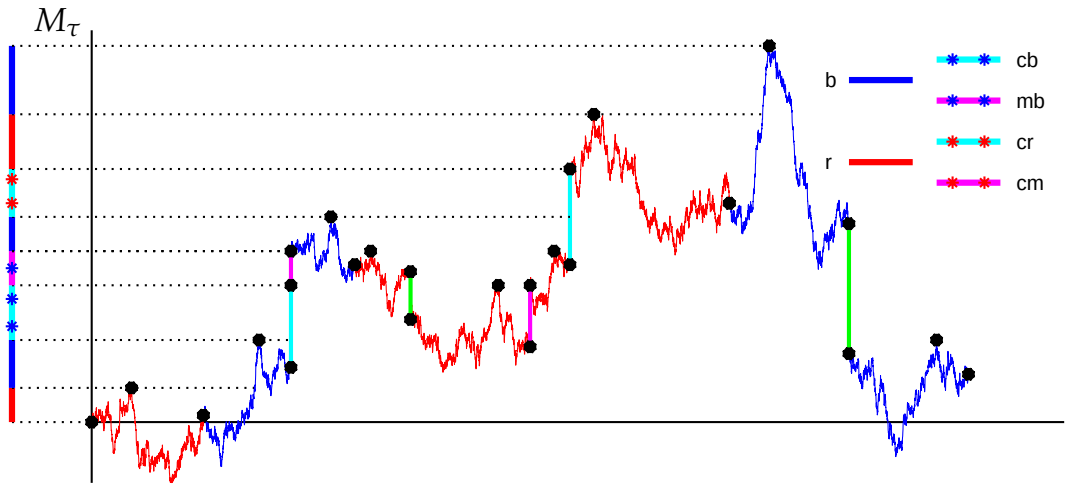



**Figure A3.** Jump diffusion and embedded Markov random walk.

The basic observation for the algorithm is that this process $\xi(x)$ of phases at maxima $x$ (and thereby, the overall maximum $M_\tau$) is the same as for the embedded discrete time killed Markov random walk marked by black dots. Its values are the values in the graph of $X$ either at the start of a Brownian segment, just before the end, or at the maximum, or at phase changes in jumps. Here, the graph of $X$ is understood to comprise the vertical parts from jumps, and Brownian segments are separated either by state changes in the Markov process underlying the phase horizon $\tau$, or by jumps. The generator $\boldsymbol{U}$ of $\xi$ is then shown to satisfy a certain fixed-point equation that is solved iteratively.

Computationally, the main effort in each step is to invert a matrix $\boldsymbol{U}_n$ whose dimension is the number $p(1+p^+)$ of states of $\xi$.

## References


Asmussen, Søren. 1995. Stationary distributions for fluid flow models with or without brownian noise. *Stochastic Models* 11: 21–49. [CrossRef]

Asmussen, Søren. 2003. Applied Probability and Queues. In *Stochastic Modelling and Applied Probability Series*, 2nd ed. Berlin: Springer, vol. 51.

Asmussen, Søren, and Hansjörg Albrecher. 2010. Ruin Probabilities. In *Advanced Series on Statistical Science and Applied Probability*, 2nd ed. Singapore: World Scientific, vol. 14.

Asmussen, Søren, Florin Avram, and Martijn R. Pistorius. 2004. Russian and american put options under exponential phase-type lévy models. *Stochastic Processes and Their Applications* 109: 79–111. [CrossRef]

Asmussen, Søren, Florin Avram, and Miguel Usabel. 2002. Erlangian approximations for finite-horizon ruin probabilities. *ASTIN Bulletin* 32: 267–81. [CrossRef]

Asmussen, Søren, and Jevgenijs Ivanovs. 2019. A factorization of a Lévy process over a phase-type horizon. *Stochastic Models*. accepted. [CrossRef]

Asmussen, Søren, Patrick J. Laub, and Hailiang Yang. 2019. Online accompaniment for "Phase-Type Models in Life Insurance: Fitting and Valuation of Equity-Linked Benefits". Available online: https://github.com/Pat-Laub/PhaseTypeLifeInsurance (accessed on 4 February 2019).

Asmussen, Søren, Olle Nerman, and Marita Olsson. 1996. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics* 23: 419–41.

Bean, Nigel G., Małgorzata M. O'Reilly, and Peter G. Taylor. 2005. Algorithms for return probabilities for stochastic fluid flows. *Stochastic Models* 21: 149–84. [CrossRef]

Bladt, Mogens. 2005. A review on phase-type distributions and their use in risk theory. *ASTIN Bulletin* 35: 145–67. [CrossRef]

Bladt, Mogens, and Bo Friis Nielsen. 2017. *Matrix-Exponential Distributions in Applied Probability*. Berlin: Springer.

Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt. 1997. *Actuarial Mathematics*, 2nd ed. Schaumburg: The Society of Actuaries.

Breuer, Lothar. 2008. First passage times for Markov additive processes with positive jumps of phase type. *Journal of Applied Probability* 45: 779–99. [CrossRef]

Cai, Jun, and Haijun Li. 2005a. Conditional tail expectations for multivariate phase type distributions. *Journal of Applied Probability* 42: 810–25. [CrossRef]

Cai, Jun, and Haijun Li. 2005b. Multivariate risk model of phase type. *Insurance: Mathematics and Economics* 36: 137–52. [CrossRef]

Carr, Peter. 1998. Randomization and the american put. *The Review of Financial Studies* 11: 597–626. [CrossRef]

Cont, Rama, and Peter Tankov. 2004. *Financial Modelling with Jump Processes*. London: Chapman & Hall, CRC.

Das, Sanjiv Ranjan, and Silverio Foresi. 1996. Exact solutions for bond and option prices with systematic jump risk. *Review of Derivatives Research* 1: 7–24. [CrossRef]

Dieker, Antonius B., and Michel Mandjes. 2011. Extremes of markov-additive processes with one-sided jumps, with queueing applications. *Methodology and Computing in Applied Probability* 13: 221–67. [CrossRef]

Egami, Masahiko, and Kazutoshi Yamazaki. 2014. Phase-type fitting of scale functions for spectrally negative Lévy processes. *Journal of Computational and Applied Mathematics* 264: 1–22. [CrossRef]

Gerber, Hans U., Elias S. W. Shiu, and Hailiang Yang. 2012. Valuing equity-linked death benefits and other contingent options: A discounted density approach. *Insurance: Mathematics and Economics* 51: 73–92. [CrossRef]

Gerber, Hans U., Elias S. W. Shiu, and Hailiang Yang. 2013. Valuing equity-linked death benefits in jump diffusion models. *Insurance: Mathematics and Economics* 53: 615–23. [CrossRef]

Gerber, Hans U., Elias S. W. Shiu, and Hailiang Yang. 2015. Geometric stopping of a random walk and its application to valuation of equity-linked death benefits. *Insurance: Mathematics and Economics* 64: 313–25.

Horváth, Gábor, and Miklós Telek. 2017. Matrix-analytic solutions of infinite, finite and level-dependent second-order fluid queues. *Queueing Systems: Theory and Applications* 87: 325–43. [CrossRef]

Jacobsen, Martin. 2005. The time to ruin for a class of markov additive risk process with two-sided jumps. *Advances in Applied Probability* 37: 963–92. [CrossRef]

Jiang, Zhengjun, and Martijn R. Pistorius. 2008. On perpetual american put valuation and first-passage in a regime-switching model with jumps. *Finance and Stochastics* 12: 331–55. [CrossRef]

Kou, Steven G. 2002. A jump diffusion model for option pricing. *Management Science* 48: 1086–101. [CrossRef]

Kuznetsov, Alexey, Andreas E. Kyprianou, Juan C. Pardo, and Kees van Schaik. 2011. A Wiener–Hopf Monte Carlo simulation technique for Lévy processes. *The Annals of Applied Probability* 21: 2171–90. [CrossRef]

Leung, Tim, Kazutoshi Yamazaki, and Hongzhong Zhang. 2015. An analytic recursive method for optimal multiple stopping: Canadization and phase-type fitting. *International Journal of Theoretical and Applied Finance* 18: 31. [CrossRef]

Lewis, Alan L., and Ernesto Mordecki. 2008. Wiener-Hopf factorization for Lévy processes having positive jumps with rational transforms. *Journal of Applied Probability* 45: 118–34. [CrossRef]

Lin, X. Sheldon, and Xiaoming Liu. 2007. Markov aging process and phase-type law of mortality. *North American Actuarial Journal* 11: 92–109. [CrossRef]

Mordecki, Ernesto. 2002. The distribution of the maximum of a Lévy process with positive jumps of phase-type. *Theory of Stochastic Processes* 8: 309–16.

Okamura, Hiroyuki, Tadashi Dohi, and Kishor S. Trivedi. 2011a. Improvement of expectation-maximization algorithm for phase-type distributions with grouped and truncated data. *Applied Stochastic Models in Business and Industry* 29: 141–56. [CrossRef]

Okamura, Hiroyuki, Tadashi Dohi, and Kishor S. Trivedi. 2011b. A refined EM algorithm for PH distributions. *Performance Evaluation* 68: 938–54. [CrossRef]

Olsson, Marita. 1996. Estimation of phase-type distributions from censored data. *Scandinavian Journal of Statistics* 23: 443–60.

Olsson, Marita. 1998. The Empht-Programme. Available online: http://home.math.au.dk/asmus/dl/EMusersguide.ps (accessed on 4 February 2019).

Pistorius, Martijn R. 2006. On maxima and ladder processes for a dense class of Lévy process. *Journal of Applied Probability* 43: 208–20. [CrossRef]

Siu, Chi Chung, Sheung Chi Yam, and Hailiang Yang. 2015. Valuing equity-linked deaths benefits in a regime-switching framework. *ASTIN Bulletin* 45: 355–95. [CrossRef]

Stanford, David A., Guy Latouche, Douglas G. Woolford, Dennis Boychuk, and Alek Hunchak. 2005. Erlangized fluid queues with application to uncontrolled fire perimeter. *Stochastic Models* 21: 631–42. [CrossRef]

Stanford, David A., Krzysztof J. Stroiński, and Karen Lee. 2000. Ruin probabilities based at claim instants for some non-Poisson claim processes. *Insurance: Mathematics and Economics* 26: 251–67. [CrossRef]

Zadeh, Amin Hassan, Bruce L. Jones, and David A. Stanford. 2014. The use of phase-type models for disability insurance calculations. *Scandinavian Actuarial Journal* 8: 714–28. [CrossRef]

Zadeh, Amin Hassan, and David A. Stanford. 2016. Bayesian and Bühlmann credibility for phase-type distributions with a univariate risk parameter. *Scandinavian Actuarial Journal* 2016: 338–55. [CrossRef]