HSC Research Report

HSC/06/02

Financial engineering methods in insurance

Jan Iwanik*

* Institute of Mathematics and Computer Science, Wrocław University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

.

Metody inżynierii finansowej w ubezpieczeniach

Jan Iwanik

Rozprawa doktorska, w.1.1

promotor: prof. dr hab. Aleksander Weron

Instytut Matematyki i Informatyki, Politechnika Wrocławska

Maj 2006

.

Financial Engineering Methods in Insurance

by

Jan Iwanik

PhD Thesis, v.1.1

supervisor: prof. dr hab. Aleksander Weron

Institute of Mathematics and Computer Science, Wrocław University of Technology

May 2006

Contents

Chapter 1

Introduction

1.1 Scope of this Paper

The aim of this paper is to apply specific statistical tools known and used in finance and risk management to the area of actuarial mathematics. The need for an interdisciplinary approach in both actuarial world and risk management has emerged and has recently been addressed by numerous publications as well as in scientific and professional events and meetings within the actuarial world. This approach is a must in a sophisticated market with complex financial instruments. Examples of such an approach include equity-linked life insurance contracts, options on mortality, and attempts to implement methodologies like Risk Adjusted Return on Capital as a principal pricing rule by more and more insurance companies.

The trend toward reapplying financial stochastic methods in insurance can also be seen in the theoretical field. There is an ongoing effort in finance and in actuarial science to learn and integrate the statistical and mathematical tools used by the two traditional streams into a single, commonly applicable, toolbox.

In this paper I want to explore two such paths. The first is the concept of failure probability that can be used as a base model for future returns in the insurance line of business. The second is an attempt to use option pricing techniques to hedge a portfolio of life insurance contracts against systematic mortality risk.

The probability of failure concept is introduced in Chapter 2. The motivation for studying this probability is given and some theoretical results are derived – a generalized version of Prabhu's equation is given. Queuing theory methods can be used for computing this probability. How this can be done is demonstrated in Chapter 3.

The probability of failure for discrete claim distributions is explored in Chapter 4. Some popular methods of exact computation of ruin probability for discrete claims are adopted to compute failure probability. Based on the formula presented in [21] a generalization of a ruin probability algorithm is proposed that can also be used for failure probability. The algorithm's computational complexity is studied and it is proved to be more effective for failure probability than for ruin probability. Finally, some numerical examples for failure probability computations are given. Chapter 4 reflects the author's earlier publication [22].

A different approach is taken in the later chapters. Life insurance is an interesting field where the methods of risk control known from advanced finance may be applied. Chapter 5 considers the problem of the random mortality intensity (also called 'hazard rate', 'force of mortality', 'mortality rate'). I examine how a discretized version of the stochastic mortality model proposed by Dahl in [10] fits the historical demographic data for some developed countries. In addition, other stochastic mortality models are proposed and are tested against the available data. Both one-dimensional and multi-dimensional statistical analyses are performed and a comparison of the models is provided that shows that the best mortality model is not necessarily the one proposed by Dahl. Analytical formulas for survival probability in the new models are given. These formulas are analogous to the ones known from interest rate modeling. Finally, a few pricing and reserving applications are given.

The idea of mortality options mentioned in Chapter 5 is further developed in Chapter 6. Two basic types of mortality-linked financial instruments are introduced and the way they should be used to protect against the systematic mortality risk is proposed. I also give explicit formulas, perform numerical simulations and propose two approximations that are modifications of methods known from the Asian options' pricing. The exactness of those approximations is examined and their usefulness is considered.

1.2 Symbols and Notation

The most important symbols used in this paper:

c	constant premium rate,
$\mathcal{C}(s)$	price of a call mortality option at time s, introduced in Section 6.2.2
$F(x)$	cumulative probability function of a single claim X,
$F^{*n}(x)$	cumulative probability function of a sum of n independent claims (n-fold convolution),
$\psi(u)$	probability of ruin with initial capital u, defined by (2.3),
$\psi(u,T)$	probability of ruin in time T with initial capital u, defined by (2.4),
$\psi(u,T,w)$	probability of failure in time T with initial capital u, see Definition 2.1.2,
$\phi(u)$	probability of survival, i.e. $1-\psi(u)$,
$\phi(u,T)$	probability of survival in time T, i.e. $1-\psi(u,T)$,
$\phi(u,T,w)$	probability of success in time T, i.e. $1-\psi(u,T,w)$,
θ	relative safety loading, i.e. $\frac{c}{\lambda\mu_X}-1$,
λ	intensity of the Poisson claim process,
μ_X	expected value of a claim, i.e. $E(X)$,
μ_t	also (in different context) mortality intensity function,
$N(t)$	number of claims occurred during $[0,t]$,
$M_X(t)$	moment generating function of X,
$\mathcal{P}(s)$	price of a put mortality option at time s, introduced in Section 6.2.2
${}_{T-t}p_t$	standard actuarial symbol for survival probability, see 6.1,
$R_u(t)$	risk process, see Definition 2.1.1,
$S(t)$	cumulated claim occurred in $[0,t]$, i.e. $\sum_{i=0}^{N(t)} X_i$,
$\mathcal{S}(s)$	time-dependent expected value of a mortality process, see (6.2.2),
Ω	set of the elementary events,
σ_i	moment when the ith claim occurs,
τ	moment of ruin, see (2.2),
τ_i	waiting time between the $(i-1)$th claim and the ith claim,
u	initial capital of an insurance company,
$V_w(t)$	claim surplus process, see Definition 3.1.1,
$W_w(t)$	virtual waiting time process, see Definition 3.1.3,
X_i	size of the ith claim, $i=1,2,3,\ldots$,
X	a random variable with the same distribution as X_1.

Chapter 2

Failure Probability – Preliminaries

2.1 Definitions

Let u be the initial capital of an insurance company, c represent the premium rate and $N(t)$ be a Poisson random variable with mean $t\lambda$. Let $N(t), X_1, X_2, \ldots$ be independent and $X_1, X_2, \ldots$ be identically distributed positive insurance claims.

Definition 2.1.1 *The standard risk process*

$$R_u(t) = u + ct - \sum_{i=1}^{N(t)} X_i \ . \tag{2.1}$$

The sum $\sum_{i=1}^{N(t)} X_i$ will be also denoted by $S(t)$.

One of the fundamental problems in both theoretical and practical approaches in actuarial literature (e.g. [2], [12], [7]) is the problem of the time of ruin of the company whose capital is described by the risk process $R_u(t)$. Let the time of ruin be denoted by

$$\tau = \begin{cases} \inf\{t \ : \ t > 0 \ \wedge \ R_u(t) < 0\} & \text{if the set is not empty,} \\ +\infty & \text{otherwise.} \end{cases} \tag{2.2}$$

Ruin probability in infinite time is defined by

$$\psi(u) = P(\tau < \infty) \tag{2.3}$$

and ruin probability in the finite time horizon $[0, T]$ is defined by

$$\psi(u, T) = P(\tau \leq T) \ . \tag{2.4}$$

Although the ruin probability problem plays a central role in insurance mathematics, another problem can be of equal practical importance for an insurance company. The company usually wishes not only to survive the next year, but also expects a reasonable rate of return. Let us define the probability of failure originally introduced in [22].

Definition 2.1.2 *For* $0 \le w$ *the probability of failure in time* T

$$\psi(u,T,w) = P\big(\tau < T \ \vee \ (T \le \tau \ \wedge \ R_u(T) < w)\big) \ . \tag{2.5}$$

For convenience we will sometimes denote the probability of non-failure (success) by

$$\phi(u,T,w) = 1 - \psi(u,T,w) \ . \tag{2.6}$$

The probability of failure can be also viewed as a natural generalization of the probability of ruin in finite time since

$$\psi(u,T) = \psi(u,T,-\infty) \ . \tag{2.7}$$

Note that – unlike ruin probability – the concept of failure probability makes sense only in the finite time case.

2.2 Motivation

The reason why the problem of failure probability is worth considering can be illustrated by a question asked by the investor with initial capital u: *what is the probability that the company will not go bankrupt and will bring interest not smaller than the risk-free financial instruments during time* T*?* This probability can be mathematically expressed as $1 - \psi(u,T,(1+i)^T u)$, where i is the risk-free interest rate. Figure 2.1 illustrates the investor's dilemma.

The popularity of the whole family of the financial '- at Risk' measures like Value at Risk (VaR), Capital at Risk (CaR), Earnings at Risk (EaR) or Cashflow at Risk (CfaR) follows from the fact that they all give some insight into the probability distribution of a future value of some investment. The knowledge of this distribution is essential for any decision maker, since it allows the investor to assess an investment or a line of business in a risk adjusted way. Wide implementation of financial measures like Risk Adjusted Return on Capital (RAROC) and other members of the 'Risk Adjusted Return' family, is another piece of evidence that the quantitative probabilistic methods in risk management are being appreciated by the business community.

Insurance is no different. The subsequent lines of insurance business must be monitored and the return on capital must meet some reasonable assumptions. If they do not meet – the investor may want to raise the insurance premiums or decide to quit from insurance in order to utilize the capital in a more efficient way. To make such decisions, however, the probabilistic distribution of future profits from engaging in insurance must be known.

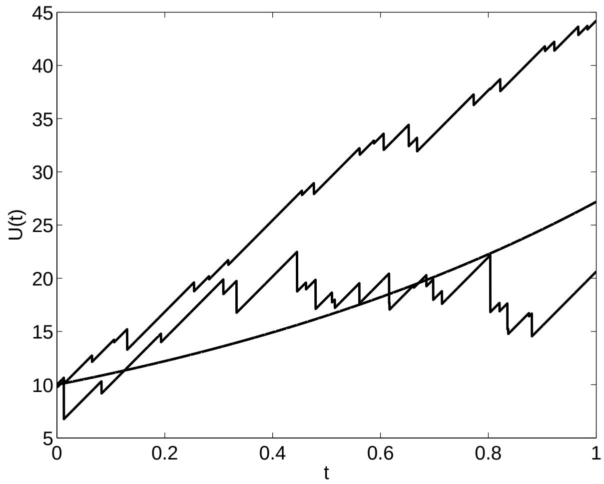


Figure 2.1: None of the above realizations of the risk process causes a ruin, however one of them causes the failure for $w = 27$ (or $i = 1.7$).

Failure probability gives the exact distribution of such profits. It is a natural generalization of the ruin probability in a finite time horizon. In many practical cases, however, it is more important to know the failure probability than to be able to determine ruin probability only.

2.3 Prabhu's Formula for $\psi(0, T, w)$

Failure probability is not only an useful concept, but also has some nice analytical features. In many cases it can be computed using methods known from ruin theory. This section provides an example of such an approach.

One of the few cases where an explicit formula for finite-time ruin probability is known is the case with zero initial capital i.e. $u = 0$. In this situation the Prabhu's or Cramér's formula holds:

$$1 - \psi(0, T) = \frac{1}{cT} \int_0^{cT} F_T'(s) ds \tag{2.8}$$

where

$$F_T'(s) = \sum_{i=0}^{\infty} P\big(N(T) = i\big) F^{*i}(s) \ , \tag{2.9}$$

see [12], Ch. I.6 or [2], Ch. IV.2 for proofs. Here F^{*i} is the i-fold convolution of the claim distribution. Note that this equation constitutes a direct link between ruin probability and the distribution of the cumulative claim $S(T)$.

In this section, we will prove a similar result for failure probability with zero initial capital. Namely

Theorem 2.3.1

$$1-\psi(0,T,w)=\frac{w}{cT}F'_T(cT-w)+\frac{1}{cT}\int_0^{cT-w}F'_T(s)ds \tag{2.10}$$

where $F'_T(s)$ is defined as in (2.9).

Proof The proof is to a large extent identical to the proof of (2.8) that can be found in [12], Ch. I.6. Similar constructs were also used by other authors and some of them are further researched in Section 4.1.1 of this paper. The proof uses the trick of replacing the interior integral with an equivalent polynomial. This trick was also proposed in [12].

Let the moment of the ith claim be denoted by σ_i. Further, let for $\sigma_i \le t$ the symbol $\pi_{t,n}=\pi_{t,n}(t_1,t_2,\ldots,t_n)$ denote the multivariate probability function for the first n claims. Note that this function is a constant with respect to $t_1,\ldots,t_n$. The formal definition is

$$\begin{aligned}\pi_{t,n} &= \pi_{t,n}(t_1,t_2,\ldots,t_n)\\ &= P\big(N(t)=n \ \wedge\ \sigma_1\in dt_1 \ \wedge\ \sigma_2\in dt_2 \ \wedge\ \ldots\ \wedge\ \sigma_n\in dt_n\big)\ .\end{aligned} \tag{2.11}$$

The first of the equations below follows from the definition of $\pi_{t,n}$ and the other one can be easily derived in a recursive way.

$$\begin{aligned}P\big(N(T)=n\big) &= \int\ldots\int_{0\le t_1\le\ldots\le t_n\le T}\pi_{t,n}dt_1\ldots dt_n\\ &= \pi_{T,n}\frac{T^n}{n!}\ .\end{aligned} \tag{2.12}$$

Now, let's assume that $N(T)=n$. To avoid failure, the following set of conditions is necessary and sufficient. Condition A is a constrain on the claim severities

$$x_1+x_2+\ldots+x_n\le cT-w\ . \tag{2.13}$$

Note that this condition – although very similar – is not identical with the one given in [12]. The other condition, denoted by B, restricts the moment of claims, given their severities.

$$\left\{\begin{array}{lcccc}(x_1+x_2+\ldots+x_n)/c & \le & t_n & \le & T\\ (x_1+x_2+\ldots+x_{n-1})/c & \le & t_{n-1} & \le & t_n\\ \ldots & & \ldots & & \ldots\\ x_1/c & \le & t_1 & \le & t_2\ .\end{array}\right. \tag{2.14}$$

We will use the notation

$$L_n(T) = P(N(T) = n \text{ and no failure occurs in time } T) \ . \tag{2.15}$$

We will derive $L_n(T)$ by integrating $\pi_{T,n}$ with respect to conditions A and B. We have

$$\begin{aligned} L_n(T) &= \int \dots \int_A \Big(\int \dots \int_B \pi_{T,n} dt_n \dots dt_1 \Big) dF(x_1) \dots dF(x_n) \\ &= \pi_{T,n} \int \dots \int_A \underbrace{\Big(\int \dots \int_B dt_n \dots dt_1 \Big)}_{inner\ integral\ is\ a\ function\ of\ x_1,\dots,x_n} dF(x_1) \dots dF(x_n) \ . \end{aligned} \tag{2.16}$$

If we use the notation $s_n = \frac{x_1+\dots+x_n}{n}$, we can see that the inner integral is a polynomial in $s_1, s_2, \dots, s_n$ and hence in $x_1, x_2, \dots, x_n$. For example if $n = 2$ we have:

$$\int \int_B dt_2 dt_1 = \int_{s_2 \le t_2 \le T} \int_{s_1 \le t_1 \le t_2} dt_2 dt_1 \ . \tag{2.17}$$

We will call two polynomials $p(x_1, \dots, x_n)$ and $q(x_1, \dots, x_n)$ equivalent if

$$\int \dots \int_A p(x_1, \dots, x_n) dx_1 \dots dx_n = \int \dots \int_A q(x_1, \dots, x_n) dx_1 \dots dx_n \ . \tag{2.18}$$

Three general facts apply and since they are identical as in [12], we will leave them without proof.

1. If $(i_1, \dots, i_n)$ is a permutation of $(1, \dots, n)$ then there is an equivalence between polynomials $p(x_1, \dots, x_n)$ and $p(x_{i_1}, \dots, x_{i_n})$.
2. From the above, it follows that $(k+1)s_k p(x_1, \dots, x_n)$ is equivalent to $k s_{k+1} p(x_1, \dots, x_n)$ if only $k + 1 \le n$ and $p(x_1, \dots, x_n)$ is symmetrical in $x_1, \dots, x_n$.
3. Applying the above in a recursive way, we get that the inner integral is equivalent to $\frac{T^n(1-s_n/T)}{n!}$.

From the last fact combined with (2.16) and (2.12) we have

$$L_n(T) = P\big(N(T) = n\big) \int \dots \int_A (1 - \frac{s_n}{T}) dF(x_1) \dots dF(x_n) \tag{2.19}$$

and by substitution $s = cs_n$ and integrating by parts, we have

$$\begin{aligned} L_n(T) &= P\big(N(T) = n\big) \int_0^{cT-w} (1 - \frac{s}{cT}) dF^{*n}(s) \\ &= P\big(N(T) = n\big) \Big(\frac{w}{cT} F^{*n}(cT - w) + \frac{1}{cT} \int_0^{cT-w} F^{*n}(s) ds \Big) \ . \end{aligned} \tag{2.20}$$

Finally, $1 - \psi(0, T, w)$ is a sum of (2.20) over n. $\square$

2.4 Generalization for $\psi(u, T, w)$

The formula from the previous section may be now generalized in the following way.

Theorem 2.4.1 *Let $F'_t(s)$ defined by (2.9) be differentiable and for a given t the density $f'_t(x) = \frac{dF'_t(x)}{dx}$. Then we have*

$$1 - \psi(u, T, w) = F'_T(u + cT - w) - \int_0^T f'_s(u + cs)\big(1 - \psi(0, T - s, w)\big)ds \ . \tag{2.21}$$

Proof The event $\{R_u(T) > w\}$ may occur in two mutually exclusive ways. Either success (non-failure) occurs or $R_u(T) > w$ but the risk process reached zero somewhere between 0 and T – in this case ruin occurs. In this case there is a moment s such that $R_u(s) = 0$ for the last time. If so, then $S(s) = u + cs$.

Probability of the event $\{R_u(T) > w\}$ is $F'_T(u + cT - w)$. The integral represents the ruin and recovery and the left-hand side of equation (2.21) is the probability of survival. □

Chapter 3

Queuing Theory Methods

3.1 Important Processes Related to the Risk Process

Some processes dual to the risk process play an important role in ruin theory. They allow us to express the ruin problem in the queuing theory language. A similar duality exist for failure probability. This section defines two important processes: the claim surplus process and the virtual waiting time process.

3.1.1 Claim Surplus Process

Let us consider the collective risk process introduced by Definition 2.1.1. Let the jump moments of this process be denoted by $\sigma_1, \sigma_2, \ldots$. We define the storage process $V_w(t)$ that, to some extent, is a mirror reflection of $R_u(t)$. $V_w(t)$ has positive jumps of size X_{N-i} at moments $\sigma_i^* = T - \sigma_i$. The 'premium' is negative for $V_w(t)$, but only as long as the process itself is positive, otherwise the premium becomes zero. Formally, if

$$\delta(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{otherwise,} \end{cases} \tag{3.1}$$

then the storage process is given by the following

Definition 3.1.1 *The storage process dual to the risk process is the solution of*

$$dV_w(t) = d(S(T) - S(T-t)) - dt\ c\delta(V_w(t)) \tag{3.2}$$

with the initial condition $V_w(0) = w$.

The storage process for $w = 0$ plays an important role in ruin theory. There is a celebrated duality between this process and the risk process. If we consider finite time T then the following is true.

Theorem 3.1.2 *If the initial capital is u then the events $\{\tau \leq T\}$ and $\{V_0(T) > u\}$ coincide.*

Proof See [2], Theorem 3.1. □

Hence, an instant corollary is hence that

$$\psi(u, T) = P(V_0(T) > u) \ . \tag{3.3}$$

3.1.2 Virtual Waiting Time Process

We also define the virtual waiting time process $W_w(t)$. In the queuing theory world, it represents the time a new client arriving at the system must wait before his service starts. The formal definition is given by

Definition 3.1.3 *The virtual waiting time process dual to the risk process is the solution of*

$$dW_w(t) = dS(t) - dt\ c\delta(W_w(t)) \tag{3.4}$$

with initial condition $W_w(0) = w$.

From Definition 3.1.3 follows that

$$dW_w(t) = -dR_{-w}(t) + dt\ c\big(1 - \delta(W_w(t))\big) \tag{3.5}$$

and finally, since $W_w(0) = -R_{-w}(0)$, the integral equivalent to Definition 3.1.3 is

$$W_w(t) = -R_{-w}(t) + c\int_0^t 1 - \delta(W_w(s))ds. \tag{3.6}$$

Note that the risk process $R_u(t)$ and the virtual waiting time process $W_w(t)$ are right-continuous, while the storage process $V_w(t)$ is left-continuous.

Figure 3.1 presents a sample path of risk process $R_u(t)$ and the corresponding paths of $V_w(t)$ and $W_w(t)$ for some given u, w and $\omega \in \Omega$.

3.2 Failure Probability and Queuing Theory

Because the moments of jumps in the risk process are governed by the homogeneous Poisson process and the size of the jumps are i.i.d., the following Lemma holds. We will leave it without formal proof.

Lemma 3.2.1 *For any t and x we have $P(V_w(t) < x) = P(W_w(t) < x)$.*

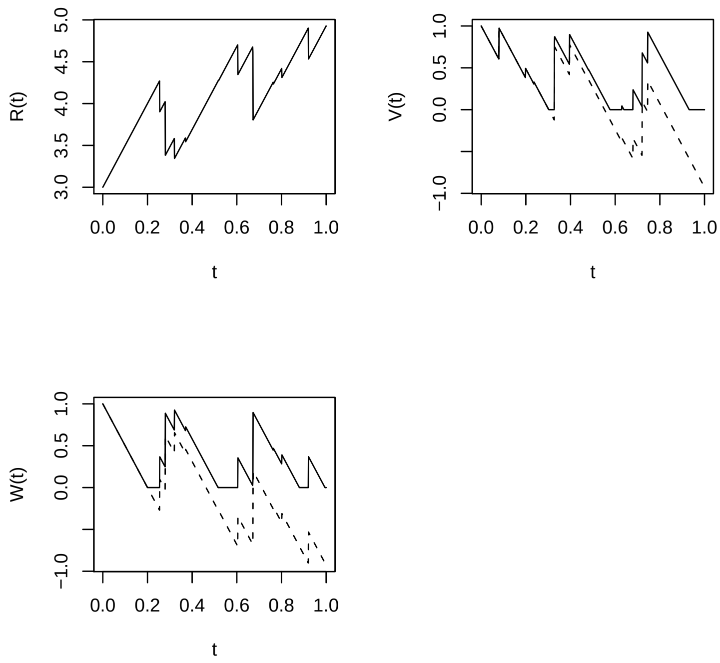


Figure 3.1: Solid lines represent sample paths of $R_u(t)$, $V_w(t)$ and $W_w(t)$ for $u = 3$, $w = 1$.

The following Theorem constitutes a link between failure probability and the behavior of the storage process. The technique used in the proof is similar to the one used in [2] in the Lemma above.

Theorem 3.2.2 *If the initial capital is u then the event of failure $\{\tau \leq T \ \vee \ R_u(T) < w\}$ and the event $\{V_w(T) > u\}$ coincide. Hence $\psi(u, T, w) = P(V_w(T) > u)$.*

Proof It suffices to prove the case when $N > 0$, since the Theorem is obvious for $N = 0$.

First, let us prove that $\{V_w(T) > u\} \Rightarrow \{\tau \leq T \ \vee \ R_u(T) < w\}$. We define $\gamma = \sup\{\sigma_i^* : \ V_w(\sigma_i^*) = 0\}$. If such γ does not exist, then

$$R_u(T) - R_u(0) = V_w(0) - V_w(T) \tag{3.7}$$

and because $V_w(0) - V_w(T) < w - u$, this means that the failure occurs. If γ exists, we have

$$V_w(\sigma_N^*) = V_w(T) + c\sigma_1 - X_1 > u + c\sigma_1 - X_1 = R_u(\sigma_1)\ . \tag{3.8}$$

This formula can be repeated to get $V_w(\sigma^*_{N-1}) > R_u(\sigma_2), \ldots$. We can continue this process until we get $0 = V_w(\gamma) > R_u(T-\gamma)$ and hence the ruin occurs at $T-\gamma$.

Next, let us prove that $\{V_w(T) \le u\} \Rightarrow \{\tau > T \ \wedge \ R_u(T) \ge w\}$. We have

$$V_w(\sigma^*_N) \le V_w(T) + c\sigma_1 - X_1 \le u + c\sigma_1 - X_1 = R_u(\sigma_1) \ , \tag{3.9}$$

and we can repeat this relation for σ^*_{N-1}, σ^*_{N-2} and so on. Since $V_w(t) \ge 0$, the ruin does not occur until T. Moreover, from the definition of $V_w(t)$, for any $t \in [0,T]$ it is true that $V_w(t) \ge v + R_u(T-t) - R_u(T)$. So that for $t = T$

$$V_w(T) - w \ge u - R_u(T). \tag{3.10}$$

Finally, from (3.10) and the assumption $V_w(T) \le u$, we can conclude that indeed $R_u(T) \ge w$. □

3.3 Application – Constant Claim

Having the relationship between failure probability and queuing theory models formulated in Theorem 3.2.2, we can see that calculating failure probability is equivalent to finding the distribution of the unserved workload in an initially non-empty single server system. The claim arrive according to a Poisson process and the service time is deterministic (and constant). Hence, we are interested in a queue that is denoted by M/D/1 in the Kendal notation.

The same problem is also known as the finite time dam problem and an elementary discussion of this issue can be found e.g. in [30]. Let us have a closer look at this model. Here, the initial level of water is $V_w(0) = w$ and the dam accumulates water that arrives in blocks of constant, deterministic size h according to a homogeneous Poisson process with intensity λ. In this model, water departs from the dam with constant rate of c.

Note that if the dam had finite capacity N (or the queue was finite with length N i.e. a M/D/1/N queue) such that $N > (u + cT)/h$, the event $\{V_w(T) > u\}$ would still occur for exactly the same $\omega \in \Omega$ as if the dam was infinite. Due to this fact, we can use the result from [16] to give the exact analytical solution of the failure probability problem for deterministic claims. Below is the formalization of this concept. We will try to stick to the notation in [16].

Let $N = \lfloor u + cT \rfloor + 1$ be the length of a M/D/1/N queue. Let us introduce a $(N+1) \times$

$(N+1)$ matrix W:

$$W = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{3.11}$$

Theorem 3.3.1 *Let h be the deterministic claim severity, $\frac{w}{h} \in \mathbb{N}$, $N = \lfloor (u+cT)/h \rfloor + 1$, matrix W be defined as in (3.11) and π^0 be a $(N+1)$-dimensional vector that has 1 at the $(\frac{w}{h}+1)$ position and 0 elsewhere. Next, let $\alpha_k = \frac{(\lambda \frac{h}{c})^k}{k!} e^{-\lambda \frac{h}{c}}$ be the probability distribution for the number of claims in time $\frac{h}{c}$. Let the vector function $\Psi(s) = \big(d_0(s), d_1(s), \dots, d_{N-1}(s), 0\big)$ be defined by*

$$d_k(s) = \begin{cases} 0 \qquad \text{for } s < \frac{h}{c},\ k = 0, \dots, N-1, \\ \\ \lambda \alpha_k \pi_0(s - \frac{h}{c}) + \sum_{i=1}^{k+1} \alpha_{k+1-i} d_i(s - \frac{h}{c}) \\ \qquad \text{for } s \geq \frac{h}{c},\ k = 0, \dots, N-2, \\ \\ \lambda \big(1 - \sum_{i=0}^{N-2} \alpha_i\big) \pi_0(s - \frac{h}{c}) + \sum_{i=1}^{N-1} \big(1 - \sum_{j=0}^{N-1-i} \alpha_j\big) d_i(s - \frac{h}{c}) \\ \qquad \text{for } s \geq \frac{h}{c},\ k = N-1\ . \end{cases} \tag{3.12}$$

Then the probability distribution $\pi_x(T) = P(\lceil \frac{V_w(T)}{h} \rceil = x)$ for $T < \frac{w}{c}$ is given by the vector

$$\pi(T) = \pi(0) \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \dots & \alpha_{N-2} & 1 - \sum_{k=0}^{N-2} \alpha_k & 0 \\ \alpha_0 & \alpha_1 & \alpha_2 & \dots & \alpha_{N-2} & 1 - \sum_{k=0}^{N-2} \alpha_k & 0 \\ 0 & \alpha_0 & \alpha_1 & \dots & \alpha_{N-3} & 1 - \sum_{k=0}^{N-3} \alpha_k & 0 \\ 0 & 0 & \alpha_0 & \dots & \alpha_{N-4} & 1 - \sum_{k=0}^{N-4} \alpha_k & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_0 & 1 - \alpha_0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}^{\lfloor \frac{Tc}{h} \rfloor} e^{\lambda W (T - \lfloor \frac{Tc}{h} \rfloor \frac{h}{c})} \tag{3.13}$$

and for $T \geq \frac{w}{c}$ is given by the vector

$$\pi(T) = \pi(0) \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \dots & \alpha_{N-2} & 1 - \sum_{k=0}^{N-2} \alpha_k & 0 \\ \alpha_0 & \alpha_1 & \alpha_2 & \dots & \alpha_{N-2} & 1 - \sum_{k=0}^{N-2} \alpha_k & 0 \\ 0 & \alpha_0 & \alpha_1 & \dots & \alpha_{N-3} & 1 - \sum_{k=0}^{N-3} \alpha_k & 0 \\ 0 & 0 & \alpha_0 & \dots & \alpha_{N-4} & 1 - \sum_{k=0}^{N-4} \alpha_k & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_0 & 1 - \alpha_0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}^{\frac{w}{h}} e^{\lambda W (T - \frac{w}{c})} - \\ - \int_{\frac{w}{c}}^{T} \Psi(s) W e^{(T-s)\lambda W} ds \ . \tag{3.14}$$

Here the matrix exponential function is defined by

$$\exp(xW) = \sum_{i=0}^{\infty} \frac{x^i}{i!} W^i \tag{3.15}$$

and a simple form of this expression is derived by Garcia et. al. Hence we have

$$1 - \sum_{i=0}^{1+u/h} \pi_i(T) \leq \psi(u, T, w) \leq 1 - \sum_{i=0}^{u/h} \pi_i(T) \ . \tag{3.16}$$

Proof See Section 3.2 of [16] for proof in the queue context. The the north-west $N \times N$ sub-matrix of the large matrices used in (3.13) and (3.14) are defined incorrectly in [16]. The last formula (3.16) is due to the fact that

$$P(V_w(T) > u) \leq P(\lceil \frac{V_w(T)}{h} \rceil > \frac{u}{h}) \leq P(V_w(T) > u - h) \tag{3.17}$$

and

$$P(\lceil \frac{V_w(T)}{h} \rceil > \frac{u}{h}) = 1 - \sum_{i=0}^{u/h} \pi_i(T) \ . \tag{3.18}$$

□

Let us illustrate this theorem with an example. Let us examine the situation of an insurer whose claims are deterministic and constant. Such a setup is a good approximation for some lines of insurance business. A good example would be theft coverage for cars in a given price range. Let us assume that each car's price is 0.1, the initial capital is 10 and the intensity of the claim process is 10. So on average the aggregated loss in a given year is one. Relative safety loading θ was set to 0.8 i.e. $c = 1.8$. We will use a good approximation of failure

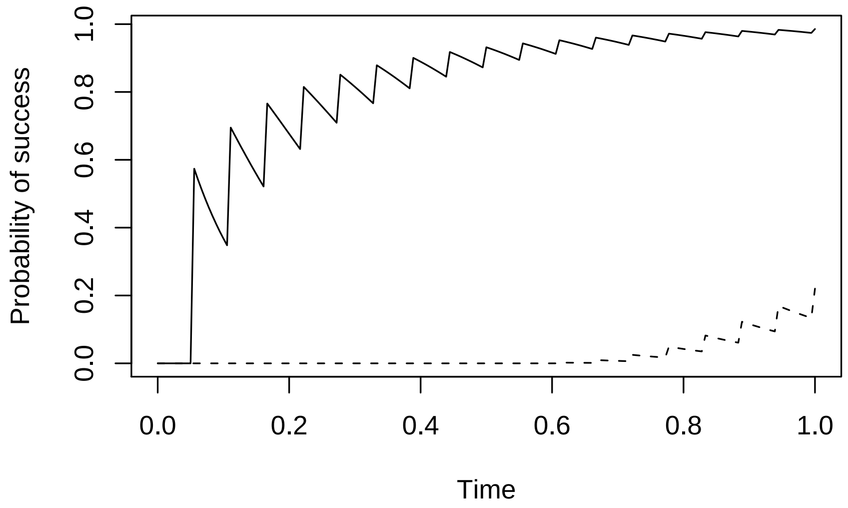


Figure 3.2: Probability of success of an insurer i.e. $1 - \psi(10, t, w)$ for $t \in [0, 1]$ where $w = 10$ (upper plot) and $w = 11$ (lower plot).

probability provided by the upper bound of (3.16). Figure 3.2 presents the situation of this insurer in the course of one year.

We can see that if the investor is satisfied with avoiding ruin and maintaining the initial capital only, he will achieve this goal almost certainly, especially in the long run. If, however the aim is to get a decent rate of return of 10%, the probability of success is merely about 1/5. Note that w (or equivalently the assumed level of return) has a great impact on the probability of failure. Even small changes in this rate will stronglly affect the probability of success.

The saw-shaped graph of $1 - \psi(10, t, w)$ might look odd at first glance. It is a consequence of the fact that success probability rapidly increases just after the cumulated capital exceeds a level that suffices to accept one more claim of size h. This effect tents to play a less important role as time passes.

Chapter 4

Probability of Failure with Discrete Claim Distribution

In this chapter, we consider failure probability with discrete claim distribution i.e. $P(X \in \mathbb{N}) = 1$. Every continuous claim distribution with probability density function $p(x)$ can be approximated by a discrete probability function $P(x)$. One way to do this is to put $P(X = n) = \int_n^{n+1} p(u)du$ for $n \in \mathbb{N}$, but of course there are many other approximation possibilities. Hence the restriction applied in this chapter is not a strong limitation. For convenience we will also assume that $u + cT \in \mathbb{N}$.

The chapter is organized as follows: in Section 4.1, generalizations of two ruin probability algorithms for discrete claims are presented that allow us to calculate failure probabilities. A brief study of computational complexity of one of them is provided. In Section 4.2, a similar generalization is proposed for the discrete time model. Finally, Section 4.3 contains two numerical examples of the application of probability of failure.

4.1 Continuous Time Models

4.1.1 Failure Probability Based on the Ignatov-Kaishev Method

An important approach to ruin probability was presented by Ignatov and Kaishev in [21]. Let $x = (x_1, x_2, \ldots)$ be subsequent discrete claims. Let the function $b_i(c_1, \ldots c_i)$ be defined as follows: $b_0 = 1$, $b_1 = c_1$ and

$$b_i(c_1, \ldots c_i) = \det \begin{pmatrix} \frac{c_1}{1!} & 1 & 0 & \ldots & 0 \\ \frac{c_2^2}{2!} & \frac{c_2}{1!} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{c_{i-1}^{i-1}}{(i-1)!} & \frac{c_{i-1}^{i-2}}{(i-2)!} & \frac{c_{i-1}^{i-3}}{(i-3)!} & \ldots & 1 \\ \frac{c_i^j}{i!} & \frac{c_i^{i-1}}{(i-1)!} & \frac{c_i^{i-2}}{(i-2)!} & \ldots & \frac{c_i}{1!} \end{pmatrix} . \tag{4.1}$$

Equation (33) from [21] states that if the vector of claims $\bar{x} = (x_i)_i$ is given, then the non-ruin probability is

$$\phi(u, T \mid \bar{X} = \bar{x}) = e^{-T} K_{\bar{x}} , \tag{4.2}$$

where

$$K_{\bar{x}} = \sum_{j=0}^{k_{\bar{x}}-1} (-1)^j b_j\left(\frac{x_1 - u}{c}, \ldots, \frac{x_1 + \ldots + x_j - u}{c}\right) \sum_{m=0}^{k_{\bar{x}}-1-j} \frac{T^m}{m!} \tag{4.3}$$

and $k_{\bar{x}}$ denotes such a value, that for $n = cT + u + 1$

$$x_1 + ... + x_{k_{\bar{x}}-1} \leq n - 1 < x_1 + ... + x_{k_{\bar{x}}} . \tag{4.4}$$

The non-conditional ruin probability can be now obtained as a sum over all possible claims' vectors i.e.

$$\phi(u, T) = e^{-T} \sum_{\substack{1 \leq x_1, \\ \ldots \\ 1 \leq x_n}} P(X_1 = x_1, \ldots, X_n = x_n) K_{\bar{x}} . \tag{4.5}$$

Note that in the recursive calculation of the $K_{\bar{x}}$ only the determinant of the largest matrix i.e. $B_{max} = b_{k_{\bar{x}}-1}$ is critical, as the other determinants are side effects of the recursive calculation of B_{max}.

The following claim shows how the Ignatov-Kaishev method can be generalized in order to be used to calculate the failure probability.

Proposition 4.1.1 *Let $n' = cT + u + 1 - w$ and let $k'_{\bar{x}}$ denote such a value, that*

$$x_1 + ... + x_{k'_{\bar{x}}-1} \leq n' - 1 < x_1 + ... + x_{k'_{\bar{x}}} \tag{4.6}$$

and let $K'_{\bar{x}}$ be defined like $K_{\bar{x}}$ but with $k_{\bar{x}}$ replaced by $k'_{\bar{x}}$. Then the probability of non-failure as defined in (2.5) can be expressed as

$$\phi(u, T, w \mid \bar{X} = \bar{x}) = e^{-T} K'_{\bar{x}} . \tag{4.7}$$

Proof Let the claims $\bar{X} = \bar{x}$ be given. Let τ_1 denote the moment when the first claim occurs and τ_i be the waiting time between the $(i-1)$-th and i-th claim for $i > 1$. Then the following holds

$$\begin{aligned} & P\big(\tau > T \ \wedge \ R_u(T) \geq w \mid \bar{X} = \bar{x}\big) \\ = \ & P\bigg(\bigcap_{i=1}^{k'_{\bar{x}}} \big(\tau_1 + \ldots + \tau_i \geq \min(T, \frac{x_1 + \ldots + x_i - u}{c})\big)\bigg) . \end{aligned} \tag{4.8}$$

Since the equality

$$\begin{aligned} & P\big(\bigcap_{i=1}^{k}(\tau_1 + \ldots + \tau_i \geq \min(T, \frac{x_1 + \ldots + x_i - u}{c}))\big) \\ & = e^{-T} \sum_{j=0}^{k-1}(-1)^j b_j(\frac{x_1 - u}{c}, \ldots, \frac{x_1 + \ldots + x_j - u}{c}) \sum_{m=0}^{k-1-j} \frac{T^m}{m!} \end{aligned} \tag{4.9}$$

was proved in [21] without any specific assumptions about k, exactly the same procedure can be used to prove the claim where k is replaced by $k'_{\bar{x}}$. □

Corollary 4.1.2 *Now the result of Kaishev and Ignatov can be generalized to deliver the probability of non-failure:*

$$\phi(u, T, w) = e^{-T} \sum_{\substack{1 \leq x_1, \\ \vdots \\ 1 \leq x_{n'}}} P(X_1 = x_1, \ldots, X_{n'} = x_{n'}) K'_{\bar{x}} . \tag{4.10}$$

The problem with the above is that it contains an infinite sum. Due to this sum, equality cannot be applied in a numerical algorithm. Therefore, a finite equivalent of formula (4.10) is needed. It is provided by the following

Theorem 4.1.3 *Let $n' = cT + u + 1 - w$. Furthermore, let the singleton $C_1^{n'} = \{(n', n', \ldots)\}$ contain an infinite sequence and for $m > 1$ let $C_m^{n'}$ be a set of sequences such that for each element $\bar{x} \in C_m^{n'}$ the following hold*

(i) $\forall i \qquad x_i \in \{1, 2, \ldots\}$,

(ii) $\sum_{i=1}^{m-1} x_i \leq n' - 1$,

(iii) $\forall i \geq m \qquad x_i = n'$.

Then

$$\phi(u,T,w) = e^{-T} \sum_{i=1}^{n'} \sum_{\bar{x} \in C_i^{n'}} P(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i \geq n') K'_{\bar{x}} \ . \tag{4.11}$$

Proof Let $k'_{\bar{x}}$ be defined as in (4.6) and let D_i be such a set that $\bar{x} \in D_i \iff k'_{\bar{x}} = i$. Since $1 \leq k'(x) \leq n'$, it is obvious that

$$\begin{aligned}
\phi(u,T,w) &= \sum_{i=1}^{n'} \sum_{\bar{x} \in D_i} P(\bar{X} = \bar{x}) \phi(u,T,w \mid \bar{X} = \bar{x}) \\
&= \sum_{i=1}^{n'} \sum_{\bar{x} \in D_i} P(\bar{X} = \bar{x}) \phi(u,T,w \mid X_1 = x_1, \ldots, X_i = x_i) \qquad (4.12) \\
&= \sum_{i=1}^{n'} \sum_{\bar{x} \in D_i} P(\bar{X} = \bar{x}) \phi(u,T,w \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = n', X_{i+1} = n', \ldots) \ .
\end{aligned}$$

If two different vectors $\bar{x}$ and $\bar{y}$ are members of D_i and $x_j = y_j$ for $j < i$ then $\phi(u,T,w \mid \bar{X} = \bar{x}) = \phi(u,T,w \mid \bar{X} = \bar{y})$. Hence

$$\begin{aligned}
&\sum_{i=1}^{n'} \sum_{\bar{x} \in D_i} P(\bar{X} = \bar{x}) \phi(u,T,w \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = n', X_{i+1} = n', \ldots) \\
&= \sum_{i=1}^{n'} \sum_{\bar{x} \in C_i^{n'}} P(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = n') \phi(u,T,w \mid \bar{X} = \bar{x}) \ . \qquad (4.13)
\end{aligned}$$

Now, applying (4.7), we have

$$\begin{aligned}
&\sum_{i=1}^{n'} \sum_{\bar{x} \in C_i^{n'}} P(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \phi(u,T,w \mid \bar{X} = \bar{x}) \\
&= e^{-T} \sum_{i=1}^{n'} \sum_{\bar{x} \in C_i^{n'}} P(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = n') K'_{\bar{x}} \ . \qquad (4.14)
\end{aligned}$$

□

4.1.2 Computational Complexity of the Ignatov-Kaishev Method

Since the infinite sum (4.10) was reduced to a finite sum in (4.11), this formula can now be used in practical failure probability applications. Now, it would be interesting to determine the numerical complexity of calculating (4.11). Because the determination of the sum $K'_{\bar{x}}$ for each claim vector is the critical numerical problem, we will consider the computational complexity of the algorithm in terms of the required number of computations of $K'_{\bar{x}}$.

Theorem 4.1.4 *The computational complexity of the naive algorithm for determining* $1 - \psi(u, T, w)$ *is* $O(2^{n'})$ *where* $n' = u + cT + 1 - w$.

Proof Let us first consider $\#C_i^{n'}$ – the size of the set $C_i^{n'}$. $\#C_i^{n'}$ equals to the number of all possible ways of packing n' indistinguishable balls into $(i-1)+1 = i$ distinguishable boxes (the extra box is added to allow the $i-1$ 'real' boxes to contain less than n balls) in such a way, that each of the $i-1$ boxes contains at least one ball. This is equal the number of possibilities of packing $n' - (i-1)$ balls into i boxes i.e.

$$\binom{n' - (i-1) + i - 1}{i-1} = \binom{n'}{i-1} . \tag{4.15}$$

We are now interested in

$$\sum_{i=1}^{n'} \#C_i^{n'} = \sum_{i=1}^{n'} \binom{n'}{i-1} . \tag{4.16}$$

The above equals the number of all possible proper subsets of a set consisting of n' elements. Hence $\sum_{i=1}^{n'} \binom{n'}{i-1} = 2^{n'} - 1$. □

It is clearly seen that the parameter $n' - 1$, the maximal allowed total claim, plays a critical role in the efficiency of the algorithm. In case of the ruin probability $n' - 1$ is chosen as the largest possible, namely $n' - 1 = u + cT$. Hence the complexity of $O(2^{n'})$ for this algorithm is not satisfactory in case of ruin probability. However, it is clear that we can use the same algorithm in a far more effective way if we are interested in computing failure probability instead of ruin probability.

4.1.3 Failure Probability Based on Appel Polynomials

In [25] Lefevre and Picard solved the classical ruin problem using the generalized Appel polynomials. For sake of simplicity we will assume that $\lambda = 1$. Let the auxiliary polynomial $e_n(x)$ be defined as

$$e_n(x) = \begin{cases} 1 & \text{if } n = 0 , \\ \sum_{i=0}^{n} \frac{x^i}{i!} P(\sum_{j=1}^{i} X_j = n) & \text{if } n > 0 . \end{cases} \tag{4.17}$$

The generalized Appel polynomial can be now defined as

$$A_n(x) = \begin{cases} e_n(x) & \text{if } 0 \le n \le u , \\ \sum_{j=0}^{u} \frac{cx-n+u}{cx-j+u} e_j(\frac{j-u}{c}) e_{n-j}(x + \frac{u-j}{c}) & \text{if } n > u . \end{cases} \tag{4.18}$$

Then, according to [25], the probability of non-ruin in the finite time T can be expressed as

$$P\big(\tau > T \ \wedge \ S_n = n\big) = e^{-T} A_n(T) \mathbb{I}_{\{T \geq v_n\}}, \tag{4.19}$$

where $v_n = \frac{n-u}{c}$. Namely

$$1 - \psi(u, T) = e^{-T} \sum_{n=0}^{\infty} A_n(T) \mathbb{I}_{\{T \geq v_n\}}. \tag{4.20}$$

Again, we can see that this non-ruin probability construction can be generalized to provide the probability of non-failure in a very intuitive way.

Proposition 4.1.5 *Non-failure probability can be expressed using the generalized Appel polynomials by*

$$1 - \psi(u, T, w) = e^{-T} \sum_{n=0}^{u+cT-w} A_n(T). \tag{4.21}$$

Proof The equality is a simple consequence of (4.19). □

The problem with this elegant result is that the Appel polynomials introduce numerical complexity, and it is not a trivial task to use them efficiently. We will not study the complexity of this approach here. Some ideas of how the Appel polynomials can be handled numerically and effectively can be found in [1].

4.2 Failure Probability in Discrete Time

In this section we will consider a discrete time model i.e. $t = 1, 2, \ldots, T$. Without loss of generality we assume that the premium revenue per time unit (say a year) is one. In this model, the ruin may occur only at the beginning of a year i.e. for $t = 1, 2, \ldots, T$. The claims are i.i.d. and the number of claims is independent of their sizes as in the previous model. Let Y_i be the aggregated claim in the i-th year. Y_is are also i.i.d. and we denote the aggregate probability function by $f(x) = P(Y_1 = x)$.

The model presented in this section is a modification of the one proposed by De Vylder and Goovaerts in [11] and recalled by Dickson in [13]. The failure in one step is simply expressed as:

$$\psi(u, 1, w) = 1 - \sum_{i=0}^{u-w+1} f(i) \ . \tag{4.22}$$

If we assume that the failure occurs, then either the ruin occurs in the first step, or the ruin does not occur in the first step, but the failure occurs during the next $T-1$ steps. This can be expressed as the recursive equation:

$$\psi(u,T,w)=\psi(u,1,0)+\sum_{j=0}^{u+1} f(j)\psi(u+1-j,T-1,w)\ . \tag{4.23}$$

To improve the numerical efficiency of this recursive algorithm, a truncation procedure similar to the one introduced in [11] can be used. The idea is to use a function $f^{\varepsilon}(x)$ instead of the original $f(x)$. Let $\varepsilon>0$ be small and k be the largest natural number such that $\sum_{i=0}^{k} f(x)\leq 1-\varepsilon$. We have

$$f^{\varepsilon}(x)=\begin{cases} f(x) & \text{if } x\leq k\\ 0 & \text{otherwise.}\end{cases} \tag{4.24}$$

Let now $\psi^{\varepsilon}(u,T,w)$ denote the modified failure probability calculated recursively using the modified function $f^{\varepsilon}(x)$ as follows:

$$\psi^{\varepsilon}(u,1,w)=\begin{cases} \psi(u,1,w) & \text{if } u\leq k\\ 0 & \text{otherwise}\end{cases} \tag{4.25}$$

and

$$\psi^{\varepsilon}(u,T,w)=\psi^{\varepsilon}(u,1,0)+\sum_{j=0}^{u+1} f(j)\psi^{\varepsilon}(u+1-j,T-1,w)\ . \tag{4.26}$$

This improvement is justified by the following

Theorem 4.2.1 *If $k>T+u$ then*

$$\psi^{\epsilon}(u,T,w)\leq\psi(u,T,w)\leq\psi^{\epsilon}(u,T,w)+T\epsilon\ . \tag{4.27}$$

Proof The first inequality is obvious. Let $R_u^{\varepsilon}(t)$ denote the modified risk process which is a copy of the original process but with the only difference that if a claim of size larger than k happens in the $R_u(t)$, then a claim of size ∞ happens in the $R_u^{\varepsilon}(t)$. While the aggregated claims are independent in each time unit, the probability that $R_u(t)=R_u^{\varepsilon}(t)$ equals $(1-\varepsilon)^T$. We have

$$(1-\varepsilon)^T\geq 1-T\varepsilon\ . \tag{4.28}$$

The above inequality is clear for $T=0$. Assuming that it is true for T, let us prove it for $T+1$:

$$
\begin{aligned}
(1-\varepsilon)^{T+1} = (1-\varepsilon)(1-\varepsilon)^T \geq \\
(1-\varepsilon)(1-T\varepsilon) = 1 - T\varepsilon - \varepsilon + T\varepsilon^2 \geq \\
1 - T\varepsilon - \varepsilon = 1 - (T+1)\varepsilon.
\end{aligned}
\tag{4.29}
$$

Hence, the probability that $R_u(t) > R_u^\varepsilon(t)$ does not exceed $T\varepsilon$. Assuming that the failure happened each time $R_u(t) > R_u^\varepsilon(t)$, the probability of failure can not exceed $\psi^\varepsilon(u,t,w) + T\varepsilon$. □

4.3 Numerical Examples

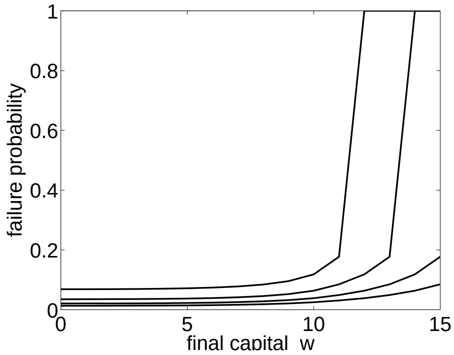
We present numerical examples for the calculation of failure probabilities in the discrete time model from the previous section. We choose one heavy-tailed truncated Pareto distribution and one light-tailed truncated exponential distribution as single claim distributions. The same claim distributions were considered in [13].

The computations were performed using the recursive algorithm expressed with formula (4.26). They were performed for different initial capitals u and different final capitals w. The results are presented as functions of w. The time horizon for all calculations was set to 10. Figure 4.1 presents failure probabilities for the risk processes and the time required to compute them for different initial and final capitals. It is not surprising that the failure probability grows with w and falls with the initial capital u and that for a small initial capital and large w the failure is sure.

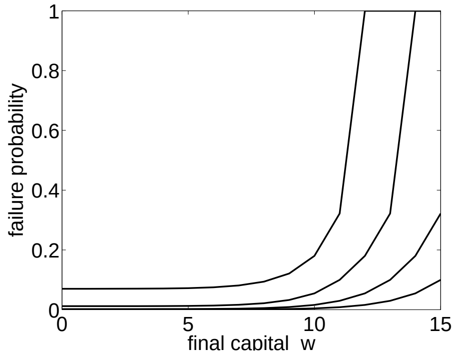
More interesting is the behavior of the computer processor (CPU) time required to compute probability as a function of w. As could have been expected, the CPU time falls rapidly with w. In fact the slope is largest for large initial capital. The empirical results show that the discrete time failure probability is computationally less expensive than the ruin probability. These results accord with the analogous result obtained in Theorem 4.1.4 for continuous time. This can be a strong motivation for using failure probabilities instead of pure ruin probabilities in some practical applications.

4.4 Conclusion

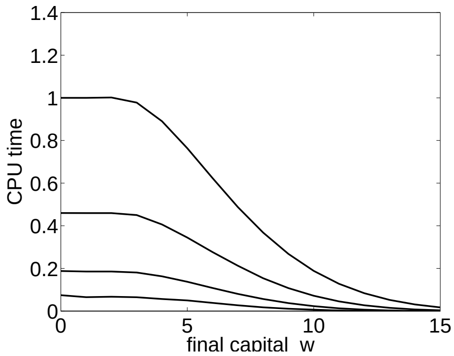
Many popular methods of solving the ruin probability problem with discrete claim distribution can be adopted to solve the failure probability problem as well. Moreover, in many cases the modified methods have better computational complexity and are less time-consuming. The above facts confirm that failure probability is an interesting and valid subject of study.



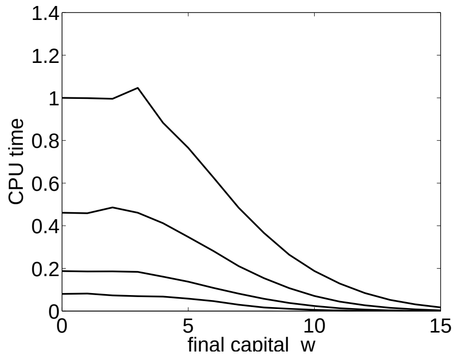
(a) *failure probability, Pareto claim*



(b) *failure probability, exponential claim*



(c) *CPU time, Pareto claim*



(d) *CPU time, exponential claim*

Figure 4.1: Failure probabilities and CPU computation times for the risk processes starting from the initial capital of one, three, five and seven respectively (from top to bottom in the upper panel and from bottom to top in the lower panel). The left panel was obtained for the Pareto claim distribution, the right panel – for exponential claim distribution.

Chapter 5

Stochastic Models for Adult Mortality Intensity

In the life insurance industry the problem concerning the unpredictable mortality intensity is of importance. It is possible that in the future the mortality parameters of the society will be far from those assumed in the actuarial plan of an insurance product. This can happen even if the assumptions were very conservative. For example, a new virus or an environmental threat may emerge that will increase the mortality of the whole population. On the other hand a new medicine may be invented and the mortality intensity will decrease. Such changes may affect the population as a whole or only selected age groups. Thus, a deeper consideration of the future mortality structure is a must. This problem is crucial for both reserving and pricing.

The aim of this chapter is to provide statistical analysis of the available demographic data. We also propose three continuous-time stochastic models that are natural generalizations of the Gompertz law in the sense, that they reduce to the Gompertz function when the volatility parameter is zero. These models have some interesting features. For example, they have few parameters only, these parameters are not functions of time, and at least one of these models can also be efficiently used for mortality option-pricing. Statistical multivariate tests for all three models are provided that allow us to decide which one fits the empirical data best. Finally, we give some practical examples for our multidimensional model.

This chapter is organized as follows. A brief introduction is provided in Section 5.1. Sections 5.2 presents three basic models that we will use in the course of this chapter. Section 5.3 presents the simplest available formula for probability of survival for the models. Section 5.4 contains a reality-check – the models are tested against the demographical data. Finally, some applications can be found in Section 5.5.

5.1 Stochastic Mortality Models

Let us consider a homogeneous cohort of people born in year y. By the standard actuarial notation ${}_{T-t}p_t^y$ we denote the probability that a $(t-y)$-years old member of this cohort survives until T. If μ_t^y is the hazard rate of a single life, we have

$$ {}_{T-t}p_t^y = e^{-\int_t^T \mu_s^y ds} \; . \tag{5.1} $$

If the environment and living conditions do not change over time, we can assume that this cohort's mortality intensity is a function of time only. In classical actuarial theory and practice, μ_t^y is often expressed by the so called Gompertz assumption (see any actuarial textbook e.g. [5]) as a function of t:

$$ \mu_t^y \approx A + Be^{Ct} \tag{5.2} $$

where A, B and C are constants. This model provides a surprisingly accurate approximation in many cases, it is commonly accepted and has been extensively used by practitioners for over a century. Despite its obvious simplicity and usefulness, this method has a serious drawback – it is deterministic and thus it cannot accomodate future randomness. Hence the need for a non-deterministic model emerges and there are a few approaches toward such models in the existing literature.

Predictions of the survival probability p_x, mortality intensity μ_x (also called: force of mortality, mortality rate, hazard rate) as well as the central mortality rate m_x are possible. Among others, the Lee-Carter model presented in [24] and further developed by many authors (e.g. [36], [37]) and the CMI recommendations [9] are broadly applied and recommended. The Lee-Carter method provides not only mortality predictions but also its confidence bounds. The fact that it provides some insight into the random nature of future mortality is of course a useful and desirable feature.

Since the Lee-Carter method is based on time series analysis, it only provides discrete analysis of the problem. Continuous-time stochastic mortality models are presented in [29] and [10]. Here the models were selected mainly to enable mortality-derivative pricing, which is the main objective of these papers. In particular the extended Cox-Ingersoll-Ross (CIR) model is used by Dahl in [10]. The Cox-Ingersoll-Ross model is important for the stochastic modeling of interest rates. In this model the mortality intensity process is described by the following SDE:

$$ d\mu_t^y = a_t(b_t - \mu_t^y)dt + c_t\sqrt{\mu_t^y}dB \; , \tag{5.3} $$

where the parameters a_t, b_t and c_t are functions of time. In this setup μ_t^y is a mean reverting process with mean b_t. Mean reversion is one of Dahl's important motivations for using this model for mortality intensity. Also [29] uses a mean reverting process to model the mortality intensity. Both papers suggest that mean reversion is desired or even required for the mortality model. It certainly important for interest rate models, which John Hull explains in [19] this way: *There are compelling economics arguments in favor of mean reversion. When rates are high, the economy tends to slow down and there is less requirement for funds on the part of borrowers. As a result, rates decline. When rates are low, there tends to be a high demand for funds on the part of borrowers. As a result rates tend to rise.* This motivation does not seem to hold for the mortality intensity, though.

Another argument against mean reversion is that usually it is difficult to estimate the mean from the data. In practical application one would probably have to assume a priori a particular form of the mean function. One possibility is the celebrated Gompertz law.

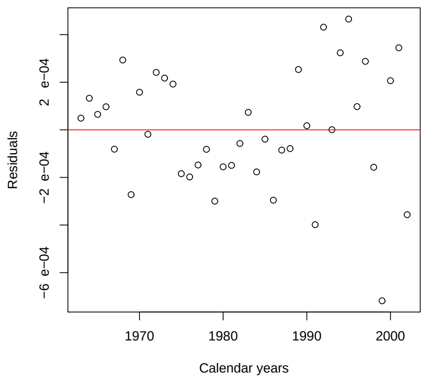


Figure 5.1: The empirical Spanish data. Residuals after fitting Gompertz law with the least squares method

Finally, there is no evidence that the demographic data is mean reverting. Figure 5.1 presents a very typical situation. The Gompertz law was fitted to the Spanish data (adults born 1937) with the least square method. If the underlying process was mean reverting with a moderate variance and a reasonably high speed of reversion, we would not expect to see many adjacent large residuals. Although this cannot be treated as a serious argument against the mean-reverting hypothesis, we can see that there are no good reasons why the mean-reverting framework should be considered the only correct approach.

We want to show that there exist a few stochastic processes that are not mean reverting but fit the data well, have nice analytical properties and have a simple structure.

In the remainder of this chapter, we will be omitting the superscripts in μ_t^y and p_t^y if this does not lead to confusion and hence we will write μ_t or p_t.

5.2 Non- Mean Reverting Models

Because there can be some reservations to the idea of mean reverting mortality models, we make a proposal to use a different group of models. These models are defined and described in this section.

5.2.1 One-dimensional Models

We suggest using the following diffusion processes for modeling mortality intensity:

$$d\mu_t = a\mu_t dt + \mu_t^\beta \sigma dB \ , \qquad t \in [t_0, T] \ , \tag{5.4}$$

for $\beta = 0$, $\beta = 0.5$ and $\beta = 1$. Here $\mu_{t_0} > 0$ is the starting value of the process μ_t. $a > 0$ and σ are constant and B_t is the Brownian motion. We also denote $G = a\mu_t$ and $H = \mu_t^\beta \sigma$. Unique solutions exist for $\beta = 0$ and $\beta = 1$ because the Lipschitz condition holds in these cases. For $\beta = 0.5$ we can apply a special case of the Yamada-Watanabe theorem and see that the weakened Lipschitz condition holds.

Models of such type have many advantages over the mean reverting or even over the Lee-Carter model. At first they are intuitive because all are very natural generalizations of the Gompertz law. Next they have a very transparent structure and are easy to simulate and test. They also have only two parameters (plus the starting value μ_{t_0}) and these parameters are constant over time, what makes them easy to calibrate and finally – apply.

Note that μ_t defined as in (5.4) does not have to follow the affine structure.

If $\beta = 0$ then the dynamics of the process is given by

$$d\mu_t = a\mu_t dt + \sigma dB \ , \qquad t \in [t_0, T] \ . \tag{5.5}$$

If the famous Vacicek interest rate model $dr = a'(b - r)dt + \sigma dB$ did not require a' and b to be strictly positive, equation (5.5) could have been viewed as a special case of the Vacicek model. Our model is no more mean reverting.

The drawback of the process (5.5) is that it can be negative. This is not desirable for the interest rates to be negative but it is even unacceptable for the mortality intensity to be so. We can overcome this problem by defining $\mu_t^* = \max(\epsilon, \mu_t)$ for some small, positive ϵ.

The second model that we propose for modeling continuous-time mortality intensity is given by the following SDE:

$$d\mu_t = a\mu_t dt + \sigma\sqrt{\mu_t} dB \ , \qquad t \in [t_0, T] \ . \tag{5.6}$$

If μ_t follows (5.6), it is positive for any t with probability one. This model could be viewed as a special case of (5.3), however formally the definition of CIR requires its coefficients to be strictly positive. Because here $b_t = 0$ and $a_t < 0$, this model is no more mean reverting. Surprisingly, we will see that this model fits the empirical data well and that there exist explicit formulas for some important functionals of μ_t in this model.

The last proposal (for $\beta = 1$) is to use the geometric Brownian motion as the stochastic replacement for the Gompertz assumption. Let the behavior of μ_t be given by the following SDE:

$$d\mu_t = a\mu_t dt + \sigma\mu_t dB_t, \qquad t \in [t_0, T] \ . \tag{5.7}$$

Of course $\ln(\mu_t)$ has the normal distribution with mean $a - \sigma^2/2$ and variance σ^2. Hence μ_t is positive for any t. This model is well known as the model for stock dynamics. In the interest rate literature (see e.g. [3], Ch. 3.2) it is known as the Dothan model but is not extensively used due to obvious limitations – in this model, the interest rates converge to infinity which is not desirable. However such behavior is reasonable in the case of mortality intensity.

Note that the mortality intensity modeling – unlike the usual interest rate modeling – takes place under the physical measure here.

5.2.2 Multi-dimensional Models

The models (5.5), (5.6) and (5.7) are one-dimensional – they describe the mortality intensity of a single cohort only. Albeit the one-dimensional models seem to be reasonable for each single cohort, one expects that there must be some dependence between the mortalities of people of different ages. For example during a war or a pandemic, the mortality of the whole population increases. The dependence between mortalities in people of like ages would be especially strong. The increase of mortality in people aged say, 82 would – intuitively – be accompanied by an increase in the mortality of those 83-years old, but not necessarily the infants.

To incorporate this common sense rule, the k-dimensional vector of Brownian motions must be used as the source of randomness in the models. This leads to vector-valued equations analogous to (5.5), (5.6) and (5.7) but where the variables μ_t, a and μ_{t_0} are replaced with their k-dimensional versions. Then the multiplications between these variables are understood as multiplications over each coefficient separately. The volatility parameter σ is replaced with a $k \times k$ matrix σ. The covariance matrix $\Sigma = \sigma\sigma^T$.

In this setup, we can describe not only the behavior of an individual cohort but we can also incorporate the dependencies between the mortality of people in different ages. Such effects can now be well modeled by the covariance matrix Σ. The values Σ_{ij} are expected to decrease with $|i - j|$ but to always stay non-negative.

5.3 Probability of Survival

Assuming we have a correct model for μ_t, we still need to be able to calculate some functionals of this process to apply the model. A functional that can be especially useful is the probability of survival.

5.3.1 Survival of a Single Cohort

Let $\{\mathcal{F}_t\}_{t\in[t_0,T]}$ be a filtration over the probability space $(\Omega, \mathbb{F}, P)$. Let μ_t be measurable w.r.t. $\mathcal{F}_t$. The stochastic process

$$p(t,T) = E(e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t) \tag{5.8}$$

denotes the probability under $\mathcal{F}_t$, that a person born in the year y and aged t will survive until the age of T. From Ito's lemma follows that $p(t,T)$ is the solution of PDE:

$$\frac{\partial}{\partial t}p(t,T) + G\frac{\partial}{\partial \mu}p(t,T) + \frac{H^2}{2}\frac{\partial^2}{\partial \mu^2}p(t,T) - \mu p(t,T) = 0\ , \tag{5.9}$$

with the condition $p(T,T) = 1$, see for instance [17], Ch VIII.5. Here G and H are the appropriate coefficients in the Ito equations (5.5), (5.6) and (5.7). For instance $G = a\mu_t$ and $H = \sigma$ if $\beta = 0$. It is useful to give simplest formula possible for (5.8) and this is done in the following

Theorem 5.3.1 *Let the force of mortality be defined by (5.5), (5.6) and (5.7) respectively. Then the probability of survival $p(t,T) = E(e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t)$ is of the form*

(i) if $\beta = 0$ then

$$p(t,T) = e^{M(t,T)+N(t,T)\mu_t}\ , \tag{5.10}$$

where $N(t,T) = \frac{1}{a}\left(1 - e^{a(T-t)}\right)$ and $M(t,T) = \frac{\sigma^2}{4a^3}(2a(T-t) - 4e^{a(T-t)} + e^{2a(T-t)} + 3)$,

(ii) if $\beta = 0.5$ then

$$p(t,T) = e^{N(t,T)\mu_t}\ , \tag{5.11}$$

where $N(t,T) = \frac{2(e^{td}-e^{Td})}{(d+a)e^{td}+(d-a)e^{Td}}$ and $d = \sqrt{a^2+2\sigma^2}$,

(iii) if $\beta = 1$ then

$$p(t,T) = \frac{r^p}{\pi^2}\int_0^\infty \sin(2\sqrt{r}\sinh y)\int_0^\infty f(z)\sin(yz)dzdy + \frac{2}{\Gamma(2p)}r^p K_{20}(2\sqrt{r})\ , \tag{5.12}$$

where $K_q()$ is the modified Bessel function of second kind of order q and

$$\begin{aligned} f(x) &= x\exp\frac{-\sigma^2(4p^2+x^2)(T-t)}{8}\left|\Gamma(i\frac{x}{2}-p)\right|^2\cosh\frac{\pi x}{2}\ , \\ r &= \frac{2\mu_t}{\sigma^2}\ , \\ p &= \frac{1}{2} - a\ . \end{aligned}$$

Proof The proof is similar to the corresponding proofs of the Vasicek and CIR models.

(i) Assume the affine structure $p(t,T) = e^{M(t,T)+N(t,T)\mu_t}$ where $M(T,T) = N(T,T) = 0$. Making use of (5.9) and separating the terms that depend on μ and those that do not, we get

$$\begin{cases} \frac{\partial}{\partial t}N(t,T) + aN(t,T) & = & 1 \ , \\ \frac{\partial}{\partial t}M(t,T) + \frac{\sigma^2}{2}N(t,T)^2 & = & 0 \ , \end{cases}$$

so that $N(t,T) = \frac{1}{a}\left(1 - e^{a(T-t)}\right)$ and finally

$$\begin{aligned} M(t,T) &= -\frac{\sigma^2}{2}\int N(t,T)^2 dt + C \\ &= \frac{\sigma^2(T-t)}{2a^2} - \frac{\sigma^2(4e^{a(T-t)} - e^{2a(T-t)} - 3)}{4a^3} \ . \end{aligned}$$

(ii) Again assume the affine structure as in (i). Making use of (5.9) yields this time

$$\begin{cases} \frac{\partial}{\partial t}N(t,T) + aN(t,T) + \frac{\sigma^2}{2}N(t,T)^2 & = & 1 \ , \\ \frac{\partial}{\partial t}M(t,T) & = & 0 \ . \end{cases}$$

From the second equation and the boundary condition follows that $M(t,T) = 0$. In the first equation the transformation $N(t,T) = \frac{2\tilde{N(t)}'}{\sigma^2 \tilde{N(t)}}$ leads to the following second-order linear equation

$$\tilde{N(t)}'' + a\tilde{N(t)}' - \frac{\sigma^2}{2}\tilde{N(t)} = 0 \ .$$

Because $a^2 + 2\sigma^2 > 0$, we can introduce an auxiliary variable $d = \sqrt{a^2 + 2\sigma^2}$. Now the general solution for $N(t,T)$ is of the form

$$\tilde{N(t)} = D_1 e^{\frac{t}{2}(d-a)} + D_2 e^{-\frac{t}{2}(d+a)} \ ,$$

for constant D_1 and D_2 do not depend on t. Hence

$$N(t,T) = \frac{D_1(d-a)e^{\frac{t}{2}(d-a)} - D_2(d+a)e^{-\frac{t}{2}(d+a)}}{\sigma^2 D_1 e^{\frac{t}{2}(d-a)} + \sigma^2 D_2 e^{-\frac{t}{2}(d+a)}} \ .$$

Applying the boundary condition yields $D_2 = D_1\frac{d-a}{d+a}e^{Td}$ so that we have the explicit formula.

(iii) The formal proof will be omitted, since the same formula can be found in [3], Ch. 3 for the interest rates. The geometric Brownian motion as a model for interest rates was originally introduced in [15]. □

Some practical applications of this theorem can be found later in Section 5.5.

One could be also interested in the conditional variance of the random variable $e^{-\int_t^T \mu_s ds}$. Since

$$Var(e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t) = E(e^{-\int_t^T 2\mu_s ds} \mid \mathcal{F}_t) - \left(E(e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t)\right)^2 \ , \tag{5.13}$$

only the expression $E(e^{-\int_t^T 2\mu_s ds} \mid \mathcal{F}_t)$ is of interest in this case. But based on Ito's lemma we can say that if μ_t is defined by (5.4), then $2\mu_t$ is given by

$$\begin{aligned} d(2\mu_t) &= \left(2a\mu_t dt + 0 + 0\mu_t^{2\beta}\sigma^2\right) dt + 2\mu_t^{\beta}\sigma dB \\ &= 2a\mu_t dt + 2\mu_t^{\beta}\sigma dB \ , \qquad t \in [t_0, T] \ . \end{aligned} \tag{5.14}$$

So to give an explicit formula for $E(e^{-\int_t^T 2\mu_s ds} \mid \mathcal{F}_t)$ it suffices to reapply Theorem 5.3.1 for μ_t with modified parameters G and H.

5.3.2 Survival Probability for Many Cohorts

Let $y = (y_0, y_1, \dots y_{k-1})$ and $m = (m_0, m_1, \dots m_{k-1})$ be vector values. Then another point of interest is the formula for the expectation of the linear combination:

$$\begin{aligned} p^{ym}(t,T) &= E(m \times e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t) \\ &= E(m_0 e^{-\int_t^T \mu_s^{y_0} ds} + \ldots + m_{k-1} e^{-\int_t^T \mu_s^{y_{k-1}} ds} \mid \mathcal{F}_t) \ . \end{aligned} \tag{5.15}$$

If the insurer has a portfolio of $\sum_{i=0}^{k-1} m_i$ pure endowment policies, where m_i policy holders were born in the year y_i, formula (5.15) will provide the expected number of claims from this portfolio at time T. This problem can be solved using results from Theorem 5.3.1 for every cohort independently.

A more interesting case is if we are interested in the variance of $m \times e^{-\int_t^T \mu_s ds}$. We have

$$\begin{aligned} & Var(m \times e^{-\int_t^T \mu_s ds} \mid \mathcal{F}_t) \\ = & \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} m_i m_j Cov\left(e^{-\int_t^T \mu_s^{y_i} ds}, e^{-\int_t^T \mu_s^{y_j} ds} \mid \mathcal{F}_t\right) \\ = & \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} m_i m_j \Bigg(E(e^{-\int_t^T \mu_s^{y_i} + \mu_s^{y_j} ds} \mid \mathcal{F}_t) - \\ & \qquad E(e^{-\int_t^T \mu_s^{y_i} ds} \mid \mathcal{F}_t) E(e^{-\int_t^T \mu_s^{y_j} ds} \mid \mathcal{F}_t)\Bigg) \ . \end{aligned} \tag{5.16}$$

The only part of (5.16) that is problematic is the expression $E(e^{-\int_t^T \mu_s^{y_i}+\mu_s^{y_j} ds} \mid \mathcal{F}_t)$. Since the process $\mu_s^{y_i} + \mu_s^{y_j}$ in not an Ito process anymore (unless the covariation matrix is trivial), we cannot apply Theorem 5.3.1 to calculate $E(e^{-\int_t^T \mu_s^{y_i}+\mu_s^{y_j} ds} \mid \mathcal{F}_t)$. Hence, in the remainder of this paper the variance of a portfolio will be determined using Monte Carlo methods.

5.4 Statistical Analysis of the Demographic Data

We examined the life tables published by The Human Mortality Database (see [20]) for the countries providing consistent datasets and sufficient long history i.e. Austria, Belgium, Bulgaria, Canada, Czech Republic, Denmark, England & Wales, Finland, France, Hungary, Italy, Japan, Latvia, Lithuania, Netherlands, Norway, Spain, Sweden, Switzerland and the USA.

5.4.1 Preliminaries

Using these life tables, the mortality intensity was recomputed from the q_x's based on the assumption of the constant mortality intensity in fractional ages. All the data was subject to the following preliminary steps:

1. All the data concerning youth (24 or younger) was removed.
2. All the data concerning the elderly (76 or older) was removed due to instabilities caused by the small size of the cohort (l_x) and the possibility of effects described in [27].
3. Only cohorts currently aged 25-75 were considered (most recent data).
4. Only the most recent 15 or 40 observations for each cohort (year of birth) was of concern.
5. If sufficient long data was not available for a cohort, this cohort was omitted.

Finally, two datasets were obtained. The first one contains the mortality intensity of people currently aged 39-75 (37 cohorts) in 15 subsequent calendar years. Hence it is a 15×37 matrix for each country. Each row is one observation and each column is one cohort. We have labeled this the 'short history data' set.

The other dataset (the 'long history data') consists of 12 cohorts observed in 40 subsequent calendar years. It concerns people currently aged 64-75. It is a 40×12 matrix for each country.

5.4.2 Extracting the White Noise

Continuing with the refined data we will test if it fits the discretized SDE of the three models proposed in Section 5.2. Note that the equations (5.17), (5.20) and (**??**) are only Euler-type approximations of (5.5), (5.6) and (5.7). This is due to the fact that we assume the transition probabilities to be normally distributed, which is not exactly true. However (5.17), (5.20) and (**??**) can be used as good approximations of the corresponding continuous models.

For $\beta = 0$ the discretized version of (5.5) i.e.

$$\mu_{i+1} - \mu_i = a\mu_i + \sigma(B_{i+1} - B_i) \tag{5.17}$$

leads to the following:

$$x_i = \mu_{i+1} - \mu_i - a\mu_i \ . \tag{5.18}$$

For each i, x_i should be normally distributed with mean zero and variance $diag(\Sigma)$. We can now test if (x_i) for $i = t_0, t_0 + 1, \ldots, T$ form (multivariate) Gaussian white noise. To do this, we have to first estimate the parameter a by matching the first moment of x_i. $E(x_i) = E(\mu_{i+1} - \mu_i - a\mu_i) = 0$ yields the following straightforward estimator:

$$a = \frac{\sum_{i=t_0}^{T-1}(\mu_{i+1} - \mu_i)}{\sum_{i=t_0}^{T-1}\mu_i} \ . \tag{5.19}$$

Having a estimated, we will further compute (x_i) and perform white-noise tests.

For $\beta = 0.5$ we will use a similar procedure as above. Hence we will test if the discretized version of (5.6) i.e.

$$\mu_{i+1} - \mu_i = a\mu_i + \sigma\sqrt{\mu_i}(B_{i+1} - B_i) \tag{5.20}$$

fits the demographic data. In this model

$$x_i = \frac{\mu_{i+1} - \mu_i - a\mu_i}{\sqrt{\mu_i}} \ . \tag{5.21}$$

should be normally distributed with mean zero and variance $diag(\Sigma)$. We will estimate the parameter a by matching the first moment of x_i analogous to the previous example. $E(x_i) = E(\frac{\mu_{i+1}-\mu_i-a\mu_i}{\sqrt{\mu_i}}) = 0$ leads to the following estimator:

$$a = \sum_{i=t_0}^{T-1} \frac{\mu_{i+1} - \mu_i}{\sqrt{\mu_i}} / \sum_{i=t_0}^{T-1} \frac{\mu_i}{\sqrt{\mu_i}} \ . \tag{5.22}$$

We will further compute (x_i) and perform white-noise tests.

If $\beta = 1$, the discretized version of (5.7) will be tested against the demographic data. The logarithm of the the sequence (μ_i) is a taken and differentiated. This way we get another sequence

$$x_i = \log \mu_{i+1} - \log \mu_i \tag{5.23}$$

that should form Gaussian white noise. We will test if this is indeed the case.

5.4.3 Hypothesis Testing for One-dimensional Models

We will perform one-dimensional analysis of (x_i) defined in (5.18), (5.21) and (5.23). For each country and for each cohort the null hypothesis is that the sequence (x_i) is one-dimensional Gaussian white noise.

To test normality, we will use the one-dimensional Shapiro-Wilk test. To test the independence of each sample, a Box-Ljung small sample test is performed for auto covariance function with lag 1, see [28] for reference. Especially for the data of length 15, the results of the Box-Ljung test can be used for orientation purposes only because this is an asymptotic test and it is recommended for large samples only. Therefore, an additional turning point test was done for each cohort.

Tables 5.1 and 5.2 list the results of all the tests for the 5% significance level for the short- and long history data respectively. The values in the first, second and third column of each block are the numbers of the non rejected tests. The last column is the number of cohorts where neither the Box-Ljung nor the Shapiro-Wilk test was rejected.

Assuming that the null hypothesis is true for each cohort and that the test for each cohort is an independent experiment, the number of passing cohorts for each test should follow the binomial model with the 95% probability of success and 5% probability of failure (probability of a type I error). The number of trials equals the number of examined cohorts in each country. For example if there were 12 cohorts examined, the number of rejected tests should not exceed 2 (with a 5% significance level). If there were 37, the number of rejected tests should not exceed 4. An asterix next to a result in Table 5.1 or 5.2 means that the number of tests rejected was not greater than 2 for short history data and 4 for long history data. A plus means that only one cohort was missing from the desired number.

For the short history data and $\beta = 0$ at least one test was not rejected for a reasonably large set of countries. However, both independence and normality tests were passed for Lithuania only. The number of countries where the tests were not rejected may seem small, but note that our hypothesis is that all 37 cohorts follow the model. In the rejected countries, only some of the cohorts do not.

We can see that the model for $\beta = 0.5$ can be applied to the Hungary, Latvia and Lithuania short history data. This is a reasonably large set and it makes this model the best of all three considered.

We can see that the geometric Brownian model ($\beta = 1$) can be applied to the Hungarian and Lithuanian short history data. This model is also applicable for not all, but for most cohorts in the short history data in each country.

For the long history data the model with $\beta = 0$ or $\beta = 1$ cannot be applied to any country as a model for all generations. However, it still fits a fair fraction of generations in all the countries.

Table 5.1: One-dimensional models – short history data (length 15). Box-Ljung, turning point and Shapiro-Wilk tests' results for 95% confidence interval. The BL+SW column shows the number of cohorts passing both the Box-Ljung test and the Shapiro-Wilk test. An asterisk means that the number of tests rejected was not greater than 4. A plus means that the number of tests rejected was not greater than 5.

		# cohorts passing $\beta = 0$				# cohorts pass. $\beta = 0.5$				# cohorts pass. $\beta = 1$			
country	coh.	BL	t.p.	SW	BL+SW	BL	t.p.	SW	BL+SW	BL	t.p.	SW	BL+SW
Austria	37	26	24	34*	23	28	24	36*	27	27	25	37*	27
Belgium	37	28	31	35*	26	29	31	33*	25	29	31	32+	24
Bulgaria	37	33*	24	35*	31	33*	23	34*	30	34*	24	35*	32+
Canada	37	29	26	34*	28	32+	25	34*	30	31	24	35*	29
Czech Rep.	37	29	31	36*	28	28	31	35*	26	27	30	33*	25
Denmark	37	27	25	35*	26	28	25	33*	25	27	25	35*	26
Engl., Wal.	37	18	33*	34*	16	20	33*	36*	19	22	33*	35*	21
Finland	37	30	28	34*	27	29	28	35*	27	27	28	36*	26
France	37	29	25	33*	26	29	25	36*	29	30	25	33*	27
Hungary	37	35*	25	33*	31	36*	23	35*	34*	35*	22	35*	33*
Italy	37	32+	28	32+	29	33*	28	34*	31	32+	28	34*	29
Japan	37	33*	19	34*	30	32+	19	33*	28	36*	21	33*	32+
Latvia	37	36*	22	33*	32+	36*	24	35*	34*	37*	24	32+	32+
Lithuania	37	37*	30	33*	33*	37*	30	36*	36*	37*	29	35*	35*
Netherl.	37	29	28	30	23	27	29	31	23	29	29	34*	26
Norway	37	23	27	35*	23	23	27	35*	23	22	27	35*	21
Spain	37	31	27	31	25	32+	27	35*	30	31	27	34*	28
Sweden	37	33*	23	32+	29	32+	24	34*	29	30	23	35*	28
Switzerl.	37	28	28	33*	25	27	30	35*	25	27	31	35*	25
USA	37	31	31	30	24	29	30	30	22	28	29	30	21

Table 5.2: One-dimensional models – long history data (length 40). Box-Ljung, turning point and Shapiro-Wilk tests' results for 95% confidence interval. The BL+SW column shows the number of cohorts passing both the Box-Ljung test and the Shapiro-Wilk test. An asterisk means that the number of tests rejected was not greater than 2. A plus means that the number of tests rejected was not greater than 3.

		# cohorts passing $\beta = 0$				# cohorts pass. $\beta = 0.5$				# cohorts pass. $\beta = 1$			
country	coh.	BL	t.p.	SW	BL+SW	BL	t.p.	SW	BL+SW	BL	t.p.	SW	BL+SW
Austria	12	2	12*	9+	1	3	12*	10*	3	3	12*	10*	2
Belgium	12	3	11*	10*	2	2	11*	10*	1	2	11*	8	1
Bulgaria	12	7	11*	4	4	6	11*	10*	4	7	11*	9+	4
Canada	12	4	11*	10*	4	6	11*	12*	6	7	11*	7	4
Czech Rep.	12	8	11*	10*	6	10*	10*	8	8	8	10*	8	7
Denmark	12	2	12*	9+	2	2	12*	12*	2	0	11*	11*	0
Engl., Wal.	12	1	10*	5	1	6	10*	12*	6	9+	10*	11*	8
Finland	12	4	11*	11*	3	4	11*	11*	4	3	11*	11*	3
France	12	5	11*	12*	5	7	11*	10*	6	6	11*	11*	5
Hungary	12	9+	11*	7	5	11*	11*	12*	11*	11*	11*	10*	9+
Italy	12	9+	11*	8	5	8	11*	11*	7	9+	11*	6	4
Japan	12	12*	12*	4	4	9+	12*	11*	8	7	12*	7	6
Latvia	12	10*	12*	3	1	12*	12*	11*	11*	8	11*	12*	8
Lithuania	12	9+	12*	10*	7	8	12*	11*	7	8	12*	10*	6
Netherl.	12	5	12*	3	0	3	12*	10*	2	5	12*	11*	5
Norway	12	4	12*	5	0	1	12*	11*	1	0	12*	12*	0
Spain	12	3	12*	12*	3	4	12*	12*	4	6	12*	4	2
Sweden	12	6	12*	9+	5	4	12*	11*	3	5	12*	11*	5
Switzerl.	12	2	12*	11*	2	3	12*	12*	3	3	12*	9+	3
USA	12	3	12*	5	2	4	12*	11*	4	7	12*	12*	7

Table 5.3: 3-dimensional models (people aged 70-72). Portmantou, short-sample portamntou and multivariate Shapiro-Wilk tests' results for long history data (length 40) and 5% conficence interval. An asterisk means that we can accept the model. A plus means that the model was not accepted, but the p-values were relatively close to the desired threshold.

	p-values for $\beta = 0$				p-values for $\beta = 0.5$				p-values for $\beta = 1$			
country	portm.	port.s.	mv-Sh.	f.	portm.	port.s.	mv-Sh.	f.	portm.	port.s.	mv-Sh.	f.
Austria	1.000	1.000	0.000		0.994	1.000	0.016		0.427	0.754	0.300	*
Belgium	1.000	1.000	0.000		0.855	0.965	0.093	*	0.950	0.992	0.055	
Bulgaria	0.999	1.000	0.000		0.814	0.969	0.025	*	0.954	0.995	0.064	
Canada	0.992	1.000	0.086		0.973	0.997	0.041		1.000	1.000	0.000	
Czech Rep.	0.545	0.865	0.001		0.320	0.678	0.222	*	0.958	0.995	0.000	
Denmark	1.000	1.000	0.000		1.000	1.000	0.010		1.000	1.000	0.028	
Engl., Wal.	1.000	1.000	0.000		0.999	1.000	0.071		1.000	1.000	0.006	
Finland	0.994	1.000	0.022		0.989	0.999	0.065		1.000	1.000	0.020	
France	1.000	1.000	0.022		0.988	0.999	0.067		0.999	1.000	0.030	
Hungary	1.000	1.000	0.050		1.000	1.000	0.019		0.506	0.731	0.023	
Italy	1.000	1.000	0.000		0.822	0.957	0.051	*	0.951	0.993	0.000	
Japan	0.500	0.762	0.061	*	0.740	0.915	0.583	*	0.997	1.000	0.000	
Latvia	0.952	0.994	0.000		0.868	0.968	0.015		0.996	1.000	0.575	
Lithuania	0.996	1.000	0.001		0.922	0.989	0.147		1.000	1.000	0.000	
Netherl.	0.991	0.999	0.000		0.948	0.993	0.000		0.799	0.943	0.077	*
Norway	0.992	0.999	0.000		0.435	0.748	0.007		1.000	1.000	0.040	
Spain	0.961	0.995	0.000		0.974	0.998	0.041		1.000	1.000	0.000	
Sweden	1.000	1.000	0.004		0.993	1.000	0.658		0.998	1.000	0.168	
Switzerl.	0.990	0.999	0.000		0.826	0.954	0.058	*	0.993	0.999	0.021	
USA	1.000	1.000	0.000		1.000	1.000	0.002		0.869	0.963	0.001	

The model with $\beta = 0.5$ can be fitted to all the cohorts in two countries Hungary and Latvia. In addition it still fits half of the generations in all other countries as well.

The results may look disappointing at first, but it is important to remember that we were testing the hypothesis that all 37 or all 12 examined cohorts follow the three models. It is possible that in some countries one or two cohorts behave in a different way. This will cause the hypothesis be rejected but it does not mean that the models cannot be used for some or even most of the cohorts in those countries. To provide a better view of this issue, we give the exact number of rejected tests in Tables 5.1 and 5.2.

5.4.4 Hypothesis Testing for Multi-dimensional Models

After a one-dimensional introduction, it is time to test the proper multi-dimensional model. We want to check if the vector sequence (x_i) defined in (5.18), (5.21) and (5.23) form a multivariate Gaussian white noise. Most multivariate tests are designed for samples of large sizes and low a number of dimensions. In our case the number of dimensions is the number of cohorts in each examined country. Therefore, we will restrict our 37-dimensional and 12-dimensional data to 3 dimensions only. We will examine the cohorts who are currently 70-, 71- and 72-years old. We will restrict ourselves to the long history data because the multivariate tests for the short data (of length 15) would not make much sense.

If $x_i = (x_i^1, x_i^2, \ldots, x_i^k)$, the matrix auto covariance function of the series (x_i) is defined by $\Gamma(h) = (\gamma_{ij})$, where

$$\gamma_{ij}(h) = E\left((x_t^i - E(x_t^i))(x_{t-h}^j - E(x_{t-h}^j))\right). \tag{5.24}$$

Two things have to be tested to decide if (x_i) forms white noise: independence and normality. We will test the multivariate normality using the multivariate Shapiro-Wilk test, see e.g. [14], [38]. For independence we will test the null hypothesis that the auto covariance function $\Gamma(h) = 0$ for $h = 1, 2, \ldots, [\frac{n}{4}]$, where n is the size of the sample. To do so, the portmantou χ^2 cross-correlation test is calculated (see [28], Ch 4.4). Because of little power of this test for small samples, [28] suggests an adjustment for short data. So, additionally, the small-sample χ^2 test is also calculated and the its p-values are summarized.

Table 5.3 summarizes the obtained results. We can see that on the ground of the Shapiro-Wilk test and the small-sample portmantou χ^2 test, the $\beta = 0$ model seems to fit to Japan only. The $\beta = 0.5$ model, however, does a better job and can be applied to Belgium, Bulgaria, Czech Republic, Italy, Japan and Switzerland. If $\beta = 1$, the model fits Austria and Netherlands.

The p-values of the portmanteau test suggest that in some cases the residuals do not form white noise but do form some self dependent sequence, maybe an autoregressive time series. However the results prove that all three models are worth considering. In general, for almost 50% of the examined countries, at least one of the considered multivariate models fits.

5.4.5 Correlation Between Cohorts in Multi-dimensional Models

Table 5.4: Three-dimensional model: correlation tests.

	$\beta = 0$	$\beta = 0.5$	$\beta = 1$
Country	accepted	accepted	accepted
Austria	*	*	*
Belgium		*	
Bulgaria	*	*	
Czech Rep.	*	*	
Italy			
Japan			
Netherl.		*	
Switzerl.		*	

We will continue with only these countries, where a model was successfully fitted. We will try to determine if a simple form of the correlation matrix between the increments of the Brownian motions driving two cohorts i and j can be assumed. As already discussed, we would expect this matrix to have non-negative values only. We also expect that values closest to the matrix's diagonal are higher. In our three-dimensional case we will test a simple hypothesis:

$$Cor(x_t^i, x_t^j) = \begin{cases} 1 & \text{for } |i-j| = 0, \\ 0.3 & \text{for } |i-j| = 1, \\ 0 & \text{for } |i-j| = 2. \end{cases} \tag{5.25}$$

Asterisks in Table 5.4 denote those countries, where all three hypotheses from (5.25) hold. We can see that e.g. for $\beta = 0.5$, the hypotheses were accepted for all the countries except Italy and Japan.

This result, together with the ones described in previous sections, provides a simple and transparent framework for modeling stochastic mortality. Randomness of cohorts is based on a multivariate Gaussian distribution and there is also a simple form of the correlation matrix between the cohorts.

5.5 Examples

In this section we will provide some numerical examples of how the systematic mortality risk models can be applied in practice.

5.5.1 Evaluating Theorem 5.3.1

First, we will review the explicit formula for $p(t,T)$ given by (5.5). We will numerically evaluate the formula based on parameters estimated from the 40-years long Austrian data, the same as used in Section 5.4. The cohort of the 70-years old will be used. Using the estimation method given by (5.18) and (5.19) against our data, we come up with $a = 0.06637$ and $\sigma = 0.00056$.

So, on ground of Theorem 5.3.1, we will use formula $p(t,T) = e^{M(t,T)+N(t,T)\mu_t}$, where $N(t,T) = \frac{1}{a}\left(1 - e^{a(T-t)}\right)$ and $M(t,T) = \frac{\sigma^2}{4a^3}(2a(T-t) - 4e^{a(T-t)} + e^{2a(T-t)} + 3)$ for $T \in [t, t+5]$. Calculation based on these simple equations will be compared with the numbers obtained from 40 thousand Monte Carlo simulations. This is summarized by Figure 5.2.

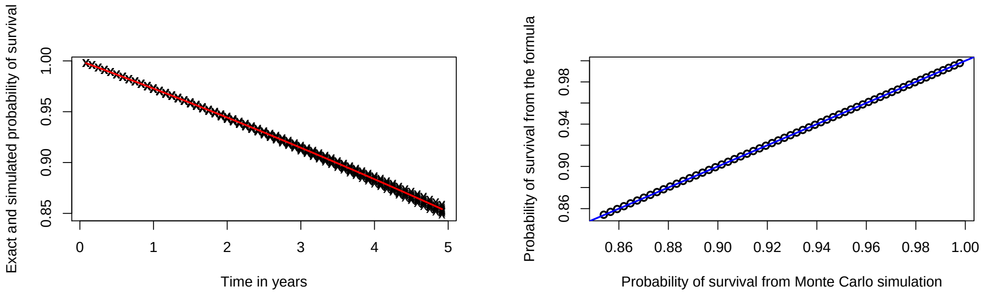


Figure 5.2: Left-hand side panel shows the exact probability of survival (red line) and 10 possible realizations of the stochastic process (black points). Right-hand side panel is the exact probability of survival obtained from the analytical formula vs. the probability based on 40 thousand Monte Carlo simulations. The blue plot is the identity line.

Both graphs show that the formula given by the theorem is confirmed by the Monte Carlo simulations. The first graph shows the exact probability of survival and 10 possible realizations of the stochastic process $e^{-\int_0^T \mu_s ds}$. The other plot shows the expected value of this process obtained from the simulations vs. the expected value obtained from the analytical formula. The sixty points (denoting the probabilities for different T) lay exactly on line $y = x$, as expected. The simplicity of the formula given by Theorem 5.3.1 is obvious and it makes the explicit formula advantageous over the time-consuming process of multiple Monte Carlo simulations.

5.5.2 Pure Endowment Portfolio

Let us consider an insurer that at time 0 sold $3n$ pure endowment contracts to people of age 70, 71 and 72. Let us assume that the contracts were equally distributed among the ages, i.e. each of the three age groups consist of n people. Using the notation from the previous section, $m = (n, n, n)$. In addition, each contract is supposed to pay $1/n$ if the policyholder is still alive at moment T. We also assume that n is large, so that only the systematic risk is an issue for the insurer.

The actuary responsible for the pure endowment product will typically be interested in estimating the value $p^{ym}(0,T)$ as defined in (5.15). Most probably, he will also be interested in the 95% confidence interval for the value $(m \times e^{-\int_0^T \mu_s ds} \mid \mathcal{F}_0)$.

We will model the mortality of this insurer's clients using the model defined by (5.20), so here $\beta = 0.5$. Parameter a and variances for individual cohorts will be estimated from the Austrian data, the same as used in Section 5.4. The 40-years long dataset will be used for the estimation. We will examine two separate scenarios and then compare the results. First, we will assume that the three cohorts in question were described by three independent stochastic processes. In the second scenario, we will assume that the correlation matrix is not an identity matrix.

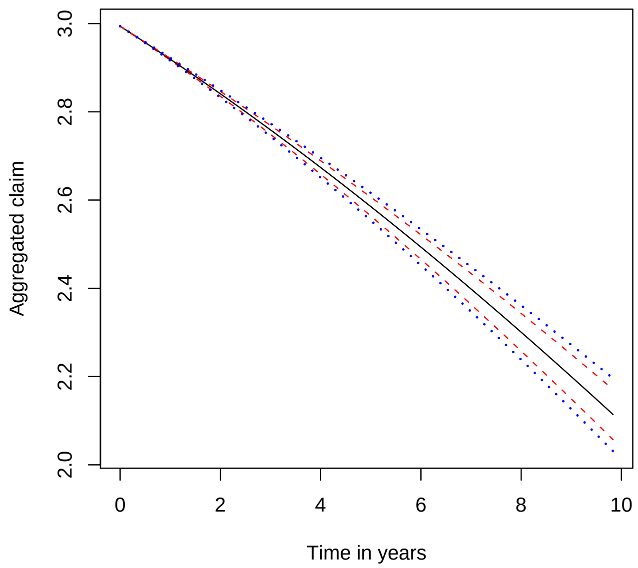


Figure 5.3: The black, solid line is $p(0,T)$ for $T \in [0,10]$. The red, dashed lines are the 95% confidence intervals if the cohorts are independent and the dependent case is marked with the blue, dotted lines.

Figure 5.3 presents the results of the analysis where the quantile lines were calculated with the Monte Carlo methods based on 40 thousand simulations with variance reduction techniques. Of course the value of $p(0,T)$ for $T = 0$ is three and it falls with time. What is essential, is that for $T = 3$ the expected value of claims is 2.11 and the 95% confidence interval is $(2.05, 2.17)$ so the level of uncertainty is remarkable. A conservative actuary would typically want to set an additional reserve to cover the risk introduced by the relativelly wide confidence intervals.

The 95% confidence interval gets even wider if the mortalities of the cohorts are related. If we assume the correlation matrix to have the form

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 2/3 & 1 & 2/3 \\ 1/3 & 2/3 & 1 \end{pmatrix}, \tag{5.26}$$

the interval becomes $(2.03, 2.20)$ so it is over 40% wider than if the mortalities of the three cohorts were not correlated. Of course, the higher the correlation of mortalities between the cohorts, the larger the amount of the systematic mortality risk the company faces. If the cohorts are strongly correlated, the insurer cannot diversify systematic risk by selling insurance to people of different ages. Since there are good reasons to believe that the cohorts' mortalities are in fact correlated (see Section 5.4.5), we can conclude that the amount of the systematic risk embeded in the pure endowment insurance can be highly significant.

5.6 Conclusion

Discretized versions of the stochastic mortality assumptions formulated in this chapter provide a surprisingly good fits to the historical demographic data. They can be used as single- and multi-cohort models for the mortality parameters in at least some countries. Statistical tests reject these models in some cases, though. However, they can still do a good job describing short term behavior of mortality parameters.

Chapter 6

Approximations for Pricing Mortality Options

6.1 Mortality Options

In the stochastic mortality environment, both mortality increase and decrease can be dangerous for a company that has an unbalanced, large portfolio of life insurances. In the first situation (mortality increases) the portfolio of life insurances with the benefit payable at the moment of death will cause unexpected losses. In the second, the portfolio of pure endowments will cause high losses. The problem with this 'systematic' mortality risk is that it cannot be handled in the usual way – by increasing the number of policies sold. It calls for a different hedging strategy.

The simplest way to defend against this risk would be to modify the assumed survival probabilities to obtain 'secure' versions for instance by a linear transformation. A better way would be to try to predict future mortality, which was done in [9]. Time series methods are used e.g. in [24] and further developed in [23], [36] and [37]. More on mortality projections, their applications and model selection can be found in [33] or [18].

In this chapter we propose a few simple continuous stochastic mortality models and then concentrate on a financial instrument that protects against systematic risk. We will call this instrument a 'mortality option'. Unlike the instrument proposed in [29] or [10], it only protects against mortality risk. The interest rate risk is not of concern and we will assume it is constant, equal r.

We will use the following notation: the non-starred symbols such as μ_t or ${}_sp_t$ denote actuarial values from the traditional, deterministic world. The symbols are compatible with the notation used in most of the actuarial literature, e.g. [5].

$$\begin{aligned} {}_{T-t}p_t &= P(\text{a person born at 0 and aged t will survive until T}) \\ &= e^{-\int_t^T \mu_u du}\,, \end{aligned} \tag{6.1}$$

where μ_u is the mortality intensity. The starred symbol μ_t^* denotes a stochastic process and

$$ {}_{T-t}p_t^* = e^{-\int_t^T \mu_u^* du} \; . \tag{6.2} $$

Moreover, we will focus on a homogeneous group of people born in year y. All actuarial symbols used in this paper will refer to this population.

This chapter is organized as follows: In Section 6.2.1 we define the underlying asset and state that this asset can be approximated by the ordinary life insurance of pure endowment insurance. Section 6.2.2 defines the derivatives and Section 6.2.3 presents a real-life situation where these derivatives are needed. The simplest pricing algorithm is proposed in Section 6.3.1 for introductory purposes and finally the continuous model is presented in 6.3.2 and 6.3.3. Sections 6.3.4 and 6.3.5 give the numerical results and a comparison for the proposed pricing approximations.

6.2 The Market

6.2.1 The Underlying Asset

Let us consider an insurance market consisting of two basic contracts. One is a pure endowment contract starting at t and paying one dollar at time T to a person that survives until T. The other one is a life insurance starting at t and paying one dollar at time T in the event the person is dies. The last one is a modification of the standard life insurance since the life insurance typically pays the benefit at the moment of death, not at fixed time T. Such a simplification approximates the normal life insurance if only the interest rate is low, the time interval $T - t$ is short, or if we assume that the death benefit increases with time according to the interest rate.

The insurance company settles the contracts in the following way: the premium is collected and a risk free instrument is bought immediately thereafter. At T the risk-free asset is sold and the benefits are paid. The remaining capital (positive or negative) is kept by the insurer. This minor settlement modification does not change the significant mechanisms of the insurance business.

In any insurance contract the mortality risk carried by the insurer is twofold. One risk is the standard unsystematic insurance risk. It is caused by the fact that the number (or moments) of deaths of the group's members will almost surely differ from the expected value of the number (moments) of deaths. This risk is especially important when the portfolio is small. There are number of actuarial techniques to deal with this kind of risk. The other risk is caused by the fact that the assumed risk parameters (for instance the survival probability) of the population can differ from the real ones. The latter type of risk can be transfered to other parties and traded. We will model it by the following setup. Let us introduce the probability space $(\Omega, \mathbb{F}, P)$. Ω represents all possible states of nature, P is the probability measure and $\mathbb{F} = \{\mathcal{F}_s\}_{t \le s \le T}$ is an increasing sequence of σ-algebras. Let the process μ_s^* be measurable w.r.t. $\mathcal{F}_s$ and let ${}_{T-t}p_t^*$ be defined as in (6.2).

In the financial language the underlying instrument is the probability of death ${}_{T-t}p_t^*$. The problem is that in the real world we cannot buy or sell ${}_{T-t}p_t^*$ directly. However, we can trade a good approximation of this instrument by selling a large portfolio of small policies on many independent lives. This can be done either through selling these policies directly or through reinsurance. Taking the long position is equivalent to selling a portfolio of life insurances. The short position means to sell the pure endowments. This way, selling (or reinsurance) a portfolio of life insurances is a substitute to buying ${}_{T-t}p_t^*$ because the expected benefit from this portfolio at T is proportional to ${}_{T-t}p_t^*$. The unsystematic risk related to this transaction must be handled by classic actuarial techniques. These techniques are out of the scope of this paper. Note that in this approach ${}_{T-t}p_t^*$ can be also viewed as a bond that pays a random amount of money at maturity.

The (actuarial) price of the underlying ${}_{T-t}p_t^*$ at moment $s \in [t, T]$ based on the equivalence rule under the physical probability measure P is

$$\begin{aligned} \mathcal{S}(s) &= e^{-r(T-s)} E^P({}_{T-t}p_t^* | \mathcal{F}_s) && (6.3) \\ &= e^{-r(T-s)}{}_{s-t}p_t^* E^P({}_{T-s}p_s^* | \mathcal{F}_s) \; . && (6.4) \end{aligned}$$

Hence at the moment s the price of ${}_{T-t}p_t^*$ is proportional to the price of ${}_{T-t'}p_{t'}^*$ for any $t' \le s$.

6.2.2 The Derivatives

In addition to $\mathcal{S}(s)$ we will consider two derivatives. One is the European call option that pays $(\mathcal{S}(T) - K)^+$ and the other one is the put option paying $(K - \mathcal{S}(T))^+$ at T.

Figure 6.1 shows sample trajectories of both the underlying asset and the corresponding trajectory of the option. In addition both panels show the 0.05 and 0.95 quantile lines. The interest rate r was set to zero. The plots were prepared assuming the de Moivre-type model described in Section 6.3.2.

The company is considered 'secured' if its only mortality risk is the traditional unsystematic risk that can be 'hedged' with the standard actuarial techniques. To secure against the systematic risk, the company that has an endowment portfolio should buy call options and the insurer that has a life insurance portfolio should buy put options. This intuitive fact is confirmed by the following

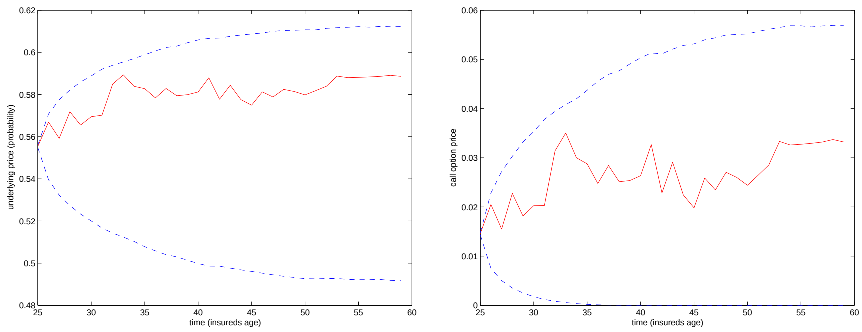


Figure 6.1: Sample trajectories of the underlying mortality instrument $\mathcal{S}(s)$ and the mortality call option $\mathcal{C}(s)$ for $T = 61$ and $b = 0.03$.

Proposition 6.2.1 *If the pure endowment policy for a person aged t with benefit payed at T for a person aged t is priced based on the assumptions that ${}_{T-t}p_t^*$ is constant and equals K, then a call mortality option that pays ${}_{T-t}p_t^* - K$ in case $K <_{T-t} p_t^*$ fully protects the insurance company against the systematic risk of the mortality intensity being lower than expected.*

Proof The company's expected financial obligations are proportional to the probability that a person survives until T. In case ${}_{T-t}p_t^* \leq K$ the company is not exposed to any additional risk so it does not need any protection. In case $K <_{T-t} p_t^*$ the company's obligations minus the contract payoff equal

$$e^{-t(T-t)}\left({}_{T-t}p_t^* - ({}_{T-t}p_t^* - K)\right) = e^{-t(T-t)}\ K\ . \tag{6.5}$$

And this equals the expected obligations calculated for the deterministic mortality intensity. The risk related to K is the standard insurance risk covered by the insurance premium. $\square$

6.2.3 Possible Risk-Trading Scenarios on the Market

The simplest scenario of risk trading between insurance companies is the following:

- company A sold N pure endowment policies to people aged t. Company A is exposed to the systematic low-mortality risk,
- company B sold N life insurances to people aged t and hence it is exposed to the systematic high-mortality risk,
- company A buys N call mortality options that at time T pay $e^{-r(T-t)}({}_{T-t}p_t^* - K)^+$ each and this way perfectly secures itself against low-mortality risk,
- company B buys N put mortality options that at time T pay $e^{-r(T-t)}(K -_{T-t} p_t^*)^+$ each and this way perfectly secures itself against high-mortality risk,

- the issuer of the put mortality options could be company B because, if the mortality is low, B has extra incomes and can pay the option benefit without loses. On the other hand if the mortality is high, the option will not be used by A,
- the issuer of the put options could be company A.

This basic scenario can be enriched by any kind of agents, middlemen and brokers. The contracts can be also settled yearly by any trusted third-party like a stock exchange in a way similar to how some derivatives are settled in the financial market.

The systematic risk can be also transferred using an instrument similar to catastrophic options. In this scenario if A wanted to secure itself, A would lend money from B. At T A would give back the money and interest under the condition that the mortality was high. So there are large variety of hedging scenarios that allow one to transfer the systematic risk to other parties.

Two additional issues are worth mentioning. All proposed contracts require both parties to agree on the way the mortality intensity is determined. Fortunately, in most developed countries, official bodies regularly release mortality tables that can be used as a base for the settlement of contracts. The other thing is that the mortality tables based on demographic data are usually released with certain delays counted in months or sometimes years so that the settlement will always be delayed.

6.3 Pricing the Mortality Call Option

At first we want to remark that if K is the strike price of the option then for the mortality put $\mathcal{P}(s)$ and call $\mathcal{C}(s)$ option the usual equation holds:

$$\mathcal{C}(s) - \mathcal{S}(s) = \mathcal{P}(s) - e^{-r(T-s)}K \ . \tag{6.6}$$

Hence it suffices to price one type of option (call or put) only.

6.3.1 Binomial Model

We present this model for introductory purposes. Let the mortality change in the yearly intervals only and be constant during the whole year. Let the probability of survival for the first year ${}_1p^*_{x-1} = 0.1$. In the oncoming year ${}_1p^*_x$ can be either larger (with probability a) or smaller than 0.1. Figure 6.2 shows this one-step situation.

It is easy to see that in such a case the value of a does not influence the price of the call option on ${}_1p^*_x$ with strike 0.1 (paying 0.02 if the probability of survival rises and nothing otherwise). Such an option can be replicated with the portfolio consisting of 0.5 underlying instruments and -0.04 bonds (here $r = 0$). I.e. the portfolio $(\Phi, \Psi) = (0.5, -0.04)$ has exactly the same payoff as the option, no matter what happens to ${}_{T-t}p^*_t$ in the next year. So the price of the call option in this case equals $\mathcal{C}(0) = \frac{1}{2}\mathcal{S}(0) - 0.04$ independently on a.

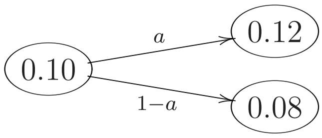


Figure 6.2: In the previous year the probability of death of a x-year old was 0.1. In the next year it will either rise to 0.12 with probability a or fall to 0.08 with probability $1 - a$.

The price of the call option in a multi-step tree can be calculated recursively starting from the right hand side. In this case, analogously to the financial market (see e.g. [19]), the self-financing hedging strategy in each step is

$$\Phi(s) = \frac{\mathcal{C}(s+1) \uparrow - \mathcal{C}(s+1) \downarrow}{\mathcal{S}(s+1) \uparrow - \mathcal{S}(s+1) \downarrow} \tag{6.7}$$

and $\Psi(s)$ is chosen to finance the position in $\mathcal{S}(s)$.

6.3.2 Continuous Mortality Proposals

In the existing literature there have been a number of mortality intensity proposals given. In this paper we will use the diffusion modifications of the classical actuarial mortality assumptions. Let the geometric Brownian motion Y_t be defined as a solution of the following Ito's equation:

$$dY_t = aY_t dt + bY_t dB_t \ , \qquad Y_0 = 1, \tag{6.8}$$

where a and b are constant. Of course $\ln(Y_t)$ has a normal distribution with mean $ta - t\frac{b^2}{2}$ and variance tb^2. The reason why we chose the geometric Brownian motion for the base for stochastic modifications is that it is continuous and it is never negative nor zero. These features are desirable and even required for the mortality intensity process. Some authors, e.g. [10] propose different stochastic processes, mostly mean reverting, for similar applications.

Let us define a martingale $\overline{Y_t}$ with the expected value equal one.

$$\overline{Y_t} = \frac{Y_t}{E(Y_t)} = Y_t e^{-ta} \ . \tag{6.9}$$

Table 6.3.2 summarizes the basic mortality intensity assumptions used in the traditional actuarial literature and their stochastic modifications proposed and examined in this paper. All the modifications are simple multiplications by $\overline{Y_t}$. The deterministic assumptions and the 0.05, 0.25, 0.75 and 0.95 quantile lines of their stochastic versions with the volatility parameter $b = 0.1$ are shown in Figure 6.3. Note that from the definition of μ_t^* follows that $E(\mu_t^*) = \mu_t$. This does not hold for the survival probabilities. From the Jensen's inequity follows that

$$e^{-\int_t^T \mu_u du} \leq E(e^{-\int_t^T \mu_u^* du}) \ . \tag{6.10}$$

Table 6.1: Deterministic mortality assumptions and their stochastic versions

Known as	Originally	Stochastic mortality
de Moivre	$\mu_t = \frac{1}{\omega - t}$,	$\mu_t^* = \frac{1}{\omega-t}\overline{Y_t}$ for sde see (6.11)
Weibull	$\mu_t = kt^n$	$\mu_t^* = kt^n\overline{Y_t}$ for sde see (6.12)
Gompertz	$\mu_t = A + Be^{Ct}$	$\mu_t^* = (A + Be^{Ct})\overline{Y_t}$ for sde see (6.13)

In practice more deterministic models could be used apart from the three mentioned above. For example one of the classic models for μ_t could be modified to comply the recommendations of the CMI Bureau [9].

The following gives the SDE for the mortality proposals and can be used for numerical simulations.

Proposition 6.3.1 *In case of the de Moivre, Weibull and Gompertz assumption's modifications, the mortality process μ_t^* satisfies the following Ito's stochastic differential equations:*

- *for de Moivre:*

$$d\mu_t^* = \left(\mu_t^*(\frac{1}{\omega - t} + a) + a\frac{e^{-ta}}{\omega - t}\right)dt + b\frac{e^{-ta}}{\omega - t}dB_t \ , \tag{6.11}$$

- *for Weibull:*

$$d\mu_t^* = \left(\mu_t^*(\frac{n}{t} - a) + a\frac{kt^n}{e^{ta}}\right)dt + b\frac{kt^n}{e^{ta}}dB_t \ , \tag{6.12}$$

- *for Gompertz:*

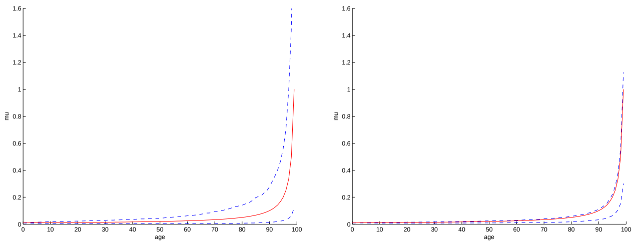
$$d\mu_t^* = \big(\mu_t^*(\frac{BCe^{tC}}{A + Be^{tC}} - a) + a\frac{A + Be^{tC}}{e^{ta}}\big)dt + b\frac{A + Be^{tC}}{e^{ta}}dB_t \ . \tag{6.13}$$

Proof For de Moivre we have

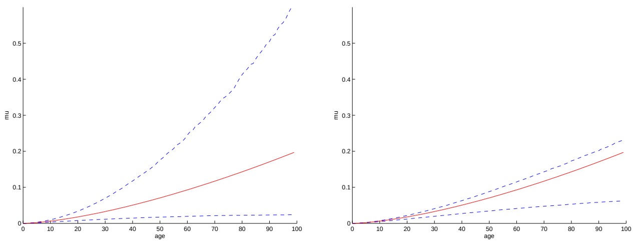
$$\mu_t^* = f(Y_t, t) = \frac{Y_t e^{-ta}}{\omega - t} \ , \tag{6.14}$$

so

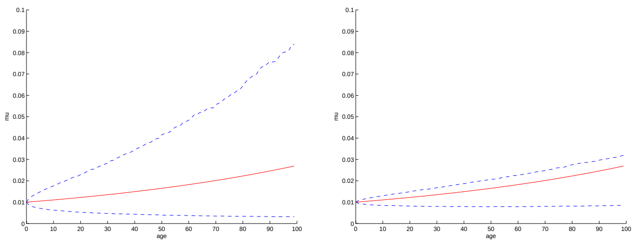
$$Y_t = \mu_t^*(\omega - t)e^{ta} \tag{6.15}$$



deMoivre



Gompertz



Weibull

Figure 6.3: Deterministic mortality assumptions and their stochastic modifications. Dashed graphs denote 0.05, 0.95 quantile lines (left panel) and 0.25 and 0.75 quantile lines (right panel).

and

$$\begin{aligned}
\frac{\delta \mu_t^*}{\delta t} &= Y_t\Big(\frac{e^{-ta}}{(\omega-t)^2} - \frac{ae^{-ta}}{\omega-t}\Big) = \mu_t^*\Big(\frac{1}{\omega-t}+a\Big) \ , \\
\frac{\delta \mu_t^*}{\delta Y_t} &= \frac{e^{-ta}}{\omega-t} \ , \\
\frac{\delta^2 \mu_t^*}{\delta Y_t^2} &= 0 \ . \qquad (6.16)
\end{aligned}$$

Hence, on the grounds of Ito's lemma, the proposition becomes obvious. The proof for the remaining mortality assumptions proceeds the same way. □

6.3.3 Pricing in the Continuous Model

At first let us prove the following

Proposition 6.3.2 *The discounted price of the underlying asset $\mathcal{S}(s)$ is an Ito process and a $\mathcal{F}_s$-martingale (w.r.t. P).*

Proof The fact that $e^{-r(s-t)}\mathcal{S}(s)$ is a martingale is rather straightforward. Let $s \in [t,T]$ and $h \in [0, T-s]$. Then the expectation of the discounted price process at $s+h$

$$\begin{aligned}
E^P \ &\big(e^{-r(s+h-t)}\mathcal{S}(s+h)|\mathcal{F}_s\big) \\
= \ &e^{-r(s+h-t)}e^{-r(T-s-h)}E^P\big({}_{s+h-t}p_t^* E^P({}_{T-s-h}p_{s+h}^*|\mathcal{F}_{s+h})|\mathcal{F}_s\big) \\
= \ &e^{-r(T-t)}E^P({}_{s+h-t}p_t^* {}_{T-s-h}p_{s+h}^*|\mathcal{F}_s) \\
= \ &e^{-r(T-t)} {}_{s-t}p_t^* E^P({}_{T-s}p_s^*|\mathcal{F}_s) \\
= \ &e^{-r(s-t)}\mathcal{S}(s) \ . \qquad (6.17)
\end{aligned}$$

Also $\forall s \in [t,T]$ we have

$$E^P\big(e^{-r(s-t)}\mathcal{S}(s)|\mathcal{F}_s\big)^2 \le 1 \ . \qquad (6.18)$$

so, on ground of the martingale representation theorem (see e.g. [34]), there exist a unique process G_t such that

$$d(e^{-r(s-t)}\mathcal{S}(s)) = G_s dB_s \ . \qquad (6.19)$$

□

So we have that $P = Q$ is the equivalent martingale measure. Hence there is no arbitrage on the market and there exists a unique replication strategy for the derivatives. So the fair market price of the options exists and the price of the call option

$$\mathcal{C}(s) = e^{-r(T-s)}E^Q(\mathcal{S}(T)-K)^+ = e^{-r(T-s)}E^P(\mathcal{S}(T)-K)^+ \ . \qquad (6.20)$$

Moreover, recalling (6.6), it is clear that it suffices to price one type of option (call or put) to find the price of the other one.

The simplest and therefore many practitioner's most beloved method of pricing exotic derivatives is the Monte Carlo method. In this chapter we will use results obtained from this method to compare the results based on the proposed approximations. Whenever the Monte Carlo results are used in this paper they are based on 10^5 simulations.

To price the call mortality option, we will concentrate on the probability distribution of $({}_{T-t}p_t^* - K)^+$ or simply the probability distribution of ${}_{T-t}p_t^*$.

$$\begin{aligned}
P({}_{T-t}p_t^* < x \mid \mathcal{F}_s) &= P\big({}_{T-s}p_s^* < \frac{x}{{}_{s-t}p_t^*} \mid \mathcal{F}_s\big) \\
&= P\big(e^{-\int_s^T \mu_u^* du} < \frac{x}{{}_{s-t}p_t^*} \mid \mathcal{F}_s\big) \\
&= P\big(\int_s^T Y_u \mu_u e^{-ua} du > -\ln \frac{x}{{}_{s-t}p_t^*} \mid \mathcal{F}_s\big) \\
&= P\big(\int_s^T Y_{u-s} \mu_u e^{-(u-s)a} du > \frac{e^{us}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{x} \mid \mathcal{F}_s\big) \\
&= P\big(A(s,T) > \frac{e^{us}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{x} \mid \mathcal{F}_s\big) \ , \qquad (6.21)
\end{aligned}$$

where

$$A(s,T) = \int_0^{T-s} Y_u' e^{-ua} \mu_{u+s} du \qquad (6.22)$$

and Y_u' is an independent copy of Y_u. The problem is that such an integral usually has an unknown distribution (in particular it is not log-normally distributed). The methods used in this section to bypass this problem are similar to the methods used in the average Asian or weighted average Asian option pricing. A comprehensive study of Asian options and the ways to price them can be found, for example, in [32], [19].

6.3.4 Levy-type Approximation

The Levy approximation was proposed in [26]. It was originally designed for pricing Asian average options. Here we will use a modification of this method that can be applied both to the weighted average options and to our purposes.

The fundamental idea is to approximate the distribution of $A(s,T)$ given in (6.22) with the log-normal distribution. Hence we assume that $\ln A(s,T)$ is normally distributed with mean $\alpha(s,T)$ and variance $\beta(s,T)^2$ and then use these parameters in Proposition 6.3.4. This approximation was proved to be accurate at least for the standard average options. Comparing first two moments of the log-normal distribution with the first two moments of the real distribution of $A(s,T)$, we obtain

$$\begin{aligned}\alpha(s,T) &= 2\ln E(A(s,T)) - \frac{\ln E(A(s,T)^2)}{2} \ , \\ \beta(s,T)^2 &= \ln E(A(s,T)^2) - 2\ln E(A(s,T)) \ . \end{aligned} \tag{6.23}$$

It remains to give the formulas for $E(A(s,T))$ and $E(A(s,T)^2)$ and this is done by the following

Lemma 6.3.3 *For $A(s,T)$ defined as in (6.22)*

$$E(A(s,T)) = \int_0^{T-s} \mu_{u+s} du \ , \tag{6.24}$$

$$E(A(s,T)^2) = \int_0^{T-s}\int_0^{T-s} \mu_{u+s}\mu_{v+s} e^{(u\wedge v)b^2} dvdu \ . \tag{6.25}$$

Proof Let us recall that

$$A(s,T) = \int_0^{T-s} \mu_{u+s} e^{-ua} Y_u du \ . \tag{6.26}$$

The equation for the first moment is apparent. As for the second moment of $A(s,T)$ if $n < m$, we have

$$\begin{aligned} E(Y_n Y_m) &= E(Y_n^2)E(\frac{Y_m}{Y_n}) \\ &= E(Y_n^2)E(Y_{m-n}) \\ &= e^{2na+nb^2} e^{(m-n)a} \\ &= e^{(n+m)a+nb^2} \end{aligned} \tag{6.27}$$

otherwise

$$E(Y_n Y_m) = e^{(n+m)a+mb^2} \ . \tag{6.28}$$

So

$$\begin{aligned} E(A(s,T)^2) &= \int_0^{T-s}\int_0^{T-s} \mu_{u+s}\mu_{v+s} e^{-(u+v)a} E(Y_u Y_v) dvdu \\ &= \int_0^{T-s}\int_0^{T-u} \mu_{u+s}\mu_{v+s} e^{-(u+v)a} e^{(u+v)a+ub^2} dvdu \\ &\quad + \int_0^{T-s}\int_0^{T-s} \mu_{u+s}\mu_{v+s} e^{-(u+v)a} e^{(u+v)a+vb^2} dvdu \\ &= \int_0^{T-s}\int_0^{T-s} \mu_{u+s}\mu_{v+s} e^{(u\wedge v)b^2} dvdu \ . \end{aligned} \tag{6.29}$$

□

Now, assuming the log-normality of $A(s,T)$ we can formulate

Proposition 6.3.4 *If we assume the log-normal distribution of $A(s,T)$ then the price at moment s of a call mortality option issued at t and maturing at T with strike price K can be expressed by*

$$e^{-r(T-s)}E^P(({}_{T-t}p_t^* - K)^+ \mid \mathcal{F}_s)$$
$$= \begin{cases} e^{-r(T-s)} \int_K^{{}_{s-t}p_t^*} \Phi(\frac{\ln(\frac{e^{as}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{u}) - \alpha(s)}{\beta(s)})du & \text{if } K < {}_{s-t}p_t^* \\ 0 & \text{otherwise} \end{cases} \tag{6.30}$$

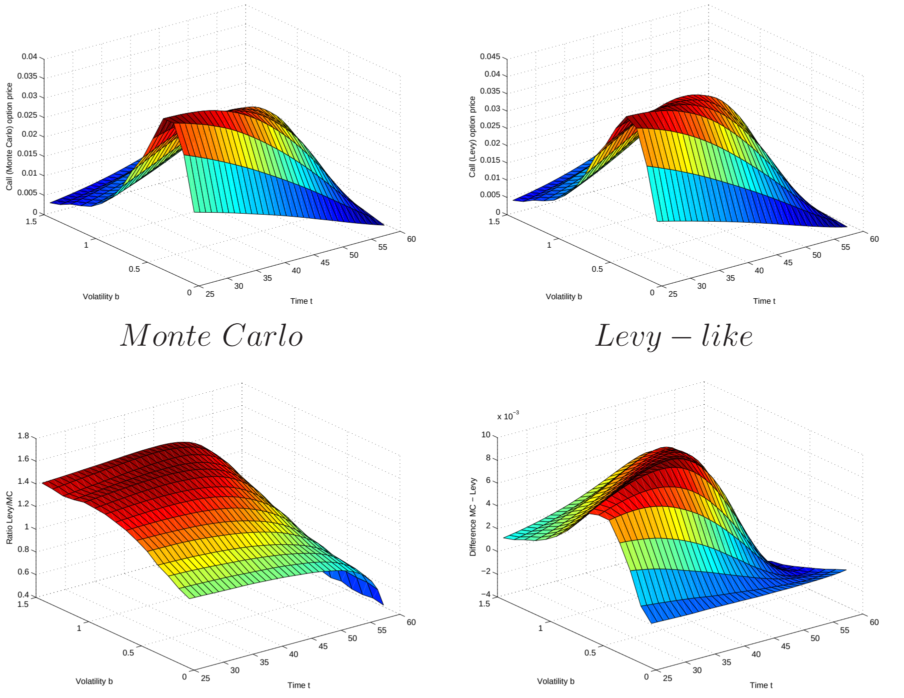
Proof If ${}_{s-t}p_t^* \leq K$ then the proposition is obvious. For $K < {}_{s-t}p_t^*$ we have:

$$\begin{aligned} E^P(({}_{T-t}p_t^* - K)^+ \mid \mathcal{F}_s) &= E^P(({}_{s-t}p_t^* \ {}_{T-s}p_s^* - K)^+ \mid \mathcal{F}_s) \\ &= \int_K^\infty P(e^{-\int_s^T \mu_v^* dv} > \frac{u}{{}_{s-t}p_t^*} \mid \mathcal{F}_s)du \\ &= \int_K^{{}_{s-t}p_t^*} P(e^{-\int_s^T \mu_v^* dv} > \frac{u}{{}_{s-t}p_t^*} \mid \mathcal{F}_s)du \\ &= \int_K^{{}_{s-t}p_t^*} P(A(s,T) < \frac{e^{as}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{u} \mid \mathcal{F}_s)du \\ &= \int_K^{{}_{s-t}p_t^*} P(\ln A(s,T) < \ln(\frac{e^{as}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{u}) \mid \mathcal{F}_s)du \\ &\approx \int_K^{{}_{s-t}p_t^*} \Phi(\frac{\ln(\frac{e^{as}}{Y_s} \ln \frac{{}_{s-t}p_t^*}{u}) - \alpha(s)}{\beta(s)})du \ . \end{aligned} \tag{6.31}$$

□

Note, that the price of the call mortality option is always less than one.

The accuracy of this approximation was checked against the result obtained with the Monte Carlo method for different volatility parameters b. All three basic stochastic mortality assumptions presented in (6.11), (6.12) and (6.13) were tested. Their parameters were estimated from the Polish mortality table for men for the year 2003, see [8]. Thus $\omega = 104.0071$ for de Moivre, $k = 1.7565 \cdot 10^{11}$, $n = 5$ for Weibull, $A = -2.4366 \cdot 10^{-5}$, $B = 7.5436 \cdot 10^{-5}$ and $C = 0.0794$ for Gompertz assumption. Here $a = 0$. We price the options at the moment of issue i.e. $t = s$ and they mature at $T = 61$. For each moment $t = s$ the strike price $K(t) = E({}_{T-t}p_t^*)$ and $\overline{Y_t} = 1$. Interest rate r is zero. Exact values and the ratio $\frac{\text{Levy price}}{\text{exact price}}$ are summarized in Tables 6.2, 6.3, 6.4. Figure 6.4 shows the price surfaces and the comparison between the exact Monte Carlo price and the approximated one.



$Monte\ Carlo$ $\quad$ $Levy - like$

$Ratio\ Levy/Monte\ Carlo$ $\quad$ $difference\ Levy - Monte\ Carlo$

Figure 6.4: Gompertz assumption. Price of the call mortality option for different volatility and moments of issue. Exact (Monte Carlo) results, Levy-like approximations and their comparison.

As can expected, the option price falls with t and grows with b, at least for small b. Such properties are known from the traditional options on the financial market, priced with the Black-Sholes formula. The option price falls again for $b > 0.5$ what may be surprising. This is because for large b the price of a single underlying instrument falls and hence the derivative's price does. The approximation seems to be sufficiently exact in the critical regions where the option price reaches its maximum. The approximation does not fit well for very large volatility (overestimates) nor for very short time to expiration (underestimates). However, in the later case, the exact price of the option is close to zero so the systematic risk can be anyway neglected. Moreover, even in those cases the Levy-like approximation can be used

Table 6.2: De Moivre assumption. Exactness of the Levy-like approximation. (m) - Monte Carlo results, (l) - Levy-like approximation, (r) - $ratio = l/m$.

	t=26			t=36			t=46			t=56		
	l	m	r	l	m	r	l	m	r	l	m	r
b=0.1	0.044	0.045	0.980	0.032	0.033	0.954	0.018	0.019	0.954	0.003	0.004	0.693
b=0.4	0.134	0.0945	1.4206	0.104	0.084	1.238	0.063	0.060	1.038	0.012	0.017	0.711
b=0.7	0.136	0.071	1.843	0.120	0.072	1.663	0.087	0.065	1.340	0.020	0.027	0.774
b=1.1	0.076	0.042	1.800	0.082	0.046	1.778	0.079	0.049	1.617	0.030	0.032	0.919

Table 6.3: Weibull assumption. Exactness of the Levy-like approximation. (m) - Monte Carlo results, (l) - Levy-like approximation, (r) - $ratio = l/m$.

	t=26			t=36			t=46			t=56		
	l	m	r	l	m	r	l	m	r	l	m	r
b=0.1	0.022	0.022	0.981	0.017	0.017	0.966	0.010	0.010	0.924	0.002	0.003	0.702
b=0.4	0.050	0.046	1.010	0.050	0.050	1.100	0.035	0.035	1.027	0.075	0.010	0.733
b=0.7	0.028	0.023	1.233	0.042	0.033	1.286	0.045	0.037	1.220	0.013	0.016	0.805
b=1.1	0.008	0.006	1.295	0.019	0.014	1.400 0.035	0.024	1.410	0.019	0.202	0.930	

Table 6.4: Gompertz assumption. Exactness of the Levy-like approximation. (m) - Monte Carlo results, (l) - Levy-like approximation, (r) - $ratio = l/m$.

	t=26			t=36			t=46			t=56		
	l	m	r	l	m	r	l	m	r	l	m	r
b=0.1	0.016	0.017	0.983	0.012	0.012	0.965	0.007	0.007	0.922	0.001	0.002	0.704
b=0.4	0.043	0.037	1.144	0.039	0.035	1.114	0.025	0.025	1.026	0.005	0.007	0.730
b=0.7	0.027	0.021	1.325	0.035	0.027	1.319	0.034	0.028	1.214	0.009	0.011	0.803
b=1.1	0.010	0.007	1.393	0.018	0.012	1.433	0.027	0.019	1.398	0.013	0.140	0.932

as a first order approximation for the call mortality option price.

6.3.5 Vorst-type Approximation

The Vorst-like approximation of the distribution of the value introduced in equation (6.22) is based on the fact that the arithmetic average can be approximated by the geometric average i.e.

$$\begin{aligned} A(s,T) &= \int_0^{T-s} Y_u e^{-ua} \mu_{u+s} du \\ &\approx (T-s) \frac{1}{T-s} \sum_{i=1}^{T-s} Y_i e^{-ia} \mu_{i+s} \\ &\approx (T-s) \prod_{i=1}^{T-s} \left(Y_i e^{-ia} \mu_{i+s} \right)^{\frac{1}{T-s}} . \end{aligned} \tag{6.32}$$

See e.g. [31], Ch. 6. Now, we have

$$\ln A(s,T) \approx \ln (T-s) + \sum_{i=1}^{T-s} \frac{\ln Y_i - ia + \mu_{i+s}}{T-s} . \tag{6.33}$$

This expression has a normal distribution with central moments

$$\begin{aligned}
m_1(s,T) &= \ln(T-s) + (T-s+1)\frac{b^2}{2} + \sum_{i=1}^{T-s} \frac{\ln \mu_{i+s}}{T-s} \ , \\
m_2(s,T) &= \sum_{i=1}^{T-s} \frac{Var(\ln Y_i)}{(T-s)^2} = \frac{T-s+1}{T-s}\frac{b^2}{2} \ . \qquad (6.34)
\end{aligned}$$

Now we can use these parameters as in Proposition 6.3.4. From Table 6.5 follows that the results are apart from the exact values. Even the usual modifications of the strike price, see [32], are not likely to help in this case.

Table 6.5: Weibull assumption. Exactness of the Vorst-like approximation. (m) - Monte Carlo results, (v) - Vorst-like approximation, (r) - $ratio = v/m$.

	t=26			t=36			t=46			t=56		
	v	m	r	v	m	r	v	m	r	v	m	r
b=0.1	0.000	0.022	0.000	0.000	0.017	0.000	0.000	0.010	0.000	0.000	0.003	0.000
b=0.4	0.000	0.046	0.000	0.000	0.050	0.002	0.098	0.035	2.826	0.047	0.010	4.665
b=0.7	0.000	0.023	0.000	0.054	0.033	1.637	0.073	0.037	1.959	0.046	0.016	2.786
b=1.1	0.000	0.006	0.000	0.020	0.014	1.449	0.040	0.024	1.619	0.040	0.020	1.988

6.4 Conclusion

We proposed a few stochastic mortality models and mortality derivatives that seem to be useful in practice to fully protect against systematic mortality risk. This way insurers can price their product not worrying about the future mortality parameters and do business on the basis of deterministic mortality models. We showed how such derivatives, called 'mortality options' can be priced in the stochastic environment and we have proposed some approximations that allow us not to use Monte Carlo simulations. The proposed Levy-like approximation is proved to be accurate at least for the cases, in which the mortality risk is large and the protection is most needed. The other proposed approximation does not seem to provide reliable results. The mortality derivatives can be uniquely and efficiently priced and might help the insurance company hedge against a potentially dangerous risk that cannot be diversified.

Bibliography

[1] Alkahby, H.; Ansong, G.; Frempong-Mireku, R.; Jalbout, A. *Symbolic computation of Appel polynomials using Maple*. Electron. J. Differ. Equ. Conf. 07 (2001), 1–13.

[2] Asmussen, S. *Ruin Probabilities*. World Sci. Publ., 2000.

[3] Brigo, D.; Mercurio, F. *Interest Rate Models: Theory and Practice*. Springer-Verlag, Berlin, 2001.

[4] Borovkov, A. *Stochastic Processes in Queueing Theory*. Springer-Verlag, 1976.

[5] Bowers, N.L.; Gerber, U.H.; Hickman, J.C.; Jones, D.A.; Nesbit C.J. *Actuarial Mathematics*. The Society of Actuaries, 1997.

[6] Boxma, O.; Cohen J.W. *The M/G/1 queue with heavy-tailed service time distribution*. IEEE journal on selected areas in communications 16 (1998), 749–763.

[7] Burnecki, K.; Miśta, P.; Weron, A. *Ruin probabilities in finite and infinite time*: in Čížek, P.; Härdle, W.; Weron, R. *Statistical Tools for Finance and Insurance*. 341–379, Springer, Berlin, 2005.

[8] Central Statistical Office (Poland), 2004. *Mortality tables 1995-2003*. available at http://www.stat.gov.pl.

[9] CMI Committee *Continuous Mortality Investigation Reports*. Institute of Actuaries, Faculty of Actuaries (CMI Bureau), 2004.

[10] Dahl, M. *Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts*. Insurance: Mathematics and Economics 35 (2004), 113–136.

[11] De Vylder, F.E.; Goovaerts, M. *Recursive calculation of finite-time ruin probabilities*. Insurance: Mathematics and Economics 7 (1988), 1–8.

[12] De Vylder, F.E. *Advanced Risk Theory. A Self-Contained Introduction*. Editions de l'Université de Bruxelles and Swiss Association of Actuaries, 1996.

[13] Dickson, D. *On Numerical Evaluation of Finite Time Survival Probabilities.* British Actuarial Journal 5 (1999), III, 575–584.

[14] Domanski, C. *Własności testu wielowymiarowej normalności Shapiro-Wilka i jego zastosowanie.* Cracow University of Economics Rector's Lectures 37, 1998.

[15] Dothan, L. *On the term structure of interest rates.* Journal of Financial Economics 6 (1978), 59–69.

[16] Garcia, J.M.; Brun, O.; Gauchard, D. *Transcient Analytical Soution of M/D/1/N Queues.* Journal of Applied Probability 39 (2002), 853–864.

[17] Gikhman, I.I.; Skorokhod, A.V. *Introduction to the Theory of Random Processes.* W. B. Saunders Company, Philadelphia, 1969.

[18] Habeman, S.; Sithole, T.; Verral, T. *An investigation into parametric models for mortality projection, with applications to immediate annuitants and life office pensioners data.* Insurance: Mathematics and Economics 27 (2000), 285–312.

[19] Hull, J. *Options, Futures and Other Derivatives.* Prentice-Hall, Englewood, New Jersey, 1989.

[20] *Human Mortality Database.* University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 7/15/2005).

[21] Ignatov, Z.; Kaishev, V. *Two-Sided Bounds for Finite Time Probability of Ruin.* Scand. Actuar. J. 1 (2000), 46–62.

[22] Iwanik, J. *Probability of failure with discrete claim distribution.* Probability and Mathematical Statistics 25 (2005), 307–316.

[23] Lee, H. *The Lee-Carter method for forecasting mortality, with various extensions and applications.* North American Actuarial Journal 4 (2000), no 1, 80–93.

[24] Lee, R.; Carter, L. *Modelling and forecasting the time series of US mortality.* Journal of the American Statistical Association 87 (1992), 659–671.

[25] Lefevre, C.; Picard, P. *The Probability of Ruin in Finite Time with Discrete Claim Size Distribution.* Scand. Actuar. J. 1 (1997), 58–69.

[26] Levy, E. *Pricing European average rate currency options.* Journal of International Money and Finance 11 (1992), 474–491.

[27] Lindbergson, M. *Mortality Among the Elderly in Sweden 1988 1997.* Scandinavian Actuarial Journal 1 (2001), 79 – 94, 2001.

[28] Lutkepohl, H. *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 1993.

[29] Milevsky, M.; Promislov, S. *Mortality derivatives and the option to annuitise*. Insurance: Mathematics and Economics 29 (2001), 299–318.

[30] Moran, P.A.P. *The Theory of Storage*. Butler & Tanner Ltd., Frome, 1959.

[31] Musiela, M.; Rutkowski, M. *Martingale Methods in Financial Modeling*. Springer, 2005.

[32] Nelken, I. (editor), *The Handbook of Exotic Options*. Irwin Professional Publishing, 1996.

[33] Olivieri, A. *Uncertainty in mortality projections: an actuarial perspective*. Insurance: Mathematics and Economics 29 (2001), 231–245.

[34] Oksendal, B. *Stochastic Differential Equations*. Springer, 2003.

[35] Polyanin, A.D.; Zaitsev, V.F.; Moussiaux, A. *Handbook of first order partial differential equations*. Taylor & Francis, London – New York, 2002.

[36] Renshaw, A.; Haberman, S. *Lee-Carter mortality forecasting with age-specific enhancement*. Insurance: Mathematics and Economics 33 (2003), 255–272.

[37] Renshaw, A.; Haberman, S. *On the forecasting of mortality reduction factors*. Insurance: Mathematics and Economics 32 (2003), 379–401.

[38] Royston, P. *An Extension of Shapiro and Wilk's W Test for Normality to Large Samples*. Applied Statistics 31 (1982), 115–124.

[39] Sen, K.; Manju A. *Transcient Busy Period Analysis of Initially Non-Empty M/G/1 Queues – Lattice Path Approach*: in Balakrishnan, N. (editor), *Advances in Combinatorial Methods and Applications to Probability Statistics*. 301 – 315, Birkhäuser, Boston – Basel – Berlin, 1997.

HSC Research Report Series 2006

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Short-term electricity price forecasting with time series models: A review and evaluation* by Rafał Weron and Adam Misiorek

02 *Financial engineering methods in insurance* by Jan Iwanik