APPLICATIONES MATHEMATICAE
34,1 (2007), pp. 53–73

Jan Iwanik (Wrocław)

MODELS FOR STOCHASTIC MORTALITY

Abstract. This paper is an attempt to present and analyse stochastic mortality models. We propose a couple of 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. We provide a statistical analysis of the available demographic data to show that the models fit historical data well. Finally, we give some practical examples for the multidimensional models.

1. Introduction. In the life insurance industry the problem of 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 paper is to address this issue. Statistical analysis of the available demographic data is provided. We also propose 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 a few parameters only, these parameters are not functions of time, and at least one of these models can also be efficiently used for mortality

2000 *Mathematics Subject Classification*: 62P05, 60H30.

Key words and phrases: stochastic mortality, systematic mortality risk, mortality option.

Part of this work was done during the author's stay at Nationwide Insurance, USA.

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.

The paper is organized as follows. An introduction of existing and new models is provided in Section 2. Section 3 contains a reality-check—the models are tested against the demographic data. Finally, some applications can be found in Section 4.

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

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

2.1. *Existing models.* If the environment and living conditions do not change over time, we can assume that the 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. [1]) as a function of t:

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

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 accommodate 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 [11] and further developed by many authors (e.g. [17], [18]) and the CMI recommendations [4] are broadly applied. The Lee–Carter method provides not only mortality predictions but also 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 a discrete analysis of the problem. Continuous-time stochastic mortality models are presented in [15] and [5]. The models there were selected mainly to enable mortality-derivative pricing, which is the main objective of those papers. In particular the extended Cox–Ingersoll–Ross (CIR) model is used by Dahl in [5]. The CIR model is important for the stochastic modeling

of interest rates. In this model the mortality intensity process is described by the following SDE:

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

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 [15] 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 is certainly important for interest rate models, which John Hull explains in [9] 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 a practical application one would probably have to assume a priori a particular form of the mean function. One possibility is the celebrated Gompertz law.

Since there is no evidence that the demographic data are mean reverting, 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 section, we will be omitting the superscripts in μ_t^y and p_t^y if this does not lead to confusion.

2.2. *New models.* Because there can be some reservations to the idea of mean reverting mortality models, we propose to use a different group of models. These models are defined and described in this subsection.

2.2.1. *One-dimensional models.* We suggest using the following diffusion processes for modeling mortality intensity:

$$d\mu_t = a\mu_t dt + \mu_t^{\beta}\sigma dB, \qquad t \in [t_0, T], \tag{1}$$

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 constants, and B_t is the Brownian motion. We also define $G = a\mu_t$ 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 a weakened Lipschitz condition holds.

Models of such type have many advantages over the mean reverting or even over the Lee–Carter model. First, they are intuitive because they are

all natural generalizations of the Gompertz law. Next, they have a 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, which makes them easy to calibrate and finally, apply.

Note that μ_t as defined in (1) does not need to have the affine structure.

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

$$d\mu_t = a\mu_t dt + \sigma dB, \qquad t \in [t_0, T]. \tag{2}$$

If the famous Vašiček interest rate model $dr = a'(b-r)dt + \sigma dB$ did not require a' and b to be strictly positive, equation (2) could have been viewed as a special case of the Vašiček model. Our model is no more mean reverting.

The drawback of the process (2) is that it can be negative. This is undesirable for the interest rates, and unacceptable for the mortality intensity. We can overcome this problem by defining $\mu_t^* = \max(\varepsilon, \mu_t)$ for some small, positive ε.

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

$$d\mu_t = a\mu_t dt + \sigma\sqrt{\mu_t}\, dB, \qquad t \in [t_0, T]. \tag{3}$$

If μ_t follows (3), it is positive for any t with probability one. This model could be viewed as a special case of the Cox–Ingersoll–Ross model, although 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 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 described by the following SDE:

$$d\mu_t = a\mu_t dt + \sigma\mu_t dB_t, \qquad t \in [t_0, T]. \tag{4}$$

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. [2, 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 undesirable. 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.

2.2.2. *Multi-dimensional models.* The models (2), (3) and (4) 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 (2)–(4) 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 for each component separately. The volatility parameter σ is replaced with a $k \times k$ matrix σ. The covariance matrix is $\Sigma = \sigma\sigma^T$.

In this setup, we can not only describe the behavior of an individual cohort but also incorporate the dependences 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.

2.3. *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.

2.3.1. *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

$$p(t,T) = E(e^{-\int_t^T \mu_s\, ds} \mid \mathcal{F}_t) \tag{5}$$

denotes the conditional probability that a person born in year y and aged t will survive until the age of T. From Ito's lemma it follows that $p(t,T)$ is the solution of the PDE:

$$\frac{\partial}{\partial t}p(t,T) + G\,\frac{\partial}{\partial \mu}p(t,T) + \frac{H^2}{2}\,\frac{\partial^2}{\partial \mu^2}p(t,T) - \mu p(t,T) = 0, \tag{6}$$

with the condition $p(T,T) = 1$ (see for instance [8, Ch. VIII.5]). Here G and H are the appropriate coefficients in the Ito equations (2), (3) and (4). For instance $G = a\mu_t$ and $H = \sigma$ if $\beta = 0$. It is useful to give a simplest formula possible for (5) and this is done in the following

THEOREM 2.1. *Let the force of mortality be defined by* (2), (3) *or* (4). *Then the probability of survival is as follows*:

(i) *if* $\beta = 0$ *then*

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

where

$$N(t,T) = \frac{1}{a}(1 - e^{a(T-t)}),$$

$$M(t,T) = \frac{\sigma^2}{4a^3}(2a(T-t) - 4e^{a(T-t)} + e^{2a(T-t)} + 3),$$

(ii) *if* $\beta = 0.5$ *then*

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

where

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

(iii) *if* $\beta = 1$ *then*

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

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

$$f(x) = x \exp \frac{-\sigma^2(4p^2 + x^2)(T-t)}{8} \left| \Gamma\left(i\,\frac{x}{2} - p\right)\right|^2 \cosh\frac{\pi x}{2},$$

$$r = \frac{2\mu_t}{\sigma^2}, \qquad p = \frac{1}{2} - a.$$

Proof. The proof is similar to the corresponding proofs for the Vašiček 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 (6) and separating the terms that depend on μ and those that do not, we get

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

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

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

(ii) Again assume the affine structure as in (i). Making use of (6) yields this time

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

From the second equation and the boundary condition it follows that $M(t,T) = 0$. In the first equation the transformation

$$N(t,T) = \frac{2\widetilde{N}(t)'}{\sigma^2\widetilde{N}(t)}$$

leads to the second-order linear equation

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

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

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

for constants D_1 and D_2 do not depending on t. Hence

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

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 [2, Ch. 3] for the interest rates. The geometric Brownian motion as a model for interest rates was originally introduced in [7]. ∎

Some practical applications of this theorem can be found in Section 4.

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

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

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 (1), then $2\mu_t$ is given by

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

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 2.1 for μ_t with modified parameters G and H.

2.3.2. *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:

$$p^{ym}(t,T) = E(m \cdot e^{-\int_t^T \mu_s\, ds} \mid \mathcal{F}_t) = E(m_0 e^{-\int_t^T \mu_s^{y_0}\, ds} + \dots + m_{k-1} e^{-\int_t^T \mu_s^{y_{k-1}}\, ds} \mid \mathcal{F}_t). \tag{7}$$

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 year y_i, formula (7) will provide the expected number of claims from this portfolio at time T. This problem can be solved using the results from Theorem 2.1 for every cohort independently.

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

$$\begin{aligned}\operatorname{Var}(m \cdot e^{-\int_t^T \mu_s\, ds} \mid \mathcal{F}_t) &= \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} m_i m_j \operatorname{Cov}(e^{-\int_t^T \mu_s^{y_i}\, ds}, e^{-\int_t^T \mu_s^{y_j}\, ds} \mid \mathcal{F}_t) \\ &= \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} m_i m_j (E(e^{-\int_t^T \mu_s^{y_i} + \mu_s^{y_j}\, ds} \mid \mathcal{F}_t) \\ &\quad - E(e^{-\int_t^T \mu_s^{y_i}\, ds} \mid \mathcal{F}_t) E(e^{-\int_t^T \mu_s^{y_j}\, ds} \mid \mathcal{F}_t)).\end{aligned} \tag{8}$$

The only part of (8) that is problematic is $E(e^{-\int_t^T (\mu_s^{y_i} + \mu_s^{y_j})\, ds} \mid \mathcal{F}_t)$. Since $\mu_s^{y_i} + \mu_s^{y_j}$ is not an Ito process any more (unless the covariance matrix is trivial), we cannot apply Theorem 2.1 to calculate this expectation. Hence, in the remainder of this paper the variance of a portfolio will be determined using Monte Carlo methods.

3. Statistical analysis of demographic data.

We examined the life tables published by The Human Mortality Database (see [10]) 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.

3.1. *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 were subject to the following preliminary steps:

1. All the data concerning youth (24 or younger) were removed.
2. All the data concerning the elderly (76 or older) were removed due to instabilities caused by the small size of the cohort (l_x) and the possibility of effects described in [13].
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) were of concern.
5. If sufficiently long data were not available for a cohort, the 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.

3.2. *Extracting the white noise.* We will test if the refined data fits the discretized SDE of the three models proposed in Section 2.2. Note that the equations (9) and (11) are only Euler-type approximations of (2) and (3). This is due to the fact that we assume the transition probabilities to be normally distributed, which is not exactly true. However, (9) and (11) can be used as good approximations of the corresponding continuous models.

For $\beta = 0$ the discretized version of (2), i.e.

$$\mu_{i+1} - \mu_i = a\mu_i + \sigma(B_{i+1} - B_i) \tag{9}$$

leads to the following:

$$x_i = \mu_{i+1} - \mu_i - a\mu_i. \tag{10}$$

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

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

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

For $\beta = 0.5$ we use a similar procedure. Hence we test if the discretized version of (3), i.e.

$$\mu_{i+1} - \mu_i = a\mu_i + \sigma\sqrt{\mu_i}(B_{i+1} - B_i), \tag{11}$$

fits the demographic data. In this model

$$x_i = \frac{\mu_{i+1} - \mu_i - a\mu_i}{\sqrt{\mu_i}} \tag{12}$$

should be normally distributed with mean zero and variance $\operatorname{diag}(\Sigma)$. We estimate the parameter a by matching the first moment of x_i analogous

to the previous example. Now, $E(x_i) = E(\frac{\mu_{i+1}-\mu_i-a\mu_i}{\sqrt{\mu_i}}) = 0$ leads to the following estimator:

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

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

If $\beta = 1$, the discretized version of (4) are tested against the demographic data. The logarithm of the sequence (μ_i) is taken and differentiated. This way we get another sequence

$$x_i = \log \mu_{i+1} - \log \mu_i \tag{13}$$

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

3.3. *Hypothesis testing for one-dimensional models.* We will perform one-dimensional analysis of (x_i) defined in (10), (12) and (13). For each country and for each cohort the null hypothesis is that the sequence (x_i) is a one-dimensional Gaussian white noise.

To test normality, we use the one-dimensional Shapiro–Wilk test. To test the independence of each sample, a Box–Ljung small sample test is performed for the auto covariance function with lag 1 (see [14]). 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.

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 a 95% probability of success and 5% probability of failure (probability of a type I error). The number of trials equals the number of cohorts examined 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.

For the short history data and $\beta = 0$ at least one test was not rejected for a reasonably large set of countries. However, only Lithuania passed both independence and normality tests. 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 short history data of Hungary, Latvia and Lithuania. 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 for 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 these countries.

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 the generations in all other countries as well.

The results may seem disappointing at first, but it is important to remember that we were testing the hypothesis that all 37 or all 12 cohorts examined follow the three models. It is possible that in some countries one or two cohorts behave in a different way. This will cause that the hypothesis is rejected but it does not mean that the models cannot be used for some or even most of the cohorts in those countries.

3.4. *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 for our three models forms a multivariate Gaussian white noise. Most multivariate tests are designed for samples of large sizes and low dimensions. In our case dimension is the number of cohorts in each country examined. Therefore, we will restrict our 37-dimensional and 12-dimensional data to three dimensions only. We will examine the cohorts that 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 $x_i = (x_i^1, x_i^2, \dots, x_i^k)$, the matrix auto covariance function of the series (x_i) is defined by $\Gamma(h) = (\gamma_{ij})$, where

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

Two things have to be tested to decide if (x_i) forms a white noise: independence and normality. We will test the multivariate normality using the multivariate Shapiro–Wilk test (see e.g. [6], [19]). For independence we will test the null hypothesis that the auto covariance function $\Gamma(h)$ is zero for $h = 1, \dots, [n/4]$, where n is the size of the sample. To do so, the portmanteau χ^2 cross-correlation test is calculated (see [14, Ch. 4.4]). Because of little power of this test for small samples, [14] suggests an adjustment for short data. So, additionally, the small-sample χ^2 test is also calculated and its p-values are summarized.

The Shapiro–Wilk test and the small-sample portmanteau χ^2 test show that the $\beta = 0$ model seems to fit 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 the Netherlands.

The p-values of the portmanteau test suggest that in some cases the residuals do not form a 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 countries examined, at least one of the multivariate models considered fits.

3.5. *Correlation between cohorts in multi-dimensional models.* We will continue with only those 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:

$$\text{Cor}(x_t^i, x_t^j) = \begin{cases} 1 & \text{for } |i-j| = 0, \\ 0.3 & \text{for } |i-j| = 1, \\ 0 & \text{for } |i-j| = 2. \end{cases} \tag{14}$$

Asterisks in Table 1 denote those countries where all three hypotheses from (14) 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 subsections, 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.

Table 1. Three-dimensional model: correlation tests

	$\beta = 0$	$\beta = 0.5$	$\beta = 1$
Austria	*	*	*
Belgium		*	
Bulgaria	*	*	
Czech Rep.	*	*	
Italy			
Japan			
Netherlands		*	
Switzerland		*	

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

4.1. *Evaluating Theorem 2.1.* First, we review the explicit formula for $p(t,T)$ given by (2). We numerically evaluate the formula based on parameters estimated from the 40-year-long Austrian data, the same as used in Section 3. The cohort of the 70-year-olds is used. Using the estimation method given in the previous section for $\beta = 0$, we come up with $a = 0.06637$ and $\sigma = 0.00056$.

By Theorem 2.1, we use the formula $p(t,T) = e^{M(t,T)+N(t,T)\mu_t}$, where

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

for $T \in [t, t+5]$. Calculation based on these simple equations is compared with the numbers obtained from 40 thousand Monte Carlo simulations. This is summarized in Figure 1.

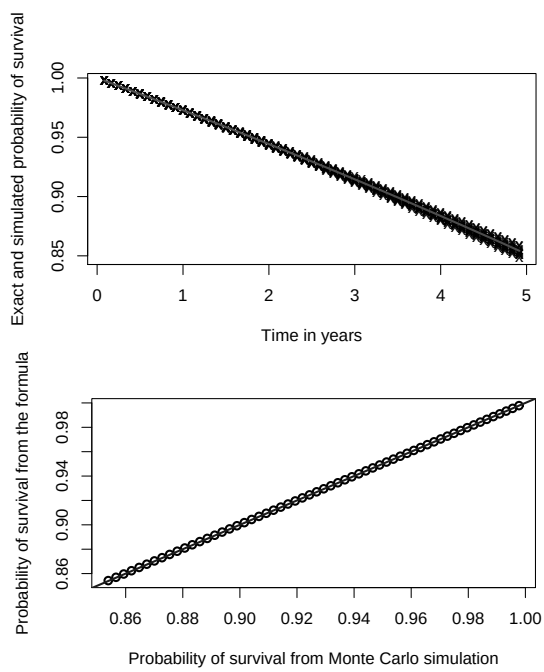

Fig. 1. The top diagram shows the exact probability of survival (the light line in the middle of the chart) and 10 possible realizations of the stochastic process (black surrounding points). The bottom diagram shows the exact probability of survival obtained from the analytical formula vs. the probability based on 40 thousand Monte Carlo simulations (black dots). The identity line is also included in the graph.

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 the line $y = x$, as expected. The simplicity of the formula given by Theorem 2.1 is obvious and it makes the explicit formula advantageous over the time-consuming process of multiple Monte Carlo simulations.

4.2. *Pure endowment portfolio.* Consider an insurer that at time 0 sold $3n$ pure endowment contracts to people of age 70, 71 and 72. Assume that the contracts were equally distributed among the ages, i.e. each of the three age groups consists of n people. Using the notation from the previous subsection, $m = (n, n, n)$. In addition, each contract is supposed to pay $1/n$ if the policyholder is still alive at time 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 (7). Most probably, he will also be interested in the 95% confidence interval for the value $E(m \cdot e^{-\int_0^T \mu_s\, ds} \mid \mathcal{F}_0)$.

We will model the mortality of this insurer's clients using the model defined by (11), so here $\beta = 0.5$. The parameter a and variances for individual cohorts will be estimated from the Austrian data, used in Section 3. The 40-year-long dataset will be used for the estimation. We examine two separate scenarios and then compare the results. First, we assume that the three cohorts in question are described by three independent stochastic processes. In the second scenario, we assume that the correlation matrix is not an identity matrix.

Figure 2 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 relatively 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

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 2/3 & 1 & 2/3 \\ 1/3 & 2/3 & 1 \end{pmatrix}, \tag{15}$$

the interval becomes $(2.03, 2.20)$ so it is over 40% wider than in the uncor-

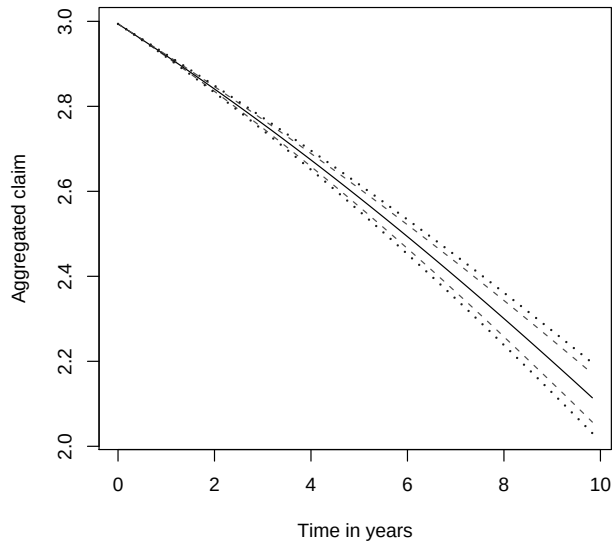

Fig. 2. The solid line is $p(0,T)$ for $T \in [0,10]$. The dashed lines are the 95% confidence intervals if the cohorts are independent, and the dependent case is marked with dotted lines

related case. 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 3.5), we conclude that the systematic risk embedded in the pure endowment insurance may be significant.

4.3. *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 time 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.

If ${}_{T-t}p_t$ denotes a stochastic process

$$ {}_{T-t}p_t = e^{-\int_t^T \mu_u \, du} $$

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

$$ \mathcal{S}(s) = e^{-r(T-s)} E^P({}_{T-t}p_t \,|\, \mathcal{F}_s) = e^{-r(T-s)} {}_{s-t}p_t E^P({}_{T-s}p_s \,|\, \mathcal{F}_s). \tag{16} $$

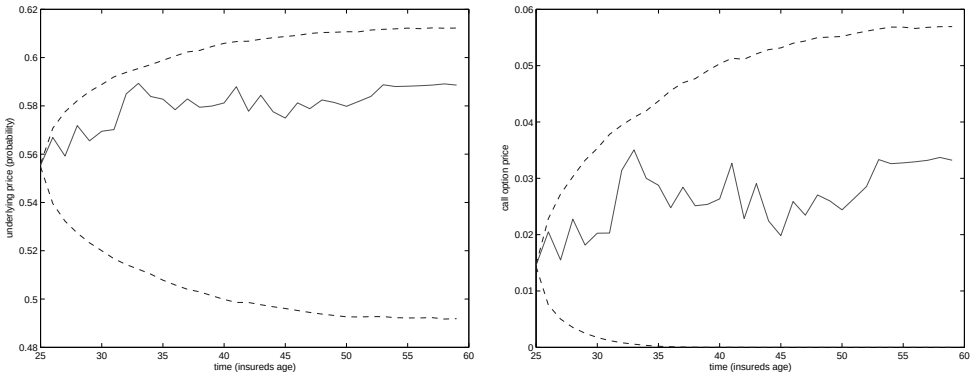

Fig. 3. Sample trajectories of the underlying mortality instrument $\mathcal{S}(s)$ and the mortality call option $\mathcal{C}(s)$ for $T = 61$ and $b = 0.03$

The easiest way to protect against the systematic mortality risk is to buy an European call option that pays $(\mathcal{S}(T) - K)^+$ at time T. Figure 3 shows sample trajectories of both the underlying asset and the corresponding trajectory of the option. In addition both diagrams show the 0.05 and 0.95 quantile lines. The interest rate r was set to zero.

In this example we will use a modification of the model defined in (4). Let Y_t be the geometric Brownian motion. Define a martingale $\overline{Y_t}$ with expected value one,

$$\overline{Y_t} = \frac{Y_t}{E(Y_t)} = Y_t e^{-ta}, \tag{17}$$

and set $\mu_t = (A + Be^{Ct})\overline{Y_t}$. It is easy to check that μ_t satisfies the following Ito stochastic differential equation:

$$d\mu_t = \left(\mu_t\left(\frac{BCe^{tC}}{A + Be^{tC}} - a\right) + a\,\frac{A + Be^{tC}}{e^{ta}}\right)dt + b\,\frac{A + Be^{tC}}{e^{ta}}\,dB_t.$$

Now, since the discounted price of the underlying asset $\mathcal{S}(s)$ is an Ito process and an $\mathcal{F}_s$-martingale, 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 is

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

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

$$\begin{aligned} P({}_{T-t}p_t < x \,|\, \mathcal{F}_s) &= P\left({}_{T-s}p_s < \frac{x}{{}_{s-t}p_t} \,\middle|\, \mathcal{F}_s\right) \\ &= P\left(e^{-\int_s^T \mu_u\,du} < \frac{x}{{}_{s-t}p_t} \,\middle|\, \mathcal{F}_s\right) \end{aligned}$$

$$= P\left(\int_s^T Y_u \mu_u e^{-ua}\, du > -\ln \frac{x}{{}_{s-t}p_t} \,\middle|\, \mathcal{F}_s \right)$$

$$= P\left(\int_s^T Y_{u-s} \mu_u e^{-(u-s)a}\, du > \frac{e^{us}}{Y_s} \ln \frac{{}_{s-t}p_t}{x} \,\middle|\, \mathcal{F}_s \right)$$

$$= P\left(A(s,T) > \frac{e^{us}}{Y_s} \ln \frac{{}_{s-t}p_t}{x} \,\middle|\, \mathcal{F}_s \right),$$

where

$$A(s,T) = \int_0^{T-s} Y'_u e^{-ua} \mu_{u+s}\, du \tag{18}$$

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 subsection 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 [16], [9].

The Levy approximation was proposed in [12]. 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 (18) 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 4.2. This approximation was proved to be accurate at least for the standard average options. Comparing the first two moments of the log-normal distribution with the first two moments of the real distribution of $A(s,T)$, we obtain

$$\alpha(s,T) = 2\ln E(A(s,T)) - \frac{\ln E(A(s,T)^2)}{2},$$
$$\beta(s,T)^2 = \ln E(A(s,T)^2) - 2\ln E(A(s,T)).$$

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

LEMMA 4.1. *For $A(s,T)$ defined in* (18),

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

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

Proof. Recall that

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

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

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

otherwise

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

So

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

Now, we can formulate

PROPOSITION 4.2. *Assume that $A(s,T)$ is log-normally distributed. Then the price at time s of a mortality call option issued at t and maturing at T with strike price K can be expressed by*

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

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

$$\begin{aligned} E^P((_{T-t}p_t^* - K)^+ \mid \mathcal{F}_s) &= E^P((_{s-t}p_t^{\ *}\ {}_{T-s}p_s^* - K)^+ \mid \mathcal{F}_s) \\ &= \int_K^\infty P\Big(e^{-\int_s^T \mu_v^*\, dv} > \frac{u}{{}_{s-t}p_t^*} \,\Big|\, \mathcal{F}_s\Big) du \end{aligned}$$

$$= \int_K^{s-tp_t^*} P\left(e^{-\int_s^T \mu_v^* dv} > \frac{u}{s-tp_t^*} \,\middle|\, \mathcal{F}_s\right) du$$

$$= \int_K^{s-tp_t^*} P\left(A(s,T) < \frac{e^{as}}{Y_s} \ln \frac{s-tp_t^*}{u} \,\middle|\, \mathcal{F}_s\right) du$$

$$= \int_K^{s-tp_t^*} P\left(\ln A(s,T) < \ln\left(\frac{e^{as}}{Y_s} \ln \frac{s-tp_t^*}{u}\right) \,\middle|\, \mathcal{F}_s\right) du$$

$$\approx \int_K^{s-tp_t^*} \Phi\left(\frac{\ln\left(\frac{e^{as}}{Y_s} \ln \frac{s-tp_t^*}{u}\right) - \alpha(s)}{\beta(s)}\right) du. \quad \blacksquare$$

Note that the price of the mortality call 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. The parameters were estimated from the Polish mortality table for men for the year 2003 (see [3]), where $A = -2.4366 \cdot 10^{-5}$, $B = 7.5436 \cdot 10^{-5}$ and $C = 0.0794$. Here $a = 0$. We price the options at issue time, i.e. $t = s$, and they mature at $T = 61$. For $t = s$ the strike price is $K(t) = E({}_{T-t}p_t)$ and $\overline{Y_t} = 1$. The interest rate r is zero. The exact values and the ratio $\frac{\text{Levy price}}{\text{exact price}}$ are summarized in Table 2. Figure 4 shows the price surfaces and the comparison between the exact Monte Carlo price and the approximate one.

Table 2. 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 can be 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–Scholes formula. The option price falls again for $b > 0.5$, which may be surprising. This is because for large b the price of a single underlying instrument falls and hence so does the derivative's price. 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) and for very short time to expiration (underestimates). However, in the latter case, the exact price of the option is close to zero so the systematic risk can anyway be neglected. Moreover, even in those cases the Levy-like approximation can be used as a first order approximation for the mortality call option price.

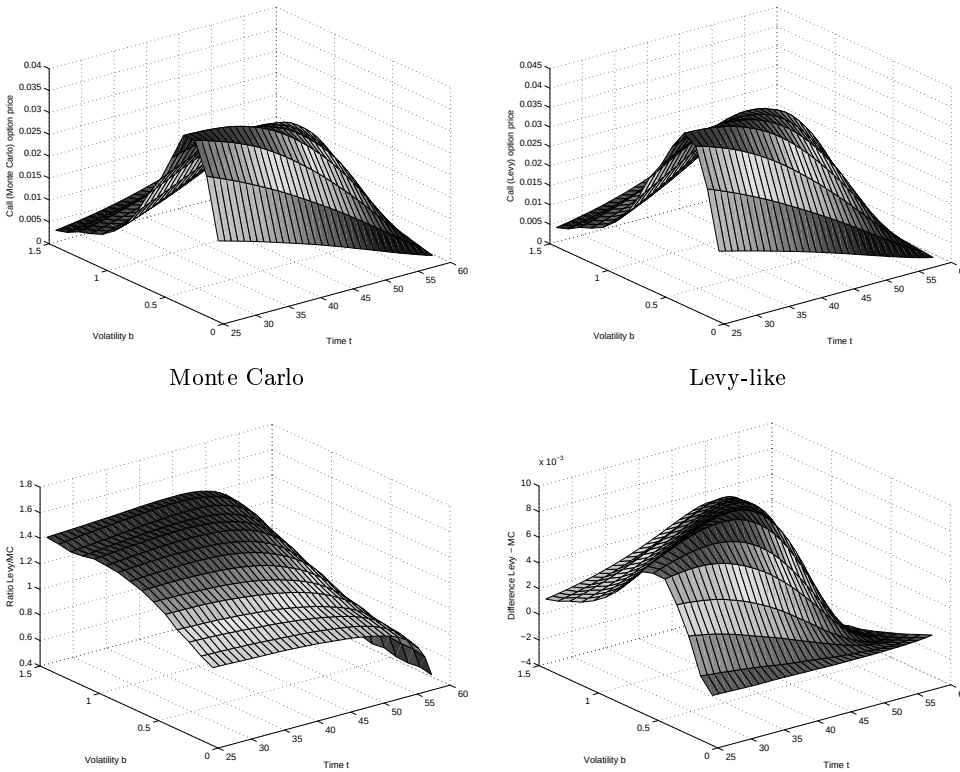

Fig. 4. Price of the mortality call option for different volatilities and issue times. Exact (Monte Carlo) results, Levy-like approximations and their comparison

5. Conclusion. We have proposed a few stochastic mortality models and proved them to fit the historical data relatively well. We have also shown how widely these models can be applied in life insurance. Mortality derivatives are good examples: they can help 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. Other applications include more reliable mortality projections and confidence intervals for future payments from a portfolio of risk in life insurance.

References

[1] N. Bowers, U. Gerber, J. Hickman, D. Jones and C. Nesbit, *Actuarial Mathematics*, The Society of Actuaries, 1997.

[2] D. Brigo and F. Mercurio, *Interest Rate Models*: *Theory and Practice*, Springer, Berlin, 2001.

[3] Central Statistical Office (Poland), *Mortality tables 1995–2003*, 2004, http://www.stat.gov.pl.

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

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

[6] C. Domański, *Properties and applications of the Shapiro–Wilk multi-dimensional normality test*, Cracow Univ. of Economics Rector's Lectures 37, 1998 (in Polish).

[7] L. Dothan, *On the term structure of interest rates*, J. Financial Econom. 6 (1978), 59–69.

[8] I. Gikhman and A. Skorokhod, *Introduction to the Theory of Random Processes*, W. B. Saunders, Philadelphia, 1969.

[9] J. Hull, *Options, Futures and Other Derivatives*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

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

[11] R. Lee and L. Carter, *Modelling and forecasting the time series of US mortality*, J. Amer. Statist. Assoc. 87 (1992), 659–671.

[12] E. Levy, *Pricing European average rate currency options*, J. Internat. Money Finance 11 (1992), 474–491.

[13] M. Lindbergson, *Mortality among the elderly in Sweden 1988–1997*, Scandinavian Actuarial J. 1 (2001), 79–94.

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

[15] M. Milevsky and S. Promislov, *Mortality derivatives and the option to annuitise*, Insurance Math. Econom. 29 (2001), 299–318.

[16] I. Nelken (ed.), *The Handbook of Exotic Options*, Irwin Professional Publ., 1996.

[17] A. Renshaw and S. Haberman, *Lee–Carter mortality forecasting with age-specific enhancement*, Insurance Math. Econom. 33 (2003), 255–272.

[18] —, —, *On the forecasting of mortality reduction factors*, ibid. 32 (2003), 379–401.

[19] P. Royston, *An extension of Shapiro and Wilk's W test for normality to large samples*, Appl. Statist. 31 (1982), 115–124.

Institute of Mathematics and Computer Science
Wrocław University of Technology
Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: jan.iwanik@pwr.wroc.pl

Received on 25.1.2007;
revised version on 6.2.2007 (1858)