Mortality risk via affine stochastic intensities: calibration and empirical relevance

LUCIANO, Elisa and VIGNA, Elena

2008

Mortality risk via affine stochastic intensities: calibration and empirical relevance *

Elisa Luciano† Elena Vigna‡

April 24, 2009

*We thank Emanuele Lucà for help with Java programming and an anonymous referee for useful comments. Any remaining errors are ours. A previous version of this paper was circulated under the title "Non mean reverting affine processes for stochastic mortality", 2005.

†University of Torino & ICER, Torino & Collegio Carlo Alberto, Moncalieri. Email: luciano at econ.unito.it. Tel. +39 011 670 5742. Fax. +39 011 670 5783.

‡Contact author. University of Torino & Collegio Carlo Alberto. Email: elena.vigna at econ.unito.it. Tel. +39 011 670 5754. Fax. +39 011 670 5783. Address: Dipartimento di Statistica e Matematica Applicata, corso Unione Sovietica 218 bis, 10134, Torino, Italy.

Mortality risk via affine stochastic intensities: calibration and empirical relevance

Abstract

In this paper, we address the mortality risk of individuals and adopt parsimonious time-homogeneous affine processes for their mortality intensities. We calibrate the models to different generations in the UK population and investigate their empirical appropriateness. We find that, in spite of their simplicity, non mean reverting processes with deterministic part that increases exponentially - which generalize the Gompertz law - seem to be appropriate descriptors of human mortality. The proposed models prove to fulfill most of the properties that a good model for stochastic mortality should have. Empirical results show that the generalization is worth exploring. Indeed, the variability of number of deaths may increase considerably due to the randomness of the mortality intensity. We show that the models are suitable for mortality forecasting and mortality trend assessment.

Keywords: stochastic mortality, affine processes, survival probability modeling, survival probability calibration.

JEL classification: G22, J11.

1 Introduction

The issue of mortality risk has been largely addressed in recent years. In the last decades significant improvements in the duration of life have been experienced in most developed countries. It is clear that continuous improvements in the mortality rates have to be allowed for when pricing insurance products that heavily depend on the duration of life at old ages, like annuities. Thus, a good model for mortality risk is a fundamental prerequisite for consistent pricing and reserving.

Traditionally, actuaries have been treating the demographic assumptions in a deterministic way. They have considered available mortality tables for describing the future evolution of mortality. More recently, stochastic models have been adopted to describe the uncertainty linked to mortality. Among these, an important stream of literature focuses on describing death arrival as the first jump time of a Poisson process with stochastic intensity. This approach, named doubly stochastic, is the one adopted in the current paper.

Within the doubly stochastic approach, we select and calibrate on UK mortality tables a number of models for the intensity of mortality. We explore those processes which lead to closed form survival probabilities, namely the affine ones. Among these, we further select time homogenous models: indeed, in the trade-off between parsimony of parameters and model flexibility or generality, we give priority to the first. We make sure that they are consistent with the criteria that a good model for mortality should have, as listed, for instance, by Cairns, Blake and Dowd (2006).

We investigate the empirical behaviour of these processes in three different ways. We simulate the number of deaths for a generation and compare it with the simulated number of deaths when the mortality is given by the Gompertz law. This permits to ascertain the impact of mortality volatility. We select and calibrate one of the models to do mortality forecasting within a given generation. We then show how the mortality trend can be captured by the behaviour of the simulated mortality intensity of different generations. The proposed models could also be used for pricing purposes, since they lend themselves nicely to a change of measure. However, we do not address the pricing issue in this paper.

The paper is organized as follows. Section 2 gives a brief review of the actuarial literature on mortality risk. Section 3 considers time-homogeneous affine processes for stochastic mortality. Section 4 discusses their calibration results and Section 5 their applications. Section 6 summarizes and outlines further research.

2 Modelling mortality risk

In this section, we make a brief review of the state of the art on mortality risk modelling. Two indicators are typically used to describe the mortality of an individual: the survival function and the death curve.

The survival function, denoted with $S(t)$, is defined as follows:

$$S(t) = P(T_0 > t) = 1 - F_{T_0}(t),$$

where T_0 is the random variable that describes the duration of life of a new-born individual, and F_{T_0} is its distribution function. The survival function indicates the probability that a new-born

individual will survive at least t years. Via the survival function, one can easily derive the distribution function of the duration of life of an individual aged x, given that he/she is alive at that age (see, for instance, Bowers, Gerber, Hickman, Jones and Nesbitt (1986), Gerber (1997)).

The death curve, ${}_{x/1}q_0$, is defined as follows:

$$ {}_{x/1}q_0 = \frac{S(x) - S(x+1)}{S(0)}, $$

and indicates the probability for a new-born individual of dying in year of age $[x, x+1]$.

An easy way of capturing the mortality trend observed in the past decades consists in looking at the graphs of the survival function and the death curves of a population in different years (for an accurate report about mortality trends, see Pitacco (2004a)). One can notice that the shape of the survival function becomes more and more "rectangular" and the mode of the death curve moves towards the right. The first phenomenon is known as rectangularization, the second as expansion. Rectangularization occurs since the volatility of the duration of life around the mode of death decreases, leading to lower dispersion of ages of death around the most likely age of death. Expansion takes place because the age when death is most likely to occur increases as time passes, due to improvements in economic and social conditions, medicine progresses etc.

The actuarial literature about modelling and forecasting mortality rates is vast and has a long history: for a detailed survey of the most significant models proposed in the literature, see for instance Pitacco (2004b).
Traditionally, a central role has been played by the "force of mortality", defined as the opposite of the derivative of the logarithm of the survival function:

$$ \mu_x = -\frac{d}{dx} \log S(x). $$

The force of mortality is a good tool for approximating the mortality of the individual at age x, since it can be shown that:

$$ P(x < T_0 \leq x + \Delta x | T_0 > x) = \mu_x \Delta x + o(\Delta x), \tag{2.1} $$

i.e. the probability of dying in a short period of time after x, between age x and age $x + \Delta x$, can be approximated by $\mu_x \Delta x$, when Δx is small. The force of mortality is normally increasing as x increases (there are some exceptions, in correspondence to very small values of x – due to the infant mortality – and values around 20-25 – due to the young mortality hump).

When allowing for mortality improvements over time, it is evident that the force of mortality has to show a dependence also on calendar year, and not only on age. Thus, the force of mortality can be described by a two variable function $\mu_x(y)$, where y indicates the calendar year. As time y increases and the age x remains fixed, the decreasing mortality rates over time translate into a decreasing function $\mu_x(y)$.

Several contributions have been proposed in the last decade in order to model and forecast the year- and age-dependent mortality, i.e. "dynamic mortality". One of the seminal works is the Lee-Carter method (Lee and Carter (1992) and Lee (2000)), that models an actuarial indicator similar to the force of mortality, the central death rate, as a two variable function. Many authors have modified the Lee-Carter method. Among these are the extensions proposed by Renshaw and

Haberman (2003) and Brouhns, Denuit and Vermunt (2002).

Another way of dealing with mortality improvements, largely adopted by insurance companies, is the use of the so-called "projected mortality tables", that incorporate (forecasts of) survival probabilities at any age for different calendar years.

Finally, a fairly recent stream of academic literature models the force of mortality as a stochastic process. Milevsky and Promislow (2001) have used a stochastic force of mortality, whose expectation at any future date – under an appropriate choice of the parameters – has a Gompertz specification. They investigate a so-called mean reverting Brownian Gompertz specification, with intensity h_t given by

$$h_t = h_0 e_t^{gt+\sigma \int_0^t e^{-b(t-u)} dW_u^h}, \tag{2.2}$$

with g, σ, b constant and the Brownian motion W uni-dimensional.

Dahl (2004), Biffis (2005), Denuit and Devolder (2006) and Schrager (2006) in modelling the stochastic force of mortality have applied the same mathematical tools used in the credit risk literature to model the time to default. Under this setting, the remaining lifetime of an individual, τ, is a doubly stochastic stopping time with intensity λ. At an intuitive level, conditionally on λ, the jump process driving mortality is a Poisson process. However, λ is itself a stochastic process. We refer the interest reader to Brémaud (1981) and Duffie (2001) for a thorough treatment of the mathematical framework. It is possible to show that if $\tau \geq t$ is a stopping time doubly stochastic with intensity λ, then:

$$P(\tau > s|\mathcal{G}_t) = E\left[e^{-\int_t^s \lambda(u)du}|\mathcal{G}_t\right], \tag{2.3}$$

where $\mathcal{G}_t$ describes the information at time t. Readers who are familiar with Mathematical Finance can easily see in the r.h.s. of equation (2.3) the price at current time t of a unitary default-free zero-coupon bond with maturity at time $s > t$, if the short-term interest rate model is given by the process λ. All the literature about interest rate models can thus be retrieved in this setting[1]. It turns out that it is convenient to specify the stochastic intensity λ as an affine process, whose dynamics is given by the stochastic differential equation (SDE):

$$d\lambda(t) = f(\lambda(t))dt + g(\lambda(t))d\tilde{W}(t) + dJ(t), \tag{2.4}$$

where $\tilde{W}$ is an n-dimensional Brownian motion, J is a pure jump process and where the drift $f(\lambda(t))$, the covariance matrix $g(\lambda(t))g(\lambda(t))'$ and the jump measure associated with J have affine dependence on $\lambda(t)$. Interest readers can find a thorough treatment of affine processes in Duffie, Filipovič and Schachermayer (2003). The convenience of adopting an affine process in modelling the intensity lies in the fact that, under technical conditions (see Duffie and Singleton (2003)), it yields:

$$E\left[e^{\int_t^T -\lambda(u)du}|\mathcal{G}_t\right] = e^{\alpha(T-t)+\beta(T-t)\lambda(t)}, \tag{2.5}$$

where the coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalized Riccati ordinary differential equations (ODEs). The latter can be solved at least numerically and in some cases analytically. Therefore, the problem of finding the survival function (2.3) becomes tractable, whenever affine processes for $\lambda(t)$ are employed. In this setting, Dahl (2004) selects an extended Cox-Ingersoll-Ross (CIR) process, i.e. a time-inhomogeneous process λ, reverting to a deterministic function of time

$$d\lambda_{x+t} = (\beta^\lambda(t,x) - \gamma^\lambda(t,x)\lambda_{x+t})dt + \rho^\lambda(t,x)\sqrt{\lambda_{x+t}}dW_t, \tag{2.6}$$

[1]Please notice that in bond pricing the expectation is taken under the risk neutral measure, while in this context it is appropriate to use the historical measure.

where x is the initial age.

Biffis (2005) chooses two different specifications for the intensity process. In the first one, the intensity λ_t is given by a deterministic function of time, $m(t)$, plus a mean reverting jump diffusion process Y_t, with dynamics given by the SDE

$$dY_t = \gamma(\overline{y}(t) - Y_t)dt + \sigma dW_t - dJ_t. \tag{2.7}$$

In the second one, which is a two factor model, the intensity λ_t is a CIR-like process, mean reverting to another process $\overline{\lambda}_t$. The dynamics of the two processes are given by

$$d\lambda_t = \gamma_1(\overline{\lambda}_t - \lambda_t)dt + \sigma_1\sqrt{\lambda_t}dW_t^1 \tag{2.8}$$

$$d\overline{\lambda}_t = \gamma_2(m(t) - \overline{\lambda}_t)dt + \sigma_2\sqrt{\overline{\lambda}_t - m^*(t)}dW_t^2. \tag{2.9}$$

Denuit and Devolder (2006) introduce various stochastic models for the force of mortality. Their models are continuous-time versions of the Lee Carter model. It should be noted that the more general form of their models includes a mean reversion effect to a limit table.

Schrager (2006) proposes an M-factor affine mortality model, whose general form is given by

$$\lambda_x(t) = g_0(x) + \sum_{i=1}^{M} Y_i(t)g_i(x), \tag{2.10}$$

where the factors Y_i are mean reverting.

3 Time-homogeneous affine processes

Consistently with the doubly stochastic approach setup, uncertainty is described by a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\{\mathcal{G}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions. We consider an individual aged x at time 0 and model his/her random future lifetime T_x as a doubly stochastic stopping time with intensity λ_x driven by the sub-filtration $\{\mathcal{F}_t : t \geq 0\}$, where $\mathcal{F}_t \subset \mathcal{G}_t$. Roughly speaking, this means that, for any time $t < s$, conditionally on the information generated by $\mathcal{G}_t \cup \mathcal{F}_s$, T_x is the first jump time of a non explosive counting process N that has a Poisson distribution with parameter $\int_t^s \lambda_u du$. Intuitively, the counting process N may be seen as a process that jumps whenever the individual dies: $N_t = 0$ if $t < T_x$, $N_t = 1$ if $t \geq T_x$.

We notice that, when t changes, the process $\lambda_x(t)$ describes the future intensity of mortality for any age $x + t$ of an individual aged x at time 0. In other words, λ captures the mortality intensity for a particular generation and a particular initial age [2].

As in the works mentioned in the previous section, we select an affine process for the intensity λ_x. Then, according to (2.3) and (2.5) the survival probability is:

$$S_x(t) = P(T_x > t|\mathcal{G}_0) = E\left[e^{-\int_0^t \lambda_x(u)du}|\mathcal{G}_0\right] = e^{\alpha(t)+\beta(t)\lambda_x(0)}, \tag{3.1}$$

where $\alpha(t)$ and $\beta(t)$ are the solutions to Riccati ODEs.

[2] This has to be allowed for when choosing the mortality table: the approach adopted here is a "diagonal" one.

The choice of λ is crucial. We are interested in the practical implementation of the model. Notice that if one chooses time-homogeneous processes, the calibration to actual data can be performed through standard procedures, such as maximum likelihood (ML). We also have in mind the need for coupling mortality modelling with interest-rate modelling for pricing insurance contracts such as annuities and longevity bonds. For these reasons, in what follows we select time-homogeneous affine processes for the mortality intensity. In addition, we observe the following:

- Cairns et al. (2006) in their list of properties that a good model for stochastic mortality should have, rule out strong mean reversion, even to a time-dependent target;
- Luciano and Vigna (2005) find that time-homogeneous mean reverting affine processes fail to fit observed mortality tables;
- the force of mortality observed and/or extrapolated from the mortality tables does not seem to present a mean reverting behaviour, but rather an exponential one.

Thus, we restrict our attention to non mean reverting processes, with deterministic part that increases exponentially. We notice that the exponential growth is the main feature of the Gompertz model, that is here extended in a natural way. Three affine models with the desired characteristics are presented and discussed below.

3.1 The Ornstein Uhlenbeck process

The first candidate for describing the intensity $\lambda_x(t)$ is an Ornstein Uhlenbeck process (from now on, we omit the initial age x for convenience), with SDE:

$$OU \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma dW(t), \tag{3.2}$$

with $a > 0$ and $\sigma \geq 0$.
By solving it, we get to the following expression for the intensity:

$$\lambda(t) = \lambda(0)e^{at} + \sigma \int_0^t e^{a(t-s)} dW(s). \tag{3.3}$$

By applying standard results on linear SDEs (see, for instance, Arnold (1974)) to the process (3.3) we have that $\lambda(t)$ is normally distributed with mean

$$E(\lambda(t)) = \lambda(0)e^{at},$$

and variance

$$Var(\lambda(t)) = \sigma^2 \cdot \frac{e^{2at} - 1}{2a}.$$

The main drawback when choosing this process for the intensity is that it becomes negative with positive probability. The probability that $\lambda(t)$ takes negative values is:

$$P(\lambda(t) \leq 0) = P\left(\lambda(0)e^{at} + \sigma\sqrt{\frac{e^{2at}-1}{2a}}\mathcal{N} \leq 0\right) = P\left(\mathcal{N} \leq -\frac{\lambda(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}\right) = \Phi(\zeta(\sigma, a)),$$

with

$$\zeta(\sigma, a) = -\frac{\lambda(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}},$$

where $\mathcal{N} \sim \mathcal{N}(0,1)$ and Φ is its distribution function. It turns out that the function $\zeta(\cdot,\cdot)$ is an increasing function of σ and a decreasing function of a, and so is the probability of negative values of λ. In practical applications to mortality modelling this probability tends to be very small, since the relevant values of σ and a are respectively small and high enough. We will come back to this point later, when presenting the calibrations.

The survival function is given by (3.1), where the functions α and β solve the system of ODEs (see Duffie, Pan and Singleton (2000) pagg. 1350–1351):

$$\begin{cases} \alpha'(t) = \frac{1}{2}\sigma^2\beta^2(t) \\ \beta'(t) = -1 + a\beta(t) \end{cases} \tag{3.4}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0. \tag{3.5}$$

By solving the system 3.4–3.5, we find that the functions $\alpha(t)$ and $\beta(t)$ are:

$$\begin{cases} \alpha(t) = \frac{\sigma^2}{2a^2}t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} \\ \beta(t) = \frac{1}{a}\left(1 - e^{at}\right). \end{cases} \tag{3.6}$$

We observe that with a strictly positive value of σ, the survival probability is decreasing for $t < T^*$ and increasing for $t > T^*$, where

$$T^* = \frac{1}{a}\ln\left[1 + \frac{a^2\lambda(0)}{\sigma^2}\left(1 + \sqrt{1 + \frac{2\sigma^2}{a^2\lambda(0)}}\right)\right]. \tag{3.7}$$

In addition, the probability of surviving forever tends to infinity. These unrealistic and undesirable features are due to the fact that the survival intensity can take negative values with positive probability. Thus, from a purely theoretical point of view, the Ornstein Uhlenbeck model can be considered inadequate to describe the intensity of mortality. However, it can be seen that in the applications this model turns out to be rather appropriate, since the calibrated values make T^* very large with respect to human survivorship and therefore make the survival probability a decreasing function of age. As a consequence, the point after which the survival probability tends to infinity is far greater than experienced human survivorship. Furthermore, as seen above, also the probability of negative values of λ turns out to be negligible with the calibrated parameters. Thus, once calibrated, the model is biologically reasonable and seems to be useful with respect to practical applications by actuaries: for instance, a modified version of this model has been used by Menoncin (2008).

3.2 The Ornstein Uhlenbeck process with jumps

In the second model we add a jump component in the stochastic part of the mortality process. The inclusion of jumps in a mortality model is motivated by the fact that the intensity of mortality can suddenly change (either improving or worsening) due to unexpected external events that can affect it. A priori, jumps can be of positive size, in which case the mortality worsens (in the case of wars, for instance), or of negative size, in which case mortality improves (in the case of medicine breakthroughs). Therefore, the process λ is given by:

$$OUj \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma dW(t) + dJ(t), \tag{3.8}$$

where J is a pure compound Poisson jump process, with Poisson arrival times of intensity $l > 0$ and exponentially distributed jump sizes with mean η. We assume independence between the Brownian motion W and the Poisson process. As in the case without jumps, the intensity can in principle become negative, even though its distribution is not Gaussian any longer.

We find a closed form solution provided that (see Duffie et al. (2000)):

$$\beta(t) < \frac{1}{\eta} \quad if \quad \eta > 0, \quad \beta(t) > \frac{1}{\eta} \quad if \quad \eta < 0. \tag{3.9}$$

Under these technical conditions, we have to solve the following system of ODEs for α and β:

$$\begin{cases} \alpha'(t) = \frac{1}{2}\sigma^2\beta^2(t) + l\frac{\eta\beta(t)}{1-\eta\beta(t)} \\ \beta'(t) = -1 + a\beta(t) \end{cases} \tag{3.10}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0. \tag{3.11}$$

The equation for β is the same as before (3.4), so is the solution. The solution for α is instead different (due to the inclusion of the jump component), and we have[3]:

$$\begin{cases} \alpha(t) = (\frac{\sigma^2}{2a^2} + \frac{l\eta}{a-\eta})t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} - \frac{l}{a-\eta}\ln(1 - \frac{\eta}{a} + \frac{\eta}{a}e^{at}) \\ \beta(t) = \frac{1}{a}\left(1 - e^{at}\right). \end{cases} \tag{3.12}$$

Observe that if $\eta > 0$, condition (3.9) is always satisfied, since $\beta(t) < 0$. On the other hand, if $\eta < 0$, the condition has to be checked. In both cases, it guarantees that the argument of the logarithm in (3.12) is positive.

Also in this model the survival probability can be an increasing or decreasing function of t. A necessary and sufficient condition for it to be decreasing is

$$\frac{1}{2}\sigma^2\beta(t)^2 + \frac{l\eta\beta(t)}{1-\eta\beta(t)} + a\lambda(0)\beta(t) - \lambda(0) < 0. \tag{3.13}$$

A sufficient condition for (3.13) in turn is that:

$$\frac{\ln(1-\beta_2 a)}{a} < t < \frac{\ln(1-\beta_1 a)}{a} \tag{3.14}$$

for $\eta > 0$ and

$$t > \frac{\ln(1-\beta_2 a)}{a} \quad \vee \quad t < \frac{\ln(1-\beta_1 a)}{a} \tag{3.15}$$

for $\eta < 0$, where

$$\beta_1 = \frac{(\sigma^2 - 2\eta a\lambda(0)) - \sqrt{(\sigma^2 + 2\eta a\lambda(0))^2 + 8\eta^2\sigma^2(\lambda(0)+l)}}{2\eta\sigma^2}$$

and

$$\beta_2 = \frac{(\sigma^2 - 2\eta a\lambda(0)) + \sqrt{(\sigma^2 + 2\eta a\lambda(0))^2 + 8\eta^2\sigma^2(\lambda(0)+l)}}{2\eta\sigma^2}.$$

[3]We thank Qin Shang for spotting a typo in a previous version of the paper.

3.3 The Feller process

The third model proposed is the Feller process:

$$FEL \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma\sqrt{\lambda(t)}dW(t), \tag{3.16}$$

where $a > 0$ and $\sigma \geq 0$.

The solution $\lambda(t)$ of the SDE (3.16) is

$$\lambda(t) = \lambda(0)e^{at} + \sigma \int_0^t e^{a(t-u)}\sqrt{\lambda(u)}dW(u) \tag{3.17}$$

and its distribution can be obtained following Feller (1951).

The application of the affine framework gives the following system of ODEs for α and β:

$$\begin{cases} \alpha'(t) = 0 \\ \beta'(t) = -1 + a\beta(t) + \frac{1}{2}\sigma^2\beta^2(t) \end{cases} \tag{3.18}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0. \tag{3.19}$$

The solution is:

$$\begin{cases} \alpha(t) = 0 \\ \beta(t) = \frac{1-e^{bt}}{c+de^{bt}} \end{cases} \tag{3.20}$$

with:

$$\begin{cases} b = -\sqrt{a^2 + 2\sigma^2} \\ c = \frac{b+a}{2} \\ d = \frac{b-a}{2} \end{cases} \tag{3.21}$$

The main advantage of this process w.r.t. the previous ones is that it does not violate the non-negativity constraint of the intensity, provided that the starting point is nonnegative. It must be said that the intensity process can reach 0 and stay there with positive probability. Nevertheless, the probability of such an event is negligible in the practical applications. Furthermore, this unfortunate event has never occurred in the simulations run with the calibrated parameters. Given that the coefficients b, c, d are negative, the survival probability is always decreasing in t - a biologically reasonable feature - if and only if

$$e^{bt}(\sigma^2 + 2d^2) > \sigma^2 - 2dc. \tag{3.22}$$

Notice that (3.22) is automatically satisfied if $\sigma^2 - 2dc < 0$, a condition which holds in our calibrations. Thus, in the calibrated model the survival probability is decreasing at every age. This is a second advantage of the model w.r.t. the OU and the OUj[4]. An example of application of this process can be found in Luciano, Spreeuw and Vigna (2008).

[4]It must be said also that for $t \to +\infty$ the survival probability tends to $e^{\frac{1}{c}}$, which in our applications turns out to be of the order of e^{-1000} or less. However, notice that this unrealistic feature would not be in contrast with some discoveries in the biological field that the force of mortality of fruit-flies increases with age up to a certain point, then reaches a plateau and may even decrease (see Thatcher (1999)).

3.4 Impact of mortality randomness

In this section, we investigate the effect on survival probabilities of the mortality intensity randomness. To this aim, it is useful to analyze the relationship between the stochastic intensity of mortality and the deterministic force of mortality. Recall that the force of mortality μ_x at age x is defined as

$$\mu_x = \lim_{h\to 0} \frac{P(x < T_0 \leq x+h | T_0 > x)}{h}.$$

In our case, we have:

$$\mu_x = \lim_{h\to 0} \frac{1}{h}\left(1 - \frac{S(x+h)}{S(x)}\right) = \lim_{h\to 0} \frac{\alpha(x) - \alpha(x+h) + \lambda_0(0)(\beta(x) - \beta(x+h))}{h} = -\alpha'(x) - \lambda_0(0)\beta'(x).$$

In the OU model, the force of mortality becomes:

$$\mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax}-1)^2. \tag{3.23}$$

In the OUj model we have:

$$OUj \qquad \mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax}-1)^2 - \frac{l}{a-\eta}\left(1 - \frac{a\eta e^{ax}}{a-\eta+\eta e^{ax}}\right). \tag{3.24}$$

In the FEL model we have:

$$\mu_x = \frac{4\lambda_0(0)b^2 e^{bx}}{[(a+b)+(b-a)e^{bx}]^2}. \tag{3.25}$$

Observe that in all cases, if $\sigma = 0$ and $\eta = l = 0$ (i.e. if the intensity is deterministic) we have:

$$\mu_x = \lambda_0(0)e^{ax} = \lambda_0(x),$$

i.e. the force of mortality at age x coincides with the intensity of mortality for a new born individual after x years. Furthermore, the force of mortality is of the Gompertz type [5]. However, the coincidence between intensity of mortality and force of mortality is clearly no longer true when the intensity is truly stochastic.

It is easy to show, by application of the Jensen inequality to the survival function, that a higher randomness in the stochastic intensity produces an improvement in the survival probabilities. Practically, the force of mortality in the OU and the OUj models (see eqs. (3.23) and (3.24)) decreases when σ or l increase: therefore, the survival probability increases when the stochastic component increases[6]. In the FEL case it can be shown that the function β of equation (3.20) is increasing in σ. This implies that the survival probability increases with the diffusive part also in the FEL model. Therefore, ceteris paribus, the force of mortality of the stochastic models is lower than its deterministic counterpart for any initial age.

4 Calibration

We have calibrated the three processes to different generations in the UK population. The mortality tables selected for the calibration are two observed generation tables, for males born in 1885 and in

[5] Indeed, if $\sigma = 0$ the evolution of $\lambda_0(t)$ is deterministic and coincides with the Gompertz specification.

[6] As an alternative, one can observe that the function α of equation (3.6) is increasing in σ and that the function α of equation (3.12) is increasing in both σ and l.

1900 respectively, and two projected mortality tables, for males born in 1935 and 1945 respectively. Initial age has been chosen equal to 65. The data relative to the observed mortality tables are taken from the Human Mortality Database (University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (2002), data downloaded on August 10, 2004). Those for the projected tables are taken from the Standard tables of mortality 1992 for UK immediate annuitants, IML92 (Institute and Faculty of Actuaries (1990)).

The death rates as well as the number of deaths at each age have been derived by the mortality tables mentioned above. In fitting the table, we have used the maximum likelihood (ML) procedure. For each given process, the corresponding log-likelihood has been maximized by numerical selection of the model parameters. The standard errors of the parameters cannot be computed because the number of exposed to risk on which these mortality tables are based is not known (see also Thatcher (1999)). However, to assess non-asymptotic robustness and reliability of the results, we have calibrated the values of the parameters also by minimization of the mean square error[7]. The values of the parameters from the two procedures are almost identical. We have also checked the conditions that make the survival function decreasing at least up to age 120, i.e. condition (3.7) for the OU model, (3.13) for the OUj and (3.22) for the FEL.

In the calibration, the optimization of the mean jump size, η, is performed on negative values only. The choice of a negative jump size is motivated by the expectation of sudden improvements in the intensity of mortality: jumps should correspond to discontinuity points of the intensity process, that can be related, for instance, to medicine progresses. Negative jumps in the intensity process render positive the probability that the intensity becomes negative. This inconvenient is also observed by Biffis (2005). However, in practical applications and calibrations the jump size and the frequency result so small that the probability of negative values can be considered negligible.

Table 1 reports for each intensity process the maximized log-likelihood, the mean square error and the optimal values of the parameters, as well as the corresponding initial value of $\lambda, \lambda_{65}(0)$ (which has been chosen equal to $-\ln(p_{65})$).

	1885	1900	1935	1945
$\lambda_{65}(0)$	0.03918	0.03739	0.01145	0.00885
$OU - MaxLoglikelihood$	-3.33277	-3.35984	-3.56148	-3.55359
$OU - error$	0.00043	0.00012	0.00085	0.00027
$OU - a$	0.0765	0.0792	0.0995	0.1094
$OU - \sigma$	0.00103	0.00335	0.0003	0.0007
$OUj - MaxLoglikelihood$	-3.33276	-3.36061	-3.56148	-3.55392
$OUj - error$	0.0001	0.00004	0.00002	0.00016
$OUj - a$	0.07653	0.07478	0.09933	0.10746
$OUj - \sigma$	0.00071	0.0001	0.0001	0.00045
$OUj - l$	0.00025	0.001	0.001	0.00098
$OUj - \eta$	-0.0015	-0.00003	-0.00003	-0.00003
$FEL - MaxLoglikelihood$	-3.33275	-3.3598	-3.56148	-3.55358
$FEL - error$	0.00044	0.00012	0.00084	0.00027
$FEL - a$	0.076	0.08	0.1	0.11
$FEL - \sigma$	0.0006	0.016	0.002	0.007

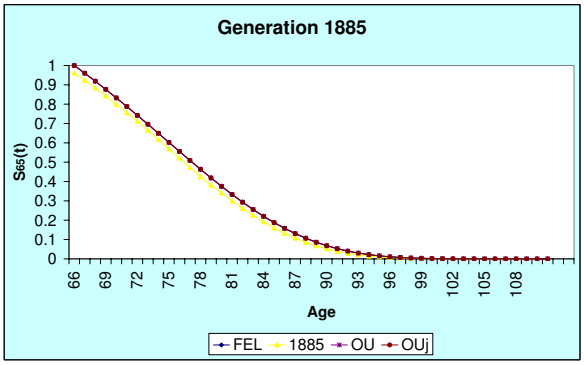
TABLE 1. Maximized log-likelihood, mean square errors and parameters of the calibrated processes.

[7]By mean square error, we mean the sum of the squares of the differences between the survival probabilities of the table and the model ones, divided by the number of data.

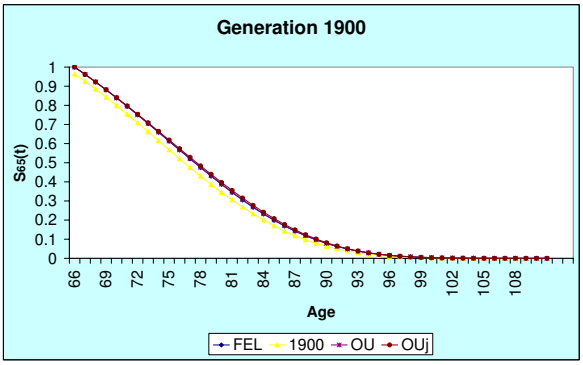
In all models, the mean square errors are very low, indicating a satisfactory fit to the relevant mortality tables. The OUj model fits always better than the OU model and in three cases out of four fits also better than the FEL model. Finally, we notice that with these values of the parameters the probability of negative intensity for the OU model can be considered negligible for all practical applications.

The change in the parameters $\lambda_{65}(0)$, which is the observed mortality intensity at age 65, and a, which is the relative rate of increase of the force of mortality of the underlying Gompertz model, gives a measure of the mortality trend, as captured by the models. When considering younger generations, $\lambda_{65}(0)$ decreases from 0.04 to 0.0088, while a increases from 0.07 to 0.11 (with the exception of the OUj model, generations 1885-1900). These results match very closely those found by Thatcher (1999).

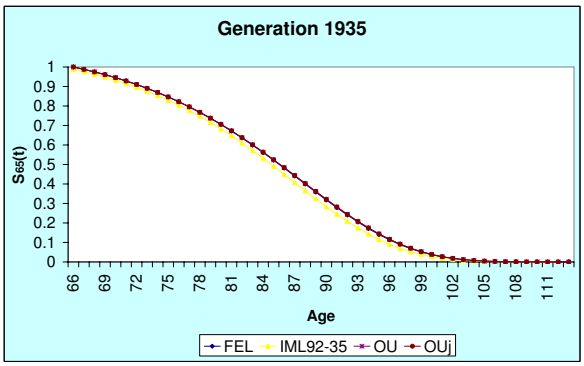
Graphs 1, 2, 3 and 4 report, for the different generations, the survival function of the three processes analyzed (OU, OUj and FEL) and that of the corresponding mortality table.



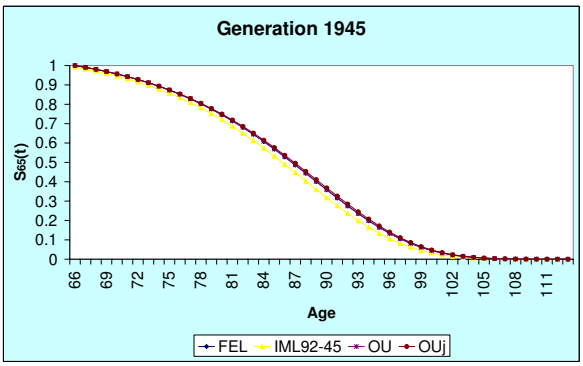
Graph 1. Survival functions, generation 1885.



Graph 2. Survival functions, generation 1900.



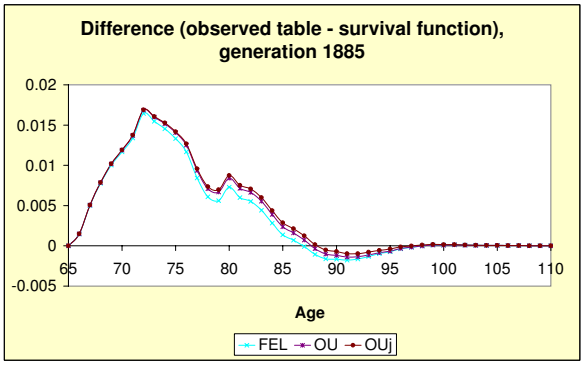
Graph 3. Survival functions, generation 1935.



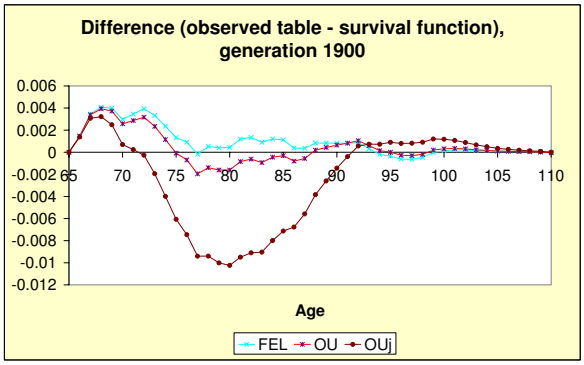
Graph 4. Survival functions, generation 1945.

The fit is very good and all the survival functions cannot be distinguished from each other. This happens also with the last two generations, indicating that these models capture the rectangularization phenomenon. To conclude, let us plot in Graphs 5, 6, 7 and 8 the calibration errors, i.e.

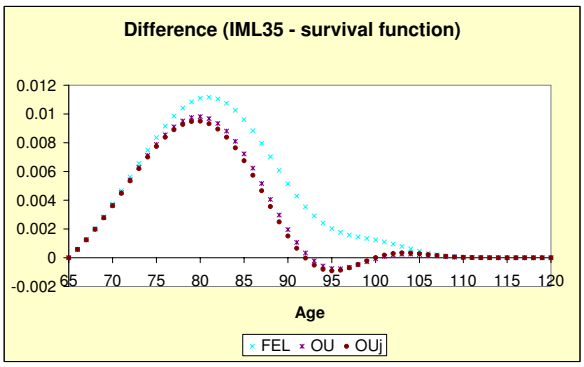
the differences between the observed survival probabilities ($_tp_{65}$) and their theoretical counterparts ($S_{65}(t)$) for all the generations.



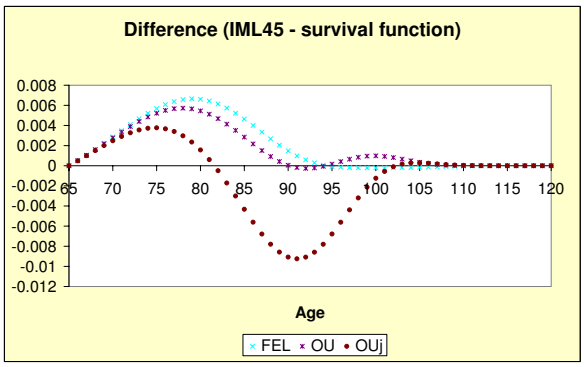
Graph 5. Calibration error, generation 1885.



Graph 6. Calibration error, generation 1900.



Graph 7. Calibration error, generation 1935.



Graph 8. Calibration error, generation 1945.

In general, the three models give similar errors for each age. It was already known from Table 1 that the global mean square error and the maximized likelihood were close to each other for all models. Graphs 5, 6, 7 and 8 confirm this result for each age. We believe that this result allows one to use interchangeably the three models[8].

4.1 Discussion of appropriateness

The aim of this section is to discuss pro and contras of the models proposed. Cairns et al. (2006) report a list of criteria that a good mortality model should meet:

[8]The difference between $_tp_{65}$ of the observed table and $S_{65}(t)$ of each model is positive for $t \leq 10$ approximately, negative between $t = 10$ and $t = 20$ approximately, then again positive for $t \geq 20$. This means that, in the case considered here, the fitted survival probabilities, in comparison with the basic table (on which the calibration is done), underestimate the survival probabilities between ages 65 and 75, overestimate them between ages 75 and 85 and underestimate them again after age 85. These considerations become quite important whenever the model were to be used for pricing purposes (under the assumption of no stochastic mortality risk premium): for example, underestimation of the survival probability between ages 65 and 75 would lead to lower than needed premiums for pure endowment policies with duration 10 years, sold to an individual aged 65, and premiums higher than needed for term assurances with the same duration sold to the same individual.

- the model should be consistent with historical data: the calibrations show that our models meet this criterium;
- the force of mortality should keep positive: our models do not meet this criterium; however, the probability of negative values of the intensity is negligible for practical applications;
- long-term future dynamics of the model should be biologically reasonable: our models meet this criterium, as the calibrated parameters fulfill the conditions for biological meaningfulness stated in section 3;
- long-term deviations in mortality improvements should not be mean-reverting to a pre-determined target, even if the target is time-dependent: our models meet this criterium by construction;
- model should be comprehensive enough to deal appropriately with pricing valuation and hedging problem: our models meet this criterium, since it is straightforward to extend them in order to deal with pricing, valuation and hedging problems;
- it should be possible to value mortality linked derivatives using analytical methods or fast numerical methods: our models meet this criterium, as they produce survival probabilities in closed form and with a very small number of parameters.

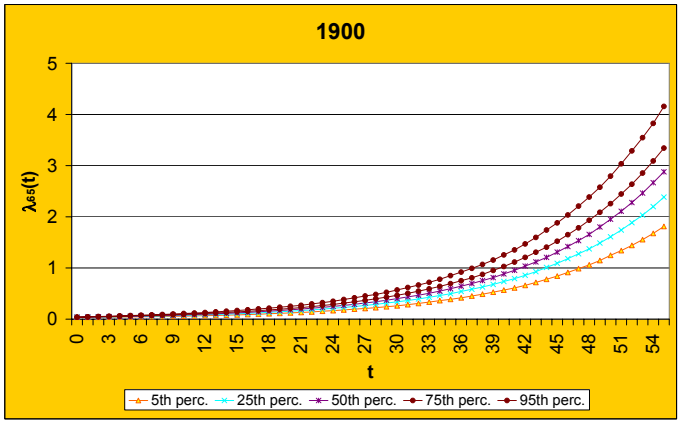
Cairns et al. (2006) add that no one of the previous criteria dominates the others. Consistently with their view, we support the validity of our models, which meet five criteria out of six. As every choice, our comes to a cost: the violation of the second criterium above is the price that we pay in order to have a simple and parsimonious model. In addition, the survival functions are given in closed form and depend on a very small number of parameters, which simplifies the calibration procedure enormously.

5 Empirical relevance

5.1 Simulated number of deaths

Optimally fitted models present low diffusion parameters. This feature is evident both in the observed mortality tables and in the projected tables. This section aims at showing that also relatively small volatility values can produce significant effects on the intensity and on the number of deaths.

To this aim, we have simulated the process $\lambda_{65}(t)$ for the generation 1900 using the calibrated parameters of the FEL model (i.e. $a = 0.08$ and $\sigma = 0.016$). We have simulated 100,000 paths of $\lambda_{65}(t)$ after having discretized each year into monthly intervals. The $5^{th}, 25^{th}, 50^{th}, 75^{th}$ and 95^{th} percentiles of the paths are reported in figure 9. The impact of volatility can be appreciated especially at old ages.



Graph 9. Percentiles of the simulated intensity for generation 1900 (FEL model).

In order to have a better understanding of the practical consequences of the volatility of the intensity, we have considered the simple case of a portfolio of 1000 ten-years term assurance policies sold to males aged 65 born in 1900. The 100,000 paths of the intensity process simulated above have been used to simulate the number of deaths within ten years in 100,000 different scenarios[9]. In order to assess the impact of volatility of the intensity process, we have simulated the number of deaths also in the case of deterministic intensity, i.e. in the case $\sigma = 0$. For consistent comparisons, the stream of pseudo random number used for simulating the death of each individual in each scenario is the same in the two cases. Table 2 reports some statistics of the distribution of the number of deaths in both cases.

	σ=0.016	σ=0
Mean	438.54	442.71
St.Dev.	50.69	15.76
Min	241	376
5th perc.	355	417
25th perc.	404	432
50th perc.	439	443
75th perc.	473	453
95th perc.	522	469
Max	648	512

TABLE 2. Statistics of simulated number of deaths: comparison between stochastic and deterministic case.

Not surprisingly, the average number of deaths is approximately the same. However, as expected, the variability of the number of deaths is substantially higher when the intensity of mortality is

[9]The methodology used for simulating the number of deaths is based on the following property: the first jump of the doubly stochastic process (in our case the death of the individual) occurs as soon as the integral of the intensity reaches a random level, which is distributed exponentially with parameter 1. In each scenario, we have simulated for each individual the realization of the exponential random variable, and then counted the number of deaths.

stochastic than in the deterministic case. Namely, the standard deviation of the number of deaths boosts up by more than 200%. To have a better understanding of the practical implications of this higher variability, let us observe that in 90% of scenarios the number of deaths varies between 417 and 469 when the intensity is deterministic, whereas it varies between 355 and 522 when the intensity is stochastic. This significant difference would have an impact on the pricing of the term-insurance policy, whenever the dispersion of number of deaths around the mean would be taken into account in the premium calculation. Based on this evidence, we believe that the deterministic Gompertz law would not have described accurately the mortality of this generation. Furthermore, it would not be appropriate for any generation with sufficiently high value of the diffusion coefficient[10].

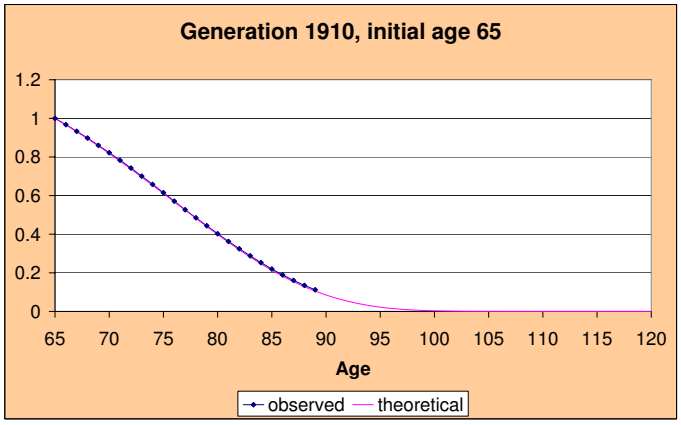
5.2 Forecasting mortality and mortality trend

The calibrated models can also be used for mortality extrapolation or future mortality prediction. There are two different ways to look at future mortality: the first one is *within* a certain generation, the second one is *between* different generations. With the first method, given a generation, one can calibrate the intensity process on the observed data and then forecast the evolution of the survival function in the future by considering its right tail after the last observation. With the second method, one can consider how the different calibrated parameters of the intensity process change when changing generation: in such a way one can consider the mortality trend (or "cohort effect"). We will differentiate between these two different methods by calling them *forecasting mortality* and *mortality trend*, respectively. For both purposes, typically one has to calibrate the survival curve when the members of a generation are not all dead. The ideal set of data one needs in order to make a calibration of the model to a relatively young cohort is a generation mortality table until the observation date. For example, if the calibration is done in 2005 and the generation under consideration is the one born in 1905, the data needed are the observed mortality rates of this particular generation for 100 years. Unfortunately, generation tables are typically available only for generations whose members are all dead. However, one can extrapolate the desired data by first collecting in a unique table all the observed mortality rates year by year (i.e. contemporaries tables) from 1905 to 2005 and then taking the diagonal starting from q_0 in 1905 to q_{100} in 2005. This procedure does not give exactly the mortality rates of a certain generation observed throughout life, but is considered a good approximation. Furthermore, it is feasible because one can easily have access to observed mortality rates year by year – for example, the Human Mortality Database mentioned above is a convenient database that provides yearly data for many countries dating back to the last century. The next step consists in calibrating the diagonal data. It is clear that the younger the generation, the lower the number of observed survival probabilities on which we make the calibration. However, the initial age x can be lowered in order to produce a sufficiently high number of data.

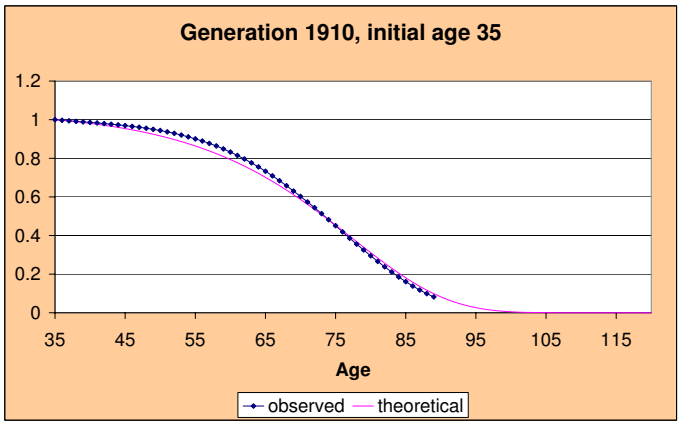
5.2.1 Mortality forecasting

As an illustration, Graphs 10 and 11 report the mortality forecast for the generation 1910 with initial ages 65 and 35. The graphs show the observed and the theoretical survival function according to the FEL model.

[10] Simulation results on the number of deaths not reported here show that the difference with the deterministic case can be appreciated in the FEL model whenever the values of σ are higher than 0.003.



Graph 10. Observed and theoretical survival function, generation 1910, initial age 65 (FEL model).



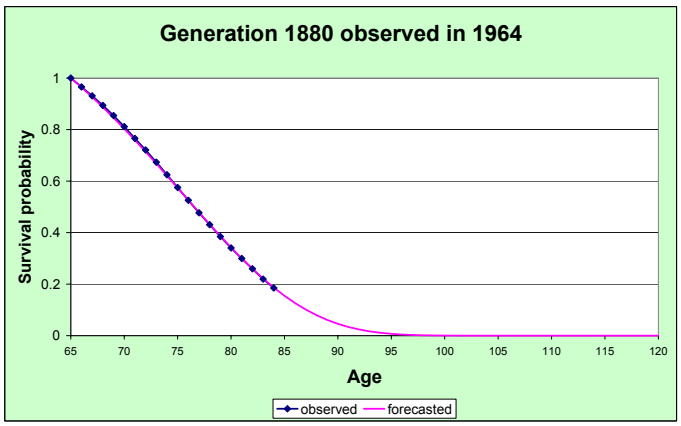
Graph 11. Observed and theoretical survival function, generation 1910, initial age 35 (FEL model).

In the graphs, the right tail of the "theoretical" curves give the forecast of the survival functions after the observation date implied by the FEL model for the same generation. The two curves are different, since both $a_{35} \neq a_{65}$ and $\sigma_{35} \neq \sigma_{65}$.

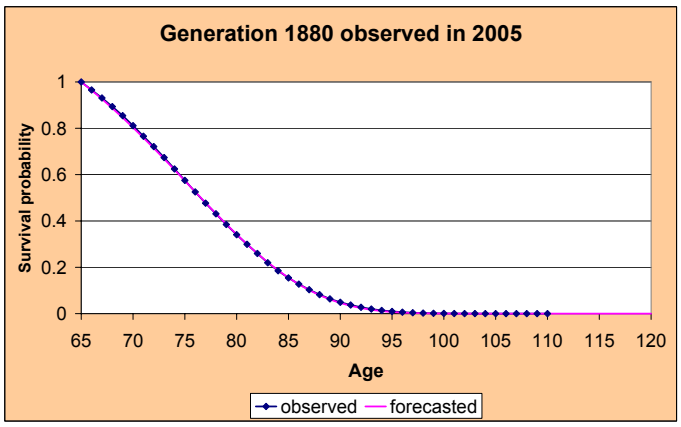
In order to check whether this forecast procedure gives reliable results (by comparing the forecasted mortality with the experienced one), we have applied it on the generation 1880, initial age 65. We have calibrated the parameters of the process taking year 1964 as observation date. It must be specified that in this case the simple ML procedure is not applicable, because the survival data observable in 1964 for the generation 1880 were not complete. For this reason, here the calibration procedure used is the minimization of the mean square error[11]. Graph 12 reports the forecasted and

[11]We are also supported by the evidence provided in section 4 that the two calibration procedures give almost identical results.

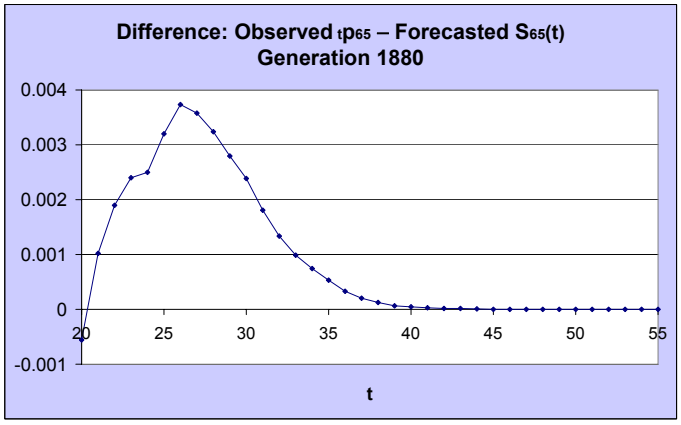
the observed survival probabilities in 1964, graph 13 reports them in 2005, and graph 14 reports the differences between the survival probabilities experienced after the observation date and the forecasted ones.



Graph 12. Observed and theoretical survival function, generation 1880, initial age 65, observation date 1964.



Graph 13. Observed and theoretical survival function, generation 1880, initial age 65, observation date 2005.

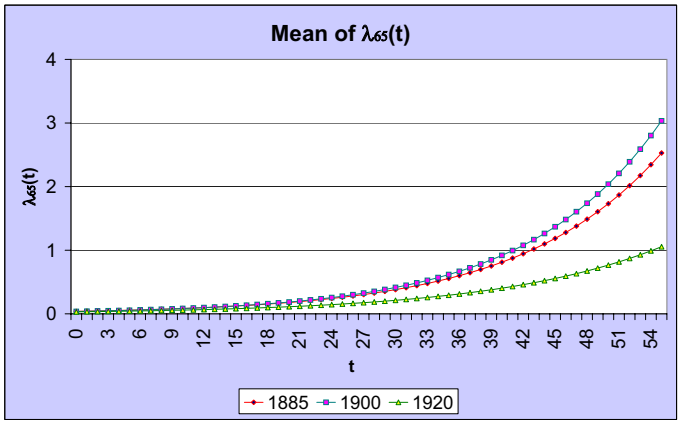


Graph 14. Difference between observed and theoretical survival function, generation 1880, initial age 65, observation date 2005.

The differences between the forecasted and the experienced survival probabilities result to be very small: we find this an encouraging result in terms of reliability of the proposed forecasting procedure.

5.2.2 Mortality trend

In section 4, when presenting the calibration results for the non mean reverting processes we have already remarked that the parameters follow a trend from one generation to the other. The mortality intensity at age 65, $\lambda_{65}(0)$, is decreasing, while the relative rate of increase of the force of mortality, a is increasing. Another way of looking at mortality trend consists in comparing the expected mortality intensity for different generations. To do this, we have simulated the process $\lambda_{65}(t)$ for the generations 1885, 1900 and 1920 with the same methodology presented in section 5.1. Graph 15 reports the mean of $\lambda_{65}(t)$ for the three generations.



Graph 15. Mean of simulated $\lambda_{65}(t)$ for generations 1885, 1900 and 1920 (FEL model).

As expected, the older the generation, the higher the mean of the intensity. Further analysis of the distribution of the path of $\lambda_{65}(t)$, here not reported, shows that the order between different generations is respected also considering the percentiles of the corresponding distributions.

6 Summary and further research

In this paper, we have described the evolution of mortality through doubly stochastic processes: namely, the time of death has been modeled as the first jump time of a Poisson process with stochastic intensity. The intensity has been described as a time-homogeneous affine process with deterministic part that increases exponentially. We have proposed three of such processes, which are different in their stochastic part and which are natural extensions of the Gompertz model. Their survival probabilities have been provided in closed form and studied analytically. The intensity processes have been calibrated to the UK population, using observed mortality tables for old generations and projected tables for younger ones. As a result, the selected processes seem to be appropriate to describe the death intensity of individuals. They satisfy most of the reasonable criteria that a good model for mortality should have. In addition, simulations show that the impact of the randomness on the actual number of deaths can be significant, even when the volatility coefficient is small. To conclude, we have proposed procedures for mortality forecasting and mortality trend assessment, which describe future evolution of mortality within a single cohort and between different cohorts, respectively.

This paper leaves scope for further research in many directions. We have considered only negative jumps in the mortality intensity in order to capture improvements in mortality (medicine progresses, better standard of life etc.). An interesting extension consists in including positive jumps, which would capture also catastrophic mortality events. In addition, positive jumps could be designed to have transitory effects, rather than permanent ones.
Extension of the doubly stochastic setup to couples of individuals has been introduced in Luciano et al. (2008) and is subject of ongoing research.
Last but not least, the adoption of these processes for pricing and reserving purposes in life insurance is also appealing.

References

Arnold, L. (1974). *Stochastic Differentials Equations: Theory and Applications*, John Wiley and Sons.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations, *Insurance: Mathematics and Economics* **37**: 443–468.

Bowers, N. L., Gerber, H. U., Hickman, Jones, J. C. and Nesbitt, C. J. (1986). *Actuarial Mathematics*, The Society of Actuaries, Itasca.

Brémaud, P. (1981). *Point Processes and Queues – Martingale Dynamics*, Springer Verlag, New York.

Brouhns, N., Denuit, M. and Vermunt, J. K. (2002). A poisson log-bilinear approach to the construction of projected lifetables, *Insurance: Mathematics and Economics* **31**: 373–393.

Cairns, A. J. G., Blake, D. and Dowd, K. (2006). Pricing death: Framework for the valuation and securitization of mortality risk, *ASTIN Bulletin* **36**: 79–120.

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

Denuit, M. and Devolder, P. (2006). Continuous time stochastic mortality and securitization of longevity risk, *Working Paper 06-02, Institut des Sciences Actuarielles, Universite' Catholique de Louvain, Louvain-la-Neuve*.

Duffie, D. (2001). *Dynamic Asset Pricing Theory, Third Edition*, Princeton University Press.

Duffie, D. and Singleton, K. J. (2003). *Credit risk*, Princeton University Press.

Duffie, D., Filipovič, D. and Schachermayer, W. (2003). Affine processes and applications in finance, *Annals of Applied Probability* **13**: 984–1053.

Duffie, D., Pan, J. and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**: 1343–1376.

Feller, W. (1951). Two singular diffusion problems, *The Annals of Mathematics* **54**: 173–182.

Gerber, H. U. (1997). *Life Insurance Mathematics*, Springer Verlag, Berlin.

Institute and Faculty of Actuaries (1990). *Standard Tables of Mortality: the "92" Series*, The Institute of Actuaries and the Faculty of Actuaries.

Lee, R. D. (2000). The Lee-Carter method for forecasting mortality, its various extensions and applications, *North American Actuarial Journal* **4**: 80–93.

Lee, R. D. and Carter, L. R. (1992). Modelling and forecasting U.S. mortality, *Journal of the American Statistical Association* **87**: 659–675.

Luciano, E. and Vigna, E. (2005). Non mean reverting affine processes for stochastic mortality, *Carlo Alberto Notebook 30/06 and ICER WP 4/05*.

Luciano, E., Spreeuw, J. and Vigna, E. (2008). Modeling stochastic mortality for dependent lives, *Insurance: Mathematics and Economics* **43**: 234–244.

Menoncin, F. (2008). The role of longevity bonds in optimal portfolios, *Insurance: Mathematics and Economics* **42**: 343–358.

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

Pitacco, E. (2004a). Longevity risk in living benefits, *in* E. Fornero and E. Luciano (eds), *Developing an Annuity Market in Europe*, Edward Elgar, Cheltenham, pp. 132–167.

Pitacco, E. (2004b). Survival models in a dynamic context: a survey, *Insurance: Mathematics and Economics* **35**: 279–298.

Renshaw, A. E. and Haberman, S. (2003). On the forecasting of mortality reductions factors, *Insurance: Mathematics and Economics* **32**: 379–401.

Schrager, D. (2006). Affine stochastic mortality, *Insurance: Mathematics and Economics* **38**: 81–97.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age, *Journal of the Royal Statistical Society A* **162**: 5–43.

University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (2002). *Human Mortality Database*, Available at www.mortality.org or www.humanmortality.de.