THE CARLO ALBERTO NOTEBOOKS

Fondazione Collegio Carlo Alberto

Non mean reverting affine processes for stochastic mortality

Elisa Luciano

Elena Vigna

No.30, October 2006

www.carloalberto.org

Non mean reverting affine processes for stochastic mortality[1]

Elisa Luciano[2] Elena Vigna[3]

October 16, 2006

[1]We thank Andrew Cairns, Bjarne Højgaard and Enrico Biffis, for interesting and stimulating comments. We especially thank Steven Haberman and Russell Gerrard, who pointed out the importance, in this context, of the exponential increase in the observed force of mortality. Any remaining errors are ours. A version of this paper was circulated as ICER working paper.

[2]University of Turin & ICER, Turin & Collegio Carlo Alberto, Moncalieri.
Email: luciano@econ.unito.it. Tel. +39 011 670 5742. Fax. +39 011 670 5783. Address: Dipartimento di Statistica e Matematica Applicata, piazza Arbarello, 8, 10122, Torino, Italy.

[3]University of Turin. Email: elena.vigna@econ.unito.it. Tel. +39 011 670 5754. Fax. +39 011 670 5783.

Abstract

In this paper we use doubly stochastic processes (or Cox processes) in order to model the random evolution of mortality of an individual. These processes have been widely used in the credit risk literature in modelling default arrival, and in this context have proved to be quite flexible, especially when the intensity process is of the affine class.

We investigate the applicability of time-homogeneous affine processes in describing the individual's intensity of mortality and the mortality trend, as well as in forecasting it. We calibrate them to the UK population. Calibrations suggest that, in spite of their popularity in the financial context, mean reverting time-homogeneous processes are less suitable for describing the death intensity of individuals than non mean reverting processes. Among the latter, affine processes whose deterministic part increases exponentially seem to be appropriate. They are natural generalizations of the Gompertz law. Stress analysis and analytical results indicate that increasing the randomness of the intensity process for a given cohort results in improvements in survivorship.

Mortality forecasts and their comparison with experienced mortality rates provide further encouraging evidence in favour of non mean reverting processes. The mortality trend is evidenced through the evolution over time of the parameters and through the intensity simulation for different generations.

JEL classification: G22, J11.

Keywords: doubly stochastic processes (Cox processes), affine processes, stochastic mortality, mortality forecasting.

1 Introduction

The issue of mortality risk – and, in particular, of longevity risk – has been largely addressed in recent years when dealing with the pricing of insurance products. It is well known from the basics of actuarial science that the price of any insurance product on the duration of life depends on two main basis: demographical and financial assumptions. Traditionally, actuaries have been treating both the demographic and the financial assumptions in a deterministic way, by considering available mortality tables for describing the future evolution of mortality and by setting the so-called "best estimate" of the rate of interest for discounting cash flows over time. More recently, stochastic models have been adopted to describe the uncertainty linked both to mortality and to financial factors. Cairns, Blake and Dowd (2005) provide a comprehensive review of existing modelling frameworks for stochastic mortality and discuss in depth the related arbitrage-free pricing issues. We focus on mortality risk and on modelling the survival function of the individual, without studying the impact of pricing and reserving. In the setting proposed here, however, the extension to stochastic interest rates is natural, under the standard assumption of independence between financial and mortality risks (see, for instance, Dahl (2004), Biffis (2005) and Cairns et al. (2005)).

The remainder of the paper is organized as follows. Section 2 gives a brief review of the actuarial literature on mortality risk. Section 3 introduces Cox processes. In section 4 there is the actuarial application of time-homogeneous Cox processes with mean reversion. Section 5 presents the application with non mean reverting processes, while section 6 discusses its calibration results. Section 7 studies the impact of mortality randomness. Section 8 introduces a procedure for mortality forecast and mortality trend, with an application to the calibrated results. Section 9 summarizes and concludes.

2 Modelling mortality risk

In the last decades significant improvements in the duration of life have been experienced in most developed countries. 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..

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. Indeed, strong or unexpected reductions in mortality rates can lead to mispricing of these products and can affect the solvency of the insurance company.

The actuarial literature about modelling and forecasting mortality rates is vaste 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 obviously increasing as x increases, as the probability of imminent death increases when ageing [1].

When allowing for mortality trends 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, known as 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 trends, 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.

[1] With some exceptions, like very small values of x – due to the infant mortality – and values around 20-25 – due to the young mortality hump.

Finally, Milevsky and Promislow (2001) have used a stochastic intensity of mortality, whose expectation at any future date – under an appropriate choice of the parameters – has a Gompertz specification. The existence of a death process which admits their stochastic intensity of mortality as arrival rate can be addressed using the doubly stochastic processes, to which the next section is devoted.

3 The mathematical framework

The theory of stochastic intensities, doubly stochastic processes and affine processes underlying the actuarial application presented here is enormous and covered in many texts about stochastic processes. A detailed and thorough treatment is clearly beyond the scope of this paper, and we limit ourselves to a brief summary of the mathematical tools used, sacrificing scientific rigor and omitting all the proofs. However, we refer the interest reader to Brémaud (1981) and Duffie (2001).

The reason why such a sophisticated mathematical framework has been used in describing the mortality risk is the great analytical tractability of the models presented, once some useful and not too restrictive assumptions are introduced. These mathematical tools have been extensively used in the credit risk literature, when modelling the time to default. The pioneering works in this field are Artzner and Delbaen (1992), Lando (1994) and Duffie and Singleton (1999). Applications of this mathematical framework to dynamic mortality modelling and to insurance products pricing can be found in Dahl (2004), Biffis (2005), Denuit and Devolder (2005) and Schrager (2005). The similarity between the time to default and the remaining duration of life is strong, and, although the factors underlying the death of an individual and the default of a firm are obviously completely different, the mathematical tools used in the two literatures are the same.

3.1 Counting processes

In describing the mathematical tools, we will mainly follow Duffie (2001) and Duffie (2002). We are given 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.

A *counting* process (or *point* process) N is defined using a sequence of increasing random variables $\{T_0, T_1, ...\}$, with values in $[0, \infty]$, s.t. $T_0 = 0$ and $T_n < T_{n+1}$ whenever $T_n < \infty$, in the following way:

$$N_t = n \qquad for \qquad t \in [T_n, T_{n+1})$$

and $N_t = \infty$ if $t \geq T_\infty = \lim_{n\to\infty} T_n$. It is easy to see T_n as the time of the n^{th} jump of the process N and N_t as the number of jumps occurred up to time t, including time t (hence the definition "counting" process). The counting process is said to be *nonexplosive* if $T_\infty = \infty$ almost surely.

3.2 Stochastic intensity

Let $\{\mathcal{F}_t : t \geq 0\}$ be a filtration satisfying the usual conditions, with $\mathcal{F}_t \subset \mathcal{G}_t$, and λ be a nonnegative $(\mathcal{F}_t)$-predictable process s.t. $\int_0^t \lambda(s)ds < \infty$ almost surely. A nonexplosive adapted counting process N is said to admit the intensity λ if the compensator of N admits the representation $\int_0^t \lambda(s)ds$, i.e. if $M_t = N_t - \int_0^t \lambda(s)ds$ is a local martingale. If the stronger condition $E(\int_0^t \lambda(s)ds) < \infty$ is satisfied, $M_t = N_t - \int_0^t \lambda(s)ds$ is a martingale.

From this, one gets:

$$E(N_{t+\Delta t} - N_t|\mathcal{F}_t) = E\left(\int_t^{t+\Delta t} \lambda(s)ds|\mathcal{F}_t\right)$$

which, after a few passages and under technical conditions, leads to:

$$E(N_{t+\Delta t} - N_t|\mathcal{F}_t) = \lambda(t)\Delta t + o(\Delta t) \tag{3.1}$$

Equation 3.1 (see the analogy with equation 2.1) stresses the importance of the process λ in giving information about the average number of jumps of the process under observation in a small period of future time. Observe that conditioning is made on the smallest filtration, therefore on the availability of poorer information. The idea is that the information at time t can give insight about the expected number of jumps in the next future or, in other words, about the likelihood of a jump in the immediate future. It cannot predict the actual occurrence of a jump, that comes as a "sudden surprise".

3.3 Doubly stochastic processes

A nonexplosive counting process N with intensity λ is said to be *doubly stochastic driven by* $\{\mathcal{F}_t : t \geq 0\}$, if for all $t < s$, conditional on the $\sigma-$algebra $\mathcal{G}_t \vee \mathcal{F}_s$, generated by $\mathcal{G}_t \cup \mathcal{F}_s$, the process $N_s - N_t$ has Poisson distribution with parameter $\int_t^s \lambda(u)du$.

As an example, we observe that any Poisson process is a doubly stochastic process driven by the filtration $\mathcal{F}_t = (\emptyset, \Omega) = \mathcal{F}_0$ for any $t \geq 0$, in that the intensity is deterministic.

A stopping time τ is said to be *doubly stochastic with intensity* λ if the underlying counting process whose first jump time is τ is doubly stochastic with intensity λ.

The mathematical arsenal presented so far is now sufficient to present the first interesting result that will be used in the applications. If $\tau \geq t$ is a stopping time doubly stochastic with intensity λ, it can be shown, by using the law of iterated expectations, that:

$$P(\tau > s|\mathcal{G}_t) = E\left[e^{-\int_t^s \lambda(u)du}|\mathcal{G}_t\right] \tag{3.2}$$

Readers who are familiar with mathematical finance can easily see in the r.h.s. of equation (3.2) 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 mathematical finance literature about interest rate models can thus be retrieved in this setting.

Another interesting result that can be used relates to the density function of a doubly stochastic stopping time τ. If we let $p(t) = P(\tau > t)$ be the *survival function*, then the density function of τ, if it exists, is given by $-p'(t)$. Under technical conditions (see for example Grandell (1976)), we have:

$$p'(t) = E\left[-e^{-\int_0^t \lambda(u)du}\lambda(t)|\mathcal{G}_t\right] \tag{3.3}$$

It is clear how these results can be naturally applied in the actuarial context: if one sees τ as the future lifetime of an individual aged x, T_x, equations 3.2 and 3.3 can be applied to find the survival function and the density function of T_x, given a model for the death intensity λ.

3.4 Affine processes

Our next step will be to show how equations like 3.2 and 3.3 can be approached. It turns out that it is convenient to specify the stochastic intensity λ as a function Λ of another process X in $\mathbf{R}$, whose dynamics is given by the SDE:

$$dX(t) = f(X(t))dt + g(X(t))d\tilde{W}(t) + dJ(t) \tag{3.4}$$

where $\tilde{W}$ is an n-dimensional Brownian motion, J is a pure jump process and where the drift $f(X(t))$, the covariance matrix $g(X(t))g(X(t))'$ and the jump measure associated with J have affine dependence on $X(t)$. Such a process is named an affine process: interest readers can find a thorough treatment of affine processes in Duffie, Filipovič and Schachermayer (2003).

The financial literature on interest rate modelling is full of examples of affine processes: the Ornstein-Uhlenbeck process, used by Vasicek (1977) for modelling interest rates, is affine, as is the Feller process, used by Cox, Ingersoll and Ross (1985).

The convenience of adopting affine processes in modelling the intensity lies in the fact that, under technical conditions (see Duffie and Singleton (2003)), it yields, for any $w \in \mathbf{R}$:

$$E\left[e^{\int_t^T -\Lambda(X(u))du + wX(T)} | \mathcal{G}_t\right] = e^{\alpha(T-t)+\beta(T-t)X(t)} \tag{3.5}$$

where the coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalized Riccati ODEs. The latter can be solved at least numerically and in some cases analytically. Therefore, the problem of finding the survival function (3.2) becomes tractable, whenever affine processes for $X(t)$ are employed. Furthermore, if one chooses time-homogeneous processes, also the calibration to actual data can be performed through standard procedures.

4 The actuarial application: mean reverting processes

Turning back to our initial problem of modelling adequately the dynamic mortality, we will now use some of the mathematical tools presented in the previous section.

As above, the 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 $\mathcal{F}_t \subset \mathcal{G}_t$. In other words, T_x is the first jump time of a nonexplosive counting process N with intensity λ_x. Intuitively, the counting process N may be seen as a process that jumps whenever the individual dies: $N_t = 0$ if $t < T_x$, $N_t > 0$ if $t \geq T_x$.

According to (3.2) the survival probability is[2]:

$$S_x(t) = P(T_x > t|\mathcal{G}_0) = E\left[e^{-\int_0^t \lambda_x(u)du} | \mathcal{G}_0\right] \tag{4.1}$$

[2] The similarity with the actuarial survival probability for t years for an individual aged x, ${}_tp_x$, expressed in terms of the force of mortality, is strong:

$${}_tp_x = e^{-\int_0^t \mu_{x+s}ds}$$

It will be addressed in section 5.

The specification of the intensity process λ_x is now crucial for the solution of equation (4.1). If we assume that the intensity itself is an affine process, then we can apply (3.5) with $w = 0$ and $\Lambda(x) = x$. Since we are interested also in the practical appropriateness of the models, in what follows we restrict our attention to time-homogeneous processes for the mortality intensity.

Recent studies on the firm's mortality (as reported in Duffie and Singleton (2003)) indicate the suitability of the following univariate time-homogeneous affine processes for modelling the intensity $\lambda_x(t)$:

$$CIR \quad process: \qquad d\lambda_x(t) = k(\gamma - \lambda_x(t))dt + \sigma\sqrt{\lambda_x(t)}dW(t)$$

$$mean\ reverting\ with\ jumps\ (m.r.j.): \qquad d\lambda_x(t) = k(\gamma - \lambda_x(t))dt + dJ(t)$$

where $W(t)$ is a standard Brownian motion, $k > 0$, $\gamma > 0$, $\sigma \geq 0$ and $J(t)$ is a compound Poisson process with intensity l and jumps exponentially distributed with expected value μ.

In addition, we consider the Vasicek process (see Vasicek (1977)) for $\lambda_x(t)$:

$$VAS \quad process: \qquad d\lambda_x(t) = k(\gamma - \lambda_x(t))dt + \sigma dW(t)$$

with the same parameter restrictions as above.

Using the result (3.5) and solving the Riccati ODEs, one gets the survival probabilities in closed form for all the specifications of the intensity process:

$$S_x(t) = e^{\alpha(t)+\beta(t)\lambda_x(0)} \tag{4.2}$$

where, in the CIR case (see for instance Duffie and Singleton (2003)):

$$\alpha(t) = -\frac{2k\gamma}{\sigma^2}\ln\left(\frac{c + de^{bt}}{b}\right) + \frac{k\gamma}{c}t$$

$$\beta(t) = \frac{1 - e^{bt}}{c + de^{bt}}$$

$$b = -\sqrt{k^2 + 2\sigma^2} \qquad c = \frac{b-k}{2} \qquad d = \frac{b+k}{2}$$

In the $m.r.j.$ case, under the conditions stated in Duffie and Singleton (2003):

$$\alpha(t) = -\gamma(t + \beta(t)) - l\frac{\mu t - \ln(1 - \mu\beta(t))}{\mu + k}$$

$$\beta(t) = \frac{e^{-kt} - 1}{k}$$

In the Vasicek (VAS) case $\beta(t)$ is defined as in the $m.r.j.$ model, while (see Vasicek (1977)):

$$\alpha(t) = -\frac{(\beta(t) + t)(k^2\gamma - \frac{\sigma^2}{2})}{k^2} - \frac{\sigma^2\beta(t)^2}{4k}$$

In all these cases, it is possible to calibrate the values of the parameters starting from a time series of survival probability data. 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, our process λ captures the mortality intensity for a particular generation and a particular initial age. This has to be allowed for when choosing the mortality table: the approach adopted here is a "diagonal" one.

4.1 Calibration to the UK population: projected and observed generation tables

As a first application, we have calibrated the three processes to the UK population.

The mortality tables selected for the calibration are two observed generation tables, for males born in 1880 and in 1900 respectively, and two projected mortality tables, for males born in 1935 and 1945 respectively. 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)).

In the calibration, we have optimized with respect to negative values of μ 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. We must say that 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.

In fitting the table, we have minimized the overall squared error, defined as the sum of the squared spreads between the different model survival probabilities and the table ones. Table 1 reports for each intensity process the optimal values of the parameters and the calibration error, as well as the corresponding initial value of $\lambda, \lambda_{65}(0)$ (which has been chosen equal to $-\ln(p_{65})$).

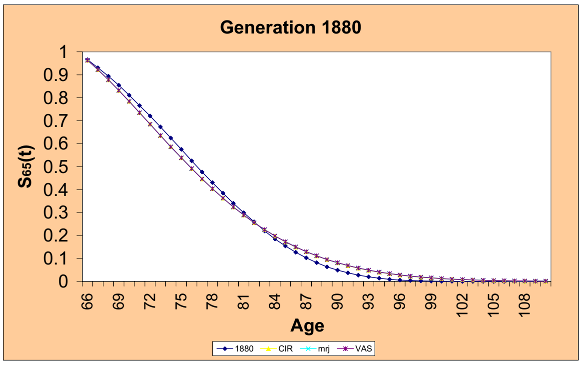
TABLE 1

	1880	1900	1935	1945
$\lambda_{65}(0)$	0.03515	0.03797	0.01145	0.00885
CIR-error	0.02182	0.01662	0.40945	0.20552
CIR-k	0.00448	0.01365	0.06494	0.0078
CIR-σ	0.00103	0.00298	0.00005	0
CIR-γ	1.24656	0.4301	0.07552	0.41711
mrj-error	0.02236	0.01327	0.15816	0.1965
mrj-k	0.00571	0.00392	0.005	0.00465
mrj-μ	-0.00246	-0.00227	-0.00249	-0.00492
mrj-l	0.00247	0.00234	0.00249	0.0099
mrj-γ	0.99382	1.31818	0.64908	0.67935
VAS-error	0.02247	0.01473	0.16191	0.1982
VAS-σ	0.00046	0.00048	0.00002	0.00002
VAS-k	0.00591	0.00835	0.00604	0.00526
VAS-γ	0.96029	0.65393	0.53278	0.59302

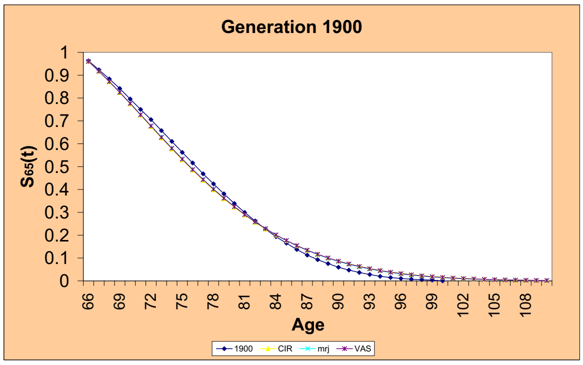
In all models, the value of the long term mean for λ, γ, lies between 0.07 and 1.3. It generally decreases when considering younger generations, which is an expected result. The speed of convergence k seems to be stable in the last two models (ranging between 0.004 and 0.008), and volatile in the first one. The size of the jumps ranges between -0.002 and -0.005, while the frequency lies between 0.002 and 0.01. The value of σ is very low in all cases, ranging between 0 and 0.003.

The most remarkable result is the change in the value of the error when passing from the observed old tables to the projected ones for younger generations: it more than decuplicates, ranging for the latter between 0.15 and 0.4 against 0.01-0.02 for the former. The different magnitude of the error can be better perceived when considering the curve of the survival function $S_{65}(t)$ implied by the three models and the survival probabilities of the relevant table.

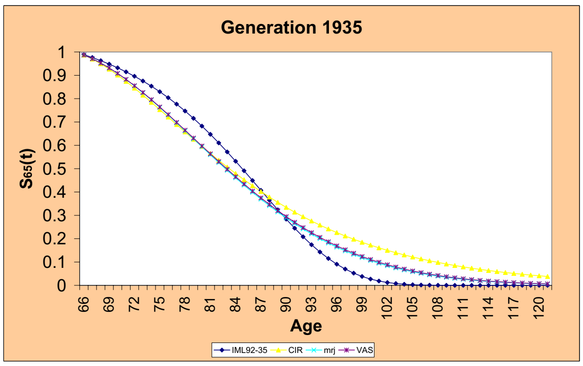
Graphs 1, 2, 3 and 4 report, for the different generations, the survival function of the three processes analyzed (CIR, m.r.j. and VAS) and the ones of the tables considered.



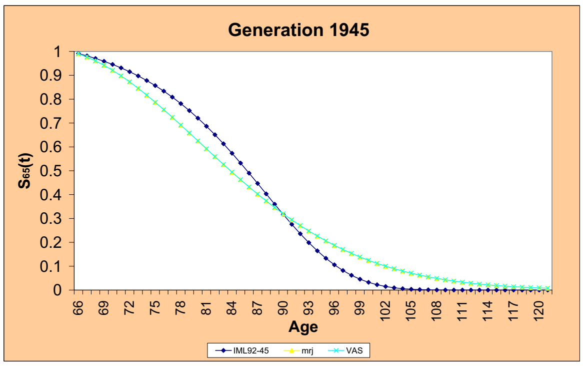
Graph 1



Graph 2



Graph 3



Graph 4

It is evident from a first inspection of the graphs that, while the fit can be considered satisfactory for the first two generations, it cannot be considered so for the last two. In particular, one can notice that, in the last two cases the survival functions implied by the three processes do not capture the rectangularization phenomenon. In addition, the survival probability at very old ages is much higher and at lower ages much lower than in the fitted tables. Therefore, although the last two tables refer to projected mortality tables and not to observed ones, these results seem to suggest that in the presence of high rectangularization phenomenon – which is an expected feature in the future generation tables – the intensity of mortality cannot be properly described by the three proposed processes.

5 The actuarial application: non mean reverting processes

The calibration of the time-homogeneous mean reverting processes presented so far gives a survival function that, with respect to the new generations, fails to capture the rectangularization phenomenon and produces unrealistic survival probabilities at very old ages. The question arises as to whether the common and disturbing element in those processes is the mean reverting term, as suggested also by Cairns et al. (2005).

Furthermore, 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. This observation, consistent with all the deterministic exponential models presented in the actuarial literature, naturally leads to the simple idea of dropping the mean reverting term in the classical affine processes used in finance and choosing processes whose deterministic part increases exponentially. Four affine models with these two desired characteristics are presented and discussed below.

As far as we know, the idea of abandoning the mean reversion feature in the intensity process is new in the actuarial literature. Other authors have proposed intensity processes which are mean reverting to a function of time instead of a constant, see for instance Dahl (2004), Biffis (2005) and Dahl and Møeller (2005). Time-inhomogeneous models are certainly much more general than the models proposed here. However, in the next sections we will show that even the simpler time-homogeneous non mean reverting special case can be an effective way to describe human mortality, as a generalization of the Gompertz law.

5.1 The Ornstein Uhlenbeck process without jumps

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

$$OU \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma dW(t) \tag{5.1}$$

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

$$\lambda(t) = \lambda(0)e^{at} + \sigma \int_0^t e^{t-s} dW(s) \tag{5.2}$$

The main drawback when choosing this process for the intensity is that it becomes negative with positive probability.

By applying standard results on linear stochastic differential equations (see, for instance, Arnold (1974)) to the process (5.2) we have that $\lambda(t)$ is normally distributed with mean

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

and variance

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

Therefore, the calculation of the probability of $\lambda(t)$ taking negative values is straightforward:

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

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 numerical applications.

By applying the framework of equation (3.5) (in particular, see Duffie, Pan and Singleton (2000) pagg. 1350–1351) we have that:

$$S_x(t) = E\left(e^{-\int_0^t \lambda(u)du}|\mathcal{G}_0\right) = e^{\alpha(t)+\beta(t)\lambda(0)} \tag{5.3}$$

where the functions α and β solve the system of ODEs':

$$\begin{cases} \alpha'(t) = \frac{1}{2}\sigma^2\beta^2(t) \\ \beta'(t) = -1 + a\beta(t) \end{cases} \tag{5.4}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0 \tag{5.5}$$

By solving the system 5.4–5.5, we find α and β:

$$\begin{cases} \alpha(t) = \frac{\sigma^2}{2a^2}t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} \\ \beta(t) = \frac{1}{a}\left(1 - e^{at}\right) \end{cases} \tag{5.6}$$

We observe that with a strictly positive value of σ, the survival probability is a decreasing function for

$$t \leq T^* = \frac{1}{a}\ln\left[1 + \frac{a^2\lambda(0)}{\sigma^2}\left(1 + \sqrt{1 + \frac{2\sigma^2}{a^2\lambda(0)}}\right)\right], \tag{5.7}$$

and increasing for $t > T^*$; 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. Furthermore, as seen above, also the probability of negative values of λ turns out to be negligible with the calibrated parameters. Thus, the evidence seems to be an encouraging one, with respect to practical application of the model by actuaries and demographers.

5.2 The Ornstein Uhlenbeck process with jumps

In the second model we add a jump component in the stochastic part of the mortality process. Therefore, the process λ is given by:

$$OUj \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma dW(t) + dJ(t) \tag{5.8}$$

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

We can apply the formulae of Duffie et al. (2000) provided that

$$\beta(t) < \frac{1}{\mu} \quad \text{if} \quad \mu > 0, \quad \beta(t) > \frac{1}{\mu} \quad \text{if} \quad \mu < 0 \tag{5.9}$$

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

$$\begin{cases} \alpha'(t) = \frac{1}{2}\sigma^2\beta^2(t) + l\frac{\mu\beta(t)}{1-\mu\beta(t)} \\ \beta'(t) = -1 + a\beta(t) \end{cases} \tag{5.10}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0 \tag{5.11}$$

The equation for β is the same as before (5.4), so is the solution. The solution for α is instead different (due to the inclusion of the jump component), and we have:

$$\begin{cases} \alpha(t) = (\frac{\sigma^2}{2a^2} + \frac{la}{a-\mu})t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} + \frac{l}{a-\mu}\ln(1 - \frac{\mu}{a} + \frac{\mu}{a}e^{at}) \\ \beta(t) = \frac{1}{a}\left(1 - e^{at}\right) \end{cases} \tag{5.12}$$

Observe that if $\mu > 0$, condition (5.9) is always satisfied, since $\beta(t) < 0$. On the other hand, if $\mu < 0$, the condition has to be checked. In both cases, it guarantees that the argument of the logarithm in (5.12) 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

$$\frac{1}{2}\sigma^2\beta(t)^2 + \frac{l\mu\beta(t)}{1-\mu\beta(t)} + a\lambda(0)\beta(t) - \lambda(0) < 0 \tag{5.13}$$

A sufficient condition for (5.13) in turn is that:

$$\frac{\ln(1-\beta_2 a)}{a} < t < \frac{\ln(1-\beta_1 a)}{a} \tag{5.14}$$

for $\mu > 0$ and

$$t > \frac{\ln(1-\beta_2 a)}{a} \quad \vee \quad t < \frac{\ln(1-\beta_1 a)}{a} \tag{5.15}$$

for $\mu < 0$, where

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

and

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

5.3 The Feller process without jumps

The third model proposed is the Feller process, already investigated in the previous section as CIR process, without the mean reverting term:

$$FEL \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma\sqrt{\lambda(t)}dW(t) \tag{5.16}$$

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

The main advantage of this process w.r.t. the previous ones is that it does not violate the nonnegativity constraint of the intensity, provided that the starting point is nonnegative.

The solution $\lambda(t)$ of the SDE (5.16) is

$$\lambda(t) = \lambda(0)e^{at} + \sigma\int_0^t e^{a(t-u)}\sqrt{\lambda(u)}dW(u) \tag{5.17}$$

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

The application of the affine framework gives the following system of ODE's for α and β:

$$\begin{cases} \alpha'(t) = 0 \\ \beta'(t) = -1 + a\beta(t) + \frac{1}{2}\sigma^2\beta^2(t) \end{cases} \tag{5.18}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0 \tag{5.19}$$

The solution is:

$$\begin{cases} \alpha(t) = 0 \\ \beta(t) = \frac{1-e^{bt}}{c+de^{bt}} \end{cases} \tag{5.20}$$

with:

$$\begin{cases} b = -\sqrt{a^2 + 2\sigma^2} \\ c = \frac{b+a}{2} \\ d = \frac{b-a}{2} \end{cases} \tag{5.21}$$

Given that the coefficients b, c, d are negative, the survival probability is always decreasing in t if and only if

$$e^{bt}(\sigma^2 + 2d^2) > \sigma^2 - 2dc \tag{5.22}$$

Notice that (5.22) is automatically satisfied if $\sigma^2 - 2dc < 0$, a condition which holds in our calibrations. 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.

5.4 The Feller process with jumps

In the fourth model, we add a jump component in the stochastic part of the Feller process. The intensity λ is given by:

$$FELj \quad process \qquad d\lambda(t) = a\lambda(t)dt + \sigma\sqrt{\lambda(t)}dW(t) + dJ(t) \tag{5.23}$$

where J is the pure jump process defined above.

Under the technical condition (5.9), the functions α and β that enter the survival probability (5.3) solve the following system of ODE's equations:

$$\begin{cases} \alpha'(t) = \frac{l\mu\beta(t)}{1-\mu\beta(t)} \\ \beta'(t) = -1 + a\beta(t) + \frac{1}{2}\sigma^2\beta^2(t) \end{cases} \tag{5.24}$$

with boundary conditions

$$\alpha(0) = 0, \beta(0) = 0 \tag{5.25}$$

Similarly to the OU case, the equation for β is the same as in the no-jump setting. The jump component enters only the equation for α. We have:

$$\begin{cases} \alpha(t) = \frac{l\mu}{c-\mu}t - \frac{l\mu(c+d)}{b(d+\mu)(c-\mu)}[\ln(\mu - c - (d+\mu)e^{bt}) - \ln(-c-d)] \\ \beta(t) = \frac{1-e^{bt}}{c+de^{bt}} \end{cases} \tag{5.26}$$

with b, c, d given by the equations (5.21) above.

As in the OUj model, if $\mu > 0$ the technical condition is always satisfied, if $\mu < 0$ it has to be checked. As before, its validity guarantees that the argument of the first logarithm in (5.26) is positive.

The survival function is decreasing for every t when $\mu > 0$. When $\mu < 0$ a necessary and sufficient condition for it to be decreasing is

$$-\frac{1}{2}\mu\sigma^2\lambda(0)\beta(t)^3 + \left(\frac{1}{2}\sigma^2\lambda(0) - a\mu\lambda(0)\right)\beta(t)^2 + (l\mu + \mu\lambda(0) + a\lambda(0))\beta(t) - \lambda(0) < 0 \tag{5.27}$$

In turn, a sufficient condition for (5.27) is

$$\beta(t) < \beta_1 \quad \vee \quad \beta(t) > \beta_2, \tag{5.28}$$

where

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

and

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

Notice that, in the case of negative jumps, the intensity $\lambda(t)$ can become negative, in which case the diffusive component could no longer be defined. Rigourously, we should study the process $\lambda(t)I_{\{min_{s\leq t}\lambda(s)\geq 0\}}$, where I_A is the indicator of the event A. However, since in actuarial applications the probability of λ becoming negative is negligible, the trade-off between analytical tractability of the model and relevance of the event $\{\lambda(t) < 0\}$ leads us to study only the process (5.23).

5.5 The link with existing models for the force of mortality

It is interesting to investigate the relationship between our models for the stochastic intensity of mortality and the deterministic force of mortality actuaries are more familiar with. 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:

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

For example, in the OU model, the force of mortality becomes:

$$\mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax} - 1)^2 \tag{5.29}$$

If $\sigma = 0$ 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. This is straightforward also observing that if $\sigma = 0$ the evolution of $\lambda_0(t)$ is deterministic and given by

$$d\lambda_0(t) = a\lambda_0(t)dt$$

However, the coincidence between intensity of mortality and force of mortality is clearly no longer true when the intensity is stochastic, and equation (5.29), compared with equation (5.2) for λ tells us that

$$\mu_x < E(\lambda_0(x)) \tag{5.30}$$

In other words, the force of mortality decreases, hence the survivorship improves, when the diffusion coefficient increases. We will come back to this feature later, when considering the impact of the random part of the process on the survival probabilities [3].

With the other three models, we have:

$$OUj \qquad \mu_x = \lambda_0(0)e^{ax} - \frac{\sigma^2}{2a^2}(e^{ax} - 1)^2 - \frac{l}{a - \mu}\left(1 - \frac{a\mu e^{ax}}{a - \mu + \mu e^{ax}}\right) \tag{5.31}$$

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

$$FELj \qquad \mu_x = \frac{4\lambda_0(0)b^2 e^{bx}}{[(a+b) + (b-a)e^{bx}]^2} + \frac{l\mu(1 - e^{bx})}{\mu - c - (d+\mu)e^{bx}}$$

It is clear (and easy to check) that also with these three models, when the coefficients σ and l of the random part are set equal to 0, there is coincidence between intensity of mortality and force of mortality, which turns out to be of the Gompertz type.

[3] Observe that the inequality (5.30) is also consistent with the fact that

$$\int_0^t \mu_s ds < \int_0^t E(\lambda_0(s))ds$$

a result that derives by application of Jensen inequality to the survival function.

6 Calibration of the non mean reverting processes

In this section we calibrate the four models introduced (OU, OUj, FEL, FELj) to the same mortality tables used for the non mean reverting processes. The calibration procedure is the same followed in section 4.1. Here too, when considering jumps, we restrict our attention to $\mu < 0$: we verify that the technical condition for the solution's appropriateness (5.9) is satisfied for both the OUj and FELj models. Furthermore, we check the conditions that make the survival function decreasing at least up to age 120, i.e. condition (5.7) for the OU model, (5.13) for the OUj, (5.22) for the FEL, (5.27) for the FELj. The results are shown in Table 2.

TABLE 2

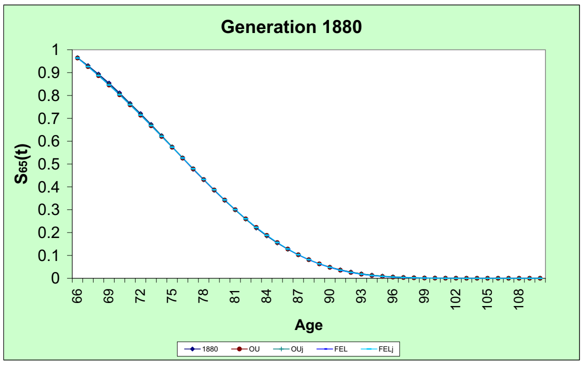
	1880	1900	1935	1945
$\lambda_{65}(0)$	0.03515	0.03797	0.01145	0.00885
OU-error	0.00043	0.00012	0.00085	0.00027
OU-a	0.0861	0.07949	0.09856	0.10859
OU-σ	0.00183	0.00341	0.0001	0.00048
OUj-error	0.0001	0.00004	0.00002	0.00016
OUj-a	0.09101	0.08192	0.10014	0.10865
OUj-σ	0.00377	0.00414	0.0001	0.00011
OUj-l	0.00173	0.00088	0.00105	0.00036
OUj-μ	-0.00003	-0.00003	-0.00003	-0.00003
FEL-error	0.00044	0.00012	0.00084	0.00027
FEL-a	0.08553	0.07896	0.09867	0.10811
FEL-σ	0.00431	0.01348	0.00005	0.0001
FELj-error	0.00043	0.00012	0.00084	0.00027
FELj-a	0.0858	0.07897	0.09867	0.10811
FELj-σ	0.00735	0.01349	0.000028	0.00001
FELj-l	0.001	0.001	0.001	0.001
FELj-μ	-0.0001	-0.0001	-0.0001	-0.0001

The main conclusion that can be drawn from the table is that the calibration errors are dramatically lower than with mean reverting intensities: they range between 0.00002 and 0.0008. In terms of calibration error, the best fitting model is the OU with jumps, though the differences between the models are quite small [4]. Models with jumps generally fit better than the corresponding ones without jumps. This result seems to suggest that negative jumps are an appropriate way to describe random variations in mortality.

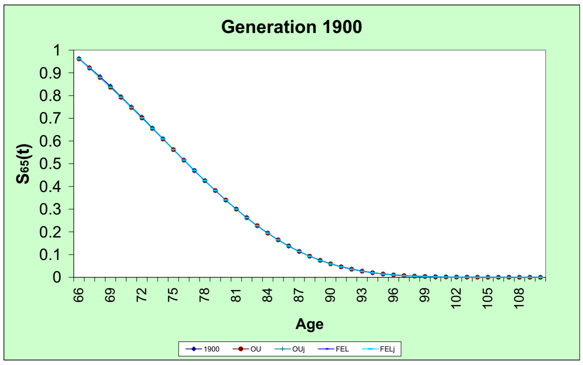
The calibrated value of the intensity diffusion, σ, is much lower in the 1935 and 1945 generations. We presume that this is due to the fact that projected tables are constructed with deterministic algorithms and in the next sections we will concentrate mainly on the observed tables, referring to the cohorts 1880, 1900.

Graphs 5, 6, 7 and 8 report the survival probabilities as from the four models analyzed and from the relevant tables.

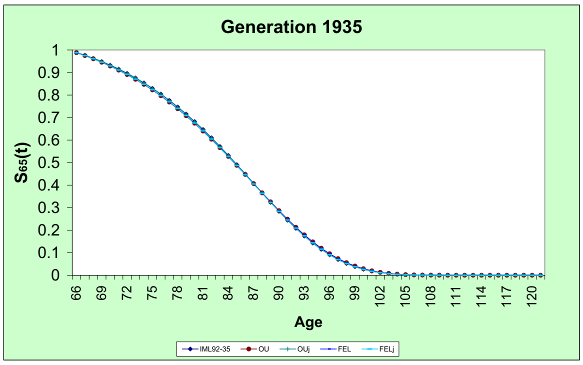
[4] We notice that with these values of the parameters the probability of negative intensity for the OU model can be considered negligible for all practical applications.



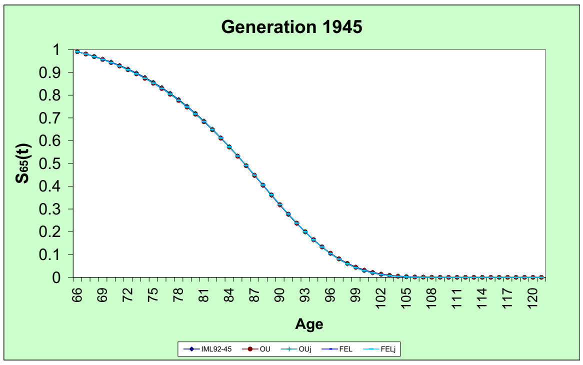
Graph 5



Graph 6



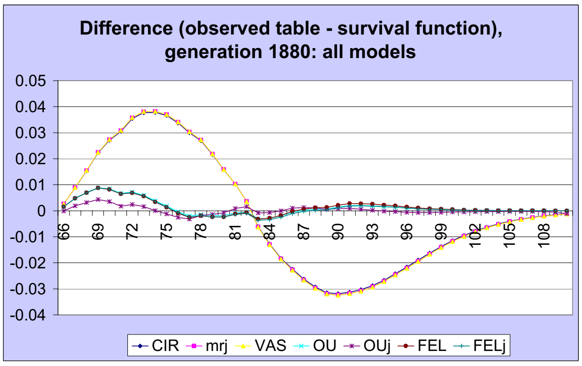
Graph 7



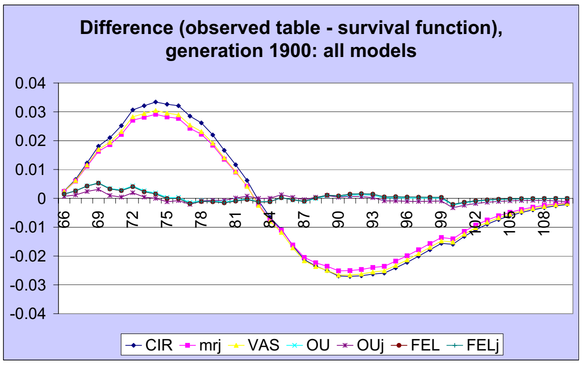
Graph 8

The fit is very good, also in the presence of strong rectangularization (the last two generations), and all the survival functions cannot be distinguished from each other.

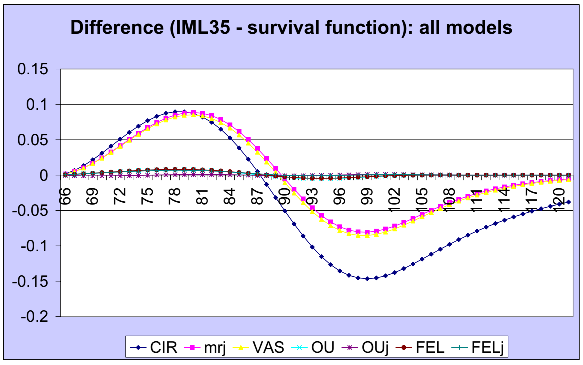
To have a better idea of the goodness of the fit, for each generation we plot the differences between the survival probabilities (${}_tp_{65}$) used as data and the survival function implied by the different models ($S_{65}(t)$). Graphs 9 to 12 report these differences for all the (seven) models considered so far, for generations 1880 to 1945.



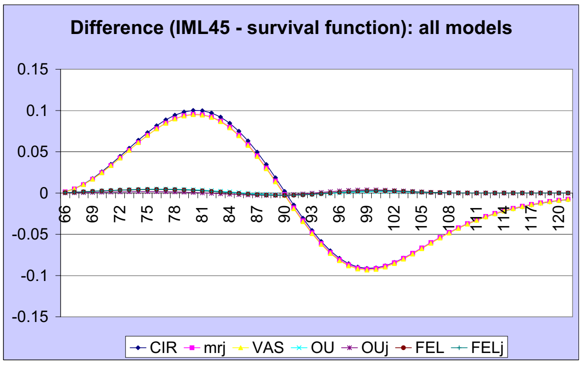
Graph 9



Graph 10



Graph 11



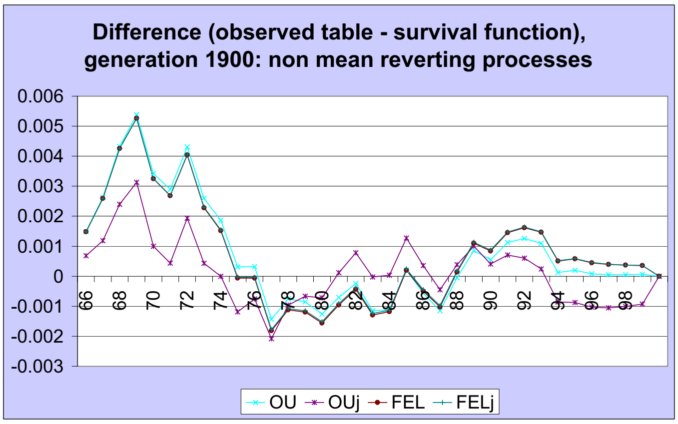
Graph 12

The improvement in the goodness of the fit when choosing a non mean reverting process for the intensity of mortality is evident.

It is not a surprising result that the differences in graphs 9 and 10 are very irregular, whereas they are smooth curves in graphs 11 and 12. Indeed, let us recall that the survival probabilities ${}_tp_{65}$ for graphs 9 and 10 are observed data, while they are projected probabilities for the generations 1935 and 1945.

To conclude, let us plot in Graph 13 the differences between the survival probabilities and their theoretical counterparts for generation 1900, for the non mean reverting processes only. Notice the difference in the scale w.r.t. the previous graphs: this phenomenon, which confirms the small overall difference in errors of Table 2, and which holds for the other generations, will allow us, in the next sections, to use interchangeably all the four models[5].

[5] 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



Graph 13

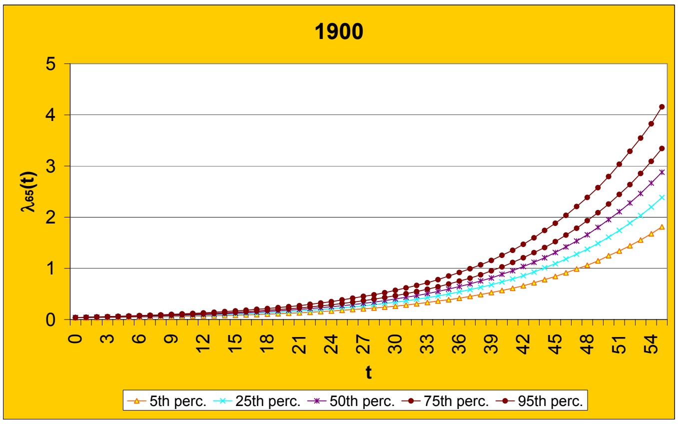
7 Impact of mortality randomness

Non mean reverting optimally fitted models present low diffusion parameters. This feature is evident both in the observed mortality tables and in the projected tables. While the explanation for the low value of σ in the latter case can be the fact that projected mortality tables are constructed in a deterministic way[6], the same explanation cannot apply for the observed generation tables. This section aims on the one side at showing that also relatively small volatility values can produce significant effects on the intensity and on the number of deaths; on the other side, it aims at investigating the effect of higher randomness, since low volatility does not need to occur also for future generations.

As for the first aim, we have simulated the process $\lambda_{65}(t)$ for the generation 1900 using the calibrated parameters of the FEL model. We have simulated 1000 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 14.

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.

[6] In UK the CMI bureau in projecting mortality rates uses a simple formula based on exponentials of polynomials.



Graph 14

In order to have a better understanding of the practical consequences of the volatility of the intensity process for an insurance company, we have considered a portfolio of 1000 ten-years term assurance policies sold to males aged 65 born in 1900. The 1000 paths of the intensity process simulated above have been used to simulate the number of deaths within ten years in 1000 different scenarios[7]. Table 3 reports some statistics of the distribution of the number of deaths.

TABLE 3

min	304
5th perc.	372
25th perc.	419
50th perc.	452
75th perc.	482
95th perc.	521
max	615
mean	450
st.dev.	46

The reader can appreciate the fact that, in spite of the small absolute values of the intensity parameters, the actual number of deaths can vary substantially. For this generation, for instance, the maximum number of death is more than the double than the minimum one.

As far as the effect of higher variability in the mortality intensity is concerned, we study it through the survival probabilities. Analytical results allow us to investigate this issue for the first two models, the OU and the OUj. Namely, the force of mortality in these two cases (see eqs. (5.29) and (5.31)) decreases when σ or l increase: therefore, the survival probability increases when the

[7]The methodology used for simulating the number of deaths in each scenario refers to an equivalent setting of the theoretical framework for Cox processes (see, for instance, Lando (2004)). According to this formulation, 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 certain 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 component increases[8]. Analytical results help to determine the behaviour of the survival function when σ increases also in the FEL case. Indeed, it can be shown that the function β of equation (5.20) is increasing in σ. This implies that the survival probability increases with the diffusive part.

As for the model with jumps, FELj, it turns out impossible to say something about the dependence of α from σ and l, since this involves the relationship between other coefficients like a and μ, which in general is not known. Therefore, for the FELj model one has to run sensitivity or stress test analysis in order to assess the impact of higher stochastic components on the survival function. As an example, we have done this for the generation 1900. The values of the parameters σ, l and μ have been increased with respect to the optimal values (collected in table 2). We have found that by increasing the values of σ and l the survival probability increases with respect to the optimal one, while changing the average magnitude of the single jumps (μ) does not lead to significant changes in the survival probabilities. The model therefore would predict a higher survivorship, if ever the stochastic components σ and l were higher than the calibrated ones.

8 Forecasting mortality and mortality trend

The calibration performed so far has been applied either to old generations, considering observed mortality tables, or to younger generations, considering projected mortality tables. The aim of the calibration has been to show the appropriateness of the non mean reverting models in describing the intensity of mortality. Once this has been done, our next step consists in making the model applicable to younger generations so as to allow forecasting.

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 following the calibration procedure used above, on 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. For instance, we have considered eleven different generations of persons born in every year from 1900 to 1910 for two different initial ages: 35 and 65. We have followed the diagonal approach described above until 1998 (the last year in which data are available in the Human Mortality Database at the time of writing the paper, for males population of England and Wales).

[8]As an alternative, one can observe that the function α of equation (5.6) is increasing in σ and that the function α of equation (5.12) is increasing in both σ and l.

Once the required observed survival probabilities are available, 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. For 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. For 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. We will differentiate between these two different methods by calling them *forecasting mortality* and *mortality trend*, respectively. We will come back again to these two different procedures in sections 8.1.1 and 8.1.2, and give some examples in sections 8.2.1 and 8.2.2.

Before proceeding with the calibration results, it is worth spending a few words on the effect of changing either the initial age or the generation.

8.1 Some considerations on the effect of changing initial age or generation

In order to discuss mortality forecast and mortality trend, it is convenient to present the whole family of intensity processes we are considering.
We first want to explain what happens to the value of the parameters of the process when we change initial age inside the same generation. This can be used for mortality forecasting at different initial ages.
Then we want to see what happens when we change generation, holding the same initial age. This is the base for mortality trend.

8.1.1 Changing initial age, given the generation

Imagine to describe the evolution of mortality intensity for a given generation[9] (in order to keep things simple, we will not introduce any index for the generation). Observe that the intensity process described so far, equation (3.4) should be written more properly as:

$$d\lambda_x(t) = f_x(\lambda_x(t))dt + g_x(\lambda_x(t))dW_x(t) + dJ_x(t) \tag{8.1}$$

where the dependence of the drift, the diffusion and the jump components on the initial age x is put into evidence by the index x. For example, in the case of the OU process, if x and y are different initial ages, we will have:

$$d\lambda_x(t) = a_x\lambda_x(t)dt + \sigma_x dW_x(t)$$

$$d\lambda_y(t) = a_y\lambda_y(t)dt + \sigma_y dW_y(t).$$

It can be shown that if $\sigma = 0$ for any age, then we will have $a_x = a$ for any age x. However, in general, the calibrated parameters are age dependent, i.e. it is $a_x \neq a_y$ and $\sigma_x \neq \sigma_y$. The same considerations apply for the other processes and the other parameters (l and μ). Therefore, when we change the initial age we expect to find different values for the optimal parameters. The fact that we do find different values when changing initial age (in fact, it is $a_{35} \neq a_{65}$ in all cases, for each generation analyzed) is a clear confirmation of the fact that it must be $\sigma \neq 0$, and that, therefore, assuming a simple Gompertz force of mortality cannot be considered appropriate.

[9] By generation we mean year of birth.

8.1.2 Changing generation, given the initial age

Let us consider the intensity of mortality for a given initial age x and different generations. A complete description of the intensity surface would be given by a two parameters-family $\lambda_{x,gen}$ [10]. However, for simplicity, here we focus only on the change of generation and omit the initial age x. We have a family of intensity processes:

$$d\lambda_{gen}(t) = f_{gen}(\lambda_{gen}(t))dt + g_{gen}(\lambda_{gen}(t))dW_{gen}(t) + dJ_{gen}(t) \tag{8.2}$$

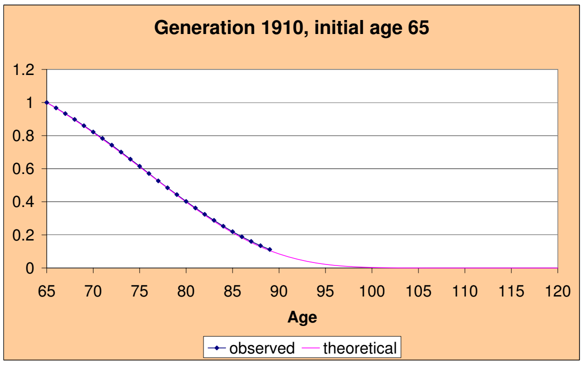
where the index gen refers to the year of birth[11].

The change in $\lambda_{gen}(0)$ and in the parameters that characterize f_{gen} and g_{gen} gives the description of the mortality trend in our setting.

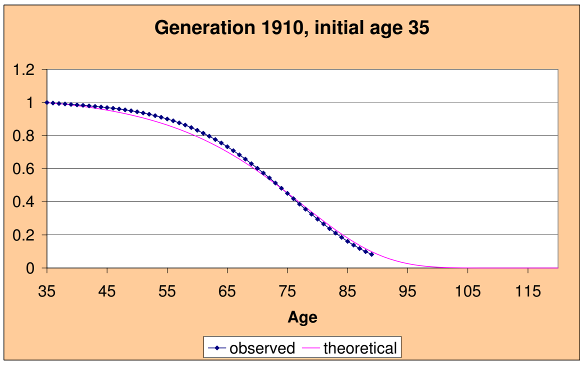
8.2 Calibration results

8.2.1 Mortality forecasting

As an illustration, Graphs 15 and 16 report the mortality forecast for the generation 1910 with initial ages 65 and 35. The graphs report the observed and the theoretical survival function according to the FELj model.



Graph 15



Graph 16

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

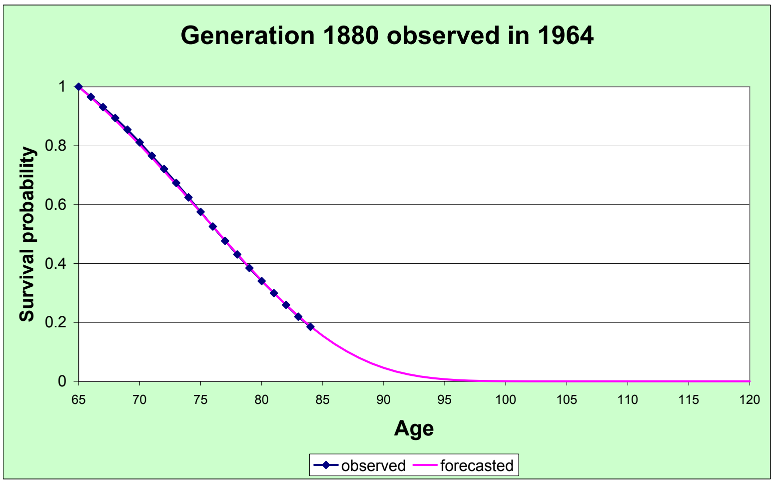
[10] For a description of the intensity mortality surface via random fields with application to the pricing of insurance products, see Biffis and Millossovich (2005).

[11] For example, in the case of the OU process, if we were considering the generations 1880 and 1905, we would have:

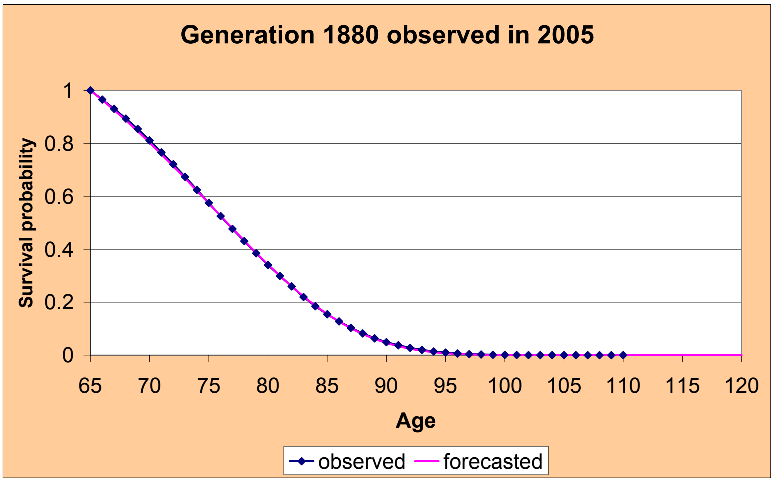
$$d\lambda_{1880}(t) = a_{1880}\lambda_{1880}(t)dt + \sigma_{1880}dW_{1880}(t)$$

$$d\lambda_{1905}(t) = a_{1905}\lambda_{1905}(t)dt + \sigma_{1905}dW_{1905}(t)$$

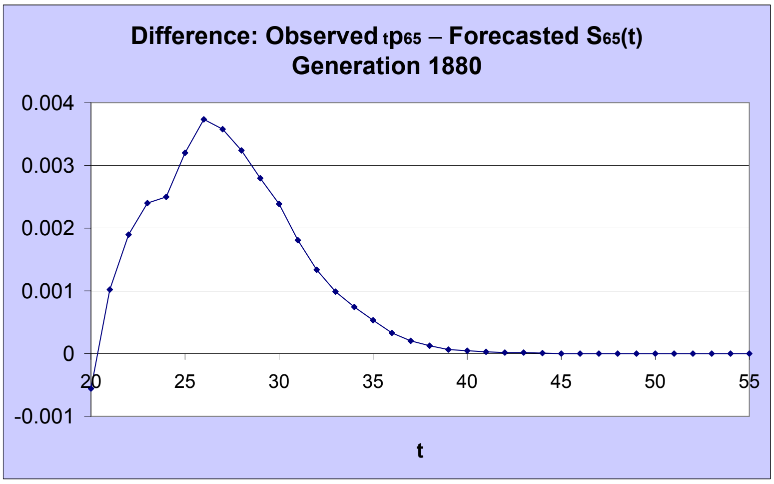
In order to check whether this forecast procedure gives reliable results (by comparing the forecasted mortality with the experienced one), we have applied it on the generation 1880, initial age 65. We have calibrated the parameters of the process taking year 1964 as observation date. Graph 17 reports the forecasted and the observed survival probabilities in 1964, graph 18 reports them in 2005, and graph 19 reports the differences between the survival probabilities experienced after the observation date and the forecasted ones.



Graph 17



Graph 18



Graph 19

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.

8.2.2 Mortality trend

In order to investigate the mortality trend, we follow two different approaches.

First approach
For the first approach, we have made the calibration for the eleven generations born in years 1900 to 1910 for initial ages 35 and 65. We show results only for $x = 65$. In what follows, we will adopt the notation:

$$\lambda_{gen}(t) \qquad gen = 1900, ..., 1910$$

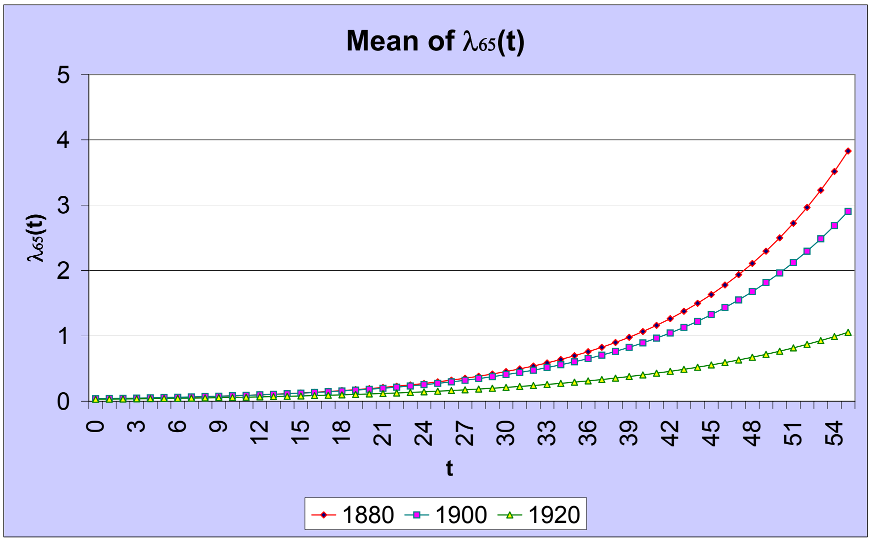
omitting the initial age 65 for notational convenience. The model selected is the FELj. For each generation we calculate the value of $\lambda(0)$ and we find a set of optimal parameters: $a_{gen}, \sigma_{gen}, l_{gen}, \mu_{gen}$, as well as the calibration error, $error_{gen}$. Table 4 reports these values:

TABLE 4

gen	a	σ	l	μ	$error$	$\lambda(0)$
1900	0.079247	0.014341	0.001	-0.0002	0.000118	0.037972
1901	0.079095	0.015006	0.000999	-0.0002	0.000085	0.037401
1902	0.084148	0.019692	0.000999	-0.0002	0.00003	0.035296
1903	0.078144	0.014075	0.000999	-0.0002	0.000062	0.036239
1904	0.07024	0.00021	0.001	-0.0002	0.000784	0.037546
1905	0.073065	0.006091	0.00025	-0.0001	0.000187	0.036228
1906	0.076096	0.013321	0.001	-0.0002	0.000033	0.034819
1907	0.071454	0.000444	0.001	-0.0002	0.000249	0.035192
1908	0.0734	0.002512	0.001	-0.0002	0.000142	0.034146
1909	0.074123	0.002528	0.001	-0.0002	0.000064	0.033443
1910	0.07376	0.000866	0.001001	-0.0002	0.000101	0.033112

We observe an evident linearly decreasing trend of $\lambda(0)$. Indeed, its linear regression on the calendar year gives an R^2 of 0.912. This is consistent with our intuition about mortality trend. The calibration errors are very small. One could try to fit a polynomial to each parameter behaviour so as to be able to extrapolate the parameters for the next generations.

Second approach
For the second approach, we have simulated the process $\lambda_{65}(t)$ with the same methodology presented in section 7 for the generations 1880, 1900 and 1920. Graph 20 reports the mean of $\lambda_{65}(t)$ for the three generations. 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.



Graph 20

9 Summary and concluding remarks

In this paper, we have described the evolution of mortality using doubly stochastic (or Cox) processes. The time of death has been modelled as a doubly stochastic stopping time: namely, as a jump time whose intensity is stochastic. The intensity has been described as a univariate time-homogeneous affine process, with two different specifications: first, as in the default risk literature, with mean reversion, then without it. For both specifications, the survival probabilities have been provided in closed form.

The intensity processes have been calibrated to the population of England and Wales, using observed mortality tables for old generations and projected tables for younger ones. Results from the calibration suggest that, in spite of their popularity in the financial context, time-homogeneous mean reverting processes are not suitable for describing the death intensity of individuals. On the contrary, affine processes whose deterministic part increases exponentially seem to be appropriate. We propose four of such processes, which are different in their stochastic part. The analysis of the relation between the stochastic intensity of mortality and the deterministic force of mortality has shown that the proposed processes can be considered natural extensions of the Gompertz model.

In the calibrations the diffusive parameter is often relatively small. However, we show through simulations that the impact of such a randomness on the actual number of deaths can be significant. In addition, stress analysis and analytical results, whenever they can be obtained, indicate that increasing the randomness of the intensity processes results in improvements in survivorship.

After having specified the dependence of the model parameters on the initial age and the generation, we provide procedures for mortality forecasting and mortality trend assessment, which describe future evolution of mortality within a single cohort and between different cohorts, respectively. In particular, we have given mortality forecasts for a given generation and different initial ages. We have also showed that an application of the same forecasting procedure to an older generation gives very satisfactory results. As far as the mortality trend is concerned, on the one side we have calibrated the same model for a sequence of consecutive generations, same initial age; on the other side we have compared the expected intensity of different generations.

To sum up, it seems to us that the proposed models are flexible and appropriate for reliable actuarial applications, in spite of their analytical simplicity and theoretical limitations with respect to time-inhomogeneous models.

References

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

Artzner, P. and Delbaen, F. (1992). Credit risk and prepayment option, *ASTIN Bulletin* **22**: 81–96.

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

Biffis, E. and Millossovich, P. (2005). A bidimensional approach to life portfolio valuation and to the insurer's future business, *working paper.*

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. (2005). Pricing death: Framework for the valuation and securitization of mortality risk, *to appear on ASTIN Bulletin.*

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–407.

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

Dahl, M. and Møeller, T. (2005). Valuation and hedging of life insurance liabilities with systematic mortality risk, *Proceedings of the* 15^{th} *AFIR Colloquium, Zurich.*

Denuit, M. and Devolder, P. (2005). Continuous time stochastic mortality and securitization of longevity risk, *working paper.*

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

Duffie, D. (2002). A short course on credit risk modeling with affine processes, *Associazione Amici della Scuola Normale Superiore.*

Duffie, D. and Singleton, K. J. (1999). Modeling term structures of defaultable bonds, *Review of Financial Studies* **12**: 687–720.

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.

Grandell, J. (1976). *Doubly Stochastic Poisson Processes*, Lecture Notes in Mathematics, Number 529, New York: Springer Verlag.

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

Lando, D. (1994). Three essays on contingent claims pricing, PhD Thesis, Cornell University.

Lando, D. (2004). *Credit Risk Modeling*, Princeton: Princeton University Press.

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.

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. (2005). Affine stochastic mortality, *to appear on Insurance: Mathematics and Economics*.

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.

Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**: 177–188.