

MEAN REVERSION IN STOCHASTIC MORTALITY: WHY AND HOW?

Zeddouk, F. and P. Devolder

REPRINT | 2020 / 18

Mean reversion in stochastic mortality: why and how?

By
Fadoua Zeddouk *and* Pierre Devolder

Abstract

Life insurance companies use stochastic models to forecast mortality. According to the literature, non-mean reversion models are more suitable for mortality modelling than mean reversion models with a fixed long-term target. In this paper, we adopt stochastic affine processes for the force of mortality and study the impact of adding a time-dependent long-term mean reversion level to two non-mean-reverting processes. We calibrate the models to different generations of the Belgian population and assess these models' abilities to predict mortality using different statistical methodologies. The backtest shows that the survival curves provided by the mean-reverting processes are closer to reality. Thus, we conclude that incorporating a time-dependent target into these considered models improves their performance significantly.

1 Introduction

For life insurance companies and pension funds, mortality risk has been traditionally estimated using a simplified approach to reality, based on deterministic life tables projected from historical mortality rates and added safety margins. While this traditional method may be prudent for risk management, it does not account for uncertainty related to mortality evolution, since a deterministic mortality scenario is applied to risk management and pricing. Ultimately, this approach may lead to a flawed evaluation of life insurance products. Moreover, under the Solvency II Framework Directive, solvency capital must be held to avoid the risk of ruin. This capital is related to the risk of deviation of future mortality based on life expectancy and, therefore, should use stochastic projections. Thus, stochastic models for mortality are needed to measure the trend of improvement in mortality over the next few decades, along with associated uncertainty in the projections. By modelling mortality as a stochastic process, reserves can expand to include uncertainty associated with future development of mortality intensity. This method should allow for a more accurate assessment of future commitments that account for the systematic risk of mortality. Moreover, the development of mortality and longevity-related securities requires the

use of stochastic models to accurately price financial instruments according to demographic risks.

To better predict longevity risk, research and development efforts have recently focused on stochastic mortality models. The Lee-Carter model [12], was the first model to consider increased life expectancy trends in age mortality dynamics and has been used for stochastic forecasts of the U.S. Social Security system and other aspects of the U.S. federal budget (see Congressional Budget Office of the United States ¹). Even though data are typically reported at discrete intervals (e.g. annually), other models have considered the development of mortality rates in continuous time.

Since the 2000s, with requirements of Solvency II for the assessment of mortality risk, both insurance companies and academic researchers have become increasingly interested in continuous-time stochastic mortality models, and various continuous-time stochastic mortality models have been proposed in the literature. Our framework focuses on mortality and modelling the survival function of individuals with one-factor short-rate models which are used in finance to model changes in the term structure of interest rates. The similarities between mortality and interest rate dynamics were exploited by Milevsky and Promislow [14], Cairns et al. [3], Biffis [2], Dahl [6], Schrager [16], and Denuit and Devolder [7] to model the force of mortality, using tools and techniques developed in interest rate modelling.

In this context, the eventual mean-reverting property of the model is essential. The analogy between interest rates and the force of mortality models motivates the issue of mean reversion models. The mean reversion concept is highly intuitive and combines long-term convergence to an equilibrium level (the trend of the process), with stochastic noise (the diffusion part of the model).

Milevsky and Promislow [14] were the first to propose a stochastic force of mortality model. Several stochastic affine models facilitate mean-reverting characteristics; however, according to Cairns et al. [5], affine stochastic models should not incorporate mean-reverting elements. Moreover, Luciano and Vigna [13] states that non-mean-reverting affine processes are better for modelling the force of mortality. These conclusions were based on fixed-target reversion models (e.g., the Vasicek model). We revisit this comparison by considering fixed-target mean-reverting models as well as moving-target mean-reverting models. Not surprisingly, a fixed-time target is inconvenient in representing the evolution of time on mortality intensity, which increases with age. Thus, examining a mean-reverting model with an increasing target is more relevant, especially with an exponential target consistent with the classical Gompertz model. Hence, we explore the impact of adding a time-dependent long-term mean element to the Ornstein-Uhlenbeck (OU) and Feller processes. In this paper, we show that introducing a moving target in the mean reversion process strongly improves the performance of these models when compared to non-mean-reverting models. This paper is organised as follows. Section 2 offers a brief review of the

¹Congressional Budget Office Of The United States, "Long-Term Budgetary Pressures and Policy Options." (1998))

longevity risk, and presents the considered mean and non-mean reversion processes. Section 3 provides an in-depth comparison between these models using statistical methodologies, including calibration, mortality trend estimation, and backtesting. Section 4 presents a multicohort vision, and Section 5 concludes this paper.

2 Models presentation

We model mortality using a single explanatory variable: the intensity of mortality μ_x at age x defined by:

$$\mu_x = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[x < \bar{T} \leq x + \Delta \mid \bar{T} > x]}{\Delta}, \quad (1)$$

where $\bar{T}$ is a random variable representing the future lifespan of an individual.

Within our framework, we model the intensity of mortality $\mu_x(t)$ as a continuous stochastic process, defined on the real probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to its natural filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$. Thus, μ_{x+t} is written as: $\mu_x(t, \omega)$, with ω an element of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Therefore, the starting point of our modelling describes the dynamics of the mortality through the specification of a stochastic differential equation (SDE). Afterwards, we focus on the survival process, which represents the index at time t of an individual initially aged x , and alive at time t to survive $T - t$ years more. For one individual, the survival process is defined as follows:

$$S(x, t, T, \omega) = e^{-\int_t^T \mu_x(u, \omega) du}. \quad (2)$$

At time t , the survival probability is obtained by accounting for the expectation of the survival process (2), i.e.:

$${}_{T-t}p_{x+t} = E_{\mathbb{P}}[S(x, t, T, \omega) \mid \mathcal{F}_t]. \quad (3)$$

Several stochastic mortality models have been proposed in the literature. According to Luciano and Vigna [13], constant mean reversion models are not suitable to describe the mortality of individuals; thus, non-mean reversion models work better.

While we aim to study two mean reversion models, we first incorporate a moving mean-reverting target, inspired by the Gompertz model, and compare the performances between a mean reversion model with a constant mean level and a non-mean reversion model. Therefore, we consider the following five affine stochastic processes:

- Ornstein-Uhlenbeck (OU)
- Feller
- Vasicek
- Hull & White (HW)

- Cox-Ingersoll-Ross extended (CIR)

For simplicity, we write $\mu_x(t, \omega) = \mu_x(t)$ to represent the intensity of mortality of an individual having the age $x + t$ at time t , and $w(t, \omega) = w(t)$, a standard Brownian motion.

The advantage of considering affine models for modelling $\mu_x(t)$ is the ease of finding an explicit expression for the survival probability (3). Therefore, for affine processes under technical conditions (see (Duffie et al. [8]), the properties of the classical stochastic process allow to write:

$$\begin{aligned} {}_{T-t}p_{x+t} &= E_{\mathbb{P}}[S(x, t, T, \omega) \mid \mathcal{F}_t] \\ &= e^{\alpha(t, T) - \beta(t, T)\mu_x(t)}, \end{aligned} \quad (4)$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are two deterministic functions and solutions of the Riccati generalized ordinary differential equations that can be determined numerically, and in some cases analytically (see for instance Duffie et al. [9]).

2.1 Non-mean reversion models

2.1.1 Ornstein-Uhlenbeck model

The force of mortality is assumed to satisfy the SDE:

$$d\mu_x(t) = a\mu_x(t)dt + \sigma dw(t), \quad (5)$$

with $a > 0$, $\sigma > 0$ and $w(t)$ a standard Brownian motion.

As the OU model is Gaussian, we can easily find the expression of the survival probability by determining its distribution. We can also use the affine framework that gives the following system of ODEs for $\alpha(t, T)$ and $\beta(t, T)$ (see for example [13]):

$$\begin{cases} \frac{d\alpha(t, T)}{dt} = \frac{1}{2}\sigma^2\beta(t, T)^2 \\ \frac{d\beta(t, T)}{dt} = a\beta(t, T) - 1 \end{cases} \quad (6)$$

with boundary conditions $\alpha(T, T) = \beta(T, T) = 0$.

By solving the system (6) we find:

$$\begin{cases} \alpha(t, T) = \frac{\sigma^2(T-t)}{2a^2} + \frac{\sigma^2}{4a^3}e^{2a(T-t)} - \frac{\sigma^2}{a^3}e^{a(T-t)} + \frac{3\sigma^2}{4a^3} \\ \beta(t, T) = \frac{1}{a}(1 - e^{a(T-t)}). \end{cases} \quad (7)$$

The solution $\mu_x(t)$ of the SDE (5) is:
($0 \leq s \leq t \leq T$)

$$\mu_x(t) = \mu_x(s)e^{a(t-s)} + \sigma \int_s^t e^{a(t-u)} dw(u). \quad (8)$$

In this model, the force of mortality has always a strictly positive probability of becoming negative. This probability is given by:

$$\mathbb{P}[\mu_x(t) < 0 \mid \mathcal{F}_s] = \Phi\left(-\frac{\mu_x(s)e^{a(t-s)}}{\sigma\sqrt{\frac{e^{2a(t-s)}-1}{2a}}}\right) = \Phi(\zeta(a, \sigma)), \quad (9)$$

where $\Phi(\cdot)$ is the distribution function of a standard normal random variable.

The function $\zeta(a, \sigma)$ is an increasing function of σ and a decreasing function of a . With this model, we can obtain negative values of $\mu_x(t)$, which is a disadvantage, but in practice, this probability tends to be very small, since the values of σ and a are respectively small and high enough. This is confirmed by calibration results in our simulation.

2.1.2 Feller model

The SDE of the Feller process describing the force of mortality dynamics is given by:

$$d\mu_x(t) = a\mu_x(t)dt + \sigma\sqrt{\mu_x(t)}dw(t), \quad (10)$$

with $a > 0$, $\sigma > 0$ and $w(t)$ a standard Brownian motion.

Compared to the OU process, the Feller process avoids negative values for the force of mortality but can reach zero with a positive probability.

In a Feller setting, $\alpha(\cdot)$ and $\beta(\cdot)$ are determined using the Feynman-Kac representation theorem [11].

The survival probability (4) is the solution of the following differential equation (PDE):

$$\begin{cases} \frac{\partial p}{\partial t} + a\mu_x \frac{\partial p}{\partial \mu_x} + \frac{1}{2}\sigma^2\mu_x \frac{\partial^2 p}{\partial \mu_x^2} - \mu_x p = 0 \\ {}_0p_{x+t} = 1. \end{cases} \quad (11)$$

By substituting (4) in the PDE (11), we obtain a system of Riccati's ODE that $\alpha(t, T)$ and $\beta(t, T)$ must verify:

$$\begin{cases} \frac{d\alpha(t, T)}{dt} = 0 \\ \frac{d\beta(t, T)}{ds} = 1 - a\beta(t, T) - \frac{1}{2}\sigma^2\beta^2(t, T) \\ \alpha(T, T) = \beta(T, T) = 0. \end{cases} \quad (12)$$

The unique solution of system (12) is given by:

$$\begin{cases} \alpha(t, T) = 0 \\ \beta(t, T) = \frac{1-e^{(b(T-t))}}{c+d \cdot e^{(b(T-t))}} \\ \beta(T, T) = 0, \end{cases} \quad (13)$$

with $b = \sqrt{a^2 + 2\sigma^2}$, $c = \frac{b+a}{2}$ and $d = \frac{b-a}{2}$.

2.2 Mean reversion models

2.2.1 Vasicek model

The Vasicek model of mortality is given by (see for instance [13]):

$$d\mu_x(t) = (a - b\mu_x(t))dt + \sigma dw(t), \quad (14)$$

with $a > 0$, $b > 0$, $\sigma > 0$, $w(t)$ a standard Brownian motion.

The drift of this model presents a mean reversion to a constant target $\frac{a}{b}$, realised with a speed b .

As the Vasicek model is affine, the survival probability takes the form:

$${}_{T-t}p_{x+t} = e^{\alpha(t,T) - \beta(t,T)\mu_x(t)}, \quad (15)$$

$\alpha(t, T)$ and $\beta(t, T)$ are given by (see for instance [13]):

$$\begin{cases} \alpha(t, T) &= (\frac{a}{b} - \frac{\sigma^2}{2b^2})(\frac{1}{b}(1 - \exp(-b(T-t))) - T + t) \\ &\quad - \frac{\sigma^2}{4b^3}(1 - \exp(-b(T-t)))^2 \\ \beta(t, T) &= \frac{1}{b}(1 - e^{-b(T-t)}). \end{cases} \quad (16)$$

Luciano and Vigna [13] considered the OU, Feller, and Vasicek models and demonstrated that non-mean-reverting models outperform mean-reverting models. This finding leads us to question the idea of mean reversion. Nevertheless, the poor performance of the Vasicek model is not surprising, not due to its mean-reverting property, but rather because of its fixed-target $\frac{a}{b}$. Indeed, contrary to interest rates models, the force of mortality has, in essence, an increasing trend with age and thus with time in equation (14). Therefore, it seems essential to recheck the performance of the mean-reverting models with an increasing target. A natural candidate in this perspective is to use as increasing target an exponential form, in line with the classical Gompertz model.

2.2.2 HW model

In order to assess the impact of using a moving target instead of a fixed target, we consider the Hull and White model [10], a time-dependent extension of the Vasicek model, where the force of mortality tends to return to a moving-target $\frac{\xi(t)}{b}$ with restoring force b . The expression of the model is as follows:

$$d\mu_x(t) = (\xi(t) - b\mu_x(t))dt + \sigma dw(t), \quad (17)$$

with $b > 0$, $\sigma > 0$, $w(t)$ a standard Brownian motion, and $\xi(t)$ a deterministic function.

The form of the target is naturally given by an exponential function, with the same kind of evolution as the trend of the non-mean-reverting models:

$$\xi(t) = Ae^{Bt}, \quad (18)$$

with A and B two strictly positive constants. $\xi(t)$ corresponds to the classical function of Gompertz, which expresses that the force of mortality increases exponentially with age, where $\frac{A}{b}$ is the baseline mortality at age x , and B is the senescent component.²

The Hull and White model is then given by:

$$d\mu_x(t) = b\left(\frac{A}{b}e^{Bt} - \mu_x(t)\right)dt + \sigma dw(t). \quad (19)$$

If we consider that $\frac{A}{b} = \mu_x(0)$, this equation becomes:

$$d\mu_x(t) = b(\mu_x(0)e^{Bt} - \mu_x(t))dt + \sigma dw(t). \quad (20)$$

Since the Hull and White model is Gaussian, we can easily find the expression of the survival probability by determining its distribution. We can also use the affine framework that gives the following system of ODEs for $\alpha(t, T)$ and $\beta(t, T)$:

$$\begin{cases} \frac{d\alpha(t, T)}{dt} = \beta(t, T)\xi(t) - \frac{1}{2}\sigma^2\beta^2 \\ \frac{d\beta(t, T)}{dt} = \beta(t, T)b - 1, \end{cases} \quad (21)$$

with the boundary conditions $\alpha(T, T) = \beta(T, T) = 0$.

$\alpha(t, T)$ and $\beta(t, T)$ are given by:

$$\begin{cases} \alpha(t, T) = \frac{A}{b} \left[e^{-bT} \frac{e^{(B+b)T} - e^{(B+b)t}}{B+b} - \frac{e^{BT} - e^{Bt}}{B} \right] - \frac{\sigma^2}{2b^2} \left[\frac{1}{b}(1 - e^{-b(T-t)}) - T + t \right] \\ \quad - \frac{\sigma^2}{4b^3} (1 - e^{-b(T-t)})^2 \\ \beta(t, T) = \frac{1}{b}(1 - e^{-b(T-t)}). \end{cases} \quad (22)$$

By short calculations we find that the force of mortality in the HW model is given by:

$$(0 \leq s \leq t \leq T)$$

$$\mu_x(t) = \mu_x(s)e^{-b(t-s)} + \frac{A}{b+B}(e^{Bt} - e^{Bs-b(t-s)}) + \sigma e^{-bt} \int_s^t e^{bu} dw(u), \quad (23)$$

and, conditionally on $\mathcal{F}_s$, normally distributed with the following expectation and variance:

²The Gompertz model is an exponential process given by:

$$\begin{aligned} \mu_x(t) &= A' e^{B(x+t)} \\ &= A' e^{Bx} e^{Bt} \\ &= \mu_x(0) e^{Bt}. \end{aligned}$$

$$\begin{cases} E_{\mathbb{P}}[\mu_x(t) \mid \mathcal{F}_s] = \mu_x(s)e^{-b(t-s)} + \frac{A}{b+B}(e^{Bt} - e^{Bs-b(t-s)}) \\ Var_{\mathbb{P}}[\mu_x(t) \mid \mathcal{F}_s] = \frac{\sigma^2}{2b}(1 - e^{-2b(t-s)}). \end{cases} \quad (24)$$

As in the OU model, for each time t , the intensity of mortality can be negative with a positive probability:

$$\Phi\left(-\frac{\mu_x(s)e^{-b(t-s)} + \frac{A}{b+B}(e^{Bt} - e^{Bs-b(t-s)})}{\sqrt{\frac{\sigma^2}{2b}(1 - e^{-2b(t-s)})}}\right) = \Phi(\psi(A, B, b, \sigma)), \quad (25)$$

which is usually negligible in practice.

2.2.3 Cox-Ingersoll-Ross (CIR) extended model

The CIR model was developed by Cox et al. [5] and is known as square root process. We consider here an extension of the CIR by considering a time-dependent long-term mean. The CIR extended process for the force of mortality is characterised by:

$$d\mu_x(t) = (\xi(t) - b\mu_x(t))dt + \sigma\sqrt{\mu_x(t)}dw(t), \quad (26)$$

with $b > 0$, $\sigma > 0$, $w(t)$ a standard Brownian motion.

To compare the two models, we choose the same expression for $\xi(t)$ as for the model of Hull and White (Gompertz model).

Unlike the model of Hull and White, this process has the advantage of ensuring the positivity of the intensity of mortality, and due to the presence of the square root, the intensity process can never be negative. The intensity process can reach zero, but in practice, this probability is negligible. In our framework, we prevent the intensity process from reaching 0 by imposing the following sufficient condition (e.g., Mishura and Posashkova [15]): $A > \frac{\sigma^2}{2}$.

Nevertheless, this improvement has a price in the sense that the intensity of mortality cannot be handled analytically.

The CIR model is affine but not Gaussian, (26) is a complicated stochastic process with a distribution difficult to evaluate. We use the Feynman-Kac representation to determine the survival function. By short calculations we get:

$$\begin{cases} \frac{d\alpha(t, T)}{dt} = \xi(t)\beta(t, T) \\ \frac{d\beta(t, T)}{dt} = b\beta(t, T) + \frac{1}{2}\sigma^2\beta^2(t, T) - 1, \end{cases} \quad (27)$$

with the boundary conditions $\alpha(T, T) = \beta(T, T) = 0$, and $\gamma = \frac{(b^2 + 2\sigma^2)^{\frac{1}{2}}}{2}$. $\alpha(t, T)$ and $\beta(t, T)$ are given by:

$$\begin{cases} \alpha(t, T) = 2A \int_t^T \frac{e^{Bs}(e^{\gamma(T-s)} - e^{-\gamma(T-s)})}{(2\gamma+b)e^{\gamma(T-s)} + (2\gamma-b)e^{-\gamma(T-s)}} ds \\ \beta(t, T) = \frac{\sinh(\gamma(T-t))}{\gamma \cosh(\gamma(T-t) + \frac{1}{2}b \sinh(\gamma(T-t)))}. \end{cases} \quad (28)$$

3 Models comparison

In this section, we evaluate the effectiveness of the aforementioned models to predict mortality using statistical methodologies. We calibrate the models on historical and projected mortality data. Next, we compare the mortality trend for the different models. Then, we assess the quality of these calibrations by performing a backtest. Finally, we use the models to predict mortality.

3.1 Calibration

We aim to calibrate our models on different historical survival processes, corresponding to four different cohorts of individuals, born in 1900, 1915, 1965, and 1970, respectively, and initially aged 50 years old. Hence, we calibrate our models on mortality data of the Belgian population (unisex data) considering the following assumptions:

- We consider the 1900 and 1915 generations, we use historical data from the Human Mortality Database³, we have unisex annual death rates for ages 0 to 110, from the years 1841 to 2015;
- We consider the 1965 and 1970 generations, we use projected data from the IABE projected generational mortality table⁴ starting from the year 2015 to 2060.
- Initial age of observation $x_0 = 50$.
- $\mu_{50}(0) = -\ln(s_{50})$, where s_{50} is the one-year survival probability for an individual initially aged $x_0 = 50$.

We have chosen these particular generations and $x_0 = 50$ as initial age of observation to have approximately the same number of observations (50 for historical data and 45 for projected data).

Let $(s_1, s_2, \dots, s_n)$ be the annual survival probabilities assumed to be functions of both the age and the calendar year extracted from both life tables. We denote by ${}_t\hat{p}_{50}$ the probability that an individual aged $x_0 = 50$ reaches age $x_0 + t$. ${}_t\hat{p}_{50}$ is computed as follows:

$${}_t\hat{p}_{50} = \prod_{i=50}^{50+t-1} s_i. \quad (29)$$

Let ${}_t p_x(\theta)$ be the function to calibrate and defined by equation (3), where θ is the vector of the model's parameters. To estimate θ , we use Least Square Estimation (LSE). While there are many ways to calibrate each of these models, different calibration techniques yield different results. For instance, we could estimate the parameters using the maximum likelihood (MLE) or method of moments techniques. However, for this study, the LSE method was the easiest to implement. We minimise the least square errors on the survival probabilities

³Available at www.mortality.org

⁴Mortality projection for the Belgian population (2015), available at <https://www.iabe.be>

computed between the observed survival probabilities (around which we want to calibrate our models) and the predicted mortality by the stochastic model:

$$R = \min_{\theta} \sum_{j=0}^n ({}_j p_x(\theta) - {}_j \hat{p}_x)^2. \quad (30)$$

For the Least Square minimisation, it is necessary to specify the initial parameters where the R routine starts searching the minimum of (30). To solve this non-linear system, we use the generalised Newton-Raphson method provided by R. Since the final result is highly sensitive to these starting values, we generate various initial parameters randomly and calculate some optimisation steps. The values with minimal residual are taken as start parameters for the final LS optimisation, which supplies the final parameters. We use this type of algorithm as well for all other calibration issues described in the following sections.

3.1.1 Non-mean reversion models

We calibrate the OU and Feller models to the data using the above described algorithm. We find the optimal parameters reported in Table 1.

	1900	1915	1965	1970
$\mu_{50}(0)$	0.007909195	0.005820909	0.002850792	0.002600332
OU-error	0.005885259	0.003403252	0.055996630	0.018659380
a -OU	0.080696570	0.082023980	0.078277481	0.074116615
σ -OU	0.000000129	0.000000130	0.000000177	0.000000036
FELL-error	0.005885342	0.003403291	0.055996670	0.018659451
a -FELL	0.080687850	0.082019520	0.078272931	0.074114700
σ -FELL	0.000001527	0.000010657	0.000001104	0.000011621

Table 1: Optimal parameter values for the survival function in the non mean-reverting models.

The values of the parameters are quite coherent in both models, and parameter a has a decreasing trend. The values of σ are very small which confirms that the volatility of the mortality is much lower as compared to the financial market. Also, the residuals are more important for younger generations.

Figures 1 and 2 present calibration results graphically. As predicted by Table 1 above, we see that the fit is acceptable for older generations and poor for younger generations:

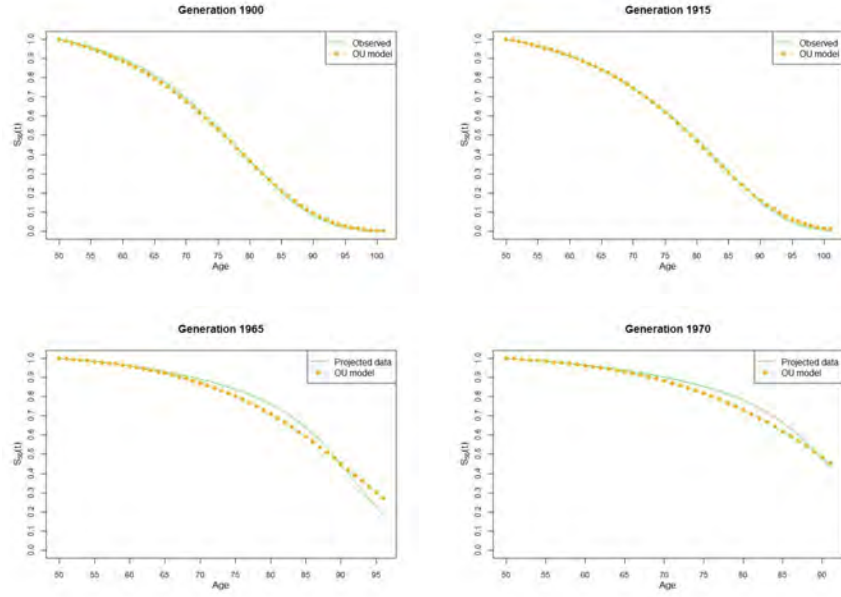


Figure 1: Survival functions, 1900,1915,1965 and 1970 generation for OU model.

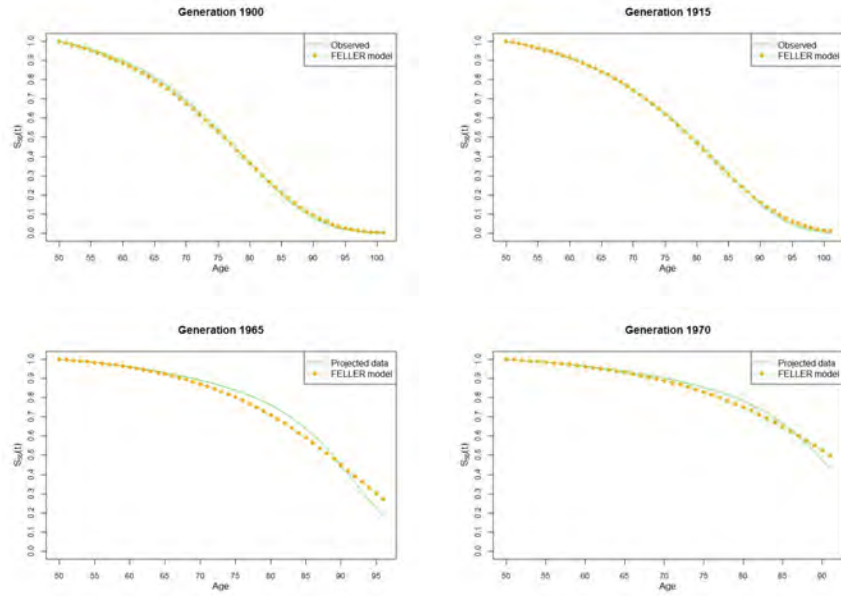


Figure 2: Survival functions, 1900,1915,1965 and 1970 generation for Feller model.

3.1.2 Mean reversion models

Table 2 reports the optimal calibration parameters for Vasiček, HW and CIR models:

	1900	1915	1965	1970
$\mu_{50}(0)$	0.007909195	0.005820909	0.002850792	0.002600332
Vasi-error	0.355228820	0.379264950	0.345576900	0.145937100
Vasi- a	0.002379388	0.001945061	0.000852597	0.000997105
Vasi- b	0.023799120	0.020618021	0.009913990	0.055305637
Vasi- σ	0.000027800	0.000009621	0.000307905	0.000652782
Vasi- a/b	0.009977980	0.094337871	0.085999420	0.018029101
HW-error	0.000049506	0.002591606	0.006926344	0.003449854
HW- A	0.004333549	0.004015638	0.000239767	0.000307570
HW- B	0.084061184	0.084626391	0.109395638	0.100627916
HW- b	0.472769826	0.650482593	0.104155542	0.138550587
HW- σ	0.030988143	0.000444231	0.000408286	0.000519610
HW- A/b	0.009166298	0.006173321	0.002302015	0.002219915
CIR-error	0.000116668	0.002568589	0.007464623	0.004277590
CIR- A	0.005041583	0.003145545	0.0002742925	0.000608089
CIR- B	0.088362257	0.084695177	0.1086040555	0.097400314
CIR- b	0.665089137	0.491036196	0.1276709251	0.329482866
CIR- σ	0.021254297	0.036289101	0.0191874838	0.021601300
CIR- A/b	0.007580312	0.006405933	0.0021484340	0.001845586

Table 2: Optimal parameter values for the survival function in the mean-reverting models.

The residuals of the Vasiček model are very large for the four generations (10^{-1}), and low for HW and CIR models as compared with the OU and Feller non-mean reversion models above. We also notice that the volatility of the Vasiček model (Vasi- σ) is very low as compared with the other two models.

As expected, the starting point of the target strongly decreases with the generations; however, less intuitively, B seems more important for the younger generations, which could be explained by the fact that the target is much lower for the younger generations.

Figures 3, 4 and 5 present the calibration results graphically. The fit is very poor for the Vasiček model and very satisfying for the HW and CIR models, as we cannot distinguish the two survival functions.

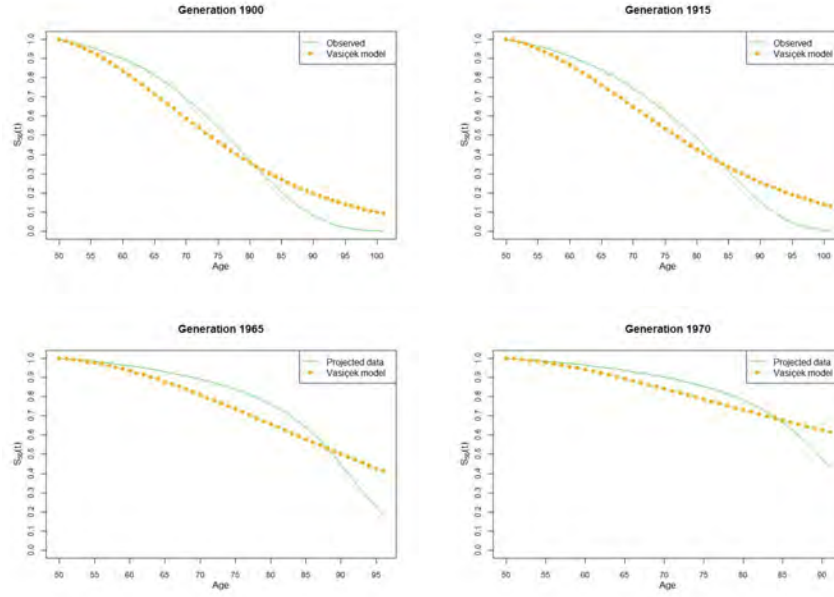


Figure 3: Survival functions, 1900,1915,1965 and 1970 generation for Vasiček model.

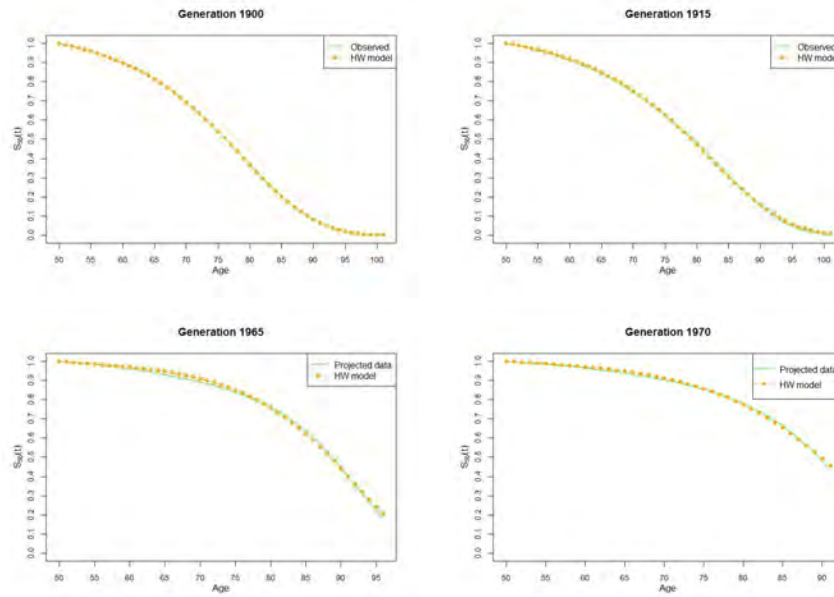


Figure 4: Survival functions, 1900,1915,1965 and 1970 generation for HW model.

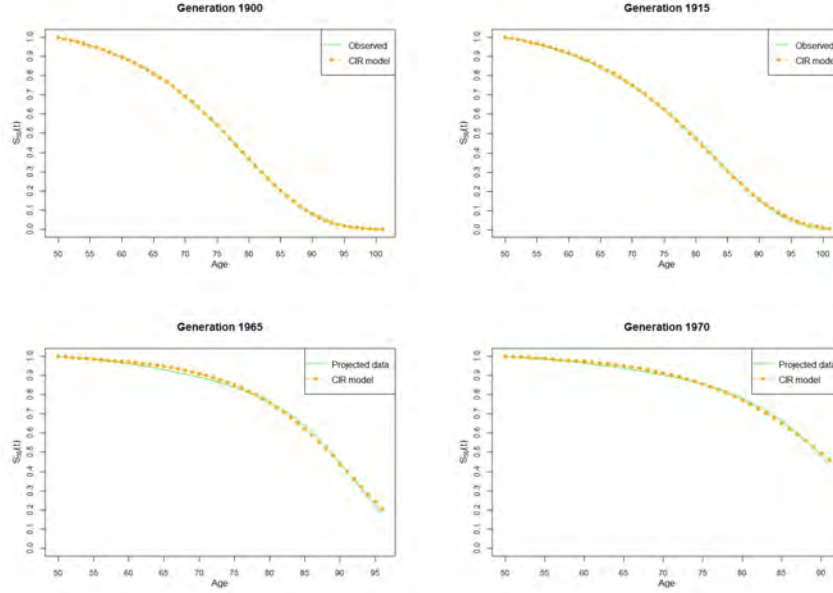


Figure 5: Survival functions, 1900,1915,1965 and 1970 generation for CIR model.

3.1.3 Calibration analysis

To conclude, we plot the calibration errors (i.e., the differences between the observed survival probabilities and their theoretical counterparts for all generations). Table 3 represents a summary of the errors for each model and each generation:

	1900	1915	1965	1970
OU	0.005885259	0.003403252	0.055996630	0.018659380
Feller	0.005885342	0.003403291	0.055996670	0.018659451
Vasiček	0.355228820	0.379264950	0.345576900	0.145937100
HW	0.000049506	0.002591606	0.006926344	0.003449854
CIR	0.000116668	0.002568589	0.007464623	0.004277590

Table 3: Calibration errors, all models.

The four graphs in figure 6 show that the Vasiček fitting error is very large, which confirms the conclusions of Luciano and Vigna [13] against mean reversion models; thus, the Vasiček model is not appropriate to describe mortality. In general, the OU and Feller models give similar acceptable errors. The HW and CIR models can fit the observed data with a smaller error. Therefore, we conclude that adding a variable long-term mean level following a Gompertz law leads to a significant improvement in the fit, and the mean reversion models

perform better than non-mean reversion models, with the condition of using a variable target.

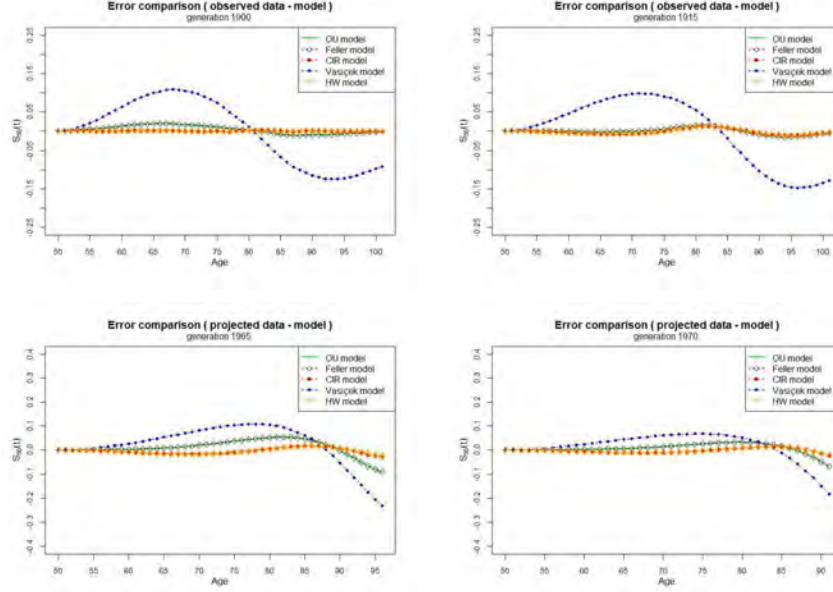


Figure 6: Calibration errors: Vasicek, OU, Feller, HW and CIR, different generations.

To compare the quality of each model relative to each of the other models, we follow the routine used by Apicella et al. [1]: we compute the Bayesian Information Criterion (BIC). This model selection method allows selecting the model that provides a good fit for the data without losing generality from over-fitting the data. This model selection method identifies the model with the best balance of goodness of fit between data and complexity. The BIC is given by:

$$BIC = n \ln\left(\frac{R}{n}\right) + k \ln(n), \quad (31)$$

where n is the number of observations, k is the number of parameters and R the residual found with the LSE method.

In Table 4 we report the BIC criteria for all models and different generations:

	1900	1915	1965	1970
BIC^{OU}	-444.5423	-471.9283	-293.3974	-342.8498
BIC^{Feller}	-444.5416	-471.9277	-293.3973	-342.8497
$BIC^{Vasiček}$	-235.6147	-232.3411	-207.6941	-246.4859
BIC^{HW}	-675.6233	-477.7269	-379.8322	-411.1972
BIC^{CIR}	-632.7620	-478.1730	-424.8321	-452.6713

Table 4: BIC values for all models and different generations.

The two non-mean reversion models, OU and Feller, provide almost identical BIC values for all generations, which is not surprising because both models have the same number of parameters and give close residuals. BIC values of the HW and CIR models are also close for the same reason. We clearly see the gap between the BIC of the two mean reversion models and non-mean reversion models. For all generations and in all the cases, the non-mean reversion models have the highest BIC values as compared with the HW and CIR models. The Vasiček model is the model with the highest BIC criteria and should be eliminated. Therefore, we keep the Hull and White and CIR models for reference, as well as the OU and Feller models for further testing.

3.2 Best estimate of future mortality

To look at the mortality trend, we compare the best estimate of $\mu_x(t)$ for different generations in order to illustrate the longevity pattern. We simulate the process $\mu_{50}(t)$ for the generations of 1900, 1915, 1965 and 1975, from age 50 until age 95, using the optimal parameters found in subsection 3.1. We use the Euler discretization for Feller and CIR processes, with a daily time step.

3.2.1 Non-mean reversion models

Figure 7 reports the mean of $\mu_{50}(t)$ for the four generations for the OU and Feller models:

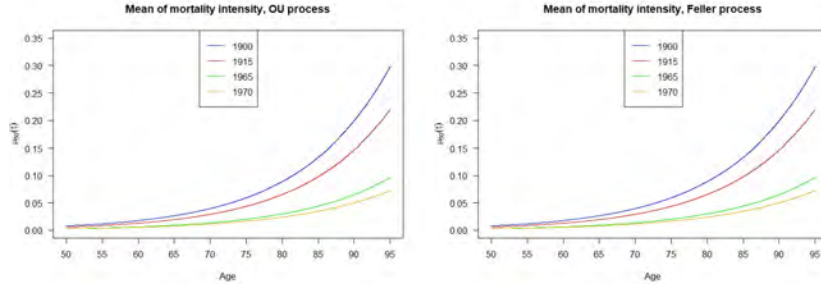


Figure 7: Mean of $\mu_{50}(t)$ for generation 1900, 1915, 1965 and 1970, OU and Feller models.

3.2.2 Mean reversion models

Figure 8 reports the mean of $\mu_{50}(t)$ for the four cohorts for the HW and CIR models:

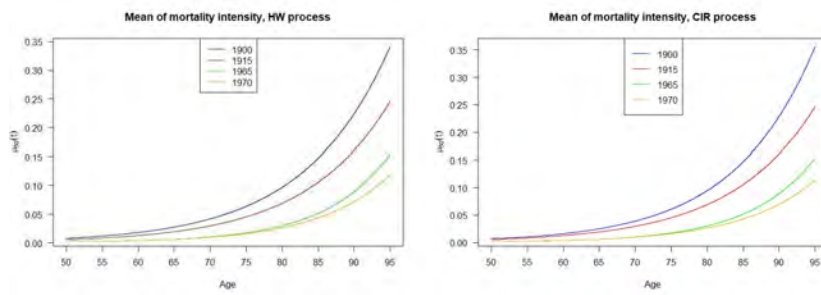


Figure 8: Mean of $\mu_{50}(t)$ for generation 1900, 1915, 1965 and 1970, HW and CIR models.

For all models, the older the generation, the higher the mean of $\mu_x(t)$; however we observe that for HW and CIR models, the mean of $\mu_x(t)$ appears to increase more rapidly with age.

3.3 Backtesting

In this section, we check if, after the calibration of our models on a portion of the historical survival curve, they can extend correctly. Hence, we perform backtesting. Using historical data of the generation 1900, we consider:

- $x_0 = 50$, the initial age of observation ;
- $N = 50$, the total amount of historical data;
- N_0 , the portion of historical data we use for calibration;
- $\delta = \frac{N_0}{N}$, the ratio of data used to total data.

The backtest consists on varying δ from 20% to 100% to see if the survival curves provided by the models meet reality.

3.3.1 Non-mean reversion models

Ornstein-Uhlenbeck model

We calibrate the models to the portion of observed survival function; then, we stock the optimal parameters and the global error, which refers to the difference between the curve projected by the model and the historical one. For the OU model, we obtain the following results:

N_0	Age	δ	global error	a	σ
10	50-60	20%	0.779266417	0.0579155861	0.000001304
20	50-70	40%	0.076667020	0.073191902	0.000000237
30	50-80	60%	0.009913895	0.078843759	0.000000006
40	50-90	80%	0.005931298	0.080496310	0.000001238
50	50-100	100%	0.005885259	0.080696570	0.000000129

Table 5: Backtesting OU model, generation 1900.

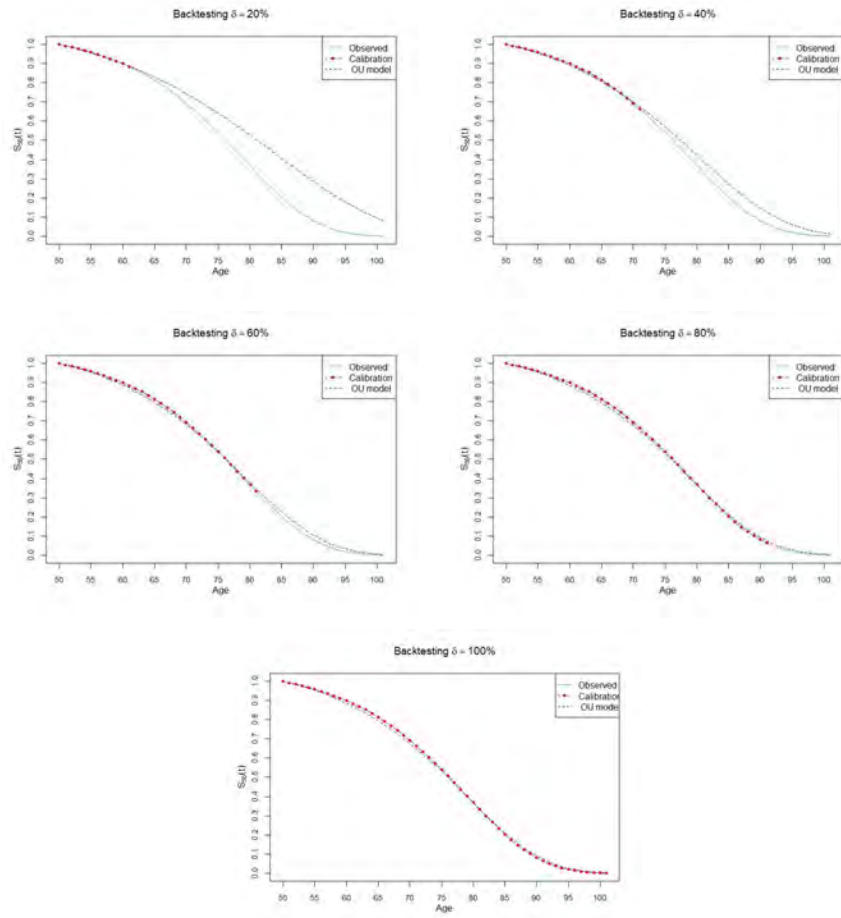


Figure 9: Evolution of the fitting quality of the OU model according to the portion of observed data considered.

Feller model

N_0	Age	δ	global error	a	σ
10	50-60	20%	0.779101200	0.057919067	0.000203210
20	50-70	40%	0.076726301	0.073188955	0.000038991
30	50-80	60%	0.009921382	0.078842051	0.000010750
40	50-90	80%	0.005931992	0.080494890	0.000024904
50	50-100	100%	0.005885342	0.080687850	0.000001527

Table 6: Backtesting Feller model, generation 1900.

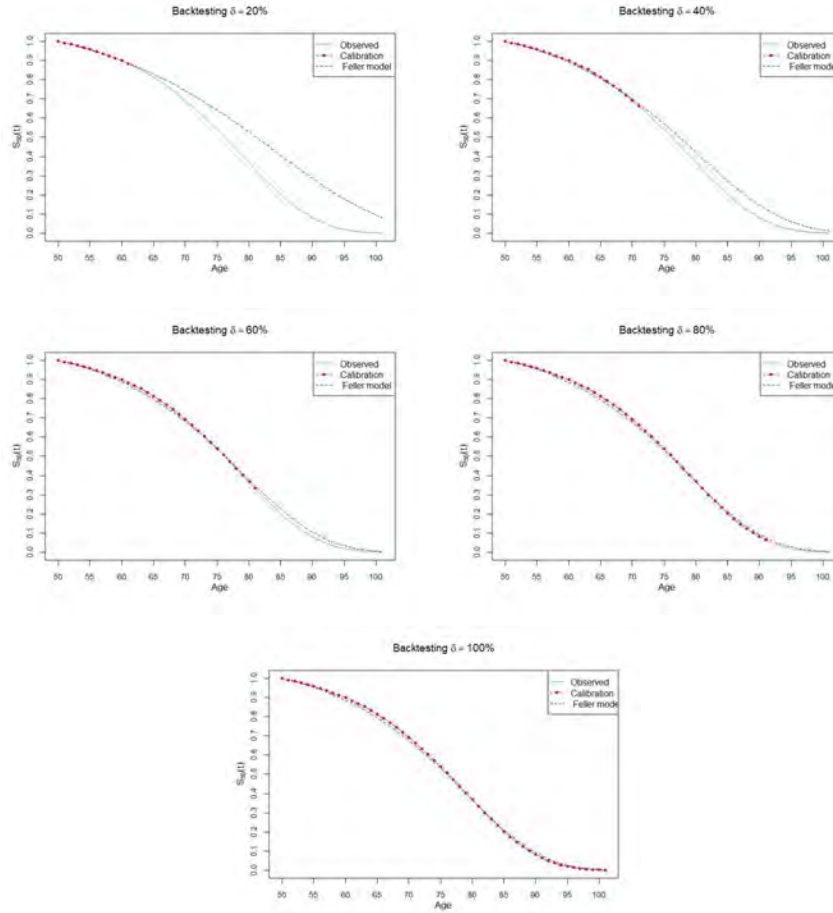


Figure 10: Evolution of the fitting quality of the Feller model according to the portion of observed data considered.

Both models OU and Feller, need an important volume ($\delta \geq 80\%$) of information to correctly complete the historical curve. Below this value, both models overestimate the survival function.

3.3.2 Mean reversion models

HW model

N_0	Age	δ	global error	A	B	b	σ
10	50-60	20%	0.003464492	0.002733086	0.088094733	0.334805279	0.001641703
20	50-70	40%	0.000594809	0.003718993	0.085592503	0.400761822	0.027073209
30	50-80	60%	0.000245888	0.006525992	0.081024406	0.690103707	0.053059593
40	50-90	80%	0.000051370	0.004631080	0.083927720	0.509788920	0.033306650
50	50-100	100%	0.000049506	0.004333549	0.084061184	0.472769826	0.030988143

Table 7: Backtesting HW model, generation 1900.

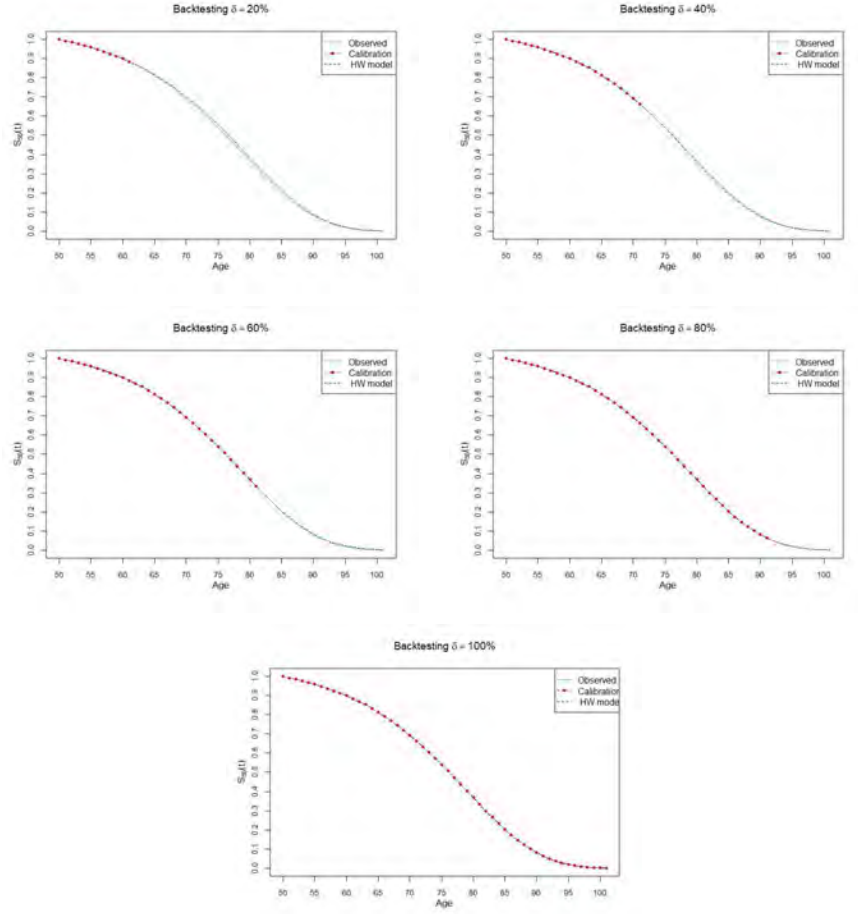


Figure 11: Evolution of the fitting quality of the HW model according to the portion of observed data considered.

CIR model

For the CIR model we obtain the following results:

N_0	Age	δ	global error	A	B	b	σ
10	50-60	20%	0.009092467	0.003205452	0.086116506	0.397389283	0.051463759
20	50-70	40%	0.003353861	0.003299728	0.092763104	0.431962054	0.023858361
30	50-80	60%	0.000321363	0.004015533	0.089673479	0.518644014	0.077440960
40	50-90	80%	0.000139469	0.003613061	0.088649272	0.450780512	0.068252594
50	50-100	100%	0.000116668	0.005041583	0.088362257	0.665089137	0.021254297

Table 8: Backtesting CIR model, generation 1900.

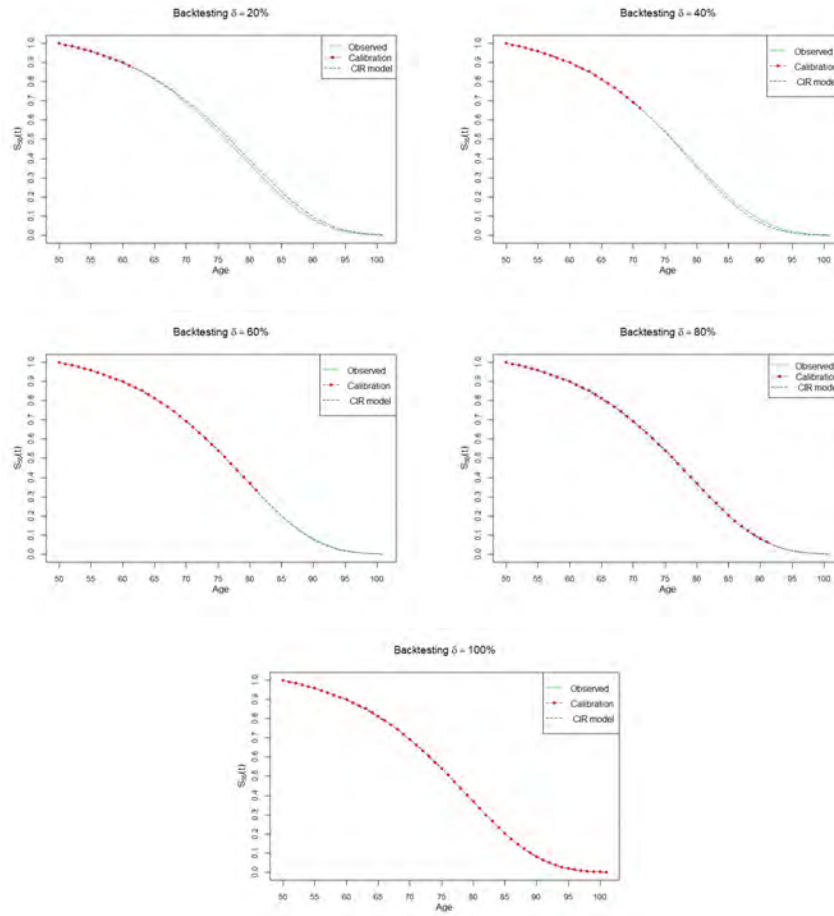


Figure 12: Evolution of the fitting quality of the CIR model according to the portion of observed data considered.

In the backtesting strategy, the HW and CIR models perform better than OU and Feller model; the fit is very satisfying when $\delta \geq 40\%$ for both models when compared with non-mean reversion models that require more data to fit the historical survival function correctly.

4 Multicohort vision

In this section, we plot the mortality surface to represent the survival functions of different cohorts simultaneously. To have enough values, we consider the projected data instead of calibrating the models to historical data.

4.1 Mortality surface

Let us define $S^f(x_0, 65, z)$, the survival forward probability between age 65 and age z for an individual aged x_0 at time 0 (2015 in our example), and assumed to be alive at age 65. We define this probability as the ratio between the two corresponding survival probabilities:

$$S^f(x_0, 65, z) = \frac{z-x_0 p_{x_0}}{65-x_0 p_{x_0}}. \quad (32)$$

This definition is similar to the concept of forward rates in the interest rate theory.

Taking into account (2), (3) and (4), this probability is explicitly given by:

$$S^f(x_0, 65, z) = S^f(x_0, z) \quad (33)$$

$$= \frac{E_{\mathbb{P}}(S(x_0, 0, z-x_0))}{E_{\mathbb{P}}(S(x_0, 0, 65-x_0))} \quad (34)$$

$$= \frac{e^{\alpha(0, z-x_0) - \beta(0, z-x_0)\mu_{x_0}(0)}}{e^{\alpha(0, 65-x_0) - \beta(0, 65-x_0)\mu_{x_0}(0)}}. \quad (35)$$

To plot the mortality surface, we follow these steps:

- We consider 10 cohorts of individuals initially aged from 55 to 65 years in 2015;
- We calibrate the models on projected data for the different cohorts (from the age x_0 until the end of the projected life table);
- We consider that all individuals survive until the age of 65;
- We use the models to project the survival functions for the 10 cohorts from the age of 65 until the age of 105;
- We plot the mortality surface in three dimensions.

4.1.1 Non-mean reversion models:

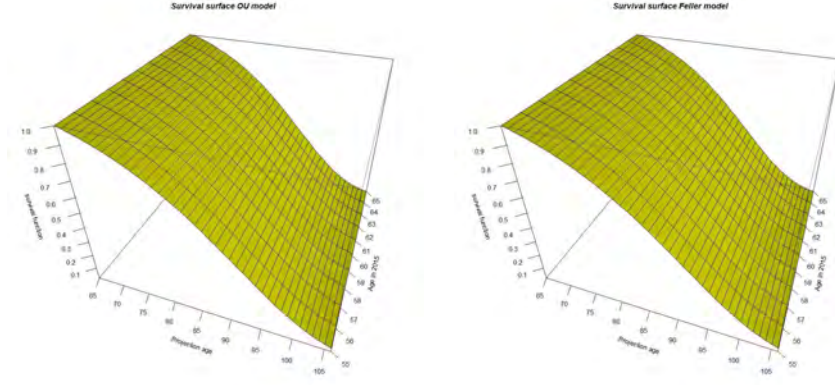


Figure 13: Forward survival surface for individuals aged from 55 to 65 in 2015, projected from the age 65 to 105, OU and Feller models.

Tables 9 and 10 present the annual survival probabilities given by the OU and Feller models, which are used to plot the three-dimensional figure above:

Projection age z	$S^f(55, z)$	$S^f(58, z)$	$S^f(61, z)$	$S^f(64, z)$
70	0.94270999	0.94086362	0.94000700	0.94073165
75	0.86324605	0.85867284	0.85584838	0.85594137
80	0.75692340	0.74867934	0.74240724	0.73964739
85	0.62208501	0.60956896	0.59843903	0.59017254
90	0.46416890	0.44786853	0.43160361	0.41629487
95	0.29981629	0.28209663	0.26297026	0.24269607
100	0.15614449	0.14104335	0.12407613	0.10538469
105	0.05896950	0.04987813	0.03973097	0.02901933

Table 9: Predicted forward survival probabilities with OU model, different cohorts.

Projection age z	$S^f(55, z)$	$S^f(58, z)$	$S^f(61, z)$	$S^f(64, z)$
70	0.94270779	0.94059967	0.93981972	0.94066064
75	0.86323974	0.85786186	0.85519102	0.85559036
80	0.75691057	0.74695390	0.74089639	0.73871205
85	0.62206344	0.60658881	0.59570571	0.58833943
90	0.46413823	0.44358153	0.42757920	0.41348398
95	0.29978022	0.27708191	0.25826448	0.23939465
100	0.15611136	0.13656292	0.12000066	0.10264470
105	0.05894795	0.04713311	0.03741897	0.02763406

Table 10: Predicted forward survival probabilities with Feller model, different cohorts.

4.1.2 Mean reversion models:

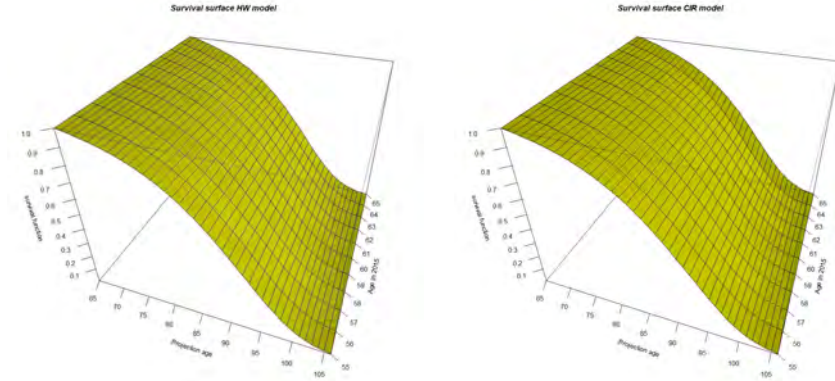


Figure 14: Forward survival surface for individuals aged from 55 to 65 in 2015, projected from age 65 to 105, HW and CIR models.

Tables 11 and 12 present the annual survival probabilities given by the HW and CIR models, which are used to plot the three-dimensional figure above:

Projection age z	$S^f(55, z)$	$S^f(58, z)$	$S^f(61, z)$	$S^f(64, z)$
70	0.962835443	0.960325946	0.957746038	0.950056248
75	0.904339900	0.899711357	0.897777534	0.884516037
80	0.811135375	0.803912194	0.801521846	0.784853820
85	0.669693860	0.659217546	0.652808569	0.635412171
90	0.477145956	0.463747219	0.451074084	0.435174359
95	0.261767768	0.248404861	0.233137192	0.220321818
100	0.090383387	0.081991900	0.072303009	0.064728824
105	0.013744286	0.011454980	0.009139222	0.007139386

Table 11: Predicted forward survival probabilities with HW model, different cohorts.

Projection age z	$S^f(55, z)$	$S^f(58, z)$	$S^f(61, z)$	$S^f(64, z)$
70	0.96289906	0.960181543	0.957353056	0.949143243
75	0.90423742	0.899746640	0.893606531	0.882913764
80	0.81075621	0.804358436	0.793837357	0.783807917
85	0.66912780	0.659968649	0.645760840	0.635573943
90	0.47668261	0.464236350	0.450115603	0.435762844
95	0.26168739	0.248033054	0.239427926	0.219823304
100	0.09058580	0.081125847	0.079370316	0.063408897
105	0.01386676	0.011058342	0.011505930	0.006615424

Table 12: Predicted forward survival probabilities with CIR model, different cohorts.

A comparison between Tables 9, 10 and Tables 11, 12 shows that HW and CIR models are once again more efficient: the forward survival probabilities provided by HW and CIR models, are more in line with the improvement of longevity for younger generations.

4.2 Parameter smoothing

Next, we examine the evolution trend of the HW and CIR models' parameters. We focus on smoothing B , $\frac{A}{b}$ and σ following these three steps:

1. we use projected data for calibration and consider individuals aged from 55 to 85 years old in 2015.
2. We draw the data points related to each parameter and each age;
3. We use smooth non-linear functions to approximate these data points.

We find the following results:

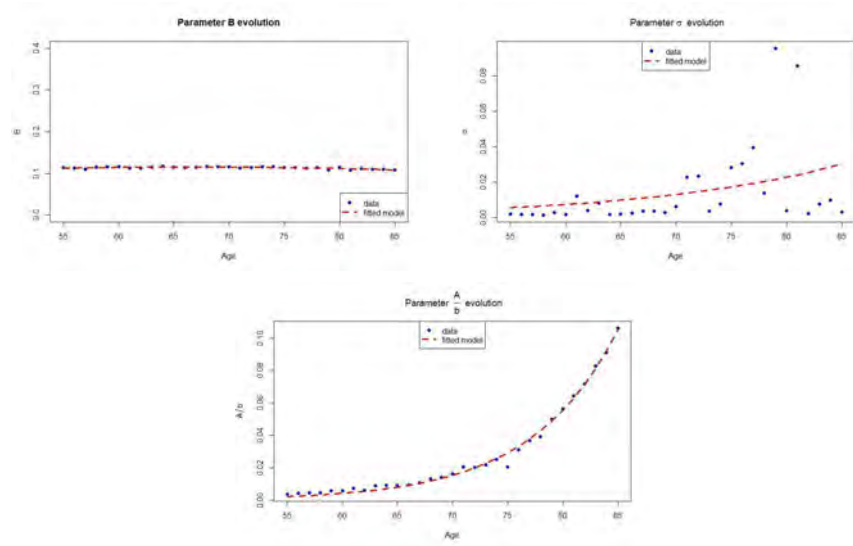


Figure 15: Non-linear regression applied to the different parameters, HW model.

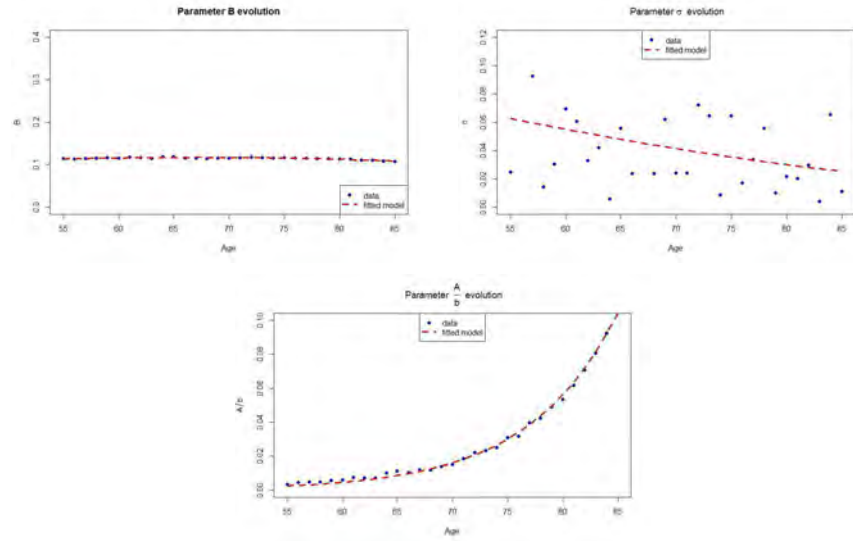


Figure 16: Non-linear regression applied to the different parameters, CIR model.

From these figures, we see that parameter B is very stable. The initial level of the target $\frac{A}{b}$ increases exponentially with age, which is also in line with the Gompertz assumption.

5 Conclusion

In this paper, we described the stochastic evolution of mortality using affine processes. We considered five kinds of stochastic processes that differ in their drift and stochastic part, and their survival probabilities were provided in closed form and studied analytically. The intensity processes were calibrated to the Belgian population, using historical data for old generations and projected mortality tables for younger generations. For each model, we performed different calibration tests and assessed their ability to forecast mortality. As a result, we showed that adding a time-dependent long-term mean to the selected processes seems to be more appropriate to describe the death intensity of individuals, satisfying most of the reasonable criteria for a good mortality model. Moreover, we proposed procedures for mortality forecasting and mortality trend assessment, which describe the future evolution of mortality within a single cohort and between different cohorts, respectively. Furthermore, we voluntarily confined ourselves to one-factor, non-jump basic models in order to focus on our research question concerning the eventual mean reversion mortality property. We conclude that while the mean reversion aspect is not suitable for mortality when using a fixed target, it is entirely appropriate when using a variable target linked to the increase in mortality with age. For further research, an interesting study would be to consider different functions for the floating target and assess the performance of the mean-reverting models.

References

- [1] Apicella G, Dacorogna M, Lorenzo E.D, Sibillo M (2019) Improving the forecast of longevity by combining models. *North American Actuarial Journal* 23:298-319.
- [2] Biffis E (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: mathematics and economics* 37:443-468.
- [3] Cairns AJ, Blake D, Dowd K (2006a) Pricing deaths: frameworks for the valuation and securitisation of mortality risk. *ASTIN Bulletin* 36:79–120.
- [4] Cairns AJ, Blake D, Dowd K (2006b) A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance* 73:687-718.
- [5] Cox JC, Ingersoll Jr JE, Ross SA (1985) An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society* 363-384.
- [6] Dahl M (2004) Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics* 35:113-136.

- [7] Denuit M, Devolder P (2006) Continuous time stochastic mortality and securitization of longevity risk. Technical report, Working Paper UCL 06-02.
- [8] Duffie D, Filipovic, D, Schachermayer W (2003) Affine processes and applications in finance. *Annals of applied probability* 984-1053.
- [9] Duffie D, Pan J, Singleton K (2000) Transform analysis and asset pricing for affine jump-diffusion. *Econometrica* 68:1343-1376.
- [10] Hull JC (2009) *Options, Futures, and Other Derivatives*. Pearson Prentice Hall.
- [11] Kac M (1949) On distributions of certain wiener functionals. *Transactions of the American Mathematical Society* 65:1-13.
- [12] Lee RD, Carter LR (1992) Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*.
- [13] Luciano E, Vigna E (2015) Non mean reverting affine processes for stochastic mortality ICER Working Papers.
- [14] Milevsky M, Promislow S (2001) Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*.
- [15] Mishura Y, Posashkova S (2008) Positivity of solution of nonhomogeneous stochastic differential equation with non-Lipschitz diffusion. *Theory of Stochastic Processes* 14.
- [16] Schrager D (2006) Affine stochastic mortality. *Insurance: Mathematics and Economics*.