Mortality Surface by Means of Continuous Time Cohort Models

(Article begins on next page)

26 April 2024

MORTALITY SURFACE BY MEANS OF CONTINUOUS TIME COHORT MODELS

PETAR JEVTIĆ, ELISA LUCIANO, AND ELENA VIGNA

ABSTRACT. We study and calibrate a cohort-based model which captures the characteristics of a mortality surface with a parsimonious, continuous-time factor approach. The model allows for imperfect correlation of mortality intensity across generations. It is implemented on UK data for the period 1900-2008. Calibration by means of stochastic search and the Differential Evolution optimization algorithm proves to yield robust and stable parameters. We provide in-sample and out-of-sample, deterministic as well as stochastic forecasts. Calibration confirms that correlation across generations is smaller than one.

JEL classification: C1, C13, C38, C53, J11.

Keywords: stochastic mortality, age effect, cohort effect, differential evolution algorithm, mortality forecasting.

1. INTRODUCTION

Insurance companies and pension funds are exposed to mortality risk and hope for the development of a liquid and transparent longevity-linked capital market. Active trading of mortality derivatives would help them assessing and hedging the risks they are exposed to, in the same manner as financial models and markets help them mutualize financial risks. Mortality-risk appraisal consisting in an accurate but easy-to-handle description of human survivorship is fundamental in this respect.

In spite of this need, no consensus has been reached yet on the best model for mortality risk modelling. A number of successful proposals have been put forward. Most of these models, starting from the celebrated Lee and Carter (1992) model and its several extensions – that include for instance Brouhns et al. (2002) and Renshaw and Haberman (2003), up to the more recent Cairns et al. (2006b) two-factors model – are discrete-time descriptions of survivorship evolution. In some cases though the adoption of a continuous–time approach proves useful. This is the case when we couple the appraisal of mortality and financial risk, and we adopt some financial model such as Black-Scholes or Duffie et al. (2000). Another motivation

for adopting a continuous-time description is the search for closed-form evaluation formulas for insurance products and their derivatives. Continuous-time stochastic mortality models for single generation were considered by a number of researchers, including Milevsky and Promislow (2001), Dahl (2004), Biffis (2005), Cairns et al. (2006a), Schrager (2006) and Luciano and Vigna (2008).

Be it discrete or continuous-time based, a proper description of mortality risk should capture several dimensions. Consider survival probabilities over a given horizon. A satisfactory model should capture their evolution when changing the horizon, for a fixed initial age and cohort (or generation), and its evolution over cohorts, for fixed initial age and horizon. Introducing the cohort dimension though adds a level of complexity to the problem, since it calls for a notion of correlation across generations, which is by no means easy to capture. In principle one has a whole "mortality surface" to model. How to do this while keeping a satisfactory trade-off between the accuracy - or the fit - and the tractability of the model is an open issue. A theoretical extension of the continuous-time single-generation model to the mortality surface appears in Biffis and Millossovich (2006). This is followed by Blackburn and Sherris (2012) who focus also on the calibration aspect. More specifically, in their attempts to calibrate the whole surface in continuous time they make the assumption of perfect correlation across generations. However, common intuition suggests that correlation among close generations is high but not perfect. This suggestion is often implemented in actuarial practice.

In order to reconcile the calibration of the whole mortality surface with common actuarial practice, this paper fits the mortality surface by means of a continuous-time cohort model, that is able to capture correlations across generations. As a relevant consequence, this model provides the actuary with a calibrated correlation among generations rather than a "best estimate" one. Given the same initial age, the intensities of several generations are written in terms of factors, identified via Principal Component Analysis (PCA). Differential Evolution algorithm is a robust stochastic search and optimization algorithm which already proved its use across a wide range of engineering applications. We use it to fit the mortality surface with an extreme precision. Provided that we fully exploit the power of this stochastic search algorithm, we discover that the fitted parameters are extremely robust, stable and lead to correlations across generations that is high but less than one.

The paper unfolds as follows. In Section 2, we review mortality modeling via affine mortality intensities for a single generation. Then, we develop ex-novo a model for the mortality intensities of several generations, i.e. we model the mortality surface. Section 3 specifies a simple two-factor model for modeling the mortality surface, which will then be calibrated. In Section 4, we discuss the criteria that a good mortality model for the mortality surface should satisfy. In Section 5, we proceed to the empirical part. We use PCA to identify the number of relevant factors and apply it to UK males data from the Human Mortality Database. We review the Evolutionary Approach to the global minimum/maximum search and use it in Section 6 to calibrate a two-factor model to a number of UK generations born between 1900 and 1950. We discuss all of the key criteria introduced in Section 4. In Section 7 we use polynomial interpolation to further improve parsimoniousness of the model. In Section 8, we conclude and outline further research.

2. The Mortality Model

In Section 2.1, we illustrate the stochastic mortality intensity setup for one generation only – as is standard in this kind of literature. In Section 2.2, we specify how to move on from the description of the mortality intensity of one generation to the mortality intensities of several generations. This procedure enables us to describe the whole mortality surface. In Section 2.3, we restrict ourselves to constant-parameter dynamics of the Ornstein-Uhlenbeck type. The general mortality model is described in Section 2.4, and a simplified version of it is presented in Section 2.5 .

2.1. The affine mortality framework for the single generation.
As in the standard unidimensional framework of stochastic mortality (see eg Biffis, 2005, Dahl, 2004) we describe the mortality of a given generation by means of a Cox or doubly stochastic counting process. Intuitively, the time of death is supposed to be the first jump time of a Poisson process with stochastic intensity.

Let us introduce a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq0}, \mathbb{P})$, where $\mathbb{P}$ is the real-world probability measure. The filtration $\{\mathcal{F}_t : 0 \leq t \leq T\}$ satisfies the usual properties of right-continuity and completeness. On this space, let us consider a predictable process $\mu(t, x)$, which represents the mortality intensity of an individual belonging to a given generation, initial age x at (calendar) time t. His death is the first stopping time of a doubly stochastic process with intensity $\mu(t, x)$.

We model the intensity $\mu(t, x)$ of the given generation and initial age x as a function $R(\mathbf{X})$ of a vector of state processes

$$\boldsymbol{X}(t) = [X_1(t), ..X_n(t)]^\top .$$

For notational simplicity, in the rest of this section we will omit the argument x. Therefore, we have that

$$(1) \qquad \mu(t) = R(\mathbf{X}(t)).$$

Moreover, in order to keep the model mathematically tractable, we put ourselves in the affine framework of Duffie et al. (2000) (sometimes referred to as DPS). In this setting $\boldsymbol{X}$ is a Markov process in some state space $D \in \mathbb{R}^n$ and it is the solution to the stochastic differential equation

$$d\mathbf{X}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{Z}(t),$$

where $\boldsymbol{Z}$ is an $(\mathcal{F}_t)$-standard Brownian motion in $\mathbb{R}^n$, $\boldsymbol{\lambda} : D \to \mathbb{R}^n$, $\boldsymbol{\sigma} : D \to \mathbb{R}^{n\times n}$, $\boldsymbol{\lambda}$, $\boldsymbol{\sigma}$, and $R : D \to \mathbb{R}$ are affine:

- $\boldsymbol{\lambda}(x) = \mathbf{K}_0 + \boldsymbol{K}_1 x$, for $\mathbf{K} = (\mathbf{K_0}, \mathbf{K_1}) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}$,

- $(\boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)^\top)_{ij} = (\boldsymbol{H}_0)_{ij} + (\boldsymbol{H}_1)_{ij} \cdot x$, for $\boldsymbol{H} = (\boldsymbol{H}_0, \boldsymbol{H}_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n}$,

- $R(x) = r_0 + \boldsymbol{r}_1 x$, where $(r_0, \boldsymbol{r}_1) \in \mathbb{R} \times \mathbb{R}^n$.

The advantage of this affine choice is that it is possible to calculate in closed form the expectation of functionals of the state variables. In fact, we have

$$(2) \qquad \mathbb{E}[e^{-\int_t^T R(\mathbf{X}(s))ds} \mid \mathcal{F}_t] = e^{\alpha(t;T)+\beta(t;T)\cdot\mathbf{X}(t)},$$

where the coefficients $\alpha(\cdot; T), \beta(\cdot; T) : \mathbb{R}^+ \to \mathbb{R}^n$ satisfy the complex-valued ODEs

$$\beta'(t;T) = \mathbf{r}_1 - \boldsymbol{K}_1^\top \beta(t;T) - \frac{1}{2}\beta(t;T)^\top \boldsymbol{H}_1 \beta(t;T),$$

$$\alpha'(t;T) = r_0 - \boldsymbol{K}_0 \beta(t;T) - \frac{1}{2}\beta(t;T)^\top \boldsymbol{H}_0 \beta(t;T),$$

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

In the actuarial context, if the intensity is given by (1), the expectation (2) is the survival probability from t to T, conditional on being alive at t:

$$(3) \qquad S(t,T) = \mathbb{E}_t\big[e^{-\int_t^T R(X(s))ds}\big] = \mathbb{E}_t\big[e^{-\int_t^T \mu(s)ds}\big].$$

2.2. Transition from single generations to the whole mortality surface.
In the previous section we have described the mortality intensity of one given generation. However, our main aim is to describe the whole mortality surface, that is composed by a number of different generations. We need then to label each generation with a proper index $i \in \mathbf{I} \subset \mathbb{N}$. Each generation has its own mortality intensity and the intensities of different generations are correlated. This implies that the straight replication of the DPS framework n times for n cohorts cannot be considered satisfactory, because it would fail to catch correlations among the intensities (unless correlation is perfect, that is typically not the case). Then, in a very natural way, we assume that the state processes which affect each generation are driven by Brownian motions that are correlated. Each generation is assigned its own instantaneous correlation matrix. As a relevant byproduct of this assumption, one obtains correlations among the generations. In order to comply with the DPS framework, we need uncorrelated Brownian motions. Hence, we need to go trough the Cholesky decomposition. This can be done as follows. Assume that the n processes driving the mortality intensity of generation i follow the process

$$(4) \qquad d\mathbf{X}^i(t) = \boldsymbol{\lambda}(\mathbf{X}^i(t))dt + \boldsymbol{\sigma}(\mathbf{X}^i(t))d\mathbf{W}^i(t),$$

where $W_1^i(t), W_2^i(t), ... W_n^i(t)$ are correlated with instantaneous correlation matrix $\boldsymbol{\rho}_{n\times n}^i = \{\rho_{lm}^i\}_{1\le l,m\le n}$, that is proper of generation i, where

$$(5) \qquad \rho_{lm}^i dt = \langle dW_l^i(t), dW_m^i(t)\rangle.$$

It is well known that – keeping the same filtration – the vector $\mathbf{W}^i(t)$ of correlated Brownian motions can be transformed into a vector of uncorrelated ones, $\mathbf{Z}(t)$, as follows (see for instance Shreve (2004), Section 4.6):

$$(6) \qquad d\mathbf{W}^i(t) = \mathbf{H}^i d\mathbf{Z}(t),$$

where

$$\mathbf{Z(t)} = [Z_1(t), Z_2(t), ..., Z_n(t)]^\top,$$

and $\mathbf{H}^i$ is obtained through Cholesky decomposition of the correlation matrix $\boldsymbol{\rho}^i$:

$$(7) \qquad \boldsymbol{\rho}^i = \mathbf{H}^i(\mathbf{H}^i)^\top.$$

If we assume that the n independent sources of risk are the same for each and every generation, the state process becomes:

$$(8) \qquad d\mathbf{X}^i(t) = \boldsymbol{\lambda}(\mathbf{X}^i(t))dt + \boldsymbol{\sigma}(\mathbf{X}^i(t))\mathbf{H}^i d\mathbf{Z}(t),$$

and the DPS framework can be applied.

This procedure can be followed for any choice of affine processes, for instance in the case of Ornstein-Uhlenbeck (OU) processes. In this case, the knowledge of the correlation matrix is all what one needs for the solution of the system of ODEs. OU processes have been used in related fields – such as interest rate modeling – for their analytical tractability. In the mortality field they are not new: they have been used for describing both deterministic and stochastic mortality. Indeed, as for deterministic mortality, the Gompertz model can be seen as an Ornstein Uhlenbeck process with null diffusion coefficient; as for stochastic mortality, when only one generation is considered, Luciano and Vigna (2008) find that the Ornstein Uhlenbeck process proves to be appropriate to describe the human mortality intensity. Motivated by analytical tractability and this evidence, in the next section we will specify the form of the state processes, and select the Ornstein-Uhlenbeck type.

2.3. **Ornstein-Uhlenbeck state processes.** We assume that in the description of the mortality intensity of generation i each state process follows an Ornstein-Uhlenbeck dynamics. Therefore, we have an n dimensional stochastic process, $\boldsymbol{X}(t)$ which has the following specification

(9)
$$
\begin{aligned}
dX_1^i(t) &= \psi_1 X_1^i dt + \sigma_1 dW_1^i(t) \\
dX_2^i(t) &= \psi_2 X_2^i dt + \sigma_2 dW_2^i(t) \\
&\cdots \\
dX_n^i(t) &= \psi_n X_n^i dt + \sigma_n dW_n^i(t),
\end{aligned}
$$

where $\psi_k \in \mathbb{R}$ and $\sigma_k > 0$ for $k = 1, 2, ..., n$ and $W_1^i(t), W_2^i(t), ...W_n^i(t)$ are correlated. The instantaneous correlation matrix for generation i, $\boldsymbol{\rho}_{n \times n}^i = \{\rho_{lm}^i\}_{1 \leq l,m \leq n}$, is given by (5). In concise form, we have

$$
d\boldsymbol{X}^i(t) = \boldsymbol{\Psi} \boldsymbol{X}^i(t) dt + \boldsymbol{\Sigma} d\mathbf{W}^i(t),
$$

where

$$
\boldsymbol{\Psi} = \mathrm{diag}[\psi_1, \psi_2, ..., \psi_n] \qquad \boldsymbol{\Sigma} = \mathrm{diag}[\sigma_1, \sigma_2, ..., \sigma_n].
$$

Following the above procedure, the state process can be transformed as follows:

(10)
$$
d\boldsymbol{X}^i(t) = \boldsymbol{\Psi} \boldsymbol{X}^i(t) dt + \boldsymbol{\Sigma} \mathbf{H}^i d\mathbf{Z}(t),
$$

where $\mathbf{H}^i$ is as in (7), and we can apply the DPS framework. We still have to specify the function $R(\cdot)$. We take the simplest affine function, namely the sum of the state variables:

(11)
$$
R(\mathbf{X}^i(t)) = \mathbf{1} \cdot \mathbf{X}^i(t).
$$

This is the DPS framework with

$$
\mathbf{K_0} = 0, \ \mathbf{K}_1 = \mathbf{K}_1^\top = \boldsymbol{\Psi}, \ r_0 = 0, \ \boldsymbol{r}_1 = \mathbf{1},
$$
$$
\mathbf{H}_0^i = \boldsymbol{\Sigma} \mathbf{H}^i (\boldsymbol{\Sigma} \mathbf{H}^i)^\top = \boldsymbol{\Sigma} \mathbf{H}^i (\mathbf{H}^i)^\top \boldsymbol{\Sigma}^\top = \boldsymbol{\Sigma} \boldsymbol{\rho}^i \boldsymbol{\Sigma}^\top, \ \mathbf{H_1} = \mathbf{0},
$$

while $\alpha^i(t; T)$ and $\beta(t; T)$ satisfy the system[1]

(12)
$$
\beta'(t; T) = \mathbf{1} - \boldsymbol{\Psi} \beta(t; T),
$$
$$
(\alpha^i)'(t; T) = -\frac{1}{2} \beta(t; T)^\top \boldsymbol{\Sigma} \boldsymbol{\rho}^i \boldsymbol{\Sigma}^\top \beta(t; T).
$$

---

[1]Notice from the system that the function $\alpha(\cdot; T)$ depends on the generation i via the correlation matrix $\boldsymbol{\rho}^i$, while the function $\beta(\cdot; T)$ does not.

Before solving it, notice that in the survivorship context it is convenient to set the valuation time $t = 0$ and to express time in terms of remaining lifetime $\tau = T - t$. It is quite natural to be interested in computing the survival probability over the (variable) horizon or duration τ. To this end, it is convenient to write the survival probability in terms of a new couple of functions, $\hat{\alpha}^i(\cdot)$ and $\hat{\beta}(\cdot)$:

$$(13) \qquad S^i(0, \tau) = \mathbb{E}\left[\exp\left(-\int_0^\tau \mathbf{1} \cdot \mathbf{X}^i(s)ds\right)\right] = e^{\hat{\alpha}^i(\tau) + \hat{\beta}(\tau) \cdot \mathbf{X}^i(0)},$$

where $\hat{\alpha}^i(\cdot)$ and $\hat{\beta}(\cdot)$ solve[2]

$$\hat{\beta}'(\tau) = -\mathbf{1} + \mathbf{\Psi}\hat{\beta}(\tau),$$

$$(\hat{\alpha}^i)'(\tau) = \frac{1}{2}\hat{\beta}(\tau)^\top \mathbf{\Sigma}\boldsymbol{\rho}^i\mathbf{\Sigma}^\top\hat{\beta}(\tau),$$

with boundary conditions

$$\hat{\alpha}^i(0) = 0 \text{ and } \hat{\beta}(0) = 0.$$

The solution for the new functions is

$$(14) \qquad \hat{\beta}(\tau) = -\int_0^\tau e^{\mathbf{\Psi}(\tau - s)} \cdot \mathbf{1}ds$$

$$(15) \qquad \hat{\alpha}^i(\tau) = \int_0^\tau \frac{1}{2}\hat{\beta}(s)^\top \mathbf{\Sigma}\boldsymbol{\rho}^i\mathbf{\Sigma}^\top\hat{\beta}(s)ds.$$

2.4. **General OU APC model.** Let us consider a life belonging to generation i, aged x at initial time 0. Recall that in the previous three sections we dropped the argument x for notational convenience. However, remind that all of the coefficients $\psi_1, ..., \psi_n, \sigma_1, ..., \sigma_n$ depend also on x. As a consequence, the functions $\hat{\alpha}^i_x(\cdot)$ and $\hat{\beta}_x(\cdot)$ depend on x. The survival curve is given by

$$(16) \qquad S^i_x(0, \tau) = e^{\hat{\alpha}^i_x(\tau) + \hat{\beta}_x(\tau) \cdot \mathbf{X}^i_x(0)}.$$

which gives us a continuous-time Age-Period-Cohort (APC) model for the mortality surface. In (16), we can distinguish the cohort effect in $\hat{\alpha}^i(\cdot)$ and $\mathbf{X}^i(0)$, the period effect in $\hat{\alpha}_{\cdot}(\tau)$ and $\hat{\beta}_{\cdot}(\tau)$, and the initial age effect in $\hat{\alpha}_x(\cdot)$, $\hat{\beta}_x(\cdot)$, and $\mathbf{X}_x(0)$.

The period or time effect is both deterministic and stochastic. Time enters in a deterministic way via the functions $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$, and is also present as a result of the random evolution of the process $\mathbf{Z}(\cdot)$ appearing in (10). In fact, if we put ourselves at time 0 when the head of generation i is aged x, the survival curve at time $t > 0$ for the same head aged $x + t$ is the random variable $S^i_{x+t}(t, T)$. Due to (2) and (3), this survival curve is given by

$$(17) \qquad S^i_{x+t}(t, T) = e^{\alpha^i_{x+t}(t) + \beta_{x+t}(t) \cdot \mathbf{X}^i_{x+t}(t)}.$$

---

[2]Indeed, we have $\hat{\alpha}^i(\tau) = \alpha(t; T)$, $\hat{\beta}(\tau) = \beta(t; T)$. Recalling that $\tau = T - t$, we have

$$(\hat{\alpha}^i)'(\tau) = \frac{d\hat{\alpha}^i(\tau)}{d\tau} = \frac{d\alpha^i(t; T)}{dt} \cdot \frac{dt}{d\tau} = -(\alpha^i)'(t; T), \text{ and}$$

$$\hat{\beta}'(\tau) = \frac{d\hat{\beta}(\tau)}{d\tau} = \frac{d\beta(t; T)}{dt} \cdot \frac{dt}{d\tau} = -\beta'(t; T).$$

Applying Ito's lemma to (17), and considering (10), we can express the dynamics of $dS^i_{x+t}(t, T)$ as a function of the independent Brownian motions $\mathbf{Z}(t)$. As a consequence, changes in $\mathbf{Z}(\cdot)$ affect directly the evolution of the initial value $\mathbf{X}^i_{x+t}(t)$. Due to (10), the evolution of the n factors $\mathbf{Z}(\cdot)$ over time affects each generation i in a different way, depending on the specific $\mathbf{H}^i$, that in turn depends on the specific, generation-based $\boldsymbol{\rho}^i$. In other words, each generation i reacts in its proper way to the period changes in the main factors driving the mortality intensity of individuals. We find this a desirable and intuitive feature.

Another consequence of (17) is that the evolution of the initial value $\mathbf{X}^i_{x+t}(t)$ – expression of the change over time of $\mathbf{Z}(\cdot)$ – has a different effect on the different initial ages, via the coefficient $\beta_{x+t}(\cdot)$. This is quite similar to the $\beta^{(2)}_x$ coefficient of the Lee-Carter model:

$$\log m(t, x) = \beta^{(1)}_x + \beta^{(2)}_x \kappa_t.$$

Differently from the existing two- or three-factors APC models in discrete-time, in this model the cohort effect is not added to the stochastic risk factors $Z(\cdot)$, but it multiplies them, via the cohort-based matrix $\mathbf{H}^i$.

2.5. **Simple OU APC model.** The general model described so far, though very rich, has the drawback of not being parsimonious. Since parsimony is a desirable property for a mortality model, we will consider and implement a simpler version of it, where all the coefficients are not age-dependent. In this simpler model the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$, as well as the sequence of the correlation matrices $\boldsymbol{\rho}^i$, $i \in I$ are constant with respect to x.

Let us consider a life belonging to generation i, aged x at initial time 0. According to this simpler model, her survival curve is given by

$$(18) \qquad S^i_x(0, \tau) = e^{\hat{\alpha}^i(\tau) + \hat{\beta}(\tau) \cdot \mathbf{X}^i_x(0)}.$$

As in the more general model, we can distinguish the cohort effect in $\hat{\alpha}^i(\cdot)$ and in $\mathbf{X}^i(0)$, the period effect in $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ and the initial age effect in $\mathbf{X}_x(0)$. The time effect can be seen both in a deterministic way, via $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$, and as a result of the stochastic evolution of the process $\mathbf{Z}(\cdot)$ appearing in (10). As before, due to (10), the evolution of the n factors $\mathbf{Z}(\cdot)$ over time affects each generation i in different way, depending on the specific $\mathbf{H}^i$, that in turn depends on the specific, generation-based $\boldsymbol{\rho}^i$. Again, each generation i reacts in its proper way to the period changes in the main factors driving the mortality intensity of individuals.

The similarity with the Lee-Carter model via the effect of the initial age on changes in $\mathbf{X}^i_{x+t}(t)$ is no longer valid.

3. Simple Two-Factor Model of Survival Probability

In most applications - including the one we are going to present below - the number of factors required to get a meaningful representation of the mortality rate for each single generation is limited to two. Moreover, in Section 5 the application of Principal Component Analysis to our data set supports the use of two-factors

model as a good starting point. For these reasons, we investigate a two-factor OU model. For generation i, we have

$$
\begin{aligned}
dX_1^i(t) &= \psi_1 X_1^i dt + \sigma_1 dW_1^i(t) \\
dX_2^i(t) &= \psi_2 X_2^i dt + \sigma_2 dW_2^i(t),
\end{aligned}
\tag{19}
$$

with

$$
\left\langle dW_1^i(t), dW_2^i(t) \right\rangle = \rho^i dt.
$$

Using (10), the process (19) is transformed into

$$
\begin{aligned}
dX_1(t) &= \psi_1 X_1 dt + \sigma_1 dZ_1(t) \\
dX_2^i(t) &= \psi_2 X_2 dt + \sigma_2 \rho^i dZ_1(t) + \sigma_2 \sqrt{1 - (\rho^i)^2} dZ_2(t),
\end{aligned}
\tag{20}
$$

with $dZ_1(t) \perp dZ_2(t)$. Notice that in this two-factor model the first process $X_1(\cdot)$ is common to all the generations, and the second one $X_2(\cdot)$ affects each generation with different weights. The mortality intensity of generation i is given by

$$
\mu^i(t) = X_1(t) + X_2^i(t).
\tag{21}
$$

The functions $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$ which appear in the survival probability (13) are

$$
\hat{\beta}(\tau) = -\int_0^\tau e^{\boldsymbol{\Psi}(\tau-s)} \cdot \mathbf{1} ds,
\tag{22}
$$

$$
\hat{\alpha}^i(\tau) = \int_0^\tau \frac{1}{2} \hat{\beta}(s)^\top \boldsymbol{\Sigma} \boldsymbol{\rho}^i \boldsymbol{\Sigma}^\top \hat{\beta}(s) ds,
\tag{23}
$$

with

$$
e^{\boldsymbol{\Psi}(\tau-s)} = \begin{bmatrix} e^{\psi_1(\tau-s)} & 0 \\ 0 & e^{\psi_2(\tau-s)} \end{bmatrix},
\tag{24}
$$

and

$$
\boldsymbol{\Sigma} \boldsymbol{\rho}^i \boldsymbol{\Sigma}^\top = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \times \begin{bmatrix} 1 & \rho^i \\ \rho^i & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}^\top = \begin{bmatrix} \sigma_1^2 & \rho^i \sigma_1 \sigma_2 \\ \rho^i \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}.
\tag{25}
$$

Then, the single components of the $\hat{\beta}(\cdot)$ function are

$$
\hat{\beta}_1(\tau) = -\int_0^\tau e^{\psi_1(\tau-s)} ds = \frac{1}{\psi_1}\left(1 - e^{\psi_1 \tau}\right),
$$

$$
\hat{\beta}_2(\tau) = -\int_0^\tau e^{\psi_2(\tau-s)} ds = \frac{1}{\psi_2}\left(1 - e^{\psi_2 \tau}\right).
$$

The function $\hat{\alpha}^i(\cdot)$ turns out to be

$$
\hat{\alpha}^i(\tau) = \sum_{j=1}^2 \frac{\sigma_j^2}{2\psi_j^3}\left(\psi_j \tau - 2e^{\psi_j \tau} + \frac{1}{2}e^{2\psi_j \tau} + \frac{3}{2}\right)
\tag{26}
$$

$$
+ \frac{\rho^i \sigma_1 \sigma_2}{\psi_1 \psi_2}\left(\tau - \frac{e^{\psi_1 \tau}}{\psi_1} - \frac{e^{\psi_2 \tau}}{\psi_2} + \frac{e^{(\psi_1+\psi_2)\tau}}{\psi_1+\psi_2} + \frac{\psi_1^2 + \psi_1 \psi_2 + \psi_2^2}{\psi_1 \psi_2 (\psi_1 + \psi_2)}\right).
$$

The model so specified for the intensity of one generation is known in the interest-rate domain as the two-factor Gaussian model, or G2++, which can be found in Brigo and Mercurio (2006). It is indeed a model in which each single intensity

is Gaussian. This can be easily seen not only from the differential representation (20)-(21), but also from the integral one, i.e.

$$(27) \quad X_1(t) = e^{\psi_1 t} X_1(0) + \sigma_1 e^{\psi_1 t} \sqrt{\frac{1}{2\psi_1}(1 - e^{-2\psi_1 t})} \, Z_1,$$

$$X_2(t) = e^{\psi_2 t} X_2(0) + \sigma_2 \rho^i e^{\psi_2 t} \sqrt{\frac{1}{2\psi_2}(1 - e^{-2\psi_2 t})} \, Z_1 +$$

$$+ \sigma_2 \sqrt{1 - (\rho^i)^2} e^{\psi_2 t} \sqrt{\frac{1}{2\psi_2}(1 - e^{-2\psi_2 t})} \, Z_2,$$

where Z_1 and Z_2 are two independent standard normal variables.

As a consequence, each intensity may become negative with positive probability:

$$(28) \quad Pr(\mu^i(\tau) < 0) = \Phi\left(-\frac{E(\mu^i(\tau))}{\sqrt{Var(\mu^i(\tau))}}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal, the expectation and the variance of $\mu^i(\tau)$ are

$$(29) \quad E(\mu^i(\tau)) = f^i(0,\tau) + \sum_{j=1}^{2} \frac{\sigma_j^2}{2\psi_j^2}\left(1 - e^{\psi_j \tau}\right)^2 + \rho^i \prod_{j=1}^{2} \frac{\sigma_j}{\psi_j}\left(1 - e^{\psi_j \tau}\right),$$

$$(30) \quad Var(\mu^i(\tau)) = -\sum_{j=1}^{2} \frac{\sigma_j^2}{2\psi_j}\left(1 - e^{\psi_j \tau}\right) - 2\rho^i \frac{\sigma_1 \sigma_2}{\psi_1 + \psi_2}\left(1 - e^{(\psi_1 + \psi_2)\tau}\right),$$

and $f^i(0,\tau)$ is the forward mortality intensity for the instant τ:

$$(31) \quad f^i(0,\tau) = -\frac{\partial \log S^i(0,\tau)}{\partial \tau} = \sum_{j=1}^{2}\left(e^{\psi_j \tau} X_j(0) - \frac{\sigma_j^2}{2\psi_j^2}\left(1 - e^{\psi_j \tau}\right)^2\right) -$$

$$- \rho^i \prod_{j=1}^{2} \frac{\sigma_j}{\psi_j}\left(1 - e^{\psi_j \tau}\right).$$

Given the availability of the closed-form expressions (29), (30) and (31), in concrete situations and practical applications it is possible to keep the probability (28) under any desired threshold, by imposing suitable restrictions on the calibrated parameters.[3] In Section 5, we implement those restrictions to keep the probability of negative intensity lower than 1%.

3.1. **Correlations.** On top of analytical tractability, a good feature of this model is that it provides formulas for instantaneous correlation among the intensities of different generations. According to (21), for any couple of generations i and j, the instantaneous mortality intensities follow the SDEs

$$d\mu^i(t) = [\psi_1 X_1(t) + \psi_2 X_2^i(t)]dt + (\sigma_1 + \rho^i \sigma_2)dZ_1(t) + \sigma_2 \sqrt{1 - (\rho^i)^2}dZ_2(t),$$

$$d\mu^j(t) = [\psi_1 X_1(t) + \psi_2 X_2^j(t)]dt + (\sigma_1 + \rho^j \sigma_2)dZ_1(t) + \sigma_2 \sqrt{1 - (\rho^j)^2}dZ_2(t).$$

---

[3]In most of the applications to the interest-rate domain the probability turns out to be low without imposing the constraints. This – together with the other benefits of the G2++ models, as listed in Brigo and Mercurio (2006) – justifies the use of this model even if it is Gaussian.

It can be shown that the instantaneous correlation between $\mu^i(\cdot)$ and $\mu^j(\cdot)$ is

$$(32) \quad \text{Corr}[d\mu(t_i), d\mu(t_j)] = \frac{(\sigma_1 + \rho^i \sigma_2)(\sigma_1 + \rho^j \sigma_2) + \sigma_2^2 \sqrt{(1 - (\rho^i)^2)(1 - (\rho^j)^2)}}{\prod_{k \in \{i,j\}} \sqrt{(\sigma_1 + \sigma_2 \rho^k)^2 + \sigma_2^2 (1 - (\rho^k)^2)}}.$$

3.2. **Stochastic forecasting.** Beside the possibility of calculating the correlation among different generations, this model enables us to make stochastic forecasting of the survival probability curve at an arbitrary future time $p > 0$. Indeed, the survival curve at time $p > 0$ for a head belonging to generation i is the random variable $S^i(p, \tau)$. In the simple two-factors model this survival curve is given by

$$(33) \qquad\qquad S^i(p, \tau) = e^{\alpha^i(\tau - p) + \beta(\tau - p) \cdot \mathbf{X}^i(p)}.$$

The random vector $\mathbf{X}^i(p) = (X_1^i(p), X_2^i(p))$ can be simulated with low computational effort using its integral form in (27).

4. Implementation Criteria

There is quite a large consensus on the criteria that a satisfactory mortality model should fulfill. Here we list and discuss them within the simple OU two-factor model. The criteria are:

(1) analytical tractability

(2) parsimoniousness

(3) fit to historical data

(4) null or low probability of negative intensities

(5) possibility and ability of deterministic forecasting

(6) possibility and ability of stochastic forecasting

(7) possibility and ability of measuring correlation among different generations

Notice that while the first six criteria are proper of any mortality model, the last one refers only to models describing the mortality surface.

Some of these criteria refer to the theoretical construction and properties of a mortality model, while other strongly depend on the results of its calibration. Analytical tractability, parsimoniousness and null probability of negative intensity are proper of a theoretical model, as is the possibility of forecasting and measuring correlation among generations. On the other hand, having a good fit to historical data and the ability of keeping the probability of negative intensity low, as well as the ability in forecasting and measuring correlation, are proper of the calibrated version.

On theoretical ground, the simple OU two-factor model satisfies all the required criteria – with the exception of null probability of negative intensity. Analytical tractability is obvious. As for parsimoniousness, it will be shown later that this model requires $2 + 4/k$ parameters for each generation, k being the number of cohorts considered. Clearly if the number of generations is large (as should be the case with mortality surface) the number of parameters approaches 2, which is exactly the case found by Luciano and Vigna (2008) in the single-generation version of this model. Notice that, in the classical Lee-Carter model, if N is number of

ages, one would need $2(N-1)$ parameters, which is in line with the number of parameters needed in this model. Moreover, in Section 7 we show that this number of parameters can be dramatically reduced at the expense of a slightly worse fit.

Deterministic forecasting at time 0 is feasible because the model provides a survival curve $S(0, \tau)$ in closed-form for any duration τ. Stochastic forecasting at time 0 consists in simulating the survival curve which will hold at time $p > 0$, $S(p, \tau)$, for any duration τ, as illustrated in Section 3.2. This stochastic forecasting is particularly manageable because the factors driving mortality are Gaussian. This allows us also to calculate confidence intervals of simulated survival curves. As for correlations, Equation (32) gives them in closed form.

The corresponding abilities to do forecasting and measuring correlation will be assessed after having calibrated the model.

5. Calibration Criteria and Method

The calibration of survival probability for either one cohort, multiple cohorts, or an entire range of cohorts, for a given age range, amounts to

- assessing how many factors are required to provide a meaningful representation of the mortality surface,
- finding the parameters which make the model-predicted survival probabilities a satisfactory approximation of the observed ones.

The first problem is addressed via Principal Component Analysis, and the second one via minimization of a standard, quadratic measurement error. The minimization of the error is performed using a powerful stochastic search algorithm of Differential Evolution (DE).

5.1. **Principal Component Analysis (PCA).** The problem of finding the number of factors high enough to capture the variability of the phenomenon under examination - in our case, survivorship - and, at the same time, low enough to ensure an accurate estimate of the parameters, given the available data is tackled in a standard way, using PCA. In this section, we explain the manner in which it is applied in the current context. In order to do this, we need to briefly introduce the dataset at our disposal.

To begin with, we collect cohort death rates for life aged x, belonging to cohort i at time t, $q_i(x, t)$, from the Human Mortality Database (2010) (HMD from now onwards), for UK males born between 1900 and 1950. The generations i span from 1900 to 1950, with a 5-year step, for a total of $k = 11$ cohorts.[4] We have chosen $x = 40$ in order to capture the middle-aged lives, and we examine them until they have reached age 59, having $t = 1, ..., 19$. From the rates $q_i(x, t)$, we compute the actual survival frequencies over the duration τ, according to the standard formula

$$\tilde{S}_x^i(0, \tau) = \prod_{s=1}^{\tau} (1 - q_i(x + s - 1, s - 1)),$$

---

[4]Unreported experiments show that all the qualitative results of the paper remain unaltered with smaller as well as larger time-intervals (such as 2, 7 and 10 years).

for $\tau = 1, ..., 19$. The choice of 19 is governed by the following reasons: we need to have a satisfactory number of observed survival frequencies for the last generation as well. This generation is born in 1950, and is aged 40 in 1990. Since the HMD has collected data until 2008, the above choice does indeed guarantee that the last generation has 19 observations. For older generations, even though more than 19 observed survival rates were available in 2008, we have opted not to use them, but have instead taken 19 observed rates for every generation, in order to keep the fit properties of the model comparable across generations.[5]

So as to apply PCA, we use observed survival frequencies to obtain the corresponding average mortality intensity for generation i, duration τ_j, and initial age x.

$$\bar{\mu}_x(0, \tau_j) = -\frac{1}{\tau_j} \log \tilde{S}_x^i(0, \tau_j) = -\frac{1}{\tau_j} \sum_{s=1}^{\tau_j} \log(1 - q_i(x + s - 1, s - 1)).$$

Figures 1 and 2 represent, respectively, the survival probability frequencies and the average mortality intensity for the generations included in our dataset, and all the observations in the HMD between 1900 and 2009, for the initial age 40.
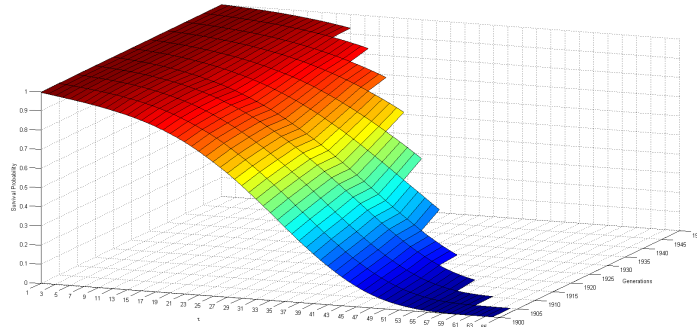

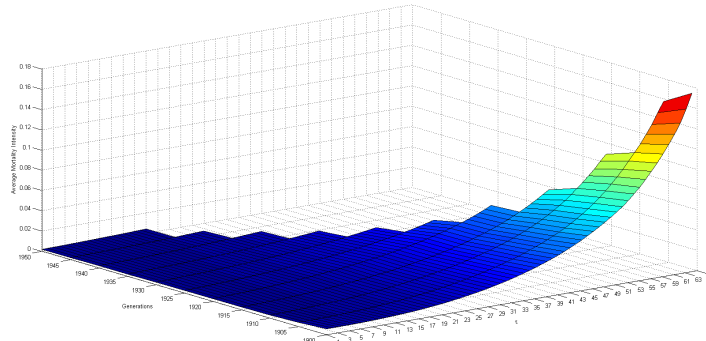FIGURE 1. Survival Probability Surface representing the dataset

FIGURE 2. Average Mortality Intensity Surface from the dataset

[5]Data availability - in particular, the fact of having the observed survival frequencies for ages beyond 59 for every generation but the 1950 one - will be helpful in assessing the out-of-sample performance of the model.

The mean and the first principal component resulting from the PCA, as applied to the average mortality intensities of all of the generations with the initial age of 40, and for durations $\tau = 1, ..., 19$, account for 95.86% of the variation in the data, while the mean, the first and the second principal component account for 99.80% of the variation in the data. This leads us to the conclusion that a 2-factor model is a reasonable candidate.

5.2. **Errors, parameter space and search regions.** Once we fix the number of relevant factors to two, we have four parameters common to all generations $\psi_1, \psi_2, \sigma_1, \sigma_2$ and three parameters for each generation $[\rho^i, X_1^i(0), X_2^i(0)]$ for total of $4 + 3k$ parameters, where $k = 11$. We collect them in a vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{4+3k}$

$$\boldsymbol{\theta} = [\psi_1, \psi_2, \sigma_1, \sigma_2, \rho^1, \rho^2, ..., \rho^k, X_1^1(0), X_1^2(0), ..., X_1^k(0), X_2^1(0), X_2^2(0), ..., X_2^k(0)].$$

In principle, we may allow for $\psi_1, \psi_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}^+$, $-1 \le \rho^i \le 1$ and $-\infty \le X_1^i(0), X_2^i(0) \le \infty$ for $i \in \{1, 2, ...k\}$. However, during the calibration procedure, for practical reasons and given the results of preliminary experiments, we restrict our search space Θ to

$$\Theta = \left\{ \psi_1, \psi_2 \in [-1, 1], \ \sigma_1, \sigma_2 \in [0, 1], \ \boldsymbol{\rho} \in [-1, 1]^k, \boldsymbol{X}_1(0) \in [-1, 1]^k \right\}.$$

Notice that the vector $\boldsymbol{X}_2(0)$ is missing from the search space. This is due to the fact that we impose the constraint that the probability of negative intensities is not greater than 1%, $\Pr(\mu^i(\tau) < 0) \le 1\%$, for every duration $\tau \in \{1, 2, ..., 69\}$, where $69 = \omega - 40 - 1$ is the extreme remaining lifetime. The constraint is respected by appropriately fixing the initial value of the second factor $X_2^i(0)$, for each generation i. This is attained in two steps. First, for each τ, we replace (31) in (28) and choose $X_2^i(0; \tau)$ so that the constraint is respected. Second, to abide by the constraint for every τ we take the maximum of the values $X_2^i(0; \tau)$ as follows

$$X_2^i(0) = \max\{X_2^i(0; \tau = 1), X_2^i(0; \tau = 2), ..., X_2^i(0; \tau = 69)\}.$$

This procedure effectively reduces the number of parameters to four parameters common to all generations and two parameters per each generation. This means a total of $4 + 2k$ parameters, meaning $2 + 4/k$ parameters for each generation, as mentioned in Section 4.

The preliminary calibration experiments we have conducted led us to conclude that for ease of calibration and model interpretation, it might be useful to separate the parameter space Θ into the following four disjoint regions, that differ from each other only in the signs of ψ_1 and ψ_2:

NN	$\psi_1 \in [-1, 0], \psi_2 \in [-1, 0], \sigma_1, \sigma_2 \in [0, 1], \boldsymbol{\rho} \in [-1, 1]^k, \boldsymbol{X}_1(0) \in [-1, 1]^k$
NP	$\psi_1 \in [-1, 0], \psi_2 \in [0, 1], \quad \sigma_1, \sigma_2 \in [0, 1], \boldsymbol{\rho} \in [-1, 1]^k, \boldsymbol{X}_1(0) \in [-1, 1]^k$
PN	$\psi_1 \in [0, 1], \quad \psi_2 \in [-1, 0], \sigma_1, \sigma_2 \in [0, 1], \boldsymbol{\rho} \in [-1, 1]^k, \boldsymbol{X}_1(0) \in [-1, 1]^k$
PP	$\psi_1 \in [0, 1], \quad \psi_2 \in [0, 1], \quad \sigma_1, \sigma_2 \in [0, 1], \boldsymbol{\rho} \in [-1, 1]^k, \boldsymbol{X}_1(0) \in [-1, 1]^k$

We address the problem of finding the estimated value of the parameters as follows: we fix the estimation error, and select an optimization algorithm for the error. As for the error itself, we minimize the mean square error between the actual and estimated parameters with the mean computed across generations and maximum duration τ. This has resulted in the following optimization problem:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\min} \sqrt{\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{\boldsymbol{\tau}} (\tilde{S}^i(j) - S^i(j; \boldsymbol{\theta}))^2}.$$

with $\tau = 19$. The choice of the objective function, by no means unique, was guided by usual practice in financial and actuarial literature.

5.3. **Differential Evolution (DE).** Given the nature of the data and our model, it was necessary for us to find very precise values for a large number of parameters, in the setting of a non-continuous, non-linear, potentially flat and multidimensional objective function. To achieve this high level of solution accuracy, we have opted to use the Differential Evolution (DE) algorithm, which is a global stochastic search tool. This algorithm was formulated in Storn and Price (1997), and its effectiveness and robustness have been proven across a wide range of engineering applications. A practical and detailed exposition of Differential Evolution can be found in Feoktistov (2006). As far as its application to finance-oriented problems is concerned, it has been used in the calibration of the Heston model in Gilli and Schumann (2010), and further elaborated in Gilli and Schumann (2011). In addition, a useful reference on its application in finance is Ardia et al. (2011), where it was implemented in the statistical programming language and environment R. The algorithm, its description and the specifics related to its implementation are provided in the Appendix.

In total, we have performed 150 calibration experiments (trials). 30 individual trials have been made for the entire search space, and for each of the regions (Θ, NN, NP, PN, PP). The number of iterations was set to be $n_I = 100000$, a choice governed by extensive preliminary experiments.

All of our experiments have been conducted using the grid platform of The Wharton School, University of Pennsylvania, PA, USA. Each trial has been conducted on a single 2.5GHz core with 4GB RAM memory, lasting 24 hours on average.

6. Calibration to the UK males population, 1900-1950

Using the DE search algorithm, we minimized the calibration error under the parameter restrictions described above, for the UK males generations from 1900 to 1950. As explained above, the PCA and the calibration were performed for all the 11 generations born in that range (with 5 years distance in birth). Before presenting the detailed results it is important to notice that the interpretation of the intensities is quite different across the four regions considered.

In the NN region, we have intensities mean-reverting to zero. Since ψ_1, ψ_2 are negative, they are indeed weighted sums of factors which mean-revert to zero. This is difficult to conceptualize, given that we are working generation by generation.

In the PP region, the intensities of every generation are straightforward extensions of the Gompertz law, in the sense that - since ψ_1, ψ_2 are positive - their factors present a Gompertz-like drift. So, the intensities are sums of factors with a positive drift. This region then lends itself to standard interpretations.

The NP region combines a first factor which affects every generation in the same manner - mean reverting to zero, with a second factor - which instead affects each generation with a weight dependent on ρ^i - of the Gompertz type. Symmetrically, the PN region combines a first factor of the Gompertz type, with a second factor mean reverting to zero. In principle, we have no reason to prefer one region to the other: intuition supports both regions. Intensities can come out from either a common and equally weighted mean reverting factor, together with a common, not-equally weighted Gompertz one, or vice versa.

Table 1 shows the minimum cost obtained across all of the regions considered - the entire surface and each of the four sub-regions.

TABLE 1. Summary of errors obtained across regions

Region	Cost (100000 iterations)
Entire	0.000620565995355
NP	0.000623213482169
PP	0.0006951410031433
PN	0.0007798886067686
NN	0.0015177865228648

6.1. **Parameter values and residuals plots.** Calibration costs, as shown in Table 1, indicate that the data at hand are best described by the parameters found by calibration in the entire region. Even if this minimum lies in the PP region, parameters obtained by exclusively searching in the PP region have slightly higher cost. This leads us to the conclusion that the algorithm is more successful when search is made in the entire region. Our extensive, non reported calibrations, show that even if we increase the number of iterations, the difference in cost does not change substantially. The fact that it turns out more efficient searching in the entire region is apparently counterintuitive. The explanation to this phenomenon is that when the optimal solution is close to boundary – which is our case – it is easier to reach it if there are no restrictions in a sufficiently large neighborhood (indeed, if there are no restrictions, jumps from one vector of parameters to another one are feasible in every direction).

The cost obtained by searching in the NP region is very close to the best cost obtained, making this region an equally good candidate, as shown in Table 2. However, we will remain focused on the results obtained in the entire region, while bearing in mind that all of the conclusions are very similar for the results obtained in the NP region, as well. In contrast to the PP and the NP regions, the costs for PN and NN regions are substantially higher.

The fact that the minimum is obtained in the PP region, that is the most natural extension of the Gompertz law, reinforces the standard actuarial intuition. The parameters of the minimum are shown in Table 3. Notice that the parameters of the second state process X_2 are in line with the parameters found by Luciano and Vigna (2008) for the single generation case. In contrast, the first state process X_1 has a negligible drift. The volatilities σ_1 and σ_2 are same order of magnitude than the values found by Luciano and Vigna (2008).

In Figure 3 we give the residuals plot for the entire region. It demonstrates that the quality of calibration is exceptionally high. Apart from one point (for which the

error is -3×10^{-3} all residuals stay in the range $[-2 \times 10^{-3}, 2 \times 10^{-3}]$. Moreover, no structural patterns can be observed.

TABLE 2. Calibration results for the NP range

		Cost	0.000623213482169	
		$\psi 1$	-0.000000759867140	
		$\psi 2$	0.094458222928995	
		$\sigma 1$	0.000809865341047	
		$\sigma 2$	0.000255507993644	
		ρ	$X_1(0)$	$X_2(0)$
	1900	0.999999999999997	0.001914489564829	0.002810307411366
	1905	0.203832575710220	-0.000173325764832	0.003149605491684
	1910	-0.184887841334923	-0.000918083027357	0.003221076614512
	1915	-0.283221855026705	-0.000972068597715	0.003158678345694
	1920	-0.078522244367503	-0.000806038099239	0.003260704168707
	1925	-0.244974348716861	-0.000400779285406	0.002834597714541
	1930	-0.420485708967922	0.000126542816362	0.002310756842044
	1935	-0.663902687935100	0.000302566907295	0.001896738812892
	1940	-0.796581809076778	0.000376485904836	0.001665360110214
	1945	-0.978405355219073	0.000412208241716	0.001354942639571
	1950	-1.000000000000000	0.000352615384141	0.001354381214249

TABLE 3. Calibration results for the entire range (parameters in PP)

	Cost	0.000620565995355	
	$\psi 1$	0.000000740897643	
	$\psi 2$	0.094424069684062	
	$\sigma 1$	0.000810431271431	
	$\sigma 2$	0.000255756355983	
	ρ	$X_1(0)$	$X_2(0)$
1900	0.999999999999999	0.001880353247158	0.002828417849011
1905	0.199224719402607	-0.000206634328718	0.003166989667529
1910	-0.192238943851852	-0.000951378602887	0.003238435491859
1915	-0.290643706099185	-0.001005127076371	0.003175910058005
1920	-0.084383114349336	-0.000839346209620	0.003278107036086
1925	-0.250316766642870	-0.000433647930847	0.002851663487249
1930	-0.424028709807915	0.000094264184875	0.002327403237757
1935	-0.667855536351835	0.000270705008126	0.001913047162555
1940	-0.800059032075857	0.000344882262500	0.001681478737687
1945	-0.981185827217655	0.000380945058314	0.001370811095613
1950	-0.999999999999999	0.000351991785546	0.001355915840057

6.2. **Correlations between generations.** In Table 4, we present the correlations table for the calibration results obtained in the entire region. We observe that the correlations are positive and high, which is in accordance with actuarial intuition. They stay between 94% and 100% and, as expected, tend to decrease with the difference in the years of birth. To the best of our knowledge, this is the first research that provides the actuary with a calibrated and sensible correlation among different generations. Notice that if we were to calibrate the model using data which include the Lexis point (i.e. the flexion point of the survival curve) we would be likely to get lower correlations. We do not present those results since we are interested in the central ages of life.
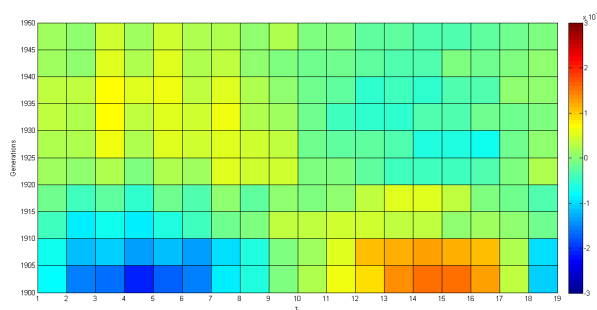
FIGURE 3. Calibration residuals plot

TABLE 4. Table of correlations

	1900	1905	1910	1915	1920	1925	1930	1935	1940	1945	1950
1900	1.0000										
1905	0.9601	1.0000									
1910	0.9497	0.9993	1.0000								
1915	0.9489	0.9992	0.9999	1.0000							
1920	0.9515	0.9995	0.9999	0.9999	1.0000						
1925	0.9491	0.9993	0.9999	1.0000	0.9999	1.0000					
1930	0.9496	0.9993	1.0000	0.9999	0.9999	0.9999	1.0000				
1935	0.9584	0.9999	0.9995	0.9994	0.9997	0.9995	0.9995	1.0000			
1940	0.9693	0.9993	0.9975	0.9973	0.9979	0.9973	0.9975	0.9991	1.0000		
1945	0.9961	0.9810	0.9735	0.9729	0.9749	0.9731	0.9735	0.9798	0.9872	1.0000	
1950	1.0000	0.9601	0.9497	0.9489	0.9515	0.9491	0.9496	0.9584	0.9693	0.9961	1.0000

6.3. **Probability of negative intensities over calendar time/age.** Figure 4 shows that, as it should be, the constraint of 1% on the probabilities of negative intensities for the for generations $1900, 1905, \ldots, 1945$ is respected for all the relevant durations $\tau \in \{1, 2, ..., 69\}$, and, for most part, well bellow the required level.
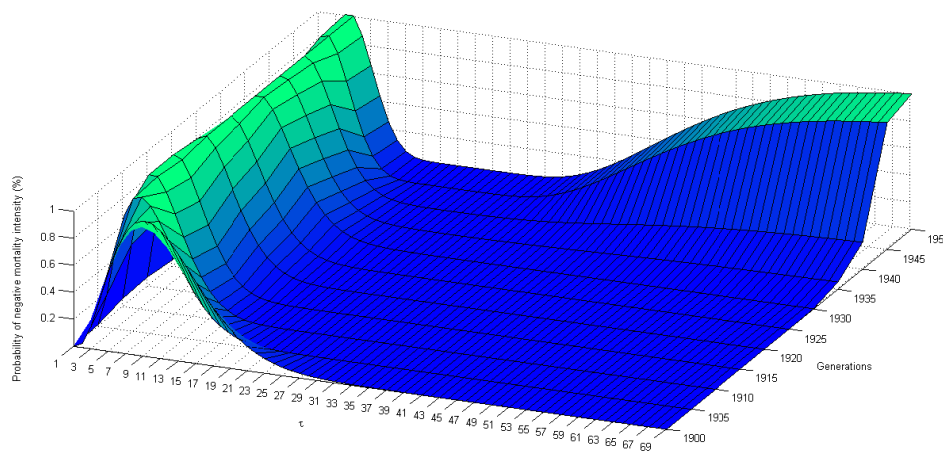

FIGURE 4. Probability of negative mortality intensities surface

6.4. **Deterministic forecasting.** Figure 5 illustrates the results of deterministic forecasting for $\tau \in \{1, 2, ..., 40\}$. In-sample data comprise of points where $\tau \leq 19$ whereas for $\tau \geq 20$ comprise out-of-sample data.

We can see that forecasting error is remarkably small, below 1% for in-sample data. However, it increases up to 26% for out-of-sample data in case of $\tau = 40$ and has the tendency of a sharp increase afterwards. This increasing tendency has been observed also by Blackburn and Sherris (2012), who remedy this by introducing a third factor – proper of very old ages – in their model.
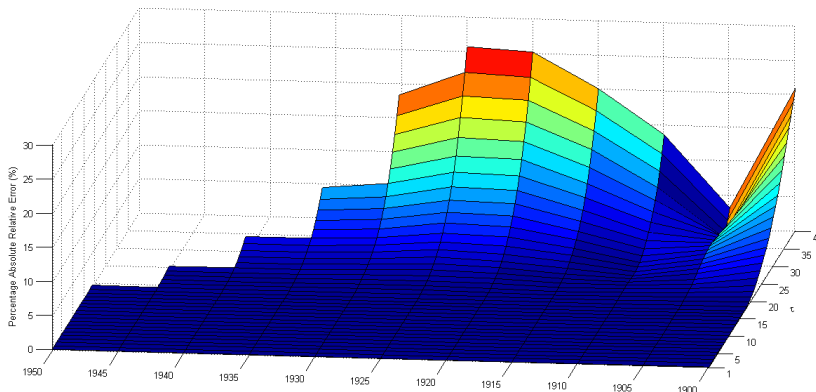

FIGURE 5. Percentage Forecasting Error

6.5. **Stochastic forecast.** In Figure 6, we plot the survival curve at time $t = 1$, $S(1, \tau)$, for generation 1950, as a function of τ. As explained above, the stochastic mortality framework is characterized by the fact that at $t = 0$ – when the calibration is performed – the survival curve which will apply one year later, at $t = 1$, is a random variable.

In the OU case, we have $S(1, \tau)$ as in (33) with $p = 1$. We make 100,000 simulations according to this formula and report in Figure 6 the median (green line), the 5th percentile (red line) and 95th percentile (blue line) of the simulations as functions of τ. The figure also contains survival probability as observed one year later (dots). The in-sample forecasting is very accurate. Almost all the survival probabilities lie on the median (or very close to it), and therefore they stay in the 90% confidence interval. For this generation, unreported in-sample forecasting ($p \leq 18$) performs well, with the survival probabilities lying in the confidence interval.

7. IMPROVEMENT OF PARSIMONIOUSNESS

Up to this point, we have had a total of $4 + 2k$ parameters to be calibrated: $\psi_1, \psi_2, \sigma_1, \sigma_2$ as well as the vectors $\boldsymbol{\rho}$ and $\boldsymbol{X}_1(0)$. The vector $\boldsymbol{X}_2(0)$ of k parameters was found as a function of the previous $4 + 2k$ parameters and the specified level of probability of negative mortality intensity.
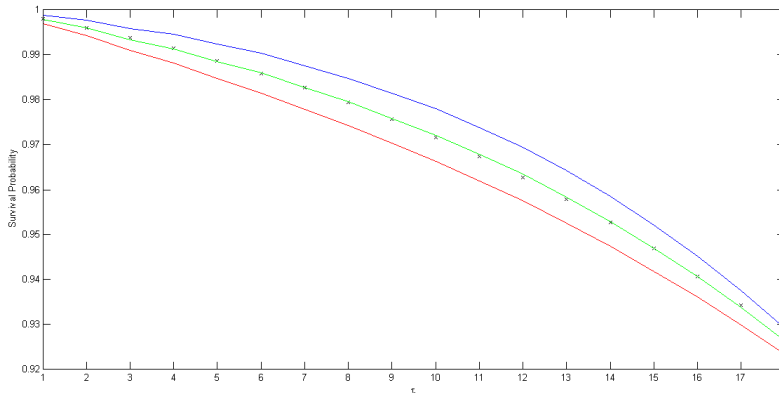
FIGURE 6. Survival probability curve at $t = 1$, $S(1, \tau)$ for generation 1950

There would be room for further parsimoniousness of the model if one were to find an adequate interpolating polynomial for the components of the vector $\boldsymbol{X}_1(0)$. Notice that, given any set of k points, it is always possible to find an interpolating polynomial with degree at most $k - 1$. A higher degree of parsimoniousness can be achieved if the degree of the interpolating polynomial is strictly lower than $k - 1$: the lower the degree of the polynomial, the higher the gain in parsimoniousness – at the expense of a worse fit.

In our calibrations, we have taken generations $i \in \{1900, 1905, \ldots, 1950\}$. With the freely calibrated values of $X_1^{1900}(0)$, $X_1^{1905}(0)$, ..., $X_1^{1950}(0)$ it is possible to see that a simple fourth-order polynomial gives a reasonably accurate interpolation.

To this aim, let us introduce $s = i - 1900$ and set

(34) $$X_1^{1900+s}(0) = as^4 + bs^3 + cs^2 + ds + e,$$

where $s \in \{0, 5, 10, \ldots, 50\}$.

By changing the implementation of our algorithm we evolve the population of parameters $\psi_1, \psi_2, \sigma_1, \sigma_2$, the vector $\boldsymbol{\rho}$, and the parameters a,b,c,d and e. In other words, we replace the k parameters of $\boldsymbol{X}_1(0)$ with the five parameters given by (34). It turns out that each member of the evolving population consists of only $9 + k$ parameters. In addition, we restrict $a, b, c, d, e \in [-1, 1]$, a choice well justified by preliminary unreported experiments, and for each iteration and each member of population we compute k values of $\boldsymbol{X}_1(0)$ and consequently, given all relevant parameter values and the specified level of probability of negative mortality intensities, we compute $\boldsymbol{X}_2(0)$.

The blue dots in Figure 7 report our freely calibrated values of $X_1^{1900}(0)$, $X_1^{1905}(0)$, ..., $X_1^{1950}(0)$, while the green line shows the polynomial fit, and Table 5 reports the polynomial coefficients for which this fit is achieved. Given the results in Table 6 obtained with this new calibration procedure, and which are the minimum cost parameters found among 30 additional experiments, we see that the cost compared to the initial results is only slightly higher. Except for ψ_1, all the other parameters

are very similar. Using the new parameters, we inspect and find that all the relevant graphs and tables shown in Section 6 are almost identical to those already presented, and hence we omit them from the text.
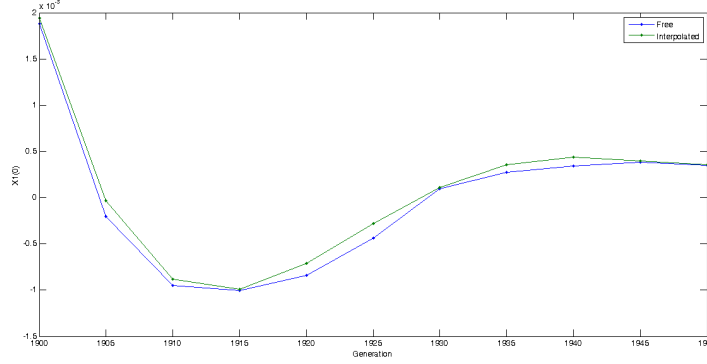

FIGURE 7. Calibrated $\boldsymbol{X}_1(0)$ and its polynomial fit.

TABLE 5. Coefficients of the interpolating polynomial.

Coefficient	Value
a	0.00000000505403258
b	-0.00000069427655836
c	0.00003225047154600
d	-0.00054030580724752
e	0.00194314865699669

TABLE 6. Calibration results with $\boldsymbol{X}_1(0)$ vector interpolated.

	Cost	0.00063804096610967	
	$\psi 1$	0.00127705925258817	
	$\psi 2$	0.09536413447854090	
	$\sigma 1$	0.00080489642885200	
	$\sigma 2$	0.00025398075547516	
	ρ	$X_1(0)$	$X_2(0)$
1900	0.99999999999998500	0.00194314865699669	0.00276410785888053
1905	0.21868230271321300	-0.00003574439002594	0.00306179447920111
1910	-0.18364914494919500	-0.00087859849347549	0.00317889885787329
1915	-0.29513107966522100	-0.00099240533914958	0.00314349059945193
1920	-0.06625258367917220	-0.00070834612419564	0.00319214918562749
1925	-0.22860527966171200	-0.00028179155711086	0.00275878455683557
1930	-0.42510345100547400	0.00010769814225771	0.00229777835759209
1935	-0.65817303178655500	0.00035637324271326	0.00185486605323652
1940	-0.79031744800522400	0.00043629450170917	0.00162261553581031
1945	-0.98129829498458300	0.00039533316534903	0.00134830311068250
1950	-0.99999999999940800	0.00035717096838659	0.00133968572813659

There are two distinct benefits from this approach. First, we reduce the total number of parameters to $9 + k$, which yields $1 + 9/k$ per cohort, and for a high

number of generations this gives a considerable savings. Second, the calibration time is greatly shortened. Now we no longer need 100,000 iterations, because 40,000 is quite adequate. This brings down the calibration time from average 24 hours to average 10 hours.

We have just shown that with our base choice of generations and initial ages it is possible to find an interpolating polynomial with a degree (four) much lower than $k-1$ (ten), giving an almost identical fit. Not reported experiments show that for the same choice of generations and different initial ages (from the range between 30 and 40 years old) a polynomial of degree four still gives a good interpolation. In principle, for different choices of cohorts/ages it is always possible to follow the same procedure. In each case the researcher has to make the desired balance between parsimoniousness and accuracy of the fit.

8. Conclusions and further research

This paper is a first attempt to construct an effective cohort-based continuous-time factor model of the mortality surface. We cast the model first in the affine framework, then specialize it to Ornstein-Uhlenbeck factors. The resulting longevity intensity model extends the G2++ interest-rate model, since the factors have different weights for each generation. The main novelty of the model with respect to existing literature is that it allows for imperfect correlation of mortality intensity across generations.

The model is implemented on UK data for the generations born between 1900 and 1950, using HMD data for the period 1900-2008. On these data, two factors are deemed as a reasonable first choice. Calibration by means of stochastic search and the Differential Evolution optimization algorithm proves to produce small errors and to yield robust and stable parameters. Standard criteria desirable for a model of the mortality surface are satisfied.

The calibration confirms that correlation across generations is very high but smaller than one. Up to our knowledge, this is the first calibration of the correlation among mortality intensities of different generations in the academic literature. The calibrated correlations turn out to be sensible and intuitive. The possibility of capturing these correlations, thanks to a generation-based model coupled with DE-driven calibrations, is our major contribution.

The ex-post performance of in-sample stochastic forecasts is very satisfactory. Both in-sample and out-of-sample deterministic forecasts have been examined. In-sample errors up to age 59 are very small, whereas out-of-sample errors remain small at least until age 65. The increase in the error at later ages could be probably amended by the introduction of a third factor. We leave this extension for further research.

References

Ardia, D., David, J., Arango, O., and Gmez, N. (2011). Jump-Diffusion calibration using differential evolution. *Wilmott*, 2011(55):76–79.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: mathematics and economics*, 37(3):443–468.

Biffis, E. and Millossovich, P. (2006). A bidimensional approach to mortality risk. *Decisions in Economics and Finance*, 29(2):71–94.

Blackburn, C. and Sherris, M. (2012). Consistent dynamic affine mortality models for longevity risk applications. *UNSW Australian School of Business Research Paper No. 2011ACTL08.*

Brigo, D. and Mercurio, F. (2006). *Interest rate models: theory and practice: with smile, inflation, and credit.* Springer Verlag.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

Cairns, A., Blake, D., and Dowd, K. (2006a). Pricing death: Frameworks for the valuation and securitization of mortality risk. *Astin Bulletin*, 36(1):79–120.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006b). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Dahl, M. (2004). Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: mathematics and economics*, 35(1):113–136.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Feoktistov, V. (2006). *Differential evolution: in search of solutions*, volume 5. Springer-Verlag New York Inc.

Gilli, M. and Schumann, E. (2010). Calibrating the heston model with differential evolution. *Applications of Evolutionary Computation*, pages 242–250.

Gilli, M. and Schumann, E. (2011). Heuristic optimisation in financial modelling. *Annals of Operations Research*, pages 1–30.

Human Mortality Database (2010). *University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany).* Available at www.mortality.org (data downloaded on 11/03/2010).

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*, 87(419):pp. 659–671.

Luciano, E. and Vigna, E. (2008). Mortality risk via affine stochastic intensities: calibration and empirical relevance. *Belgian Actuarial Bulletin*, 8(1):5–16.

Milevsky, M. and Promislow, D. (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*, 29(3):299–318.

Renshaw, A. and Haberman, S. (2003). Leecarter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255 – 272.

Schrager, D. (2006). Affine stochastic mortality. *Insurance: mathematics and economics*, 38(1):81–97.

Shreve, S. (2004). *Stochastic calculus for finance: Continuous-time models.* Springer Verlag, New York, NY.

Storn, R. and Price, K. (1997). Differential evolutiona simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

APPENDIX: THE DIFFERENTIAL EVOLUTION ALGORITHM

Algorithm 1: DE Algorithm Source: Gilli and Schumann (2010)

---

initialize parameters n_P, n_I, F and CR;

initialize population $P_{j,i}^{(1)}, j = 1, ..., D, i = 1, ..., n_p$;

for $m = 1$ **to** n_I **do**
$\quad P^{(0)} = P^{(1)}$
$\quad$ **for** $i = 1$ **to** n_p **do**
$\quad\quad$ generate $l_1, l_2, l_3 \in \{1, ..., n_p\}, l_1 \neq l_2 \neq l_3 \neq i$
$\quad\quad$ compute $P_{.,i}^{(v)} = P_{.,l_1}^{(0)} + F \times (P_{.,l_2}^{(0)} - P_{.,l_3}^{(0)})$
$\quad\quad$ **for** $j = 1$ **to** D **do**
$\quad\quad\quad$ **if** randb(j) $< CR$ **then** $P_{j,i}^{(u)} = P_{j,i}^{(v)}$ **else** $P_{j,i}^{(u)} = P_{j,i}^{(0)}$
$\quad\quad$ **end**
$\quad\quad$ **if** Cost($P_{.,i}^{(u)}$) $<$ Cost($P_{.,i}^{(0)}$) **then** $P_{.,i}^{(1)} = P_{.,i}^{(u)}$ **else** $P_{.,i}^{(1)} = P_{.,i}^{(0)}$
$\quad$ **end**
end

DE is a parallel direct search and optimization algorithm, using n_P D-dimensional vectors of parameters. In this setting, we refer to n_P as the number of members of the population, while D is the number of parameters to be estimated. In our case, given the restrictions on parameters, we have $D = 4 + 2 * k$ to be the length of the vector $\boldsymbol{\theta}$, while the size of the population remains constant throughout the entire minimization, and is equal to $n_P = 10D$, a number chosen based on a similarly complex calibration, presented in Gilli and Schumann (2011). All of the parameters were initially sampled uniformly from the predetermined parameter space.

In Storn and Price (1997), the algorithm is presented as follows. We denote the number of iterations to be constructed by the algorithm by n_I. In every iteration m, where $m \in \{1, \ldots, n_I\}$ we arrive to a new population $P^{(m)}$, the size of which is always n_P. DE generates new parameter vectors by adding the weighted difference between two population vectors to a third vector. This operation is called *mutation*, and is characterized by the empirically determined constant $F \in [0, 2]$. Mutated vector's parameters are mixed with the parameters of another predetermined vector, the target vector, to create so-called trial vector u. This parameter mixing is referred to as a *crossover* operation. This process is characterized by a constant CR, and its value is also best determined empirically. In the figure reporting the Algorithm 1 above, *randb(j)* is the jth evaluation of a uniform random number generator with outcome $\in [0, 1]$. If the trial vector has the cost function value Cost(.) lower than the target vector, the trial vector replaces the target vector in the following iteration. This operation is called *selection*. Finally, each population vector must once serve as the target vector. In our setting, the search ends after the n_I-th iteration has been reached.

There are many implementations of the Differential Evolution algorithm. We have chosen to make use of the Matlab implementation which is freely available at http://www.icsi.berkeley.edu/~storn/code.html, and have adjusted it for the purposes of our problem and data structures. During our extensive unreported

preliminary calibration experiments, we have found that it is best to choose the parameters $F = 0.5$ and $CR = 0.85$, and, since this algorithm has many variations, to use the DE/rand/1/bin strategy, which is in line with the findings of (Gilli and Schumann, 2011). In addition, since the range of our parameters is constrained to Θ, we have implemented a re-sampling approach when the parameters of the newly constructed population vectors fall out of bounds.