Bauer, Daniel, Matthias Börger, Jochen Ruß and Hans-Joachim Zwiesler (2008). “The Volatility of Mortality,” *Asia-Pacific Journal of Risk and Insurance*, Volume 3, Issue 1: 172-199.

The Volatility of Mortality

Daniel Bauer†
Matthias Börger†
Jochen Ruß†
Hans-Joachim Zwiesler†

Abstract

The use of forward models for the future development of mortality has been proposed by several authors. In this article, we specify adequate volatility structures for such models. We derive a Heath-Jarrow-Morton drift condition under different measures. Based on demographic and epidemiological insights, we then propose two different models with a Gaussian and a non-Gaussian volatility structure, respectively. We present a Maximum Likelihood approach for the calibration of the Gaussian model and develop a Monte Carlo Pseudo Maximum Likelihood approach that can be used in the non-Gaussian case. We calibrate our models to historic mortality data and analyze and value certain longevity-dependent payoffs within the models.

I. Introduction

In classical actuarial mathematics, future mortality rates and survival probabilities are assumed to be known. Life insurance premiums are determined by the principle of equivalence, which is based on the insight that, for a sufficiently large portfolio of insureds, mortality risk is diversifiable (see Bowers et al. (1997) for an introduction to classical actuarial mathematics). However, recent studies show that mortality improvements occur in an unpredictable manner (see, e.g., Currie et al. (2004) or Macdonald et al. (2003)), and it is now a widely accepted fact that *systematic mortality risk*, i.e., the risk that aggregate mortality trends might differ from those anticipated, is an important risk factor in both life insurance and pensions (cf. Cairns et al. (2005b)). In particular, for annuity providers *longevity risk*, i.e., the risk that mortality improvements are higher than anticipated, constitutes a source of risk which is disregarded by classical actuarial mathematics. In contrast to the “classical" mortality risk resulting from small sample deviations (henceforth *unsystematic mortality risk*), the risk arising from non-deterministic mortality improvements is systematic in the sense that it can not be diversified away by an increasing number of insureds within an insurer’s portfolio. Thus, other methods of managing this risk have to be applied, e.g., securitization (see Cowley and Cummins

†Daniel Bauer [dbauer@gsu.edu] is an assistant professor of risk management and insurance in the Robinson College of Business at Georgia State University, Atlanta, Georgia, USA. Matthias Börger [m.boerger@ifa-ulm.de] is a PhD student at the Institute of Insurance, Ulm University and a consultant at the Institute of Finance and Actuarial Science at Ulm, Germany. Jochen Ruß [j.russ@ifa-ulm.de] is managing partner of the Institute of Finance and Actuarial Science at Ulm, Germany. Hans-Joachim Zwiesler [hans-joachim.zwiesler@uni-ulm.de] is professor of Actuarial Sciences at Ulm University, Germany. The authors are very grateful to the *Institute for Epidemiology*, Ulm University, for their time and advice. In particular, we appreciate the help from their director Prof. Dr. Weiland, who passed away suddenly and unexpectedly on 03/19/2007 – we are in mourning for this outstanding scientist and exceptional colleague, whose wisdom and friendliness will remain unforgotten.

(2005) for an introduction to securitization in life insurance and Blake et al. (2006a) for a survey of mortality linked securities for managing longevity risk).

In order to account for systematic mortality risk in life insurance or annuity portfolios and for modeling mortality linked securities, models are needed which consider the stochastic characteristics of the mortality evolution. In recent literature, a number of such models have been presented (see Cairns et al. (2005b) for an overview and a categorization of stochastic mortality models).

Most of the proposed models are so-called spot force models as they model the spot mortality rate $q_x(t)$ or the spot force of mortality $\mu_t^s(x)$ (cf. Cairns et al. (2005b)). For example, several authors have presented stochastic versions of well-known deterministic mortality laws by replacing some or all of the parameters by stochastic processes (see, e.g., Milevsky and Promislow (2001) or Schrager (2006)). Loosely speaking, these spot force models are a stochastic version of period life tables, where mortality rates for all ages for a certain year are noted.

However, annuity products are typically priced based on generation tables rather than period tables, where some (expected) mortality trend is already considered, i.e., mortality rates for each cohort of x-year olds for each future year are provided. These correspond to so-called forward mortality models, which have been introduced by Milevsky and Promislow (2001) and Dahl (2004), and studied in more detail by Cairns et al. (2005b), Miltersen and Persson (2005), and Bauer (2008). Here, in contrast to spot force models, the whole age/term structure of mortality is modeled. However, to the authors' knowledge, no concrete forward models have been presented thus far.

As pointed out by Bauer and Russ (2006), given the initial term structure of mortality, for the valuation of mortality contingent securities it is sufficient to fix an adequate volatility structure since the drift term is implied by the so-called Heath-Jarrow-Morton drift condition (see Cairns et al. (2005b) or Miltersen and Persson (2005)). The goal of this paper is therefore to derive an adequate model for the *volatility of mortality*, which, on the one hand, is capable of capturing empirical effects as well as epidemiological and demographic insights and, on the other hand, is still tractable in terms of calibration.

The text is organized as follows: In Section II, we introduce the forward force mortality modeling framework based on Bauer (2008); aside from providing the necessary definitions, we present and discuss Heath-Jarrow-Morton drift conditions for different situations.

When building their well-known model for the interest rate term structure, Nelson and Siegel (1987) relied on their economic insights on how the yields' structure may be transformed. Similarly, when building a model for the term structure of mortality, one should take into account insights how mortality may evolve. However, this is not primarily an economic or mathematical question. Therefore, in Section III, we present opinions and insights regarding possible future evolutions for the term structure of mortality from demographic and epidemiological points of view and discuss how these insights can be reflected in mathematical mortality models. Based on this discussion, we derive and specify our model for the volatility of the forward force of mortality in the subsequent section.

Section V describes the calibration procedures: In the Gaussian case, i.e., for deterministic volatility structures, we can determine the distribution of the forward force of mortality and, on this basis, are able to calibrate the model using Maximum Likelihood estimation. For non-Gaussian specifications, the distribution may not be known or may be hard to derive. Hence, we propose a "Pseudo Maximum Likelihood" approach, which does not rely on a specific structure of the volatility. Based on described data, illustrative numerical results as well as some applications are presented in the next section. We close with a summary and discuss limitations as well as possible extensions of our approach.

II. Forward Force of Mortality Modeling

Similar to Cairns et al. (2005b), we consider the definition of stochastic mortality rates via a (hypothetical) exogenously given longevity bond market. However, the respective quantities may also be defined using an intensity based modeling approach, for example by using so-called doubly stochastic processes (see, e.g., Biffis et al. (2005) or Miltersen and Persson (2005)) – under some assumptions on the *market price of mortality risk*, which are satisfied when modeling and pricing mortality linked securities, the approaches are equivalent (see Bauer (2008)). We do not focus on generality, but we only provide the necessary ideas for our application – for a deeper study we refer to Bauer (2008).

The Forward Force of Mortality For the remainder of the paper, we fix a time horizon T^* and a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$, where $\mathbf{F} = (\mathcal{F})_{0 \leq t \leq T^*}$ is assumed to satisfy the usual conditions, i.e., P-completeness and right continuity. Furthermore, we fix a (large) underlying population of individuals at inception, where each individual has a certain age denoted by x_0.

Following Cairns et al. (2005b), we denote by $\Pi_t(T, x_0)$ the price at time t of a (T, x_0) (Longevity) Bond, i.e., a financial security paying ${}_Tp_{x_0}^{(T)}$ at maturity T, where ${}_Tp_{x_0}^{(T)}$ denotes the proportion of x_0 year olds at inception who are still alive at time T (the survival rate or the "realized" survival probability). Denoting the price of a zero coupon bond with maturity T at time t by $p(t, T)$, we can define the forward force of mortality at time t with maturity T by:

$$\tilde{\mu}_t(T, x_0) := -\frac{\partial}{\partial T} \log \left\{ \frac{\Pi_t(T, x_0)}{p(t, T)} \right\}. \tag{1}$$

This implies:

$$\Pi_t(T, x_0) = {}_tp_{x_0}^{(t)}\, p(t, T)\, e^{-\int_t^T \tilde{\mu}_t(s, x_0)\, ds}, \tag{2}$$

which motivates the definition and clarifies the similarity to forward interest rates.

Assuming independence of financial and biometric events[1] and assuming that there exists a risk-neutral measure Q for the market, Equations (1) and (2) yield:

$$\begin{aligned}
\frac{\Pi_t(T, x_0)}{{}_tp_{x_0}^{(t)}} &= p(t, T) E_Q \left[\left. \frac{{}_Tp_{x_0}^{(T)}}{{}_tp_{x_0}^{(t)}} \right| \mathcal{F}_t \right] \\
&= p(t, T) E_Q \left[\left. {}_{T-t}p_{x_0+t}^{(T)} \right| \mathcal{F}_t \right] \\
\Rightarrow E_Q \left[\left. {}_{T-t}p_{x_0+t}^{(T)} \right| \mathcal{F}_t \right] &= e^{-\int_t^T \tilde{\mu}_t(s, x_0)\, ds}.
\end{aligned} \tag{3}$$

Now, we further assume that for every fixed $T > 0$ and x_0, the forward forces $\tilde{\mu}_\cdot(T, x_0)$ have stochastic differentials, which, under Q, are given by:[2]

$$d\tilde{\mu}_t(T, x_0) = \tilde{\alpha}(t, T, x_0)\, dt + \tilde{\sigma}(t, T, x_0)\, d\tilde{W}_t,\ \tilde{\mu}_0(T, x_0) > 0, \tag{4}$$

[1] This assumption has been challenged by, among others, Miltersen and Persson (2005) and Bauer and Russ (2006). At this stage, we will nevertheless rely on it for the sake of simplicity, but we intend to consider correlations in future work.

[2] We omit technical assumptions on the structure, but implicitly assume that all quantities are regular enough such that all following operations are well-defined (see Bauer (2008) for details).

where $\tilde{W}_t$ is a finite dimensional Brownian motion under Q. Therefore, in order to specify a model for the term structure of mortality based on these generics, we have to:

1. Find (or define) the initial forward plane $\tilde{\mu}_0(\cdot,\cdot) : \mathbb{R}^2 \to \mathbb{R}$.
2. Specify the drift term $\tilde{\alpha}(t,T,x_0)$ and the volatility structure $\tilde{\sigma}(t,T,x_0)$.

When directly modeling under the risk-neutral measure Q, the situation simplifies as the drift term is implied by the volatility structure:[3]

$$\tilde{\alpha}(t,T,x_0) = \tilde{\sigma}(t,T,x_0) \int_t^T \tilde{\sigma}(t,s,x_0)'\, ds. \tag{5}$$

This is the well-known Heath-Jarrow-Morton (HJM) drift condition, which – for mortality modeling – has been derived by Cairns et al. (2005b) and Miltersen and Persson (2005). For pricing interest rate derivatives, the corresponding equation for forward interest rates can be handily applied for calibrating the volatility structure to market prices. A similar idea for mortality modeling has been presented by Bauer and Russ (2006), who propose to calibrate the mortality volatility structure to mortality contingent options within life insurance products, for example certain Guaranteed Minimum Benefits within Variable Annuity contracts. However, at this point the market is not liquid enough to find adequate data for deriving meaningful parameterizations. It is thus necessary to rely on historical mortality data – and hence, we need to model the real world dynamics under the physical measure P. Therefore, the question arises whether, under this measure, it is also sufficient to specify a volatility structure in order to build a model.

The Best Estimate Forward Force of Mortality In the previous subsection, the forward forces have been defined based on longevity bond prices or, in other words, based on expected values of survival rates under a risk-neutral measure; from Equations (1) and (3), it particularly follows that:

$$\begin{aligned} \tilde{\mu}_t(T,x_0) &= -\frac{\partial}{\partial T} \log\left\{ E_Q\left[{}_T p_{x_0}^{(T)} \middle| \mathcal{F}_t \right] \right\} \\ &\overset{T\geq t}{=} -\frac{\partial}{\partial T} \log\left\{ E_Q\left[{}_{T-t} p_{x_0+t}^{(T)} \middle| \mathcal{F}_t \right] \right\}. \end{aligned} \tag{6}$$

However, typically only best estimate generation tables are available which do not provide “risk-neutral" expectations, but rather expectations under the real world measure P. Thus, we define forward rates based on these best estimates:

$$\begin{aligned} \hat{\mu}_t(T,x_0) &:= -\frac{\partial}{\partial T} \log\left\{ E_P\left[{}_T p_{x_0}^{(T)} \middle| \mathcal{F}_t \right] \right\} \\ &\overset{T\geq t}{=} -\frac{\partial}{\partial T} \log\left\{ E_P\left[{}_{T-t} p_{x_0+t}^{(T)} \middle| \mathcal{F}_t \right] \right\}. \end{aligned} \tag{7}$$

In oder to differentiate between the “risk-neutral" forward rates from the previous subsection and forward intensities from Equation (7), we will refer to the latter as *best estimate* forward forces.

Let us now consider the dynamic evolution of best estimate forward forces by assuming P-differentials of the form:

$$d\hat{\mu}_t(T,x_0) = \hat{\alpha}(t,T,x_0)\, dt + \hat{\sigma}(t,T,x_0)\, dW_t,\ \hat{\mu}_0(T,x_0) > 0, \tag{8}$$

[3] Here, x' denotes the transpose of the vector x.

where W_t is a Brownian motion under P of the same dimension as $\tilde{W}_t$. Two natural questions arise:

- Can we make any inference about the drift term of these dynamics?
- How do the dynamics of the different forward forces relate?

The first question is straight forward to answer: By virtually the same arguments as under the risk-neutral measure, the HJM drift condition also holds for best estimate forward intensities under the physical measure P (see Bauer (2008)). Therefore, we have (cf. Equation (5)):

$$\hat{\alpha}(t, T, x_0) = \hat{\sigma}(t, T, x_0) \int_t^T \hat{\sigma}(t, s, x_0)'\, ds. \tag{9}$$

Hence, given initial forward rates, it is again sufficient to specify the volatility structure. Based on the volatility structure, we would then be able to, e.g., determine the expected payoff of a mortality contingent security under the measure P. However, in order to price the security, we need to determine the expected (discounted) payoff under some risk-neutral or risk-adjusted measure, which brings us back to the second question.

Here, the answer is not that easy. However, we know that the spot force implied by the "risk-neutral" and the best estimate forward dynamics must be the same:

$$\begin{aligned}
\tilde{\mu}_t^s(x_0) := \tilde{\mu}_t(t, x_0) \quad &= \quad -\frac{\partial}{\partial T} \log\left\{E_Q\left[{}_T p_{x_0}^{(T)} \middle| \mathcal{F}_t\right]\right\}\Bigg|_{T=t} = -\frac{\partial}{\partial t}\log\left\{{}_t p_{x_0}^t\right\} \\
&= \quad -\frac{\partial}{\partial T} \log\left\{E_P\left[{}_T p_{x_0}^{(T)} \middle| \mathcal{F}_t\right]\right\}\Bigg|_{T=t} = \hat{\mu}_t(t, x_0).
\end{aligned}$$

Furthermore, the forward dynamics imply the spot force dynamics. Hence, Proposition 2.1.2 in Bauer (2008) together with (8) and (9) yields that, under P, we have:

$$d\tilde{\mu}_t^s(x_0) = \frac{\partial}{\partial T}\hat{\mu}_t(t, x_0)\, dt + \hat{\sigma}(t, t, x_0)\, dW_t,\ \ \tilde{\mu}_0^s(x_0) = \hat{\mu}_0(0, x_0) > 0. \tag{10}$$

Analogously, the spot force dynamics under Q are implied by the dynamics of the "risk-neutral" forward intensities (see (4) and (5)):

$$d\tilde{\mu}_t^s(x_0) = \frac{\partial}{\partial T}\tilde{\mu}_t(t, x_0)\, dt + \tilde{\sigma}(t, t, x_0)\, d\tilde{W}_t,\ \ \tilde{\mu}_0^s(x_0) = \tilde{\mu}_0(0, x_0) > 0. \tag{11}$$

As the equivalent measure Q is given by its Girsanov density, say:

$$\frac{dQ}{dP}\bigg|_{\mathcal{F}_t} = \exp\left\{-\int_0^t \lambda(s)'\, dW_s - \frac{1}{2}\int_0^t \|\lambda(s)\|^2\, ds\right\},$$

we can also determine the Q-dynamics taking Equation (10) as the starting point. By applying Girsanov's Theorem (see, e.g., Bingham and Kiesel (2003), Theorem 5.7.1), together with (10) we get for the Q-dynamics of $\tilde{\mu}_t^s(x_0)$:

$$\begin{aligned}
d\tilde{\mu}_t^s(x_0) \quad &= \quad \frac{\partial}{\partial T}\hat{\mu}_t(t, x_0)\, dt - \hat{\sigma}(t, t, x_0)\lambda(t)\, dt + \hat{\sigma}(t, t, x_0)\, d\tilde{W}_t, \\
\tilde{\mu}_0^s(x_0) \quad &= \quad \hat{\mu}_0(0, x_0) > 0.
\end{aligned} \tag{12}$$

Finally from (11) and (12), it follows that:

$$\begin{aligned}\tilde{\sigma}(t,t,x_0) &= \hat{\sigma}(t,t,x_0), \\ \frac{\partial}{\partial T}\tilde{\mu}_t(t,x_0) &= \frac{\partial}{\partial T}\hat{\mu}_t(t,x_0) - \hat{\sigma}(t,t,x_0)\lambda(t),\end{aligned} \tag{13}$$

almost surely. This means that we are able to determine $\tilde{\mu}_t(t,x_0)$ if the structure of *the market price of risk* $\lambda(t)$, the best estimate forward dynamics, i.e., $\hat{\sigma}(t,T,x_0)$ and, thus, implicitly $\hat{\mu}_t(T,x_0)$ as well as $\frac{\partial}{\partial T}\hat{\mu}_t(t,x_0)$ are given. However, in general, we cannot conclude $\tilde{\sigma}(t,T,x_0) = \hat{\sigma}(t,T,x_0)$ for all T.

When considering the *Gaussian case*, that is a deterministic choice of $\hat{\sigma}(t,T,x_0)$ and a deterministic market price of risk $\lambda(t)$, we are able to derive a stronger result: In this case, it is easy to verify that:

$$E_Q\left[{}_{T-t}p^{(T)}_{x_0+t}\middle|\mathcal{F}_t\right] = e^{-\int_t^T\int_t^s \hat{\sigma}(u,s,x_0)\lambda(u)\,du\,ds}\, E_P\left[{}_{T-t}p^{(T)}_{x_0+t}\middle|\mathcal{F}_t\right],$$

which yields:

$$\tilde{\mu}_t(t,T,x_0) = \hat{\mu}_t(t,T,x_0) + \int_t^T \hat{\sigma}(s,T,x_0)\lambda(s)\,ds,$$

and hence:

$$\hat{\sigma}(t,T,x_0) = \tilde{\sigma}(t,T,x_0)\ \forall t, T > t, x_0. \tag{14}$$

To conclude this section, we summarize the theoretical results and their practical implications which are important for the remainder of the text:

- Given initial best estimate forward intensities, e.g., from some generation life table, we can build a model for the dynamic evolution of the form (8) for the forward plane under the physical measure P by simply specifying the volatility structure $\hat{\sigma}(t,T,x_0)$. The drift term is given by the P-HJM condition (9).
- The model under the physical measure can be calibrated to observed forward planes for the same underlying population at different points in time.
- Once calibrated, the resulting model may be used to determine expected payoffs, quantiles, etc. of mortality contingent claims. However, it is not yet a pricing model, even though under the assumption of risk-neutrality with respect to mortality risk, i.e., $\lambda(t) \equiv 0$, the models coincide.
- In general, in order to specify a corresponding pricing model, we also have to specify the market price of risk $\lambda(t)$.
- In the Gaussian case and under the assumption of a deterministic market price of risk, Equation (14) yields that the volatilities of risk-neutral and best estimate forward rates coincide. Therefore, given the initial risk-neutral forward plane from (hypothetical) longevity bond prices or as implied by insurance prices (see Bauer and Russ (2006)), the Q-HJM condition (5) and a calibrated volatility structure can yield a pricing model.
- The resulting pricing model can be used to determine the values of mortality securitization products such as longevity bonds or, under some additional assumptions, to determine values of mortality contingent options within insurance products such as Guaranteed Annuity Options (GAOs) or Guaranteed Minimum Benefits (GMBs) within Variable Annuities.

In order to specify a forward model, we are nevertheless still left with a wide range of choices, the “fundamental" ones being the number of risk drivers (i.e., the dimension of the Brownian motion) and the selection of volatility shapes (cf. Angelini and Herzel (2005)). A standard approach would be to conduct a *Principal Component Analysis* (PCA) to determine the number of explaining factors. For forward interest rate models, two factors are usually considered enough for modeling the entire term structure (see Angelini and Herzel (2005)). By analyzing UK yields, Rebonato (1998) even concludes that one factor is enough to explain 92 percent of the variance. For (periodic) mortality data on the other side, Lee and Carter (1992) show that the first singular value explains about 93 percent of the total deviation.

However, these insights are only of limited use for our considerations, as we want to describe forward forces for different terms to maturity and different ages at a time, such that a specific PCA for our situation would be necessary. The first technical problem we are facing is that we need a stationary formulation of the dynamics, which may be achieved by a reformulation of our specification similarly to the well-known Musiela parameterization for forward interest rate models. But aside from this technical peculiarity, we are faced with a more severe practical problem: While interest rate or derivative data is available daily or even intra-daily and periodic mortality rates are quoted at least annually, generation life tables are only compiled in relatively large and irregular intervals. Hence, the data basis for calibrating a given model is very limited and definitely too slim for conducting a meaningful PCA. Furthermore, it is questionable that we will find the same patterns in future mortality evolution as were observed in the past, so that it may be adequate to allow for “additional structures".

Therefore, as indicated in the Introduction, we rely on demographic and epidemiological insights for the specification of our model.

III. Epidemiological Considerations for Mortality Modeling

We focus on the evolution of mortality for industrialized countries and restrict ourselves to one representative population only. In the first subsection, we discuss possible interpretations of the stochastic drivers in a mortality model. In particular, our argumentation is not yet directly linked to the specific modeling framework presented in the previous section. However, we conclude the subsection with a discussion of the adequacy of Brownian motions for modeling the evolution of mortality. In the subsequent subsection, we then focus on the “translation" of epidemiological insights in order to derive a suitable volatility structure within our model framework.

Interpretation of Stochastic Mortality Drivers The most obvious approach to stochastic mortality modeling from an epidemiological background would be to simply consider models already applied in epidemiology. The most prominent models used for analyzing mortality (and also morbidity) effects, are so-called *risk factor models*, *cause of death models*, or combinations of these (see, e.g., Van den Berg Jeths et al. (2001)). An intuitive idea is then to adopt risk factors and/or causes of death from such models for the specification of a stochastic mortality model.

However, there are certain limitations to this approach: With respect to the risk factor approach, Christensen and Vaupel (1996) assess the *determinants of longevity* and conclude that “*probably, a large number of environmental and genetic factors interact to determine lifespan*", for instance smoking, diet, physical activity, medicine, weight, socio-economic status, education etc. It is hardly possible to list all relevant risk factors, let alone to consider all of them in a stochastic mortality model. Associating each factor with a stochastic driver would yield an extremely complex and intractable model.

Alternatively, one could focus on the most crucial factors and thereby reduce the number of stochastic drivers. But, in that case, during the calibration process, some mortality changes would be arbitrarily assigned to some risk factors which is not desirable because the model would loose parts of its interpretability. In general, the problem of always describing only parts of the observed mortality is a critical shortcoming of risk factor models as the results may be misinterpreted (see Van den Berg Jeths et al. (2001)). Furthermore, it is often difficult to assess how changes in (the exposure to) a certain risk factor will change the expectation of future mortality. In particular, due to the same risk factor, mortality can improve or decline for all ages or improve for some ages and decline for others. It is not possible to account for such expectations in a risk factor based mortality model without significantly increasing the number of stochastic drivers.

If, on the other hand, causes of death, such as ischemic heart disease, cerebrovascular accidents, chronic obstructive lung disease, lung cancer, breast cancer, dementia, traffic and other accidents (see Van den Berg Jeths et al. (2001)), are used as drivers for the model, a similar argumentation holds as for the risk factors. Again, it is difficult to determine which causes might be the crucial ones in the future and which ages and points in time will particularly be affected.

Since – for the sake of feasibility – one could not consider all causes one would also need a category “remaining causes". But for that group of causes of death, it is virtually impossible to assess the impact of its changes on particular ages and points in future time. Additionally, the cause of death interpretation is not very suitable for modeling the evolution of very senior age mortality because of the numerous causes of death affecting those ages (see Tabeau et al. (2001)).

Thus, both, a risk factor and a cause of death model do not seem practical. Similar arguments apply when considering other *determinants of mortality*.

Therefore, in what follows, we consider a different, more practical possibility: We combine determinants into groups which affect the mortality for similar ages and at a similar time. For such an approach, the individual random effects do not correspond to a certain determinant, but rather the aggregate effect. We will refer to those determinant groups simply as (age and time dependent) effects and the occurrence of each effect will be modeled by one stochastic driver, in our case a Brownian motion.

While it is difficult to assess whether Brownian motions are adequate stochastic drivers for our purposes, the following heuristic argument shows that the application of diffusion processes might be a reasonable choice within such a modeling approach. Newman and Wright (1981) and later Birkel (1988) for the more general, non-stationary case show that the partial sum process of identically distributed random variables, under some assumptions on their covariance structure, converges in distribution to a standard Brownian motion. As stated above, regarding the evolution of mortality, we have a lot of determinants, i.e., random variables, which are influential, but they are not identically distributed. However, by combining different classes of determinants and transforming them appropriately, the mortality changes resulting from these groups may be close to identically distributed. The distribution of their partial sum will then, by the theorems in Newman and Wright (1981) and later Birkel (1988), converge in distribution to a standard Brownian motion.

Epidemiological Insights and Practical Implementation Even though modeling approaches used in epidemiological research do not seem appropriate for our purpose, qualitative epidemiological and demographic insights are helpful for building specifications. Clearly, there is a trade-off between the “resolution" or “flexibility" of considered effects and the tractability. The question of how to cluster ages and time intervals for which the assumption of a similar mortality evolution is acceptable is certainly a demographic/epidemiological one. Note that we explicitly deal with age and time in combination as there clearly seems to be a connection. At least in the past, the strongest changes in mortality have been observed for varying ages at different times (see, e.g., Weiland et al. (2006) and Vaupel (1986)). For instance, at the beginning of the 20th century, the strongest improvements were observed in infant mortality. Toward the end of the century, this completely shifted to improvements for older ages (see Weiland et al. (2006)).

We allow for ages between 20 and 140 and, thus, have a time horizon of $T^* = 120$ years. A limiting age of 140 may seem rather high from today's point of view, but considering the current increase of life expectancy (cf., e.g., Oeppen and Vaupel (2002)), this number does not appear to be high for a person aged 20 today – in particular in tail scenarios. As a compromise between tractability and flexibility, we divide the age span of 120 years in three groups. The so-called old age group contains the ages 80 to 140. In the past decades, the largest mortality improvements have been observed for these ages (see Christensen and Vaupel (1996)). Wiesner (2001) even states that (at least for the German population), the ages above 65 cannot be treated as a homogeneous group because there are significant differences between what he calls "old" and "oldest-old"[4] people in terms of health status, life style, education and family status. Also, the increase in force of mortality slows down for the oldest-old with the for younger ages exponential trend becoming rather linear then (see Boleslawski and Tabeau (2001)), which again distinguishes the oldest-old from younger ages. Hence, it seems reasonable to treat these ages separately from the others. We divide the remaining ages in two groups: "young ages" from age 20 to 55 and "middle ages" from age 55 to 80.

With respect to time effects, we also use three effects: A short-term, a mid-term and a long-term effect. Clearly, there is a need for a short-term effect that models events that lead to a change of mortality for a limited time only, e.g., a wave of influenza that will lead to an increase in expected mortality for the next couple of months. However, the considered diffusion processes only allow for smooth mortality evolutions, which is not appropriate for modeling e.g., catastrophes. Instead, a jump component might be preferable for modeling short-term effects. This is left for future research. Apart from the short-term effect, longer termed effects are included. Since with only one longer-term effect, the model would be fairly inflexible, we distinguish between changes in the nearer future (mid-term), e.g., the next few decades, and the far future thereafter (long-term).

There is an ongoing scientific debate on whether an upper limit for human life expectancy may exist. Olshansky et al. (1990) and Fries et al. (1989) believe that there is such an upper bound. The converse view is held by, e.g., Oeppen and Vaupel (2002) and Manton et al. (1991). In particular, Oeppen and Vaupel (2002) show that worldwide maximum life expectancy[5] has increased almost linearly over the last 160 years and they state that there is no obvious reason why this trend should not prevail in the future. As, so far, no country's life expectancy has exceeded this trend, in a certain sense, they also postulate some upper bound on worldwide life expectancy – a bound that is time dependent and linearly increasing. So, no matter which expectation will turn out to be true in the future, both claim that, at any given point in time, there is some upper limit for life expectancy and, thus, also a lower limit for mortality. On the other hand, there seems to be no limit on potential increases in mortality. For example, in Russia, life expectancy for men sank from 63.8 years to only 57.7 years between 1990 and 1994 (see Notzon et al. (1998)). That means age-adjusted mortality rose by about 33 percent in just four years (see Notzon et al. (1998)). Although such a development seems rather unlikely in Western industrialized countries, it is still possible. Hence, regarding the distributional property, we expect the forward force of mortality to have a positively skewed distribution with rather light tails.

[4] In general and also in Wiesner (2001), oldest-old refers to ages above 80.

[5] Maximum life expectancy means that for each year, the country with the largest life expectancy is considered.

IV. Modeling the Volatility of Mortality

Specification of the Mortality Effects Based on the qualitative insights presented in Section III, in this section, we derive a quantitative model for the volatility terms $\tilde{\sigma}(t, T, x_0)$ and $\hat{\sigma}(t, T, x_0)$, respectively, from Section II, where, in order to simplify notation, we simply denote the volatility by $\sigma(t, T, x_0)$ as long as no ambiguity arises. Based on the description above, we consider three mid-term effects: one for young ages, one for middle ages and one for old ages. These effects are complemented by a short-term effect and a long-term effect, which both are independent of age, since we expect strong correlations between the short-termed effects for different ages, and the long-term effect is only of importance for relatively young ages as seen from today.

While there is already some correlation incorporated into the model thus far for ages and points in time that are combined in the same mortality effect, there exists no "inter-effect" correlation yet. Such a correlation is important as there exist many mortality determinants which similarly affect the expected mortality evolution for all ages and points in time, for example, improvements in medication, political changes, or pollution. Thus, we introduce another mortality effect which is independent of age and time; it simply models the general tendency of mortality evolution.

We suggest the following functional structure for $\sigma(t, T, x_0)$ (to simplify notation, we denote the three mid-term effects only by their corresponding age group and omit the phrase "mid-term" as only these effects are age dependent):

$$\begin{array}{ll}
\text{general:} & \sigma_1(t, T, x_0) = c_1 \\
\text{short-term:} & \sigma_2(t, T, x_0) = c_2 \cdot \exp\left(-w_2(T-t)\right) \\
\text{young age:} & \sigma_3(t, T, x_0) = c_3 \cdot \exp\left(-w_{31}(T-t-20)^2 - w_{32}(x_0+T-37.5)^2\right) \\
\text{middle age:} & \sigma_4(t, T, x_0) = c_4 \cdot \exp\left(-w_{41}(T-t-20)^2 - w_{42}(x_0+T-67.5)^2\right) \\
\text{old age:} & \sigma_5(t, T, x_0) = c_5 \cdot \exp\left(-w_{51}(T-t-20)^2 - w_{52}(x_0+T-110)^2\right) \\
\text{long-term:} & \sigma_6(t, T, x_0) = c_6 \cdot \exp\left(-w_6(T-t-120)^2\right).
\end{array}$$

Here, $c_1, ..., c_6$ and $w_2,\ w_{31}, ..., w_{52},\ w_6$ are parameters still to be determined. Depending on the choice for w_2, $\sigma_2(t, T, x_0)$ is quickly decreasing in time and, thus, suitable for modeling short-termed mortality. The next three functional representations are two-dimensional bell-shaped curves with maximum at the center of the particular age group and time interval. Obviously, these functions overlap which is useful because, given a reasonable choice of width parameters, that results in a smooth transition from one age (group) and time interval to the next. Furthermore, this leads to additional correlation between adjacent[6] mortality effects. Finally, $\sigma_6(t, T, x_0)$ increases over the whole time frame and is particularly large in the long-term future.

Parameter Elimination In order to reduce the number of free parameters, we impose restrictions on the effects' functional representations. We leave the scaling parameters c_i variable because they are crucial for determining how strong mortality changes for a particular age and time combination are – both absolutely and compared to the other effects. We impose conditions for the width parameters w_i and w_{ij}.

The parameter w_2 determines how long the short-term effect influences the forward force of mortality. We fix this parameter such that, after one year, this effect is only 10 percent of its immediate value, i.e., $w_2 = -\log\left(\frac{1}{10}\right)$.

For the other mortality effects, we require smooth transitions from one effect to the next one, in

[6] Adjacent with respect to age groups and time intervals under consideration.

terms of age as well as in terms of time. We set each effect to be 50 percent of its maximum value at the boundary to adjacent effects. Because of the symmetric construction of the functional representations, this is possible for all effects. By adding the effects' values at the boundaries we automatically receive a smooth transition. Exemplarily, for the middle age effect the conditions are:

$$\text{age condition:} \quad \exp\left(-w_{42}(x_0+T-67.5)^2\right)\big|_{x_0+T=67.5} = 2 \cdot \exp\left(-w_{42}(x_0+T-67.5)^2\right)\big|_{x_0+T=80},$$

$$\text{time condition:} \quad \exp\left(-w_{41}(T-t-20)^2\right)\big|_{T-t=20} = 2 \cdot \exp\left(-w_{41}(T-t-20)^2\right)\big|_{T-t=40}.$$

Hence, the effects in our model read as follows:

$$\begin{array}{ll}
\text{general:} & \sigma_1(t,T,x_0) = c_1 \\
\text{short-term:} & \sigma_2(t,T,x_0) = c_2 \cdot \exp\left(\log(0.1)(T-t)\right) \\
\text{young age:} & \sigma_3(t,T,x_0) = c_3 \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{17.5^2}(x_0+T-37.5)^2\right) \\
\text{middle age:} & \sigma_4(t,T,x_0) = c_4 \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{12.5^2}(x_0+T-67.5)^2\right) \\
\text{old age:} & \sigma_5(t,T,x_0) = c_5 \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{30^2}(x_0+T-110)^2\right) \\
\text{long-term:} & \sigma_6(t,T,x_0) = c_6 \cdot \exp\left(\frac{\log(0.5)}{80^2}(T-t-120)^2\right).
\end{array}$$

Distributional Properties and a Correction Term Up to this point, the implied changes in the forward force of mortality $\hat{\mu}_t(T,x_0)$ are absolute values, i.e., they do not depend on the current level of $\hat{\mu}_t(T,x_0)$. This has several implications. First of all, in extreme scenarios, $\hat{\mu}_t(T,x_0)$ could become negative, although depending on calibration, this may be seen as an acceptable shortcoming of our model. Moreover, the resulting Normal distribution for the forward forces of mortality $\hat{\mu}_t(T,x_0)$ may not be desirable since, in Section **III**, we motivated a light tailed and positively skewed distribution. However, we believe that a Gaussian model can still generate reasonable results and it has some major advantages regarding tractability in terms calibration and applications (see Sections V and VI).

A third implication of the absolute changes in $\hat{\mu}_t(T,x_0)$ is that the total standard deviation will significantly decrease with age relative to the level of $\hat{\mu}_t(T,x_0)$ since the force of mortality is increasing with age. This is not a desirable feature. Hence, we introduce a correction term in the form of a Gompertz curve. While we could also apply other functional forms as correction terms, we choose the Gompertz form due to its simplicity.

Finally, our Gaussian mortality model has the following volatility structure:

$$\boxed{\begin{array}{ll}
\text{general:} & \sigma_1(t,T,x_0) = c_1 \cdot \exp\left(a(x_0+T)+b\right) \\
\text{short-term:} & \sigma_2(t,T,x_0) = c_2 \cdot \exp\left(a(x_0+T)+b\right) \cdot \exp\left(\log(0.1)(T-t)\right) \\
\text{young age:} & \sigma_3(t,T,x_0) = c_3 \cdot \exp\left(a(x_0+T)+b\right) \\
 & \quad \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{17.5^2}(x_0+T-37.5)^2\right) \\
\text{middle age:} & \sigma_4(t,T,x_0) = c_4 \cdot \exp\left(a(x_0+T)+b\right) \\
 & \quad \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{12.5^2}(x_0+T-67.5)^2\right) \\
\text{old age:} & \sigma_5(t,T,x_0) = c_5 \cdot \exp\left(a(x_0+T)+b\right) \\
 & \quad \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{30^2}(x_0+T-110)^2\right) \\
\text{long-term:} & \sigma_6(t,T,x_0) = c_6 \cdot \exp\left(a(x_0+T)+b\right) \\
 & \quad \cdot \exp\left(\frac{\log(0.5)}{80^2}(T-t-120)^2\right)
\end{array}}$$

An Alternative Model with Different Distributional Properties An alternative mortality model which better complies with the distributional properties we established in Section **III**, i.e., a positive skewness and light tails, can be achieved by multiplying $\sigma(t, T, x_0)$ by $\sqrt{\hat{\mu}_t(T, x_0)}$. This leads to a square-root diffusion process for each $\hat{\mu}_t(T, x_0)$ and, hence, implies a distribution similar to the χ^2 distribution (cf. Glasserman (2004, Section 3.4)) which is in fact positively skewed and light tailed.

At the same time, the problem of a decreasing volatility addressed in the previous subsection is eased since the volatility is now proportional to $\sqrt{\hat{\mu}_t(T, x_0)}$. While a correction term, for example, in form of the square-root of a Gompertz curve could be inserted to make the volatility approximately proportional to $\hat{\mu}_t(T, x_0)$, we consider the scaling by $\sqrt{\hat{\mu}_t(T, x_0)}$ sufficient for our purposes and hence, we do not consider any correction term.

Thus, our *Square-root model* has the following volatility structure:

general:	$\sigma_1(t, T, x_0) = c_1 \cdot \sqrt{\hat{\mu}_t(T, x_0)}$
short-term:	$\sigma_2(t, T, x_0) = c_2 \cdot \sqrt{\hat{\mu}_t(T, x_0)} \cdot \exp\left(\log(0.1)(T-t)\right)$
young age:	$\sigma_3(t, T, x_0) = c_3 \cdot \sqrt{\hat{\mu}_t(T, x_0)} \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{17.5^2}(x_0+T-37.5)^2\right)$
middle age:	$\sigma_4(t, T, x_0) = c_4 \cdot \sqrt{\hat{\mu}_t(T, x_0)} \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{12.5^2}(x_0+T-67.5)^2\right)$
old age:	$\sigma_5(t, T, x_0) = c_5 \cdot \sqrt{\hat{\mu}_t(T, x_0)} \cdot \exp\left(\frac{\log(0.5)}{20^2}(T-t-20)^2 + \frac{\log(0.5)}{30^2}(x_0+T-110)^2\right)$
long-term:	$\sigma_6(t, T, x_0) = c_6 \cdot \sqrt{\hat{\mu}_t(T, x_0)} \cdot \exp\left(\frac{\log(0.5)}{80^2}(T-t-120)^2\right)$

V. Calibration of the Models

As pointed out in Section II, we rely on historical data of the mortality age/term structure. Generation life tables report forward survival probabilities and, usually, the risk margins are explicitly noted such that best estimate forward expectations can be conveniently derived. Thus, we can make use of the results from Section II, in particular of the drift condition under the physical measure P from Equation (9). Problematic is the fact that generation tables are not frequently and not regularly compiled, and hence, our data basis is rather sparse.

The task for this section is to find procedures for calibrating parameters of our volatility model or another volatility specification in the same framework, i.e., to find parameters which match the given mortality data in an optimal way. For the *Gaussian case* as introduced in Section II, we can explicitly determine the distributions of the quantities in view and, hence, we are able to present a *Maximum Likelihood* approach in the first subsection. For the general case, the situation is more complex – we present a “Pseudo Maximum Likelihood" approach which is based on Monte Carlo simulation in the second subsection; while the computation is very time-consuming and tedious, it can be applied for very general specifications and, contingent on certain requirements on the data, provides stable results.

The Gaussian Case As we are modeling the dynamic evolution of the age/term-structure of the mortality intensity, the natural approach would be to directly consider a data-set of different mortality intensities for the calibration procedure. However, these are not explicitly quoted in life tables and need to be approximated from the given data (see Section VI). Another possibility without the necessity of using approximations would be to take expected future survival probabilities:

$$E\left[p_{x_0+T}^{(T+1)} \middle| \mathcal{F}_{t_i}\right] = E\left[\exp\left\{-\int_0^1 \hat{\mu}_T(T+u, x_0)\, du\right\} \middle| \mathcal{F}_{t_i}\right]$$

as denoted in the generation table compiled at time $t_i \in \{t_0, ..., t_N\}$, i.e., we assume we are given N generation tables at different times. In practical use, it is more convenient to work with their logarithms:

$$Y_{t_i}(T, x_0 + T) := \log\left\{E\left[p^{(T+1)}_{x_0+T}\,\middle|\,\mathcal{F}_{t_i}\right]\right\}.$$

Note that we have $T \geq t_i$, such that, e.g., $p^{(50)}_{30}$ would imply a negative x_0, namely $x_0 = -19$. As we want to employ all available data, we accept this slight abuse of notation and accept “negative x_0" such that $x_0 + t_i \geq 0$. This shortcoming may be fixed by changing the notation, but the notation used here proves to be handy in other situations.

Some computations yield:

$$\begin{aligned} Y_{t+\tau}(T, x_0 + T) &= Y_{t+\tau}(T, x_0 + T) - \frac{1}{2}\int_t^{t+\tau}\left\|\int_T^{T+1}\sigma(u, s, x_0)\,ds\right\|^2 du \\ &\quad - \int_t^{t+\tau}\int_T^{T+1}\sigma(u, s, x_0)\,ds\,dW_u, \end{aligned}$$

and hence:

$$\Delta(t_i, t_j, T, x_0 + T) := Y_{t_j}(T, x_0 + T) - Y_{t_i}(T, x_0 + T),\ \ t_i \leq t_j \leq T \tag{15}$$

is Normally (Gaussian) distributed with mean:

$$E\left[\Delta(t_i, t_j, T, x_0 + T)\right] = -\frac{1}{2}\int_{t_i}^{t_j}\left\|\int_T^{T+1}\sigma(u, s, x_0)\,ds\right\|^2 du \tag{16}$$

and variance:

$$VAR\left[\Delta(t_i, t_j, T, x_0 + T)\right] = \int_{t_i}^{t_j}\left\|\int_T^{T+1}\sigma(u, s, x_0)\,ds\right\|^2 du. \tag{17}$$

As the Gaussian distribution is completely determined by the first two moments, we are thus given the distribution of $\Delta(t_i, t_j, T, x_0 + T)$. However, since our data-set does not consist of independent quantities, we are interested in the joint distribution for different $t_i,\ t_j,\ T,\ x_0$, which, on the other hand, is completely determined by Equations (16), (17), and the correlation structure. Therefore, we need to compute the covariances between $\Delta(t_i, t_j, T_1, x_1 + T_1)$ and $\Delta(t_k, t_l, T_2, x_2 + T_2)$ for $t_i \leq t_k$. By some algebra and an application of Itô's product rule, we obtain:

$$\begin{aligned} & COV\left[\Delta(t_i, t_j, T_1, x_1 + T_1), \Delta(t_k, t_l, T_2, x_2 + T_2)\right] \\ = \;& 1_{\{t_k \leq t_j\}} E\left[\int_{t_k}^{t_j \wedge t_l}\int_{T_1}^{T_1+1}\sigma(u, s, x_1)\,ds\,dW_u \right.\\ & \qquad\qquad \left. \times \int_{t_k}^{t_j \wedge t_l}\int_{T_2}^{T_2+1}\sigma(u, s, x_2)\,ds\,dW_u\right] \\ = \;& 1_{\{t_k \leq t_j\}}\int_{t_k}^{t_j \wedge t_l}\left(\int_{T_1}^{T_1+1}\sigma(u, s, x_1)\,ds\right)\left(\int_{T_2}^{T_2+1}\sigma'(u, s, x_2)\,ds\right) du, \end{aligned}$$

where $a \wedge b := \min\{a, b\},\ \ a, b \in \mathbb{R}$.

So for different generation tables labeled by their compilation times $\{t_0, ..., t_N\}$ and choices for ages $\{x_0, ..., x_M\}$ as well as maturities $\{T_0, ..., T_K\}$, we have determined the distribution of the random vector $(\Delta(t_i, t_j, T_k, x_l + T_k))$ of dimension D. In particular, the probability density is given by:

$$f(y) = \frac{1}{\sqrt{(2\pi)^D \det\{\Sigma\}}} \exp\left\{-\frac{1}{2}(y-m)'\,\Sigma^{-1}\,(y-m)\right\},\ y \in \mathbb{R}^D,$$

where:

$$\begin{aligned} m &= \left(m_{t_i,t_j,T,x_0}\right) \\ &= \left(-\frac{1}{2}\int_{t_i}^{t_j}\left\|\int_T^{T+1}\sigma(u,s,x_0)\,ds\right\|^2\,du\right) \end{aligned}$$

and:

$$\begin{aligned} \Sigma &= \left(\Sigma_{(t_i,t_j,T_1,x_1),(t_k,t_l,T_2,x_2)}\right) \\ &= \left(1_{\{t_k\le t_j\}}\int_{t_k}^{t_j\wedge t_l}\left(\int_{T_1}^{T_1+1}\sigma(u,s,x_1)\,ds\right)\left(\int_{T_2}^{T_2+1}\sigma'(u,s,x_2)\,ds\right)\,du\right). \end{aligned}$$

However, so far we have neglected that $\sigma(\cdot)$ is only given contingent on a certain parameter vector $c \in \mathbb{R}^d$, i.e., we are in fact only given a parametric form of the density:

$$f(y;c) = \frac{1}{\sqrt{(2\pi)^D \det\{\Sigma(c)\}}} \exp\left\{-\frac{1}{2}(y-m(c))'\,(\Sigma(c))^{-1}\,(y-m(c))\right\},\ y \in \mathbb{R}^D,$$

and our goal is to estimate the parameter vector c from our sample data based on given generation tables. So, given our realization vector $\Delta \in \mathbb{R}^D$, we are given a *likelihood function* by:

$$f(\Delta;\cdot) : \mathbb{R}^d \to [0,\infty)$$

and we want to determine a Maximum Likelihood estimation $\hat{c}$ for the parameter vector such that $\hat{c}$ maximizes the likelihood function, i.e.:

$$f(\Delta;\hat{c}) = \sup\left\{f(\Delta;c) \mid c \in \mathbb{R}^D\right\}.$$

Instead of the likelihood function, we may also consider the log-likelihood function:[7]

$$L(\Delta,c) = -\log\left(\det\{\Sigma(c)\}\right) - (\Delta - m(c))'\,(\Sigma(c))^{-1}\,(\Delta - m(c)), \qquad (18)$$

and, thus, our calibration problem narrows down to the determination of the maximum value for $L(\Delta,\cdot)$, which can be carried out numerically.

This calibration routine as well as all other numerical calculations are implemented in $\mathtt{C}++$ using routines from the GNU Scientific Library (GSL)[8]. In particular, for the numerical computation of integrals we make use of a routine based on the *QNG non-adaptive Gauss-Kronrod* method and for the numerical optimization we apply a routine using the Simplex algorithm of Nelder and Mead (1965). For the latter, it is worth noting that minimization algorithms find local minima – and there is no way to determine whether a local minimum is a global one. We consider this problem by choosing different sets of starting parameters and comparing the resulting values.

[7] Note that we disregard additive terms and positive factors.
[8] See *www.gnu.org* or The GSL Team (2007) for details.

The General Case By the “general case", we denote a volatility structure of the form:

$$\sigma_i(t, T, x_0) = c_i \cdot g_i\left(t, T, x_0, \hat{\mu}_t(T, x_0)\right),$$

where g_i is a non-negative, non-trivial deterministic function, $i = 1, ..., d$, such that the system of stochastic differential equations (8) has a unique strong solution. In this case, the joint probability density is generally not given explicitly. Hence, implementing a similar approach as in the previous subsection is not possible. However, we may still use the same intuition as for Maximum Likelihood estimation: For a scalar random variable X with density f_c, where c is some parameter vector, we know that:

$$P_c\left(|X - x| < \varepsilon\right) \approx f_c(x)2\varepsilon \tag{19}$$

for some “small" $\varepsilon > 0$, where P_c is the (parametrical) probability distribution implied by f_c. Therefore, maximizing $c \to f_c(\bar{x})$ for a given sample outcome $\bar{x}$ is equivalent to maximizing the probability that X is close to $\bar{x}$ under a certain parameterization c.

Since the $\hat{\mu}_{t_i}(\cdot, \cdot)$ are assumed to be continuous, we can consider the Banach space of continuous functions with the supremum-norm $\|f\| = \sup_{x \in D_f} |f(x)|$, where D_f is the domain of the continuous function f.[9] Thus, the probability that $\hat{\mu}_{t_i}$ is close to the sample outcome, say $\bar{\mu}_{ti}$, given some parameter vector c, reads as follows:

$$P_c\left(\|\hat{\mu}_{t_i} - \bar{\mu}_{t_i}\| < \varepsilon\right) = P_c\left(\sup_{T \geq t_i,\, x_0+T \geq x_0+t_i} |\hat{\mu}_{t_i}(T, x_0) - \bar{\mu}_{t_i}(T, x_0)| < \varepsilon\right) \tag{20}$$

for some $\varepsilon > 0$ and given an initial term structure $\hat{\mu}_0 \equiv \bar{\mu}_0$. In order to estimate the parameter vector c, we may now choose the $\hat{c}$ that maximizes these probabilities for all $t_1,\ t_2, ...$, which could, e.g., be determined numerically by Monte Carlo simulation. However, there are some immediate questions:

- Does the algorithm converge, i.e., can we find such an estimate $\hat{c}$?
- Does it depend on the choice of ε, and, if so, how should ε be chosen?

With respect to the first question, we find that in order for the algorithm to converge, the target planes $\bar{\mu}_{t_i}$ must be attainable for some c_ε:

> **Proposition 1.** *Let $\varepsilon > 0$ be fixed and consider only one target plane at $t_1 > t_0 = 0$, $\bar{\mu}_{t_1}$. Then, the algorithm converges, i.e., there exists an estimate $\hat{c}$ that maximizes the probability from (20), if and only if the following condition is satisfied:*
>
> $$\exists c:\ P_c\left(\|\hat{\mu}_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) > 0. \tag{21}$$

For a proof, see the Appendix.

For relatively large choices of ε, this condition will be satisfied, but the resulting estimate may be biased: For example, when choosing ε big enough such that the deterministic model (i.e., $c \equiv 0$) lies in the ε-corridor, the algorithm would automatically yield this deterministic choice as then:

$$P_0\left(\|\hat{\mu}_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) = 1 \geq P_c\left(\|\hat{\mu}_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right)\ \forall c.$$

Thus, and also due to the motivation behind the approach from Equation (19), we should prefer rather small corridors. However, for small choices of ε, condition (21) may not be satisfied – and it is not possible to provide general conditions as we would need to impose conditions

[9] For simplicity, we only allow for intervals respectively products of intervals as domains.

on $\bar{\mu}_{t_i}$, which are given exogenously.[10] Furthermore, for small choices of ε, the probabilities in Equation (21) may become very small; this could lead to problems when determining these quantities numerically by Monte Carlo simulations.

We resolve these problems by considering a more general distance function in (21). In particular, we relax the condition of the simulation outcomes $\hat{\mu}_{t_i}$ being uniformly close to the "target curves" $\bar{\mu}_{t_i}$ by considering the integrated relative distance between simulated curve and target curve:

$$d(\hat{\mu}_{t_i}, \bar{\mu}_{t_i}) = \frac{1}{d_i^0} \int_{20}^{140} \int_{t_i}^{140-x_0} \frac{|\hat{\mu}_{t_i}(T, x_0) - \bar{\mu}_{t_i}(T, x_0)|}{\bar{\mu}_{t_i}(T, x_0)} \, dT \, dx_0$$

where:

$$d_i^0 = \int_{20}^{140} \int_{t_i}^{140-x_0} \frac{|\hat{\mu}_0(T, x_0) - \bar{\mu}_{t_i}(T, x_0)|}{\bar{\mu}_{t_i}(T, x_0)} \, dT \, dx_0.$$

Hence, if a target curve strongly deviates from the initial curve, we allow for a larger deviation (in absolute value) of the simulation outcomes than if a target curve is already fairly close to the starting curve. In this setting, the convergence of our algorithm is not guaranteed by Proposition 1 anymore, but we found that meaningful values for the parameter vector c can still be obtained.

In order to include several target planes simultaneously, for a fixed choice of ε,[11] for each simulation we count the number of times the simulated curve is "close" to a target curve and sum up the results of each simulation. In principle, the larger this sum is, the better should the model with the corresponding parameter vector c fit the data. Thus, this sum is the likelihood expression we maximize.

For the realization of this calibration procedure, we use Monte Carlo simulations: Based on an iterated procedure with the range and the grid of possible values for the components of c becoming smaller with each step, we repeatedly simulate the development of the forward force of mortality for c and apply the criterion specified above. The calibration procedure has been implemented in C++ following the algorithm for the simulation of HJM-models proposed by Glasserman (2004, Section 3.6) where the time steps were set to a quarter year. For T and x_0, we deploy yearly approximations of the continuous forward force of mortality.

Due to the long time lag between two consecutive "data points" (generation tables) and the specification of the distance function, the calibration of the short-term effect based on the presented procedure is problematic. Hence, we use an approximative refinement approach based on annually available period mortality data. Under the assumption that the deviations within the period mortality data are close to alike, i.e., that the underlying populations are sufficiently similar, this is achieved by re-calibrating the short-term parameter c_2, on the basis of the corresponding spot force model. Proposition 2.1.2 in Bauer (2008) and a yearly Euler discretization yield the following approximate distributional relation:

$$\frac{\tilde{\mu}_{t+1}^s(x_0) - \tilde{\mu}_t^s(x_0) - \frac{\partial}{\partial T}\hat{\mu}_t(t, x_0)}{\sqrt{\sigma_1(t, t, x_0)^2 + \cdots + \sigma_6(t, t, x_0)^2}} \sim \mathcal{N}(0, 1).$$

Given values for all other parameters from the Monte Carlo algorithm, an estimate for c_2 can be obtained for each x_0 by matching the second moment. Finally, averaging these values for c_2 yields our overall estimate. In order to carry out this procedure, it is necessary to determine values for $\frac{\partial}{\partial T}\hat{\mu}_t(t, x_0)$. To keep the inconsistency in considering two data-sets as small as possible, we first derive a mortality trend from the spot force data. Subsequently, based on

[10]This question is related to so-called *consistency problems* as the target plane needs to be consistent with the model. See Björk (2003), Filipović (2001) or references therein for detailed discussions of consistency problems for forward interest rate models or Bauer (2008) for an application of their ideas to mortality modeling.

[11]In the calibration process, we tested several values for ε in order to confirm the stability of the results.

this trend, we determine a Gompertz fit $\exp\{\hat{a}(x_0+T)+\hat{b}\}$ for the mortality intensity for each initial age x_0 by exponential regression, and set:

$$\frac{\partial}{\partial T}\hat{\mu}_t(T,x_0)\bigg|_{T=t} = \frac{\partial}{\partial T}e^{\hat{a}(x_0+T)+\hat{b}}\bigg|_{T=t} = \hat{a}\, e^{\hat{a}(x_0+t)+\hat{b}}.$$

VI. Data and Calibration Results

For a practical application of the model, of course, the choice of data should cohere with the application in view, i.e., the considered population for the application should be similar to the population underlying the data used for the calibration procedures. If generation tables for different points in time for the considered population are readily available, these could be used for determining adequate parameterizations for the volatility structure. But as already mentioned earlier in the text, generation life tables are not frequently published and may not even exist for the population under consideration.

However, in contrast to the compilation of life tables themselves, it is not necessary to assess mortality rates or even trends for the calibration of the volatility, but solely fluctuations around the trends. If mortality rates differ for certain populations, this may not be the case for the respective trends, and, even if the trends differ, this may not be the case for the volatility. For example, even though it is well-known that mortality rates differ considerably between the general population and the insured population, this is not the case for mortality improvements: Even though mortality improvements were somewhat higher for insureds in the recent past, the deviation was rather slim; particularly, for the incorporation of mortality trends into the *Valuation Basic Table 2001* for the insured population, the *American Academy of Actuaries* reports that they relied on the observed improvements for the general population (cf. American Academy of Actuaries (2002)). Therefore, assuming a similar structure for deviations from this trend for the respective populations should also be adequate. An additional problem with the use of generation life tables is that they are often compiled in an unsophisticated manner. Hence, the best estimate mortality rates denoted in a generation life table, which serve as the basis for our approach, may not be very accurate.

In this article, we omit analyzing trends, their deviations for different populations, or the quality of underlying forecasts, but for illustrating our ideas we will rely on the *Group Annuity Mortality tables* (GAM tables) published by the *Society of Actuaries*, which are mainly employed for the calculation of reserves in U.S. companies' pension schemes. The reasons for our choice are simple: On the one hand, they have been published fairly regularly (our data-set consists of the 1951, 1971, 1983, and 1994 generation table) and, on the other hand, they are publicly available, e.g., from *Table Manager*.[12]

From each table, for some time t, approximations of the forward forces of mortality $\hat{\mu}_t$ can be determined in the following way for each x_0, t, and k such that $x_0+t+k \in \{20, 21, ..., 109\}$ at time t:

- For $k=0$:

$$\begin{aligned} E\left[p_{x_0+t}^{(t+1)}\middle|\mathcal{F}_t\right] &= e^{-\int_t^{t+1}\hat{\mu}_t(s,x_0)\,ds} \\ \Rightarrow\ -\log\left\{E\left[p_{x_0+t}^{(t+1)}\middle|\mathcal{F}_t\right]\right\} &= \int_t^{t+1}\hat{\mu}_t(s,x_0)\,ds \approx \hat{\mu}_t(t+\frac{1}{2},x_0) \end{aligned}$$

[12] *Table Manager* can be obtained from the website of the *Society of Actuaries* (*www.soa.org/professional-interests/technology/tech-table-manager.aspx*).

- For $k \geq 1$:

$$\frac{E\left[\left. {}_{k+1}p_{x_0+t}^{(t+k+1)} \right| \mathcal{F}_t\right]}{E\left[\left. {}_{k}p_{x_0+t}^{(t+k)} \right| \mathcal{F}_t\right]} = e^{-\int_{t+k}^{t+k+1} \hat{\mu}_t(s,x_0)\,ds}$$

$$\Rightarrow \; -\log\left\{\frac{E\left[\left. {}_{k+1}p_{x_0+t}^{(t+k+1)} \right| \mathcal{F}_t\right]}{E\left[\left. {}_{k}p_{x_0+t}^{(t+k)} \right| \mathcal{F}_t\right]}\right\} = \int_{t+k}^{t+k+1} \hat{\mu}_t(s,x_0)\,ds \approx \hat{\mu}_t(t+k+\frac{1}{2}, x_0).$$

It is worth noting that there is a slight inconsistency hidden in generation life tables with basically any stochastic mortality model as it is generally assumed that:

$$\frac{E\left[{}_{k}p_{x_0}^{(k)}\right]}{E\left[{}_{k-1}p_{x_0}^{(k-1)}\right]} \stackrel{!}{=} E\left[\frac{{}_{k}p_{x_0}^{(k)}}{{}_{k-1}p_{x_0}^{(k-1)}}\right] = E\left[p_{x_0+k-1}^{(k)}\right],$$

which aside from some pathological cases, implies a deterministic evolution of mortality (see Bauer (2008)). We disregard this feature for now as all considered quantities are sufficiently well-defined.

Based on the derived approximations, we may now calibrate the volatility structure of the Square-root model using the algorithm introduced in Section V. But due to the enormous computational time necessary for the procedure, we disregard the first table (1951), since the long time lag of 20 years between the first and the second table would slow down the calculations considerably. For the refinement of the short-term effect based on the approximative method, we use annual period mortality data for the general US population as available from the *Human Mortality Database*[13], assuming that the respective volatilities are similar. The resulting parameter estimates are provided in Table 4.

For the calibration in the Gaussian case (see Section V), we could employ all available data as the computation is very fast: Results are available in just a few seconds. However, when incorporating all data, the algorithm is not stable as the correlation matrix Σ is poorly conditioned and, hence, numerically singular. Even if employing a reduced data-set, the different influences interact in such a way that the resulting volatility takes values which are inconsistent with other analyses. These problems may be overcome by appropriately smoothing the data, but here the question arises which smoothing function would be consistent with our model. Another possibility would be the inclusion of more random effects, maybe even infinitely many, i.e., to use a Gaussian field rather than a finite dimensional Brownian motion as proposed in Biffis and Millossovich (2006) for spot force models. However, it is arguable whether this would be adequate and a statistical assessment of the principal components does not seem feasible given the limited amount of data. While we intend to consider these possibilities in more detail in future work, we “solve" these problems here by simply using a reduced data-set of six data points per period (as many as there are random factors). But this puts us in the situation of having to choose appropriate data points, and the resulting estimates differ considerably, particularly for choices of Δ which are clustered in the (T, x_0)-triangle of available data. However, for choices which are evenly spread among the triangle, the estimates are fairly close and seem reasonable regarding analyses in the Square-root model or analyses of the variation observable in period mortality data. In order to reduce subjectivity, we will use parameter estimations for data points which are widely spread in the triangle of available data,[14] i.e., $(T-t, x_0+T) \in \{(0,20),(0,60),(0,95),(35,60),(35,95),(75,95)\}$.

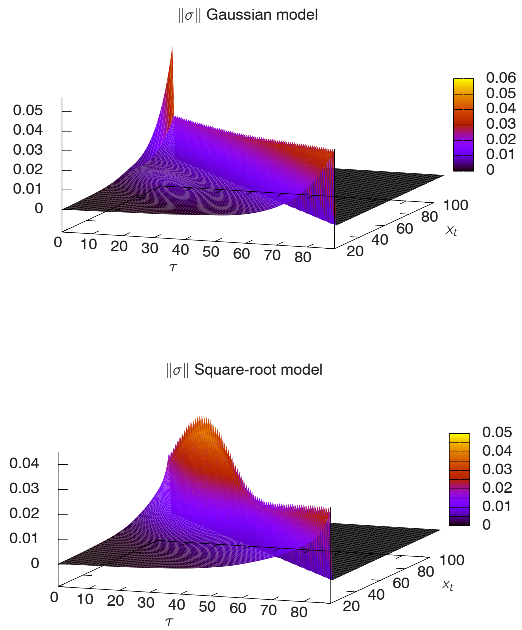
[13] *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at *www.mortality.org* or *www.humanmortality.de* (downloaded 04/11/2007).

[14] Note that, since the mortality trends incorporated in the GAM tables are equal to zero for very old ages, we consider only ages up to 95 in order to facilitate mortality changes in time.

Table 4: Resulting parameterizations

	Gaussian model	Square-root model
c_1	0.09371	0.00007
c_2	0.2358	0.00467
c_3	0.4470	0.0063
c_4	0.5343	0.0005
c_5	0.01714	0.032
c_6	0.1647	0.024

Figure 1: Volatility structure for the Gaussian and the Square-root model



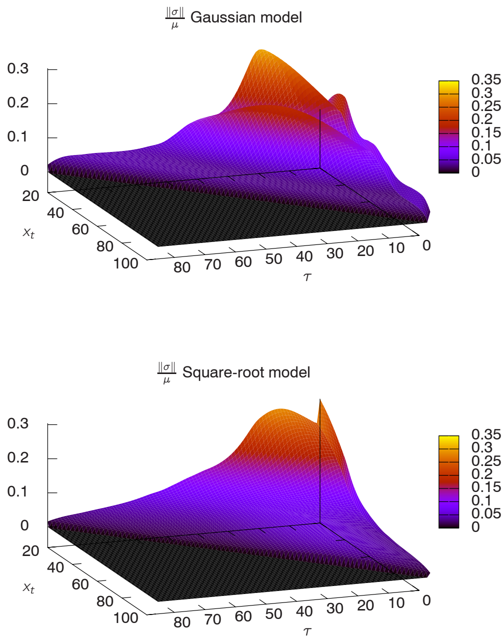
Moreover, we choose $a = 0.078$ as this maximizes the likelihood of the resulting estimate and $b = -10$. Obviously, the multiplicative term e^b is superfluous from a theoretical perspective as it is automatically incorporated in the parameter vector c, but we decided to choose a value such that the resulting estimates are in a reasonable range. The resulting parameterization is also displayed in Table 4.

Due to the differences in the structure of the models, the respective values for $c_1, \ldots, c_6$ cannot be directly compared. In order to illustrate the resulting structures, in Figure 1 the norm $\|\sigma(t, T, x_0)\|$ for $t = 20$, i.e., for the spot volatility at the 1971 table is plotted for both models, where $x_t := x_0 + t$ and $\tau = T - t$.[15]

For both models, we can clearly see the exponential trend in τ for each x_t and in x_t for each fixed τ. Furthermore, the volatilities for the same age at maturity generally increase in the time to maturity τ in both models: Each diagonal parallel to the line through $(\tau, x_t) = (0, 109)$ and

[15]Note that since our proposed volatility structure only depends on $x_0 + T$ and $T - t$, at least in the Gaussian model the spot volatility is the same for each point in time t. For the Square-root model, the local time t is only incorporated by changing values of $\sqrt{\hat{\mu}_t(T, x_0)}$.

Figure 2: Relative volatility structure for the Gaussian and the Square-root model



$(\tau, x_t) = (89, 20)$ represents volatilities for intensities with the same age at maturity and these diagonals are mostly upward sloping in τ. In contrast, the short-term effect seems to be by far more pronounced in the Gaussian model in comparison to the Square-root model, whereas the old age effect for the Square-root model seems to be considerably larger. However, these observations are somewhat misleading as different weightings are applied in both models and, due to the exponential trend, effects for smaller choices of (τ, x_t) are hardly noticeable.

Thus, in Figure 2, the relative spot volatility structure $\frac{\|\sigma(t,T,x_0)\|}{\hat{\mu}_t(T,x_0)}$ is displayed and it clearly appears that the exponential trend overshadowed several structures. Firstly, we notice that the short-term effect is now more pronounced in the Square-root model in comparison to the Gaussian model. However, this may again be a consequence of the different weights as especially for small values of x_t, the remaining $\sqrt{\hat{\mu}_t(T, x_0)}$ in the denominator of the relative spot volatilities for the Square-root case plays a dominant role. Furthermore, we observe that the young age effect is very large in both models, whereas the middle age effect is hardly noticeable for the Square-root case but quite pronounced for the Gaussian calibration. The overall relative volatility now decreases in the age at maturity $x_t + \tau$.

All in all, the general structure and the size of the volatilities seem to be fairly similar when comparing both models, but some detailed structures differ. This may in part also be a consequence of the slightly different data-set underlying the calibration as the 1951 table was not considered when determining the Square-root model parameterization. It is hard to judge which structure serves as a “better" model for the actual volatility and which calibration is “more accurate": On the one hand, due to numerical limitations, the Square-root parameters may be biased, but, on the other hand, the limited amount of data points used in the calibration procedure for the Gaussian model could also have led to skewed outcomes. Moreover, the a priori structure fixed for the different effects still plays a dominant role, too.

Despite these potential shortcomings and problems, our model allows for more sophisticated effects and complex structures to be considered than most other stochastic mortality models proposed in the literature. Moreover, due to similarities of the outcomes from the (different) calibration procedures for the two models, we believe that our results are fairly reliable.

VII. Illustrative Application of the Models

As mentioned in the introduction, forward mortality models differ from spot mortality models in that the future evolution of mortality is not forecasted, but the expected mortality structure is incorporated as the input. The stochasticity arises from the fact that these forecasts are uncertain, i.e., the realized evolution of mortality may well deviate from the expected one.

This means that, within the forward modeling approach, mortality contingent claims only depending on the expected evolution may be valued without considering the stochasticity. For example, under the assumption that an insurer is risk-neutral with respect to unsystematic mortality risk arising from finite portfolios of insureds, for the value of an immediate single premium annuity paying \$1 per year for an x_0-year old, we have (cf. Bauer and Russ (2006)):[16]

$$\begin{aligned} a_{x_0} &= E_Q\left[\sum_{k=0}^{\infty} e^{-\int_0^k r_s\,ds}\, {}_k p_{x_0}^{(k)}\right] = \sum_{k=0}^{\infty} p(0,k) E_Q\left[{}_k p_{x_0}^{(k)}\right] \\ &= \sum_{k=0}^{\infty} p(0,k)\, e^{-\int_0^k \tilde{\mu}_0(s,x_0)\,ds}, \end{aligned}$$

i.e., for immediate annuities, insurers may apply the forward mortality model similar to regular life tables. However, for annuity products with mortality contingent options such as Guaranteed Annuitization Options (GAOs) or Guaranteed Minimum Income Benefits (GMIBs) within Variable Annuities, the stochastic effects need to be considered.

Similarly, for the valuation of longevity securitization products which solely depend on the expected mortality structure, it is not necessary to take the stochasticity into account. For example, the coupon payments of the so-called *EIB Bond* announced by the European Investment Bank in 2004 approximately are of the form $C\,{}_t p_{65}^{(t)}$, $t = 1, 2, ..., 25$, where C is some nominal amount. Based on the forward mortality model, the value of each coupon payment at time zero can thus be determined by:

$$E_Q\left[e^{-\int_0^t r_s\,ds}\, C\, {}_t p_{65}^{(t)}\right] = C p(0,t) e^{-\int_0^t \tilde{\mu}_0(s,65)\,ds}.$$

There are several reasons why the EIB Bond was not successfully placed on the market (see Cairns et al. (2005a) for a discussion and a detailed description of the deal); one possible explanation is the payoff structure: *"The upfront capital may be too large compared with the risk being hedged, leaving no capital to hedge other risks"* (cf. Cairns et al. (2005a)). Other instruments and payoff structures have been proposed, for example so-called *Survivor Swaps* (see Dowd et al. (2006)) or longevity derivatives with an option-type payoff structure (see Lin and Cox (2005)), respectively.

In order to illustrate how our model may be applied to valuate such more sophisticated structures and to obtain an idea about the implications of our model regarding the risk inherent in the given estimates, we consider a security with an option-type payoff of the form:

$$C_T = \left({}_T p_{x_0}^{(T)} - K\right)^+ = \max\left\{{}_T p_{x_0}^{(T)} - K, 0\right\},$$

[16] Here r_t denotes the short rate of interest at time t.

Table 5: Comparative statistics of the longevity derivative

Model	Gaussian model		Square-root model	
	$P\left({}_Tp_{x_0}^{(T)} > K\right)$	$E\left[\left({}_Tp_{x_0}^{(T)} - K\right)^+\right]$	$P\left({}_Tp_{x_0}^{(T)} > K\right)$	$E\left[\left({}_Tp_{x_0}^{(T)} - K\right)^+\right]$
$x_0 = 65$	0.4076	0.0491	0.4214	0.0200
$x_0 = 70$	0.4021	0.0334	0.4268	0.0194
$x_0 = 75$	0.3877	0.0174	0.4190	0.0122
$x_0 = 80$	0.3567	0.0058	0.3961	0.0042
$x_0 = 85$	0.3053	0.0011	0.3670	0.0008

i.e., the longevity derivative C_T is triggered if the realized T-year survival probability exceeds some strike level:

$$K := (1+\alpha)E_P\left[{}_Tp_{x_0}^{(T)}\right] = (1+\alpha)e^{-\int_0^T \hat{\mu}_0(s,x_0)\,ds},$$

with $\alpha > 0$.

Table 5 shows the probability for the derivative to be triggered as well as the expected payoff under the physical measure P for the Gaussian and the Square-root model for $\alpha = 2.5\%$, maturity $T = 20$ and different initial ages x_0 at $t = 0$ (1994). The quantities for the Square-root model have been derived via Monte Carlo simulations (30,000 paths), whereas in the Gaussian case they can be derived analytically with a Black-type formula (see Bauer (2008)).

We find that the trigger probabilities implied by the Square-root model slightly exceed the respective values for the Gaussian model, whereas the expected payoff is considerably larger in the Gaussian case. The first observation results from the distributional differences of the models: While for Gaussian random variables the mean and the median coincide, for the distribution of the mortality intensities in the Square-root model, the mean exceeds the median (see Section IV), implying a higher probability for intensities just below the expected value, which in turn yields an increase of the probability of survival probabilities slightly above the expected value compared to the Gaussian case. Regarding the second observation, we notice that the relative difference between the expected payoffs implied by both models decreases in x_0. Thus, this deviation appears to be a consequence of the far more pronounced middle age effect in the Gaussian model (cf. Section VI). As explained earlier, it is hard to judge whether this difference is a peculiarity of the data used in the Gaussian model or due to a bias in the calibration procedure applied for the Square-root model.

In order to determine the value, i.e., the expected discounted payoff under a risk-neutral or risk-adjusted measure Q, of a longevity derivative we need to determine the dynamics of the risk-neutral forward mortality intensities $\tilde{\mu}_t$. As explained in Section II, in general, it is necessary to specify a market price of risk to determine these dynamics based on the given dynamics of the best estimate forward rates. As pointed out by Blake et al. (2006b), "*sophisticated assumptions about the dynamics of the market price of risk are pointless given that, at the time, there is just a single item of price data available for a single date* (i.e., the EIB Bond) *and even that is no longer valid.*" However, in the Gaussian model, under the assumption of a deterministic market price of risk, the volatilities of best estimate and risk-neutral forward intensities coincide (cf. Section II). Hence, in this case it is not necessary to explicitly fix the market price of risk, but it is sufficient to determine the initial configuration of the risk-neutral or risk-adjusted forward rates $\tilde{\mu}_0$.[17]

[17] Note that the market price of risk is then implied by the risk-adjustment, i.e., the two approaches are theoretically equivalent.

Table 6: Discounted expected values of the longevity derivative

	$p(0,T)\,E_P\left[{}_Tp_{x_0}^{(T)}\right]$	$p(0,T)\,E_P\left[\left({}_Tp_{x_0}^{(T)}-K\right)^+\right]$	$p(0,T)\,E_Q\left[\left({}_Tp_{x_0}^{(T)}-K\right)^+\right]$
$T=10$	0.3906	0.0083	0.0210
$T=20$	0.1006	0.0112	0.0214
$T=30$	0.0101	0.0029	0.0055

As shown by Bauer and Russ (2006) based on an idea from Lin and Cox (2005), under certain conditions, it is possible to derive the initial risk-neutral forward rates based on annuity quotes; however, instead of considering multiple annuity quotes, here, similarly to Lin and Cox (2005), we rely on an approximation by the so-called Wang transform (see Wang (2000)), the parameter λ of which can be calibrated to just a single quote. Using the price of a single premium annuity for a 65-year old from 1996, Lin and Cox (2005) obtain a parameter $\lambda = 0.1792$, and hence, we can approximate the risk-adjusted expectation of the survival probability by:

$$E_Q\left[{}_Tp_{65}^{(T)}\right] = 1 - \Phi\left(\Phi^{-1}\left(1 - E_P\left[{}_Tp_{65}^{(T)}\right]\right) - \lambda\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the Standard Normal distribution. However, it is important to note that, in general, the Wang transform is not a consistent transformation of the best estimate survival probabilities within our modeling framework (see Bauer (2008)). Nevertheless, we rely on their approach here for illustrative purposes.

Thus, based on the distorted expectations and the determined parameterization of the volatility structure for the Gaussian model, we can valuate the considered longevity derivative analytically. The respective values and the discounted expected values under the physical measure as well as the "actuarially fair" values of the payoff ${}_Tp_{x_0}^{(T)}$, $p(0,T)\,E_P\left[{}_Tp_{x_0}^{(T)}\right]$, for $x_0 = 65$, $\alpha = 2.5\%$ and different choices for T are displayed in Table 6.[18]

It appears that the value of the derivative considerably exceeds the corresponding discounted expected value under the physical measure – for all maturities, the risk premium accounts for an increase by at least 90 percent. Furthermore, for larger maturities, the derivative's value is rather large when compared to the "actuarially fair" value of an endowment payment for a 65-year old maturing after T years.[19] Therefore, from this perspective, such longevity derivatives do not seem to be attractive for insurers to hedge their exposure to longevity risk arising from their annuity business. However, clearly the annuity provider will not charge an actuarially fair value, but the annuity premium also contains a risk premium (see Bauer and Weber (2007)). Hence, when being short in the endowment payment but having a long position in the longevity derivative, the insurer would benefit from outcomes where the realized survival probability is lower than the (P-)expected value. In particular, an insurer could reduce the cost of the hedge by also assuming a "short position" in longevity, for instance, by simultaneously selling life insurance policies or by investing in a so-called Catastrophe Mortality Bonds (see, e.g., Bauer and Kramer (2007) for details on such securities).

[18]For the computation of the T-year zero coupon bond $p(0,T)$ we use the T-Year Treasury Constant Maturity Rates on 07/01/1994 as denoted in FRED (Federal Reserve Economic Data) provided by the St. Louis Federal Reserve Bank (*www.research.stlouisfed.org*).

[19]Note that, for illustrative purposes, we simply assume that the underlying population for the derivative coincides with the population of annuitants.

VIII. Summary and Conclusion

Most stochastic mortality models presented in the literature focus on the proper assessment of mortality trends. However, from an actuarial perspective, trends are already incorporated in annuity products. Hence, for actuarial applications, the central problem is not to assess trends, but rather to assess the structure of deviations from these trends. As pointed out by, among others, Dahl (2004) or Cairns et al. (2005b), forward approaches similar to Heath-Jarrow-Morton models for the term structure of interest rates may constitute an appropriate base for such models in a no-arbitrage framework. However, to the authors' knowledge, no explicit models have been proposed.

One potential reason may be that (at this point) there is no liquid market for longevity securities, and consequently, it is not possible to directly apply the same ideas as for forward interest rate models. However, we show that, when considering best estimate forward structures, i.e., forward structures implied by the real world measure, similar conditions hold as for classical forward rates implied by a risk-neutral measure. Moreover, there are interrelations between these different notions of forward rates, and in some special cases, the dynamics under different measures even coincide.

This observation allows us to build best estimate forward models which may be calibrated using best estimate expectations observed at different points in time, e.g., generation life tables. However, due to limited data availability, it does not seem feasible to directly derive a model structure from the data. Therefore, we first derived a suitable structure based on epidemiological and demographic insights. Of course, such an approach bears a certain arbitrariness and requires compromises, but we are convinced that such insights should not be neglected. While there are some structural shortcomings of our specification, we believe that it serves as a good starting point for future forward mortality models.

We presented two calibration procedures which do not depend on our specification and may also be applied for other parametric models. The first approach based on Maximum Likelihood estimation yields fast results but demands a deterministic volatility structure, whereas the second algorithm works for very general choices, but is rather time consuming as it relies on Monte Carlo simulation. When applying both procedures to our particular specifications, the resulting parameterizations yield similar volatility structures. However, there remain open issues; for example, the first algorithm does not yield reasonable values when employing "too many" data points, which is a consequence of the incoherence of the structure implied by our model and the structure of the data. Nevertheless, the resulting volatility appears adequate when comparing it to the variation within period mortality data, and the similarities between the two approaches indicate a fair quality of the results.

One of the big advantages of the framework lies in the application: Although the models seem complex at first glance, it is a natural extension of classical actuarial approaches. Basic annuity contracts or simple mortality-linked securities may be valuated without having to use the stochasticity since the expected trend is incorporated in the input. For the valuation of more complex mortality contingent claims, Monte Carlo simulation can be used and for deterministic volatilities even analytical solutions exist for the value of several payoff structures. We illustrate the model application by considering longevity securities with an option-type payoff structure.

Even though we presented two calibrated models, there are several points which should be considered in future research:

- For both calibration algorithms, there have been difficulties arising from the inconsistency of the observed data with the structures implied by the models – these could be overcome by smoothing the data appropriately. However, it is unclear which smoothing function to use, i.e., which functions, if any, are (approximately) consistent with our model.

- Another solution could be to consider more general driving factors for our models such as Lévy models or random fields. However, it is arguable whether such approaches are reasonable. Empirical investigations are necessary to assess this question.

Despite these open issues, we believe that our particular approach presents a good starting point toward practical applications of forward mortality models.

Appendix

Proof of Proposition 1

- Necessity of the condition:
  Assume the contrary, i.e.:

$$\forall c \; P_c\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) = 0.$$

  Clearly, no (real) global maximum exists.

- Sufficiency of the condition:
  We assume that $P_0\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) = 0$; if not:

$$P_0\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) = 1 > P_c\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) \forall c$$

  as g is assumed to be non-trivial.

  First we show that for:

$$\|c\| \to \infty : \; P_c\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right) \to 0.$$

  Thus, it is sufficient to show that:

$$c \longmapsto P_c\left(\|\mu_{t_1} - \bar{\mu}_{t_1}\| < \varepsilon\right)$$

  is continuous, and, hence, a non-constant function in $C_0\left(\mathbb{R}^d\right)$ which has a maximum $\hat{c}$.

  In order to keep the notation simple, we only consider the one-dimensional case, that is $d = 1$. The multidimensional case may be treated analogously.

  For fixed (T, x_0), we have:

$$\begin{aligned}
\mu_{t_1}^{(c)}(t, x_0) &= \mu_0(T, x_0) \\
&\quad + c^2 \underbrace{\int_0^{t_1} g(u, T, x_0, \mu_u(T, x_0)) \int_u^T g(u, s, x_0, \mu_u(s, x_0))\, ds\, du}_{=:A_{T,x_0}} \\
&\quad + c \underbrace{\int_0^{t_1} g(u, T, x_0, \mu_u(T, x_0)) dW_u,}_{=:B_{T,x_0}}
\end{aligned}$$

where A_{T,x_0} is a positive random variable on $\mathcal{F}$ and B_{T,x_0} is a random variable with zero mean and positive variance independent of c. Thus, we have:

$$\begin{aligned} & P\left(\left|\mu_{t_1}^{(c)}(T,x_0)-\bar{\mu}_{t_1}(T,x_0)\right|<\varepsilon\right) \\ = \quad & P\left(\frac{\bar{\mu}_{t_1}(T,x_0)-\mu_0(T,x_0)-\varepsilon}{|c|}<\right. \\ & \qquad \left.|c|\,A_{T,x_0}+sgn(c)\;B_{T,x_0}<\frac{\bar{\mu}_{t_1}(T,x_0)-\mu_0(T,x_0)+\varepsilon}{|c|}\right)\longrightarrow 0 \end{aligned}$$

as $|c|\to\infty$ and:

$$0\leq P_c\left(\|\mu_{t_1}-\bar{\mu}_{t_1}\|<\varepsilon\right)\leq P\left(\left|\mu_{t_1}^{(c)}(T,x_0)-\bar{\mu}_{t_1}(T,x_0)\right|<\varepsilon\right)\to 0\ \ (|c|\to\infty),$$

which completes the first part.

Let now $(c)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ with $\lim_{n\to\infty}=c_\infty\in\mathbb{R}$. We want to show:

$$P_{c_n}\left(\|\mu_{t_1}-\bar{\mu}_{t_1}\|<\varepsilon\right)\to P_{c_\infty}\left(\|\mu_{t_1}-\bar{\mu}_{t_1}\|<\varepsilon\right)$$

and, by the Theorem of dominated convergence, it is sufficient to note that:

$$\begin{aligned} & \left\|\mu_0(T,x_0)+c_n^2\,A_{T,x_0}(\omega)+c_n\,B_{T,x_0}(\omega)-\bar{\mu}_{t_1}(T,x_0)\right\| \\ \longrightarrow\; & \left\|\mu_0(T,x_0)+c_\infty^2\,A_{T,x_0}(\omega)+c_\infty\,B_{T,x_0}(\omega)-\bar{\mu}_{t_1}(T,x_0)\right\| \end{aligned}$$

for almost every ω being a continuous function as a composition of continuous functions.

$\square$

References

American Academy of Actuaries (2002). "Final Report of the American Academy of Actuaries Commissioners' Standard Ordinary Task Force," Presented in Draft to the National Association of Insurance Commissioners' Life and Health Actuarial Task Force, Philadelphia, PA.

Angelini, F., and S. Herzel (2005). "Consistent Calibration of HJM Models to CAP Implied Volatilities," *Journal of Futures Markets* 25: 1093-1120.

Bauer, D. (2008). "Stochastic Mortality Modeling and Securitization of Mortality Risk," Ulm: ifa-Verlag.

Bauer, D., and F. Kramer (2007). "Risk and Valuation of Mortality Contingent Catastrophe Bonds," Working Paper, Georgia State University and Ulm University, [www.mortalityrisk.org].

Bauer, D., and J. Russ (2006). "Pricing Longevity Bonds using Implied Survival Probabilities," Working Paper, Ulm University and Institut für Finanz- und Aktuarwissenschaften, [www.mortalityrisk.org].

Bauer, D., and F. Weber (2007). "Assessing Investment and Longevity Risks within Immediate Annuities," *Asia-Pacific Journal of Risk and Insurance* 3: 90-112.

Biffis, E., M. Denuit, and P. Devolder (2005). "Stochastic Mortality under Measure Change," Working Paper, CASS Business School and Université Catholique de Louvain.

Biffis, E., and P. Millossovich (2006). "A Bidimensional Approach to Mortality Risk," *Decisions in Economics and Finance* 29: 71-94.

Bingham, N., and R. Kiesel (2003). *Risk-Neutral Valuation*, London: Springer Finance.

Birkel, T. (1988). “The Invariance Principle for Associated Processes,” *Stochastic Processes and their Applications* 27: 57-71.
Björk, T. (2003). “On the Geometry of Interest Rate Models,” *Lecture Notes in Mathematics*, Berlin: Springer.
Blake, D., A.J. Cairns, and K. Dowd (2006a). “Living with Mortality: Longevity Bonds and other Mortality-Linked Securities,” Presented to the Faculty of Actuaries, 16 January 2006.
Blake, D., A.J. Cairns, K. Dowd, and R. MacMinn (2006b). “Longevity Bonds: Financial Engineering, Valuation and Hedging,” *The Journal of Risk and Insurance* 73: 647-672.
Boleslawski, L., and E. Tabeau (2001). “Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80,” in E. Tabeau, A. van den Berg Jeths, and C. Heathcote (eds.), *Forecasting Mortality in Developed Countries*. Dordrecht: Kluwer Academic Publishers, 127-157.
Bowers, N.L., H.U. Gerber, J.C. Hickman, D.A. Jones, and C.J. Nesbitt (1997). *Actuarial Mathematics*, Schaumburg, Illinois: The Society of Actuaries.
Cairns, A.J., D. Blake, P. Dawson, and K. Dowd (2005a). “Pricing the Risk on Longevity Bonds,” *Life and Pensions*, October 2005: 41-44.
Cairns, A.J., D. Blake, and K. Dowd (2005b). “Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk,” *ASTIN Bulletin* 36: 79-120.
Christensen, K., and J. Vaupel (1996). “Determinants of Longevity: Genetic, Environmental and Medical Factors,” *Journal of Internal Medicine* 240: 333-341.
Cowley, A., and J.D. Cummins (2005). “Securitization of Life Insurance Assets and Liabilities,” *The Journal of Risk and Insurance* 72: 193-226.
Currie, I.D., M. Durban, and P.H. Eilers (2004). “Smoothing and forecasting mortality rates,” *Statistical Modelling* 4: 279-298.
Dahl, M. (2004). “Stochastic Mortality in Life Insurance: Market Reserves and Mortality-Linked Insurance Contracts,” *Insurance: Mathematics and Economics* 35: 113-136.
Dowd, K., D. Blake, A.J. Cairns, and P. Dawson (2006). “Survivor Swaps,” *The Journal of Risk and Insurance* 73: 1-17.
Filipović, D. (2001). “Consistency Problems for Heath Jarrow Morton Interest Rate Models,” *Lecture Notes in Mathematics* 1760, Berlin: Springer.
Fries J., L. Green, and S. Levine (1989). “Health Promotion and the Compression of Morbidity,” *The Lancet* i: 481-483.
Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, New York: Springer.
Lee, R.D, and L.R. Carter (1992). “Modeling and forecasting US mortality,” *Journal of the American Statistical Association* 87: 659-675.
Lin, Y., and S. Cox (2005). “Securitization of Mortality Risks in Life Annuities,” *The Journal of Risk and Insurance* 72: 227-252.
Macdonald, A.S., D. Bartlett, C. Berman, C. Daykin, D. Grimshaw, and P. Savill (2003). “Mortality improvements and the cohort effect,” *CMM Working Papers 1 and 2*, [www.sias.org.uk].
Manton, K., E. Stallard, and H. Tolley (1991). “Limits to Human Life Expectancy: Evidence, Prospects and Implications,” *Population and Development Review* 17: 603-637.
Milevsky, M.A., and S.D. Promislow (2001). “Mortality Derivatives and the Option to Annuitize,” *Insurance: Mathematics and Economics* 29: 299-318.
Miltersen, K.R., and S.A. Persson (2005). “Is mortality dead? Stochastic force of mortality determined by no arbitrage,” Working Paper, University of Bergen.
Nelder, J.A., and R. Mead (1965). “A simplex method for function minimization,” *Computer Journal* 7: 308-315.
Nelson, C.R., and A.F. Siegel (1987). “Parsimonious modelling of yield curves,” *Journal of Business* 60: 473-489.
Newman, C.M., and A.L. Wright (1981). “An Invariance Principle for Certain Dependent Sequences,” *The Annals of Probability* 9: 671-675.
Notzon, F., Y. Komarov, S. Ermakov, C. Sempos, J. Marks, and E. Sempos (1998). “Causes of Declining Life Expectancy in Russia,” *Journal of the American Medical Association* 279: 793-800.

Oeppen, J., and J.W. Vaupel (2002). “Broken Limits to Life Expectancy,” *Science* 269: 1029-1031.

Olshansky, S., B. Carnes, and C. Cassel (1990). “In Search of Methuselah: Estimating the Upper Limits of Human Longevity,” *Science* 250: 634-640.

Rebonato, R. (1997). *Interest Rate Option Models*, Chichester: Wiley Series in Financial Engineering.

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

Tabeau, E., A. van den Berg Jeths, and C. Heathcote (2001). “Towards an Integration of the Statistical, Demographic and Epidemiological Perspectives in Forecasting Mortality,” in E. Tabeau, A. van den Berg Jeths, and C. Heathcote (eds.), *Forecasting Mortality in Developed Countries*. Dordrecht: Kluwer Academic Publishers, 281-303.

The GSL Team (2007). *GNU Scientific Library – Reference Manual*, Edition 1.0, for GSL Version 1.9, 20 February 2007.

Van den Berg Jeths, A., R. Hoogenveen, G. De Hollander, and E. Tabeau (2001). “A Review of Epidemiological Approaches to Forecasting Mortality and Morbidity,” in E. Tabeau, A. van den Berg Jeths, and C. Heathcote (eds.), *Forecasting Mortality in Developed Countries*. Dordrecht: Kluwer Academic Publishers, 33-56.

Vaupel, J. (1986). “How Change in Age-specific Mortality Affects Life Expectancy,” *Population Studies* 40: 147-157.

Wang, S.S. (2000). “A Class of Distortion Operations for Pricing Financial and Insurance Risks,” *The Journal of Risk and Insurance* 67: 15-36.

Weiland, S., K. Rapp, J. Klenk, and U. Keil (2006). “Zunahme der Lebenserwartung,” *Deutsches Ärzteblatt* 103: 1072-1077.

Wiesner, G. (2001). “Der Lebensverlängerungsprozess in Deutschland,” *Beiträge zur Gesundheitsberichterstattung des Bundes*, Berlin: Robert Koch-Institut.