Modeling the Forward Surface of Mortality*

Daniel Bauer†
Department of Risk Management and Insurance, Georgia State University
35 Broad Street, 11th Floor; Atlanta, GA 30303; USA
Email: dbauer@gsu.edu

Fred Espen Benth
Centre of Mathematics for Applications, University of Oslo
PO Box 1053 Blindern; N-0316 Oslo; Norway
Email: fredb@math.uio.no

Rüdiger Kiesel
Chair for Energy Trading and Financial Services, University of Duisburg-Essen
Universitätsstraße 12; 45141 Essen; Germany
Email: ruediger.kiesel@uni-due.de

January 2010

Abstract

Longevity risk constitutes an important risk factor for insurance companies and pension plans. For its analysis, but also for evaluating mortality-contingent structured financial products, modeling approaches allowing for uncertainties in mortality projections are needed.

One model class that has attracted interest in applied research as well as among practitioners are *forward mortality models*, which are defined based on forecasts of survival probabilities as can be found in generation life tables and infer dynamics on the entire age/term-structure – or forward surface – of mortality. However, thus far, there has been little guidance on identifying suitable specifications and their properties.

The current paper provides a detailed analysis of forward mortality models driven by a finite-dimensional Brownian motion. In particular, after discussing basic properties, we present an infinite-dimensional formulation, and we examine the existence of finite-dimensional realizations for time-homogenous Gaussian forward models, which are shown to possess important advantages for practical applications.

Keywords: Stochastic mortality, HJM-framework, Musiela parametrization, translation semigroups, finite-dimensional realization.

*An earlier version of the paper was entitled "An Arbitrage-Free Family of Longevity Bonds", and parts of this paper are taken from the first author's doctoral dissertation (cf. Bauer (2008)). The authors are grateful for helpful comments from participants at the 2009 CMA workshop on Insurance mathematics and longevity risk, the 2009 DGVFM Scientific Conference, the Fifth International Longevity Risk and Capital Markets Solutions Conference, the ISBIS-2008 Conference, the LSE Risk and Stochastics Day 2009 as well as from seminar participants at Heriot-Watt University, Princeton University, the Georgia Institute of Technology, and Georgia State University. In particular, we are indebted to René Carmona, Ragnar Norberg, Rocío Sotomayor, and Ajay Subramanian for valuable input. All remaining errors are ours.

†Corresponding author. Phone: +1 (404)-413-7490. Fax: +1 (404)-413-7499.

1 Introduction

Standard mathematical theory for the actuarial analysis and valuation of life contingencies assumes that future survival probabilities are known. This assumption, however, is inconsistent with demographic research, which documents high levels of uncertainty in mortality projections (see e.g. Booth (2006)). Clearly, the possibility that future mortality trends deviate from current forecasts induces a systematic type of mortality risk for insurance companies and pension plans. For its analysis, but also for assessing and pricing structured financial products depending on the evolution of mortality, modeling approaches allowing for uncertainties in mortality projections are needed.[1]

One model class that is particularly suited for such an analysis and, therefore, has attracted interest in applied research as well as among practitioners are forward mortality models (see e.g. Bauer et al. (2010), Dawson et al. (2010), or Duchassing and Suter (2009)). However, these models entail a high degree of complexity because they simultaneously capture the evolution of the entire age/term-structure – or forward surface – of mortality, and, thus far, there has been little guidance on identifying tractable specifications with desirable properties.

The current paper closes this gap in literature by presenting the first thorough theoretical investigation of forward mortality models driven by a finite-dimensional Brownian motion. More specifically, after introducing relevant definitions and discussing elementary properties, we demonstrate that while the quantities in view generally depend on the underlying probability measure, Gaussian models stand out as here the (forward) model structure is preserved under changes of measure with a deterministic choice of the market price of risk process. Hence, Gaussian models present a particularly expedient subclass since pricing models can be developed from the best-estimate model by simply adjusting the initial mortality surface. To identify tractable specifications of time-homogenous Gaussian models, we provide a necessary condition on the volatility structure for the existence of finite-dimensional realizations, even though – depending on the underlying function space – the volatility does not uniquely determine the representation. At the technical level, our paper presents the first infinite-dimensional formulation of forward mortality models, which reveals several novel features of this model class. These features are illustrated based on a detailed example with an affine specification towards the end of the paper.

Some of the basic results are derived by exploiting analogies to modeling the term structure of interest rates, which allows for an expedited treatment (see e.g. Björk (1999) for a comprehensive and accessible account). For the more advanced discussion, we also partially draw on similarities to yield curve modeling, where we especially rely on ideas from Björk and Gombani (1999), Björk and Svensson (2001), and Filipović (2001). However, there are some important distinguishing characteristics of mortality modeling, which lead to key differences in the structure of our framework as well as the in results of our analysis. More precisely, with age entering the model as an additional dimension, we are considering surfaces rather than curves, which require different function spaces. Within the dynamics of these surfaces, "new generations" enter the system at every point in time and, depending on the considered function space, the mortality trend for these newborns is not necessarily captured by past surfaces. For instance, this is the case if we

[1] Such *mortality derivatives* have been issued recently as potential instruments for diversifying this type of risk (see Blake et al. (2008) for an overview).

allow for "kinks" in the age/term-structure of mortality. As a consequence, the semigroup occurring in the infinite-dimensional equation that describes the evolution of mortality – in contrast to interest rate modeling – may not be uniquely determined but, in addition to the volatility structure, can be a degree of freedom chosen by the modeler. We show that this choice has direct implications for the existence and form of finite-dimensional realizations.

Aside from their theoretical interest, our results can find immediate practical applications. In view of recent developments regarding the solvency regulation of life insurance companies, a particularly relevant and timely application is *Asset-Liability management*. More specifically, within so-called *internal* modeling approaches for determining solvency capital, insurers face the problem of how to stochastically forecast their liability side, i.e. how to apply risk factors to the life table underlying their reserve calculations. Our results not only provide guidance on how to carry out this extrapolation consistently, but they imply the explicit functional form for incorporating finite-dimensional Normal distributed risk factors (see the Conclusion for more details and other potential applications).

The remainder of the paper is organized as follows: In Section 2, we introduce all relevant definitions and analyze basic properties of different modeling approaches; Section 3 presents the infinite-dimensional formulation of forward mortality models; the existence of finite-dimensional realizations for time-homogenous Gaussian forward models is addressed in Section 4; and a detailed example illustrating our ideas is provided in Section 5. Finally, Section 6 concludes and provides an outlook on future research.

2 Stochastic Mortality Modeling

2.1 Stochastic Time of Death

Similar to classical actuarial theory, we begin by considering the *time of death* or *future lifetime* τ_x of an x-year old individual. However, in contrast to classical theory, instead of modeling τ_x as a random variable with known distribution, we interpret τ_x as a totally inaccessible stopping time on the filtered probability space $\left(\Omega, \mathcal{H}, \mathbf{H} = (\mathcal{H}_t)_{0 \le t < T^*}, \mathbb{P}\right)$, where the filtration $\mathbf{H}$ models the information flow and satisfies the usual conditions, and $T^* \in [0, \infty) \cup \{\infty\}$ is the time horizon. More precisely, following Lando (1998), it is convenient to assume that we are given an $\mathbf{H}$-adapted d-dimensional RCLL semi-martingale $Z = (Z_t)_{0 \le t < T^*}$ as well as a positive continuous function $\mu : [0, \infty) \times \mathbb{R}^d \to [0, \infty)$. Then τ_x is defined as the first jump time of a Cox process with intensity $(\mu(x + t, Z_t))$, i.e.

$$\tau_x = \inf\left\{ t > 0 : \int_0^t \underbrace{\mu(x+s, Z_s)}_{=:\mu_s(x)} \, ds \ge E \right\}, \tag{1}$$

where $E \sim \mathrm{Exp}(1)$ is a unit-Exponential random variable. Similar frameworks for stochastic mortality have been considered by, among others, Biffis (2005), Dahl (2004), or Miltersen and Persson (2005); for a more general setup, we refer to Biffis et al. (2010).

We now let the subfiltrations $\mathbf{F} = (\mathcal{F}_t)_{0 \le t < T^*}$ and $\mathbf{G} = (\mathcal{G}_t)_{0 \le t < T^*}$ be given as the augmentations of the filtrations generated by Z and $\left(\mathbf{1}_{\{\tau_x \le t\}}\right)_{0 \le t < T^*}$, respectively, and write $\mathcal{F}_{T^*-} = \mathcal{F}$. As in Lando (1998),

p. 104, we obtain

$$_{T-t}P_{x+t}(T) := \mathbb{E}^{\mathbb{P}}\left[\mathbf{1}_{\{\tau_x > T\}}\middle|\,\mathcal{F}_T \vee \{\tau_x > t\}\right] = \exp\left\{-\int_t^T \mu_s(x)\,ds\right\}$$

for the *realized* $(T-t)$-year survival probability for an $(x+t)$-year old at time t. Note that in comparison to standard actuarial notation, we write "capital" P to indicate that we are dealing with a random variable and $\cdot(T)$ to specify the time interval. Moreover, we define (forward) survival probabilities

$$_{T_2-T_1}p_{x+T_1}(t;T_2) := \mathbb{E}^{\mathbb{P}}\left[{}_{T_2-T_1}P_{x+T_1}(T_2)\middle|\,\mathcal{F}_t\right] = \mathbb{E}^{\mathbb{P}}\left[\exp\left\{-\int_{T_1}^{T_2}\mu_s(x)\,ds\right\}\middle|\,\mathcal{F}_t\right],\ \ T_1 \leq T_2. \tag{2}$$

Within most actuarial applications, due to the usually large number of individuals, the risk associated with a single individual's death is negligible. Hence, for these applications, but also for the financial analysis of index-linked mortality derivatives, it is sufficient to consider the uncertainty related to the realized survival probabilities ${}_TP_x(T)$, and to disregard the idiosyncratic component. In doing so, we may base our modeling considerations on the family of stochastic processes (2), which could be given in form of survival probabilities in a continuous sequence of traditional generation life tables. In particular, as already pointed out by Milevsky and Promislow (2001), the "traditional rates used by actuaries" really are *forward rates*, so such an approach can be viewed as the natural extension of traditional actuarial theory.

2.2 An Alternative Approach to Modeling Stochastic Mortality

We initially restrict ourselves to the consideration of survival probabilities as seen from time zero, i.e. we consider the family of processes $({}_Tp_x(t;T))_{0\leq t\leq T}$ for $T > 0$ and a selection of *generations* $x \in \mathcal{X}_0 \subseteq [0,\infty)$. Akin to classical theory, based on the given survival probabilities, we define the *force of mortality*:

Definition 2.1. *For $x \in \mathcal{X}_0$, $t \leq T < T^*$, we define the (forward) force of mortality for time T as from time t by*

$$\mu_t(T,x) = -\frac{\partial}{\partial T}\log\left\{{}_Tp_x(t;T)\right\}$$

and the spot force of mortality at time t by

$$\mu_t(x) = \mu_t(t,x).$$

It is easy to verify that $\mu_t(x)$ is well-defined, i.e. that the redefinition here does not add ambiguity with regards to Equation (1).

A first consequence of this definition is that different model classes arise automatically in this setup: To construct a stochastic mortality model, we may specify dynamics of the family of all possible survival probabilities, specify the dynamics of all spot forces, or specify the dynamics of all possible forward forces. In what follows, we will focus on models driven by a finite-dimensional Brownian motion although some of the results remain valid for a more general class of driving processes (see e.g. Eberlein and Raible (1999) in the case of interest rate modeling). We consider the following systems of generic dynamics for $x \in \mathcal{X}_0$:

$$
\begin{aligned}
d_T p_x(t;T) &= {}_T p_x(t;T)\left(m(t,T,x)\,dt + v(t,T,x)\,dW_t\right), \ 0 \le t \le T < T^*, \ {}_T p_x(0;T) > 0; &(3)\\
d\mu_t(x) &= a(t,x)\,dt + b(t,x)\,dW_t, \ 0 \le t < T^*, \ \mu_0(x) > 0; \text{ and} &(4)\\
d\mu_t(T,x) &= \alpha(t,T,x)\,dt + \sigma(t,T,x)\,dW_t, \ 0 \le t \le T < T^*, \ \mu_0(T,x) > 0. &(5)
\end{aligned}
$$

Here, W is a d-dimensional standard Brownian motion, and we assume that all processes $(m(t,T,x))_{0\le t\le T}$, $(b(t,x))_{0\le t<T^*}$ etc. are adapted, of adequate dimension, and satisfy conditions such that a unique strong solution exists to every stochastic differential equation (SDE). Moreover, we assume that for almost every sample outcome $\omega \in \Omega$, all objects $m(t,T,x)$, $v(t,T,x)$, $\alpha(t,T,x)$, and $\sigma(t,T,x)$ are continuously differentiable in T.

Similar to Proposition 20.5 in Björk (1999) for interest rate models, we may now assess the relationship among the different modeling approaches:

Proposition 2.1. *Let $x \in \mathcal{X}_0$, $0 \le t \le T < T^*$.*

1. *If the ${}_T p_x(t;T)$ satisfy (3), then we have (5) for the forward force dynamics where*[2]

$$
\begin{aligned}
\alpha(t,T,x) &= -m_T(t,T,x) + v_T(t,T,x) \times v(t,T,x)',\\
\sigma(t,T,x) &= -v_T(t,T,x).
\end{aligned}
$$

2. *If the $\mu_t(T,x)$ satisfy (5), then we have (4) for the spot force dynamics where*

$$
\begin{aligned}
a(t,x) &= \partial/\partial T\, \mu_t(T,x)|_{T=t} + \sigma(t,t,x),\\
b(t,x) &= \sigma(t,t,x).
\end{aligned}
$$

3. *If the $\mu_t(T,x)$ satisfy (5), then*

$$
d_T p_x(t;T) = {}_T p_x(t;T)\left(\left(A(t,T,x) + 1/2\,\|S(t,T,x)\|^2\right)\,dt + S(t,T,x)\,dW_t\right),
$$

where $A(t,T,x) = -\int_t^T \alpha(t,s,x)\,ds$ and $S(t,T,x) = -\int_t^T \sigma(t,s,x)\,ds$.

The proof is analogous to that of Proposition 20.5 in Björk (1999).

It is important to note that we have not required that the dynamics (3), (4), and (5) are specified under a certain measure. In fact, if W is a Brownian motion under the underlying measure $\mathbb{P}$, by the martingale property of ${}_T p_x(t;T)$ (cf. Eq. (2)), the drift will be fixed implicitly:

Proposition 2.2. *If W is a Brownian motion under $\mathbb{P}$, then we have $m(t,T,x) \equiv 0$, $T > 0$, $x \in \mathcal{X}_0$, $0 \le t \le T$ in (3) and*

$$
\alpha(t,T,x) = \sigma(t,T,x) \times \int_t^T \sigma(t,s,x)'\,ds, \ T > 0, \ x \in \mathcal{X}_0, \ 0 \le t \le T \tag{6}
$$

[2] As usual, if no ambiguity arises, we denote partial derivatives by subscripts.

in (5) almost surely.

Proof. Let $T > 0$, $x \in \mathcal{X}_0$. From the martingale property of ${}_Tp_x(t;T)$, we immediately obtain

$$m(t,T,x) \equiv 0,\ 0 \le t \le T.$$

Proposition 2.1.3 then yields

$$A(t,T,x) = -1/2\ \|S(t,T,x)\|^2,$$

and the second claim follows by differentiating with respect to T. □

The latter drift condition is similar to the well-known *Heath-Jarrow-Morton* (HJM) drift condition for forward interest rate models (cf. Heath et al. (1992)), and – for slightly different model setups – has been ascertained by Cairns et al. (2006) and Miltersen and Persson (2005).

From Definition 2.1, we obtain for $x \in \mathcal{X}_0,\ T > 0$, that

$${}_Tp_x(t;T) = \exp\left\{-\int_0^T \mu_t(s,x)\,ds\right\},\ t \le T,$$

and, therefore,

$${}_{T-t}p_{x+t}(t;T) = \frac{{}_Tp_x(t;T)}{{}_tP_x(t)} = \exp\left\{-\int_t^T \mu_t(s,x)\right\}.$$

This means that we could have alternatively defined the (forward) force of mortality as

$$\mu_t(T,x) = -\frac{\partial}{\partial T}\log\left\{{}_{T-t}p_{x+t}(t;T)\right\} = -\frac{\partial}{\partial T}\log\left\{{}_{T-t}p_{x_t}(t;T)\right\}$$

for $x_t = x + t$. In contrast to the initial Definition 2.1, the latter remains meaningful for $x < 0$ as long as $x_t \ge 0 \Leftrightarrow x \ge -t$, i.e. for $x \in \mathcal{X}_t \subseteq [-t,\infty)$. Whence, in what follows, we will adopt this generalized definition, and consider forward mortality models of the form

$$d\mu_t(T,x) = \alpha(t,T,x)\,dt + \sigma(t,T,x)\,dW_t,\ t_0 \le t \le T,\ \mu_{t_0}(T,x) > 0, \tag{7}$$

for $0 \le t_0 \le T < T^*$, $x \in \mathcal{X}_{t_0}$.[3]

However, in contrast to classical actuarial theory, in general we have

$${}_{T_2-T_1}p_{x+T_1}(t;T_2) \ne \exp\left\{-\int_{T_1}^{T_2} \mu_t(s,x)\,ds\right\}$$

for $t < T_1 \le T_2$. In fact, the following proposition shows that equality necessarily yields a deterministic evolution of mortality:

[3]Clearly, we can choose $x \in \mathcal{X}_t$ in Propositions 2.1 and 2.2 when relying on this generalized definition.

Proposition 2.3. *If for* $0 \leq t < T_1 \leq T_2 < T^*$ *and* $x \in \mathcal{X}_t$*,*

$$_{T_2-T_1}p_{x+T_1}(t;T_2) = \exp\left\{-\int_{T_1}^{T_2} \mu_t(s,x)\,ds\right\},$$

then we have $\sigma(t,T,x) = 0\ \forall\, T \geq t \geq 0$.

Proof. Let $0 \leq t < T_1 \leq T_2 < T^*$ and $x \in \mathcal{X}_0$. From Equation (2) we see that $({}_{T_2-T_1}p_{x+T_1}(t;T_2))$ is a martingale. For the right-hand side of the above equation, on the other hand, we have

$$\exp\left\{-\int_{T_1}^{T_2} \mu_t(s,x)\,ds\right\} = \frac{{}_{T_2}p_x(t;T_2)}{{}_{T_1}p_x(t;T_1)} =: g(t).$$

By Itô's formula, we obtain

$$\begin{aligned} dg(t) &= g(t)\ \big((m(t,T_2,x) - m(t,T_1,x) - v(t,T_2,x) \times v(t,T_1,x)' + v(t,T_1,x) \times v(t,T_1,x)')\ dt \\ &\qquad + (v(t,T_2,x) - v(t,T_1,x))\,dW_t\big) \\ &= g(t)\ \big((v(t,T_1,x) - v(t,T_2,x))\ v(t,T_1,x)'\,dt + (v(t,T_2,x) - v(t,T_1,x))\,dW_t\big) \end{aligned}$$

if W is a $\mathbb{P}$-Brownian motion. Now, in case equality holds, $g(t)$ will be a martingale, and therefore:

$$(v(t,T_1,x) - v(t,T_2,x))\ v(t,T_1,x)' = \int_{T_1}^{T_2} \sigma(t,s,x)\,ds \times \int_t^{T_1} \sigma(t,s,x)'\,ds = 0,$$

which, by differentiating in T_2 and T_1, yields

$$\sigma(t,T_2,x) \times \sigma(t,T_1,x) = 0,$$

and the claim follows by taking the limit $T_1 \to T_2$. $\square$

2.3 Valuation of Mortality-Contingent Payoffs

As is well-known from arbitrage pricing theory, the value of a contingent claim is determined by its expected discounted payoff under some *equivalent martingale measure* $\mathbb{Q}$ (see Harrison and Kreps (1979) or Delbaen and Schachermayer (1994)), which in turn is given by its Radon-Nikodym density, say

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \exp\left\{-\int_0^t \lambda(s)'\,dW_s - 1/2\int_0^t \|\lambda(s)\|^2\,ds\right\}.$$

As pointed out by Delbaen and Schachermayer (1994), p. 463f, this technique has a long history in life insurance, too: "*The actual* [or best-estimate] *mortality table is replaced* [...] *by a table reflecting a lower mortality rate if e.g. a lump sum buying a pension is calculated. Changing probabilities is common practice in actuarial sciences*."

In order to determine the value of some life contingency or some mortality-linked security $C_T \in \mathcal{H}_T$, we hence need to determine its expected discounted value under $\mathbb{Q}$. By Girsanov's Theorem, our forward model equation will then take the form (7) with

$$\alpha(t,T,x) = \sigma(t,T,x) \times \int_t^T \sigma(t,s,x)'\, ds + \sigma(t,T,x)\, \lambda(t) \tag{8}$$

and W now is a Brownian motion under $\mathbb{Q}$, i.e. the *market price of risk process* λ enters the drift.

However, it is important to note that Equation (8) only holds for forward forces defined on the basis of $\mathbb{P}$-survival probabilities. If $\mu_t(\cdot,\cdot)$ is defined based on risk-neutral or risk-adjusted survival probabilities as e.g. implied by market prices of some mortality-linked securities or some valuation life table, i.e. if we have $\mathbb{P} = \mathbb{Q}$ in Section 2.2, then clearly $\lambda \equiv 0$ in (8) as in Proposition 2.2. In particular, the definition of the forward force depends on the probability measure, while clearly the corresponding spot forces $\mu_t(x) = \mu_t(t,x)$ coincide.

Thus, in order to specify a pricing ($\mathbb{Q}$) model given a best-estimate ($\mathbb{P}$) forward model – i.e. initial surface and the volatility structure – it is necessary to specify the market price of risk process λ and to determine risk-neutral (risk-adjusted) forward forces as well as their dynamics via the corresponding spot force model, cf. Proposition 2.1.2. An exception are Gaussian forward models with a deterministic specification of λ:[4]

Proposition 2.4. *If $t \mapsto \sigma(t,T,x)$ and $t \mapsto \lambda(t)$, $0 \le t \le T$, $x \in \mathcal{X}_t$, are sufficiently regular deterministic functions, we have*

$$\mathbb{E}^{\mathbb{Q}}\left[{}_{T-t}P_{x+t}(T)\middle|\, \mathcal{F}_t\right] = e^{-\int_t^T \int_t^s \sigma(u,s,x)\,\lambda(u)\,du\,ds}\; {}_{T-t}p_{x+t}(t;T)$$

and the volatility structures of best-estimate and risk-neutral (risk-adjusted) forward models coincide.

Proof.

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[{}_{T-t}P_{x+t}(T)\middle|\, \mathcal{F}_t\right] &= \mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_t^T \mu_s(x)\,ds}\middle|\, \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_t^T \mu_s(s,x)\,ds}\middle|\, \mathcal{F}_t\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_t^T \mu_t(s,x)+\int_t^s \alpha(u,s,x)\,du+\int_t^s \sigma(u,s,x)\,dW_u\,ds}\middle|\, \mathcal{F}_t\right] \\
&\overset{\text{Eq. (8)}}{=} \mathrm{e}^{-\int_t^T \int_t^s \sigma(u,s,x)\,\lambda(u)\,du\,ds} \times {}_{T-t}p_{x+t}(t;T),
\end{aligned}$$

where W is a Brownian motion under $\mathbb{Q}$. The second claim then follows by Itô's Lemma. □

In particular, for this class of Gaussian models, in order to specify a pricing model based on a given best-estimate forward model, due to Proposition 2.2 holding under $\mathbb{P}$ as well as under $\mathbb{Q}$, it is solely necessary to specify a risk-neutral or risk-adjusted initial mortality surface, e.g. "*by a table reflecting a lower mortality rate.*" Hence, Gaussian models are especially convenient for practical applications (see e.g. Bauer et al. (2010) for applications to the pricing of longevity derivatives).

The most obvious approach to building such models would be to replace deterministic or constant parameters in *analytic mortality laws*, i.e. parametric functions describing human mortality, by stochastic processes. However, in actuarial practice, discretized representations of the entire forward mortality surface in

[4]Clearly, Gaussian models suffer the theoretical disadvantage that mortality rates could become negative. However, since the probability is very small for suitable parameter values, this problem is usually considered negligible for practical applications (cf. Schrager (2006)).

the form of generation life tables rather than low-dimensional mortality laws are used for risk-management and pricing purposes. From an application perspective, it would nevertheless be convenient if transitions from one life table to another one could be realized by a parametric transformation parametrized by a (finite-dimensional) random vector or process. Hence, we are interested in the question when the – inherently infinite-dimensional – forward mortality model allows for a *finite-dimensional representation* or *realization*. In order to assess this problem, it proves to be conducive to formulate the forward mortality setup in terms of infinite-dimensional stochastic equations in the next section.

3 Infinite-Dimensional Formulation of Mortality Models

So far, we have considered the dynamics of the forward force of mortality separately for each cohort of x-year olds and each maturity T, i.e. we imposed the *system of stochastic differential equations* (7) for the evolution of mortality. A more coherent view is to consider each *forward surface* $(\mu_t(T,x))_{t\leq T<T^*,x\in\mathcal{X}_t}$ in its entireness as an infinite-dimensional stochastic object and model its evolution by an infinite-dimensional stochastic equation (see Da Prato and Zabczyk (1992) for a thorough treatment of stochastic equations in infinite dimensions).

However, under the current parametrization, the domain for each $\mu_t(\cdot,\cdot)$, $\{t \leq T < T^*, x \in \mathcal{X}_t\}$, depends on t. In order to obtain uniform objects over time, which can be modeled as objects in the same space, it is necessary to "standardize" the parametrization. This is similar to the so-called *Musiela parametrization* of forward interest rate models (cf. Musiela (1993)). For the remainder of this text, we will generally assume that $T^* = \infty$ and $\mathcal{X}_t = [-t,\infty)$, i.e. we choose the largest possible domains for each object. Then, a uniform parametrization can be achieved by letting $\tau = T - t \in [0,\infty)$, $x_t = x + t \in [0,\infty)$,

$$\begin{aligned}\bar{\mu}_t(\tau,x_t) &:= \mu_t(t+\tau,x_t-t) = \mu_t(T,x), &(9)\\ \bar{\alpha}_t(\tau,x_t) &:= \alpha(t,t+\tau,x_t-t) = \alpha(t,T,x), &(10)\\ \text{and } \bar{\sigma}_t(\tau,x_t) &:= \sigma(t,t+\tau,x_t-t) = \sigma(t,T,x).\end{aligned}$$

Alternatively, we could choose to express Equations (9) and (10) in terms of the "age at maturity" $x_T = x + T = x + t + \tau \in [\tau,\infty)$ rather than the "current age" x_t, but we consider our choice the more natural one (see Bauer (2008) for a discussion in the alternative setup).

3.1 The Function Space

In order to reformulate the dynamics (7), we need to choose an appropriate function space in which $\bar{\mu}_t$, $\bar{\alpha}_t$, and $\bar{\sigma}_t$ exist. Due to difficulties when defining the stochastic integral on Banach spaces or, more generally, metric spaces, we resort to finding a suitable Hilbert space $\mathfrak{H}$. In general, similarly to Carmona and Tehranchi (2006), instead of assuming a particular space, we simply impose certain conditions on our choice of $\mathfrak{H}$:

Assumption 3.1. *1. The space $\mathfrak{H}$ is a separable Hilbert space and the elements of $\mathfrak{H}$ are continuous functions*

$$f : [0,\infty)^2 \longrightarrow \mathbb{R},\ (\tau,x) \mapsto f(\tau,x).$$

2. *For every* $(\tau, x) \in [0, \infty)^2$, *the evaluation functional* $\delta_{(\tau,x)}$ *with* $\delta_{(\tau,x)}(f) = f(\tau, x)$ *is a continuous linear functional on* $\mathfrak{H}$.

3. *There exists a strongly continuous semigroup* $\{S_t\}_{t \geq 0}$ *with infinitesimal generator denoted by* A *that coincides with the translation semigroup of left shifts in the first and right shifts in the second variable for* $x \geq t \geq 0$, *i.e.*
$$(S_t\, f)\,(\tau, x) = f(\tau + t, x - t),\ 0 \leq t \leq x.$$

We obtain the following result for the infinitesimal generator:

Lemma 3.1. *We have* $A = \frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}$ *on the domain of* A, $dom(A)$. *If* $\left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\right)$ *is bounded in* $\mathfrak{H}$, *we may choose* $S_t = \exp\left\{t\,\left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\right)\right\}$. *In particular,* $\{S_t\}_{t\geq 0}$ *is uniformly continuous in this case.*

Proof. Assuming that the limits exist, we have for every $f \in \mathfrak{H}$, $\tau \geq 0$, $x > 0$:

$$\begin{aligned}
\lim_{h\downarrow 0} \frac{S_h\, f(\tau,x) - f(\tau,x)}{h} &= \lim_{h\downarrow 0} \frac{f(\tau+h, x-h) - f(\tau,x)}{h} \\
&= \lim_{h\downarrow 0} \frac{f(\tau+h, x-h) - f(\tau, x-h) + f(\tau, x-h) - f(\tau,x)}{h} \\
&= \frac{\partial +}{\partial \tau} f(\tau,x) - \frac{\partial -}{\partial x} f(\tau,x) \\
\Rightarrow \lim_{h\downarrow 0} \delta_{(\tau,x)} \left(\frac{S_h\, f - f}{h}\right) &= \delta_{(\tau,x)} \left(\frac{\partial +}{\partial \tau} f - \frac{\partial -}{\partial x} f\right),\ \tau \geq 0, x > 0.
\end{aligned}$$

Now, if $f \in dom(A)$, $\lim_{h\downarrow 0} \frac{S_h\, f - f}{h}$ exists in $\mathfrak{H}$ and, in particular, is continuous. Hence, we have equality for the functions and the one-sided derivatives are two-sided, i.e.

$$A = \frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}$$

on $dom(A)$.

In case $\left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\right)$ is bounded in $\mathfrak{H}$, by Corollary II.1.5 in Engel and Nagel (2000), we have

$$A \text{ continuous } \Leftrightarrow dom(A) = \mathfrak{H} \Leftrightarrow \{S_t\}_{t\geq 0} \text{ uniformly continuous},$$

where $S_t = \exp\{t\, A\}$. □

An example of a space satisfying Assumption 3.1 – similar to the univariate space proposed by Björk and Svensson (2001) – in which the semigroup is uniformly continuous is provided by the following proposition:

Proposition 3.1. *For* $\beta > 1$ *and* $\gamma > 0$ *define*

$$\mathfrak{H}_{\beta,\gamma} = \left\{ f : [0,\infty)^2 \longrightarrow \mathbb{R} \text{ infinitely differentiable} \middle| \langle f, f\rangle_{\beta,\gamma} < \infty \right\},$$

where $\langle \cdot, \cdot \rangle_{\beta,\gamma} : \mathfrak{H}_{\beta,\gamma} \times \mathfrak{H}_{\beta,\gamma} \to \mathbb{R}$ *with*

$$\langle f, g \rangle_{\beta,\gamma} = \sum_{n,m=0}^{\infty} \beta^{-(n+m)} \int_0^\infty \int_0^\infty \left(\frac{\partial^{n+m} f}{\partial \tau^n \partial x^m}(\tau, x) \right) \left(\frac{\partial^{n+m} g}{\partial \tau^n \partial x^m}(\tau, x) \right) e^{-\gamma\,(\tau+x)}\, d\tau\, dx.$$

Then $\left(\mathfrak{H}_{\beta,\gamma}, \langle \cdot, \cdot \rangle_{\beta,\gamma} \right)$ *is a Hilbert space satisfying Assumption 3.1.*

A proof can be found in the Appendix.

Since $\mathfrak{H}_{\beta,\gamma}$ is a space of real-analytic functions in both variables (cf. Björk and Svensson (2001)), its elements can be extended to the entire complex plane in both variables, such that the semigroup $\{S_t\}_{t \geq 0}$ is in fact uniquely determined in this case. In order to allow for some degree of modeling freedom in the choice of the semigroup, it is necessary to allow for more general spaces in which $\{S_t\}_{t \geq 0}$ is not uniformly continuous, although this requires the consideration of non-strong solutions of our model equation (14) in the following subsection.

As in Vargiolu (1999), natural candidates are the (exponentially) weighted Sobolev spaces $H^1_\gamma \left([0, \infty)^2 \right)$, $\gamma > 0$. However, due to the additional dimension, the Sobolev embedding theorem does not assert continuity in this case. Thus, we turn to Sobolev spaces with different orders of differentiability:

Proposition 3.2. *For* $\beta > 1$ *and* $\gamma > 0$ *define*

$$\begin{aligned} \tilde{\mathfrak{H}}_{\beta,\gamma} \;=\; & \Big\{ f : [0, \infty)^2 \longrightarrow \mathbb{R} \in H^1_\gamma \left([0, \infty)^2 \right) \Big|\; f \text{ continuous}, \\ & \quad f \text{ and } D^{(0,1)} f \text{ infinitely differentiable in the first variable and } [f, f]_{\beta,\gamma} < \infty \Big\} \end{aligned}$$

where $D^{(0,1)}$ *denotes the weak derivative in the second variable and* $[\cdot, \cdot]_{\beta,\gamma} : \tilde{\mathfrak{H}}_{\beta,\gamma} \times \tilde{\mathfrak{H}}_{\beta,\gamma} \to \mathbb{R}$ *with*

$$\begin{aligned} [f, g]_{\beta,\gamma} \;=\; & \sum_{n=0}^{\infty} \beta^{-n} \int_0^\infty \int_0^\infty \left(\frac{\partial^n f}{\partial \tau^n}(\tau, x) \right) \left(\frac{\partial^n g}{\partial \tau^n}(\tau, x) \right) e^{-\gamma\,(\tau+x)}\, d\tau\, dx \\ & + \sum_{n=0}^{\infty} \beta^{-(n+1)} \int_0^\infty \int_0^\infty \left(\left(\frac{\partial^n}{\partial \tau^n} D^{(0,1)} f \right)(\tau, x) \right) \left(\left(\frac{\partial^n}{\partial \tau^n} D^{(0,1)} g \right)(\tau, x) \right) e^{-\gamma\,(\tau+x)}\, d\tau\, dx. \end{aligned}$$

Then $\left(\tilde{\mathfrak{H}}_{\beta,\gamma}, [\cdot, \cdot]_{\beta,\gamma} \right)$ *is a Hilbert space satisfying Assumption 3.1.*

A proof is provided in the Appendix. In particular, we show that in this case, we may set

$$(S_t f)(\tau, x) = \begin{cases} f(\tau + t, x - t) & , x \geq t, \\ f(\tau + x, 0) & , x < t. \end{cases} \tag{11}$$

This assumption essentially entails that newborns at all times $t > 0$ are expected to enter the world with the same expectations regarding future mortality as newborns at time zero (see Sections 3.2 and 5 below).

3.2 Forward Mortality Models

In order to ensure that the newly defined mortality surfaces from (9) are elements of $\mathfrak{H}$, we make the following assumption (cf. Section 6.1 in Da Prato and Zabczyk (1992)):

Assumption 3.2. *1. The initial forward surface, $\bar{\mu}_0$, is an element of $\mathfrak{H}$, i.e. $\bar{\mu}_0 \in \mathfrak{H}$.*

2. The processes $(\bar{\alpha}_t)$ and $\left(\bar{\sigma}_t^{(i)}\right)$, $i = 1, ..., d$, are $\mathfrak{H}$-valued, predictable and for $T^{**} > 0$ satisfy

$$\mathbb{P}\left(\int_0^{T^{**}} \|\bar{\alpha}_s\|_{\mathfrak{H}} + \sum_{i=1}^{d} \left\|\bar{\sigma}_s^{(i)}\right\|_{\mathfrak{H}}^2 ds < \infty\right) = 1. \tag{12}$$

Then, under a slight abuse of notation, for $t \le T^{**}$ and $t \le x$,

$$\begin{aligned}
\delta_{(\tau,x)}\left(\bar{\mu}_t\right) &= \mu_t(t+\tau, x-t) \\
&= \mu_0(t+\tau, x-t) + \int_0^t \alpha(s, s+(t-s)+\tau, x-s-(t-s))\, ds \\
&\quad + \sum_{i=1}^{d} \int_0^t \sigma^{(i)}(s, s+(t-s)+\tau, x-s-(t-s))\, dW_s^{(i)} \\
&= \delta_{\tau,x}\left(S_t\bar{\mu}_0\right) + \int_0^t \delta_{(\tau,x)}\left(S_{t-s}\,\bar{\alpha}_s\right)\, ds + \sum_{i=1}^{d}\int_0^t \delta_{(\tau,x)}\left(S_{t-s}\,\bar{\sigma}_s^{(i)}\right)\, dW_s^{(i)} \\
&= \delta_{(\tau,x)}\left(S_t\bar{\mu}_0 + \int_0^t S_{t-s}\,\bar{\alpha}_s\, ds + \sum_{i=1}^{d}\int_0^t S_{t-s}\,\bar{\sigma}_s^{(i)}\, dW_s^{(i)}\right), \text{ a.s.}
\end{aligned}$$

By the same reasoning as for Equation (4.3) in Filipović (2001), for a fixed t, the exceptional $\mathbb{P}$-Null-set depends on (τ, x), and since both sides are continuous in (τ, x), we have equality $\mathbb{P}$-almost surely and simultaneously for all $(\tau, x) \in [0, \infty) \times [t, \infty)$. This will be satisfied exactly if

$$\bar{\mu}_t = S_t\bar{\mu}_0 + \int_0^t S_{t-s}\,\bar{\alpha}_s\, ds + \sum_{i=1}^{d}\int_0^t S_{t-s}\,\bar{\sigma}_s^{(i)}\, dW_s^{(i)} \;\mathbb{P}\text{-a.s.},\; t \in [0, T^{**}] \tag{13}$$

for $\{S_t\}_{t\ge 0}$ as in Assumption 4.1.3. By applying Proposition 6.2 from Da Prato and Zabczyk (1992), it is then clear that the right-hand side of (13) has a predictable version, so that $(\bar{\mu}_t)$ is a *mild solution* (cf. Section 6.1 in Da Prato and Zabczyk (1992)) to the infinite-dimensional SDE

$$\left\{\begin{array}{lcl} d\bar{\mu}_t &=& \left(A\,\bar{\mu}_t + \bar{\alpha}_t\right)\, dt + \sum_{i=1}^{d} \bar{\sigma}_t^{(i)}\, dW_t^{(i)}, \\ \bar{\mu}_0 &=& \mu_0(\cdot,\cdot). \end{array}\right. \tag{14}$$

It is important to note that in contrast to interest rate models, the choice of $\{S_t\}_{t\ge 0}$ or, equivalently, A (cf. Theorem II.1.4 in Engel and Nagel (2000)) may not be unique, and we will obtain different model dynamics (13) or (14) for different choices. From a modeling perspective, this choice determines how mortality evolves for future generations. As an illustration, Figure 1 displays the action of the semigroup for the case $t = 10$. For generations alive at time zero – i.e. if $x \ge t$ – the function value $(S_t f)(\tau, x) = f(\tau + t, x - t)$ is obtained via translation. For "new generations" – i.e. if $x < t$ – on the other hand, in case $\{S_t\}_{t\ge 0}$ is uniformly continuous, we have $(S_t f)(\tau, x) = (\exp\{t\,A\}f)(\tau, x)$ (see Proposition 3.1 for an example). However, if $dom(A) \ne \mathfrak{H}$, there may be some modeling discretion. For example, as

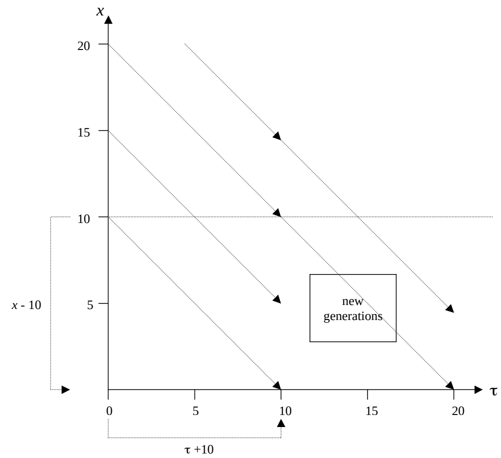

Figure 1: Action of the semigroup S_t for $t = 10$.

pointed out, an admissible choice for the space $\widetilde{\mathfrak{H}}_{\beta,\gamma}$ introduced in Proposition 3.2 is given by (11) where $(S_t f)(\tau, x) = f(\tau + x, 0)$, $x < t$. This implies that

$$f(\tau + x, 0) = (S_x\, S_{t-x}\, f)(\tau, x) = (S_{t-x} f)(\tau + x, 0),$$

so the function value for a $(\tau+x)$-year old at time $(\tau+t) > (\tau+x)$ as seen from time $(t-x) > 0$ coincides with the function value for a $(\tau + x)$-year old at time $(\tau + x)$ as seem from time zero, that is disregarding random fluctuations, future newborns enter the world with the same expectations regarding future mortality as newborns at time zero (see also Section 5).

We omit discussing the existence and uniqueness of the solution to (14) in the general case, which, e.g., could be ascertained by imposing the conditions from Theorem 7.4 in Da Prato and Zabczyk (1992). In what follows, we simply assume that there exists a continuous version of $\bar{\mu}_t$. However, we note that in the time-homogenous, deterministic case, i.e. if

$$\bar{\alpha}_t \equiv \bar{\alpha} \in \mathfrak{H} \text{ and } \bar{\sigma}_t^{(i)} \equiv \bar{\sigma}^{(i)} \in \mathfrak{H},\ t \in [0, T^{**}],\ i = 1, 2, \ldots, d,$$

which is important for applications (cf. Section 2.3), the situation simplifies considerably:

Lemma 3.2. *If $\bar{\alpha}_\cdot = \bar{\alpha} \in \mathfrak{H}$, $\bar{\sigma}_\cdot^{(i)} = \bar{\sigma}^{(i)} \in \mathfrak{H}$, $i = 1, ..., d$, and $\bar{\mu}_0 \in \mathfrak{H}$, then there exists a mild solution to (14), which – up to equivalence – is unique among the processes satisfying*

$$\mathbb{P}\left(\int_0^{T^{**}} \|\bar{\mu}_t\|_{\mathfrak{H}}^2\, dt < \infty\right) = 1$$

and has a continuous modification.

Proof. The claim is a direct consequence of Theorem 7.4 in Da Prato and Zabczyk (1992): The Lipschitz and linear growth conditions on $\bar{\alpha}$ and $\bar{\sigma}$ are trivially satisfied. □

We may now "translate" the previous results, in particular the drift condition, to the "new" formulation. Similar to Björk and Christensen (1997), for a function $f \in \mathfrak{H}$ we let

$$(J\,f)\,(\tau,x) = \int_0^\tau f(s,x)\,ds$$

and require:

Assumption 3.3. *The processes* $\left(\bar{\sigma}_t^{(i)}\ \left(J\,\bar{\sigma}_t^{(i)}\right)\right)_{t\in[0,T^{**}]}$, $i = 1, ..., d$, *are* $\mathfrak{H}$*-valued predictable.*

Then the drift condition from Proposition 2.2 reads as follows:

Corollary 3.1. *If* W *is a* $\mathbb{P}$*-Brownian motion, we have*

$$\bar{\alpha}_t = \sum_{i=1}^{d} \bar{\sigma}_t^{(i)}\ \left(J\,\bar{\sigma}_t^{(i)}\right)\ \forall t \in [0,T^{**}]\ \ a.s.$$

under $\mathbb{P}$.

Proof. By Proposition 2.2, we have for given (τ, x)

$$\delta_{(\tau,x)}\left(\bar{\alpha}_t\right) \quad = \quad \sigma\left(t,t+\tau,x-t\right) \times \int_0^\tau \left(\sigma\left(t,t+s,x-t\right)\right)'\,ds = \delta_{(\tau,x)}\left(\sum_{i=1}^{d} \bar{\sigma}_t^{(i)}\ \left(J\,\bar{\sigma}_t^{(i)}\right)\right)$$

almost surely and the claim follows by continuity. □

4 Finite-Dimensional Realizations of Time-Homogenous Gaussian Forward Models

The problem we are dealing with is to find conditions under which the forward mortality model (14) can be realized by a finite-dimensional system of the form

$$\begin{cases} dZ_t = a(Z_t)\,dt + b(Z_t)\,dW_t,\ Z_0 \in \mathbb{R}^m, \\ \bar{\mu}_t(\tau,x) = G(\tau,x,Z_t), \end{cases} \tag{15}$$

where Z is a finite-dimensional diffusion process and $G : [0,\infty)^2 \times \mathbb{R}^m \to \mathbb{R}$ is a deterministic function. For interest rate models, this problem has been studied for very general setups (see e.g. Björk and Svensson (2001) or Filipović and Teichmann (2002)). Here, we restrict ourselves to the case of deterministic, time-homogenous, and differentiable volatility structures, i.e. $\bar{\sigma}_\cdot^{(i)} \equiv \bar{\sigma}^{(i)} \in \mathfrak{H}$ and $\bar{\sigma}^{(i)}\left(J\bar{\sigma}^{(i)}\right) \in \mathfrak{H}$, $i = 1, \ldots, d$, (Ass. 3.3) as it is important for applications (cf. Section 2.3).

Note that in this case, Equation (13) yields

$$\bar{\mu}_t(\tau, x) = \delta_{(\tau,x)} \underbrace{\left(\int_0^t S_{t-s} \sum_{i=1}^d \bar{\sigma}^{(i)} \, dW_s^{(i)} \right)}_{=:\bar{\mu}_t^0} + \underbrace{\delta_{(\tau,x)} \left(S_t \, \bar{\mu}_0 + \int_0^t S_{t-s} \sum_{i=1}^d \bar{\sigma}^{(i)} \, \left(J \, \bar{\sigma}^{(i)} \right) ds \right)}_{=:\xi(t,\tau,x)},$$

where ξ is a deterministic function. Therefore, it is sufficient to study $\bar{\mu}_t^0$ with $\bar{\mu}_0^0 \equiv 0$.

We commence by considering the case $\mathfrak{H} = \mathfrak{H}_{\beta,\gamma}$ and $S_t = \exp\left\{ \left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \right) t \right\}$, where we write $f(\tau + t, x - t) := (S_t f)(\tau, x)$ for $t > x$, $f \in \mathfrak{H}$. Then, we can follow Björk and Gombani (1999), who address the corresponding problem for Gaussian forward interest rate models. Applying Itô's formula, we write

$$\begin{aligned}
\bar{\mu}_t^0(\tau, x) &= \sum_{i=1}^d \int_0^t \bar{\sigma}(\tau + t - s, x - t + s) \, dW_s^{(i)} \\
&= \sum_{i=1}^d W_t^{(i)} \, \bar{\sigma}^{(i)}(\tau, x) + \int_0^t W_s^{(i)} \left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \right) \bar{\sigma}^{(i)}(\tau + t - s, x - t + s) \, ds \\
&= \left(O_{(\tau,x)} W(\omega) \right)(t),
\end{aligned}$$

where the operator $O_{(\tau,x)} : (C[0,\infty))^d \to C[0,\infty)$ is given by

$$O_{(\tau,x)} v = \sum_{i=0}^d v^{(i)}(\cdot) \, \bar{\sigma}^{(i)}(\tau, x) + \int_0^{\cdot} v^{(i)}(s) \, \left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \right) \bar{\sigma}^{(i)}(\tau + \cdot - s, x - \cdot + s) \, ds.$$

Similarly to Björk and Gombani (1999), we observe that for (τ, x) fixed, $O_{(\tau,x)}$ is continuous in the topology of uniform convergence on compacts, and thus is completely characterized by its behavior on the dense subset $\left(C^1[0,\infty) \right)^d$. For differentiable v, on the other hand, we obtain

$$\frac{\partial}{\partial t} \left(O_{(\tau,x)} v \right)(t) = \left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \right) \left(\left(O_{(\tau,x)} v \right)(t) \right) + \sum_{i=1}^d v^{(i)\prime}(t) \, \bar{\sigma}^{(i)}(\tau, x). \tag{16}$$

If, by a slight abuse of notation, we write $\left(O_{(\tau,x)} v \right)(t) = \bar{\mu}_t^0(\tau, x)$, (16) reads as

$$\frac{\partial}{\partial t} \bar{\mu}_t^0(\tau, x) = \left(A \bar{\mu}_t^0 \right)(\tau, x) + \sum_{i=1}^d v^{(i)\prime}(t) \, \bar{\sigma}^{(i)}(\tau, x). \tag{17}$$

Hence, $\bar{\mu}_t^0$ will have a finite-dimensional realization if and only if the deterministic system can be represented in terms of that realization (cf. Lemma 3.1 in Björk and Gombani (1999)). Since we have a linear dynamic system, it seems natural to look for a linear realization (cf. Björk (2003))

$$\left\{ \begin{array}{l} dZ_t = M \, Z_t \, dt + N \, dW_t, \; Z_0 = 0 \\ \bar{\mu}_t^0(\tau, x) = C(\tau, x) \times Z_t \end{array} \right. \quad \longleftrightarrow \quad \left\{ \begin{array}{l} (\partial/\partial t) \, z_t = M \, z_t + N \, v'(t), \; z_0 = 0 \\ \bar{\mu}_t^0(\tau, x) = C(\tau, x) \times z_t \end{array} \right. \tag{18}$$

where $M \in \mathbb{R}^{m\times m}$, $N \in \mathbb{R}^{m\times d}$, and $C : [0,\infty)^2 \to \mathbb{R}^m$. We have

$$\begin{aligned}
\bar{\mu}^0_t(\tau,x) &= \int_0^t \bar{\sigma}(\tau+(t-s), x-(t-s))\, v'(s)\, ds = \left[\bar{\sigma}(\tau+\cdot, x-\cdot) * v'\right](t) \\
\Rightarrow \mathcal{L}\left[\bar{\mu}^0_\cdot(\tau,x)\right](y) &= \underbrace{\mathcal{L}\left[\bar{\sigma}(\tau+\cdot, x-\cdot)\right](y)}_{=:F(y,\tau,x)} \times \mathcal{L}\left[v'\right](y),
\end{aligned}$$

where $\mathcal{L}$ denotes the Laplace transform and we call F the *transfer function* of (17). Therefore, the question of whether (18) is a finite-dimensional realization simplifies to the question of whether (18) has the same transfer function as (17).

For the system (18), we find

$$\begin{aligned}
\bar{\mu}^0_t(\tau,x) &= \int_0^t C(\tau,x)\, \exp\{M\,(t-s)\}\, N\, v'(s)\, ds \\
\Rightarrow \mathcal{L}\left[\bar{\mu}^0_\cdot(\tau,x)\right](y) &= \int_0^\infty e^{-yz} \int_0^z C(\tau,x)\, \exp\{M\,(z-s)\}\, N\, v'(s)\, ds\, dz \\
&= \underbrace{C(\tau,x)\, \left[y\, I - M\right]^{-1} N}_{=:T(y,\tau,x)}\, \mathcal{L}[v'](y),
\end{aligned}$$

where $I \in \mathbb{R}^{m\times m}$ is the identity matrix and T is the transfer function of (18).

Hence, (17) can be represented in terms of (18) if and only if we can find M, N and $C(\cdot,\cdot)$ such that

$$\begin{aligned}
F(y,\tau,x) &\overset{!}{=} T(y,\tau,x) = C(\tau,x)\, \left[y\, I - M\right]^{-1}\, N \\
\Leftrightarrow \bar{\sigma}(\tau+t, x-t) &= \mathcal{L}^{-1}\left[C(\tau,x)\, \left[\cdot\, I - M\right]^{-1}\, N\right](t) = C(\tau,x)\, \exp\{M\, t\}\, N \\
\Leftrightarrow \bar{\sigma}(\tau,x) &= C(x+\tau)\, \exp\{M\, \tau\}\, N,
\end{aligned}$$

where $C(x) = C(0,x)$.

Consider now the case $\mathfrak{H} = \widetilde{\mathfrak{H}}_{\beta,\gamma}$ and $\{S_t\}_{t\geq 0}$ given by (11). Then,

$$\begin{aligned}
\bar{\mu}^0_t(\tau,x) &= \sum_{i=1}^d \int_0^t \left(S_{t-s}\, \bar{\sigma}^{(i)}\right)(\tau,x)\, dW^{(i)}_s \\
&= \sum_{i=1}^d W^{(i)}_{(t-x)^+} \bar{\sigma}(\tau+x,0) + \sum_{i=1}^d \int_{(t-x)^+}^t \bar{\sigma}(\tau+(t-s), x-(t-s))\, dW^{(i)}_s \\
\Leftrightarrow \quad & \bar{\mu}^0_t(\tau,x) - \sum_{i=1}^d W^{(i)}_{(t-x)^+} \bar{\sigma}(\tau+x,0) \\
&= \sum_{i=1}^d \int_0^{t-(t-x)^+} \bar{\sigma}(\tau+t-(u+(t-x)^+), x-(t-(u+(t-x)^+)))\, dW^{(i)}_{u+(t-x)^+} \\
&= \sum_{i=1}^d \int_0^{t\wedge x} \bar{\sigma}(\tau+((x\wedge t)-u), x-((x\wedge t)-u))\, d\widetilde{W}^{(i)}_u, \qquad (19)
\end{aligned}$$

where $\widetilde{W}$ with $\widetilde{W}_u = W_{u+(t-x)^+} - W_{(t-x)^+}$ is a Brownian motion. Now the results above assert that a finite-dimensional realization of the right-hand side of (19) exists if and only if

$$\bar{\sigma}(\tau, x) = C(x+\tau)\,\exp\{M\,\tau\}\,N,$$

and it is given by

$$C(x+\tau)\,\exp\{M\,\tau\}\,\check{Z}_{t\wedge x} = C(x+\tau)\,\exp\{M\,\tau\}\,\left(Z_t - Z_{(t-x)^+}\right),$$

where

$$d\check{Z}_u = M\,\check{Z}_u\,du + N\,d\widetilde{W}_u,\ \check{Z}_0 = 0,$$

and $(Z_t)_{t\geq 0}$ as in (18).

The following proposition summarizes the results:

Proposition 4.1. *1. The (Gaussian) forward mortality model (14) has a finite-dimensional realization (18) in $\mathfrak{H}_{\beta,\gamma}$ if and only if $\bar{\sigma}$ can be represented as*

$$\bar{\sigma}(\tau, x) = C(x+\tau)\,\exp\{M\,\tau\}\,N \tag{20}$$

for $M \in \mathbb{R}^{m\times m}$, $N \in \mathbb{R}^{m\times d}$, and $C \in C^1([0,\infty), \mathbb{R}^m)$.

In this case, a concrete realization is given by

$$\begin{cases} dZ_t &= M\,Z_t\,dt + N\,dW_t,\ Z_0 = 0, \\ \bar{\mu}_t(\tau, x) &= \xi(t,\tau,x) + C(x+\tau)\,\exp\{M\,\tau\}\,Z_t. \end{cases}.$$

2. If $\mathfrak{H} = \widetilde{\mathfrak{H}}_{\beta,\gamma}$ and $\{S_t\}_{t\geq 0}$ is defined via (11), representation (20) is still necessary and sufficient for the existence of a finite-dimensional realization. In this case, a concrete realization is given by

$$\begin{cases} dZ_t &= M\,Z_t\,dt + N\,dW_t,\ Z_0 = 0\,/\,(W_t)_{t\geq 0}, \\ \bar{\mu}_t(\tau, x) &= \xi(t,\tau,x) + C(x+\tau)\,\exp\{M\,\tau\}\,\left(Z_t - Z_{(t-x)^+}\right) \\ & \quad + C(x+\tau)\,\exp\{M(\tau+x)\}\,N\,W_{(t-x)^+}. \end{cases}$$

Note that while in Proposition 4.1.2, the realization of $\bar{\mu}_t$ is given in terms of the $(m+d)$-dimensional process $(Z', W')'$, it depends on the entire path rather than the terminal value only, and hence it is not quite of the form postulated in (15). In particular, Proposition 4.1 asserts that no such realization exists in general. Therefore, both the existence and the form of the finite-dimensional realization depend on the choice of the semigroup and, thus, indirectly on the considered space.

5 Example: A Gaussian Gompertz Model

We proceed by providing a "simple" example, namely a Gaussian Gompertz model, to illustrate our findings:

$$\begin{aligned}
\mu_t(x_0) &= Z_t \exp\{a(x_0+t)\}, \\
dZ_t &= \zeta\,(\beta - Z_t)\,dt + \psi\,dW_t,\ Z_0>0,
\end{aligned} \tag{21}$$

where a, ζ, β, and ψ are positive constants with $a \neq \zeta$. Similar Gaussian examples have been proposed by Schrager (2006). By an application of Itô's Lemma, we obtain

$$d\mu_t(x_0) = (\zeta - a)\left[\frac{\zeta\beta}{\zeta-a}e^{a(x_0+t)} - \mu_t(x_0)\right]dt + e^{a(x_0+t)}\psi\,dW_t. \tag{22}$$

In particular, we observe that the dynamics of $\mu_t(x_0)$ have an affine structure (cf. Duffie et al. (2000, 2003)), so we have

$$\mathbb{E}\left[\exp\left\{-\int_t^T \mu_s(x_0)\,ds\right\}\middle|\,\mathcal{F}_t\right] = \exp\left\{u_{T,x_0}(T-t) + v(T-t)\,\mu_t(x_0)\right\},$$

where u and v satisfy the following Riccati ordinary differential equations (ODEs):

$$\begin{aligned}
v'(y) &= -(\zeta-a)\,v(y) - 1,\ v(0)=0, \\
u'_{T,x_0}(y) &= \zeta\,\beta\,v(y)e^{x_0+T-y} + \frac{1}{2}\psi^2\,v^2(y)\,e^{2a(x_0+T-y)},\ u(0)=0.
\end{aligned}$$

For the first equation, it is well-known that the unique solution is

$$v(y) = -\frac{1}{\zeta-a}\left(1-e^{-(\zeta-a)y}\right),$$

and thus, by some simple calculations,

$$\begin{aligned}
u_{T,x_0}(y) &= \int_0^y \zeta\,\beta\,v(s)\,e^{a(x_0+T-s)}\,ds + \int_0^y \frac{1}{2}\psi^2\,v^2(s)\,e^{2a(x_0+T-s)}\,ds \\
&= -\frac{\zeta\beta}{\zeta(\zeta-a)}e^{x_0+T}\left[1-e^{ay}\right] + \frac{\beta}{\zeta-a}e^{x_0+T}\left[1-e^{\zeta y}\right] + \frac{\psi^2}{4(\zeta-a)^2 a}e^{2a(x_0+T)}\left[1-e^{2ay}\right] \\
&\quad -\frac{\psi^2}{(\zeta-a)^2(\zeta+a)}e^{2a(x_0+T)}\left[1-e^{-(\zeta+a)y}\right] + \frac{\psi^2}{4(\zeta-a)^2\zeta}e^{2a(x_0+T)}\left[1-e^{-2\zeta y}\right].
\end{aligned}$$

With some calculus, we obtain

$$\begin{aligned}
\mu_t(T,x_0) &= -\frac{\partial}{\partial T}\log\left\{\mathbb{E}\left[\exp\left\{-\int_t^T \mu_t(x_0)\right\}\middle|\,\mathcal{F}_t\right]\right\} \\
&= \beta e^{a(x_0+T)}\left[1-e^{-\zeta(T-t)}\right] - \frac{\psi^2}{2\zeta(\zeta+a)}e^{2a(x_0+T)} + \frac{\psi^2}{(\zeta-a)(\zeta+a)}e^{2a(x_0+T)}e^{-(\zeta+a)(T-t)} \\
&\quad -\frac{\psi^2}{2(\zeta-a)\zeta}e^{2a(x_0+T)}e^{-2\zeta(T-t)} + e^{a(x_0+T)}e^{-\zeta(T-t)}Z_t,
\end{aligned} \tag{23}$$

and, again by some simple calculations,

$$d\mu_t(T,x_0) = e^{2a(x_0+T)}\frac{\psi^2}{\zeta-a}e^{-\zeta(T-t)}\left[e^{-a(T-t)}-e^{-\zeta(T-t)}\right]dt + \psi e^{a(x_0+T)}e^{-\zeta(T-t)}\,dW_t. \tag{24}$$

In order to double-check our calculations, we may apply Proposition 2.1 to recover the spot force dynamics. We get

$$\begin{aligned}
b(t,x_0) &= \sigma(t,t,x_0) = \psi e^{a(T+x_0)}e^{-\zeta(T-t)}\Big|_{T=t} = \psi e^{a(x_0+t)},\\
\alpha(t,t,x_0) &= e^{2a(x_0+T)}\frac{\psi^2}{\zeta-a}e^{-\zeta(T-t)}\left[e^{-a(T-t)}-e^{-\zeta(T-t)}\right]\Big|_{T=t} = 0\\
\Rightarrow a(t,x_0) &= \frac{\partial}{\partial T}\mu_t(T,x_0)\Big|_{T=t} = (\zeta-a)\left[\frac{\zeta\beta}{\zeta-a}e^{a(x_0+t)}-\mu_t(t,x_0)\right],
\end{aligned}$$

i.e. we have (22). Therefore, $\sigma(t,T,x_0)=\psi e^{a(T+x_0)}e^{-\zeta(T-t)}$, and, by the drift condition (Prop. 2.2), we should also have

$$\begin{aligned}
\alpha(t,T,x_0) &\stackrel{!}{=} \sigma(t,T,x_0)\int_t^T \sigma(t,s,x_0)'\,ds\\
&= e^{2a(x_0+T)}\frac{\psi^2}{\zeta-a}e^{-\zeta(T-t)}\left[e^{-a(T-t)}-e^{-\zeta(T-t)}\right],
\end{aligned}$$

which verifies the proposition for this special case.

Let us now consider the infinite-dimensional formulation. By Equations (9) and (10), we have

$$\begin{aligned}
\delta_{(\tau,x_t)}(\bar{\alpha}) &= \bar{\alpha}(\tau,x_t) = \alpha\left(t,t+\tau,x_t-t\right) = e^{2a(x_t+\tau)}\frac{\psi^2}{\zeta-a}e^{-\zeta\tau}\left[e^{-a\tau}-e^{-\zeta\tau}\right],\\
\delta_{(\tau,x_t)}(\bar{\sigma}) &= \bar{\sigma}(\tau,x_t) = \sigma\left(t,t+\tau,x_t-t\right) = \psi e^{a(x_t+\tau)}e^{-\zeta\tau},\\
\delta_{(\tau,x_t)}(\bar{\mu}_t) &= \bar{\mu}_t(\tau,x_t) = \mu_t\left(t+\tau,x_t-t\right),
\end{aligned}$$

and the dynamics as an SDE on $\mathfrak{H}$ read as (cf. (14))

$$d\bar{\mu}_t = \left(A\bar{\mu}_t + \bar{\alpha}\right)dt + \bar{\sigma}\,dW_t. \tag{25}$$

In particular, $\bar{\mu}_t$, $\bar{\alpha}$, and $\bar{\sigma}$ are in $\mathfrak{H}_{\beta,\gamma}$ for $\gamma > 2a$ and $\beta > 1$, with A defined accordingly.

Since $\bar{\sigma}$ is time-homogenous, we can apply the results from Section 4 to recover the finite-dimensional realization from (21). Here, the case is particularly simple as the volatility structure is already given in the form (20):

$$\bar{\sigma}(\tau,x_t) = \underbrace{e^{a(x_t+\tau)}}_{=C(x_t+\tau)}\quad\underbrace{e^{-\zeta\tau}}_{\Rightarrow M=-\zeta}\quad\underbrace{\psi}_{=N},$$

and the finite-dimensional realization is thus given by

$$\begin{cases} d\bar{Z}_t = -\zeta \bar{Z}_t \, dt + \psi \, dW_t, \ \bar{Z}_0 = 0, \\ \bar{\mu}_t(\tau, x) = \xi(t,\tau,x) + C(x_t+\tau)\, e^{-M\tau}\, \bar{Z}_t = \xi(t,\tau,x) + e^{a(x_t+\tau)} e^{-\zeta\tau} \bar{Z}_t, \end{cases}$$

where by Corollary 3.1,

$$\begin{aligned} \xi(t,\tau,x_t) &= \bar{\mu}_0(t+\tau, x_t - t) + \int_0^t \delta_{(\tau+t-s, x-t+s)} \left(\bar{\sigma}\left(J\bar{\sigma}\right)\right) \, ds \\ &= \bar{\mu}_0(t+\tau, x_t - t) + \int_0^t e^{2a(x_t+\tau)} \frac{\psi^2}{\zeta - a} e^{-\zeta(\tau+t-s)} \left[e^{-a(\tau+t-s)} - e^{-\zeta(\tau+t-s)} \right] ds, \end{aligned}$$

and hence, after some calculations,

$$\begin{aligned} \bar{\mu}_t(\tau, x_t) &= \bar{\mu}_0(t+\tau, x_t - t) + e^{2a(x_t+\tau)} \frac{\psi^2}{\zeta - a} e^{-\zeta\tau} \left[\frac{1}{\zeta + a} e^{-a(\tau+t)} \left[e^{at} - e^{-\zeta t} \right] \right. \\ & \left. \quad - \frac{1}{2\zeta} e^{-\zeta(\tau+t)} \left[e^{\zeta t} - e^{-\zeta t} \right] \right] + e^{a(x_t+\tau)} e^{-\zeta\tau} \bar{Z}_t. \end{aligned}$$

In order to recover the original equation, we have to incorporate the "original" forward surface $\bar{\mu}_0$ into the last equation. From (23) for $t = 0$ we have:

$$\begin{aligned} \bar{\mu}_0(t+\tau, x_t - t) &= \beta e^{a\,(x_t+\tau)} \left[1 - e^{\zeta(t+\tau)} \right] - \frac{\psi^2}{2(\zeta+a)\zeta} e^{2a(x_t+\tau)} + \frac{\psi^2}{(\zeta-a)(\zeta+a)} e^{2a(x_t+\tau)} e^{-(\zeta+a)(\tau+t)} \\ & \quad - \frac{\psi^2}{2(\zeta-a)\zeta} e^{2a(x_t+\tau)} e^{-2\zeta(t+\tau)} + e^{a(x_t+\tau)} e^{-\zeta(t+\tau)} Z_0, \end{aligned} \tag{26}$$

and hence,

$$\begin{aligned} \bar{\mu}_t(\tau, x_t) &= \beta e^{a(x_t+\tau)} \left[1 - e^{-\zeta\tau} \right] - \frac{\psi^2}{2\zeta(\zeta+a)} e^{2a(x_t+\tau)} + \frac{\psi^2}{(\zeta-a)(\zeta+a)} e^{2a(x_t+\tau)} e^{-(\zeta+a)\tau} \\ & \quad - \frac{\psi^2}{2(\zeta-a)\zeta} e^{2a(x_t+\tau)} e^{-2\zeta\tau} + e^{a(x_t+\tau)} e^{-\zeta\tau} \underbrace{\left[\bar{Z}_t + Z_0 e^{-\zeta t} - \beta e^{-\zeta t} + \beta \right]}_{=:\tilde{Z}_t}. \end{aligned} \tag{27}$$

So, we have (23) for $t \geq 0$ with $Z_t = \tilde{Z}_t$. By Itô's Lemma, on the other hand, we obtain

$$\begin{aligned} d\tilde{Z}_t &= -\zeta \underbrace{\left(\bar{Z}_t + Z_0 e^{-\zeta t} - \beta e^{-\zeta t} + \beta \right)}_{=\tilde{Z}_t} dt + \zeta\beta \, dt + \psi \, dW_t = \zeta \left(\beta - \bar{Z}_t \right) \, dt + \psi \, dW_t, \\ \tilde{Z}_0 &= Z_0, \end{aligned} \tag{28}$$

and thus $\tilde{Z}_t = Z_t$ a.s., so we come full circle.

However, by considering the dynamics on $\mathfrak{H}_{\beta,\gamma}$, we implicitly made an assumption about the evolution of mortality for future generations determined by the choice of the semigroup. Since $\mathfrak{H}_{\beta,\gamma} \subset \widetilde{\mathfrak{H}}_{\beta,\gamma}$, we may alternatively consider the dynamics (25) on $\widetilde{\mathfrak{H}}_{\beta,\gamma}$ with the semigroup (and hence, A) specified according to

(11). By Proposition 4.1, in this case,

$$\begin{aligned}\bar{\mu}_t(\tau, x_t) &= \xi(t,\tau,x_t) + e^{a(x_t+\tau)} e^{-\zeta\tau} \left(\bar{Z}_t - \bar{Z}_{(t-x_t)^+}\right) \\ &\quad + e^{a(x_t+\tau)}\, e^{-\zeta(x_t+\tau)}\, \psi\, W_{(t-x_t)^+},\end{aligned}$$

where, after some calculus,

$$\begin{aligned}\xi(t,\tau,x_t) &= \delta_{(\tau,x_t)} \left(S_t\,\bar{\mu}_0 + \int_0^t S_{t-s}\,\bar{\sigma}\,\left(J\,\bar{\sigma}\right) ds \right) \\ &= \bar{\mu}_0(\tau + (x_t \wedge t), (x_t - t)^+) + e^{2a(x_t+\tau)} \frac{\psi^2}{\zeta - a} e^{-\zeta\tau} \Big[e^{-\zeta\, x_t} \Big(e^{-a\,(\tau+x_t)} (t - x_t)^+ \\ &\quad - e^{-\zeta\,(\tau+x_t)} (t-x_t)^+ \Big) + \frac{e^{-a(\tau+t)}}{\zeta + a} \left[e^{a\,(x_t \wedge t) - \zeta(t-x_t)^+} - e^{-\zeta t} \right] \\ &\quad - \frac{e^{-\zeta(\tau+t)}}{2\zeta} \left[e^{\zeta(t\wedge x_t) - \zeta(t-x_t)^+} - e^{-\zeta t} \right] \Big],\end{aligned}$$

and, thus, by (26) and several calculations,

$$\begin{aligned}\bar{\mu}_t(\tau, x_t) &= \frac{\psi^2}{\zeta - a} e^{(a-\zeta)(x_t+\tau)} (t-x_t)^+ \left[1 - e^{(a-\zeta)(x_t+\tau)} \right] + e^{(a-\zeta)(x_t+\tau)}\, \psi\, W_{(t-x_t)^+} \\ &\quad + \beta e^{a(x_t+\tau)} \left[1 - e^{-\zeta\tau - \zeta(t-x_t)^+} + e^{-\zeta(\tau+t)} - e^{-\zeta(\tau + (t\wedge x_t))} \right] - \frac{\psi^2}{2\zeta(\zeta+a)} e^{2a(x_t+\tau)} \\ &\quad + \frac{\psi^2}{(\zeta-a)(\zeta+a)} e^{2a(x_t+\tau)} e^{-(\zeta+a)\tau} \left[e^{-(a+\zeta)(t-x_t)^+} - e^{-(a+\zeta)t} + e^{-(a+\zeta)(t\wedge x_t)} \right] \\ &\quad - \frac{\psi^2}{2(\zeta-a)\zeta} e^{2a(x_t+\tau)} e^{-2\zeta\tau} \left[e^{-2\zeta(t-x_t)^+} - e^{-2\zeta t} + e^{-2\zeta(t\wedge x_t)} \right] \\ &\quad + e^{a(x_t+\tau)} e^{-\zeta\tau} \left[Z_0 \left(e^{-\zeta(t-x_t)^+} - e^{-\zeta t} + e^{-\zeta(t\wedge x_t)} + \tilde{Z}_t - \tilde{Z}_{(t-x_t)^+} \right) \right]. \qquad (29)\end{aligned}$$

In particular, for $x_t \geq t \Leftrightarrow x_0 = x_t - t \geq 0$, we have (27), so that (29)/(28) is a finite-dimensional realization for (24).

If $x_t < t$, for example $x_t = 0$, on the other hand, we obtain

$$\bar{\mu}_t(\tau, 0) = \frac{\psi^2}{\zeta - a} e^{(a-\zeta)\tau} t \left[1 - e^{(a-\zeta)\tau} \right] + e^{(a-\zeta)\tau}\, \psi\, W_t + \bar{\mu}_0(\tau, 0),$$

and therefore,

$$\begin{aligned}{}_{T-t}p_0(t;T) &= \exp\left\{ - \int_0^{T-t} \bar{\mu}_t(s,0)\, ds \right\} \\ &= {}_{T-t}p_0(0;T-t) \underbrace{\exp\left\{ -\frac{1}{2} \frac{\psi^2}{(a-\zeta)^2} t \left[e^{(a-\zeta)\,(T-t)} - 1 \right]^2 - \frac{\psi}{a-\zeta} \left[e^{(a-\zeta)\,(T-t)} - 1 \right] W_t \right\}}_{=Y},\end{aligned}$$

where obviously Y is a random variable with expected value $\mathbb{E}[Y] = 1$. This illustrates that under (11), newborns at time t are expected to enter the world with the same expectations regarding future mortality as newborns at time zero (cf. Section 3.1).

6 Conclusion

Forward mortality models specify dynamics of the entire age/term-structure of mortality. In particular, mortality forecasts are included as inputs, whereas the variability of these projections is modeled. The current paper presents a theoretical investigation of forward mortality models driven by a finite-dimensional Brownian motion. However, our findings also have direct practical implications and applications documenting their relevance. Examples include:

- *Valuation and analysis of mortality-contingent embedded options*

  For the numerical analysis and valuation of non-linear mortality-contingent payoffs in embedded options within life insurance contracts such as Guaranteed Minimum Income Benefits (GMIBs) within Variable Annuities, forward models possess the advantage that no "nested simulations" are necessary since the simulated forward surface usually allows for a closed-form valuation at expiration. Our results help to identify and build suitable models.

- *Asset-liability management*

  In asset-liability management, life insurers face the problem of how to stochastically forecast their liability side, i.e. how to apply *risk factors* to the life table underlying their reserve calculations. This question is addressed in Section 4: Given the initial surface $\bar{\mu}_0$ corresponding to the initial mortality table, Proposition 4.1 shows how to adjust it given a sampled "risk factor" $Z_t(\omega)$ to obtain a consistent sample surface – or life table – at time t.

Aside from operationalizing these applications, future research directions include the survey of finite-dimensional realizations for more general volatility structures, the analysis of more general spaces for the infinite-dimensional formulation of forward mortality models, and the generalization of the framework towards a larger class of driving processes.

References

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

Bauer, D., Börger, M., Ruß, J., 2010. On the Pricing of Longevity-Linked Securities. Forthcoming in Insurance: Mathematics and Economics.

Biffis, E., 2005. Affine processes for dynamic mortality and actuarial valuation. Insurance: Mathematics and Economics, 37: 443-468.

Biffis, E., Denuit, M., Devolder, P., 2010. Stochastic Mortality under Measure Change. Forthcoming in the Scandinavian Actuarial Journal.

Björk, T., 1999. Arbitrage Theory in Continuous Time. Oxford University Press, Oxford.

Björk, T., 2003. On the Geometry of Interest Rate Models. Lecture Notes in Mathematics. Springer, Berlin.

Björk, T., Christensen, B.J., 1997. Interest Rate Dynamics and Consistent Forward Rate Curves. Mathematical Finance, 9: 323-348.

Björk, T., Gombani, A., 1999. Minimal realizations of interest rate models. Finance and Stochastics, 3: 413-432.

Björk, T., Svensson, L., 2001. On the Existence of Finite-Dimensional Realizations for Nonlinear Forward Rate Models. Mathematical Finance, 11: 205-243.

Blake, D., Cairns, A.J.G., Dowd, K., 2008. The Birth of the Life Market. Alternative Investments Quarterly, Fourth Quarter 2008.

Booth, H., 2006. Demographic Forecasting: 1980 to 2005 in Review. International Journal of Forecasting, 22: 547-581.

Cairns, A.J.G., Blake, D., Dowd, K., 2006. Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. ASTIN Bulletin, 36: 79-120.

Carmona, R., Tehranchi, M., 2005. Interest Rate Models: an Infinite Dimensional Stochastic Analysis Perspective. Springer Finance. Springer, Berlin.

Da Prato, G., Zabczyk, J., 1992. Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge.

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

Dawson, P., Dowd, K., Cairns, A.J.G., Blake, D., 2010. Survivor Derivatives: A Consistent Pricing Framework. Forthcoming in the Journal of Risk and Insurance.

Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Mathematische Annalen, 300: 463-520.

Duchassing, J., Suter, F., 2009. Longevity: A "simple" stochastic modelling of mortality. Presentation at the 2009 LIFE Colloquium on behalf of Partner RE.

Duffie, D., Filipović, D., Schachermayer, W., 2003. Affine processes and applications in finance. Annals of Applied Probability, 13: 984-1053.

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

Eberlein, E., Raible, S., 1999. Term structure models driven by general Lévy processes. Mathematical Finance, 9: 31-54.

Engel, K.-J., Nagel, R., 2000. One-Parameter Semigroups for Linear Evolution Equations. Graduate Texts in Mathematics, 194. Springer, New York.

Filipović, D., 2001. Consistency Problems for Heath Jarrow Morton Interest Rate Models. Lecture Notes in Mathematics, 1760. Springer, Berlin.

Filipović D., Teichmann, J., 2002. On finite-dimensional term structure models. Working Paper, Princeton University.

Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod security markets. Journal of Economic Theory, 20: 381–408.

Heath, D., Jarrow, R., Morton, R., 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation. Econometrica, 60: 77-105.

Lando, D., 1998. On Cox processes and credit risky securities. Review of Derivatives Research, 2: 99-120.

Milevsky, M.A., Promislow, S.D., 2001. Mortality Derivatives and the Option to Annuitize. Insurance: Mathematics and Economics, 29: 299-318.

Miltersen, K.R., Persson, S.A., 2005. Is mortality dead? Stochastic force of mortality determined by no arbitrage. Working Paper, Norwegian School of Economics and Business Administration.

Musiela, M., 1993. Stochastic PDEs and term structure models. Journées Internationales de Finance. IGR-AFFI, La Baule.

Schrager, D., 2006. Affine stochastic mortality. Insurance: Mathematics and Economics, 38: 81-97.

Vargiolu, T., 1999. Invariant measures for the Musiela equation with deterministic diffusion term. Finance and Stochastic, 3: 483-492.

Appendices

Proof of Proposition 3.1

Let $\beta > 1$ and $\gamma > 0$. Proceeding similarly to Björk and Svensson (2001), we start by noting that for $A = \frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}$, we have

$$\|A\, f\|^2_{\beta,\gamma} \;=\; \left\| \left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x} \right) f \right\|^2_{\beta,\gamma} \leq \left(\left\| \frac{\partial}{\partial \tau} f \right\|_{\beta,\gamma} + \left\| \frac{\partial}{\partial x} f \right\|_{\beta,\gamma} \right)^2 \leq \beta\, 4\, \|f\|^2_{\beta,\gamma}.$$

In particular, $A : \mathfrak{H}_{\beta,\gamma} \to \mathfrak{H}_{\beta,\gamma}$ defines a continuous operator such that $S_t := \exp\{t\,A\} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}$ defines a uniformly continuous semigroup satisfying Assumption 3.1.3. Moreover, it is easily verified that $\langle\cdot,\cdot\rangle_{\beta,\gamma}$ defines an inner product, so $\left(\mathfrak{H}_{\beta,\gamma}, \langle\cdot,\cdot\rangle_{\beta,\gamma}\right)$ is a pre-Hilbert (inner product) space.

For the proof of completeness, we introduce the measures ν and π via $\nu(E) = \sum_{(k,l)\in E} \beta^{-(k+l)}$, $E \subseteq \mathbb{N}_0^2$, and $\pi(F) = \int_0^\infty \int_0^\infty \mathbf{1}_F(\tau,x)\, e^{-\gamma(\tau+x)}\, d\tau\, dx$, $F \in \mathcal{B}\left([0,\infty)^2\right)$, respectively. Denote by $\pi^\tau(F) = \int_F e^{-\gamma\tau}\, d\tau$ and $\pi^x(F) = \int_F e^{-\gamma x}\, dx$, $F \in \mathcal{B}([0,\infty))$, so that π can be understood as $\pi^\tau \times \pi^x$. Moreover, for $f \in C^\infty\left([0,\infty)^2\right)$, let $Tf : \mathbb{N}_0^2 \times [0,\infty)^2 \to \mathbb{R}$ be given by $Tf\,((k,l),(\tau,x)) = f^{k,l}(\tau,x)$, where the first superscript denotes the k^{th} derivative with respect to τ and the second superscript denotes the l^{th} derivative with respect to x, so

$$\left(\mathfrak{H}_{\beta,\gamma}, \langle\cdot,\cdot\rangle_{\beta,\gamma}\right) = \left\{ f \in C^\infty\left([0,\infty)^2\right) \,\middle|\, Tf \in L^2(\nu\times\pi) \right\}.$$

Now let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\left(\mathfrak{H}_{\beta,\gamma}, \langle\cdot,\cdot\rangle_{\beta,\gamma}\right)$ and define $f_{n,m} := f_n - f_m$ as well as $f_{n,m}^{k,l} = f_n^{k,l} - f_m^{k,l}$. Then as $n,m\to\infty$, $Tf_{n,m} \to 0$ in the (complete) space $L^2(\nu\times\pi)$, and there exists g with $Tf_n \to g$ in $L^2(\nu\times\pi)$ and a subsequence $(Tf_{n'})_{n'\in\mathbb{N}}$ such that $Tf_{n'} \to g$ $(\nu\times\pi)$-a.e. We have

$$\begin{aligned} f_n^{k,l} &\longrightarrow g_{k,l} \text{ in } L^2(\pi), && (30)\\ f_n^{k+1,l}(\cdot,x) &\longrightarrow g_{k+1,l}(\cdot,x) \text{ in } L^2(\pi^\tau) \text{ for } \pi^x\text{-almost every } x, && (31)\\ f_n^{k,l+1}(\tau,\cdot) &\longrightarrow g_{k,l+1}(\tau,\cdot) \text{ in } L^2(\pi^x) \text{ for } \pi^\tau\text{-almost every } \tau. && (32) \end{aligned}$$

For $T, X > 0$, let $\tau_0 \in [0,T]$ and $x_0 \in [0,X]$ such that (31) and (32) are satisfied. Again following Björk and Svensson (2001), we shall now prove that for every (k,l), $f_n^{k,l}$ is Cauchy in $C([0,T]\times[0,X])$: We commence by noting that for $f_{n,m}^{k,l}$,

$$f_{n,m}^{k,l}(\tau,x) = f_{n,m}^{k,l}(\tau_0,x_0) + \int_{x_0}^{x} f_{n,m}^{k,l+1}(\tau_0,s)\,ds + \int_{\tau_0}^{\tau} f_{n,m}^{k+1,l}(t,x_0)\,dt + \int_{\tau_0}^{\tau}\int_{x_0}^{x} f_{n,m}^{k+1,l+1}(t,s)\,ds\,dt.$$

Now for $(\tau,x) \in [0,T]\times[0,X]$,

$$\begin{aligned} \left|\int_{x_0}^{x} f_{n,m}^{k,l+1}(\tau_0,s)\,ds\right|^2 &\le \left(\frac{e^{\gamma X}-1}{\gamma}\right) \times \left\| f_{n,m}^{k,l+1}(\tau_0,\cdot)\right\|_{L^2(\pi^x)}^2,\\ \left|\int_{\tau_0}^{\tau} f_{n,m}^{k+1,l}(t,x_0)\,dt\right|^2 &\le \left(\frac{e^{\gamma T}-1}{\gamma}\right) \times \left\| f_{n,m}^{k+1,l}(\cdot,x_0)\right\|_{L^2(\pi^\tau)}^2,\\ \left|\int_{\tau_0}^{\tau}\int_{x_0}^{x} f_{n,m}^{k+1,l+1}(t,s)\,ds\,dt\right|^2 &\le \left(\frac{e^{\gamma X}-1}{\gamma}\right) \times \left(\frac{e^{\gamma T}-1}{\gamma}\right) \times \left\| f_{n,m}^{k+1,l+1}\right\|_{L^2(\pi)}^2\\ \Rightarrow \left\| f_{n,m}^{k,l} - f_{n,m}^{k,l}(\tau_0,x_0)\right\|_{C([0,T]\times[0,X])} &\le K(\gamma,T,X) \times \left[\left\| f_{n,m}^{k,l+1}(\tau_0,\cdot)\right\|_{L^2(\pi^x)}^2 + \left\| f_{n,m}^{k+1,l}(\cdot,x_0)\right\|_{L^2(\pi^\tau)}^2 \right.\\ &\qquad \left. + \left\| f_{n,m}^{k+1,l+1}\right\|_{L^2(\pi)}^2 \right], && (33) \end{aligned}$$

where $K(\gamma,T,X)$ is a constant solely depending on γ, T and X, and the right-hand side goes to zero as

$n, m \to \infty$ by (30), (31), and (32).

From this, analogously to Björk and Svensson (2001), we are able to deduce that

$$f_{n,m}^{k,l}(\tau_0, x_0) \to 0,\ n, m \to \infty,$$

i.e. $f_n^{k,l}(\tau_0, x_0)$ is Cauchy in $\mathbb{R}$: If it were not, we would have a contradiction to $f_{n,m}^{k,l} \to 0$ in $L^2(\pi)$. Moreover,

$$\left\|f_{n,m}^{k,l}\right\|_{C([0,T]\times[0,X])} \leq \left|f_{n,m}^{k,l}(\tau_0, x_0)\right| + \left\|f_{n,m}^{k,l} - f_{n,m}^{k,l}(\tau_0, x_0)\right\|_{C([0,T]\times[0,X])} \to 0,\ n, m \to \infty,$$

i.e. $\left(f_n^{k,l}\right)_{n\in\mathbb{N}}$ is Cauchy in $C\left([0,T] \times [0,X]\right)$. In particular, this implies

$$\delta_{(\tau,x)}\left(f_n\right) = f_n(\tau, x) \to f(\tau, x) = \delta_{(\tau,x)}\left(f\right),\ n \to \infty,\ (\tau, x) \in [0, \infty)^2,$$

i.e. the evaluation functional is continuous (Ass. 3.1.2).

To finally show completeness, our goal is now to show that $f_n \to g\left((0,0), \cdot\right)$ in $\left(\mathfrak{H}_{\beta,\gamma}, \langle\cdot,\cdot\rangle_{\beta,\gamma}\right)$, for which it suffices to show that $\frac{\partial^{k+l}}{\partial\tau^k\partial x^l} g\left((0,0), \cdot\right) = g\left((k,l), \cdot\right)$, or inductively

$$\frac{\partial}{\partial\tau} g\left((k,l), \cdot\right) = g\left((k+1,l), \cdot\right) \text{ and } \frac{\partial}{\partial x} g\left((k,l), \cdot\right) = g\left((k,l+1), \cdot\right)\ \forall\, (k,l) \in \mathbb{N}_0^2. \tag{34}$$

Since $\left(f_n^{k,l}\right)_{n\in\mathbb{N}}$ and $\left(f_n^{k+1,l}\right)_{n\in\mathbb{N}}$ are Cauchy sequences in the topology of uniform convergence on compacts, $f_n^{k,l} \to h_{(k,l)}$ and $f_n^{k+1,l} \to h_{(k+1,l)}$ uniformly on every compact K for some $h_{(k,l)},\ h_{(k+1,l)} \in C(K)$, which by the uniqueness of the limit (subsequence (n')) and the continuity implies $h_{(k,l)} = g\left((k,l), \cdot\right)$ and $h_{(k+1,l)} = g\left((k+1,l), \cdot\right)$ on K. Since the convergence is uniform on K, we obtain

$$g\left((k+1,l), \cdot\right) = \lim_{n\to\infty} f_n^{k+1,l} = \frac{\partial}{\partial\tau} \lim_{n\to\infty} f_n^{k,l} = \frac{\partial}{\partial\tau} g\left((k,l), \cdot\right)$$

on K and, hence, on $[0, \infty)^2$. This shows the left-hand side of (34). The right-hand side follows analogously.

Proof of Proposition 3.2

The proof of completeness in great parts is analogous to the one of Proposition 3.1. In particular, define π, π^x and π^τ in an analogous fashion, but let

$$\nu(E) = \sum_{k\in E'} \beta^{-(k+i)},\ E = (E', i) \subseteq \mathbb{N}_0 \times \{0, 1\}$$

and let $Tf : \mathbb{N}_0 \times \{0,1\} \times [0,\infty)^2 \longrightarrow \mathbb{R}$ with

$$Tf\left(k, i, (\tau, x)\right) = \frac{\partial^k}{\partial\tau^k} D^{(0,i)} f(\tau, x),$$

where $D^{(0,0)} = \mathbf{Id}$.

For a Cauchy sequence in $\tilde{\mathfrak{H}}_{\beta,\gamma}$, say $(f_n)_{n\in\mathbb{N}}$, similarly to before let $f_{n,m}^{0,0} = f_n - f_m$, $f_{n,m}^{k,0} = \frac{\partial^k}{\partial\tau^k} f_{n,m}^{0,0}$ and $f_{n,m}^{k,1} = \frac{\partial^k}{\partial\tau^k} D^{(0,1)} f_{n,m}^{0,0}$, and $Tf_{n,m} \to 0$ in $L^2(\pi\times\nu)$ as $n,m\to\infty$ implies that there exists $g \in L^2(\pi\times\nu)$ with $Tf_n \to g$ in $L^2(\pi\times\nu)$ as $n\to\infty$ and $Tf_{n'} \to g$ $(\pi\times\nu)$-a.e. for some subsequence $(n')_{n\in\mathbb{N}}$. Moreover, we have analogous properties to (30), (31) and (32) in the proof of Proposition 3.2.

For $T, X > 0$, $\tau_0 \in [0,T]$ and $x_0 \in [0,X]$ such that corresponding versions of (31) and (32) are satisfied, we obtain

$$\begin{aligned}
f_{n,m}^{k,0}(\tau,x) &= f_{n,m}^{k,0}(\tau_0,x_0) + \int_{x_0}^{x} D^{(0,1)} f_{n,m}^{k,0}(\tau_0,s)\,ds + \int_{\tau_0}^{\tau} f_{n,m}^{k+1,0}(t,x_0)\,dt \\
&\quad + \int_{\tau_0}^{\tau}\int_{x_0}^{x} D^{(0,1)} f_{n,m}^{k+1,0}(t,s)\,ds\,dt \\
&= f_{n,m}^{k,0}(\tau_0,x_0) + \int_{x_0}^{x} f_{n,m}^{k,1}(\tau_0,s)\,ds + \int_{\tau_0}^{\tau} f_{n,m}^{k+1,0}(t,x_0)\,dt + \int_{\tau_0}^{\tau}\int_{x_0}^{x} f_{n,m}^{k+1,1}(t,s)\,ds\,dt,
\end{aligned}$$

where the latter equality derives from

$$\begin{aligned}
-\int_a^b \phi(s) D^{(0,1)} f_{n,m}^{k,0}(\tau_0,s)\,ds &= \int_a^b \frac{\partial^k}{\partial\tau^k}\phi'(s) f_{n,m}(\tau_0,s)\,ds = -\frac{\partial^k}{\partial\tau^k}\int_a^b \phi(s) D^{(0,1)} f_{n,m}(\tau_0,s)\,ds \\
&= -\int_a^b \phi(s) f_{n,m}^{k,1}(\tau_0,s)\,ds
\end{aligned}$$

for every test-function ϕ by dominated convergence since f and $D^{(0,1)}f$ are continuously k-times differentiable in the first argument. Whence, analogously to the former proof,

$$\begin{aligned}
\left\| f_{n,m}^{k,0} - f_{n,m}^{k,0}(\tau_0,x_0) \right\|_{C([0,T]\times[0,X])} &\le K(\gamma,T,X) \times \left[\left\| f_{n,m}^{k,1}(\tau_0,\cdot) \right\|_{L^2(\pi^x)}^2 + \left\| f_{n,m}^{k+1,0}(\cdot,x_0) \right\|_{L^2(\pi^\tau)}^2 \right. \\
&\qquad \left. + \left\| f_{n,m}^{k+1,1} \right\|_{L^2(\pi)}^2 \right],
\end{aligned}$$

where the right-hand side goes to zero as $n,m\to\infty$. This means that $\left(f_n^{k,0}(\tau,x_0)\right)_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{R}$ and, consequently, $\left(f_n^{k,0}\right)_{n\in\mathbb{N}}$ is Cauchy in $C\left([0,T]\times[0,X]\right)$, i.e. the convergence is uniform on compacts, which in turn implies pointwise convergence (Ass. 3.1.2).

Now,

$$\begin{aligned}
f_{n,m}^{k,1}(\tau,x_0) &= f_{n,m}(\tau_0,x_0) + \int_{\tau_0}^{\tau} f_{n,m}^{k+1,1}(t,x_0)\,dt \\
\Rightarrow \left\| f_{n,m}^{k,l}(\cdot,x_0) \right\|_{C[0,T]} &\le |f_{n,m}(\tau_0,x_0)| + K(\gamma,T,x_0) \left\| f_{n,m}^{k+1,1}(\cdot,x_0) \right\|_{L^2(\pi^\tau)},
\end{aligned}$$

i.e. $f_{n,m}^{k,1}(\cdot,x_0)$ is Cauchy in $C[0,T]$. Therefore,

$$\frac{\partial}{\partial\tau} g((k,i),\cdot) = g((k+1,i),\cdot),\ k\in\mathbb{N}_0,\ i\in\{0,1\},$$

and all there is left to show for $f_n \to g((0,0),\cdot)$ in $\tilde{\mathfrak{H}}_{\beta,\gamma}$ is

$$D^{(0,1)} g((0,0),\cdot) = g((0,1),\cdot).$$

This, however, simply follows from

$$\underbrace{f_n(\tau_0, x)}_{\to g((0,0),(\tau_0,x))} = f_n^{k,0}(\tau_0, x_0) + \int_{x_0}^{x} \underbrace{D^{(0,1)} f_n(\tau_0, s)}_{\to g((0,1),(\tau_0,s))} \, ds$$

and dominated convergence. This shows completeness, and since it is easily seen that $[\cdot,\cdot]_{\beta,\gamma}$ is a scalar product, $\left(\tilde{\mathfrak{H}}_{\beta,\gamma}, [\cdot,\cdot]_{\beta,\gamma}\right)$ is a Hilbert space satisfying Assumption 3.1.1.

Now define $\{S_t\}_{t\geq 0}$ via

$$S_t f(\tau, x) = \begin{cases} f(\tau + t, x - t) & , 0 \leq t \leq x, \\ f(\tau + x, 0) & , \text{otherwise.} \end{cases}$$

Then each S_t is a bounded linear operator on $\tilde{\mathfrak{H}}_{\beta,\gamma}$ and satisfies $S_t\, S_u = S_{t+u}$, $t, u \geq 0$, i.e. $\{S_t\}_{t\geq 0}$ is a semigroup. The strong continuity is a direct consequence of the strong continuity of translation semigroups on L^2-spaces (see e.g. Example I.5.4 in Engel and Nagel (2000)), which completes the proof.