Stochastic Mortality Models:
An Infinite-Dimensional Approach

Stefan Tappe   Stefan Weber

Leibniz Universität Hannover [a]

June 21, 2013

Abstract

Demographic projections of future mortality rates involve a high level of uncertainty and require stochastic mortality models. The current paper investigates forward mortality models driven by a (possibly infinite dimensional) Wiener process and a compensated Poisson random measure. A major innovation of the paper is the introduction of a family of processes called forward mortality improvements which provide a flexible tool for a simple construction of stochastic forward mortality models. In practice, the notion of mortality improvements are a convenient device for the quantification of changes in mortality rates over time that enables, for example, the detection of cohort effects.

We show that the forward mortality rates satisfy Heath-Jarrow-Morton-type consistency conditions which translate to the forward mortality improvements. While the consistency conditions of the forward mortality rates are analogous to the classical conditions in the context of bond markets, the conditions of the forward mortality improvements possess a different structure: forward mortality models include a cohort parameter besides the time horizon; these two dimensions are coupled in the dynamics of consistent models of forwards mortality improvements. In order to obtain a unified framework, we transform the systems of Itô-processes which describe the forward mortality rates and improvements: in contrast to term-structure models, the corresponding stochastic partial differential equations (SPDEs) describe the random dynamics of two-dimensional surfaces rather than curves.

Key words: Mortality, longevity, forward mortality, Heath-Jarrow-Morton, mortality improvements, dynamic point processes, stochastic partial differential equations (SPDEs)

AMS Subject Classification (2010): 97D20, 60H15

1   Introduction

Actuarial mathematics is often a pragmatic and simplified approach to reality. Classical life insurance mathematics, for example, is concerned with the valuation of insurance products and the computation of reserves and is based on the idea of pooling (formalized by the equivalence principle). It requires the availability of suitable projections of the mortality of the insured individuals. In practice, insurance companies typically use deterministic mortality tables which are constructed from past mortality data and include safety margins. For standard life insurance products like term life insurance or annuities, projections are needed that stretch over various decades.

While deterministic mortality tables constitute a rather substantial simplification of reality, insurers and actuaries were always aware of the fact that demographic projections of future mortality rates involve a high level of uncertainty. Instead of attempting to correctly predict future mortality rates, actuarial practice implements a prudent risk management scheme that involves substantial safety margins. While customers of life insurance companies are typically overcharged, the mechanism is fair in the sense that surpluses are redistributed to the insured. Actuarial life tables must not be interpreted as models of actual mortality rates, but rather as specific technical tools inside actuarial practice.

In reality, demographic projections of future mortality rates involve a high level of uncertainty. This has, for example, been observed by Booth (2006). Understanding mortality is thus not the same as analyzing or

---

[a]Leibniz Universität Hannover, Institut für Mathematische Stochastik, Welfengarten 1, 30167 Hannover, Germany. E-mail: Stefan Tappe <tappe@stochastik.uni-hannover.de>, Stefan Weber <sweber@stochastik.uni-hannover.de>.

constructing insurance life tables and requires stochastic models of mortality rates and of mortality projection mechanisms. The current article focuses on stochastic forward mortality models (see e.g. Milevsky & Promislow (2001), Dahl (2004), Miltersen & Persson (2005), Cairns, Blake & Dowd (2006), Bauer (2008), Barbarin (2008), Norberg (2010), Bauer, Benth & Kiesel (2012), Zhu & Bauer (2011), Zhu & Bauer (2012)). These are closely related to intensity models of mortality which were discussed by Biffis (2005), Biffis, Denuit & Devolder (2010), Hainaut & Devolder (2008), Luciano & Vigna (2008), and Schrager (2006), among others. Our main contribution is to provide a mathematically rigorous and transparent framework for this approach that generalizes and substantially clarifies previous contributions in the literature.

Stochastic mortality and mortality projection models can be employed as a framework for analyzing the reliability, robustness and cost of current actuarial practice. They might also improve demographic projections and provide a better basis for the management of mortality and longevity risk. Finally, stochastic mortality models are needed for the computation of the market-consistent value of insurance liabilities – a quantity that is particularly important for management and reporting purposes. Stochastic mortality and mortality projection models constitute also essential ingredients for the construction of various re-insurance or capital market solutions that facilitate the mitigation of mortality and longevity risk. Recent product innovations include mortality swaps, longevity bonds, and q-forwards.

Contribution and Outline: The current paper investigates stochastic forward mortality models. Section 2 provides a motivation for our approach by introducing a dynamic point process model of the mortality of individuals (cf. Brémaud (1981), Bielecki & Rutkowski (2002)). We define forward mortality processes and rates. A major innovation of the paper is the introduction of a family of processes called forward mortality improvements which constitute a flexible tool that enables a simple construction of stochastic forward mortality models. Additionally, in practice, the notion of mortality improvements provides a convenient instrument to quantify changes in mortality rates over time that allows, for example, the detection of cohort effects (Prévôt, Rinke & Stollmann (2011)).

Section 3 provides conditional laws of large numbers as a rational for the significance of forward mortality models. Although implicit in many papers on the subject, to the best of our knowledge, these theorems have never rigorously been proven in the literature. A special case of forward mortality models are intensity-based models which allow an alternative description of their probabilistic dynamics via compensators. This point of view (that is preferred by some authors to the approach taken in the current paper) is explained in Section 4.

The suggested forward mortality models can either be interpreted as describing the real-world (see e.g. Zhu & Bauer (2012)) or risk-neutral dynamics (see e.g. Biffis (2005), Biffis & M. (2006), Biffis et al. (2010)). In both cases, we can identify a martingale condition, see Remark 2.1 below, that implies consistency conditions on the dynamics of any 'code-book' in the sense of Carmona (2007). Section 5 describes consistency conditions for the forward mortality rates and improvements. We also present a unified modeling framework on the basis of SPDEs. The proofs of the results are deferred to the Appendix.

While the consistency conditions of the forward mortality rates are analogous to the classical conditions in the context of bond markets, the conditions of the forward mortality improvements possess a different structure: forward mortality models include a cohort parameter besides the time horizon; these two dimensions are coupled in the dynamics of consistent models of forwards mortality improvements.

In order to obtain a unified framework, we transform the systems of Itô-processes which describe the forward mortality rates and forward mortality improvements. In contrast to term-structure models, the corresponding stochastic partial differential equations (SPDEs) describe the random dynamics of two-dimensional surfaces rather than curves. These surfaces are parametrized by cohorts and time horizons (see also Biffis & M. (2006) for a related study in the context of random fields). Most interesting are consistent models of forward mortality improvements which induce stochastic forward mortality models. Moreover, the shape of the forward mortality surfaces requires Hilbert spaces of functions that are not covered by the literature on interest rate models, see Definition 5.7 and Example 5.8.

Our results are illustrated in the context of a Lévy driven Gompertz-Makeham model of forward mortality in Section 6.

2 Definitions and Elementary Properties

We start by introducing the basic notions of our stochastic mortality model. We denote by $(\Omega, \mathcal{G}, \mathbb{P})$ a sufficiently rich probability space on which all random variables and processes are defined. The probability measure $\mathbb{P}$ can either be interpreted as the real-world measure or a pricing measure depending on the context in which the theory will be applied.

The lifetime of an individual is characterized by its date of birth c and its random time of death. As common in life insurance mathematics, we encode the cohorts of individuals by their age and thus define $x = -c \in \mathbb{R}$ which can be interpreted as the (hypothetical) age at time 0. The death of the individual occurs at a $\mathcal{G}$-measurable random time $\tau^x : \Omega \to (-x, \infty)$. Equivalently, the time of death is described by the survival indicator, a stochastic process that is defined by

$$N_t(x)(\omega) = \begin{cases} 1, & t < \tau^x(\omega), \\ 0, & t \geq \tau^x(\omega) \end{cases} \qquad (\omega \in \Omega, t \in \mathbb{R}_+).$$

An established approach to modeling the probabilistic evolution of the occurrence of events are intensity-based models. A particularly convenient case are Cox process models, see Brémaud (1981), that assume that intensities are driven by stochastic covariates. In the current paper, we focus on such models, but introduce these on a slightly more abstract level as described in Brémaud (1981) and Bielecki & Rutkowski (2002). For a detailed analysis of filtration enlargements and further references see also Jeanblanc, Yor & Chesney (2009).

We restrict attention to the time period $\mathbb{R}_+ = [0, \infty)$. *Systematic information* is modeled by a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ – which is sometimes also called the *background information*. In a classical Cox process model this filtration corresponds to the family of sigma-algebras that are generated by the history of the stochastic covariate processes. Intuitively, $\mathbb{F}$ contains all information that determines the likelihood of death events. As we will see below, it shall, however, be assumed that $\mathbb{F}$ does not include information about the exact times of death of specific individuals. For technical reasons, we assume that $(\Omega, \mathcal{G}, \mathbb{F}, \mathbb{P})$ satisfies the usual conditions.

- The conditional probability that an individual born at time $-x$ survives until time $t \geq 0$ given the background information is described by the $\mathbb{F}$-*survival process* of τ^x:

$$G_t(t, x) := \mathbb{P}(\tau^x > t \mid \mathcal{F}_t), \quad t \geq -x \vee 0.$$

  This process is known to be an important ingredient in the theory of point processes. Under suitable technical assumptions, $G_t(t, x)$ equals the fraction of individuals born at date $-x$ that survive until date t. The precise result will be stated in Theorem 3.4.

- Another object of particular interest for firms that are exposed to mortality and longevity risk – in particular for pension funds and reinsurance companies, cf. Prévôt et al. (2011) – is the best prediction at date t of the fraction of individuals born at date $-x$ that survive until a future date T. Again, in Theorem 3.4 we will show under suitable technical conditions that this prediction equals the $\mathbb{F}$-*forward survival process* that we define as

$$G_t(T, x) := \mathbb{P}(\tau^x > T \mid \mathcal{F}_t) = \mathbb{E}[G_T(T, x) \mid \mathcal{F}_t], \quad T \geq -x \vee t. \tag{1}$$

Remark 2.1. *It is apparent from its definition that, for fixed x and T, the forward survival process $(G_t(T, x))_t$ is a martingale with respect to the probability measure $\mathbb{P}$. This martingale property is very natural:*

- *If $\mathbb{P}$ is the real-world measure, then the random variable $G_t(T, x)$ describes the conditional probability that an individual born at date $-x$ survives until date T given the information at date t; if the available information grows as time t increases, the corresponding process of such conditional probabilities is, of course, a $\mathbb{F}$-martingale. A detailed discussion of this point of view can be found in Zhu & Bauer (2012).*

  In addition, Theorem 3.4 provides a slightly different interpretation of the process. As mentioned before, under technical conditions, $G_t(T, x)$ is the best time-t-prediction of the fraction of individuals born at date $-x$ that survive until the future date T (the survival ratio). If t increases, the available information grows which apparently implies the martingale property.

- *If the money market account is chosen as the numéraire with deterministic interest rates $r(t)$, $t \geq 0$, and $\mathbb{P}$ is a pricing measure, then $\exp\left(-\int_t^T r(s)ds\right) \cdot G_t(T, x)$ can be interpreted as the price at time t of a survivor bond that pays at time T an amount equal to the survival ratio.*

Standard technical tools from the theory of point processes are hazard processes and intensities. We adapt these notions in the context of mortality models. We start with a technical assumption.

Assumption 2.2. *For all $t \in \mathbb{R}_+$, $T \geq -x \vee t$, we have that $G_t(T, x) > 0$.*

Remark 2.3. *Assumption 2.2 guarantees that the random times τ^x are not stopping times with respect to $\mathbb{F}$ and is standard in the context of many intensity models, see e.g. Bielecki & Rutkowski (2002). The filtration $\mathbb{F}$ contains background information about the likelihood of death events, but not their actual occurrence; in the context of Cox processes this typically means that $\mathbb{F}$ is generated by the covariates, but not by the individual death events. Within our context of mortality models, we will actually demonstrate in Section 3 that also predictions of survival ratios based on all available information lead to the $\mathbb{F}$-forward survival process, see Theorem 3.4.*

Note that Assumption 2.2 implies that a maximal age does not exist at which all individuals are necessarily dead. While unrealistic, this limitation is mitigated, if conditional survival properties are very low for old individuals, and does therefore not pose any serious restriction.

Definition 2.4. *Let $-x \in \mathbb{R}$ be arbitrary. The family of $\mathbb{F}$-forward hazard processes is defined by*

$$[0, T] \ni t \mapsto \Gamma_t(T, x) = -\ln G_t(T, x), \quad T \geq -x \vee 0.$$

The family of $\mathbb{F}$-conditional hazard processes is defined by $t \mapsto \Gamma_t(t, x), t \geq -x \vee 0$.

Definition 2.5. *Let $-x \in \mathbb{R}$ be arbitrary.*

(i) *If $t \mapsto \Gamma_t(t, x)$, $t \geq -x \vee 0$, is absolutely continuous with respect to Lebesgue measure, i.e.,*

$$\Gamma_t(t, x) = \Gamma_0(0, x) + \int_{-x \vee 0}^{t} \gamma_s(x) ds \tag{2}$$

for a $\mathbb{F}$-optional process $\gamma(x) = (\gamma_t(x))_{t \geq -x \vee 0}$, then $\gamma(x)$ is called the $\mathbb{F}$-spot mortality rate (or sometimes intensity). In (2), we set $\Gamma_0(0, x) := 0$ for $x \leq 0$.

(ii) *If $T \mapsto \Gamma_t(T, x)$, $T \geq -x \vee t$ for $t \in \mathbb{R}_+$, is absolutely continuous with respect to Lebesgue measure, i.e.,*

$$\Gamma_t(T, x) = \Gamma_t(-x \vee t, x) + \int_{-x \vee t}^{T} \mu_t(s, x) ds \tag{3}$$

for $\mathbb{F}$-optional processes $\mu(T, x) = (\mu_t(T, x))_{t \in [0,T]}$, $T \geq -x \vee 0$, then the processes $\mu(T, x)$ are called the $\mathbb{F}$-forward mortality rates. In (3), we set $\Gamma_t(-x, x) := 0$ for $t \leq -x$.

(iii) *If for $t \in \mathbb{R}_+$ the $\mathbb{F}$-forward mortality rates $\mu(T, x)$ exist for $T \geq -x \vee t$, and if $h \mapsto \mu_t(T + h, x - h)$ is absolutely continuous with respect to Lebesgue measure, i.e.,*

$$\mu_t(T + h, x - h) - \mu_t(T, x) = -\int_0^h j_t(T + u, x - u) du$$

for $\mathbb{F}$-optional processes $j(T, x) = (j_t(T, x))_{t \in [0,T]}$, $T \geq -x \vee 0$, then the processes $j(T, x)$ are called $\mathbb{F}$-forward mortality improvements.

We will investigate the dynamics of the forward mortality rates and improvements in Section 5. The martingale condition in Remark 2.1 imposes restrictions on their evolution that will be studied in detail.

Remark 2.6.

(i) *The conventions $\Gamma_0(0, x) = 0$ for $x \leq 0$ and $\Gamma_t(-x, x) = 0$ for $t \leq -x$ in (2) and (3), respectively, encode that all individuals are supposed to be alive before their date of birth $-x$ with probability 1. By Definition 2.4 the forward hazard process needs thus to be 0 in these cases.*

(ii) *Although we assume only the existence of a $\mathbb{F}$-optional $\mathbb{F}$-spot mortality rate, one could require that the spot mortality rate is predictable. This is implied by the continuity of the process $(\Gamma_t(t, x))_t$. A given $\mathbb{F}$-optional measurable spot mortality rate γ is not necessarily predictable, but defines a measure $\gamma_s(x)(\omega) ds \mathbb{P}(d\omega)$ on the optional σ-algebra $\mathcal{O}$, which contains the predictable σ-algebra $\mathcal{P}$. The Radon-Nikodym-derivative of its restriction to $\mathcal{P}$ with respect to $\mathbb{P}(d\omega) dt$ is a $\mathbb{F}$-predictable spot mortality rate satisfying equation (2).*

(iii) *The spot and forward mortality rates can be interpreted as the infinitesimal rate at which individuals die given the current information. In order to be more precise, if $T \geq -x \vee t$ and the conditions of Definition 2.5 are satisfied, then*

$$\mathbb{P}\left(T < \tau^x \leq T + \epsilon \mid \mathcal{F}_t\right) = \mu_t(T, x)\epsilon + o(\epsilon) \quad \text{as } \epsilon \to 0.$$

4

(iv) The $\mathbb{F}$-forward mortality improvements $j(T, x)$ quantify the infinitesimal improvements of $\mathbb{F}$-forward mortality rates across cohorts. Intuitively, the increment

$$\mu_t(T + du, x - du) - \mu_t(T, x) = -j_t(T, x)du$$

describes the changes of forward mortalities for two cohorts with identical age at two time horizons; cohort $x - du$ is of age $T + x$ at time $T + du$, while cohort x is of the same age at time T. The forward mortality improvement $j_t(T, x)$ is positive, if the forward mortality rate decreases, and negative, if the forward mortality rate increases.

Stochastic dynamics of the $\mathbb{F}$-forward mortality improvements j can capture future random cohort effects. A framework for modeling their stochastic evolution is discussed in Section 5 below.

3 A Conditional Law of Large Numbers

The principle of pooling is key to insurance mathematics. It states that in a very large population idiosyncratic risk per insured almost vanishes. This is traditionally used as the basis for the computation of risk premia of insurance products. In this section we apply the idea of pooling within the context of our model and prove a conditional law of large numbers showing that the $\mathbb{F}$-survival and $\mathbb{F}$-forward survival processes capture the systematic risk associated with stochastic mortality.

We fix a date of birth $-x \in \mathbb{R}$ and consider a large homogeneous family of individuals born at this date. Our goal is to compute the fraction of individuals alive at a future date t as well as the best time-t-prediction of the fraction of individuals alive at time $T > t$; the best prediction will be based on the full information available at t that does not only include the background information, but also the information about all death events. We first state an assumption that formalizes that we are considering a homogeneous population.

Assumption 3.1. Let τ^x and $\hat{\tau}^x$ be the times of death of two arbitrary individuals born at time $-x \in \mathbb{R}$. Then $\mathbb{P}(\tau^x > t \,|\, \mathcal{F}_t) = \mathbb{P}(\hat{\tau}^x > t \,|\, \mathcal{F}_t)$ almost surely for all $t \in \mathbb{R}_+$.

Limit theorems for a population require a large pool of individuals. We thus consider a countably infinite family of individuals born at date $-x$.

Remark 3.2. *Mathematically, the existence of an infinite, but countable collection of random times on a sufficiently rich probability space $(\Omega, \mathcal{G}, \mathbb{P})$ with given hazard process satisfying Assumption 3.1 can easily be shown using the canonical construction of random times, a notion well-known in the literature on reduced-form credit risk models. Suppose that $\Gamma_t(t, x)$ is an increasing $\mathbb{F}$-adapted process. Letting ε^n be a sequence of independent unit exponentially distributed random variables independent of $\mathcal{F}_\infty := \bigvee_{t \in \mathbb{R}_+} \mathcal{F}_t$, the random times*

$$\tau^{x,n} := \inf\{t : \Gamma_t(t, x) > \varepsilon^n\}$$

are conditionally independent given $\mathcal{F}_\infty$, each with hazard process $\Gamma_t(t, x)$.

Definition 3.3. *Letting $x \in \mathbb{R}$, a family of associated death times $(\tau^{x,n})_{n \in \mathbb{N}}$ is called $\mathbb{F}$-doubly stochastic conditionally independent ($\mathbb{F}$-DSCI), if*

(i) *the sequence $(\tau^{x,n})_{n \in \mathbb{N}}$ is doubly stochastic, i.e.,*

$$\mathbb{P}(\tau^{x,n} > t \,|\, \mathcal{F}_t) = \mathbb{P}(\tau^{x,n} > t \,|\, \mathcal{F}_\infty) \text{ for all } t \in \mathbb{R}_+, n \in \mathbb{N};$$

(ii) *the sequence $(\tau^{x,n})_{n \in \mathbb{N}}$ is $\mathcal{F}_\infty$-conditionally independent, i.e., for any finite $J \subseteq \mathbb{N}$, $t_j \in \mathbb{R}_+$, $j \in J$:*

$$\mathbb{P}(\tau^{x,j} > t_j \text{ for all } j \in J \,|\, \mathcal{F}_\infty) = \prod_{j \in J} \mathbb{P}(\tau^{x,j} > t_j \,|\, \mathcal{F}_\infty).$$

This definition can canonically be extended to families of death times of countably many different cohorts.

Property (i) of Definition 3.3 states that the probability of death of an individual up to time t depends only on the background information up to time t, but not on background information arriving later. In the special context of Cox process intensities this means, for example, that the intensity at time t is a function of the paths of the factor processes only up to time t.

Property (ii) formalizes that death times are independent given the background information. Note that this excludes contagion effects in the sense of local or global (mean-field) interaction. This assumption is relatively innocent in the context of mortality modeling, as long as large-scale epidemic outbreaks are neglected.

Theorem 3.4. *Letting $x \in \mathbb{R}$, we denote by $(N^n(x))_{n \in \mathbb{N}}$ the survival indicators associated to a family of individuals born at date $-x$ with $\mathbb{F}$-DSCI death times. Then for all $t \geq -x \vee 0$:*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} N^n(x)_t = G_t(t,x) \quad \mathbb{P}\text{--almost surely.}$$

Setting $\mathcal{G}_t := \sigma\{N^n(x)_s : s \leq t, n \in \mathbb{N}\} \vee \mathcal{F}_t$, $t \in \mathbb{R}_+$, $\mathbb{G} = (\mathcal{G}_t)_{t \in \mathbb{R}_+}$ is the full information filtration. Then for all $t \in \mathbb{R}_+$, $T \geq -x \vee t$:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[N^n(x)_T \,|\, \mathcal{G}_t] = G_t(T,x) \quad \mathbb{P}\text{--almost surely.} \tag{4}$$

Proof. See Appendix A. $\qquad \square$

Theorem 3.4 is a law of large numbers. The $\mathbb{F}$-conditional survival process describes mortality on the aggregate level of large populations. The quantity $G_t(t,x)$ equals the fraction of individuals born at time $-x$ that survive until t. The $\mathbb{F}$-forward survival process describes the best predictions of these fractions. As defined in the Theorem, $\mathbb{G} = (\mathcal{G}_t)_{t \in \mathbb{R}_+}$ models the full information that is available at time t. It includes both the background information as well as information about the occurrence of all death events up to time t. The quantity $G_t(T,x)$ is the time t best estimate of the fraction of individuals born at time $-x$ that survive until T. We emphasize that $G_t(T,x)$ is $\mathcal{F}_t$-measurable, thus depends only on the background information. The rational behind this property is pooling: idiosyncratic risks are not relevant anymore on the aggregate level.

Theorem 3.4 does also allow us to extend the martingale property of $(G_t(T,x))_t$ from the background filtration $\mathbb{F}$ to the full filtration $\mathbb{G}$. The bounded convergence theorem allows us to interchange the order of the expectation and the limit in (4). This shows that $G_t(T,x) = \mathbb{E}[G_T(T,x)|\mathcal{G}_t]$ proving the $\mathbb{G}$-martingale property.

Finally observe that the stochastic process $(G_t(t,x))_{t \geq -x \vee 0}$ is $\mathbb{P}$--almost surely decreasing by Theorem 3.4. This is also apparent from Definition 3.3(i), since $G_t(t,x) = \mathbb{P}(\tau^{x,1} > t \,|\, \mathcal{F}_t) = \mathbb{E}(N^1(x)_t \,|\, \mathcal{F}_\infty)$ and $(N^1(x)_t)_{t \in \mathbb{R}_+}$ is decreasing.

Remark 3.5. *The laws of large numbers refer to large homogeneous populations with cohort x fixed. However, in applications, liabilities and mortality derivatives are usually linked to inhomogeneous pools that include, for example, various cohorts. Suppose that we are interested in market-consistent valuation and that $\mathbb{P}$ is interpreted as a reference measure. In this case, the relevant cohorts need to be weighted in a suitable way. These weights can easily be encoded by Borel measures on the real line, as explained in Section 4 of Biffis & M. (2006).*

4 Compensators

Although quite convenient, the approach described in Section 2 is not always employed in the literature on doubly-stochastic point processes. An alternative procedure for describing the probabilistic properties of the death times considers the $\mathbb{G}$-compensators. In order to facilitate a translation between both concepts, we briefly summarize some basic relationships.

For simplicity, we restrict attention to individuals born after time 0, i.e. we assume that $-x \geq 0$. Let $(\tau^{x,n})_{n \in \mathbb{N}}$ again be a $\mathbb{F}$-DSCI family of death times with $-x \in \mathbb{R}_+$. The corresponding survival indicators are denoted by $(N^n(x))_{n \in \mathbb{N}}$. As defined in Theorem 3.4, $\mathbb{G}$ denotes the full information filtration. We start by defining the notion of a compensator. Its existence follows from the Doob-Meyer-decomposition theorem. The compensator is unique, up to indistinguishability.

Definition 4.1. *A $\mathbb{G}$-predictable, right-continuous, increasing process $A^n(x)$ is a $\mathbb{G}$-compensator of $\tau^{x,n}$, if $A^n(x)_t = 0$, $t \leq -x$, and*
$$1 - N^n(x)_t - A^n(x)_t, \quad t \geq -x,$$
is a $\mathbb{G}$-martingale.

The relationship between the $\mathbb{G}$-compensators of all death times and the $\mathbb{F}$-survival process is now described by the following proposition.

Proposition 4.2. *Assume that $(\tau^{x,n})_{n \in \mathbb{N}}$ is a $\mathbb{F}$-DSCI family of death times with $-x \in \mathbb{R}_+$. Then there exists a $\mathbb{F}$-predictable, right-continuous, increasing process $\Lambda(x)$ with the following property:*

$$\Lambda(x)^{\tau^{x,n}} = A^n(x) \text{ for all } n \in \mathbb{N},$$

where $\Lambda(x)^{\tau^{x,n}}$ signifies the process $\Lambda(x)$ stopped at $\tau^{x,n}$. The process $\Lambda(x)$ is unique, up to indistinguishability.

In other words, $\Lambda(x)$ is a $\mathbb{F}$-predictable, right-continuous, increasing process such that

$$1 - N^n(x)_t - \Lambda(x)_{t \wedge \tau^{x,n}}, \quad t \geq -x,$$

is a $\mathbb{G}$-martingale for all $n \in \mathbb{N}$, and $\Lambda(x)_t = 0$, $t \leq -x$. This means that $\Lambda(x)$ is a $(\mathbb{F}, \mathbb{G})$-martingale hazard process of $\tau^{x,n}$ for any $n \in \mathbb{N}$ according to Definition 6.1.1 in Bielecki & Rutkowski (2002).

Letting $(\tilde{F}_t(x))_t$ be the unique $\mathbb{F}$-predictable, increasing process with $\tilde{F}_t(x) = 0$, $t \leq -x$, such that

$$(1 - G_t(t,x) - \tilde{F}_t(x))_{t \geq -x}$$

is a $\mathbb{F}$-martingale, then $\Lambda(x)$ is given by

$$\Lambda(x)_t = \int_{(-x,t]} \frac{1}{G_{t-}(t-,x)} d\tilde{F}_t(x). \tag{5}$$

If $t \mapsto G_t(t,x)$ is almost surely continuous, then $\tilde{F}_t(x) = 1 - G_t(t,x)$, thus

$$\Lambda(x)_t = -\int_{(-x,t]} \frac{1}{G_t(t,x)} dG_t(t,x).$$

Proof. See Appendix B. $\qquad\square$

The following proposition states that under regularity conditions the joint $(\mathbb{F}, \mathbb{G})$-martingale hazard process coincides with the $\mathbb{F}$-conditional hazard process.

Proposition 4.3. *Let $-x \in \mathbb{R}_+$. If the $\mathbb{F}$-conditional hazard process $(\Gamma_t(t,x))_{t \geq -x}$ is continuous, then $\Gamma_t(t,x) = \Lambda(x)_t$ almost surely for $t \geq -x$.*

Proof. See Appendix B. $\qquad\square$

5 Infinite-Dimensional Formulation

In this section, we provide a model for the stochastic evolution of the $\mathbb{F}$-forward mortality rates μ and the $\mathbb{F}$-forward mortality improvements j on the time interval $\mathbb{R}_+ = [0, \infty)$. We will define these quantities for each fixed (T, x) as an Itô process driven by a (possibly infinite dimensional) Wiener process and a compensated Poisson random measure. In a second step, we shall transform these systems of Itô processes to obtain a single infinite-dimensional stochastic process having values in an adequate function space. In the framework of bond markets, this idea originates from Musiela (1993). As we pointed out in Remark 2.1, the $\mathbb{F}$-survival processes $G(T, x)$ must necessarily be martingales. This property leads to consistency conditions which we will describe for all considered quantities.

We shall now present the general stochastic framework. We begin with the driving Wiener process W taking values in some separable Hilbert space U with covariance operator $Q \in L(U)$. For details we refer to Chapter 4 in Da Prato & Zabczyk (1992). The standard space of stochastic integrands with respect to W consists of stochastic processes with values in the space of Hilbert-Schmidt operators $L_2^0(H) := L_2(U_0, H)$ from $U_0 := Q^{1/2}(U)$ into some separable Hilbert space H. The space H is the state-space of the model; adequate choices will be discussed later. The (possibly infinite-dimensional) integrands can be decomposed into one-dimensional components on the basis of the spectral decomposition of Q. To be precise, we denote by $(\lambda_k)_{k \in \mathbb{N}} \subset (0, \infty)$ the sequence of non-zero eigenvalues of Q and by $(e_k)_{k \in \mathbb{N}}$ the corresponding orthonormal basis of eigenvectors. Then the one-dimensional components of a $L_2^0(H)$-valued integrand Φ are given by

$$\Phi^k := \Phi(\sqrt{\lambda_k} e_k), \quad k \in \mathbb{N}. \tag{6}$$

As a second stochastic driver of the evolution we introduce a time-homogeneous Poisson random measure $\mathfrak{p}$ that allows to include jumps. For details we refer to Jacod & Shiryaev (2003, Def. II.1.20). The mark space $(E, \mathcal{E})$ of the Poisson random measure $\mathfrak{p}$ is a measurable space. For technical reasons we assume that $(E, \mathcal{E})$ is a Blackwell space (see Dellacherie & Meyer (1982) or Getoor (1975)). Blackwell spaces include Polish spaces as a special case. The compensator of $\mathfrak{p}$ is of the form $dt \otimes \nu(d\xi)$ for a σ-finite measure ν on $(E, \mathcal{E})$. For further reference, we set $\mathcal{L}_\nu^2(H) := \mathcal{L}^2(E, \mathcal{E}, \nu; H)$.

We present our main results in the following Sections 5.1–5.4. For convenience of the reader, technical assumptions and proofs are deferred to Appendix C. An illustrative example is provided in Section 6.

5.1 Consistent HJM type dynamics of the forward mortality rates

In this section, we will specify the dynamics of the $\mathbb{F}$-forward mortality rates $\mu_t(T, x)$. We begin by choosing a suitable parameter domain Θ for the values of (t, T, x). The variable t represents the running time, the parameter $-x$ signifies the date of birth of an individual, and T denotes a future date. Natural parameter restrictions are thus provided by $0 \leq t \leq T$, $T \geq -x$. The relevant domain $\Theta \subset \mathbb{R}_+^2 \times \mathbb{R}$ is hence given by

$$\Theta := \{(t, T, x) \in \mathbb{R}_+^2 \times \mathbb{R} : (T, x) \in \Xi \text{ and } t \in [0, T]\}, \tag{7}$$

with

$$\Xi := \{(T, x) \in \mathbb{R}_+ \times \mathbb{R} : -x \leq T\}. \tag{8}$$

Next, we specify the stochastic dynamics of the forward mortality rates. Suppose that $\mu_0 : \Xi \to \mathbb{R}$ is an initial surface of $\mathbb{F}$-forward mortality rates. We suppose that the $\mathbb{F}$-forward mortality rates $\mu(T, x)$, $(T, x) \in \Xi$, follow an Itô process:

$$
\begin{aligned}
\mu_t(T, x) := \mu_0(T, x) &+ \int_0^t \alpha_s(T, x) ds + \int_0^t \sigma_s(T, x) dW_s \\
&+ \int_0^t \int_E \delta_s(T, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds), \quad t \in [0, T].
\end{aligned}
\tag{9}
$$

Here, $\alpha : \Omega \times \Theta \to \mathbb{R}$, $\sigma : \Omega \times \Theta \to L_2^0(\mathbb{R})$ and $\delta : \Omega \times \Theta \times E \to \mathbb{R}$ are stochastic processes that satisfy the technical Assumption C.1 (see below) that guarantees that all stochastic integrals in equation (9) exist.

For fixed $x \in \mathbb{R}$ we introduce the $\mathbb{F}$-spot mortality rates $\gamma(x)$ as

$$\gamma_t(x) := \mu_t(t, x) \mathbb{1}_{\{t \geq -x\}}, \quad t \geq 0. \tag{10}$$

The spot mortality rates follow an Itô process as characterized in the following proposition.

Proposition 5.1. *We suppose that:*

- *For all $(t, x) \in \Xi$ and $\xi \in E$ the mappings $T \mapsto \mu_0(T, x)$, $T \mapsto \alpha_t(T, x)$, $T \mapsto \sigma_t(T, x)$ and $T \mapsto \delta_t(T, x, \xi)$ are differentiable on their domains.*

- *Assumption C.1 holds as stated as well as for the derivatives $\partial_T \mu_0$, $\partial_T \alpha$, $\partial_T \sigma$ and $\partial_T \gamma$ in place of μ_0, α, σ and γ.*

Then, for each $x \in \mathbb{R}_+$ the process $\gamma(x)$ of $\mathbb{F}$-spot mortality rates is an Itô process of the form

$$
\begin{aligned}
\gamma_t(x) = \gamma_0(x) &+ \int_0^t \zeta_u(x) du + \int_0^t \sigma_u(u, x) dW_u \\
&+ \int_0^t \int_E \delta_u(u, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds), \quad t \geq 0,
\end{aligned}
$$

where the process $\zeta : \Omega \times \Xi \to \mathbb{R}$ is given by

$$
\begin{aligned}
\zeta_u(x) = \alpha_u(u, x) &+ \partial_u \mu_0(u, x) + \int_0^u \partial_u \alpha_s(u, x) ds + \int_0^u \partial_u \sigma_s(u, x) dW_s \\
&+ \int_0^u \int_E \partial_u \delta_s(u, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds), \quad u \geq 0.
\end{aligned}
$$

Proof. The proof is analogous to that of Filipović (2009, Prop. 6.1), and therefore omitted. □

As pointed out in Remark 2.1, the forward survival process satisfies a martingale condition which is key to the following analysis. The condition implies a crucial consistency condition for the dynamics:

Definition 5.2. *The $\mathbb{F}$-forward mortality rates μ are called* consistent, *if for all $(T, x) \in \Xi$ the $\mathbb{F}$-survival process $G(T, x)$ is a $\mathbb{F}$-martingale.*

The following theorem provides a precise criterion for the consistency of the mortality rates. Technical assumptions are again deferred to Appendix C. In particular, we require an exponential integrability condition on the Lévy measure ν which is stated in Assumption C.5.

Theorem 5.3. *Suppose that Assumptions C.1 and C.5 are satisfied. Then the $\mathbb{F}$-forward mortality rates μ are consistent if and only if*

$$
\alpha_t(T, x) = \sum_{k \in \mathbb{N}} \sigma_t^k(T, x) \int_{-x \vee t}^{T} \sigma_t^k(u, x) du
$$

$$
- \int_E \delta_t(T, x, \xi) \left[\exp\left(-\int_{-x \vee t}^{T} \delta_t(u, x, \xi) du \right) - 1 \right] \nu(d\xi) \tag{11}
$$

$$
\text{for all } (T, x) \in \Xi \text{ with } T \geq t, \quad d\mathbb{P} \otimes dt\text{-almost surely on } \Omega \times \mathbb{R}_+.
$$

Proof. See Appendix C.1. □

Remark 5.4. *Theorem 5.3 provides a criterion for the consistency of the $\mathbb{F}$-forward mortality rates μ showing that the drift α is completely determined by the volatilities σ and γ. This resembles the HJM drift condition for bond markets, cf. Heath, Jarrow & Morton (1992) for the Wiener driven case, and Björk, Di Masi, Kabanov & Runggaldier (1997) for the general situation with an additional Poisson random measure. Note that our mortality model is specified with respect to the reference measure $\mathbb{P}$ while the HJM drift condition for interest rate models holds only with respect to an equivalent martingale measure. In the current paper, $\mathbb{P}$ may also play the role of the statistical measure.*

5.2 Consistent HJM type dynamics of the forward mortality improvements

In this section we specify the dynamics of the forward mortality improvements and derive consistency conditions. Suppose that $j_0 : \Xi \to \mathbb{R}$ is an initial surface of $\mathbb{F}$-forward mortality improvements. We assume that the $\mathbb{F}$-forward mortality improvements $j(T, x)$, $(T, x) \in \Xi$, follow an Itô process:

$$
j_t(T, x) := j_0(T, x) + \int_0^t a_s(T, x) ds + \int_0^t b_s(T, x) dW_s
$$

$$
+ \int_0^t \int_E c_s(T, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi) ds), \quad t \in [0, T]. \tag{12}
$$

Here, $a : \Omega \times \Theta \to \mathbb{R}$, $b : \Omega \times \Theta \to L_2^0(\mathbb{R})$, and $c : \Omega \times \Theta \times E \to \mathbb{R}$ are stochastic processes satisfying the technical Assumption C.8 (see below). Assumption C.8 ensures that the stochastic integrals in Definition (12) exist.

Letting $\gamma_0 : \mathbb{R}_+ \to \mathbb{R}$ be an initial curve of $\mathbb{F}$-spot mortality rates, we extend this curve to an initial surface $\mu_0 : \Xi \to \mathbb{R}$ of $\mathbb{F}$-forward mortality rates by setting

$$
\mu_0(T, x) := \gamma_0(T + x) - \int_0^T j_0(u, T + x - u) du. \tag{13}
$$

Now, using the dynamics of the forward mortality improvements as the starting point, we redefine the stochastic processes $\alpha : \Omega \times \Theta \to \mathbb{R}$, $\sigma : \Omega \times \Theta \to L_2^0(\mathbb{R})$ and $\delta : \Omega \times \Theta \times E \to \mathbb{R}$ as

$$
\alpha_t(T, x) := -\int_t^T a_t(u, T + x - u) du, \quad \sigma_t(T, x) := -\int_t^T b_t(u, T + x - u) du \quad \text{and}
$$

$$
\delta_t(T, x, \xi) := -\int_t^T c_t(u, T + x - u, \xi) du. \tag{14}
$$

For each $(T, x) \in \Xi$ we redefine the $\mathbb{F}$-forward mortality rates $\mu(T, x)$ by (9). Again, we impose an exponential integrability condition on the Lévy measure ν formalized by Assumption C.9.

Theorem 5.5. *Suppose that Assumption C.8 is satisfied. Then the following statements are true:*

(i) For all $x \in \mathbb{R}_+$ the $\mathbb{F}$-spot mortality rates $\gamma(x)$ follow the dynamics

$$
\gamma_t(x) = \mu_0(t, x) + \int_0^t \alpha_s(t, x) ds + \int_0^t \sigma_s(t, x) dW_s
$$

$$
+ \int_0^t \int_E \delta_s(t, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi) ds), \quad t \geq -x \vee 0. \tag{15}
$$

9

(ii) For all $(T, x) \in \Xi$ we have

$$\mu_t(T, x) = \gamma_t(T + x - t) - \int_t^T j_t(u, T + x - u)du, \quad t \in [0, T]. \tag{16}$$

(iii) If, in addition, Assumption C.9 is satisfied, and the drift a is given by

$$
\begin{aligned}
a_t(T, x) = &-\sum_{k \in \mathbb{N}} \left(\int_t^T b_t^k(u, T + x - u)du \right) \left(\int_{-x \vee t}^T b_t^k(u, x)du \right) \\
&- \sum_{k \in \mathbb{N}} b_t^k(T, x) \int_{-x \vee t}^T \int_t^u b_t^k(v, u + x - v)dvdu \\
&- \int_E \left(\int_t^T c_t(u, T + x - u, \xi)du \right) \left(\int_{-x \vee t}^T c_t(u, x, \xi)du \right) \\
&\quad \times \exp \left(\int_{-x \vee t}^T \int_t^u c_t(v, u + x - v, \xi)dvdu \right) \nu(d\xi) \\
&- \int_E c_t(T, x, \xi) \left[\exp \left(\int_{-x \vee t}^T \int_t^u c_t(v, u + x - v, \xi)dvdu \right) - 1 \right] \nu(d\xi),
\end{aligned}
\tag{17}
$$

then the $\mathbb{F}$-forward mortality rates μ are consistent.

Proof. See Appendix C.2. $\qquad \square$

Remark 5.6. *The previous results entail the following relations between the initial surfaces and the volatilities of consistent dynamic evolutions of forward mortality rates and improvements:*

- *For a given initial surface j_0 of $\mathbb{F}$-forward mortality improvements and an initial curve γ_0 of $\mathbb{F}$-spot mortality rates, we can compute the initial surface μ_0 of $\mathbb{F}$-forward mortality rates by (13).*

- *Conversely, for a given initial surface μ_0 of $\mathbb{F}$-forward mortality rates, we can compute the initial surface j_0 of $\mathbb{F}$-forward mortality improvements as*

$$j_0(T, x) := -(\partial_T - \partial_x)\mu_0(T, x).$$

- *For given volatilities a, b, c in (12) we can compute the volatilities α, σ, δ in (9) by (14).*

- *For given volatilities α, σ, δ in (9) with $\alpha_t(t, x) = \sigma_t(t, x) = \delta_t(t, x, \xi) = 0$ we can compute the volatilities a, b, c in (12) by*

$$
\begin{aligned}
&a_t(T, x) := -(\partial_T - \partial_x)\alpha_t(T, x), \quad b_t(T, x) := -(\partial_T - \partial_x)\sigma_t(T, x) \quad and \\
&c_t(T, x, \xi) := -(\partial_T - \partial_x)\delta_t(T, x, \xi).
\end{aligned}
$$

- *Note that, for consistency, the drift term a is given by (17), and the drift term α is given by (11).*

5.3 Consistent Musiela type dynamics of the forward mortality rates

In this section, we shall transform the parameter domain Θ in order to get a unified framework. The $\mathbb{F}$-forward mortality rates will be described by one infinite-dimensional stochastic process with values in a Hilbert space H consisting of functions $h : \Xi \to \mathbb{R}$.
For this procedure, we introduce the mapping

$$\phi : \Theta \to \mathbb{R}_+ \times \Xi, \quad \phi(t, T, x) = (t, T - t, x + t),$$

which is bijective with inverse

$$\phi^{-1} : \mathbb{R}_+ \times \Xi \to \Theta, \quad \phi^{-1}(t, s, y) = (t, s + t, y - t).$$

Definition 5.7. *We call a separable Hilbert space $(H, \| \cdot \|)$ a* forward mortality space, *if the following conditions are satisfied:*

(i) *H consists of continuous functions $h : \Xi \to \mathbb{R}$.*
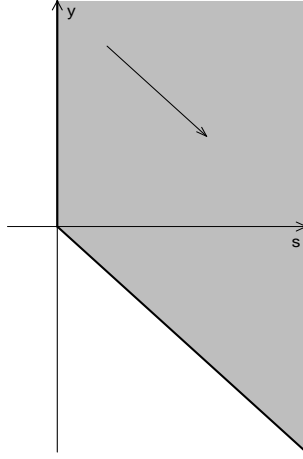
Figure 1: The domain Ξ and the action of the shift semigroup $(S_t)_{t \geq 0}$.

(ii) For each $(s, y) \in \Xi$ the point evaluation $\ell_{(s,y)} : H \to \mathbb{R}$, $\ell_{(s,y)}(h) := h(s, y)$ is a continuous linear functional.

(iii) For each bounded Borel set $B \subset \Xi$ there is a constant $C > 0$ such that

$$\|\ell_{(s,y)}\| \leq C \quad \text{for all } (s, y) \in B. \tag{18}$$

(iv) The shift semigroup $(S_t)_{t \geq 0}$ given by

$$S_t h(s, y) := h(s + t, y - t), \quad (s, y) \in \Xi \tag{19}$$

is a C_0-semigroup on H.

The domain Ξ of functions in a forward mortality space is shown in Figure 1. The variable s represents the length of the time horizon for which survival probabilities are computed; y denotes the current age of individuals of a given cohort. The sum $s + y$ is the age of individuals of cohort y at the end of the time horizon s. Note that the current age y is allowed to be negative, labeling individuals of future generations for which forecasts are made. For each $(s, y) \in \Xi$ we necessarily have $s + y \geq 0$ by the choice of the domain Ξ; i.e. at the end of the prediction time horizon, individuals for which predictions are made will already have been born.

Figure 1 also illustrates the action of the shift semigroup, where the vector $(s, y) \in \Xi$ is mapped to $(s + t, y - t) \in \Xi$. For positive t this mapping shifts quantities to younger generations with the same age.

An example of a forward mortality space can be constructed as follows:

Example 5.8. Let $w_1, w_2 : \mathbb{R}_+ \to (0, \infty)$ and $w_3 : \mathbb{R}_+^2 \to (0, \infty)$ be strictly positive, continuous weight functions. A possible specific example is provided by the following parametric choice:

$$\begin{aligned} w_1(s) &= e^{-\beta s}, \quad w_2(z) = e^{-\beta z} \quad \text{and} \\ w_3(s, z) &= e^{-\beta(s+z)} \quad \text{for some constant } \beta > 0. \end{aligned} \tag{20}$$

Let H be the linear space consisting of all functions $h : \Xi \to \mathbb{R}$ satisfying the following conditions:

- For all $s \in \mathbb{R}_+$ the mappings $z \mapsto h(s, z - s)$, $z \mapsto \partial_s h(s, z - s)$ are absolutely continuous, and for all $z \in \mathbb{R}$ the mappings $s \mapsto h(s, z - s)$, $s \mapsto \partial_z h(s, z - s)$ are absolutely continuous (and hence, almost everywhere differentiable).

- We have $\partial_{sz} h(s, z - s) = \partial_{zs} h(s, z - s)$ for almost all $(s, z) \in \mathbb{R}_+^2$.

- We have

$$\begin{aligned} \|h\| := \Bigg(& |h(0,0)|^2 + \int_0^\infty |\partial_s h(s, -s)|^2 w_1(s) ds + \int_0^\infty |\partial_z h(0, z)|^2 w_2(z) dz \\ & + \int_0^\infty \int_0^\infty |\partial_{sz} h(s, z - s)|^2 w_3(s, z) dz ds \Bigg)^{1/2} < \infty. \end{aligned} \tag{21}$$

11

Then $(H, \| \cdot \|)$ is a forward mortality space. The arguments are similar to those from Section 5 in Filipović (2001): A straightforward calculation shows that for all $h \in H$ and $s_0, z_0 \in \mathbb{R}_+$ we have

$$h(s_0, z_0 - s_0) = h(0,0) + \int_0^{s_0} \partial_s h(s, -s) ds + \int_0^{z_0} \partial_z h(0, z) dz \tag{22}$$
$$+ \int_0^{s_0} \int_0^{z_0} \partial_{sz} h(s, z - s) dz ds.$$

The representation (22) shows that every $h \in H$ is a continuous function. Moreover, by (22) and the Cauchy-Schwarz inequality, we obtain (18). Using the estimate (18), for each $h \in H$ with $\|h\| = 0$ we have $h = 0$, showing that $\| \cdot \|$ is a norm (not just a seminorm) on H. By Definition (21) of the norm $\| \cdot \|$ we have

$$H \cong \mathbb{R} \times L^2(\mathbb{R}_+) \times L^2(\mathbb{R}_+) \times L^2(\mathbb{R}_+^2),$$

showing that $(H, \| \cdot \|)$ is a separable Hilbert space. Finally, similar arguments as in Filipović (2001, Section 5) show that $(S_t)_{t \geq 0}$ is a C_0-semigroup on H.

From now on, let H be a forward mortality space. Denoting by A the infinitesimal generator of the shift semigroup $(S_t)_{t \geq 0}$, for each $h \in \mathcal{D}(A)$ we obtain

$$Ah(s,y) = \lim_{t \to 0} \frac{S_t h(s,y) - h(s,y)}{t} = \lim_{t \to 0} \frac{h((s,y) + t(1,-1)) - h(s,y)}{t}$$
$$= Dh(s,y)(1,-1) = (\partial_s - \partial_y) h(s,y), \quad (s,y) \in \Xi.$$

Therefore, we have $A = \partial_s - \partial_y$ and

$$\mathcal{D}(\partial_s - \partial_y) \subset \{ h \in H : (\partial_s - \partial_y) h \text{ exists with } (\partial_s - \partial_y) h \in H \}.$$

Let $\mu_0 \in H$ be an initial surface of forward mortality rates. We define the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ as the H-valued process

$$\bar{\mu}_t := S_t \mu_0 + \int_0^t S_{t-s} \bar{\alpha}_s ds + \int_0^t S_{t-s} \bar{\sigma}_s dW_s \tag{23}$$
$$+ \int_0^t \int_E S_{t-s} \bar{\delta}_s(\xi) (\mathfrak{p}(ds, d\xi) - \nu(d\xi) ds), \quad t \geq 0.$$

Here, $\bar{\alpha} : \Omega \times \mathbb{R}_+ \to H$, $\bar{\sigma} : \Omega \times \mathbb{R}_+ \to L_2^0(H)$ and $\bar{\delta} : \Omega \times \mathbb{R}_+ \times E \to H$ are stochastic processes satisfying Assumption C.11 (see below) which ensures that all stochastic integrals in the variation of constants formula (23) exist.

Remark 5.9. *As we have seen in Section 2, forward mortality rates should be positive processes. In the context of the current paper, we do, however, not impose a general positivity assumption on the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ defined in (23) resp. on the $\mathbb{F}$-forward mortality rates μ defined in (9). Positivity depends on the choice of the specific model. Even if positivity is violated, in practice, one can ensure by an adequate choice of the parameters that the mortality rates become negative only with low probability. Such a model may then be regarded as a good approximation of reality; this modeling philosophy conceptually parallels the approach of the Vasiček model and its Hull-White extension in the context of interest rates.*

In order to rigorously investigate positivity in a general framework, one would, as in Filipović, Tappe & Teichmann (2010b) for classical HJM models, define the closed, convex cone

$$P := \bigcap_{(s,y) \in \Xi} \{ h \in H : h(s,y) \geq 0 \}$$

of nonnegative mortality surfaces and derive appropriate conditions for stochastic invariance of P for (23).

Definition 5.10. *The $\mathbb{F}$-forward mortality rates $\bar{\mu}$ are called* consistent, *if the transformed $\mathbb{F}$-forward mortality rates $\mu := \bar{\mu} \circ \phi$ are consistent in the sense of Definition 5.2.*

In the following theorem we impose again an exponential integrability condition on the Lévy measure ν.

Theorem 5.11. *Suppose that Assumptions C.11 and C.12 are fulfilled. Then the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ are consistent if and only if*

$$\bar{\alpha}_t(s,y) = \sum_{k \in \mathbb{N}} \bar{\sigma}_t^k(s,y) \int_{-y \vee 0}^s \bar{\sigma}_t^k(u,y) du$$
$$- \int_E \bar{\delta}_t(s,y,\xi) \left[\exp \left(- \int_{-y \vee 0}^s \bar{\delta}_t(u,y,\xi) du \right) - 1 \right] \nu(d\xi) \tag{24}$$

for all $(s,y) \in \Xi$, $d\mathbb{P} \otimes dt$–almost surely on $\Omega \times \mathbb{R}_+$.

Proof. See Appendix C.3. □

Remark 5.12. *In the spirit of Da Prato & Zabczyk (1992), the process $\bar{\mu}$ of $\mathbb{F}$-forward mortality rates in (23) is a mild solution to the stochastic partial differential equation (SPDE)*

$$\begin{cases} d\bar{\mu}_t &= \big((\partial_s - \partial_y)\bar{\mu}_t + \bar{\alpha}_t\big)dt + \bar{\sigma}_t dW_t + \bar{\delta}_t(\xi)(\mathfrak{p}(dt,d\xi) - \nu(d\xi)dt) \\ \bar{\mu}_0 &= \mu_0. \end{cases}$$

Condition (24) is necessary and sufficient for consistency of the $\mathbb{F}$-forward mortality rates. It resembles the HJM drift condition for bond markets with Musiela parametrization.

5.4 Consistent Musiela type dynamics of the forward mortality improvements

In this section, we derive a unified framework for the $\mathbb{F}$-forward mortality improvements. Stochastic processes will take values in an appropriate function space. The implied $\mathbb{F}$-forward mortality rates will be consistent under suitable conditions.

Let H be a forward mortality space (see Definition 5.7) and $j_0 \in H$ an initial surface of forward mortality improvements. We define the $\mathbb{F}$-forward mortality improvements $\bar{j}$ as the H-valued process

$$\begin{aligned} \bar{j}_t &:= S_t j_0 + \int_0^t S_{t-s}\bar{a}_s ds + \int_0^t S_{t-s}\bar{b}_s dW_s \\ &\quad + \int_0^t \int_E S_{t-s}\bar{c}_s(\xi)(\mathfrak{p}(ds,d\xi) - \nu(d\xi)ds), \quad t \geq 0. \end{aligned} \tag{25}$$

Here, $\bar{a}: \Omega \times \mathbb{R}_+ \to H$, $\bar{b}: \Omega \times \mathbb{R}_+ \to L_2^0(H)$ and $\bar{c}: \Omega \times \mathbb{R}_+ \times E \to H$ are stochastic processes satisfying the technical Assumption C.13 (see below). Assumption C.13 ensures that the stochastic integrals in the variation of constants formula (25) exist.

Letting $\gamma_0: \mathbb{R}_+ \to \mathbb{R}$ be an initial curve of $\mathbb{F}$-spot mortality rates, we extend this curve to an initial surface $\mu_0: \Xi \to \mathbb{R}$ of $\mathbb{F}$-forward mortality rates by setting

$$\mu_0(s,y) := \gamma_0(s+y) - \int_0^s j_0(u, s+y-u)du. \tag{26}$$

Furthermore, we redefine the stochastic processes $\bar{\alpha}: \Omega \times \Xi \to \mathbb{R}$, $\bar{\sigma}: \Omega \times \Xi \to L_2^0(\mathbb{R})$ and $\bar{\delta}: \Omega \times \Xi \times E \to \mathbb{R}$ as

$$\bar{\alpha}_t(s,y) := -\int_0^s \bar{a}_t(u, s+y-u)du, \quad \bar{\sigma}_t(s,y) := -\int_0^s \bar{b}_t(u, s+y-u)du \quad \text{and}$$
$$\bar{\delta}_t(s,y,\xi) := -\int_0^s \bar{c}_t(u, s+y-u, \xi)du. \tag{27}$$

Suppose that $\mu_0 \in H$, and that $\bar{\alpha}$, $\bar{\sigma}$ and $\bar{\delta}$ are H-valued processes such that Assumption C.11 is fulfilled. The H-valued process $\bar{\mu}$ is redefined by the variation of constants formula (23), and for $y \geq 0$ we define the $\mathbb{F}$-spot mortality rates $\bar{\gamma}(y)$ as

$$\bar{\gamma}_t(y) := \bar{\mu}_t(0,y), \quad t \geq 0. \tag{28}$$

Again, we impose an exponential integrability condition on the Lévy measure ν.

Theorem 5.13. *Suppose that Assumption C.13 is satisfied. Then the following statements are true:*

(i) For all $y \in \mathbb{R}_+$ the $\mathbb{F}$-spot mortality rates $\bar{\gamma}(y)$ have the dynamics

$$\begin{aligned} \bar{\gamma}_t(y) &= S_t\mu_0(0,y) + \int_0^t S_{t-s}\bar{\alpha}_s(0,y)ds + \int_0^t S_{t-s}\bar{\sigma}_s(0,y)dW_s \\ &\quad + \int_0^t \int_E S_{t-s}\bar{\delta}_s(0,y,\xi)(\mathfrak{p}(ds,d\xi) - \nu(d\xi)ds), \quad t \geq 0. \end{aligned} \tag{29}$$

(ii) For all $(s,y) \in \Xi$ we have

$$\bar{\mu}_t(s,y) = \bar{\gamma}_t(s+y) - \int_0^s \bar{j}_t(u, s+y-u)du, \quad t \geq 0. \tag{30}$$

(iii) If, in addition, Assumption C.14 is satisfied, and the drift $\bar{a}$ is given by

$$
\begin{aligned}
\bar{a}_t(s,y) = & -\sum_{k\in\mathbb{N}} \left(\int_0^s \bar{b}_t^k(u, s+y-u) du \right) \left(\int_{-y\vee 0}^s \bar{b}_t^k(u,y) du \right) \\
& - \sum_{k\in\mathbb{N}} \bar{b}_t^k(s,y) \int_{-y\vee 0}^s \int_0^u \bar{b}_t^k(v, u+y-v) dv du \\
& - \int_E \left(\int_0^s \bar{c}_t(u, s+y-u, \xi) du \right) \left(\int_{-y\vee 0}^s \bar{c}_t(u,y,\xi) du \right) \\
& \qquad \times \exp\left(\int_{-y\vee 0}^s \int_0^u \bar{c}_t(v, u+y-v, \xi) dv du \right) \nu(d\xi) \\
& - \int_E \bar{c}_t(s,y,\xi) \left[\exp\left(\int_{-y\vee 0}^s \int_0^u \bar{c}_t(v, u+y-v, \xi) dv du \right) - 1 \right] \nu(d\xi),
\end{aligned}
\tag{31}
$$

then the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ are consistent.

Proof. See Appendix C.4. $\qquad\qquad\square$

Remark 5.14. *In the spirit of Da Prato & Zabczyk (1992), the process $\bar{j}$ of $\mathbb{F}$-forward mortality improvements in (25) is a mild solution to the stochastic partial differential equation (SPDE)*

$$
\begin{cases}
d\bar{j}_t = \left((\partial_s - \partial_y)\bar{j}_t + \bar{a}_t \right) dt + \bar{b}_t dW_t + \bar{c}_t(\xi)(\mathfrak{p}(dt, d\xi) - \nu(d\xi) dt) \\
\bar{j}_0 = j_0.
\end{cases}
$$

Similar to Remark 5.6, we can derive the following relations between the initial surfaces and the volatilities of consistent dynamic evolutions of forward mortality rates and improvements:

- *For a given initial surface j_0 of $\mathbb{F}$-forward mortality improvements and an initial curve γ_0 of $\mathbb{F}$-spot mortality rates, we can compute the initial surface μ_0 of $\mathbb{F}$-forward mortality rates by (26).*

- *Conversely, for a given initial surface μ_0 of $\mathbb{F}$-forward mortality rates, we can compute the initial surface j_0 of $\mathbb{F}$-forward mortality improvements as*

$$
j_0(s,y) := -(\partial_s - \partial_y)\mu_0(s,y).
$$

- *For given volatilities $\bar{a}, \bar{b}, \bar{c}$ in (25) we can compute the volatilities $\bar{\alpha}, \bar{\sigma}, \bar{\delta}$ in (23) by (27).*

- *For given volatilities $\bar{\alpha}, \bar{\sigma}, \bar{\delta}$ in (23) with $\bar{\alpha}_t(0,y) = \bar{\sigma}_t(0,y) = \bar{\delta}_t(0,y,\xi) = 0$ we can compute the volatilities $\bar{a}, \bar{b}, \bar{c}$ in (25) by*

$$
\bar{a}_t(s,y) := -(\partial_s - \partial_y)\bar{\alpha}_t(s,y), \quad \bar{b}_t(s,y) := -(\partial_s - \partial_y)\bar{\sigma}_t(s,y) \quad \text{and}
$$
$$
\bar{c}_t(s,y,\xi) := -(\partial_s - \partial_y)\bar{\delta}_t(s,y,\xi).
$$

- *Note that, for consistency, the drift term $\bar{a}$ is given by (31), and the drift term $\bar{\alpha}$ is given by (24).*

6 Example: A Lévy process driven Gompertz-Makeham model

In order to illustrate our previous results, we present a Lévy process driven version of the Gompertz-Makeham model, and compute consistent dynamics of this model. In Section 6.1 we consider the general situation, where the $\mathbb{F}$-forward mortality rates and the $\mathbb{F}$-forward mortality improvements are driven by a Lévy process; in Section 6.2 we focus on the particular case of the Gompertz-Makeham model.

6.1 Lévy process driven mortality models

Let X be a real-valued Lévy process with Gaussian part $C \geq 0$ and Lévy measure ν. We assume that there exist constants $M, \epsilon > 0$ such that

$$
\int_{\{|\xi|>1\}} e^{z\xi} \nu(d\xi) < \infty \quad \text{for all } z \in [-(1+\epsilon)M, (1+\epsilon)M].
\tag{32}
$$

Then the cumulant generating function $\Psi(z) := \ln \mathbb{E}[e^{zX_1}]$ exists on $[-(1+\epsilon)M, (1+\epsilon)M]$ and is of class C^∞ on the interior $(-(1+\epsilon)M, (1+\epsilon)M)$ with representations

$$\Psi(z) = Bz + \frac{C}{2}z^2 + \int_{\mathbb{R}}(e^{z\xi} - 1 - z\xi)\nu(d\xi),$$

$$\Psi'(z) = B + Cz + \int_{\mathbb{R}}\xi(e^{z\xi} - 1)\nu(d\xi),$$

$$\Psi''(z) = C + \int_{\mathbb{R}}\xi^2 e^{z\xi}\nu(d\xi),$$

where $B \in \mathbb{R}$ denotes the drift of X. We shall directly switch to Musiela type dynamics. Let H be a forward mortality space, see Definition 5.7. Suppose that the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ and the $\mathbb{F}$-forward mortality improvements $\bar{j}$ are given by the variation of constants formulas

$$\bar{\mu}_t = S_t \mu_0 + \int_0^t S_{t-s}\hat{\alpha}_s ds + \int_0^t S_{t-s}\hat{\sigma}_s dX_s, \quad t \geq 0, \tag{33}$$

$$\bar{j}_t = S_t j_0 + \int_0^t S_{t-s}\hat{a}_s ds + \int_0^t S_{t-s}\hat{b}_s dX_s, \quad t \geq 0, \tag{34}$$

with initial surfaces $\mu_0, j_0 \in H$ and appropriate H-valued processes $\hat{\alpha}, \hat{\sigma}$ and $\hat{a}, \hat{b}$. Note that these are particular cases of the variation of constant formulas (23), (25): the state space U of the Wiener process W and the mark space E of the Poisson random measure $\mathfrak{p}$ are $U = E = \mathbb{R}$; the mapping $\bar{\sigma}$ in (23) is provided by $\sqrt{C}\hat{\sigma}$, and $\bar{\delta}(\xi)$ in (23) is given by $\xi\hat{\sigma}$; the mapping $\bar{b}$ in (25) by $\sqrt{C}\hat{b}$, and $\bar{c}(\xi)$ in (25) by $\xi\hat{b}$.

In this case, the drift condition (24) in Theorem 5.11 for the $\mathbb{F}$-forward mortality rates ensuring consistency becomes

$$\hat{\alpha}_t(s,y) = -\hat{\sigma}_t(s,y)\Psi'\left(-\int_{-y\vee 0}^s \hat{\sigma}_t(u,y)du\right), \tag{35}$$

and the drift condition (31) in Theorem 5.13 for the $\mathbb{F}$-forward mortality improvements translates to

$$
\begin{aligned}
\hat{a}_t(s,y) = &-\left(\int_0^s \hat{b}_t(u, s+y-u)du\right)\left(\int_{-y\vee 0}^s \hat{b}_t(u,y)du\right) \\
&\cdot \Psi''\left(\int_{-y\vee 0}^s \int_0^u \hat{b}_t(v, u+y-v)dvdu\right) \\
&- \hat{b}_t(s,y)\Psi'\left(\int_{-y\vee 0}^s \int_0^u \hat{b}_t(v, u+y-v)dvdu\right).
\end{aligned} \tag{36}
$$

6.2 The Gompertz-Makeham model

As an illustrating example of a stochastic mortality model, we describe in the current section a Lévy process driven Gompertz-Makeham model. Letting $\theta_1 > 1$ and $\theta_2, \theta_3, \theta_4, \theta_5 > 0$ be real constants, the initial surface $j_0 : \Xi \to \mathbb{R}$ of $\mathbb{F}$-forward mortality improvements, the initial surface $\mu_0 : \Xi \to \mathbb{R}$ of $\mathbb{F}$-forward mortality rates and the initial curve $\gamma_0 : \mathbb{R}_+ \to \mathbb{R}$ of $\mathbb{F}$-spot mortality are provided by

$$j_0(s,y) = \theta_2 e^{-\theta_2 s}(\theta_3 e^{\theta_4(s+y)} + \theta_5),$$
$$\mu_0(s,y) = (\theta_1 + e^{-\theta_2 s})(\theta_3 e^{\theta_4(s+y)} + \theta_5),$$
$$\gamma_0(y) = (\theta_1 + 1)(\theta_3 e^{\theta_4 y} + \theta_5).$$

These initial surfaces satisfy the relationships described in Remark 5.14. For every $z \in \mathbb{R}_+$ we observe that

$$\lim_{s\to\infty} \mu_0(s, z-s) = \theta_1(\theta_3 e^{\theta_4 z} + \theta_5),$$

i.e., if the length s of the prediction time horizon tends to ∞ for predictions about the mortality of individuals at age z at the end of the time horizon, then the initial forward mortality rates are again described by a classical Gompertz-Makeham model.

Remark 6.1. *The initial surfaces j_0 and μ_0 belong to the forward mortality space H defined in Example 5.8, with appropriate choices of weight functions (20).*

Finally, we describe three examples of volatility structures $\hat{a}, \hat{b}$ and $\hat{\alpha}, \hat{\sigma}$, which are computed according to Remark 5.14 and drift conditions (35), (36).

Example 6.2. *If the volatility $\hat{b}$ of the $\mathbb{F}$-forward mortality improvements is constant and equal to 1, then we compute*

$$\hat{a}(s,y) = -s(s + y\mathbb{1}_{\{y<0\}})\Psi''\left(\frac{s^2}{2} - \frac{y^2}{2}\mathbb{1}_{\{y<0\}}\right) - \Psi'\left(\frac{s^2}{2} - \frac{y^2}{2}\mathbb{1}_{\{y<0\}}\right),$$

$$\hat{b}(s,y) = 1,$$

$$\hat{\alpha}(s,y) = s\Psi'\left(\frac{s^2}{2} - \frac{y^2}{2}\mathbb{1}_{\{y<0\}}\right),$$

$$\hat{\sigma}(s,y) = -s.$$

In particular, the volatility of the forward mortality rates is proportional to the length of the prediction time horizon.

Example 6.3. *In this example, the volatility $\hat{b}$ of the $\mathbb{F}$-forward mortality improvements equals the age $s + y$ of individuals at the end of the time period. In this case, we obtain*

$$\hat{a}(s,y) = -s(s+y)\left(\frac{s^2}{2} + sy + \frac{y^2}{2}\mathbb{1}_{\{y<0\}}\right)\Psi''\left(\frac{3s^2y + 2s^3}{6} - \frac{y^3}{6}\mathbb{1}_{\{y<0\}}\right)$$
$$- (s+y)\Psi'\left(\frac{3s^2y + 2s^3}{6} - \frac{y^3}{6}\mathbb{1}_{\{y<0\}}\right),$$

$$\hat{b}(s,y) = s + y,$$

$$\hat{\alpha}(s,y) = s(s+y)\Psi'\left(\frac{3s^2y + 2s^3}{6} - \frac{y^3}{6}\mathbb{1}_{\{y<0\}}\right),$$

$$\hat{\sigma}(s,y) = -s(s+y).$$

In particular, the volatility of the forward mortality rates is proportional to the length of the prediction time horizon times the age of the individuals at the end of the time horizon.

Example 6.4. *Finally, the volatility $\hat{b}$ of the $\mathbb{F}$-forward mortality improvements is assumed to equal the age $s + y$ at the end of the time horizon multiplied by $1 - e^{-s}$, a factor which models increasing uncertainty for longer prediction horizons. Introducing the auxiliary functions $g, h : \Xi \to \mathbb{R}$ as*

$$g(s,y) = \frac{(2se^s + 2)y + s^2e^s + 2s + 2}{2}e^{-s} - (y+1)\mathbb{1}_{\{y<0\}} - \frac{2e^y - y^2}{2}\mathbb{1}_{\{y\geq0\}},$$

$$h(s,y) = \frac{((3s^2 - 6s)e^s - 6)y + (2s^3 - 3s^2)e^s - 6s - 6}{6}e^{-s} + (y+1)\mathbb{1}_{\{y\geq0\}}$$
$$+ \frac{6e^y - y^3 - 3y^2}{6}\mathbb{1}_{\{y<0\}},$$

we obtain the drift and volatility terms

$$\hat{a}(s,y) = -(s+y)(1 + s - e^{-s})g(s,y)\Psi''(h(s,y)) - (s+y)(1 - e^{-s})\Psi'(h(s,y)),$$

$$\hat{b}(s,y) = (s+y)(1 - e^{-s}),$$

$$\hat{\alpha}(s,y) = (s+y)(1 + s - e^{-s})\Psi'(h(s,y)),$$

$$\hat{\sigma}(s,y) = -(s+y)(e^{-s} + s - 1).$$

The $\mathbb{F}$-forward mortality rates μ described in Examples 6.2–6.4 may generally become negative with non zero probability; they should, thus, be interpreted as approximations of real mortality rates, see Remark 5.9.

Remark 6.5. *For the volatilities $\hat{b}$ from Examples 6.2–6.4 the double integrals appearing in (36), which are evaluated by Ψ' and Ψ'', take values in $\mathbb{R}_+$. Therefore, in the context of these examples one should assume – besides condition (32) – that the cumulant generating function Ψ exists even on $\mathbb{R}_+$. This hypothesis is, for example, satisfied if the Lévy process is a jump diffusion $X = W + N$ that can be described as the sum of a Wiener process W and a Poisson process N. In this particular case, the cumulant generating function equals*

$$\Psi(z) = \frac{z^2}{2} + e^z - 1.$$

For appropriate choices of weight functions w_1, w_2 and w_3, the drift and volatility terms in Examples 6.2–6.4 thus belong to the forward mortality space H defined in Example 5.8.

A   Appendix to Section 3: A Conditional Law of Large Numbers

In this appendix, we provide the proof of Theorem 3.4.

Proof of Theorem 3.4. The random variables $N^n(x)_t$, $n \in \mathbb{N}$, are conditionally independent given $\mathcal{F}_\infty$ with identical $\mathcal{F}_\infty$-conditional Bernoulli-distributions by Assumption 3.1 and Definition 3.3. By a conditional strong law of large numbers, see e.g. Theorem 7 in Prakasa Rao (2009), we have

$$\frac{1}{N} \sum_{n=1}^{N} N^n(x)_t \longrightarrow \mathbb{E}[N^1(x)_t \,|\, \mathcal{F}_\infty] = \mathbb{P}(\tau^{x,1} > t \,|\, \mathcal{F}_t) = G_t(t,x) \qquad \mathbb{P}\text{--almost surely.}$$

By a conditional version of Lebesgue's dominated convergence theorem, see e.g. Theorem 23.8 in Jacod & Protter (2004), we obtain

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[N^n(x)_T \,|\, \mathcal{G}_t] = \mathbb{E}\left[\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} N^n(x)_T \,\bigg|\, \mathcal{G}_t \right] = \mathbb{E}[G_T(T,x) \,|\, \mathcal{G}_t]$$

$$\overset{(*)}{=} \mathbb{E}[G_T(T,x) \,|\, \mathcal{F}_t] = G_t(T,x).$$

Equality $(*)$ follows by Proposition 6.6 in Kallenberg (2002), if $\mathcal{F}_T \perp\!\!\!\perp_{\mathcal{F}_t} \sigma\{N^n(x)_s : s \leq t, n \in \mathbb{N}\}$. Since $\mathcal{F}_T \subset \mathcal{F}_\infty$, this follows from Lemma A.1. $\qquad \square$

Lemma A.1. *Consider the setting of Theorem 3.4. Then for $t \in \mathbb{R}_+$: $\mathcal{F}_\infty \perp\!\!\!\perp_{\mathcal{F}_t} \sigma\{N^n(x)_s : s \leq t, n \in \mathbb{N}\}$.*

Proof. If $B \in \sigma(N^n(x)_s)$ for some $s \leq t$, then $\mathbb{1}_B = f(N^n(x)_s)$ with $f : \{0,1\} \to \{0,1\}$. Letting $A \in \mathcal{F}_\infty$, $K \in \mathbb{N}$, $s_k \leq t$, $f_k : \{0,1\} \to \{0,1\}$, $k = 1, \ldots, K$, we have

$$\mathbb{P}\left(A \cap \bigcap_{k=1}^{K} \{f_k(N^k(x)_{s_k}) = 1\} \,\bigg|\, \mathcal{F}_t \right) = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_A \cdot \prod_{k=1}^{K} f_k(N^k(x)_{s_k}) \,\bigg|\, \mathcal{F}_\infty \right] \,\bigg|\, \mathcal{F}_t \right]$$

$$= \mathbb{E}\left[\mathbb{1}_A \cdot \mathbb{E}\left[\prod_{k=1}^{K} f_k(N^k(x)_{s_k}) \,\bigg|\, \mathcal{F}_\infty \right] \,\bigg|\, \mathcal{F}_t \right] = \mathbb{E}\left[\mathbb{1}_A \cdot \prod_{k=1}^{K} \mathbb{E}\left[f_k(N^k(x)_{s_k}) \,|\, \mathcal{F}_t \right] \,\bigg|\, \mathcal{F}_t \right]$$

$$= \mathbb{E}\left[\mathbb{1}_A \,|\, \mathcal{F}_t \right] \cdot \prod_{k=1}^{K} \mathbb{E}\left[f_k(N^k(x)_{s_k}) \,|\, \mathcal{F}_t \right],$$

and hence

$$\mathbb{P}\left(A \cap \bigcap_{k=1}^{K} \{f_k(N^k(x)_{s_k}) = 1\} \,\bigg|\, \mathcal{F}_t \right) = \mathbb{P}\left(A \,|\, \mathcal{F}_t \right) \cdot \mathbb{E}\left[\prod_{k=1}^{K} \mathbb{E}\left[f_k(N^k(x)_{s_k}) \,|\, \mathcal{F}_t \right] \,\bigg|\, \mathcal{F}_t \right]$$

$$= \mathbb{P}\left(A \,|\, \mathcal{F}_t \right) \cdot \mathbb{E}\left[\prod_{k=1}^{K} \mathbb{E}\left[f_k(N^k(x)_{s_k}) \,|\, \mathcal{F}_\infty \right] \,\bigg|\, \mathcal{F}_t \right]$$

$$= \mathbb{P}\left(A \,|\, \mathcal{F}_t \right) \cdot \mathbb{E}\left[\mathbb{P}\left(\bigcap_{k=1}^{K} \{f_k(N^k(x)_{s_k}) = 1\} \,\bigg|\, \mathcal{F}_\infty \right) \,\bigg|\, \mathcal{F}_t \right]$$

$$= \mathbb{P}\left(A \,|\, \mathcal{F}_t \right) \cdot \mathbb{P}\left(\bigcap_{k=1}^{K} \{f_k(N^k(x)_{s_k}) = 1\} \,\bigg|\, \mathcal{F}_t \right).$$

The class of subsets

$$\mathcal{C} := \left\{ \bigcap_{k=1}^{K} \{f_k(N^k(x)_{s_k}) = 1\} : s_k \leq t, \ f_k : \{0,1\} \to \{0,1\}, \ k \in \{1, \ldots, K\}, \ K \in \mathbb{N} \right\}$$

is a π-system (i.e. closed under finite intersections) that generates the σ-algebra $\sigma\{N^n(x)_s : s \leq t, n \in \mathbb{N}\}$. The lemma now follows from a conditional analogue of Lemma 3.6 in Kallenberg (2002). $\square$

B Appendix to Section 4: Compensators

In this appendix, we provide the proofs of Propositions 4.2 and 4.3.

Proof of Proposition 4.2. Define the filtration $\mathbb{G}^n = (\mathcal{G}_t^n)_{t \in \mathbb{R}_+}$ by setting $\mathcal{G}_t^n = \mathcal{F}_t \vee \sigma(N^n(x)_s : s \leq t)$, $t \in \mathbb{R}_+$. The $\mathbb{G}^n$-compensator $B^n(x)$ of $\tau^{x,n}$ coincides with the $\mathbb{G}$-compensator $A^n(x)$. This can be seen as follows: the $\mathbb{G}^n$-compensator $B^n(x)$ of $\tau^{x,n}$ is a $\mathbb{G}$-predictable, right-continuous, increasing process. If $1 - N^n(x) - B^n(x)$ was a $\mathbb{G}$-martingale, then $B^n(x)$ would be equal to the $\mathbb{G}$-compensator $A^n(x)$. The martingale property follows from Lemma B.1.

Formula (5) thus defines a $(\mathbb{F}, \mathbb{G})$-martingale hazard process of $\tau^{x,n}$ by Proposition 6.1.2 in Bielecki & Rutkowski (2002). Since formula (5) does not depend on n, the existence of a process $\Lambda(x)$ with the desired properties is proven. It remains to show uniqueness. Since $\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} N^n(x)_t = G_t(t, x) > 0$ $\mathbb{P}$–almost surely by Theorem 3.4 and Assumption 2.2, we know that the increasing sequence of $\mathbb{G}$-stopping times $\tilde{\tau}^{x,n} := \max\{\tau^{x,i} : i = 1, \ldots, n\}$, $n \in \mathbb{N}$, diverges $\mathbb{P}$–almost surely to ∞. Observe that $\Lambda(x)^{\tilde{\tau}^{x,n}} = A^i(x)$ on $\{\tilde{\tau}^{x,n} = \tau^{x,i}\}$ and $\Omega = \bigcup_{i=1}^{n} \{\tilde{\tau}^{x,n} = \tau^{x,i}\}$. Since the $\mathbb{G}$-compensators $A^i(x)_{i \in \mathbb{N}}$ are unique, it follows that the stopped process $\Lambda(x)^{\tilde{\tau}^{x,n}}$ is uniquely specified. Since $\tilde{\tau}^{x,n} \to \infty$ $\mathbb{P}$–almost surely for $n \to \infty$, this implies the uniqueness of $\Lambda(x)$.

If $t \mapsto G_t(t, x)$ is almost surely continuous, then it is also predictable, whence the last statement follows. $\square$

For the proof of Proposition 4.3 we prepare an auxiliary result:

Lemma B.1. *Assume that $(\tau^{x,j})_{j \in \mathbb{N}}$ is a $\mathbb{F}$-DSCI family of death times of individuals born at date $-x \in \mathbb{R}_+$. Define $\mathcal{G}_t := \sigma\{N^j(x)_s : s \leq t, j \in \mathbb{N}\} \vee \mathcal{F}_t$, $\mathcal{G}_t^n = \sigma(N^n(x)_s : s \leq t) \vee \mathcal{F}_t$, $t \in \mathbb{R}_+$, $\mathcal{G}_\infty^n = \bigvee_{t \in \mathbb{R}_+} \mathcal{G}_t^n$ for some $n \in \mathbb{N}$.*

Let Y be a $\mathcal{G}_\infty^n$-measurable, integrable random variable. Then for $t \in \mathbb{R}_+$:

$$\mathbb{E}[Y \mid \mathcal{G}_t] = \mathbb{E}[Y \mid \mathcal{G}_t^n] \qquad \mathbb{P}\text{–almost surely.}$$

Proof. Let $\mathcal{H}_t^n = \sigma(N^n(x)_s : s \leq t)$, $\mathcal{H}_t^{\neq n} = \sigma(N^j(x)_s : s \leq t, j \in \mathbb{N}, j \neq n)$, $\mathcal{H}_t = \sigma(N^j(x)_s : s \leq t, j \in \mathbb{N})$ and $\mathcal{H}_\infty^n = \bigvee_{t \in \mathbb{R}} \mathcal{H}_t^n$. By Lemma A.1 we have $\mathcal{F}_\infty \underset{\mathcal{F}_t}{\perp\!\!\!\perp} \mathcal{H}_t = \mathcal{H}_t^n \vee \mathcal{H}_t^{\neq n}$. This implies by Proposition 6.8 in Kallenberg (2002):

$$\mathcal{F}_\infty \underset{\mathcal{F}_t \vee \mathcal{H}_t^n}{\perp\!\!\!\perp} \mathcal{H}_t^{\neq n}. \tag{37}$$

It follows from Definition 3.3(ii) that $\mathcal{H}_t^n \vee \mathcal{H}_\infty^n = \mathcal{H}_\infty^n \underset{\mathcal{F}_\infty}{\perp\!\!\!\perp} \mathcal{H}_t^{\neq n}$, thus by Proposition 6.8 in Kallenberg (2002):

$$\mathcal{H}_\infty^n \underset{\mathcal{F}_\infty \vee \mathcal{H}_t^n}{\perp\!\!\!\perp} \mathcal{H}_t^{\neq n}. \tag{38}$$

Equations (37) and (38) imply by Proposition 6.8 in Kallenberg (2002) that

$$\mathcal{G}_\infty^n = \mathcal{F}_\infty \vee \mathcal{H}_\infty^n \underset{\mathcal{F}_t \vee \mathcal{H}_t^n}{\perp\!\!\!\perp} \mathcal{H}_t^{\neq n}.$$

Since $\mathcal{F}_t \vee \mathcal{H}_t^n = \mathcal{G}_t^n$ and $\mathcal{H}_t^{\neq n} \vee \mathcal{G}_t^n = \mathcal{G}_t$, Corollary 6.7(i) in Kallenberg (2002) shows that $\mathcal{G}_\infty^n \underset{\mathcal{G}_t^n}{\perp\!\!\!\perp} \mathcal{G}_t$. We obtain $\mathbb{E}[Y \mid \mathcal{G}_t] = \mathbb{E}[Y \mid \mathcal{G}_t \vee \mathcal{G}_t^n] = \mathbb{E}[Y \mid \mathcal{G}_t^n]$, where the last equality follows from Proposition 6.6 in Kallenberg (2002). $\square$

Proof of Proposition 4.3. Lemma B.1 shows that $\Lambda(x)$ is a $(\mathbb{F}, \mathbb{G}^n)$-martingale hazard process of $\tau^{x,n}$ for any $n \in \mathbb{N}$, where $\mathbb{G}^n = (\mathcal{G}_t^n)_{t \in \mathbb{R}_+}$ is the filtration introduced in the proof of Proposition 4.2. Hence, by Proposition 6.2.1(ii) in Bielecki & Rutkowski (2002) we obtain

$$\Lambda(x)_t = -\ln(1 - \tilde{F}_t(x)) = -\ln G_t(t, x) = \Gamma_t(t, x), \quad t \geq -x,$$

proving the claim. $\square$

C Appendix to Section 5: Infinite-Dimensional Formulation

In this appendix, we provide the proofs of Section 5 as well as technical assumptions.

C.1 Proof of Theorem 5.3

In this appendix, we provide the proof of Theorem 5.3. Motivated by Definitions (7) and (8) of Θ and Ξ, for each $x \in \mathbb{R}$ we define the sets $\Xi_x \subset \mathbb{R}_+$ and $\Theta_x \subset \mathbb{R}_+ \times \mathbb{R}$ as

$$\Xi_x := \{T \in \mathbb{R}_+ : (T, x) \in \Xi\} \quad \text{and} \quad \Theta_x := \{(t, T) \in \mathbb{R}_+^2 : (t, T, x) \in \Theta\}.$$

Assumption C.1. *We suppose that the following conditions are satisfied:*

(i) *μ_0 is $\mathcal{B}(\Xi)$-measurable, α and σ are $\mathcal{F}_\infty \otimes \mathcal{B}(\Theta)$-measurable, and δ is $\mathcal{F}_\infty \otimes \mathcal{B}(\Theta) \otimes \mathcal{E}$-measurable.*

(ii) *For all $(T, x) \in \Xi$ the processes $\alpha(T, x)$ and $\sigma(T, x)$ are optional, and $\delta(T, x)$ is predictable.*

(iii) *For each $x \in \mathbb{R}$ and every bounded Borel set $B \subset \Xi_x$ we have $\int_B |\mu_0(T, x)| dT < \infty$.*

(iv) *For each $x \in \mathbb{R}$ and every bounded Borel set $B \subset \Theta_x$ there are random variables $X^\alpha : \Omega \to \mathbb{R}$ and $X^\sigma, X^\delta \in \mathcal{L}^2(\Omega)$ such that*

$$
\begin{aligned}
|\alpha_t(T, x)| \leq X^\alpha, \quad \|\sigma_t(T, x)\|_{L_2^0(\mathbb{R})} &\leq X^\sigma \quad and \\
\|\delta_t(T, x)\|_{\mathcal{L}_\nu^2(\mathbb{R})} \leq X^\delta \quad &for \ all \ (t, T) \in B.
\end{aligned}
\tag{39}
$$

(v) *For all $t \in \mathbb{R}_+$ and $\xi \in E$ the mappings $(T, x) \mapsto \alpha_t(T, x)$, $(T, x) \mapsto \sigma_t(T, x)$ and $(T, x) \mapsto \delta_t(T, x, \xi)$ are continuous on their domains.*

Condition (39) ensures that all subsequent stochastic integrals regarding α, σ and δ exist. In particular, for each $x \in \mathbb{R}$ and every bounded Borel set $B \subset \Theta_x$ we have

$$
\begin{aligned}
\iint_B |\alpha_t(T, x)| dt dT < \infty, \quad \mathbb{E}\left[\iint_B \|\sigma_t(T, x)\|_{L_2^0(\mathbb{R})}^2 dt dT\right] &< \infty \quad \text{and} \\
\mathbb{E}\left[\iint_B \|\delta_t(T, x)\|_{\mathcal{L}_\nu^2(\mathbb{R})}^2 dt dT\right] &< \infty.
\end{aligned}
\tag{40}
$$

This ensures that we may apply the classical Fubini theorem and the stochastic Fubini theorems (Theorem 2.8 in Gawarecki & Mandrekar (2011) and Theorem A.2 in Filipović, Tappe & Teichmann (2010*a*)) later on.

Remark C.2. *Definition (1) shows that for all $(T, x) \in \Xi$ we have $|G(T, x)| \leq 1$. Therefore, the $\mathbb{F}$-forward mortality rates μ are consistent if and only if for all $(T, x) \in \Xi$ the $\mathbb{F}$-survival process $G(T, x)$ is a local $\mathbb{F}$-martingale, see e.g. Jacod & Shiryaev (2003, Prop. I.1.47).*

For the proof of Theorem 5.3, it will be useful to extend the $\mathbb{F}$-forward mortality rates μ and $\mathbb{F}$-spot mortality rates γ as follows. We define the process $\tilde{\mu} : \Omega \times \mathbb{R}_+^2 \times \mathbb{R} \to \mathbb{R}$ as

$$
\tilde{\mu}_t(T, x) := \begin{cases} \mu_t(T, x), & (t, T, x) \in \Theta, \\ 0, & -x > T, \\ \mu_T(T, x), & -x \leq T \text{ and } t > T, \end{cases}
$$

and the process $\tilde{\gamma} : \Omega \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ as

$$\tilde{\gamma}_t(x) := \tilde{\mu}_t(t, x).$$

A straightforward calculation shows that for all $(T, x) \in \Xi$ we have

$$G_t(T, x) = \exp(-\Gamma_0(0, x)) \exp\left(-\int_0^t \tilde{\gamma}_s(x) ds\right) \exp\left(-\int_t^T \tilde{\mu}_t(s, x) ds\right), \quad t \in [0, T]. \tag{41}$$

We shall now determine the dynamics of the processes $\tilde{\mu}(T, x)$ and $\tilde{\gamma}(x)$ in (41). For this purpose, we extend the initial surface μ_0 and the processes α, σ and δ in (9) as follows. We define the mapping

$$\tilde{\mu}_0 : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}, \quad \tilde{\mu}_0(T, x) := \mu_0(T, x) \mathbb{1}_\Xi(T, x),$$

and the processes $\tilde{\alpha} : \Omega \times \mathbb{R}_+^2 \times \mathbb{R} \to \mathbb{R}$, $\tilde{\sigma} : \Omega \times \mathbb{R}_+^2 \times \mathbb{R} \to L_2^0(\mathbb{R})$ and $\tilde{\delta} : \Omega \times \mathbb{R}_+^2 \times \mathbb{R} \times E \to \mathbb{R}$ by

$$\tilde{\alpha}_t(T,x) := \alpha_t(T,x)\mathbb{1}_\Theta(t,T,x), \quad \tilde{\sigma}_t(T,x) := \sigma_t(T,x)\mathbb{1}_\Theta(t,T,x) \quad \text{and}$$
$$\tilde{\delta}_t(T,x,\xi) := \delta_t(T,x,\xi)\mathbb{1}_\Theta(t,T,x).$$

Then for each $(T,x) \in \Xi$ we have

$$\tilde{\mu}_t(T,x) = \tilde{\mu}_0(T,x) + \int_0^t \tilde{\alpha}_s(T,x)ds + \int_0^t \tilde{\sigma}_s(T,x)dW_s$$
$$+ \int_0^t \int_E \tilde{\delta}_s(T,x,\xi)(\mathfrak{p}(ds,d\xi) - \nu(d\xi)ds), \quad t \in [0,T], \tag{42}$$

and for each $x \in \mathbb{R}$ we have

$$\tilde{\gamma}_t(x) = \tilde{\mu}_0(t,x) + \int_0^t \tilde{\alpha}_s(t,x)ds + \int_0^t \tilde{\sigma}_s(t,x)dW_s$$
$$+ \int_0^t \int_E \tilde{\delta}_s(t,x,\xi)(\mathfrak{p}(ds,d\xi) - \nu(d\xi)ds), \quad t \geq 0. \tag{43}$$

By virtue of (39), we may define the processes $A : \Omega \times \Theta \to \mathbb{R}$, $\Sigma : \Omega \times \Theta \to L_2^0(\mathbb{R})$ and $\Delta : \Omega \times \Theta \times E \to \mathbb{R}$ as

$$A_t(T,x) := -\int_t^T \tilde{\alpha}_t(u,x)du = -\int_{-x\vee t}^T \alpha_t(u,x)du, \tag{44}$$

$$\Sigma_t(T,x) := -\int_t^T \tilde{\sigma}_t(u,x)du = -\int_{-x\vee t}^T \sigma_t(u,x)du, \tag{45}$$

$$\Delta_t(T,x,\xi) := -\int_t^T \tilde{\delta}_t(u,x,\xi)du = -\int_{-x\vee t}^T \delta_t(u,x,\xi)du. \tag{46}$$

The integrals in (45) are Bochner integrals over the state space $L_2^0(\mathbb{R})$. Using the notation (6), for each $k \in \mathbb{N}$ we have

$$\Sigma_t^k(T,x) = -\int_t^T \tilde{\sigma}_t^k(u,x)du = -\int_{-x\vee t}^T \sigma_t^k(u,x)du.$$

Remark C.3. *By Assumption C.1, the mappings $(T,x) \mapsto A_t(T,x)$, $(T,x) \mapsto \Sigma_t(T,x)$ and $(T,x) \mapsto \Delta_t(T,x,\xi)$ are continuous, and for all $t \in \mathbb{R}_+$ and $x \in \mathbb{R}$ the mappings $T \mapsto A_t(T,x)$, $T \mapsto \Sigma_t(T,x)$ and $T \mapsto \Delta_t(T,x,\xi)$ are continuously differentiable on $[-x \vee t, \infty)$.*

Proposition C.4. *For all $(T,x) \in \Xi$ the $\mathbb{F}$-survival process $G(T,x)$ is an Itô process with dynamics*

$$G_t(T,x) = G_0(T,x) + \int_0^t G_s(T,x)\Big(A_s(T,x) + \frac{1}{2}\|\Sigma_s(T,x)\|_{L_2^0(\mathbb{R})}^2\Big)ds$$
$$+ \int_0^t G_s(T,x)\Sigma_s(T,x)dW_s$$
$$+ \int_0^t \int_E G_{s-}(T,x)\Delta_s(T,x,\xi)(\mathfrak{p}(ds,d\xi) - \nu(d\xi)ds) \tag{47}$$
$$+ \int_0^t \int_E G_{s-}(T,x)\big(e^{\Delta_s(T,x,\xi)} - 1 - \Delta_s(T,x,\xi)\big)\mathfrak{p}(ds,d\xi), \quad t \in [0,T].$$

Proof. By virtue of equation (41) and the dynamics (42), (43), we may argue as in the proof of Björk et al. (1997, Prop. 5.2). For the calculations, we may use the classical Fubini theorem and the stochastic Fubini theorems (see, e.g., Theorem 2.8 in Gawarecki & Mandrekar (2011) and Theorem A.2 in Filipović et al. (2010a)) by virtue of (40). $\qquad\square$

In addition, we require the following assumption. For $n \in \mathbb{N}$ we denote by $\Theta_n \subset \Theta$ the compact set $\Theta_n := \{\theta \in \Theta : \|\theta\| \leq n\}$.

Assumption C.5. *We suppose that for each $n \in \mathbb{N}$ there exist a measurable mapping $\rho_n : E \to \mathbb{R}_+$ and a constant $\epsilon_n > 0$ such that*

$$\int_{\{\rho_n \leq 1\}} \rho(\xi)^2 \nu(d\xi) + \int_{\{\rho_n > 1\}} e^{(1+\epsilon_n)\rho_n(\xi)} \nu(d\xi) < \infty, \tag{48}$$

$$|\delta_t(T, x, \xi)| \leq \rho_n(\xi) \quad \textit{for all } (t, T, x) \in \Theta_n \textit{ and } \xi \in E. \tag{49}$$

Lemma C.6. *For each $n \in \mathbb{N}$ there exists a measurable mapping $\pi_n : E \to \mathbb{R}_+$ such that*

$$\int_E \pi_n(\xi) \nu(d\xi) < \infty, \tag{50}$$

$$\left| e^{\Delta_t(T,x,\xi)} - 1 - \Delta_t(T, x, \xi) \right| \leq \pi_n(\xi) \quad \textit{for all } (t, T, x) \in \Theta_n \textit{ and } \xi \in E, \tag{51}$$

$$\left| \delta_t(T, x, \xi) \left(e^{\Delta_t(T,x,\xi)} - 1 \right) \right| \leq \pi_n(\xi) \quad \textit{for all } (t, T, x) \in \Theta_n \textit{ and } \xi \in E. \tag{52}$$

Proof. By Assumption C.5 and Definition (46) of Δ, for each $n \in \mathbb{N}$ there exist a measurable mapping $\rho_n : E \to \mathbb{R}_+$ and a constant $\epsilon_n > 0$ such that conditions (48), (49) are fulfilled and we have

$$|\Delta_t(T, x, \xi)| \leq \rho_n(\xi) \quad \text{for all } (t, T, x) \in \Theta_n \text{ and } \xi \in E. \tag{53}$$

Let $n \in \mathbb{N}$ be arbitrary. We define the measurable mapping

$$\pi_n : E \to \mathbb{R}_+, \quad \pi_n(\xi) := \exp(1)\rho_n(\xi)^2 \mathbb{1}_{\{\rho_n \leq 1\}} + \frac{2}{\epsilon_n^2} e^{(1+\epsilon_n)\rho_n(\xi)} \mathbb{1}_{\{\rho_n > 1\}}.$$

Then the integrability condition (50) is satisfied due to (48). Note that for each $m \in \mathbb{N}_0$ we have the estimate

$$\left| e^x - \sum_{k=0}^{m-1} \frac{x^k}{k!} \right| \leq \sum_{k=m}^{\infty} \frac{|x|^k}{k!} = |x|^m \sum_{k=0}^{\infty} \frac{|x|^k}{(k+m)!} \leq |x|^m e^{|x|}, \quad x \in \mathbb{R}.$$

Moreover, for all $\epsilon > 0$ and $x > 1$ we have

$$x^2 e^x = \frac{2}{\epsilon^2} \frac{(\epsilon x)^2}{2} e^x \leq \frac{2}{\epsilon^2} e^{\epsilon x} e^x = \frac{2}{\epsilon^2} e^{(1+\epsilon)x}.$$

Consequently, by (49), (53) we deduce (51), (52). □

Proposition C.7. *For all $(T, x) \in \Xi$ the $\mathbb{F}$-survival process $G(T, x)$ is an Itô process with dynamics*

$$\begin{aligned}
G_t(T, x) = G_0(T, x) \\
+ \int_0^t G_s(T, x) \bigg(A_s(T, x) + \frac{1}{2} \|\Sigma_s(T, x)\|_{L_2^0(\mathbb{R})}^2 \\
+ \int_E \left(e^{\Delta_s(T,x,\xi)} - 1 - \Delta_s(T, x, \xi) \right) \nu(d\xi) \bigg) ds \\
+ \int_0^t G_s(T, x) \Sigma_s(T, x) dW_s \\
+ \int_0^t \int_E G_{s-}(T, x) \left(e^{\Delta_s(T,x,\xi)} - 1 \right) (\mathfrak{p}(ds, d\xi) - \nu(d\xi) ds), \quad t \in [0, T].
\end{aligned} \tag{54}$$

Proof. Let $(T, x) \in \Xi$ be arbitrary. Then there exists $n \in \mathbb{N}$ such that $(t, T, x) \in \Theta_n$ for all $t \in [0, T]$. Since $|G(T, x)| \leq 1$, by estimate (51) from Lemma C.6 we obtain

$$\int_0^t \int_E \left| G_s(T, x) \left(e^{\Delta_s(T,x,\xi)} - 1 - \Delta_s(T, x, \xi) \right) \right| \nu(d\xi) ds \leq \int_E \pi_n(\xi) \nu(d\xi), \quad t \in [0, T]$$

showing that this process belongs to $\mathcal{A}_{\text{loc}}^+$, see Jacod & Shiryaev (2003, Sec. I.3). Now, applying Jacod & Shiryaev (2003, Prop. II.1.28), by dynamics (47) we obtain (54). □

Now, we are ready to provide the proof of Theorem 5.3:

Proof of Theorem 5.3. By Assumption 3.1, Remark C.2 and Proposition C.7, the forward mortality rates (9) are consistent if and only if for each $(T, x) \in \Xi$ we have

$$A_t(T, x) + \frac{1}{2}\|\Sigma_t(T, x)\|^2_{L^0_2(\mathbb{R})} + \int_E \left(e^{\Delta_t(T, x, \xi)} - 1 - \Delta_t(T, x, \xi)\right)\nu(d\xi) = 0 \tag{55}$$

$d\mathbb{P} \otimes dt$–almost surely on $\Omega \times [0, T]$, for all $(T, x) \in \Xi$.

By Remark C.3, Lemma C.6 and Lebesgue's dominated convergence theorem, the left-hand side of (55) is continuous in (T, x). Thus, condition (55) is equivalent to

$$A_t(T, x) + \frac{1}{2}\|\Sigma_t(T, x)\|^2_{L^0_2(\mathbb{R})} + \int_E \left(e^{\Delta_t(T, x, \xi)} - 1 - \Delta_t(T, x, \xi)\right)\nu(d\xi) = 0 \tag{56}$$

for all $(T, x) \in \Xi$ with $T \geq t$, $d\mathbb{P} \otimes dt$–almost surely on $\Omega \times \mathbb{R}_+$.

By Remark C.3, Lemma C.6 and Lebesgue's dominated convergence theorem, the left-hand side of (56) is even continuously differentiable in $T \geq -x \vee t$. Since the left-hand side of (56) evaluated at $T = -x \vee t$ vanishes, condition (56) is satisfied if and only if

$$\alpha_t(T, x) = -\langle \sigma_t(T, x), \Sigma_t(T, x)\rangle_{L^0_2(\mathbb{R})} - \int_E \delta_t(T, x, \xi)\left(e^{\Delta_t(T, x, \xi)} - 1\right)\nu(d\xi) \tag{57}$$

for all $(T, x) \in \Xi$ with $T \geq t$, $d\mathbb{P} \otimes dt$–almost surely on $\Omega \times \mathbb{R}_+$,

and this is equivalent to (11). $\qquad\qquad\square$

C.2   Proof of Theorem 5.5

In this appendix, we provide the proof of Theorem 5.5. We impose the following assumption:

Assumption C.8. *We suppose that the following conditions are satisfied:*

  (i) j_0 *is* $\mathcal{B}(\Xi)$*-measurable,* a *and* b *are* $\mathcal{F}_\infty \otimes \mathcal{B}(\Theta)$*-measurable, and* c *is* $\mathcal{F}_\infty \otimes \mathcal{B}(\Theta) \otimes \mathcal{E}$*-measurable.*

  (ii) For all $(T, x) \in \Xi$ *the processes* $a(T, x)$ *and* $b(T, x)$ *are optional, and* $c(T, x)$ *is predictable.*

  (iii) For each $x \in \mathbb{R}$ *and every bounded Borel set* $B \subset \Xi_x$ *we have* $\int_B |j_0(T, x)|dT < \infty$.

  (iv) For each $x \in \mathbb{R}$ *and every bounded Borel set* $B \subset \Theta_x$ *there are random variables* $X^a : \Omega \to \mathbb{R}$ *and* $X^b, X^c \in \mathcal{L}^2(\Omega)$ *such that*

$$|a_t(T, x)| \leq X^a, \quad \|b_t(T, x)\|_{L^0_2(\mathbb{R})} \leq X^b \quad and$$
$$\|c_t(T, x)\|_{\mathcal{L}^2_\nu(\mathbb{R})} \leq X^c \quad for\ all\ (t, T) \in B. \tag{58}$$

  (v) For all $t \in \mathbb{R}_+$ *and* $\xi \in E$ *the mappings* $(T, x) \mapsto a_t(T, x)$, $(T, x) \mapsto b_t(T, x)$, $(T, x) \mapsto c_t(T, x, \xi)$ *are continuous.*

Condition (58) ensures that all subsequent stochastic integrals regarding a, b and c exist. In particular, for each $x \in \mathbb{R}$ and every bounded Borel set $B \subset \Theta_x$ we have

$$\iint_B |a_t(T, x)|dtdT < \infty, \quad \mathbb{E}\left[\iint_B \|b_t(T, x)\|^2_{L^0_2(\mathbb{R})}dtdT\right] < \infty \quad and$$
$$\mathbb{E}\left[\iint_B \|c_t(T, x)\|^2_{\mathcal{L}^2_\nu(\mathbb{R})}dtdT\right] < \infty. \tag{59}$$

This ensures that we may apply the classical Fubini theorem and the stochastic Fubini theorems (Theorem 2.8 in Gawarecki & Mandrekar (2011) and Theorem A.2 in Filipović et al. (2010a)) later on. Furthermore, Assumption C.8 guarantees that the processes α, σ and δ defined in (14) fulfill Assumption C.1. In addition, we require the following assumption:

Assumption C.9. *We suppose that for each* $n \in \mathbb{N}$ *there exist a measurable mapping* $\rho_n : E \to \mathbb{R}_+$ *and a constant* $\epsilon_n > 0$ *such that (48) is satisfied and we have*

$$|c_t(T, x, \xi)| \leq \rho_n(\xi) \quad for\ all\ (t, T, x) \in \Theta_n\ and\ \xi \in E. \tag{60}$$

Lemma C.10. *For each $n \in \mathbb{N}$ there exists a measurable mapping $\pi_n : E \to \mathbb{R}_+$ such that conditions (50)–(52) are fulfilled and we have*

$$\left| c_t(T, x, \xi)\left(e^{\Delta_t(T, x, \xi)} - 1\right)\right| \leq \pi_n(\xi) \quad \text{for all } (t, T, x) \in \Theta_n \text{ and } \xi \in E, \tag{61}$$

$$\left| \left(\int_t^T c_t(u, T + x - u, \xi)du\right)\left(\int_{-x \vee t}^T c_t(u, x, \xi)du\right)e^{\Delta_t(T, x, \xi)}\right| \leq \pi_n(\xi)$$

$$\text{for all } (t, T, x) \in \Theta_n \text{ and } \xi \in E. \tag{62}$$

Proof. By Assumption C.9 and Definition (14) of δ, for each $n \in \mathbb{N}$ there exist a measurable mapping $\rho_n : \Omega \times E \to \mathbb{R}_+$ and a constant $\epsilon_n > 0$ such that conditions (48), (60) are fulfilled and we have (49). Proceeding as in the proof of Lemma C.6 yields the desired estimates (61) and (62). $\qquad \square$

Now, we are ready to provide the proof of Theorem 5.5:

Proof of Theorem 5.5. Let $x \in \mathbb{R}$ be arbitrary. By Definition (10) of $\gamma(x)$ and the dynamics (9) of $\mu(T, x)$ we have (15), proving the first statement. In particular, for all $(T, x) \in \Xi$ we have

$$\gamma_t(T + x - t) = \mu_0(t, T + x - t) + \int_0^t \alpha_s(t, T + x - t)ds + \int_0^t \sigma_s(t, T + x - t)dW_s$$
$$+ \int_0^t \int_E \delta_s(t, T + x - t, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds), \quad t \geq 0. \tag{63}$$

Let $(T, x) \in \Xi$ be arbitrary. We also fix an arbitrary $t \in [0, T]$. By the dynamics (12) of $j(T, x)$ we have

$$-\int_t^T j_t(u, T + x - u)du = -\int_t^T j_0(u, T + x - u)du - \int_t^T \int_0^t a_s(u, T + x - u)dsdu$$
$$- \int_t^T \int_0^t b_s(u, T + x - u)dW_sdu \tag{64}$$
$$- \int_t^T \int_0^t \int_E c_s(u, T + x - u, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds)du.$$

We shall now consider the terms in (63) and (64) separately. By Definition (13) we have

$$\mu_0(t, T + x - t) - \int_t^T j_0(u, T + x - u)du$$

$$= \gamma_0(T + x) - \int_0^t j_0(u, T + x - u)du - \int_t^T j_0(u, T + x - u)du$$

$$= \gamma_0(T + x) - \int_0^T j_0(u, T + x - u)du = \mu_0(T, x).$$

By (59) we may apply the classical Fubini theorem for the following calculation. Incorporating Definition (14) we obtain

$$\int_0^t \alpha_s(t, T + x - t)ds - \int_t^T \int_0^t a_s(u, T + x - u)dsdu$$

$$= \int_0^t \left(\alpha_s(t, T + x - t) - \int_t^T a_s(u, T + x - u)du\right)ds$$

$$= \int_0^t \left(-\int_s^t a_s(u, T + x - u)du - \int_t^T a_s(u, T + x - u)du\right)ds$$

$$= -\int_0^t \int_s^T a_s(u, T + x - u)duds = \int_0^t \alpha_s(T, x)ds.$$

Analogous calculations yield

$$\int_0^t \sigma_s(t, T + x - t)dW_s - \int_t^T \int_0^t b_s(u, T + x - u)dW_sdu = \int_0^t \sigma_s(T, x)dW_s,$$

$$\int_0^t \int_E \delta_s(u, T + x - u, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds)$$

$$- \int_t^T \int_0^t \int_E c_s(u, T + x - u, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds)du$$

$$= \int_0^t \int_E \delta_s(T, x, \xi)(\mathfrak{p}(ds, d\xi) - \nu(d\xi)ds).$$

Note that we may apply the respective stochastic Fubini theorems (Gawarecki & Mandrekar (2011, Thm. 2.8) and Filipović et al. (2010a, Thm. A.2)) due to condition (59). Consequently, by (63), (64) and the previous identities we arrive at identity (16), establishing the second statement.

Now, suppose that, in addition, Assumption C.9 is satisfied, and the drift a is given by (17). By Definitions (14), for all $\xi \in E$ we have

$$\sigma_t(t, x) = 0 \quad \text{and} \quad \delta_t(t, x, \xi) = 0 \quad \text{for all } (t, x) \in \Xi, \tag{65}$$

$$b_t(T, x) = -(\partial_T - \partial_x)\sigma_t(T, x) \quad \text{and} \quad c_t(T, x, \xi) = -(\partial_T - \partial_x)\delta_t(T, x, \xi). \tag{66}$$

Let $(T, x) \in \Xi$ be arbitrary. We also fix an arbitrary $t \in [0, T]$. In view of Definition (45) of Σ, if $-x \geq t$, we obtain by (66), (58) and Lebesgue's dominated convergence theorem

$$(\partial_T - \partial_x)\Sigma_t(T, x) = \partial_h\big|_{h=0}\Sigma_t(T + h, x - h) = -\partial_h\big|_{h=0} \int_{-(x-h)}^{T+h} \sigma_t(u, x - h)du$$

$$= -\partial_h\big|_{h=0} \int_{-x+h}^{T+h} \sigma_t(u, x - h)du = -\partial_h\big|_{h=0} \int_{-x}^{T} \sigma_t(u + h, x - h)du$$

$$= -\int_{-x}^{T} \partial_h\big|_{h=0}\sigma_t(u + h, x - h)du = -\int_{-x}^{T} (\partial_T - \partial_x)\sigma_t(u, x)du = \int_{-x}^{T} b_t(u, x)du;$$

and, if $t \geq -x$, we observe by (65), (66) and (58) together with Lebesgue's dominated convergence theorem

$$(\partial_T - \partial_x)\Sigma_t(T, x) = \partial_h\big|_{h=0}\Sigma_t(T + h, x - h) = -\partial_h\big|_{h=0} \int_{t}^{T+h} \sigma_t(u, x - h)du$$

$$= -\partial_h\big|_{h=0} \int_{t-h}^{T} \sigma_t(u + h, x - h)du = \sigma_t(t, x) - \int_{t}^{T} \partial_h\big|_{h=0}\sigma_t(u + h, x - h)du$$

$$= -\int_{t}^{T} (\partial_T - \partial_x)\sigma_t(u, x)du = \int_{t}^{T} b_t(u, x)du.$$

Performing analogous calculations with Δ_t, we deduce that for each $\xi \in E$ the directional derivative $\partial_T - \partial_x$ of the mappings $(T, x) \mapsto \Sigma_t(T, x)$ and $(T, x) \mapsto \Delta_t(T, x, \xi)$ in (T, x) exists, and we have

$$(\partial_T - \partial_x)\Sigma_t(T, x) = \int_{-x \vee t}^{T} b_t(u, x)du \quad \text{and} \quad (\partial_T - \partial_x)\Delta_t(T, x) = \int_{-x \vee t}^{T} c_t(u, x, \xi)du. \tag{67}$$

We define the stochastic process $\beta : \Omega \times \Theta \to \mathbb{R}$ as

$$\beta_t(T, x) := -\langle \sigma_t(T, x), \Sigma_t(T, x) \rangle_{L_2^0(\mathbb{R})} - \int_E \delta_t(T, x, \xi)\big(e^{\Delta_t(T, x, \xi)} - 1\big)\nu(d\xi).$$

Taking into account (66), (67), Definition (14) of σ and δ, and estimates (61), (62) together with Lebesgue's dominated convergence theorem, the directional derivative $\partial_T - \partial_x$ of β in (T, x) exists, and we have

$$- (\partial_T - \partial_x)\beta_t(T, x)$$

$$= \langle \sigma_t(T, x), (\partial_T - \partial_x)\Sigma_t(T, x) \rangle_{L_2^0(\mathbb{R})} + \langle (\partial_T - \partial_x)\sigma_t(T, x), \Sigma_t(T, x) \rangle_{L_2^0(\mathbb{R})}$$

$$+ \int_E \delta_t(T, x, \xi) \cdot (\partial_T - \partial_x)\big(e^{\Delta_t(T, x, \xi)} - 1\big)\nu(d\xi)$$

$$+ \int_E (\partial_T - \partial_x)\delta_t(T, x, \xi) \cdot \big(e^{\Delta_t(T, x, \xi)} - 1\big)\nu(d\xi),$$

and hence

$$-(\partial_T - \partial_x)\beta_t(T,x) = \langle \sigma_t(T,x), (\partial_T - \partial_x)\Sigma_t(T,x)\rangle_{L_2^0(\mathbb{R})} - \langle b_t(T,x), \Sigma_t(T,x)\rangle_{L_2^0(\mathbb{R})}$$
$$+ \int_E \delta_t(T,x,\xi) \cdot (\partial_T - \partial_x)\Delta_t(T,x,\xi) \cdot e^{\Delta_t(T,x,\xi)}\nu(d\xi)$$
$$- \int_E c_t(T,x,\xi)\big(e^{\Delta_t(T,x,\xi)} - 1\big)\nu(d\xi),$$

so that we arrive at

$$-(\partial_T - \partial_x)\beta_t(T,x) = -\Big\langle \int_t^T b_t(u,T+x-u)du, \int_{-x\vee t}^T b_t(u,x)du \Big\rangle_{L_2^0(\mathbb{R})}$$
$$+ \Big\langle b_t(T,x), \int_{-x\vee t}^T \sigma_t(u,x)du \Big\rangle_{L_2^0(\mathbb{R})}$$
$$- \int_E \Big(\int_t^T c_t(u,T+x-u,\xi)du \Big)\Big(\int_{-x\vee t}^T c_t(u,x,\xi)du \Big)$$
$$\exp\Big(-\int_{-x\vee t}^T \delta_t(u,x,\xi)du \Big)\nu(d\xi)$$
$$- \int_E c_t(T,x,\xi)\Big[\exp\Big(-\int_{-x\vee t}^T \delta_t(u,x,\xi)du \Big) - 1 \Big]\nu(d\xi).$$

Applying Definition (14) of σ and δ once again, we obtain by Definition (17)

$$-(\partial_T - \partial_x)\beta_t(T,x) = a_t(T,x),$$

and hence, by Definition (14) of α we deduce that

$$\alpha_t(T,x) = -\int_t^T a_t(u,T+x-u)du = \beta_t(T,x) - \beta_t(t,T+x-t) = \beta_t(T,x).$$

Consequently, by Theorem 5.11 the $\mathbb{F}$-forward mortality rates μ are consistent, completing the proof. $\qquad\square$

C.3   Proof of Theorem 5.11

In this appendix, we provide the proof of Theorem 5.11. We impose the following assumption:

Assumption C.11. *We suppose that the following conditions are satisfied:*

(i) *The processes $\bar{\alpha}$ and $\bar{\sigma}$ are optional, and $\bar{\delta}$ is predictable.*

(ii) *For each bounded Borel set $B \subset \mathbb{R}_+$ there are random variables $X^{\bar{\alpha}} : \Omega \to \mathbb{R}$ and $X^{\bar{\sigma}}, X^{\bar{\delta}} \in \mathcal{L}^2(\Omega)$ such that*

$$\|\bar{\alpha}_t\| \leq X^{\bar{\alpha}}, \quad \|\bar{\sigma}_t\|_{L_2^0(H)} \leq X^{\bar{\sigma}} \quad and \quad \|\bar{\delta}_t\|_{\mathcal{L}_\nu^2(H)} \leq X^{\bar{\delta}} \quad for\ all\ t \in B.$$

Note that we may regard these processes as mappings $\bar{\alpha} : \Omega \times \mathbb{R}_+ \times \Xi \to \mathbb{R}$, $\bar{\sigma} : \Omega \times \mathbb{R}_+ \times \Xi \to L_2^0(\mathbb{R})$ and $\bar{\delta} : \Omega \times \mathbb{R}_+ \times \Xi \times E \to \mathbb{R}$. We define the stochastic processes $\alpha : \Omega \times \Theta \to \mathbb{R}$, $\sigma : \Omega \times \Theta \to L_2^0(\mathbb{R})$ and $\delta : \Omega \times \Theta \times E \to \mathbb{R}$ as

$$\alpha := \bar{\alpha} \circ \phi, \quad \sigma := \bar{\sigma} \circ \phi \quad and \quad \delta := \bar{\delta} \circ \phi. \tag{68}$$

By Assumption C.11 and estimate (18), the processes α, σ and δ also fulfill Assumption C.1. In addition, we require the following assumption:

Assumption C.12. *We suppose that for each $n \in \mathbb{N}$ there exist a measurable mapping $\rho_n : E \to \mathbb{R}_+$ and a constant $\epsilon_n > 0$ such that (48) is satisfied and we have*

$$|\bar{\delta}_t(s,y,\xi)| \leq \rho_n(\xi) \quad for\ all\ (t,s,y) \in \phi(\Theta_n)\ and\ \xi \in E.$$

Note that Assumption C.12 implies Assumption C.5. Now, we are ready to provide the proof of Theorem 5.11:

Proof of Theorem 5.11. We define the transformed $\mathbb{F}$-forward mortality rates $\mu : \Omega \times \Theta \to \mathbb{R}$ as $\mu := \bar{\mu} \circ \phi$. For all $(t, T, x) \in \Theta$ and $(t, s, y) = \phi(t, T, x) \in \mathbb{R}_+ \times \Xi$ we have

$$\mu_0(T, x) = \mu_0(s + t, y - t) = S_t \mu_0(s, y),$$

and, by Definition (68) we obtain

$$\int_0^t \alpha_s(T, x)ds = \int_0^t \bar{\alpha}_s(T - s, x + s)du = \int_0^t S_{t-s}\bar{\alpha}_s(T - t, x + t)du$$
$$= \int_0^t S_{t-s}\bar{\alpha}_s(s, y)ds.$$

Analogous calculations for σ and δ show that for all (T, x) the $\mathbb{F}$-forward mortality rates $\mu(T, x)$ have the dynamics (9). Moreover, using (68), a straightforward calculation shows that conditions (11) and (24) are equivalent. Therefore, applying Theorem 5.3 completes the proof. $\qquad\square$

C.4   Proof of Theorem 5.13

In this appendix, we provide the proof of Theorem 5.13. We impose the following assumption:

Assumption C.13. *We suppose that the following conditions are satisfied:*

(i) *The processes $\bar{a}$ and $\bar{b}$ are optional, and $\bar{c}$ is predictable.*

(ii) *For each bounded Borel set $B \subset \mathbb{R}_+$ there are random variables $X^{\bar{a}} : \Omega \to \mathbb{R}$ and $X^{\bar{b}}, X^{\bar{c}} \in \mathcal{L}^2(\Omega)$ such that*

$$\|\bar{a}_t\| \leq X^{\bar{a}}, \quad \|\bar{b}_t\|_{L_2^0(H)} \leq X^{\bar{b}} \quad and \quad \|\bar{c}_t\|_{\mathcal{L}_\nu^2(H)} \leq X^{\bar{c}} \quad for\ all\ t \in B.$$

Note that we may also regard these processes as mappings $\bar{a} : \Omega \times \mathbb{R}_+ \times \Xi \to \mathbb{R}$, $\bar{b} : \Omega \times \mathbb{R}_+ \times \Xi \to L_2^0(\mathbb{R})$ and $\bar{c} : \Omega \times \mathbb{R}_+ \times \Xi \times E \to \mathbb{R}$. We define the stochastic processes $a : \Omega \times \Theta \to \mathbb{R}$, $b : \Omega \times \Theta \to L_2^0(\mathbb{R})$ and $c : \Omega \times \Theta \times E \to \mathbb{R}$ as

$$a := \bar{a} \circ \phi, \quad b := \bar{b} \circ \phi \quad and \quad c := \bar{c} \circ \phi. \tag{69}$$

By Assumption C.13 and estimate (18), the processes a, b and c also fulfill Assumption C.8. In addition, we require the following assumption:

Assumption C.14. *We suppose that for each $n \in \mathbb{N}$ there exist a measurable mapping $\rho_n : E \to \mathbb{R}_+$ and a constant $\epsilon_n > 0$ such that (48) is satisfied and we have*

$$|\bar{c}_t(s, y, \xi)| \leq \rho_n(\xi) \quad for\ all\ (t, s, y) \in \phi(\Theta_n)\ and\ \xi \in E.$$

Now, we are ready to provide the proof of Theorem 5.13:

Proof of Theorem 5.13. The identity (29) follows from the variation of constants formula (23) and the Definition (28) of the $\mathbb{F}$-spot mortality rates $\bar{\gamma}$.

We define the transformed $\mathbb{F}$-forward mortality rates $\mu : \Omega \times \Theta \to \mathbb{R}$ as $\mu := \bar{\mu} \circ \phi$ and the transformed $\mathbb{F}$-forward mortality improvements $j : \Omega \times \Theta \to \mathbb{R}$ as $j := \bar{j} \circ \phi$. Using Definitions (68) and (69), analogous calculations as in the proof of Theorem 5.11 show that for all (T, x) the $\mathbb{F}$-forward mortality rates $\mu(T, x)$ have the dynamics (9) and the $\mathbb{F}$-forward mortality rates $j(T, x)$ have the dynamics (12).

Now, identity (30) follows from identity (16) of Theorem 5.5. Furthermore, if Assumption C.14 is satisfied, then condition (31) and Definitions (69) imply that condition (17) holds true. Applying Theorem 5.5 yields that the $\mathbb{F}$-forward mortality rates $\bar{\mu}$ are consistent, which concludes the proof. $\qquad\square$

References

Barbarin, J. (2008), 'Heath-Jarrow-Morton modelling of longevity bonds and the risk minimization of life insurance portfolios', *Insurance Math. Econom.* **43**(1), 41–55.

Bauer, D. (2008), *Stochastic mortality modeling and securitization of mortality risk*, IFA-Schriftenreihe, Ulm.

Bauer, D., F. E. Benth & R. Kiesel (2012), 'Modeling the forward surface of mortality', *SIAM Journal on Financial Mathematics* **3**(1), 639–666.

Bielecki, T. R. & M. Rutkowski (2002), *Credit risk: Modeling, valuation and hedging*, Springer Finance, Springer-Verlag, Berlin.

Biffis, E. (2005), 'Affine processes for dynamic mortality and actuarial valuations', *Insurance Math. Econom.* **37**(3), 443–468.

Biffis, E., M. Denuit & P. Devolder (2010), 'Stochastic mortality under measure changes', *Scand. Actuar. J.* (4), 284–311.

Biffis, E. & Pietro M. (2006), 'A bidimensional approach to mortality risk', *Decisions in Economics and Finance* **29**(2), 71–94.

Björk, T., G. Di Masi, Y. Kabanov & W. Runggaldier (1997), 'Towards a general theory of bond markets', *Finance and Stochastics* **1**(2), 141–174.

Booth, H. (2006), 'Demographic forecasting: 1980 to 2005 in review', *International Journal of Forecasting* **22**, 547–581.

Brémaud, P. (1981), *Point processes and queues*, Springer-Verlag, New York. Martingale dynamics, Springer Series in Statistics.

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

Carmona, R. A. (2007), HJM: a unified approach to dynamic models for fixed income, credit and equity markets, *in* 'Paris-Princeton Lectures on Mathematical Finance 2004', Vol. 1919 of *Lecture Notes in Math.*, Springer, Berlin, pp. 1–50.

Da Prato, G. & J. Zabczyk (1992), *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge.

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

Dellacherie, C & P. A. Meyer (1982), *Probabilités et potentiel*, Hermann, Paris.

Filipović, D. (2001), *Consistency problems for Heath–Jarrow–Morton interest rate models*, number 1760 *in* 'Lecture Notes in Mathematics', Springer-Verlag, Berlin.

Filipović, D. (2009), *Term-structure models: A graduate course*, Springer Finance, Springer-Verlag, Berlin.

Filipović, D., S. Tappe & J. Teichmann (2010*a*), 'Jump-diffusions in Hilbert spaces: Existence, stability and numerics', *Stochastics* **82**(5), 475–520.

Filipović, D., S. Tappe & J. Teichmann (2010*b*), 'Term structure models driven by Wiener processes and Poisson measures: Existence and positivity', *SIAM Journal on Financial Mathematics* **1**(1), 523–554.

Gawarecki, L. & V. Mandrekar (2011), *Stochastic differential equations in infinite dimensions*, Probability and Its Applications, Springer-Verlag, Berlin.

Getoor, R. K. (1975), On the construction of kernels, *in* 'Séminaire de Probabilités IX', number 465 *in* 'Lecture Notes in Mathematics', pp. 443–463.

Hainaut, D. & P. Devolder (2008), 'Mortality modelling with Lévy processes', *Insurance Math. Econom.* **42**(1), 409–418.

Heath, D., R. Jarrow & A. Morton (1992), 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation', *Econometrica* **60**(1), 77–105.

Jacod, J. & A. N. Shiryaev (2003), *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften, 2nd edn, Springer-Verlag, Berlin.

Jacod, J. & P. Protter (2004), *Probability essentials*, Universitext, corrected second edn, Springer-Verlag, Berlin.

Jeanblanc, M., M. Yor & M. Chesney (2009), *Mathematical methods for financial markets*, Springer Finance, Springer-Verlag London Ltd., London.

Kallenberg, O. (2002), *Foundations of modern probability*, Probability and its Applications (New York), second edn, Springer-Verlag, New York.

Luciano, E. & E. Vigna (2008), 'Mortality risk via affine stochastic intensities: Calibration and empirical relevance', *Belg. Actuar. Bull.* **8**(1), 5–16.

Milevsky, M. A. & S. D. Promislow (2001), 'Mortality derivatives and the option to annuitise', *Insurance Math. Econom.* **29**(3), 299–318. 4th IME Conference (Barcelona, 2000).

Miltersen, K. & S.-A. Persson (2005), Is mortality dead? Stochastic force of mortality determined by arbitrage? Working Paper, University of Bergen, www.mathematik.uni-ulm.de/carfi/vortraege/downloads/DeadMort.pdf.

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

Norberg, R. (2010), 'Forward mortality and other vital rates—are they the way forward?', *Insurance Math. Econom.* **47**(2), 105–112.

Prakasa Rao, B. L. S. (2009), 'Conditional independence, conditional mixing and conditional association', *Ann. Inst. Statist. Math.* **61**(2), 441–460.

Prévôt, C., C.-R. Rinke & M. Stollmann (2011), 'Mortality and longevity in reinsurance'. Personal Discussion, Hannover Re.

Schrager, D. F. (2006), 'Affine stochastic mortality', *Insurance Math. Econom.* **38**(1), 81–97.

Zhu, N. & D. Bauer (2011), 'Applications of forward mortality factor models in life insurance practice', *The Geneva Papers* **36**, 567–594.

Zhu, N. & D. Bauer (2012), Coherent modeling of the risk in mortality projections: A semi-parametric approach, *in* 'Essays on Lifetime Uncertainty: Models, Applications, and Economic Implications', PhD Thesis, Georgia State University. http://digitalarchive.gsu.edu/rmi_diss/30/.