The Annals of Applied Probability
2002, Vol. 12, No. 2, 710–729

MARKOVIAN TERM STRUCTURE MODELS IN DISCRETE TIME

BY DAMIR FILIPOVIĆ[1] AND JERZY ZABCZYK[2]

ETH Zurich and Polish Academy of Science

In this article we discuss Markovian term structure models in discrete time and with continuous state space. More precisely, we are concerned with the structural properties of such models if one has the Markov property for a part of the forward curve. We investigate the two cases where these parts are either a true subset of the forward curve, including the short rate, or the entire forward curve. For the former case we give a sufficient condition for the term structure model to be affine. For the latter case we provide a version of the Heath, Jarrow and Morton drift condition. Under a Gaussian assumption a Heath–Jarrow–Morton–Musiela type equation is derived.

1. Forward curve models in discrete time. We consider a discrete trading economy with trading times $t \in \mathbb{N}_0$. Denote by $P(t,T)$ the price of the *zero-coupon bond* at time t that pays a sure unit at time of maturity $T \geq t$. Thus in particular $P(T,T) = 1$. The time t continuously compounded *forward rate* $r(t,k)$ for the period $[t+k, t+k+1]$ is defined by

$$r(t,k) := \log \frac{P(t,t+k)}{P(t,t+k+1)}, \qquad k \in \mathbb{N}_0.$$

Equivalently,

$$P(t,T) = \exp\left(-\sum_{j=0}^{T-t-1} r(t,j)\right), \qquad t = 0,\dots,T-1. \tag{1}$$

The *short rate* $R(t)$ is the continuously compounded rate contracted at time t on a one-period loan starting immediately. By definition, hence $R(t) = r(t,0)$. This defines the *savings account*,

$$B(0) := 1, \qquad B(t) := \exp\left(\sum_{s=0}^{t-1} R(s)\right), \qquad t \in \mathbb{N}.$$

We denote by $K \leq +\infty$ the maximal time to maturity of those bonds which are traded at each calender time t. If K is finite then the forward rates $r(t,j)$,

Received July 2000; revised November 2001.

[1]Supported in part by Credit Suisse.

[2]Supported in part by Levehulm Trust and Polish Committee for Research Grant PBZ KBN-016/PO3/99.

AMS 2000 subject classifications. Primary 91B28; secondary 60J05.

Key words and phrases. Affine term structure, forward curve models, Gaussian term structure, Markov chains in discrete time with continuous state space.

given by (1), are only defined for $j = 0, \dots, K-1$. In any case the sequence $r(t) = (r(t, j))_{0 \le j < K}$ is called the *forward curve* at time t.

Here and subsequently, we let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{N}_0}, \mathbb{P})$ denote a filtered probability space. We suppose that the forward rate processes $r(t, j)$, for all $0 \le j < K$, and thus the bond prices $P(t, T)$, are $(\mathcal{F}_t)$-adapted. The savings account $B(t)$ accordingly is $(\mathcal{F}_t)$-predictable.

Suppose for the moment that $\mathbb{P}$ were the physical measure. The first fundamental theorem of asset pricing in discrete time states that, on a finite time horizion $t = 0, 1, \dots, T < \infty$, the existence of an equivalent martingale measure $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}_T$ is equivalent to the absence of arbitrage. In general, $\mathbb{Q}$ is not unique and there are various ways to distinguish a particular equivalent martingale measure. A detailed exposition of the arbitrage-theory in discrete time can be found in [11], Chapter V. We do not further discuss the issues of incompleteness here. In what follows we are interested in the dynamics of the forward curve process under a generic martingale measure on $\mathcal{F}$, which we shall denote by $\mathbb{P}$. This is expressed by the following assumption.

(NA) For arbitrary $T \in \mathbb{N}_0$ the sequence

$$\frac{P(t, T)}{B(t)}, \qquad t = (T-K)^+, \dots, T,$$

is a martingale.

In this article we will analyze the interplay of **(NA)** and various Markov hypotheses imposed on the forward curve process $r(t)$.

To clarify the terminology we recall some basic concepts. First, we establish the convention that all equalities between random variables hold $\mathbb{P}$-almost surely. Let $(E, \mathcal{E})$ be a measurable space. We write $B_b(E)$ for the space of bounded measurable functions [and, if E is equipped with a topology, $C_b(E)$ for the space of bounded continuous functions]. An $(\mathcal{F}_t)$-adapted sequence $(X(t))$ of E-valued random variables is called a *Markov chain with respect to the filtration* $(\mathcal{F}_t)$ if, for any $\varphi \in B_b(E)$,

$$\mathbb{E}[\varphi(X(t+1)) \mid \mathcal{F}_t] = \mathbb{E}[\varphi(X(t+1)) \mid \sigma(X(t))],$$

where $\sigma(X(t))$ denotes the σ-field of events generated by $X(t)$. Then also

$$\mathbb{E}[\varphi(X(t+1)) \mid \sigma(X(0), \dots, X(t))] = \mathbb{E}[\varphi(X(t+1)) \mid \sigma(X(t))], \tag{2}$$

and if (2) holds, $X(t)$ is simply a *Markov chain*. If, in addition, there exists a sequence of *transition kernels* $P_t(x, \Gamma)$, $x \in E$, $\Gamma \in \mathcal{E}$, such that

$$\mathbb{E}[\varphi(X(t+1)) \mid \sigma(X(0), \dots, X(t))] = P_t\varphi(X(t)),$$

then the sequence $(X(t))$ is called a *Markov chain with transition kernels* P_t. Here we used the notation

$$P_t\varphi(x) = \int_E \varphi(y) P_t(x, dy).$$

A Markov chain is *time-homogeneous* if for all $t \in \mathbb{N}_0$, $P_t = P_0$, and then P_0 is denoted by P.

The remainder of the article is as follows. In Section 2 we consider the case where a finite subset $r^\Gamma(t) := (r(t,0), \dots, r(t,\gamma))$ of the forward curve is a time-homogeneous Markov chain with transition kernel P. Here $\Gamma = \{0, \dots, \gamma\}$ for some $0 \le \gamma < K$. Assumption **(NA)** yields a representation of the forward curve as a function of $r^\Gamma(t)$ (Theorem 1). For $\gamma \ge 1$ this imposes arbitrage restrictions for P (Corollary 2). If P is generated by a continuous convolution semigroup then the term structure is affine (Theorem 5). This is the discrete time analogue to the results in [3] and [5]. Section 2.3 is devoted to the study of affine short rate models. We characterize the shapes of the implied forward curves and examine the limiting behavior of $R(t)$ when t tends to infinity. A concrete example is given in Section 2.3.1.

In Section 3 the entire forward curve $r(t)$ is viewed as a Markov chain on $E \subset \mathbb{R}^K$. From general Markov theory it follows that $r(t)$ will always admit a representation of the form $r(t+1) = F(t, r(t), \xi_{t+1})$, where the noise terms ξ_t are i.i.d. Under **(NA)** there has to be some kind of "drift condition." Theorem 10 gives this condition in terms of the mapping F. In Section 3.2 the Gaussian case is studied. The main result is Theorem 13 which shows that the forward curve can be represented as the solution to the discrete time analogue of the Heath–Jarrow–Morton–Musiela equation [10].

The Appendix contains some classical results for conditional Gaussian distributions in infinite dimension.

We write $\mathbb{R}^n_+ = [0, +\infty)^n$, $\mathbb{R}^n_{++} = (0, +\infty)^n$ and $\mathbb{N}_0 = \{0, 1, \dots\}$. Whenever working with a Hilbert space H, we denote by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the scalar product and the norm, respectively.

We emphasize that the various Markov hypotheses on $r(t)$ are always imposed under the measure $\mathbb{P}$ which is not the physical measure.

2. Partly Markovian forward curves. In this section we consider the case where a finite subset of the forward curve is Markovian. We derive a representation of the forward curve as a function of its Markovian part. A focus will be on affine term structure (in particular short rate) models, where our results can be made more explicit. For the latter we provide a concrete example. For simplicity of presentation we suppose that $K = +\infty$. Most of the following results can be carried over without problem to finite K.

2.1. *Generalities.* Let $\gamma \in \mathbb{N}_0$ and set $\Gamma = \{0, \dots, \gamma\}$. We assume that $r^\Gamma(t) = (r(t,0), \dots, r(t,\gamma))$ follows a time-homogeneous Markov chain on $E \subset \mathbb{R}^{\gamma+1}$ with transition kernel P. We further assume that

$$P\phi = \int_E \phi(y)\, P(\cdot, dy) \in B_b(E), \tag{3}$$

where $\phi(y) := e^{-y_0}$, for $y = (y_0, \dots, y_\gamma) \in E$.

We mention that, without any problems, Γ could be replaced by an arbitrary finite subset of $\mathbb{N}_0$ which contains 0. That is, the short rate $R(t) = r(t, 0)$ has to be a component of the Markovian part in any case.

Define inductively the functions

$$\phi_0 := 1, \qquad \phi_{k+1} := P(\phi\phi_k), \qquad k \in \mathbb{N}_0. \tag{4}$$

Notice that $\phi_k \in B_b(E)$, for $k \in \mathbb{N}_0$, by (3).

THEOREM 1. *Assume that* **(NA)** *holds and that $r^\Gamma(t)$ is a time-homogeneous Markov chain with respect to the filtration $(\mathcal{F}_t)$. Then*

$$r(t, k+1) = \ln\left(\frac{\phi_k}{\phi_{k+1}}\left(r^\Gamma(t)\right)\right) \qquad \forall\, t,\ k \in \mathbb{N}_0. \tag{5}$$

PROOF. Let $T \in \mathbb{N}$. From **(NA)** we have

$$\frac{P(t,T)}{B(t)} = \mathbb{E}\left[\frac{1}{B(T)} \,\Big|\, \mathcal{F}_t\right], \qquad t \le T.$$

Therefore

$$P(t,T) = \exp\left(-\sum_{j=0}^{T-t-1} r(t,j)\right) = \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-1} R(s)\right) \Big|\, \mathcal{F}_t\right], \qquad t \le T-1.$$

Hence

$$\begin{aligned} P(t,T) &= \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-1} R(s)\right) \Big|\, \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-2} R(s)\right) \mathbb{E}\left(e^{-R(T-1)} \mid \mathcal{F}_{T-2}\right) \Big|\, \mathcal{F}_t\right] \end{aligned}$$

and by the Markov property

$$P(t,T) = \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-2} R(s)\right) P\phi\left(r^\Gamma(T-2)\right) \Big|\, \mathcal{F}_t\right].$$

In the same way, taking into account that $P\phi = \phi_1$,

$$\begin{aligned} P(t,T) &= \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-3} R(s)\right) \mathbb{E}\left(e^{-R(T-2)}\phi_1\left(r^\Gamma(T-2)\right) \mid \mathcal{F}_{T-3}\right) \Big|\, \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\exp\left(-\sum_{s=t}^{T-3} R(s)\right) \phi_2\left(r^\Gamma(T-3)\right) \Big|\, \mathcal{F}_t\right]. \end{aligned}$$

By induction,

$$P(t,T) = e^{-R(t)}\phi_{T-t-1}(r^\Gamma(t)), \qquad t \le T-1.$$

Consequently, for arbitrary $t, k \in \mathbb{N}_0$,

$$\exp\left(-\sum_{j=0}^{k} r(t,j)\right) = e^{-R(t)}\phi_k(r^\Gamma(t)).$$

Or equivalently, since $R(t) = r(t,0)$,

$$\sum_{j=1}^{k} r(t,j) = -\ln\phi_k(r^\Gamma(t)),$$

which yields the assertion. □

It is easily seen that, for $\gamma \ge 1$, equation (5) imposes arbitrage constraints on the transition kernel P.

COROLLARY 2. *Suppose $\gamma \ge 1$ and that the assumptions of Theorem* 1 *hold for any initial point $r^\Gamma(0) = x \in E$. Then necessarily*

$$\ln\phi_k(x) = -\sum_{j=1}^{k} x_j \qquad \forall\, x = (x_0, \dots, x_\gamma) \in E,\ \forall\, k = 1, \dots, \gamma. \tag{6}$$

In (6) there are γ conditions for the transition kernel P to be satisfied. They can be made explicit as we shall see in Theorem 5 below. Conditions (6) are implied by the fact that the dynamics of the bond prices $P(t,T)$, for $0 \le t \le T$ and $1 \le T \le \gamma + 1$ are directly specified by P via (1). This requires consistency with condition **(NA)**. In contrast, the bond prices $P(t,T)$ with time to maturity $T - t > \gamma + 1$ are *defined by*—and hence consistent with—**(NA)**. Notice that for $\gamma = 0$ (i.e., Markovian short rates) there are no constraints since in this case, condition **(NA)** is trivially satisfied:

$$\frac{P(t,t+1)}{B(t)} = \frac{e^{-R(t)}}{\exp(\sum_{s=0}^{t-1} R(s))} = \exp\left(-\sum_{s=0}^{t} R(s)\right) = \frac{1}{B(t+1)}.$$

In other words, in a Markovian short rate model every bond price is given as a derivative by **(NA)**. We also refer to the discussion in [1], Section 16.1.

2.2. *Affine term structure.* We now shall determine a class of transition kernels P for which the functions ϕ_k can be calculated explicitly. In view of (4), a candidate is given by any P which transforms exponential functions into exponential functions. The Lévy–Khintchine formula (see Proposition 3 below) tells us that such measures $P(x,\cdot)$ are infinitely divisible and have the convolution semigroup property with respect to x. We arrive this way at the so called *affine term structure models*. Let $m, n \in \mathbb{N}$.

DEFINITION 1. A family of probability measures $(\mu_x)_{x\in\mathbb{R}^n_+}$ on $\mathbb{R}^m_+$ is called a *continuous convolution semigroup* if

$$\mu_{x+y} = \mu_x * \mu_y \qquad \forall\, x, y \in \mathbb{R}^n_+$$

and $x \mapsto \mu_x$ is weakly continuous. That is,

$$\int_{\mathbb{R}^m_+} f\, d\mu_{x^k} \to \int_{\mathbb{R}^m_+} f\, d\mu_x \qquad \forall f \in C_b(\mathbb{R}^m_+)$$

whenever $x^k \to x$.

In particular, each μ_x is infinitely divisible and $\mu_0 = \delta_0$. It is easy to see that $(\mu_x)_{x\in\mathbb{R}^n_+}$ is a continuous convolution semigroup if and only if $\mu_x = \mu^1_{x_1} * \cdots * \mu^n_{x_n}$, for $x = (x_1, \dots, x_n) \in \mathbb{R}^n_+$, where each $(\mu^i_t)_{t\in\mathbb{R}_+}$ is a continuous convolution semigroup. In fact, $\mu^i_t = \mu_{te_i}$, where e_i is the ith standard basis vector in $\mathbb{R}^n$.

The following result is a corollary of the classical Lévy–Khintchine formula (see [4], Section XIII.7).

PROPOSITION 3. *A family of probability measures $(\mu_x)_{x\in\mathbb{R}^n_+}$ on $\mathbb{R}^m_+$ is a continuous convolution semigroup if and only if the Laplace transform of μ_x is of the form*

$$\widetilde{\mu_x}(\lambda) := \int_{\mathbb{R}^m_+} e^{-\langle\lambda, y\rangle}\, \mu_x(dy) = e^{-\langle\psi(\lambda), x\rangle}, \qquad x \in \mathbb{R}^n_+,\ \lambda \in \mathbb{R}^m_+,$$

where $\psi = (\psi_1, \dots, \psi_n)$ with

$$\psi_i(\lambda) = \langle\beta_i, \lambda\rangle + \int_{\mathbb{R}^m_{++}} \big(1 - e^{-\langle\lambda, y\rangle}\big) m_i(dy), \tag{7}$$

for $\beta_i \in \mathbb{R}^m_+$ and nonnegative measures $m_i(dy)$ on $\mathbb{R}^m_{++}$ such that

$$\int_{\mathbb{R}^m_{++}} (1 \wedge \|y\|) m_i(dy) < +\infty, \qquad 1 \le i \le n.$$

Based on these facts we now construct a Markov chain model $r^\Gamma(t)$. To be consistent with the notation in Section 2.1 we set $m = n = \gamma + 1 \in \mathbb{N}$ and let $E = \mathbb{R}^m_+$. Accordingly, we write $x = (x_0, \dots, x_\gamma) \in \mathbb{R}^m_+$, and $\{e_0, \dots, e_\gamma\}$ for the standard basis in $\mathbb{R}^m$. Suppose $(\mu_x)_{x\in\mathbb{R}^m_+}$ is a continuous convolution semigroup. Then $P(x, dy) = \mu_x(dy)$ is a Markov transition kernel on $\mathbb{R}^m_+$. Since $P(0, dy) = \delta_0(dy)$ the point 0 is absorbing. This can be relaxed as shown in the next lemma.

LEMMA 4. *Let ν be a probability measure on $\mathbb{R}^m_+$ with Laplace transform $\tilde{\nu} = e^{-\varphi}$ and $(\mu_x)_{x\in\mathbb{R}^m_+}$ as above with ψ given by* (7). *Define*

$$P(x, dy) := \nu * \mu_x(dy), \qquad x \in \mathbb{R}^m_+. \tag{8}$$

Then

$$\phi_k(x) = \exp(-A_k - \langle B_k, x\rangle), \qquad k \in \mathbb{N}_0, \tag{9}$$

where $A_0 := 0$, $B_0 := 0$ *and*

$$A_{k+1} := A_k + \varphi(B_k + e_0), \qquad B_{k+1} := \psi(B_k + e_0), \qquad k \in \mathbb{N}_0. \tag{10}$$

PROOF. We proceed inductively. By definition (4), the statement is true for $k = 0$. Now let $k \in \mathbb{N}_0$ and calculate

$$\begin{aligned}\phi_{k+1}(x) = P(\phi\phi_k)(x) &= e^{-A_k} \int_{\mathbb{R}_+^m} e^{-\langle B_k+e_0, y\rangle} P(x, dy) \\ &= e^{-A_k} e^{-\varphi(B_k+e_0)-\langle\psi(B_k+e_0),x\rangle},\end{aligned}$$

which yields the assertion. □

In the present setup, (6) reads as follows.

THEOREM 5. *Suppose* **(NA)** *holds for every initial point* $r^\Gamma(0) = x \in \mathbb{R}_+^m$. *If* $\gamma \geq 1$ *then necessarily*

$$\varphi\left(\sum_{j=0}^{k-1} e_j\right) = 0, \qquad \psi\left(\sum_{j=0}^{k-1} e_j\right) = \sum_{j=1}^{k} e_j \qquad \forall\, k = 1, \ldots, \gamma. \tag{11}$$

Accordingly, we have

$$\operatorname{supp} \nu \subset \{y \in \mathbb{R}_+^m \mid y_0 = \cdots = y_{\gamma-1} = 0\} \tag{12}$$

and, for all $t \geq 0$,

$$\operatorname{supp} \mu_t^j \subset \begin{cases} \{y \in \mathbb{R}_+^m \mid y_0 = \cdots = y_{\gamma-1} = 0\}, & j = 0, \\ \{y \in \mathbb{R}_+^m \mid y_0 = \cdots = y_{j-2} = 0\}, & j = 2, \ldots, \gamma. \end{cases} \tag{13}$$

The resulting forward curve is an affine function of $r^\Gamma(t)$,

$$r(t, k+1) = A_{k+1} - A_k + \langle B_{k+1} - B_k, r^\Gamma(t)\rangle, \qquad k \in \mathbb{N}_0, \tag{14}$$

with A_k, B_k *as in Lemma* 4. *Here in particular*,

$$A_k = 0, \qquad B_k = \sum_{j=1}^{k} e_j \qquad \forall\, k = 1, \ldots, \gamma. \tag{15}$$

If $\gamma = 0$ *then we only conclude* (14).

PROOF. Equation (14) follows directly by Theorem 1 and (9), for all $\gamma \geq 0$. Now suppose that $\gamma \geq 1$. According to (6) we have

$$A_k + \langle B_k, x \rangle = \sum_{j=1}^{k} x_j \qquad \forall\, x \in \mathbb{R}_+^m,\ \forall\, k = 1, \dots, \gamma.$$

This yields (15). Equation (11) is now a direct consequence of (10). From (11) we conclude that

$$\int_{\mathbb{R}_+^m} \exp\left(-\sum_{i=0}^{\gamma-1} y_i\right) \nu(dy) = 1$$

and, by the property $\widetilde{\mu_x} = \widetilde{\mu_{x_0}^0} \cdots \widetilde{\mu_{x_\gamma}^\gamma}$,

$$\int_{\mathbb{R}_+^m} \exp\left(-\sum_{i=0}^{k-1} y_i\right) \mu_t^j(dy) = \begin{cases} 1, & j = 0 \text{ or } k < j \leq \gamma, \\ e^{-t}, & 2 \leq j \leq k. \end{cases}$$

This yields (12) and (13). □

REMARK 6. We have noticed in the preceding proof that (9) implies (14). Conversely, if (14) holds for every initial point $r^\Gamma(0) = x \in \mathbb{R}_+^m$ then this yields (9). Equation (9) holds since, by (8), the Laplace transform of the transition kernel is exponential-affine in x,

$$\widetilde{P}(x, \lambda) = e^{-\varphi(\lambda) - \langle \psi(\lambda), x \rangle}. \tag{16}$$

Hence we have the implications

$$(8) \Rightarrow (16) \Rightarrow (9) \Leftrightarrow \big((14)\ \forall\, r^\Gamma(0) \in \mathbb{R}_+^m\big).$$

We will show in the next section (Proposition 9) that, for $m = 1$ and under some mild conditions, (16) and (9) are equivalent. But (16) does not imply (8), in general. A counterexample has been found by F. Hubalek [8]. This is in contrast to continuous time Markov models, where (16) and (8) are equivalent (see [5]).

Finally, we give a slightly alternative description of the process $r^\Gamma(t)$. Let X be a random variable with distribution ν. Let L^j be the (increasing) Lévy process specified in distribution by μ_t^j (that is, $L_t^j \sim \mu_t^j$), for $j = 0, \dots, \gamma$. We assume that X and $L^0, \dots, L^\gamma$ are mutually independent. Then we have, in distribution,

$$r^\Gamma(t+1) = X + L_{r(t,0)}^0 + \cdots + L_{r(t,\gamma)}^\gamma, \qquad t \in \mathbb{N}_0. \tag{17}$$

Of course, for each t we have to chose an independent copy of the family $X, L^0, \dots, L^\gamma$. Representation (17) clarifies the interplay between the different components in the dynamics of $r^\Gamma(t)$.

In the case of $\gamma \geq 1$, (17) gives a better understanding of (12) and (13). Indeed, by (12) only the last component of X is different from 0. By (13), the above Lévy processes are of the form $L_t^0 = (0, \dots, 0, *)$ and

$$\begin{pmatrix} L_t^1 \\ L_t^2 \\ \vdots \\ L_t^\gamma \end{pmatrix} = \begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & * \end{pmatrix},$$

where $*$ stands for a generic nonnegative number. To give an illustration of this particular structure, suppose that the present Markovian part of the forward curve is flat zero, $r^\Gamma(0) = (r(0,0), \dots, r(0,\gamma)) = (0, \dots, 0)$. By simple arbitrage considerations it is clear that $r(1, j) = 0$ for all $0 \leq j \leq \gamma - 1$. Indeed, since $P(0, 1) = P(0, \gamma + 1) = 1$, we can enter the following strategy at time $t = 0$ at zero cost: buy one bond maturing at 1, sell one bond maturing at $\gamma + 1$. At time $t = 1$ we get one dollar from the first bond. We immediately reinvest this dollar and buy $1/P(1, \gamma + 1)$ bonds maturing at $\gamma + 1$. If $r(1, j) > 0$ for some $0 \leq j \leq \gamma - 1$ then

$$1/P(1, \gamma + 1) = \exp\left(\sum_{j=0}^{\gamma-1} r(1, j)\right) > 1,$$

see (1). Hence we realize a net gain at time $t = \gamma + 1$, which means arbitrage. By the same reasoning one shows that

$$r(t, j) = 0 \qquad \forall\, 0 \leq j \leq \gamma - t,\ t = 1, \dots, \gamma.$$

It is now easy to comprehend this phenomenon by simply looking at (17), given the particular structure of X and $L^0, \dots, L^\gamma$.

2.3. *Affine short rate models.* In this section we investigate the case where $\gamma = 0$ (that is, $m = 1$) in more detail. First, we discuss the possible shapes of the implied forward curve (14). Let A_k and B_k be given as in Lemma 4. Write

$$a_k := \begin{cases} 0, & \text{for } k = 0, \\ A_k - A_{k-1}, & \text{for } k \geq 1, \end{cases} \qquad b_k := \begin{cases} 1, & \text{for } k = 0, \\ B_k - B_{k-1}, & \text{for } k \geq 1. \end{cases}$$

Now (14) reads

$$r(t, k) = a_k + b_k R(t), \qquad k \in \mathbb{N}_0. \tag{18}$$

If ν is the Dirac measure at 0 then $\varphi \equiv 0$ and therefore $a_k = 0$, for all $k \in \mathbb{N}_0$. In the sequel we shall exclude this trivial case and suppose that $\nu((0, +\infty)) > 0$ and similarly $\mu_x((0, +\infty)) > 0$, for $x > 0$. We use the notation as in Proposition 3 but skip the indices since now $m = 1$. Write

$$\Delta := \beta + \int_{\mathbb{R}_{++}} y\, m(dy) \leq +\infty.$$

By monotone convergence we have

$$\lim_{\lambda\downarrow 0}\psi'(\lambda)=\beta+\lim_{\lambda\downarrow 0}\int_{\mathbb{R}_{++}} ye^{-\lambda y}m(dy)=\Delta.$$

PROPOSITION 7. *The sequence (a_k) is strictly increasing with*

$$\lim_{k\to+\infty} a_k \begin{cases} =+\infty, & \text{if } \beta\geq 1,\\ <+\infty, & \text{if } \beta<1.\end{cases}$$

If $\beta>1$ then (b_k) is strictly increasing and $\lim_{k\to+\infty} b_k=+\infty$.

If $\beta=1$ then (b_k) is nondecreasing with finite limit.

If $\beta<1$ then $\lim_{k\to+\infty} b_k=0$, and there exists $k^*\in\mathbb{N}_0$ such that $(b_k)_{k\leq k^*}$ is nondecreasing and $(b_k)_{k\geq k^*}$ is strictly decreasing. Necessary for $k^*\geq 1$ is $\Delta\geq 1$. Then (b_k) has a hump.

PROOF. Since $\psi(\lambda)$ is strictly increasing in λ, the sequence (B_k) is strictly increasing. If $\beta<1$ then its limit is finite. If $\beta\geq 1$, the limit is infinite. Since $a_k=\varphi(B_{k-1}+1)$, for $k\in\mathbb{N}$, the first part of the proposition is established. We claim that

$$b_{k+1}\geq\beta\, b_k \qquad \forall\, k\in\mathbb{N}_0.$$

Indeed, $b_1=\psi(1)\geq\beta$ and for $k\geq 1$ we have

$$b_{k+1}=\psi(B_k+1)-\psi(B_{k-1}+1)\geq\beta(B_k-B_{k-1})=\beta\, b_k.$$

Taking into account that $\psi'(\lambda)=\beta+\int_{\mathbb{R}_{++}} ye^{-\lambda y}m(dy)\to\beta$ for $\lambda\to+\infty$, the rest of the proposition follows. $\square$

We now examine the limiting behavior of the short rate process $R(t)$. Denote by P^n the nth iterate transition kernel ($P^n(x,\cdot)$ is the distribution of $R(n)$ given that $R(0)=x$).

PROPOSITION 8. *If $\Delta>1$ then P^n converges weakly to $\delta_{+\infty}$ on $\overline{\mathbb{R}_+}$ (the one-point compactification of $\mathbb{R}_+$).*

If $\Delta<1$ and $\int_{\mathbb{R}_+} y\nu(dy)<\infty$ then P^n converges weakly to an invariant measure μ^* on $\mathbb{R}_+$. Hence the Markov chain $R(t)$ is strongly mixing.

PROOF. By the Chapman–Kolmogorov equation, the Laplace transform of P^n is

$$\widetilde{P^n}(x,\lambda)=e^{-\varphi(n,\lambda)-\psi(n,\lambda)x},$$

where

$$\begin{aligned}&\varphi(n+1,\lambda):=\varphi(n,\lambda)+\varphi\bigl(\psi(n,\lambda)\bigr), \qquad \varphi(1,\lambda):=\varphi(\lambda),\\ &\psi(n+1,\lambda):=\psi\bigl(\psi(n,\lambda)\bigr), \qquad \psi(1,\lambda):=\psi(\lambda),\ n\in\mathbb{N}.\end{aligned}\tag{19}$$

Notice that $\psi''(\lambda) = -\int_{\mathbb{R}_{++}} y^2 e^{-\lambda y} m(dy) \le 0$, hence ψ is concave. We thus have

$$0 \le \psi'(\lambda) \le \Delta \qquad \forall\, \lambda \in \mathbb{R}_+.$$

If $\Delta > 1$ then $\lim_{n\to\infty} \psi(n,\lambda) = \lambda^*$, for all $\lambda > 0$, for some $\lambda^* \in (0, +\infty]$. Hence $\lim_{n\to\infty} \varphi(n,\lambda) = +\infty$, for all $\lambda > 0$. Therefore $\lim_{n\to\infty} \widetilde{P^n}(x,\lambda) = 0$, for all $\lambda > 0$, and the first part of the proposition is proved.

Suppose $\Delta < 1$ and $\int_{\mathbb{R}_+} y\nu(dy) = \lim_{\lambda\downarrow 0} \varphi'(\lambda) < \infty$. Then ψ is contracting on $\mathbb{R}_+$,

$$\psi(n+1,\lambda) = \psi\big(\psi(n,\lambda)\big) \le \Delta\psi(n\lambda) \le \Delta^n \psi(\lambda), \qquad n \in \mathbb{N}_0.$$

In particular, $\lim_{n\to\infty} \psi(n,\lambda) = 0$ uniformly in λ on compacts. On the other hand,

$$\begin{aligned} \big|\varphi(n+k,\lambda) - \varphi(n,\lambda)\big| &\le \sum_{j=1}^{k} \big|\varphi(n+j,\lambda) - \varphi(n+j-1,\lambda)\big| \\ &\le \sum_{j=1}^{k} C\Delta^{n+j-1}\psi(\lambda) = C\Delta^n \frac{1-\Delta^k}{1-\Delta}\psi(\lambda), \end{aligned}$$

for some $C < \infty$, for n large enough. Hence $\varphi(n,\cdot)$ converges uniformly on compacts to a function φ^* and

$$\lim_{n\to\infty} \widetilde{P^n}(x,\lambda) = e^{-\varphi^*(\lambda)}.$$

This specifies μ^*. Since $P^n f \in C_b(\mathbb{R}_+)$ for $f \in C_b(\mathbb{R}_+)$ and $\sup_{x\in\mathbb{R}_+} |P^n f(x)| \le \sup_{x\in\mathbb{R}_+} |f(x)|$, it follows by dominated convergence that

$$\int_{\mathbb{R}_+} \left(P^n f(x) - \int_{\mathbb{R}_+} f(y)\mu^*(dy) \right)^2 \mu^*(dx) \to 0 \qquad \text{for } n \to \infty.$$

Hence the Markov chain $R(t)$ is strongly mixing. $\square$

The next proposition was announced in Remark 6.

PROPOSITION 9. *If*

$$\sum_{k\in\mathbb{N}} \frac{1}{B_k} = +\infty \tag{20}$$

then (16) *and* (9) *are equivalent*.

PROOF. The implication (16) $\Rightarrow$ (9) is trivial. Now suppose (9) holds. An easy calculation shows that

$$\int_{\mathbb{R}_+} e^{-(B_k+1)y}\, P(x,dy) = e^{-(A_{k+1}-A_k) - B_{k+1}x}, \qquad k \in \mathbb{N}_0.$$

Thus, for $x, y \in \mathbb{R}_+$ fixed,

$$\widetilde{P}(x+y,\lambda)\widetilde{P}(0,\lambda) = \widetilde{P}(x,\lambda)\widetilde{P}(y,\lambda), \tag{21}$$

for all $\lambda = B_k + 1$, $k \in \mathbb{N}_0$. But the product of two Laplace transforms is again a Laplace transform. Since (20) is equivalent to $\sum_k (B_k+1)^{-1} = +\infty$, Müntz' theorem applies (see [4], Section XIII.1). It states that a Laplace transform is uniquely determined by the sequence $B_k + 1$. Hence (21) holds for all $\lambda \in \mathbb{R}_+$. Now we fix $\lambda \in \mathbb{R}_+$ and define $g(x) := \widetilde{P}(x,\lambda)/\widetilde{P}(0,\lambda)$. This function is measurable, positive, bounded and satisfies the functional equation $g(x)g(y) = g(x+y)$. Hence there exists $\psi(\lambda) \in \mathbb{R}_+$ such that $g(x) = \exp(-\psi(\lambda)x)$. We can write $\phi(\lambda) = -\ln \widetilde{P}(0,\lambda)$, and (16) follows. □

2.3.1. *Examples.* We illustrate some possible choices of $P(x, dy)$ (see also [4], Chapter XIII.7).

Compound Poisson distributions. Let F be a probability distribution on $\mathbb{R}_+$ and $\alpha > 0$. Then

$$\mu_x := e^{-\alpha x} \sum_{n \in \mathbb{N}_0} \frac{(\alpha x)^n}{n!} F^{n*}$$

defines a continuous convolution semigroup on $\mathbb{R}_+$. Here $\psi(\lambda) = \alpha(1 - \tilde{F}(\lambda))$, and (7) is true with $m(dy) = \alpha F(dy)$ and $\beta = 0$.

If $F = \delta_1$, then μ_x is the ordinary Poisson distribution with expectation αx and $\psi(\lambda) = \alpha(1 - e^{-\lambda})$.

As a concrete example consider $\psi(\lambda) = 3/4\lambda + 1/5(1 - e^{-\lambda})$ and $\varphi(\lambda) = \ln(1 + \lambda/5)$. That is, $\nu(dy) = 5e^{-5y}\,dy$. Then the assumptions of the second part of Proposition 8 are satisfied and the Markov chain is strongly mixing. The representation (17) reads

$$R(t+1) = X + L_{R(t)},$$

where X is exponentially distributed with expectation $1/5$ and $L_t = 3/4t + N_t$, where N is a Poisson process with intensity $1/5$. The components of the forward curve (18) are shown in Figure 1.

Gamma distributions. For $a, b > 0$ let $f_{a,b}(y) := \frac{1}{\Gamma(b)} a^b y^{b-1} e^{-ay}$ denote the density of the corresponding gamma distribution. Then

$$\mu_x(dy) := f_{a,bx}(y)\,dy, \qquad x > 0,$$

defines a continuous convolution semigroup on $\mathbb{R}_+$. Here $\psi(\lambda) = b\log(1 + \frac{\lambda}{a})$ and (7) holds with $m(dy) = abe^{-ay}\frac{dy}{y}$ and $\beta = 0$, which is seen by differentiation. The choice of a, b is made according to Propositions 7 and 8 and with regard to $\psi'(0) = b/a$.

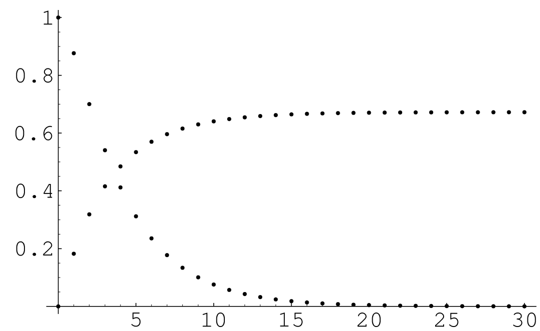

FIG. 1. *The sequences a_k (increasing) and b_k (decreasing), for $k = 0, \dots, 30$.*

Stable distributions on $\mathbb{R}_+$. For $0 < \alpha < 1$ the function $\psi(\lambda) = \lambda^\alpha$ can be expressed with $m(dy) = \frac{\alpha}{\Gamma(1-\alpha)} \frac{dy}{y^{\alpha+1}}$. This is seen again by differentiation. Hence ψ defines a continuous convolution semigroup $(\mu_x)_{x\in\mathbb{R}_+}$ and each μ_x is stable. However, since here $\Delta = \lim_{\lambda\downarrow 0} \alpha\lambda^{\alpha-1} = +\infty$, stable distributions are not really convenient for our setup (see Propositions 7 and 8).

3. Markovian forward curves. We now consider the case where the *entire* forward curve is a Markov chain. We also give an alternative, equivalent description of the Markov chain as a dynamical system.

It is well known that Markov chains on rather general measurable spaces can be regarded (have the same laws) as solutions of stochastic difference equations,

$$X(t+1) = F\big(t, X(t), \xi_{t+1}\big), \qquad t \in \mathbb{N}_0, \tag{22}$$

where $\xi_1, \xi_2, \dots$ is a sequence of independent identically distributed random variables taking values in $E_0 = \mathbb{R}^d$, or even in [0, 1], independent of X_0. Moreover, for each $t \in \mathbb{N}_0$, $F(t, \cdot, \cdot)$ is a measurable mapping from $E \times E_0$ into E. For a representation of this type it is sufficient that the state space E is a Borel subset of a separable, complete metric space (see, e.g., [12]). If the Markov chain is time-homogeneous, the function on the right-hand side of (22) does not explicitly depend on t.

3.1. *General Markovian term structure.* In the present subsection K might be finite or infinite. We assume that $r(t)$ is a Markov chain on a Borel set $E \subset \mathbb{R}^K$ and is a solution of (22). If $x = (x_0, \dots) \in E$ and $z \in \mathbb{R}^d$ then we set

$$F(t, x, z) = \big(F_0\big(t, (x_0, \dots), z\big), \dots, F_j\big(t, (x_0, \dots), z\big), \dots\big).$$

The following theorem is the analogue of Theorem 1.

THEOREM 10. *Assume that $r(t)$ is a Markov chain, given by* (22), *with respect to the filtration $(\mathcal{F}_t)$ and that the σ-fields $\mathcal{F}_t$ are independent of $\xi_{t+1}, \xi_{t+2}, \dots$.*

Define

$$\mathbb{F}_J(t,x) := \mathbb{E}\left[\exp\left(-\sum_{j=0}^{J} F_j(t,x,\xi)\right)\right], \qquad x \in E,\ t,\ J = 0,1,\ldots,$$

where ξ *is a random variable with the same distribution as all* ξ_t. *If*

$$\mathbb{F}_J(t,x) = \exp\left(-\sum_{j=0}^{J} x_{j+1}\right), \qquad x \in E,\ t = 0,1,\ldots, \tag{23}$$

for $J = 0,1,\ldots,K-2$ *if* K *is finite*, *and for* $J \in \mathbb{N}_0$ *if* K *is infinite*. *Then the martingale hypothesis* **(NA)** *holds*.

PROOF. Assume for instance that $K < +\infty$. Note that for $2 \le T - t \le K$,

$$\begin{aligned}\mathbb{E}\left[\frac{P(t+1,T)}{B(t+1)} \,\Big|\, \mathcal{F}_t\right] &= \mathbb{E}\left[\exp\left(-\sum_{s=0}^{t} R(s)\right)\exp\left(-\sum_{j=0}^{T-t-2} r(t+1,j)\right) \Big|\, \mathcal{F}_t\right]\\ &= \exp\left(-\sum_{s=0}^{t} R(s)\right)\mathbb{E}\left[\exp\left(-\sum_{j=0}^{T-t-2} F_j(t,r(t),\xi_{t+1})\right) \Big|\, \mathcal{F}_t\right]\\ &= \frac{1}{B(t)} e^{-r(t,0)}\mathbb{F}_{T-t-2}\big(t,r(t)\big).\end{aligned}$$

The final identity is a consequence of the imposed properties on $(\mathcal{F}_t)$. Since (23) holds, the result follows. □

REMARK 11. If **(NA)** holds for any initial state $r(0) = x \in E$, the condition of Theorem 10 is also necessary. It is important to remark that if $K < +\infty$, then the function $\mathbb{F}_{K-1}$ is not determined by the theorem and in fact it can be arbitrary. It means, in practical terms, that the dynamics of the long rate, $r(t, K-1)$, has to be additionally specified.

REMARK 12. The case of binomially distributed random variables ξ_t was analyzed in particular in Jarrow's book [9] (see also the references therein).

3.2. *Markov Gaussian term structure.* In this section we regard $r(t)$ as a process on $E = \mathbb{R}^K$. If $K = +\infty$ we treat $E = \mathbb{R}^{+\infty}$ as a metric space with coordinate-wise convergence. An E-valued random variable $X = (X_0, X_1, \ldots)$ is called *Gaussian* if any arbitrary finite subset of the random variables $\{X_0, X_1, \ldots\}$ is Gaussian. The definition can be extended to any family of random variables X_t.

Let H be a separable Hilbert space. An H-valued random variable ξ is said to be Gaussian with *mean vector* m and *covariance operator* Q if for arbitrary $h \in H$, $\langle \xi, h\rangle$ is a real-valued Gaussian random variable and

$$\mathbb{E}[\langle \xi, h\rangle] = \langle m, h\rangle, \qquad \mathbb{E}[\langle \xi, h\rangle\langle \xi, g\rangle] = \langle Qh, g\rangle, \qquad h, g \in H.$$

Arbitrage-free Markov Gaussian forward curve processes can be nicely characterized. For simplicity of representation we write $\sum_{k=0}^{-1} \cdots := 0$.

THEOREM 13. *Let $K = +\infty$ and assume that $r(t)$ is a sequence of E-valued Gaussian random variables, which is Markovian with respect to $(\mathcal{F}_t)$.*

(i) *If* **(NA)** *holds then there exists a sequence of independent E-valued Gaussian random variables $\xi_1, \xi_2, \dots$ with coordinates $\xi_t(j)$, $j = 0, 1, \dots$, such that*

$$r(t+1) = Ar(t) + \tfrac{1}{2}a_t + \xi_{t+1}, \qquad t = 0, 1, \dots, \tag{24}$$

where $A = (\alpha_{ij})$ is the left-shift operator with

$$\alpha_{ij} = \begin{cases} 1, & \text{if } j = i+1, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

and

$$a_t(j) = \mathbb{E}\left[\xi_{t+1}(j)\left(2\sum_{k=0}^{j-1} \xi_{t+1}(k) + \xi_{t+1}(j)\right)\right]. \tag{26}$$

(ii) *Conversely, if $r(t)$ is defined by* (24)–(26), *where $\xi_1, \xi_2, \dots$ are E-valued, independent and Gaussian such that $\mathcal{F}_t$ is independent of $\xi_{t+1}, \xi_{t+2}, \dots$, then* **(NA)** *is satisfied.*

PROOF. Part (ii) follows from Theorem 10. For the proof of part (i) we need some well-known results for Gaussian Markov chains, which are sketched in the Appendix. On a finite time horizon we can regard $r(t)$ as a Hilbert space valued Gaussian sequence, choosing as the Hilbert space the set $H = l^2_\rho$ of of all sequences $x = (x_0, x_1, \dots)$ such that

$$\|x\|^2_\rho = \sum_{j=0}^{+\infty} \rho_j x_j^2 < +\infty,$$

equipped with the norm $\|\cdot\|_\rho$. Here (ρ_j) is a sequence of positive numbers tending to 0 sufficiently fast. Consequently, by Proposition 17,

$$r(t+1) = \zeta_t + m_{t+1} + \xi_{t+1},$$

where ζ_t is $\sigma(r(t))$-measurable and ξ_{t+1} is an H-valued Gaussian random variable independent of $\sigma(r(0), \dots, r(t))$. Component-wise we thus have

$$r(t+1, j) = \zeta_t(j) + m_{t+1}(j) + \xi_{t+1}(j), \qquad j = 0, 1, \dots. \tag{27}$$

Hypothesis **(NA)** is equivalent to the identities

$$\mathbb{E}\left[\exp\left(-\sum_{j=0}^{J} r(t+1, j)\right) \Big| \mathcal{F}_t\right] = \exp\left(-\sum_{j=1}^{J+1} r(t, j)\right), \qquad J = 0, 1, \dots. \tag{28}$$

Taking into account (27) and the Markov property of the process r one obtains that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\sum_{j=0}^{J} r(t+1,j)\right)\Big|\,\mathcal{F}_t\right]\\
&\quad=\mathbb{E}\left[\exp\left(-\sum_{j=0}^{J} r(t+1,j)\right)\Big|\,\sigma\big(r(t)\big)\right]\\
&\quad=\mathbb{E}\left[\exp\left(-\sum_{j=0}^{J}\big(m_{t+1}(j)+\zeta_t(j)+\xi_{t+1}(j)\big)\right)\Big|\,\sigma\big(r(t)\big)\right]\\
&\quad=\exp\left(-\sum_{j=0}^{J}\big(m_{t+1}(j)+\zeta_t(j)\big)\right)\mathbb{E}\left[\exp\left(\sum_{j=0}^{J}\xi_{t+1}(j)\right)\right]\\
&\quad=\exp\left(-\sum_{j=0}^{J}\big(m_{t+1}(j)+\zeta_t(j)\big)+\tfrac{1}{2}\mathbb{E}\left[\left(\sum_{j=0}^{J}\xi_{t+1}(j)\right)^2\right]\right).
\end{aligned}$$

Then we arrive at the following identities:

$$\sum_{j=0}^{J}\big(m_{t+1}(j)+\zeta_t(j)\big)-\tfrac{1}{2}\mathbb{E}\left[\left(\sum_{j=0}^{J}\xi_{t+1}(j)\right)^2\right]=\sum_{j=1}^{J+1} r(t,j),\qquad J=0,\dots,$$

from which

$$\zeta_t(j)=r(t,j+1)-m_{t+1}(j)+\tfrac{1}{2}\mathbb{E}\left[\left(\sum_{k=0}^{j}\xi_{t+1}(k)\right)^2-\left(\sum_{k=0}^{j-1}\xi_{t+1}(k)\right)^2\right].\tag{29}$$

Inserting (29) into (27) we obtain the required result. □

REMARK 14. Thus it follows from rather general conditions on the evolution of the forward curve that $r(t)$ necessarily satisfies a linear stochastic difference equation with the left-shift matrix operator A and the drift vector linked to the driving noise through the generalized HJM drift condition (26) (see [6], [7]).

REMARK 15. A similar result can be obtained for $K<+\infty$ with the exception that the component $r(t,K-1)$ is not determined by hypothesis **(NA)** (cf. Remark 11). This is related to the fact that the stochastic difference equation for $r(t)$ is only well posed with a boundary condition.

APPENDIX

Here we provide the material needed for the proof of Theorem 13, part (i).

First, we have to define images of Gaussian random variables by unbounded linear transformations. Let ξ be a Gaussian random variable with mean vector 0 and covariance operator Q, taking values in a separable Hilbert space U. Assume that B is a linear operator with domain $\mathcal{D}(B) \supset \operatorname{Im} Q^{1/2}$ and values in some separable Hilbert space H. Let $\{e_k : k = 1, 2, \dots\}$ be the orthonormal sequence of all eigenvectors of Q corresponding to nonzero eigenvalues of Q and let P_k be the orthogonal projection onto the finite-dimensional space spanned by $e_1, \dots, e_k$,

$$P_k u = \sum_{j=1}^{k} \langle u, e_j \rangle e_j.$$

One easily checks that if the operator $BQ^{1/2}$ is Hilbert–Schmidt then the formula

$$B\xi = \lim_{k \to \infty} BP_k \xi,$$

with limit in the $L^2(\Omega, H)$ norm, defines an H-valued Gaussian random variable with mean zero and covariance operator $(BQ^{1/2})(BQ^{1/2})^*$. In the sequel, if ξ is a Hilbert space valued Gaussian random variable with mean 0 and B a linear operator, $B\xi$ is a Gaussian random variable defined in the above way.

The pseudo-inverse $Q^{-1/2}$ of $Q^{1/2}$ has domain $\mathcal{D}(Q^{-1/2}) = \operatorname{Im} Q^{1/2}$ and $Q^{-1/2}u$ is defined as the element of $Q^{-1/2}(\{u\})$ with the minimal norm. We point out that if $B : U \to H$ is a Hilbert–Schmidt operator then $BQ^{-1/2}\xi$ is a well-defined H-valued Gaussian random variable with covariance operator BB^* [this is immediate from the above discussion since $(BQ^{-1/2})Q^{1/2} = B$ is Hilbert–Schmidt].

We recall a result on *conditional Gaussian distributions*, see [13]. A Gaussian measure with mean m and covariance Q is denoted by $\mathcal{N}_{m,Q}$.

PROPOSITION 16. *Suppose (X, Y) is a Gaussian random variable with values in a separable Hilbert space $U \times V$, with mean vector (m_X, m_Y) and covariance operator*

$$\begin{pmatrix} Q_{XX}, & Q_{XY} \\ Q_{YX}, & Q_{YY} \end{pmatrix}.$$

Then $\operatorname{Im} Q_{YX} \subset \operatorname{Im} Q_{YY}^{1/2}$ *and the operator* $Q_{YY}^{-1/2} Q_{YX}$ *is Hilbert–Schmidt. Moreover for an arbitrary Borel map $\phi : U \to \mathbb{R}_+$,*

$$\mathbb{E}[\phi(X)|Y] = \int_U \phi(x) \mathcal{N}_{\hat{X}, \hat{Q}}(dx),$$

where

$$\hat{X} = \mathbb{E}[X|Y] = m_X + (Q_{YY}^{-1/2} Q_{YX})^* Q_{YY}^{-1/2}(Y - m_Y)$$

and

$$\hat{Q} = Q_{XX} - (Q_{YY}^{-1/2} Q_{YX})^* (Q_{YY}^{-1/2} Q_{YX}).$$

PROOF. Denote by $\{e_k\}$ an orthonormal basis in U. Let $v \in V$. Then we have

$$\begin{aligned}\|Q_{XY}v\|_U^2 &= \sum_{k\in\mathbb{N}} \langle Q_{XY}v, e_k\rangle_U^2 = \sum_{k\in\mathbb{N}} \big(\mathbb{E}[\langle X, e_k\rangle_U \langle Y, v\rangle_V]\big)^2 \\ &\le \sum_{k\in\mathbb{N}} \mathbb{E}[\langle X, e_k\rangle_U^2]\mathbb{E}[\langle Y, v\rangle_V^2] \\ &= \mathbb{E}[\|X\|_U^2]\langle Q_{YY}v, v\rangle_V = \mathbb{E}[\|X\|_U^2]\|Q_{YY}^{1/2}v\|_V^2.\end{aligned}$$

By [2], Proposition B.1, we conclude that $\operatorname{Im} Q_{YX} \subset \operatorname{Im} Q_{YY}^{1/2}$.

Now let $\{f_k\}$ be an orthonormal basis of V, consisting of eigenvectors of Q_{YY}, such that

$$Q_{YY} f_k = \lambda_k f_k, \qquad k \in \mathbb{N}.$$

We may assume that $\lambda_k > 0$ (otherwise we chose V smaller by skipping the corresponding f_k). Then $\{\langle Y, f_k\rangle_V / \sqrt{\lambda_k}\}$ is an orthonormal basis of $L^2(\Omega)$,

$$\mathbb{E}\left[\frac{\langle Y, f_k\rangle_V}{\sqrt{\lambda_k}} \frac{\langle Y, f_l\rangle_V}{\sqrt{\lambda_l}}\right] = \delta_{kl}.$$

Let $u \in U$. We now have

$$\begin{aligned}\langle Q_{YY}^{-1/2} Q_{YX}u, Q_{YY}^{-1/2} Q_{YX}u\rangle_V &= \sum_{k\in\mathbb{N}} \langle Q_{YY}^{-1/2} Q_{YX}u, f_k\rangle_V^2 \\ &= \sum_{k\in\mathbb{N}} \left\langle Q_{YX}u, \frac{1}{\sqrt{\lambda_k}} f_k\right\rangle_V^2 \\ &= \sum_{k\in\mathbb{N}} \left(\mathbb{E}\left[\langle X, u\rangle_U \frac{\langle Y, f_k\rangle_V}{\sqrt{\lambda_k}}\right]\right)^2 \\ &= \mathbb{E}[\langle X, u\rangle_U^2] = \langle Q_{XX}u, u\rangle_U.\end{aligned}$$

However, Q_{XX} is trace class, hence $Q_{YY}^{-1/2} Q_{YX}$ is Hilbert–Schmidt.

The rest of the proposition follows as in the finite-dimensional case (see, e.g., [12], Theorem 11.1.1). □

PROPOSITION 17. *Assume that an H-valued Gaussian sequence (X_t) is a Markov chain on H. Then there exists a sequence of closed (in general unbounded) linear operators $A_0, A_1, \dots$ on H and a sequence of independent H-valued Gaussian random variables $\xi_1, \xi_2, \dots$, independent of X_0, such that*

$$X_{t+1} = A_t(X_t - m_t) + m_{t+1} + \xi_{t+1}.$$

Sequences as described in Proposition 17 are called *Ornstein–Uhlenbeck processes*. For related results see, for example, [12], where a finite dimensional case is treated.

PROOF. Suppose η is a U-valued Gaussian random variable with covariance operator Q. Denote by $\mathbb{L}^2_H(\eta)$ the closed subspace of $L^2(\Omega, H)$, consisting of all H-valued Gaussian random variables of the form $a + B(\eta - \mathbb{E}[\eta])$ where $a \in H$ and $BQ^{1/2}$ is a Hilbert–Schmidt operator. The norm on $\mathbb{L}^2_H(\eta)$ is induced by $L^2(\Omega, H)$,

$$\|\zeta\|^2_{\mathbb{L}^2_H(\eta)} = \mathbb{E}\big[\|\zeta\|^2_H\big] = \|a\|^2_H + \|BQ^{1/2}\|^2_{HS},$$

for $\zeta = a + B(\eta - \mathbb{E}[\eta]) \in \mathbb{L}^2_H(\eta)$.

Since the sequence (X_t) is Markovian, $\mathbb{E}[X_{t+1}|X_0, \dots, X_t] = \mathbb{E}[X_{t+1}|X_t]$. Moreover, by Proposition 16, the random variable $\mathbb{E}[X_{t+1}|X_0, \dots, X_t]$ is the $L^2(\Omega, H)$-orthogonal projection of X_{t+1} onto $\mathbb{L}^2_H(X_0, \dots, X_t)$. Consequently, the Gaussian random variable $\xi_{t+1} = X_{t+1} - \mathbb{E}[X_{t+1}|X_0, \dots, X_t]$ is orthogonal to $\mathbb{L}^2_H(X_0, \dots, X_t)$, hence in particular independent of $X_0, \dots, X_t$. By induction ξ_{t+1} is independent of $X_0, \xi_1, \dots, \xi_t$. From Proposition 16 we have

$$\mathbb{E}[X_{t+1}|X_t] = \mathbb{E}[X_{t+1}] + \big(Q^{-1/2}_{X_t X_t} Q_{X_t X_{t+1}}\big)^* Q^{-1/2}_{X_t X_t}(X_t - \mathbb{E}[X_t]),$$

where $Q_{X_t X_t}$ and $Q_{X_t X_{t+1}}$ denote the covariance operators of (X_t, X_t) and (X_t, X_{t+1}), respectively. We conclude that

$$X_{t+1} = \mathbb{E}[X_{t+1}|X_t] + \xi_{t+1} = A_t(X_t - m_t) + m_{t+1} + \xi_{t+1},$$

where $A_t = (Q^{-1/2}_{X_t X_t} Q_{X_t X_{t+1}})^* Q^{-1/2}_{X_t X_t}$. $\square$

Acknowledgments. We thank F. Hubalek for providing the counterexample which is mentioned in Remark 6. We also thank the anonymous referees for some useful comments.

REFERENCES

[1] BJÖRK, T. (1998). *Arbitrage Theory in Continuous Time*. Oxford Univ. Press.

[2] DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions*. Cambridge Univ. Press.

[3] DUFFIE, D. and KAN, R. (1996). A yield-factor model of interest rates. *Math. Finance* **6** 379–406.

[4] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

[5] FILIPOVIĆ, D. (2001). A general characterization of one-factor affine term structure models. *Finance and Stochastics* **5** 389–412.

[6] HEATH, D., JARROW, R. and MORTON, A. (1989). Bond pricing and the term structure of interest rates: a discrete time approximation. *J. Financial Qualitative Anal.* **25** 419–440.

[7] HEATH, D., JARROW, R. and MORTON, A. (1992). Bond pricing and the term structure of interest rates: A new methodology. *Econometrica* **60** 77–101.

[8] HUBALEK, F. (2001). A counterexample involving exponential-affine Laplace transforms. Working paper, TU Vienna.

[9] JARROW, R. (1996). *Modelling Fixed Income Securities and Interest Rate Options*. McGraw-Hill, New York.

[10] MUSIELA, M. (1993). Stochastic PDEs and term structure models. *Journées Internationales de Finance*, IGR-AFFi, La Boule.
[11] SHIRYAEV, A. N. (1999). *Essentials of Stochastic Finance*. World Scientific, Singapore.
[12] ZABCZYK, J. (1996). *Chance and Decision*. Quaderni, Scuola Normale Superiore di Pisa.
[13] ZABCZYK, J. (2002). *Selected Topics in Stochastic Processes*. Appunti dei Corsi, Scuola Normale Superiore di Pisa, 1997–1999. To appear.

DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
E-MAIL: dfilipov@princeton.edu

POLISH ACADEMY OF SCIENCE
P-00-950 WARSAW
POLAND
E-MAIL: zabczyk@panim.impan.gov.pl