

Article

# Applying Heath-Jarrow-Morton Model to Forecasting the US Treasury Daily Yield Curve Rates

Valerii Maltsev <sup>1,†</sup>  and Michael Pokojovy <sup>2,\*,†</sup> 

<sup>1</sup> Department of Mathematics and Statistics, University of Konstanz, 78464 Konstanz, Germany; valerii.maltsev@uni-konstanz.de

<sup>2</sup> Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX 79968, USA

\* Correspondence: mpokojovy@utep.edu

† These authors contributed equally to this work.

**Abstract:** The Heath-Jarrow-Morton (HJM) model is a powerful instrument for describing the stochastic evolution of interest rate curves under no-arbitrage assumption. An important feature of the HJM approach is the fact that the drifts can be expressed as functions of respective volatilities and the underlying correlation structure. Aimed at researchers and practitioners, the purpose of this article is to present a self-contained, but concise review of the abstract HJM framework founded upon the theory of interest and stochastic partial differential equations in infinite dimensions. To illustrate the predictive power of this theory, we apply it to modeling and forecasting the US Treasury daily yield curve rates. We fit a non-parametric model to real data available from the US Department of the Treasury and illustrate its statistical performance in forecasting future yield curve rates.

**Keywords:** Heath-Jarrow-Morton model; zero-coupon bonds; forward rates; SPDE; arbitrage-free

**MSC:** 35R60; 60H15; 60H20; 60H35; 62M20; 65C30



**Citation:** Maltsev, V.; Pokojovy, M. Applying Heath-Jarrow-Morton Model to Forecasting the US Treasury Daily Yield Curve Rates. *Mathematics* **2021**, *9*, 114. <https://doi.org/10.3390/math9020114>

Received: 28 November 2020

Accepted: 29 December 2020

Published: 6 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Stocks and bonds play a crucial role in making investment decisions. However, there is a significant difference between stock and bond markets. A stock market is a platform enabling companies to issue their shares and trade them on a stock exchange. Buying these shares and having ownership stake, investors can benefit from the performance of the issuer in the future. In contrast, a bond market is used for issuing and trading debt securities. Usually, an investor and a bond issuer make a contract enabling the former to benefit from the interest paid on a regular basis and full nominal amount received when reaching maturity. The bond issuer may be a government, a corporation, etc.

Besides, there are bonds that do not pay an intermediate interest but are traded at a deep discount, rendering profit at the maturity when they are redeemed for their full face value. These bonds are called zero-coupon bonds. For a maturity  $T \geq 0$ , let  $P(t, T)$  denote the price of a zero-coupon bond at time  $t \in [0, T]$ . It is known from the theory of interest rates (viz. Definition 10 in Section 2) that the instantaneous forward rate of a zero-coupon bond is given by

$$f(t, T) = -\frac{\partial \ln P(t, T)}{\partial T} \quad \text{for all } t \in [0, T], \quad T > 0. \quad (1)$$

Under appropriate assumptions, Heath, Jarrow and Morton [1] demonstrated that  $f(t, T)$  can be modeled as a Markovian diffusion that evolves according to the following stochastic differential equation (SDE):

$$\begin{cases} df(t, T) = \alpha(t, T)dt + \sigma(t, T) \cdot dW(t), \\ f(0, T) = f^0(T), \end{cases} \quad (2)$$

where  $(W(t))_{t \in [0, T]}$  is a  $d$ -dimensional standard Brownian motion,  $T \mapsto f^0(T)$  is the market instantaneous forward curve at time  $t = 0$ ,  $(\alpha(t, T))_{t \in [0, T]}$  and  $(\sigma(t, T))_{t \in [0, T]}$  are 1- and  $d$ -dimensional stochastic processes adapted to the natural filtration of  $(W(t))_{t \in [0, T]}$  and  $\sigma(t, T) \cdot dW(t) = \sum_{k=1}^d \sigma_k(t, T) dW_k(t)$  is formally defined as the scalar product. Under suitable regularity assumptions, integrating Equation (2) over  $[0, t]$  for  $t \in [0, T]$ , the instantaneous forward rate  $f(t, T)$  can be represented as an Itô-process:

$$f(t, T) = f(0, T) + \int_0^t \alpha(s, T) ds + \int_0^t \sigma(s, T) \cdot dW(s). \tag{3}$$

However, the dynamics in Equation (2) is not necessarily arbitrage-free. In the spirit of [2], we understand the presence of arbitrage as the existence of a self-financing strategy which yields a positive income in the future without investing a positive amount today (cf. Section 4). It was further proved in [3] that the existence of an equivalent martingale measure implies the absence of arbitrage opportunities. It was also shown that the unique equivalent martingale measure exists if the drift  $\alpha$  depends on the volatility  $\sigma$  (see ([2], Chapter 2) for details). In particular, the requirement that the evolution in Equation (2) is taking place under a risk-neutral measure  $\mathbb{Q}$  necessitates the following representation:

$$\alpha(t, T) = \sigma(t, T) \cdot \int_t^T \sigma(t, s) ds = \sum_{k=1}^d \sigma_k(t, T) \int_t^T \sigma_k(t, s) ds \quad \text{for all } t \in [0, T].$$

Therefore, for all  $t \in [0, T]$ , we conclude that  $f(t, T)$  is fully specified (as a stochastic process) as long as the volatility vector  $\sigma$  is provided:

$$f(t, T) = f(0, T) + \sum_{k=1}^d \int_0^t \sigma_k(\tau, T) \int_\tau^T \sigma_k(\tau, s) ds d\tau + \sum_{k=1}^d \int_0^t \sigma_k(s, T) dW_k(s).$$

Recalling Equation (1), Itô’s lemma furnishes the following SDE for the zero-coupon bond price  $P(t, T)$  for all  $t \in [0, T]$ :

$$dP(t, T) = P(t, T) \left( r_t dt - \left( \int_t^T \sigma(t, s) ds \right) dW(t) \right),$$

where  $r_t$  is the instantaneous short-term interest rate at  $t$  given by

$$\begin{aligned} r_t &= f(t, t) = f(0, t) + \int_0^t \sigma(\tau, t) \int_\tau^t \sigma(\tau, s) ds d\tau + \int_0^t \sigma(s, t) dW(s) \\ &= f(0, t) + \sum_{k=1}^d \int_0^t \sigma_k(\tau, t) \int_\tau^t \sigma_k(\tau, s) ds d\tau + \sum_{k=1}^d \int_0^t \sigma_k(s, t) dW_k(s). \end{aligned}$$

In his dissertation [4], Filipović proposed to extend this approach to an infinite-dimensional framework and consider the following stochastic *partial* differential equation (SPDE):

$$\begin{cases} df_t(x) &= \left( \frac{d}{dx} f_t(x) + \alpha_t(x) \right) dt + \sigma_t(x) \cdot dW(t), \\ f_t(x) &= f^0(x), \end{cases} \tag{4}$$

with the so-called Musiela’s notation (cf. [5])  $f_t(x) := f(t, t + x)$ ,  $x \in [0, T - t]$  being employed (see also [6], Chapter 6). Here,  $(W(t))_{t \in [0, T]}$  is an infinite-dimensional vector of (countably many) independent real-valued Brownian motions, the forward rate curve  $x \mapsto f_t(x)$  is a random variable taking its values in a function space  $\mathcal{H}$  and the drift  $\alpha$  is given by the no-arbitrage condition

$$\alpha_t(x) := \alpha(t, t + x) = \sum_{k=1}^{\infty} \sigma_k(t, t + x) \left( \int_0^x \sigma_k(t, t + s) ds \right).$$

To facilitate the analytic treatment of Equation (4), we put it into the framework of abstract operator SDEs with stochastic coefficients:

$$\begin{cases} dX_t &= (AX_t + F(t, X_t))dt + B(t, X_t)dW(t), \\ X_0 &= X^0, \end{cases} \tag{5}$$

where  $X$  is an  $\mathcal{H}$ -valued random process,  $X^0$  is a random  $\mathcal{H}$ -valued  $\mathcal{F}_0$ -measurable initial value,  $A: \mathcal{H} \rightarrow \mathcal{H}$  is a linear (possibly unbounded) operator,  $(W(t))_{t \in [0, T]}$  is a standard cylindrical Wiener process over  $\mathcal{U}$ ,  $F: [0, T] \times \Omega \times \mathcal{H} \rightarrow \mathcal{H}$ ,  $B: [0, T] \times \Omega \times \mathcal{H} \rightarrow \mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})$  and  $\mathcal{H}, \mathcal{U}$  are separable Hilbert spaces. Here and in the sequel,  $\mathcal{L}_{HS}(\cdot, \cdot)$  denotes the space of Hilbert-Schmidt operators (cf. Definition A8). Letting

$$A := \frac{d}{dx}, \quad X_t := f_t, \quad F := \alpha_t, \quad B := \langle \sigma_t, \cdot \rangle.$$

we can easily see that Equation (4) is a special case of Equation (5). A straightforward integration of Equation (5) over  $[0, t]$  yields the integral formulation of Equation (5):

$$X_t = X^0 + \int_0^t (AX_s + F(s, X_s))ds + \int_0^t B(s, X_s)dW(s) \quad \text{for all } t \in [0, T]. \tag{6}$$

The rest of the paper is structured as follows. Sections 2–4 are meant to provide a concise review of the HJM framework and related results. With most of the results being scattered over a number of articles and monographs, our thrust was to give a coherent presentation on this matter with consistent terminology and notation to make the theory better accessible to a broad scientific and professional audience. In Section 2, a brief summary of the interest rate theory is given, in particular, a relation between the price of a zero-coupon bond and its instantaneous forward rate from Equation (1) is established. In Section 3, we introduce the abstract HJM model first presented and studied by [4,6], discuss major advantages of this approach and protocol some known well-posedness results. In Section 4, we summarize central properties of the abstract HJM model, in particular, we give sufficient conditions under which the abstract HJM model defines an arbitrage-free market, describe the long rates associated with the model, and give an example of a state space compatible with the instantaneous forward rate dynamics. Section 5 contains the main contribution of this paper. A (semi-)implicit numerical scheme is developed and applied to Equation (4). Using real data from the US Department of the Treasury [7], the resulting non-parametric model is calibrated using a custom-developed least-squares-like procedure involving principal component analysis (PCA) for parameter estimation and dimension reduction. The simplicity of our approach is one of its main advantages. Indeed, it does not rely on any sophisticated optimization techniques or third-party libraries, thus, making itself attractive and accessible to a wide audience. The model is then deployed to forecast future yield curve rates, estimate the mean future yield curve and compute prediction regions. Using the predictive power of the model as a benchmark measure in lieu of the mere calibration quality is another novelty of our contribution. To the best of our knowledge, this is the first time an infinite-dimensional model is applied to analyze the US Treasury’s daily yield curve rates. We refer [8–10] for the finite-dimensional treatise. In Section 6, conclusions are presented. Appendices A–C contain all seminal probabilistic and functional-analytic concepts and results [6,11–15] used in the article.

## 2. Interest Rate Theory

Before presenting the abstract Heath-Jarrow-Morton model in Section 3, we give a concise summary of the theory of interest rates based on ([2] pp. 2–18). The goal of this section is twofold. First, we want to give motivation for using the theory of interest rates in this paper. Second, we aim to prove relation (1) between the price and the instantaneous forward rate of a zero-coupon bond in order to define the abstract HJM model afterwards.

As pointed out in ([16], Chapter 25), multivariate (or multidimensional) models have numerous advantages over single-state short rate models, in particular, in describing the intrinsic volatility structure or explaining the inversion of the yield curve, etc. The HJM framework, being at “the far end of this [multivariate] spectrum” (cf. [16], p. 388), is, probably, the most comprehensive one as it employs the forward rate curve as an infinite-dimensional state variable. We refer to the aforementioned monograph as well as ([17], pp. 242–245) for further discussions.

Application areas of the HJM framework are very diverse. They include, but are not limited to, modeling of interest futures [18], risk management [19], American option pricing [20], credit risk modeling [21], investment management [22], etc.

### 2.1. Bank Account and Short Rate

Depositing a certain amount of money in a bank account is typically expected to guarantee its growth at some rate. We need appropriate definitions to express this concept in mathematical terms. Consider a bank account which represents a riskless investment with the profit accruing continuously.

**Definition 1** (Bank account). *Let  $B(t)$  be the value of a bank account at time  $t \geq 0$ . We assume that the bank account evolves according to the differential equation*

$$\begin{cases} dB(t) = r_t B(t) dt, \\ B(0) = B_0, \end{cases} \tag{7}$$

where  $r_t$  is a positive function of time. Letting  $B_0 = 1$ , we have

$$B(t) = \exp\left(\int_0^t r_s ds\right). \tag{8}$$

Definition 1 suggests that a unit amount invested at  $t = 0$  has value (8) at  $t > 0$ , implying also that  $B(t)$  accrues at the short rate  $r_t$ .

In fact, the first-order Taylor expansion of  $B(t + \Delta t)$  is

$$B(t + \Delta t) = B(t)(1 + r_t \Delta t),$$

whence

$$\frac{B(t + \Delta t) - B(t)}{B(t)} = r_t \Delta t \quad \text{as } \Delta t \rightarrow 0.$$

**Definition 2** (Numéraire). *A numeraire is any positive non-dividend-paying asset.*

**Remark 1.** *The concept of bank account numéraire is important for relating amounts of currencies available at different times. Assume for simplicity that  $r_t$  is deterministic. If  $K$  units of currency are deposited into a bank account at the time  $t = 0$ , the amount of units of currency at time  $t \geq 0$  is  $KB(t)$ . Analogously, the amount is  $KB(T)$  at time  $T \geq t$ . If we want to have one unit of currency at time  $T$ , i.e.,  $KB(T) = 1$ , we have to initially invest  $K = 1/B(T)$ . Hence, the value of the invested amount  $K$  at time  $t$  is*

$$KB(t) = \frac{B(t)}{B(T)}.$$

Remark 1 motivates the following definition (cf. [2], Chapter 1).

**Definition 3** (Stochastic discount factor). *The amount at time  $t$  “equivalent” to one unit of currency payable at time  $T$*

$$D(t, T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r_s ds\right).$$

*is referred to as stochastic discount factor.*

## 2.2. Zero-Coupon Bonds and Spot Interest Rates

**Definition 4** (Zero-coupon bond). *A zero-coupon bond is a contract that entitles its holder to a (single) payment of one unit of currency at maturity  $T$ . Denoting the contract value at time  $t \leq T$  by  $P(t, T)$ , we observe  $P(T, T) = 1$  for all  $T \geq 0$ .*

**Definition 5** (Time to maturity). *The amount of time  $T - t$  from the present moment of time  $t$  to the maturity time  $T > t$  is referred to as time to maturity  $T - t$ .*

Obviously,  $T - t$  is well-defined if  $t$  and  $T$  are real numbers associated with some time instants. However, if  $t$  and  $T$  denote two dates  $D_1, D_2$ , e.g., of format (month/day/year), we can define the amount of time between them in terms of the number of days. Moreover, the number of days between  $D_1$  and  $D_2$  is calculated according to appropriate market convention, which specifies how these days are to be counted.

**Definition 6** (Year fraction). *Let  $\tau(t, T)$  denote a time measure between  $t$  and  $T$ , which is usually referred to as year fraction between the dates  $t$  and  $T$ . If  $t$  and  $T$  are less than one day apart (typically, occurring when dealing with limiting quantities which involve time to maturities tending to zero),  $\tau(t, T)$  is roughly given by  $T - t$ . A particular choice is known as the day-count convention (see [2,23] for examples).*

Additionally, we have to exploit another feature of the zero-coupon bonds called the compounding type in order to switch from the price of a zero-coupon bond and its instantaneous forward rate and vice versa.

**Definition 7** (Continuously-compounded spot interest rate). *The (constant) rate*

$$R(t, T) := -\frac{\ln P(t, T)}{\tau(t, T)}$$

*at which an investment of amount  $P(t, T)$  at time  $t$  accrues continuously to produce a unit amount of currency at maturity  $T$  is referred to as continuously-compounded spot interest rate.*

**Definition 8** (Simply-compounded spot interest rate). *Given an initial capital of  $P(t, T)$  at time  $t$  and assuming accruing takes place proportionally to the investment time, the (constant) rate*

$$L(t, T) := \frac{1 - P(t, T)}{\tau(t, T)P(t, T)}$$

*at which money needs to be invested to yield a single unit of currency at maturity  $T$  is called the simply-compounded spot interest.*

## 2.3. Forward Rates

Forward rates are interest rates that can be locked in on a particular day for an investment in a future time period. They are characterized by three time instants: the time  $t$  at which the rate is considered, its expiry  $T$  and its maturity  $S$ ,  $t < T < S$ .

The forward rate can be also viewed as a typical forward rate agreement represented by a contract involving three time characteristics: the current time  $t$ , the expiry time  $T > t$  and the maturity  $S > T$ . It enables the investor to benefit from the interest paid

between  $T$  and  $S$ . Moreover, it is possible to fix the interest between  $T$  and  $S$  at rate  $K$  with simply-compounded contract rates, implying a payment of  $\tau(T, S)L(T, S)N$  and a receipt of  $\tau(T, S)KN$ , with  $N$  denoting the contract nominal value. This renders the contract value to be  $N\tau(T, S)(K - L(T, S))$ . Recalling Definition 8, the former quantity rewrites it as

$$N\left(\tau(T, S)K - \frac{1}{P(T, S)} + 1\right). \tag{9}$$

Let  $1/P(T, S)$  be the amount of currency held at time  $S$ . Its value at time  $T$  is obtained by multiplying this amount with the zero coupon price  $P(T, S)$ , i.e.,

$$P(T, S)(1/P(T, S)) = 1,$$

so this term is equivalent to holding one unit of currency at the time  $T$ . Besides, one unit of currency at the time  $T$  is worth  $P(t, T)$  units of currency at the time  $t$ . Therefore, the amount  $1/P(T, S)$  in  $S$  is equivalent to an amount of  $P(t, T)$  in  $t$ . Now, consider the remaining terms in Equation (9). The worth of  $W := \tau(T, S)K + 1$  at the maturity  $S$  at the time  $t$  is

$$P(t, S)W = P(t, S)\tau(T, S)K + P(t, S).$$

Therefore, the total value of the contract at the time  $t$  is

$$N(P(t, S)\tau(T, S)K - P(t, T) + P(t, S)). \tag{10}$$

**Remark 2.** The value of  $K$  rendering the contract in Equation (10) fair at the time  $t$  is unique.

**Definition 9** (Simply-compounded forward interest rate). Given expiry  $T > t$  and maturity  $S > T$ , the simply-compounded forward interest rate at time  $t$  reads as

$$F(t; T, S) := \frac{1}{\tau(T, S)} \left( \frac{P(t, T)}{P(t, S)} - 1 \right).$$

As the maturity of the forward rate approaches its expiry, the concept of instantaneous forward rate emerges. If  $S$  is sufficiently close to  $T$ , then  $\tau(T, S) = S - T$  and we have:

$$\begin{aligned} \lim_{S \rightarrow T} F(t; T, S) &= - \lim_{S \rightarrow T} \frac{1}{P(t, S)} \frac{P(t, S) - P(t, T)}{S - T} \\ &= - \frac{1}{P(t, T)} \frac{\partial P(t, T)}{\partial T} \\ &= - \frac{\partial \ln P(t, T)}{\partial T}. \end{aligned}$$

This motivates the following definition.

**Definition 10** (Instantaneous forward interest rate). The instantaneous forward interest rate prevailing at time  $t$  for the maturity  $T > t$  is denoted by  $f(t, T)$  and given as

$$f(t, T) = \lim_{S \rightarrow T} f(t; T, S) = - \frac{\partial \ln P(t, T)}{\partial T}.$$

**Remark 3.** Using Musiela’s notation (viz. [5]),  $f_t(x) := f(t, t + x)$  for  $x = T - t$  and  $t, x \geq 0$ . The space of functions  $f_t$  will be denoted by  $\mathcal{S}$  and called the state space.

### 3. The Abstract Heath-Jarrow-Morton Model

Filipović [4] proposed to model the instantaneous forward rate of a zero-coupon bond via an infinite-dimensional SDE driven by the standard cylindrical Wiener process. In contrast to the standard finite-dimensional Heath-Jarrow-Morton (HJM) framework, the

underlying bond market captures the entire (unobservable) infinite-dimensional interest rate term structure. Among many other advantages, the infinite-dimensional HJM model mitigates the so-called “maturity specific risk” typically present in finite-dimensional models. For a more detailed discussion of infinite-dimensional models, the reader may be referred to the monograph ([6], Chapter 6.5). Also, it gives a natural way to “extrapolate” outside of the original finite maturity grid. From the empirical point of view, the infinite-dimensional structure of the model not only helps better mimic the observed dynamics by merely increasing the “number” of model parameters, but makes it possible to capture the non-trivial topological aspects of the dynamics, thus, improving the predictive power and allowing for natural dimension reduction or regularization without resorting to *ad hoc* penalties. Arguably, this justifies the increased price associated with the infinite-dimensional functional-analytic and probabilistic framework.

In this section, we introduce the abstract HJM model following the approach of [4,6]. The resulting model is a semilinear stochastic PDE (SPDE) driven by the standard cylindrical Wiener process (see Appendix C). The well-posedness of these abstract SPDEs in the class of mild solutions is obtained using the abstract results from ([11], Chapter 7) and discussed in Section 3.1. The results are then applied to justify the well-posedness of the abstract HJM model to be introduced in Section 3.2.

### 3.1. Existence and Uniqueness of Mild Solutions to Semilinear Spdes

We consider the abstract SPDE

$$\begin{cases} dX_t &= (AX_t + F(t, X_t))dt + B(t, X_t)dW(t), \\ X_0 &= X^0 \end{cases} \tag{11}$$

and investigate the existence and uniqueness of mild solutions. Depending on the context, the solution process will be denoted as  $(X_t)_{t \in [0, T]}$  or  $(X(t))_{t \in [0, T]}$ . Let  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  be a filtered probability space and let  $T > 0$ . Let  $\mathcal{P}_T$  and  $\mathcal{P}_\infty$  denote the predictable  $\sigma$ -algebras on  $[0, T] \times \Omega$  and  $[0, \infty) \times \Omega$ , respectively. We also assume that  $\mathcal{U}, \mathcal{H}$  are separable Hilbert spaces and  $(W(t))_{t \in [0, T]}$  is a standard cylindrical Wiener process over  $\mathcal{U}$  given in Definition A7. Therefore, using the notation of Appendix C.1, we can let  $Q = I$  and  $\mathcal{U}_0 = Q^{\frac{1}{2}}\mathcal{U} \equiv \mathcal{U}$ . The spaces  $\mathcal{U}, \mathcal{H}, \mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})$  induce corresponding Borel  $\sigma$ -algebras  $\mathcal{B}(\mathcal{U}), \mathcal{B}(\mathcal{H}), \mathcal{B}(\mathcal{L}_{HS}(\mathcal{U}, \mathcal{H}))$ . (Recall that  $\mathcal{L}_{HS}(\cdot, \cdot)$  denotes the space of Hilbert-Schmidt operators.) Lastly,  $X^0$  is a random  $\mathcal{H}$ -valued  $\mathcal{F}_0$ -measurable initial value. See Appendices A and C for a brief review of probabilistic concepts and results used in this section.

**Assumption 1.** Let  $A, F$  and  $B$  in Equation (11) satisfy:

- $F: [0, T] \times \Omega \times \mathcal{H} \rightarrow \mathcal{H}$  is measurable from  $([0, T] \times \Omega \times \mathcal{H}, \mathcal{P}_T \times \mathcal{B}(\mathcal{H}))$  to  $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ ,
- $B: [0, T] \times \Omega \times \mathcal{H} \rightarrow \mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})$  is measurable from  $([0, T] \times \Omega \times \mathcal{H}, \mathcal{P}_T \times \mathcal{B}(\mathcal{H}))$  to  $(\mathcal{L}_{HS}(\mathcal{U}, \mathcal{H}), \mathcal{B}(\mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})))$ ,
- There exists a constant  $C > 0$  such that for all  $u, v \in \mathcal{H}, t \in [0, T], \omega \in \Omega$  we have:

$$\begin{aligned} \|F(t, \omega, u) - F(t, \omega, v)\|_{\mathcal{H}} + \|B(t, \omega, u) - B(t, \omega, v)\|_{\mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})} &\leq C\|u - v\|_{\mathcal{H}}, \\ \|F(t, \omega, u)\|_{\mathcal{H}}^2 + \|B(t, \omega, u)\|_{\mathcal{L}_{HS}(\mathcal{U}, \mathcal{H})}^2 &\leq C^2(1 + \|u\|_{\mathcal{H}}^2), \end{aligned}$$

- $A$  is the generator of a strongly continuous operator semigroup  $(S_t)_{t \geq 0}$  on  $\mathcal{H}$ .

**Definition 11** (Mild Solution). An  $\mathcal{H}$ -valued predictable stochastic process  $(X(t))_{t \in [0, T]}$  with

$$\mathbb{P}\left(\int_0^T \|X(s)\|_{\mathcal{H}}^2 ds < \infty\right) = 1$$

is referred to as a mild solution to Equation (11) if it satisfies the integral equation (well-defined on the strength of Proposition A9):

$$X_t = S_t X^0 + \int_0^t S_{t-s} F(s, X_s) ds + \int_0^t S_{t-s} B(s, X_s) dW(s) \quad \text{for all } t \in [0, T]$$

$\mathbb{P}$ -almost surely.

Note that the predictability of  $(X(t))_{t \in [0, T]}$  implies it possesses a continuous modification, i.e.,  $\mathbb{P}(X \in C^0([0, T], \mathcal{H})) = 1$ , which (in addition to squared-integrability) is essential for the latter integral to be well-defined.

**Theorem 1** ([11], Section 7.1). *Under Assumption 1, there exists a unique (up to indistinguishability)  $\mathcal{H}$ -valued mild solution  $(X(t))_{t \in [0, T]}$  to Equation (11). Moreover, for any  $p \geq 2$ , there exists a positive constant  $C$ , which only depends on  $p$  and  $T$ , such that*

$$\sup_{t \in [0, T]} \mathbb{E}[\|X(t)\|_{\mathcal{H}}^p] \leq C \left(1 + \mathbb{E}[\|X^0\|_{\mathcal{H}}^p]\right).$$

### 3.2. Definition of Abstract HJM Model

Let  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  be a filtered probability space. Further, let  $\mathcal{I}$  either be a compact interval of the form  $[0, t_0]$  for some positive  $t_0$  or the half-line  $[0, \infty)$  and let  $(W(t))_{t \in \mathcal{I}}$  be a standard cylindrical Wiener process over a separable Hilbert space  $\mathcal{U}$ . Let  $\mathcal{S}$  be a state space and  $\mathcal{P}_{\mathcal{I}}$  be a predictable  $\sigma$ -algebra on  $\mathcal{I} \times \Omega$ . We make the following assumption (cf. [6], Chapter 6).

**Assumption 2.** *Suppose:*

- $\mathcal{S}$  is a separable Hilbert space consisting of continuous, real-valued functions  $f: \mathcal{I} \rightarrow \mathbb{R}$ . Additionally, we assume that for every  $x \in \mathcal{I}$  the functionals

$$\delta_x(f) = f(x), \quad \mathcal{J}_x(f) = \int_0^x f(s) ds$$

are continuous on  $\mathcal{S}$ . (The former assumption can also be encountered in statistics, for example, in the context of Reproducing Kernel Hilbert Spaces (RKHS)).

- The left shift semigroup  $(S_t)_{t \geq 0}$  defined via

$$(S_t f)(x) = f(t + x)$$

is a strongly continuous operator semigroup on  $\mathcal{S}$  generated by a (linear) operator  $A$ .

- $F_{\text{HJM}}$  is a measurable mapping from a non-empty subset of  $\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S})$  into  $\mathcal{S}$  and defined by

$$F_{\text{HJM}}(\sigma)(x) = \langle \sigma^* \delta_x, \sigma^* \mathcal{J}_x \rangle_{\mathcal{U}}$$

for all  $\sigma \in \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S})$ .

The crucial property of the space  $\mathcal{S}$  is that its elements  $f \in \mathcal{S}$  are (locally) integrable functions so that  $P(t, T) = e^{\int_0^{T-t} f_t(s) ds}$  is well-defined. Indeed, the local integrability holds true as the linear functional  $\mathcal{J}_x$  is bounded on  $\mathcal{S}$  for each  $x \in \mathcal{I}_t$  since

$$\left| \int_0^x f(s) ds \right| \leq x \sup_{s \in [0, x]} |f(s)| \leq x \sup_{s \in [0, x]} \|\delta_s\|_{\mathcal{S}^*} \|f\|_{\mathcal{S}} < \infty$$

where  $\sup_{s \in [0, x]} \|\delta_s\|_{\mathcal{S}^*}$  is finite on the strength of Banach-Steinhaus theorem.

The financial implication of the first part of Assumption 2 is the identity  $r_t = f_t(0)$  for the short interest rate. The bank account is then defined via  $B(t) = e^{\int_0^t r_s ds}$ . The semigroup



$(S_t)_{t \geq 0}$  given in the second part of Assumption 2 enables us to switch to Musiela’s notation  $f_t(x)$ , where  $f_t(x) = f(t, t + x)$ . The third part of Assumption 2 is intimately related to the no-arbitrage principle and is needed to define the drift term in the HJM equation. Further, it is generally false that  $F_{\text{HJM}}(\sigma)$  is an element of  $\mathcal{S}$  unless  $\sigma$  is an element of a proper subspace of  $\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S})$ .

**Assumption 3.** Assume there exists a subspace  $\mathcal{S}_*$  of  $\mathcal{S}$  such that the binary operator  $\star$  defined by

$$(f \star g)(x) = f(x) \int_0^x g(s) ds,$$

mapping from  $\mathcal{S}_* \times \mathcal{S}_*$  into  $\mathcal{S}$  is bounded, i.e., there exists  $C > 0$  such that the following bound

$$\|f \star g\|_{\mathcal{S}} \leq C \|f\|_{\mathcal{S}} \|g\|_{\mathcal{S}}$$

holds for all  $f, g \in \mathcal{S}_*$ .

The following estimate for  $F_{\text{HJM}}$  is known after ([6], p. 167).

**Proposition 1.** Let the space  $\mathcal{S}$  satisfy Assumption 3. Then the mapping  $F_{\text{HJM}}$  satisfies the estimate

$$\|F_{\text{HJM}}(\sigma_1) - F_{\text{HJM}}(\sigma_2)\|_{\mathcal{S}} \leq C \|\sigma_1 + \sigma_2\|_{\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S})} \|\sigma_1 - \sigma_2\|_{\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S})}$$

for all Hilbert-Schmidt operators  $\sigma_1, \sigma_2 \in \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S}_*)$  with the range contained in  $\mathcal{S}_*$ . In particular, the mapping  $F_{\text{HJM}}$  is measurable from  $\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S}_*)$  to  $\mathcal{S}$ .

Now we can give a rigorous formulation of the abstract HJM model adopted from ([6], Section 6.3). It is rooted in the standard HJM framework [1], which describes a certain class of arbitrage-free finite-dimensional models for the instantaneous forward rate. We provide a more powerful framework, referred to as the abstract HJM model, which incorporates arbitrage-free infinite-dimensional models.

**Definition 12** (The abstract HJM model). The abstract HJM model in a functional space  $\mathcal{S}$  is a pair of functions  $(\lambda, \sigma)$  such that

- $\lambda$  is a measurable function from  $(\mathcal{I} \times \Omega \times \mathcal{S}, \mathcal{P}_{\mathcal{I}} \times \mathcal{B}(\mathcal{S}))$  into  $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ ,
- $\sigma$  is a measurable function from  $(\mathcal{I} \times \Omega \times \mathcal{S}, \mathcal{P}_{\mathcal{I}} \times \mathcal{B}(\mathcal{S}))$  into  $(\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S}_*), \mathcal{B}(\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{S}_*)))$

so that there exists a non-empty set of  $\mathcal{S}$ -valued  $\mathcal{F}_0$ -measurable initial conditions  $\Phi^0$ , which admits a unique time-continuous mild  $\mathcal{S}$ -valued solution  $(\Phi_t)_{t \in \mathcal{I}}$  to the HJM equation:

$$\begin{cases} d\Phi_t &= (A\Phi_t + \alpha(t, \Phi_t))dt + \sigma(t, \Phi_t)dW_t, \\ \alpha(t, \Phi_t) &= F_{\text{HJM}}(\sigma(t, \Phi_t)) + \sigma(t, \Phi_t)\lambda(t, \Phi_t), \\ \Phi_0 &= \Phi^0. \end{cases} \tag{12}$$

The mild formulation of Equation (12) is given via:

$$\begin{cases} \Phi_t &= S_t\Phi_0 + \int_0^t S_{t-s}\alpha(s, \Phi_s)ds + \int_0^t S_{t-s}\sigma(s, \Phi_s)dW_s, \\ \alpha(t, \Phi_t) &= F_{\text{HJM}}(\sigma(t, \Phi_t)) + \sigma(t, \Phi_t)\lambda(t, \Phi_t), \\ \Phi_0 &= \Phi^0. \end{cases} \tag{13}$$

Imposing typical Lipschitzianity and boundedness assumptions, as a direct corollary of Theorem 1, we obtain (cf. [6], p. 169):

**Proposition 2** (Sufficient existence condition). *Let  $\mathcal{S}$  satisfy Assumption 3 with a closed subspace  $\mathcal{S}_* \subset \mathcal{S}$  such that  $\|f \star g\|_{\mathcal{S}} \leq C\|f\|_{\mathcal{S}}\|g\|_{\mathcal{S}}$  for  $f, g \in \mathcal{S}_*$ . Let the operator  $\sigma$  map into  $\mathcal{S}_*$  for every  $(t, \omega, f) \in \mathcal{I} \times \Omega \times \mathcal{S}$ . Assume  $\sigma$  is bounded and the Lipschitzian*

$$\begin{aligned} \|\sigma(t, f) - \sigma(t, g)\|_{\mathcal{L}_{HS}(\mathcal{U}, \mathcal{S})} &\leq C\|f - g\|_{\mathcal{S}}, \\ \|\sigma(t, f)\lambda(t, f) - \sigma(t, g)\lambda(t, g)\|_{\mathcal{S}} &\leq C\|f - g\|_{\mathcal{S}} \end{aligned}$$

for some constant  $C > 0$  and all  $t \in \mathcal{I}, f, g \in \mathcal{S}_*$ , then the pair  $(\lambda, \sigma)$  is an abstract HJM model on  $\mathcal{S}$ . Furthermore, for any initial  $\mathcal{S}$ -valued forward curve  $\Phi^0$ , there exists a unique time-continuous mild solution  $(\Phi_t)_{t \in \mathcal{I}}$  to Equation (12) such that

$$\mathbb{E} \left[ \sup_{t \in \mathcal{I}} \|\Phi_t\|_{\mathcal{S}}^p \right] < \infty \quad \text{for all finite } p \geq 0.$$

#### 4. Absence of Arbitrage

A fundamental economic assumption behind any adequate financial model is the absence of arbitrage opportunities in the associated financial market (cf. [24], pp. 637–654). The absence of arbitrage means impossibility to invest nothing today and receive a positive amount in the future. We restrict ourselves to continuous-time economies (cf. [25], pp. 381–408).

Consider a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{Q}_0)$  with a right-continuous filtration  $\{\mathcal{F}_t\}_{t \in [0, T]}$ . Assume  $(K + 1)$  non-dividend paying securities are traded from time 0 until time  $T$  continuously with their modeled by an adapted  $(K + 1)$ -dimensional semimartingale  $M = \{M_t \mid 0 \leq t \leq T\}$  having non-negative components  $M^0, M^1, \dots, M^K$ . The asset indexed by 0 is a bank account. We refer to Appendix B for a brief review of the martingale theory.

**Definition 13** (Trading strategy). *A trading strategy is a  $((K + 1)$ -dimensional) process  $\phi = \{\phi_t \mid 0 \leq t \leq T\}$  with locally bounded, predictable components  $\phi^0, \phi^1, \dots, \phi^K$ . The value process associated with a strategy  $\phi$  is defined by*

$$V_t(\phi) = \phi_t \cdot M_t = \sum_{k=0}^K \phi_t^k M_t^k \quad \text{for all } t \in [0, T],$$

and the gains process associated with a strategy  $\phi$  are given by

$$G_t(\phi) = \int_0^t \phi_s \cdot dM_s = \sum_{k=0}^K \int_0^t \phi_s^k dM_s^k \quad \text{for all } t \in [0, T].$$

The quantity  $\phi_t^k$  corresponds to the number of units of  $k$ -th security held by an investor at time  $t$ . Predictability of  $\phi^k$  suggests that  $\phi_t^k$  is known directly before  $t$ .

**Definition 14** (Self-financing trading strategy). *We refer to a trading strategy  $\phi$  as self-financing if  $V_T(\phi) \geq 0$  and  $V_t(\phi) = V_0(\phi) + G_t(\phi)$  for  $t \in [0, T)$ .*

**Definition 15** (Equivalent martingale measure). *A probability measure  $\mathbb{Q}$  on  $(\Omega, \mathcal{F})$  is referred to as equivalent martingale measure if it satisfies:*

- $\mathbb{Q}$  and  $\mathbb{Q}_0$  are equivalent martingale measures, i.e.,  $\mathbb{Q}(A) = 0$  if and only if  $\mathbb{Q}_0(A) = 0$  for all  $A \in \mathcal{F}$ ,
- The Radon-Nikodym derivative  $\frac{d\mathbb{Q}}{d\mathbb{Q}_0}$  belongs to  $L^2(\Omega, \mathcal{F}, \mathbb{Q}_0)$ ,
- The “discounted asset price” process  $D(0, \cdot)M$  is an  $(\{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{Q})$ -martingale, i.e.,  $\mathbb{E}_{\mathbb{Q}}[D(0, t)M_t^k | \mathcal{F}_s] = D(0, s)M_s^k$  for all  $k \in \{1, 2, \dots, K\}$  and all  $0 \leq s \leq t \leq T$ .

**Definition 16** (Locally equivalent martingale measure). *We say that  $\mathbb{Q}$  is a locally equivalent martingale measure if for all  $t > 0$  the restriction of  $\mathbb{Q}$  to  $\mathcal{F}_t$  is equivalent to the restriction of  $\mathbb{P}$  to  $\mathcal{F}_t$ .*

An arbitrage opportunity is defined by a self-financing strategy  $\phi$  such that  $V_0(\phi) = 0$  but  $\mathbb{Q}_0(V_T(\phi) > 0) > 0$ . It is known (cf. ([3], pp. 215–260) and ([26], pp. 17–19)) that the existence of an equivalent martingale measure implies the absence of arbitrage opportunities:

**Theorem 2** (Fundamental theorem of asset pricing ([6], Theorem 2.1)). *There are no arbitrage strategies if there exists a probability measure  $\mathbb{Q}$ , equivalent to  $\mathbb{P}$ , such that the discounted asset prices process  $(P_D(t))_{t \geq 0}$  is a local martingale under  $\mathbb{Q}$ .*

Chapter 3.1.1 of [6] provides sufficient conditions for the abstract HJM model from Section 3 to be free of arbitrage. These results are summarized below. Consider an abstract HJM model  $(\sigma, \lambda)$  with an initial condition  $\Phi^0$  and denote by  $(\Phi_t)_{t \in [0, T]}$  the unique mild solution to the HJM equation. For brevity, denote  $\lambda_t := \lambda(t, \omega, f_t)$  and  $\sigma_t := \sigma(t, \omega, f_t)$ .

**Theorem 3** (Conditions of no arbitrage for the abstract HJM model). *Provided*

$$\mathbb{E} \left[ \exp \left( -\frac{1}{2} \int_0^t \|\lambda_s\|_{\mathcal{U}}^2 ds + \int_0^t \lambda_s dW(s) \right) \right] = 1 \quad \text{and}$$

$$\int_0^t \left( \mathbb{E} \left[ \int_0^t \|\sigma_s^* \delta_{s-\tau}\|_{\mathcal{U}}^2 d\tau \right] \right)^{\frac{1}{2}} ds < \infty$$

for all  $t \in [0, T]$ , then the market given by the abstract HJM model  $(\sigma, \lambda)$  admits no arbitrage.

In its current form, the setup of Theorem 3 appears to be too general for practical use. Indeed, it would suffice to show the Novikov condition for  $\mathbb{Q}$

$$\mathbb{E}_{\mathbb{Q}} \left[ \exp \left( \frac{1}{2} \int_0^T \|\sigma_s^* \mathcal{J}_{T-s}\|_{\mathcal{U}}^2 ds \right) \right] < \infty$$

is satisfied in order to prove that  $(P_D(t))_{t \geq 0}$  is a martingale (cf. [6], Theorem 6.1) since every martingale is a local martingale.

#### 4.1. Long Rates

Modeling the forward rate as a function on a bounded interval  $\mathcal{I}$  is different from modeling it on the real half-line. Namely, we encounter an unexpected phenomenon that the long rate defined by the limit

$$l_t = \lim_{x \rightarrow \infty} f_t(x) =: f_t(\infty)$$

never decreases. This phenomenon for long forward and zero-coupon rates was previously studied in [27]. The intrinsic nature of this phenomenon is rooted in the fact that the (appropriately defined) discounted bond prices  $(\tilde{P}(t, T))_{t \in [0, T]}$  are local martingales for all  $T > 0$  (cf. [6], p. 172). The following result is known after ([6], Chapter 6.3.2).

**Assumption 4.** *Every  $f \in \mathcal{S}$  is a function  $f: \mathbb{R}^+ := [0, \infty) \rightarrow \mathbb{R}$  such that the limit  $\lim_{x \rightarrow \infty} f(x)$  exists. The functional  $\delta_{\infty}: f \mapsto \lim_{x \rightarrow \infty} f(x)$  is an element of  $\mathcal{S}^*$ .*

**Theorem 4.** *If  $0 \leq s \leq t$ , then  $l_s \leq l_t$   $\mathbb{P}$ -almost surely.*

#### 4.2. Example of State Space

An example of a state space  $\mathcal{S}$  was first proposed in ([4], Chapter 3) and further investigated in ([6], Chapter 3.3). Obviously, this space should fulfill the first and the

second parts of Assumption 2 under which the abstract HJM model is well-defined, i.e., it should consist of real-valued continuous functions, for which the functionals  $\delta_x, \mathcal{J}_x$  should be continuous on  $\mathcal{S}$  and the strongly continuous semigroup  $(S_t)_{t \geq 0}$  should exist on  $\mathcal{S}$ . Parenthetically, we point out that, though the following presentation assumes  $\mathcal{I} = [0, \infty)$ , our empirical example in Section 5 will rather utilize a bounded interval  $[0, x_{\max}]$ .

**Definition 17** (Example of state space). *Let  $w: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be a monotonically increasing function such that*

$$\int_0^\infty \frac{1}{w(s)} ds < \infty. \tag{14}$$

denote by  $\mathcal{H}_w$  the space

$$\mathcal{H}_w = \{f \in L^1_{\text{loc}}(\mathbb{R}^+) | f' \in L^1_{\text{loc}}(\mathbb{R}^+) \text{ and } \|f\|_{\mathcal{H}_w} < \infty\},$$

where  $f'$  is the distributional derivative of  $f$  and

$$\|f\|_{\mathcal{H}_w}^2 := \langle f, f \rangle_{\mathcal{H}_w} \text{ with } \langle f, g \rangle_{\mathcal{H}_w} = f(0)g(0) + \int_0^\infty f'(s)g'(s)w(s)ds.$$

If  $f \in L^1_{\text{loc}}(\mathbb{R}^+)$  has a weak derivative  $f' \in L^1_{\text{loc}}(\mathbb{R}^+)$ , e.g., some function  $\phi$  such that

$$\int_0^\infty f(s)\phi'(s)ds = - \int_0^\infty f'(s)\phi(s)ds \quad \text{for all } \phi \in C^1_c(\mathbb{R}^+),$$

then there exists an absolutely continuous representative  $f$ , still denoted by  $f$ , such that

$$f(t_2) - f(t_1) = \int_{t_1}^{t_2} f'(s)ds \quad \text{for all } t_1, t_2 \in \mathbb{R}^+$$

(cf. [28], Section VIII.2). In particular, this implies that the trace  $\delta_x: f \mapsto f(x)$  is well-defined.

**Proposition 3** ([6], Chapter 3.3). *If  $w$  satisfies inequality (14), then the inner product space  $\mathcal{H}_w$  is a separable Hilbert space. The evaluation functional  $\delta_x$  and the definite integration functional  $\mathcal{J}_x$  defined by*

$$\delta_x(f) = f(x), \quad \mathcal{J}_x(f) = \int_0^x f(s)ds$$

respectively, are continuous on  $\mathcal{H}_w$  for all positive  $x$ . Further, the operator semigroup  $(S_t f)(x) = f(t + x)$  on  $\mathcal{H}_w$  is strongly continuous.

### 5. Example: Modeling and Predicting US Treasuries

Over the past two decades, the problem of parameter estimation, also referred to as model calibration or estimation, for the classical *finite-dimensional* HJM model [1] (viz. Equation (2)) has attracted major attention from various researchers in econometrics, quantitative finance, stochastic analysis, numerical analysis, etc. In the finite-dimensional context, estimation for the HJM model is typically performed using the method of least squares, the method of moments, maximum likelihood estimation and penalized versions thereof, stochastic filtration, kriging, parametric and nonparametric Bayesian inference, etc. See, e.g., [29–34] for relevant studies in both parametric and non-parametric settings. We also refer to “classic” papers [8–10] as well as recent contributions [35,36] and references therein for closely related results for the Vasicek model [37].

As for the abstract HJM model (viz. Equation (4)) in the *infinite-dimensional* context, with the exception of the numerical paper [38], no results are known in the literature. With the primary focus on studying the convergence properties of the proposed numerical splitting schemes, ([38], Section 2.3.1) presents a brief empirical study based on a dataset from

the dissertation [39]. As opposed to the finite-dimensional situation, infinite-dimensional parameter estimation or calibration constitutes a (possibly ill-posed) statistical inverse problem. To the best of our knowledge, no rigorous investigations of HJM model calibration in the context of statistical inverse problems are currently available in the literature. A promising angle of attack is though provided in [40]. Also, no calibration and/or forecasting studies on the infinite-dimensional HJM model in connection with real-world data are also available up to date. In this section, we investigate a non-parametric calibration problem for the linear version of Equation (4). To this end, we discretize the equation with an implicit/explicit “splitting scheme” and propose a PCA- and least-squares based parameter estimation/dimension reduction procedure to fit the discretized model. We illustrate our approach by applying our proposed scheme to a real-world dataset presented below.

Daily US Treasury yield curve rates are available from the US Department of the Treasury [7]. For any given business day (i.e., all days excluding holidays and weekends), the yield curves relate the closing market bid yields on actively traded Treasury securities in the over-the-counter market to their time to maturity. Further details on the methodology used to calculate the yield rates are provided in [7]. The time-to-maturity grid includes 1 mo, 2 mo (starting 16 October 2018), 3 mo(nths), 6 mo, 1 y(ea)r, 2 yr, 3 yr, 5 yr, 7 yr, 10 yr, 20 yr and 30 yr. Figure 1 displays historical yield rates for the year 2018. Since the 2-month rates were missing for the dates between 1 January 2018 and 15 October 2018, linear interpolation was used to estimate these values.

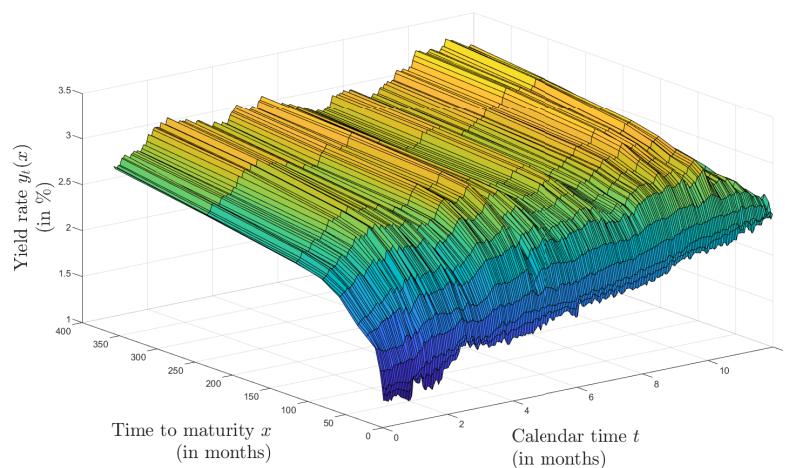


Figure 1. Historical yield rate curves for 2018.

Employing Musiela’s notation [5], let  $y_t(x)$  denote the yield rate of a US treasury traded at time  $t$  with the time to maturity  $x$ . According to ([4], Definition 3.13),  $y_t(x)$  can be expressed with the aid of the instantaneous forward rate  $f_t(x)$  as

$$y_t(x) = \frac{1}{x} \int_0^x f_t(x) dx \equiv \frac{1}{x} \mathcal{I}_x(f_t). \tag{15}$$

In particular, invoking the mean value theorem, the former equation suggests

$$y_t(0) = f_t(0) \equiv r_t, \tag{16}$$

where  $r_t$  stands for the short rate. Multiplying both sides of Equation (15) by  $x$  and differentiating in  $x$ , we can solve for  $f_t(x)$ , i.e.,

$$f_t(x) = \partial_x(x y_t(x)) = y_t(x) + x \partial_x y_t(x). \tag{17}$$

Unfortunately, no “natural” boundary condition is readily available for  $f_t(x)$ . In contrast, since  $y_t(x)$  can be assumed bounded in  $x = 0$ , a homogeneous Dirichlet boundary

condition emerges for  $xy_t(x)$ . Therefore, in spirit of ([4], Chapter 3.3), we will rather work with the (observable) quantity

$$Y(t, x) = xy_t(x), \tag{18}$$

referred to as the “integrated forward rate” since

$$Y(t, x) = \mathcal{I}_x(f_t) \equiv \int_0^x f_t(\eta) d\eta.$$

This quantity,  $Y(t, x)$ , satisfies the homogeneous Dirichlet boundary condition  $Y(t, 0) = 0$  at  $x = 0$ .

Since the shortest time-to-maturity available in our dataset is  $x_{\min} = 1$  month,  $y_t(0)$  will be approximated/extrapolated as  $y_t(x_{\min})$ . Further, as Figure 1 suggests, the  $x$ -grid becomes very coarse for  $x > 3$  years (= 36 month). In particular, this means that the infinitesimal effects of the operator  $A$  in the abstract HJM model in Equation (13) will become negligible as the semigroup  $S_t(\cdot)$  will practically “collapse” to the identity operator, essentially reducing the model (for large  $x$ ) to the classic HJM model. Therefore, the maturities were truncated at 36 months. Lastly, the observed yield rates are interpolated onto  $[0, 36]$  month using cubic splines. Figure 2 displays the historical  $Y$ -curves for the year 2018 over the truncated 36-month  $x$ -horizon.

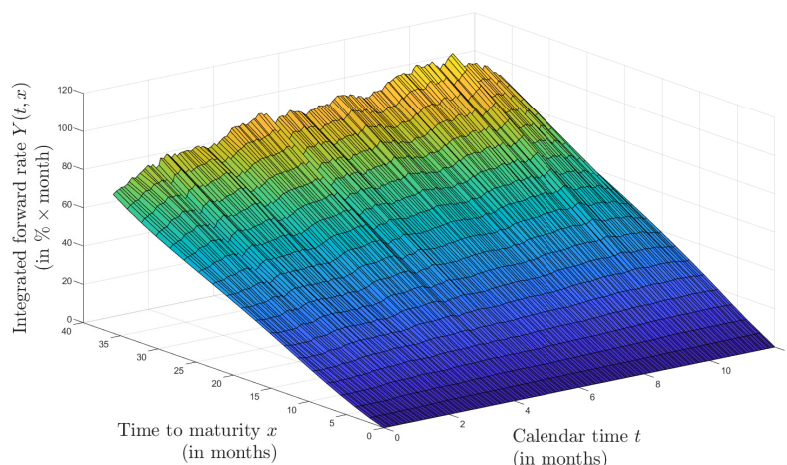


Figure 2. Historical values for integrated forward rates  $Y(t, x)$ .

### 5.1. The Model

Consider the space  $\mathcal{H}_w$  defined in Section 4.2. The “spatio-”temporal evolution of the forward rates will be assumed to follow the linear autonomous HJM model from Equation (4), i.e.,

$$\begin{cases} df_t(x) &= \left( \frac{d}{dx} f_t(x) + \tilde{\alpha}_t(x) \right) dt + \sum_{k=1}^{\infty} \sigma^k(x) \cdot d\tilde{W}_k(t), \\ f_0(x) &= f^0(x), \end{cases} \tag{19}$$

with some initial datum  $f^0$  and  $\tilde{\alpha}_t(x)$  determined via the “no arbitrage” condition

$$\tilde{\alpha}_t(x) = \sum_{k=1}^{\infty} \sigma^k(x) \left( \int_0^x \sigma^k(\eta) d\eta \right),$$

where the evolution in Equation (19) is taking place under the risk-neutral measure  $\tilde{\mathbb{P}}$ . To assure the model is identifiable, similar to [8,10],  $\sigma^k(x)$ ’s will be assumed orthogonal with respect to an appropriate inner product to be discussed later.

Following the results of Section 4, we can write

$$\tilde{W}_t^k = W_t^k + \int_0^t \lambda_s^k ds$$

for a standard cylindrical Wiener process  $(W_t)_{t \geq 0}$  with respect to the physical measure  $\mathbb{P}$ , where  $(\lambda_t)_{t \geq 0}$  must satisfy the conditions of Theorem 3. See also ([10], Section 2). In the spirit of [8], assuming an “affine specification” for the market price of risk

$$\lambda_t^k \equiv \left( \int_0^x \sigma^k(\eta) d\eta \right) \lambda^k \quad \text{with some constant } \lambda^k\text{'s}$$

the former conditions are trivially satisfied. Thus, Equation (19) can be equivalently expressed under the physical measure  $\mathbb{P}$ :

$$\begin{cases} df_t(x) &= \left( \frac{d}{dx} f_t(x) + \alpha(x) \right) dt + \sum_{k=1}^{\infty} \sigma^k(x) \cdot dW_k(t), \\ f_0(x) &= f^0(x), \end{cases} \tag{20}$$

with

$$\alpha(x) = \sum_{k=1}^{\infty} \sigma^k(x) \left( \int_0^x \sigma^k(x) ds + \lambda_t^k \right) = \sum_{k=1}^{\infty} (1 + \lambda^k) \sigma^k(x) \left( \int_0^x \sigma^k(x) ds \right).$$

Since the parameter vectors  $(\sigma^k(x))_{k \in \mathbb{N}}$  and  $(\lambda^k)_{k \in \mathbb{N}}$  are infinite-dimensional, the model (20), though being linear and autonomous, is genuinely non-parametric from the statistical point of view.

Arguing similar to ([4], p. 42), the stochastic process  $(Y(t, x))_{t \geq 0}$  is expected to solve the SPDE:

$$\begin{cases} dY(t, x) &= \left( - \left( \frac{\partial}{\partial x} Y(t, x) - r_t \right) + \frac{1}{2} \sum_{k=1}^{\infty} (1 + \lambda_k) (\mathcal{I}_x \sigma^k)^2(x) \right) dt \\ &\quad + \sum_{k=1}^{\infty} (\mathcal{I}_x \sigma^k)(x) \cdot dW_k(t) \quad \text{for } t > 0, x \in (0, x_{\max}), \\ Y(t, 0) &= 0 \quad \text{for } t > 0, \\ Y(0, x) &= xy^0(x) \quad \text{for } x \in [0, x_{\max}], \end{cases} \tag{21}$$

where  $y^0$  is the initial value for the yield curve  $y_t(\cdot)$  at  $t = 0$  and

$$r_t = f_t(0) \equiv y_t(0)$$

denotes the short rate. We would like to draw the reader’s attention to the minus sign in front of the first term on the right-hand side of the first equation in (21). Since the boundary condition is now prescribed at  $x = 0$ , the information propagates from left to right, i.e., the abstract operator is now  $A = -\frac{\partial}{\partial x}$  subject to homogeneous Dirichlet boundary conditions at  $x = 0$ .

On the strength of Hille & Yosida theorem, the operator  $A$  with the maximal domain

$$D(A) = \{Y \in \mathcal{H} \mid AY \in \mathcal{H}, Y(0) = 0\}$$

(note that the point evaluation is well-defined in view of Sobolev embedding theorem) generates a strongly continuous semigroup  $(S_t)_{t \geq 0}$  of contractions on

$$\mathcal{H} = L^2(0, x_{\max}; \mathbb{R})$$

endowed with the standard scalar product. The semigroup  $(S_t)_{t \geq 0}$  is now the right-shift semigroup given as

$$(S_t Y)(x) = \begin{cases} 0, & 0 \leq x \leq t \leq x_{\max}, \\ Y(x - t), & 0 \leq t \leq x \leq x_{\max}. \end{cases}$$

Assuming  $r_t$  is a scalar Itô process independent of the cylindrical Wiener process  $(W(t))_{t \geq 0}$ , the concept of a mild solution can be developed similar to Section 3.1.

To make Equation (21) useful for prediction purposes, we need to specify the evolution of  $r_t$ . To this end, we utilize the classical Vasiček’s short rate model [37]:

$$\begin{cases} dr_t &= \alpha(\tilde{r}^* - r_t)dt + \beta d\tilde{W}_t^0 \quad \text{for } t > 0, \\ r_0 &= r^0, \end{cases} \tag{22}$$

where  $\alpha, \beta, \tilde{r}^* > 0$  are parameters,  $r_0$  is a given random variable independent of  $(\tilde{W}^0(t))_{t \geq 0}$  and  $(\tilde{W}^0(t))_{t \geq 0}$  is, in turn, independent of  $y^0$  and  $(\tilde{W}(t))_{t \geq 0}$ . Since the evolution in Equation (22) is taking place under the risk-neutral measure  $\mathbb{P}^0$ , we perform a change of measures using Girsanov’s theorem assuming

$$\tilde{W}_t^0 = W_t^0 + \int_0^t \lambda^0 ds = W_t^0 + \lambda^0 t.$$

Thus, under the physical measure  $\mathbb{P}^0$ , the evolution in Equation (22) assumes the form

$$\begin{cases} dr_t &= \alpha(r^* - r_t)dt + \beta dW_t^0 \quad \text{for } t > 0, \\ r_0 &= r^0 \end{cases} \tag{23}$$

with  $r^* = \tilde{r}^* + \frac{\lambda^0}{\alpha}$  and a standard Wiener process  $(W_t^0)_{t \geq 0}$ . In the following, we will work with Equations (21) and (23) under the physical measures  $\mathbb{P}$  and  $\mathbb{P}^0$ . The key feature of Equations (21) and (23) is that they are autonomous, which is crucial requirement if a model is to be used in forecasting.

Since our dataset is discrete, it is natural to discretize Equations (21) and (23) as well. To this end, we employ (semi-)implicit and explicit Euler & Maruyama schemes for time integration and a two-point finite-difference difference quotient approximation for  $A$ . Selecting a time grid  $\{t_j | j = 0, \dots, m - 1\}$  and a “space” grid  $\{x_i | i = 0, \dots, n - 1\}$  with  $t_j = j(\Delta t)$ ,  $x_i = i(\Delta x)$ ,  $\Delta t = \frac{T_{\max}}{m-1}$ ,  $\Delta x = \frac{x_{\max}}{n-1}$ , we obtain

$$\begin{cases} \frac{Y(t_{j+1}, x_i) - Y(t_j, x_i)}{\Delta t} &= \left( - \left( \frac{Y(t_{j+1}, x_{i+1}) - Y(t_{j+1}, x_i)}{\Delta x} - r_{t_j} \right) + \frac{1}{2} \sum_{k=1}^{\infty} (1 + \lambda^k) (\mathcal{I}_{x_i} \sigma^k)^2(x_i) \right) \\ &+ \sum_{k=1}^{\infty} (\mathcal{I}_{x_i} \sigma^k)(x_i) \frac{W_k(t_{j+1}) - W_k(t_j)}{\Delta t} \\ &\quad \text{for } j = 0, \dots, m - 2, \quad i = 0, \dots, n - 2, \\ Y(t_j, 0) &= 0 \quad \text{for } j = 0, \dots, m - 1, \\ Y(t_0, x_i) &= x_i y^0(x_i) \quad \text{for } i = 0, \dots, n - 1 \end{cases} \tag{24}$$

and

$$\begin{cases} \frac{r_{t_{j+1}} - r_{t_j}}{\Delta t} &= \alpha(r^* - r_{t_j}) + \beta \frac{W_{t_{j+1}}^0 - W_{t_j}^0}{\Delta t} \quad \text{for } j = 0, \dots, m - 2, \\ r_{t_0} &= r^0. \end{cases} \tag{25}$$

Using standard results from numerical analysis of S(P)DEs, both numerical schemes can be proved to be unconditionally feasible and stable (i.e., for any  $\Delta t, \Delta x > 0$ ). The consistency also holds true, but is irrelevant for our purposes as  $\Delta t, \Delta x$  remain fixed. A numerical implementation in Matlab® is given in the Supplementary Materials.

The historical yield rate dataset [7] used in this paper contains  $m = 249$  time periods for the year 2018. Restricting the times to maturity to 0 to 12 months, the equidistant  $x$ -grid contains  $n = 769$  nodes ranging from 0 to 32 months. Thus,  $\Delta t = \frac{12}{m-1} = 0.0484$  and  $\Delta x = \frac{32}{n-1} = 0.0417$ .

### 5.2. Parameter Estimation

Before the model in Equations (24) and (25) can be used for prediction purposes, the unknown functional parameter vector  $\sigma \equiv (\sigma_k(\cdot))_{j \in \mathbb{N}}$  and the scalar parameters  $\alpha, \beta$  need to be estimated. Since  $\sigma$  is infinite-dimensional, the underlying estimation problem is



a genuinely non-parametric and will be regularized through a type of PCA (Principal Component Analysis) truncation ([6], Chapter 1.7.1). In the PDE optimization literature, this approach is commonly known as POD (Proper Orthogonal Decomposition) and used as a dimension reduction technique [41].

We begin with Equation (25). The maximum likelihood (ML) estimators for  $\alpha, r^*$  and  $\beta$  are given as

$$\hat{\alpha} = -\frac{1}{\Delta t} \log \left( \frac{m \sum_{j=1}^m r_{t_j} r_{t_{j-1}} - (\sum_{j=1}^m r_{t_j})(\sum_{j=1}^m r_{t_{j-1}})}{m \sum_{j=1}^m R_{t_{j-1}}^2 - (\sum_{j=1}^m r_{t_{j-1}})^2} \right),$$

$$\hat{r}^* = \frac{1}{m(1 - \exp(-\hat{\alpha}\Delta t))} \left( \sum_{j=1}^m r_{t_j} - \exp(\hat{\alpha}\Delta t) \sum_{j=1}^m r_{t_{j-1}} \right),$$

$$\hat{\beta}^2 = \frac{2\hat{\alpha}}{m(1 - \exp(-2\hat{\alpha}\Delta t))} \sum_{j=1}^m \left( r_{t_j} - r_{t_{j-1}} \exp(-\hat{\alpha}\Delta t) - \hat{r}(1 - \exp(-\hat{\alpha}\Delta t)) \right)^2$$

(cf. [42–44]). The ML estimates for the 2018 historic data are displayed in Table 1 below.

**Table 1.** ML estimates for the Vasicek model based on 2018 historic short rate data.

$\hat{\alpha}$	$\hat{r}^*$	$\hat{\beta}$
0.079473	-0.63952	0.15449

Since no maximum likelihood estimator is known for  $\sigma$  in Equation (24), we will develop a custom least-squares-like method instead. To this end, we rewrite Equation (24) as

$$\begin{cases} V_{t_j} = \frac{1}{2} \sum_{k=1}^{\infty} (1 + \lambda^k) (\mathcal{I}_{x_i} \sigma^k)^2(x_i) + \sum_{k=1}^{\infty} (\mathcal{I}_x \sigma^k)(x_i) \frac{W_k(t_{j+1}) - W_k(t_j)}{\Delta t} \\ \text{for } j = 0, \dots, m-2, \quad i = 0, \dots, n-2, \\ Y(t_j, 0) = 0 \quad \text{for } j = 0, \dots, m-1, \\ Y(t_0, x_i) = x_i y^0(x_i) \quad \text{for } i = 0, \dots, n-1 \end{cases} \tag{26}$$

with

$$V_{t_j} = \frac{Y(t_{j+1}, x_i) - Y(t_j, x_i)}{\Delta t} + \left( \frac{Y(t_{j+1}, x_{i+1}) - Y(t_{j+1}, x_i)}{\Delta x} - r_{t_j} \right).$$

According to Equation (26),  $V_{t_j}$ 's are i.i.d. Gaussian elements of  $\mathcal{H}$ . In view of the fact that we interested in predicting the yields  $y_t(x)$  given as  $y_t(x) = \frac{1}{x} Y(t, x)$ , for regularization purposes, we will rather employ the stronger topology of  $D(A)$ .

As mentioned before, to render  $\sigma^k(x)$ 's identifiable, we assume the latter are orthogonal in  $\mathcal{H}$ . This implies  $\mathcal{I}_x \sigma^k$  are orthogonal in  $D(A)$  since

$$\langle \sigma^k, \sigma^l \rangle_{\mathcal{H}} = \langle A \mathcal{I}_x \sigma^k, A \mathcal{I}_x \sigma^l \rangle_{\mathcal{H}} = \langle \mathcal{I}_x \sigma^k, \mathcal{I}_x \sigma^l \rangle_{D(A)}. \tag{27}$$

Also note that  $(\mathcal{I}_x \sigma)^2 \in \mathcal{H}$  since  $\mathcal{I}_x \sigma^k \in D(A) \hookrightarrow L^\infty((0, x_{\max}))$  for  $\sigma^k \in \mathcal{H}$  on the strength of Sobolev's imbedding theorem and Hölder's inequality.

We proceed by forming the sample  $\{A^{\Delta x} V(t_0), \dots, A^{\Delta x} V(t_{m-2})\}$  with the discretized negative derivative operator

$$A^{\Delta x} = -\frac{1}{\Delta x} \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix}$$

and compute the covariance operator in the scalar product of

$$\mathcal{H}^{\Delta x} = \{Y = (Y_{t_1}, \dots, Y_{t_{m-1}})\} \text{ with } \|Y\|_{\mathcal{H}^{\Delta x}}^2 = (\Delta x) \sum_{i=1}^{m-1} Y^2(x_i).$$

Let  $C^{\Delta x}$  denote the resulting (bounded) covariance operator/matrix given as

$$C^{\Delta x} = \frac{1}{m-2} \sum_{j=0}^{m-1} (A^{\Delta x} V(t_j) - m^{\Delta x}) \otimes (A^{\Delta x} V(t_j) - m^{\Delta x})$$

with the sample mean element/vector

$$m^{\Delta x} = \frac{1}{m-1} \sum_{j=0}^{m-2} A^{\Delta x} V(t_m), \tag{28}$$

Performing spectral decomposition

$$C^{\Delta x} = \sum_{k=1}^p v_k (\phi_k \otimes \phi_k)$$

for some ‘‘principal variances’’  $v_1 \geq v_2 \geq \dots \geq v_p \geq 0$  and unit  $\phi_k \in \mathcal{H}^{\Delta x}$  (cf. [6], Chapter 1.7.1), we choose the first  $p' \leq p$  ‘‘principal modes’’ responsible for, say, 99% of variation, i.e., we choose the smallest  $p'$  such that  $\frac{\sum_{k=1}^{p'} v_k}{\sum_{k=1}^p v_k} \geq 0.99$ .

This truncation step, in addition to the  $D(A)$ -topology employed, renders the procedure numerically and statistically well-posed. Figure 3 displays the six principal modes responsible for 99.32% of variation in  $AV(t_j)$ ’s for the 2018 integrated forward rate data. Summarizing,  $\mathcal{I}_x \sigma^k$ ’s and, thus,  $\sigma^k$  can now be estimated via

$$\begin{aligned} \widehat{\mathcal{I}_x \sigma}^k &= \begin{cases} (\Delta t) v_k^{1/2} \mathcal{I}^{\Delta x} \phi_k, & 1 \leq k \leq p', \\ 0, & k > p' \end{cases} \quad \text{and} \\ \widehat{\sigma}^k &= \begin{cases} (\Delta t) v_k^{1/2} \phi_k, & 1 \leq k \leq p', \\ 0, & k > p', \end{cases} \end{aligned} \tag{29}$$

where the discrete negative antiderivative operator

$$-\mathcal{I}^{\Delta x} : Y \mapsto \left( -(\Delta x) \sum_{l=0}^i Y(t_l) \right)_{i=0, \dots, n-1} \tag{30}$$

comprises the (isometric isomorphic) inverse of  $A^{\Delta x}$ . Since the covariance operator  $C^{\Delta x}$  is (in general) non-spherical, we expect that our method is consistent, but does not necessarily need to exhibit optimal efficiency. The latter can be improved with a type of ‘‘backfitting’’ procedure, which is beyond the scope of this article.

Once  $\sigma^k(x)$ ’s are estimated, we can estimate  $\lambda_k$ ’s using the least squares method. Computing the ‘‘design matrix’’

$$X^{\Delta x} = (A^{\Delta x} \widehat{\mathcal{I}_x \sigma}^1, \dots, A^{\Delta x} \widehat{\mathcal{I}_x \sigma}^{p'})$$

and recalling the sample mean  $m^{\Delta x}$  from Equation (28), the Gauss & Markov theorem yields an estimate for the ‘‘truncated’’  $\lambda$  vector:

$$\hat{\lambda} = 2((X^{\Delta x})' X^{\Delta x})^{-1/2} (X^{\Delta x})' m^{\Delta x} - 1.$$

(Note that we first estimated  $\frac{1}{2}(\lambda^k - 1)$  and then solved for  $\lambda^k$ .) Applied to the 2018 data, the estimates for the  $p' = 6$  “principal”  $\lambda^k$ 's in given in Table 2.

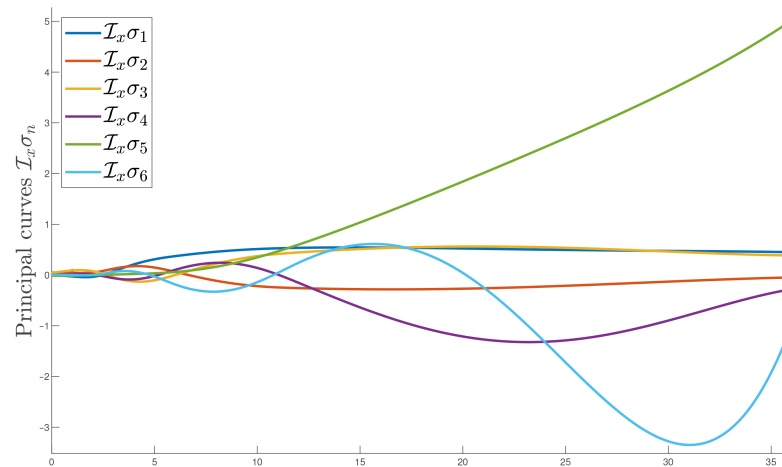


Figure 3. Six principal modes responsible for over 99% of variation.

Table 2. Least-squares estimates for the first  $p' = 6$  parameters  $\lambda^k$ .

$\hat{\lambda}^1$	$\hat{\lambda}^2$	$\hat{\lambda}^3$	$\hat{\lambda}^4$	$\hat{\lambda}^5$	$\hat{\lambda}^6$
-0.71078	6.88	0.48587	-0.74969	-0.73313	-0.84993

### 5.3. Prediction

Now, that the estimates  $\hat{\alpha}, \hat{r}^*, \hat{\beta}$  and  $\hat{\sigma}^k, \hat{\lambda}^k$  are at hand, Equations (24) and (25) can be used to predict the dynamics of  $r_t$  and  $Y(t, x)$  and, thus, that of  $y_t(x) = \frac{Y(t,x)}{x}$ . 1 January 2019 was used as  $t = 0$  with the yield curve from 31 December 2018 (multiplied by  $x$ ) playing the role of the initial value. We generated 10,000 sample paths for the standard Wiener increments  $(W^k(t_{j+1}) - W^k(t_j))_{0 \leq j \leq m' - 2, 1 \leq k \leq 6}$  and  $(W^0(t_{j+1}) - W^0(t_j))_{0 \leq j \leq m' - 2}$ . Here, we chose  $m' = 22$  which corresponds to predicting the integrated forward rates over a one-month time horizon. Further, 10,000 sample paths of the discretized solution processes  $r^{\Delta t, \Delta x, l}, Y^{\Delta t, \Delta x, l}, l = 1, \dots, 10,000$  were computed by solving the difference Equations (24) and (25). The predicted yield curves were then formed as  $y_{t_j}^{\Delta t, \Delta x, l}(x_i) = x_i Y^{\Delta t, \Delta x, l}$ .

To empirically assess the statistical performance of our method, we chose  $t = 1$  February 2020 and compared our predicted curves with the historically observed one. Figure 4 displays the observed yield rate curve (solid) and the estimated mean yield curve  $\bar{y}_{t_j}^{\Delta t, \Delta x, l}(x_i) = \frac{1}{10,000} \sum_{l=1}^{10,000} y_{t_j}^{\Delta t, \Delta x, l}(x_i)$  (dotted). Further, the dashed curves represent the lower and upper 99% pointwise prediction bounds. Note that the overall confidence level of the resulting tube region is less than 99% since no Bonferroni correction was performed. Figure 4 further contains 5 sample solution paths, which are all (essentially) contained in the prediction region. The estimated mean yield curve is also contained in the prediction region and does not substantially deviate from the observed curve. The prediction region is relatively tight suggesting good predictive power of our model.

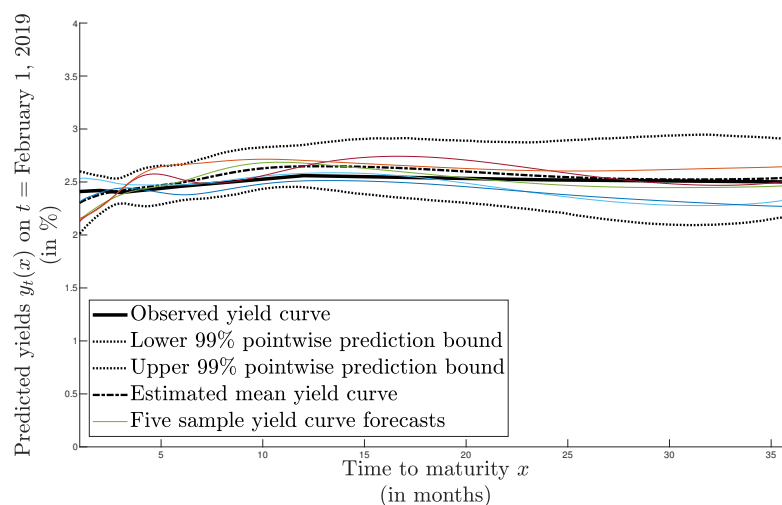


Figure 4. Predicting the yield rates on 1 February 2019 based on 2018 historical data.

The left panel of Figure 5 displays the historical yield curve rates for January 2019 (22 business days), while the right panel gives the estimated mean yield curve rate for the same time window based on 10,000 replications. We conjecture that the slight overestimation of the yield curve for large maturities may be related with the non-decreasing behavior of the long rates (see Section 4.1).

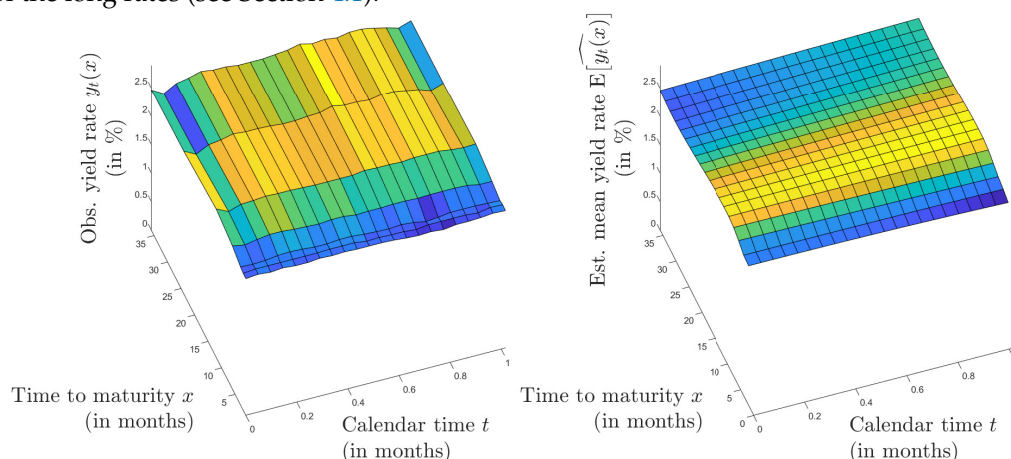


Figure 5. Observed yield curve rates (left) and estimated mean yield curve rates (right) for January 2019.

In sum, the presented model can be used both for prediction and estimation purposes as well as uncertainty quantification. Further potential applications include stochastic optimization and optimal stopping time problems, etc. The full set of Matlab® codes used to produce the results of Section 5 are available at <https://github.com/mpokojovy/abstract.HJM>.

6. Discussion and Conclusions

High level of abstraction and numerous prerequisites in various areas of probability theory, stochastic processes and martingale theory, functional analysis and operator semi-group theory, etc., are among the barriers encountered by researchers and practitioners when working with the abstract Heath-Jarrow-Morton (HJM) framework. We attempted to alleviate these difficulties by giving an accessible self-contained review of mathematical results related to the abstract HJM model. Based on real data, we also demonstrated how an infinite-dimensional HJM model can be calibrated and applied to forecasting US Treasury yield curve rates. Our future work will include empirical investigation of the nonlinear models as well as analysis of multiple zero coupon bonds.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2227-7390/9/2/114/s1>: Matlab<sup>®</sup> codes implementing the example in Section 5.

**Author Contributions:** Formal analysis, V.M.; Investigation, V.M.; Methodology, M.P.; Software, M.P. and V.M.; Visualization, M.P.; Writing—original draft, V.M. (Sections 2–4 and Appendices A–C) and M.P. (Sections 1 and 5); Writing—review & editing, M.P. and V.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset analyzed in this paper is available at <https://github.com/mpokojovy/abstract.HJM>.

**Acknowledgments:** The Authors thank the anonymous reviewers for their valuable comments and improvement suggestions.

**Conflicts of Interest:** The authors declare no conflict of interests.

## Appendix A. Q-Wiener Process

### Appendix A.1. Representation of Q-Wiener Process

**Definition A1** (Q-Wiener process). Let  $\mathcal{U}$  be a Hilbert space and  $Q \in \mathcal{L}(\mathcal{U})$  be symmetric, non-negative and with finite trace. A  $\mathcal{U}$ -valued stochastic process  $(W(t))_{t \in [0, T]}$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a Q-Wiener process if

- $W(0) = 0$ ,
- $W(t)$  has continuous trajectories  $\mathbb{P}$ -almost surely,
- $W(t)$  has independent increments, e.g., the random variables  $W(t_1), W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1})$  are independent for all  $0 \leq t_1 < t_2 < \dots < t_n \leq T$  and  $n \in \mathbb{N}$ ,
- The increments have Gaussian laws, e.g.,  $(W(t) - W(s)) \sim \mathcal{N}(0, (t - s)Q)$  for all  $0 \leq s \leq t \leq T$ .

**Proposition A1** (Representation of Q-Wiener process). Let  $Q \in \mathcal{L}(\mathcal{U})$  be symmetric, non-negative and with finite trace.

Let  $\{e_k\}_{k \in \mathbb{N}}$  denote an orthonormal basis of  $\mathcal{U}$  comprised of eigenvectors of  $Q$  associated with eigenvalues  $(\lambda_k)_{k \in \mathbb{N}}$ . For a stochastic  $\mathcal{U}$ -valued process  $(W(t))_{t \in [0, T]}$  to be Q-Wiener process, it is necessary and sufficient that

$$W(t) = \sum_{k \in \mathbb{N}} \sqrt{\lambda_k} \beta_k(t) e_k \quad \text{for all } t \in [0, T],$$

where  $\{\beta_k\}_{k \in \mathbb{N}}$  is a vector of independent real-valued Brownian motions with respect to a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The latter series is convergent in the topology of  $L^2((\Omega, \mathcal{F}, \mathbb{P}); C([0, T], \mathcal{U}))$ .

### Appendix A.2. Q-Wiener Process with Respect to Filtration

**Definition A2** (Filtration). Let  $T \in [0, \infty)$ . A filtration on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is a family of  $\sigma$ -algebras  $\{\mathcal{F}_t\}_{t \in [0, T]}$  with the property  $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$  for all  $0 \leq s \leq t \leq T$ .

**Definition A3** (Adapted process). A stochastic process  $(X(t))_{t \in [0, T]}$  is said to be adapted to a filtration  $\{\mathcal{F}_t\}_{t \in [0, T]}$  if for every  $t \in [0, T]$  the random variable  $X_t$  is  $\mathcal{F}_t$ -measurable.

**Definition A4** (Q-Wiener process). A stochastic process  $(W(t))_{t \in [0, T]}$  satisfying

- $(W(t))_{t \in [0, T]}$  is adapted to  $\{\mathcal{F}_t\}_{t \in [0, T]}$ ,
- the increments  $W(t) - W(s)$  are independent of  $\mathcal{F}_s$  for all  $0 \leq s \leq t \leq T$

is referred to as a Q-Wiener process.

**Proposition A2.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. Any  $\mathcal{U}$ -valued  $Q$ -Wiener process  $(W(t))_{t \in [0, T]}$  is a  $Q$ -Wiener process with respect to the normal filtration  $\{\mathcal{F}_t\}_{t \in [0, T]}$ .

**Appendix B. Conditional Expectation and Martingales in Banach Spaces**

**Proposition A3** (Existence of conditional expectation). Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $B$  a separable Banach space,  $X$  a Bochner integrable  $B$ -valued random variable and  $\mathcal{Y}$  a  $\sigma$ -algebra contained in  $\mathcal{F}$ . Then, up to a  $\mathbb{P}$ -negligible set, there exists a unique integrable  $B$ -valued,  $\mathcal{Y}$ -measurable random variable  $Z$  such that

$$\int_A X d\mathbb{P} = \int_A Z d\mathbb{P} \quad \text{for all } A \in \mathcal{Y}.$$

The random variable  $Z$  will be denoted by  $\mathbb{E}(X|\mathcal{Y})$  and referred to as the conditional expectation of  $X$  given  $\mathcal{Y}$ . Moreover, we have  $\|\mathbb{E}(X|\mathcal{Y})\|_B \leq \mathbb{E}(\|X\|_B|\mathcal{Y})$ .

Proposition A3 allows us to extend the definition of a martingale to an  $\mathcal{F}_t$ -martingale.

**Definition A5** ( $\mathcal{F}_t$ -martingale). A stochastic process  $(X(t))_{t \geq 0}$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with values in a separable Banach space  $B$  is called a martingale with respect to a filtration  $\{\mathcal{F}_t\}_{t \geq 0}$  (or an  $\mathcal{F}_t$ -martingale) if

- $\mathbb{E}(\|X(t)\|_B) < \infty$ ,
- $X(t)$  is  $\mathcal{F}_t$ -measurable for all  $t \geq 0$ ,
- $\mathbb{E}(X(t)|\mathcal{F}_s) = X(s)$   $\mathbb{P}$ -almost surely for all  $0 \leq s \leq t < \infty$ .

**Definition A6** ( $\mathcal{F}_t$ -submartingale). A real-valued integrable adapted stochastic process  $(X(t))_{t \geq 0}$  is called a submartingale if  $\mathbb{E}(X(t)|\mathcal{F}_s) \geq X(s)$  for all  $0 \leq s \leq t < \infty$ .

**Proposition A4.** If  $(X(t))_{t \geq 0}$  is a  $B$ -valued  $\mathcal{F}_t$ -martingale and  $p \in [1, \infty)$ , then  $\|X\|_B^p$  is a real-valued  $\mathcal{F}_t$ -submartingale.

**Proposition A5.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $T > 0$ . Then any  $B$ -valued  $Q$ -Wiener process  $(W(t))_{t \in [0, T]}$  with respect to a normal filtration  $\{\mathcal{F}_t\}_{t \in [0, T]}$  is a continuous square integrable martingale.

**Appendix C. Stochastic Integrals Driven by a Standard Cylindrical Wiener Process**

Appendix C.1. Cylindrical  $Q$ -Wiener Process

**Definition A7** (Cylindrical  $Q$ -Wiener process). Let  $Q$  be a linear bounded operator, but not necessarily with finite trace, on a Hilbert space  $\mathcal{U}$ . We call  $(W(t))_{t \in [0, T]}$  a cylindrical  $Q$ -Wiener process over  $\mathcal{U}$  if

$$W(t) = \sum_{k \in \mathbb{N}} \lambda_k^{\frac{1}{2}} \beta_k(t) e_k,$$

where  $\{(\beta_k(t))_{t \in [0, T]}\}_{k \in \mathbb{N}}$  is a family of independent Brownian motions,  $\{\lambda_k\}_{k \in \mathbb{N}}$  is a set of eigenvalues of  $Q$  with corresponding eigenvectors  $\{e_k\}_{k \in \mathbb{N}}$  which form a basis in  $\mathcal{U}$ . If  $Q = I$ ,  $(W(t))_{t \in [0, T]}$  is called a standard cylindrical Wiener process.

Since this series may not converge in  $L^2((\Omega, \mathcal{F}, \mathbb{P}); \mathcal{U})$  if  $Q$  has no finite trace, we will use the following propositions.

**Proposition A6.** Let  $Q$  be a linear bounded operator, but not necessarily with finite trace, on a Hilbert space  $\mathcal{U}$ . Let  $\mathcal{U}_1$  be a Hilbert space such that there exists a Hilbert-Schmidt embedding  $\mathcal{J}: \mathcal{U}_0 := Q^{\frac{1}{2}}(\mathcal{U}) \rightarrow \mathcal{U}_1$ . If  $\{e_k\}_{k \in \mathbb{N}}$  is an orthonormal basis of  $\mathcal{U}_0$  and  $\{(\beta_k(t))_{t \in [0, T]}\}_{k \in \mathbb{N}}$  is a

family of independent real-valued Brownian motions, then there exists  $Q_1 \in \mathcal{L}(\mathcal{U}_1)$  symmetric, non-negative and with finite trace such that the series

$$W(t) = \sum_{k \in \mathbb{N}} \beta_k(t) \mathcal{J}(e_k) \quad \text{for all } t \in [0, T],$$

converges in  $\|\cdot\|_{\mathcal{M}_T^2(\mathcal{U}_1)} = \left\{ \mathbb{E} \left( \sup_{t \in [0, T]} \|(\cdot)(t)\|_{\mathcal{U}_1}^2 \right) \right\}^{\frac{1}{2}}$  and defines a  $Q_1$ -Wiener process on  $\mathcal{U}_1$ .

Appendix C.2. Integration for Predictable Processes

**Definition A8** (Hilbert-Schmidt operator). Given an orthonormal basis  $\{e_k\}_{k \in \mathbb{N}}$  of  $\mathcal{U}$ , we refer to an operator  $A \in \mathcal{L}(\mathcal{U}, \mathcal{H})$  as a Hilbert-Schmidt operator if

$$\sum_{k \in \mathbb{N}} \langle Ae_k, Ae_k \rangle_{\mathcal{H}} < \infty.$$

The space of all Hilbert-Schmidt operators from  $\mathcal{U}$  to  $\mathcal{H}$  will be denoted by  $\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H})$ .

**Definition A9** (Elementary process). Let  $T$  be positive. A stochastic process  $(\Phi(t))_{t \in [0, T]}$  is said to be an elementary process on a Hilbert space  $\mathcal{U}$  if there is a finite family of disjoint intervals

$$(s_k, t_k] \subset [0, T], k \in \{1, 2, \dots, n\} \text{ for some } n \in \mathbb{N}, \bigcup_{k=1}^n (s_k, t_k] = (0, T],$$

and a corresponding finite family of  $\mathcal{F}_{s_k}$ -measurable random variables  $\{\Phi_k\}_{k \in \{1, 2, \dots, n\}}$  taking a finite number of values in the space of Hilbert-Schmidt operators  $\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H})$  such that for all  $t \in [0, T]$

$$\Phi(t, \omega) = \sum_{k=1}^n \Phi_k(\omega) \cdot \mathbb{1}_{(s_k, t_k]}(t) \quad \mathbb{P}\text{-almost surely.}$$

**Definition A10** (Stochastic integral for elementary process). For a given elementary process  $\Phi \in L^2([0, T] \times \Omega, \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H}))$  the stochastic integral is defined by

$$\int_0^T \Phi(t) dW(t) = \sum_{k=1}^n \Phi_k(W(t_k) - W(s_k)) \quad \text{for some } n \in \mathbb{N},$$

where  $(W(t))_{t \in [0, T]}$  is a standard cylindrical  $\mathcal{U}$ -valued Wiener process.

**Definition A11** (Predictable  $\sigma$ -algebra). A  $\sigma$ -algebra over  $[0, T] \times \Omega$  generated by all subsets of the form  $(s, t] \times A, A \in \mathcal{F}_s$  with  $\{0\} \times A_0$  for  $A_0 \in \mathcal{F}_0$  is called a predictable  $\sigma$ -algebra and is denoted by  $\mathcal{P}_T$ .

**Definition A12** (Predictable process). A stochastic process  $(X(t))_{t \geq 0}$  with values in a Hilbert space  $\mathcal{U}$  is called predictable if it is  $\mathcal{P}_T$ -measurable.

**Proposition A7.** The stochastic integral can be uniquely defined for every predictable process  $\Phi \in L^2([0, T] \times \Omega, \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H}))$ .

**Definition A13** ( $\mathcal{F}_t$ -stopping time). An  $\mathcal{F}_t$ -stopping time is a non-negative real-valued random variable  $\tau$  such that for all times  $t \geq 0$  the event  $\{\tau \leq t\} \in \mathcal{F}_t$ .

**Definition A14** (Simple  $\mathcal{F}_t$ -stopping time). An  $\mathcal{F}_t$ -stopping time  $\tau$  is said to be simple if it takes only a finite number of values.

**Theorem A1 (Itô isometry).** If  $(\Phi(t))_{t \in [0, T]} \in L^2([0, T] \times \Omega, \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H}))$  is predictable and  $(W(t))_{t \in [0, T]}$  is a standard  $\mathcal{U}$ -valued cylindrical Wiener process, then we have

$$\mathbb{E} \left( \left\| \int_0^T \Phi(s) dW(s) \right\|_{\mathcal{H}}^2 \right) = \mathbb{E} \left( \int_0^T \|\Phi(s)\|_{\mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H})}^2 ds \right) = \mathbb{E} \left( \int_0^T (\Phi^*(s) \Phi(s)) ds \right).$$

**Proposition A8.** Let  $(\Phi(t))_{t \in [0, T]} \in L^2([0, T] \times \Omega, \mathcal{L}_{\text{HS}}(\mathcal{U}, \mathcal{H}))$  be predictable,  $(W(t))_{t \in [0, T]}$  be a standard cylindrical  $\mathcal{U}$ -valued Wiener process and  $\tau$  be an  $\mathcal{F}_t$ -stopping time such that  $\mathbb{P}(\tau \leq T) = 1$ . Then we have  $\mathbb{P}$ -almost surely

$$\int_0^t \Phi(s) \mathbb{1}_{[0, \tau]}(s) dW(s) = \int_0^{\tau \wedge t} \Phi(s) dW(s).$$

**Proposition A9.** The stochastic integral defined by

$$\int_0^t \Phi(s) dW(s) := \lim_{n \rightarrow \infty} \int_0^t \mathbb{1}_{[0, \tau_n]}(s) \Phi(s) dW(s) \quad \text{for all } t \in [0, T]$$

does not depend on the choice of the sequence of stopping times  $\{\tau_n\}_{n \in \mathbb{N}}$  if  $\lim_{n \rightarrow \infty} \tau_n = T$ .

## References

1. Heath, D.; Jarrow, R.; Morton, A. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* **1992**, *60*, 77–105. [CrossRef]
2. Brigo, D.; Mercurio, F. *Interest Rate Models—Theory and Practice: With Smile, Inflation and Credit*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006.
3. Harrison, J.; Pliska, S. Martingales and Stochastic Integrals in the Theory of Continuous Trading. *Stoch. Proc. Appl.* **1981**, *11*, 215–260. [CrossRef]
4. Filipović, D. Consistency Problems for HJM Interest Rate Models. Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland, 2001.
5. Musiela, M. Stochastic PDEs and Term Structure Models. In *Journées Internationales de Finance*; IGR-AFFI: La Baule, France, 1993.
6. Carmona, R.; Tehranchi, M. *Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
7. US Department of the Treasury. Daily Treasury Yield Curve Rates. Available online: <https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield> (accessed on 14 April 2020).
8. Dai, Q.; Singleton, K. Specification analysis of affine term structure models. *J. Financ.* **2000**, *55*, 1943–1978. [CrossRef]
9. Duffee, G. Term premia and interest rate forecasts in affine models. *J. Financ.* **2002**, *57*, 405–443. [CrossRef]
10. Cheridito, P.; Filipović, D.; Kimmel, R. Market price of risk specifications for affine models: Theory and evidence. *J. Financ. Econ.* **2007**, *83*, 123–170. [CrossRef]
11. Da Prato, G.; Zabczyk, J. *Stochastic Equations in Infinite Dimensions*, 2nd ed.; Cambridge University Press: New York, NY, USA, 1992.
12. Reed, M.; Simon, B. *Methods of Modern Mathematical Physics*, 1st ed.; Academic Press Inc.: Cambridge, MA, USA, 1972.
13. Capinski, M.; Kopp, E. *Measure, Integral and Probability*, 2nd ed.; Springer Undergraduate Mathematics Series; Springer: Berlin/Heidelberg, Germany, 2007.
14. Rudin, W. *Functional Analysis*, 2nd ed.; International Series in Pure and Applied Mathematics; McGraw-Hill: New York, NY, USA, 1991.
15. Knoche, C.; Frieler, K. Solutions of Stochastic Differential Equations in Infinite Dimensional Hilbert Spaces and Their Dependence on Initial Data. Ph.D. Thesis, University of Bielefeld, Bielefeld, Germany, 2001.
16. Björk, T. *Arbitrage Theory in Continuous Time*, 4th ed.; Oxford Finance Series; Oxford University Press: Oxford, UK, 2020.
17. Cheyette, O. Interest Rates Models. In *Investment Management for Insurers*; Babel, D., Fabozzi, F., Eds.; Frank J. Fabozzi Associates: New Hope, PA, USA, 1999.
18. Kamizono, K.; Kariya, T. An implementation of the HJM model with application to Japanese interest futures. *Financ. Eng. Jpn. Market.* **1996**, *3*, 151–170. [CrossRef]
19. Sabelli, C.; Pioppi, M.; Sitzia, L.; Borretti, G. Multi-curve HJM modelling for risk management. *Quant. Financ.* **2018**, *18*, 563–590. [CrossRef]
20. Fernando, W. American Options Pricing Using HJM Approach. Ph.D. Thesis, UNC Charlotte, Charlotte, NC, USA, 2017.
21. Schmidt, T. An infinite factor model for credit risk. *Int. J. Theor. Appl. Financ.* **2006**, *9*, 43–68. [CrossRef]



22. Babbel, D.; Fabozzi, F. *Investment Management for Insurers*; Babbel, D., Fabozzi, F., Eds.; Frank J. Fabozzi Associates: New Hope, PA, USA, 1999.
23. Miron, P.; Swannell, P. *Pricing and Hedging Swaps*, 1st ed.; Euromoney Publications PLC. Nestor House, Playhouse Yard: London, UK, 1992.
24. Black, F.; Scholes, M. The Pricing of Options and Corporate Liabilities. *J. Pol. Econ.* **1973**, *81*, 637–654. [[CrossRef](#)]
25. Harrison, J.; Kreps, D. Martingales and arbitrage in multiperiod securities markets. *J. Econ. Theor.* **1979**, *20*, 381–408. [[CrossRef](#)]
26. Delbaen, F.; Schachermayer, W. *The Mathematics of Arbitrage*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
27. Dybvig, P.H.; Ingersoll, J.E.; Ross, S.A. Long forward and zero-coupon rates can never fall. *J. Bus.* **1996**, *59*, 1–25. [[CrossRef](#)]
28. Brezis, H. *Analyse Fonctionnelle*, 4th ed.; Théorie et Applications (French) [Theory and Applications]; Masson: Paris, France, 1993.
29. Bhar, R.; Chiarella, C. The estimation of the Heath-Jarrow-Morton model by use of Kalman filtering techniques. In *Computational Approaches to Economic Problems*; Springer: Boston, MA, USA, 1997; pp. 113–126. [[CrossRef](#)]
30. Pang, K. Calibration of Gaussian Heath, Jarrow and Morton and random field interest rate term structure models. *Rev. Derivat. Res.* **1998**, *2*, 315–345. [[CrossRef](#)]
31. Jeffrey, A.; Kristensen, D.; Linton, O.; Nguyen, T.; Phillips, P.C. Nonparametric estimation of a multifactor Heath-Jarrow-Morton model: An integrated approach. *J. Financ. Econom.* **2004**, *2*, 251–289. [[CrossRef](#)]
32. Broszkiewicz-Suwaj, E.; Weron, A. Calibration of the multi-factor HJM model for energy market. *Acta Phys. Polonica B* **2006**, *37*, 1455–1466.
33. Bhar, R.; Chiarella, C.; Tô, T.D. A Maximum Likelihood Approach to Estimation of a Class of the Heath-Jarrow-Morton Models. Available online: <https://econpapers.repec.org/paper/utsrpaper/80.htm> (accessed on 28 October 2020).
34. Falcò, L.N.; Nave, J. On the calibration of a Gaussian Heath-Jarrow-Morton model using consistent forward rate curves. *Quant. Financ.* **2011**, *11*, 495–504. [[CrossRef](#)]
35. Li, S.; Dong, Y. Parametric estimation in the Vasicek-type model driven by sub-fractional Brownian motion. *Algorithms* **2018**, *11*, 197. [[CrossRef](#)]
36. Funahashi, H. An approximate swaption formula in Heath-Jarrow-Morton models. *J. Derivat.* **2020**, *27*, 1–21.10.3905/jod.2020.1.101. [[CrossRef](#)]
37. Vasicek, O. An equilibrium characterization of the term structure. *J. Financ. Econom.* **1977**, *5*, 177–188. [[CrossRef](#)]
38. Dörsek, P.; Teichmann, D. Efficient simulation and calibration of general HJM models by splitting schemes. *SIAM J. Financ. Math.* **2013**, *4*, 575–598. [[CrossRef](#)]
39. Kluge, W. Time-Inhomogeneous Lévy Processes in Interest Rate and Credit Risk Models. Ph.D. Thesis, University of Freiburg, Freiburg, Germany, 2005.
40. Kim, S.; Pokojovy, M.; Wan, X. The taut string approach to statistical inverse problems: Theory and applications. *J. Comput. Appl. Math.* **2021**, *382*, 113098. [[CrossRef](#)]
41. Sachs, E.W.; Volkwein, S. POD-Galerkin approximations in PDE-constrained optimization. *GAMM-Mitteilungen* **2010**, *33*, 194–208. [[CrossRef](#)]
42. Tang, C.; Chen, S. Parameter estimation and bias correction for diffusion processes. *J. Econom.* **2009**, *149*, 65–81. [[CrossRef](#)]
43. Fergusson, K.; Platen, E. Application of maximum likelihood estimation stochastic short-rate models. *Ann. Financ. Econom.* **2015**, *10*, 1–26.
44. Ballestra, L.; Pacelli, G.; Radi, D. A note on Fergusson and Platen: “Application of maximum likelihood estimation stochastic short-rate models”. *Ann. Financ. Econom.* **2016**, *11*, 1–7. [[CrossRef](#)]