ROBUST CALIBRATION IN YIELD CURVE MODELLING

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
DAVID STEFANOVITS
MSc ETH in Applied Mathematics
born on 1. January 1986
citizen of Lugano (TI), Switzerland

accepted on the recommendation of
Prof. Dr. Mario V. Wüthrich, examiner
Prof. Dr. Josef Teichmann, co-examiner
Prof. Dr. Damir Filipović, co-examiner

2016
Abstract

Robust approaches to yield curve modelling are important for banks, asset managers and insurance companies. This thesis deals with several conceptual and practical difficulties associated with yield curve modelling.

A first issue is that models need to be recalibrated over time as new market information becomes available. Widely used and tractable affine models with static parameters cannot be recalibrated consistently, since new market yield curves may (and will) lie outside the set of possible realisations of the model. We show that one way to handle inconsistencies in the recalibration procedure is to allow for stochastic parameter changes (without necessarily sacrificing analytic tractability). Consistent recalibration models, as introduced in this thesis, turn static parameters of affine models into state variables such that the analytical tractability is maintained. We show that these provide flexible models which can be recalibrated consistently to new market information and yield curves.

Second, solvency directives in insurance rely on market consistent valuation, which means that liabilities must be valued by means of replication with liquid financial instruments. This is challenging because the value of long-term cash flows depends on yields beyond the maturities available in the liquid bond market. Hence, standard models, which assume unlimited availability of maturity dates, do not provide an appropriate framework for real-world solvency assessments. A yield curve model with limited availability of maturity dates is discussed in this thesis. Based on the model we develop a computational methodology to value and hedge long-term cash flows trading in short- and medium-term bonds which also includes reinvestment risks.

Third, empirical evidence suggests that short-term interest rates, which exhibit long-range dependence, may not be captured well by low-dimensional affine models. Accordingly, models driven by fractional processes have gained popularity. The non-Markovianity of these processes gives, however, rise to difficulties in computation and calibration. We introduce a class of fractional processes which can be represented as linear functionals of an infinite-dimensional affine process. The affine structure allows one to construct tractable models with fractional features.
Sintesi

Approcci robusti alla modellizzazione della curva dei rendimenti sono importanti per banche, gestori patrimoniali e società assicurative. Questa tesi tratta diverse problematiche concettuali e pratiche associate ai modelli della curva dei rendimenti.

La prima questione riguarda la necessità di ricalibrare i modelli nel tempo a seguito della disponibilità di nuove informazioni sul mercato. I modelli affini con parametri statici, che sono trattabili analiticamente e vengono largamente utilizzati, non permettono una ricalibrazione consistente. Le curve dei rendimenti costruite sulla base delle nuove informazioni sul mercato possono infatti trovarsi (ed in generale si trovano) al di fuori dell’insieme delle possibili realizzazioni del modello. Dimostriamo che una soluzione all’inconsistenza nella procedura di ricalibrazione consiste nella sostituzione dei parametri con processi stocastici (senza necessariamente sacrificare la trattabilità analitica). I modelli presentati in questa tesi trasformano i parametri statici dei modelli affini in variabili di stato mantenendo la trattabilità analitica. Illustriamo la flessibilità di questi modelli che consentono una ricalibrazione consistente con le curve dei rendimenti costruite sulla base delle nuove informazioni disponibili sul mercato.

La seconda questione riguarda la necessità di modelli per la curve dei rendimenti con una limitata disponibilità di scadenze. Le direttive di solvibilità per le società assicurative si basano sul principio della valutazione consistente con il mercato. Ciò in particolare significa che gli obblighi contrattuali devono essere valutati mediante la replicazione degli stessi con strumenti finanziari liquidi. Questo è complicato dato che il valore dei flussi di cassa a lungo termine dipende da rendimenti con scadenze non disponibili sul mercato obbligazionario liquido. Per questo motivo modelli convenzionali, che assumono la disponibilità illimitata di scadenze, non forniscono un quadro teorico appropriato per delle valutazioni di solvibilità adatte al mondo reale. Dopo aver proposto un modello adeguato a queste circostanze, sviluppiamo una metodologia computazionale per valutare e replicare flussi di cassa a lungo termine, effettuando contrattazioni sui mercati obbligazionari a corta e media scadenza, ciò comprende rischi dovuti ai reinvestimenti.
Infine, studi empirici suggeriscono che i tassi di interesse a corta scadenza che presentano dipendenza di lungo termine, possono non essere modellati adeguatamente da modelli affini di dimensione bassa. Di conseguenza, i modelli guidati da processi frazionari hanno acquisito popolarità. Tuttavia l’assenza della proprietà di Markov per questi processi dà luogo a difficoltà per quanto riguarda i calcoli e la calibrazione. Nella tesi introduciamo una classe di processi frazionari che possono essere rappresentati come funzionali lineari di un processo affine di dimensione infinita. La struttura affine permette di costruire modelli trattabili analiticamente e con proprietà frazionarie.
Acknowledgements

I would like to express my sincere gratitude to Mario Wüthrich and Josef Teichmann for supervising this thesis. I would also like to thank Philipp Harms who has worked with me on several projects. My research over the past years has greatly profited from their helpful contributions, thoughts and critical comments. I extend my gratitude to Damir Filipović for acting as co-examiner.

During my time at ETH Zurich I was generously supported by RiskLab and the Stochastic Finance Group which provided an exceptionally stimulating research environment. I benefited from many discussions with current and former members of these research groups.

I am very thankful to my former colleagues at Wegelin & Co. for allowing me to work part-time during the first part of my thesis. In particular, I would like to thank Urs Schubiger, Thomas Brusa, Daniel Leveau, Patrick Gander and Michael Calonder.

Financial support by the Swiss National Science Foundation (grant 149879) is gratefully acknowledged.
Contents

Abstract 3
Acknowledgements 7

1 Introduction 13

2 Consistent recalibration models 19
  2.1 Hull-White extended affine short rate models 19
  2.2 Construction of consistent recalibration models 23
  2.3 Properties of consistent recalibration models 26

3 Hedging in bond markets with reinvestment risk 33
  3.1 Bond market model with reinvestment risk 33
  3.2 Acceptable hedging 35
  3.3 Galerkin approximation of the hedging strategy 37

4 Affine models with fractional features 41
  4.1 $L^1$-valued Ornstein-Uhlenbeck process 41
  4.2 Connection to fractional Brownian motion 44
  4.3 Fractional short rate model 46
  4.4 Fractional Stein & Stein model 48

5 Conclusions 51

Accompanying papers 51
  Paper A – Consistent recalibration of yield curve models 51
  Paper B – Consistent re-calibration of the discrete-time multifactor Vasiček model 95
  Paper C – Exponential moments of time-inhomogeneous affine processes 131
  Paper D – Hedging of long term zero-coupon bonds in a market model with reinvestment risk 144
CONTENTS

Paper E – Affine representations of fractional processes with applications in mathematical finance 173

Cumulative Bibliography 217

Curriculum Vitae 223
Accompanying papers

A Philipp Harms, David Stefanovits, Josef Teichmann, Mario V. Wüthrich. 
Consistent recalibration of yield curve models. 

B Philipp Harms, David Stefanovits, Josef Teichmann, Mario V. Wüthrich. 
Consistent re-calibration of the discrete-time multifactor Vasiček model. 

C Philipp Harms, David Stefanovits, Josef Teichmann. 
Exponential moments of time-inhomogeneous affine processes.

D David Stefanovits, Mario V. Wüthrich. 
Hedging of long term zero-coupon bonds in a market model with reinvestment risk. 

E Philipp Harms, David Stefanovits. 
Affine representations of fractional processes with applications in mathematical finance. 
1. Introduction

In financial markets, contracts on the same underlying asset which differ in the maturity date trade at different prices. The function which relates the price of the contract to its maturity is called **term structure**. For instance, debt securities under which the issuer is obliged to pay interest and to repay the principal at the maturity date constitute the term structure of **interest rates**, also known as **yield curve**. Further examples of term structures include option implied volatilities, commodity futures, variance swaps and credit spreads. This thesis concentrates on yield curve modelling but many ideas apply to other term structure problems, as well.

Problems in risk management, pricing and hedging associated with the uncertainty in the shape of future yield curves are important for banks, asset managers and insurance companies. Two main approaches to yield curve modelling are **factor models** and Heath-Jarrow-Morton (HJM) models. Factor models are usually based on a finite-dimensional Markovian process, which describes selected market variables. No-arbitrage arguments then provide the corresponding yield curve, see Brigo and Mercurio [10] for an overview. Widely used and highly tractable are **affine factor models**, see e.g. Duffie, Filipović, and Schachermayer [30]. In Markovian HJM models the yield curve itself is taken as state variable and the state space is generally no longer finite-dimensional. A potentially difficult drift condition is necessary to guarantee absence of arbitrage, see Heath, Jarrow, and Morton [44]. HJM models are very general and flexible. However, without strong assumptions on the yield curve evolution, one usually encounters severe lack of analytic tractability and several interesting quantities for applications cannot be computed efficiently.

Yield curve modelling involves the following challenges:

(i) Affine factor models with static parameters generally allow to solve (at least numerically) complicated pricing and hedging questions. On the other hand in these models the potentially infinite-dimensional setup immediately gives rise to conceptual and practical difficulties. Affine factor models with static parameters generate yield curve processes with finite-dimensional support. Therefore, new market yield curves may (and will) lie outside the set of possible realisations of the model. This means that
recalibration of the model to the new market data implies a rejection of the previously selected model. This inconsistency is called \textit{recalibration problem}. The recalibration problem is most critical in low-dimensional affine factor models.

(ii) Affine factor models need to exhibit time-dependent, or even stochastic parameters, in order to achieve a reasonable fit of the market dynamics. This in turn may break their analytic tractability.

(iii) Standard yield curve models, which assume unlimited availability of maturity dates, do not provide the right framework for solvency modelling in insurance. In practice, the value of long-term liabilities depends on yields for maturity dates which are not available in the liquid bond market. Hence, dynamic hedging strategies need to be developed which account for reinvestment risks.

(iv) Low-dimensional affine factor models are not able to capture the statistical properties of short-term interest rates. In particular, empirical evidence of long-range dependence has been reported in the literature. Dependent increments are characteristic features of fractional processes. The non-Markovianity of fractional processes gives, however, rise to difficulties in computation and calibration.

In this thesis (i)-(iv) are addressed as described below.

\textbf{Consistent recalibration models}

A model can be recalibrated consistently if, loosely speaking, every neighbourhood of a state can be reached with positive probability. Therefore, the recalibration problem amounts to finding \textit{irreducible} Markovian models on sufficiently large state spaces (think of open subsets of a Hilbert space of yield curves). Additionally, the yield curve increments of the model should be sufficiently generic to accommodate changing market conditions, yet tractable enough to allow for efficient simulation. This results in finding models which allow one to replace static parameters by stochastic processes without losing tractability.

This thesis addresses (i) and (ii) by introducing a new class of HJM models which are called \textit{consistent recalibration} (CRC) models. These HJM models are constructed from \textit{tangent affine models} in a similar spirit as tangent Lévy models for the term structure of option prices, see Carmona and Nadtochyi [16], and Kallsen and Krühner [55]. As the coefficients of the tangent affine models are allowed to change over time and may even be stochastic, the conditions for finite-dimensional realisations are typically not fulfilled. Therefore, CRC models can
be irreducible. Additionally, CRC models can be simulated efficiently. This is a remarkable property because irreducible models on large state spaces are generally not tractable. The key is to exploit the *infinitesimally affine structure* of the models. This can be done by splitting the generator of the CRC semigroup into a part describing the yield curve dynamics under a fixed affine model and a second part describing the dynamic choice of the affine model. It turns out that this first order *splitting scheme* results in an efficient algorithm for simulation.

**Hedging in bond markets with reinvestment risk**

The issue of illiquidity (iii) is motivated by the problem of valuation and hedging of insurance liabilities. Solvency directives in insurance require to value liabilities in a *market consistent* way, see Wüthrich [81]. That is, insurance cash flows should be optimally replicated by an appropriate investment strategy of liquid financial instruments. The replication of long-term cash flows can be difficult because their value typically depends on interest rates beyond the maturities available in the liquid bond market. In that case, securities with shorter times to maturity need to be rolled over and the insurer is subject to *reinvestment risk*. Moreover, the cash flows may depend on stochastic factors, such as mortality rates, which cannot be traded in a liquid market. This thesis proposes a valuation approach to a specific sub-problem of the one outlined above. Deterministic long-term cash flows are replicated trading in short- and medium-term bonds. The considerations provide a fully specified computational methodology for pricing and hedging such deterministic long-term cash flows in a market model with reinvestment risk.

In practice, only bonds up to a certain maturity date are traded, and at each point in time a new bond is issued into the market with a maturity date previously unavailable. In the literature there are only a few papers dealing with the problem of modelling bonds up to a finite maturity date, see Dahl [25, 26]. An HJM model in discrete time is chosen in the spirit of Teichmann and Wüthrich [77]. In particular, the model presented in [77] is modified by introducing the restriction of finitely many available maturity dates. The model has two essential features for the problem of valuing long-term liabilities: it has been shown to model yield curves appropriately over a long period of time, and yield curves can be simulated efficiently.

Due to reinvestment risks there is no perfect hedging strategy available for long-term liabilities. Therefore, a suitable replication principle (which is tractable at least numerically) needs to be defined. Solvency directives also require to account for the level of *shortfall risk* in covering the liabilities. For this reason the *acceptable hedging* approach is used: illiquid long-term bonds are valued as the minimal capital which must be invested in liquid bonds to cover the cash
I N T R O D U C T I O N

flows at an acceptable level of risk, see Artzner, Delbaen, and Koch-Medina [2], and Hilli, Koivu, and Pennanen [47].

Affine models with fractional features

The last part of this thesis studies fractional processes in financial modelling which have recently gained popularity due to the dependence structure of their increments and the roughness of their sample paths. Currently, much attention is devoted to the study of asset pricing models with fractional stochastic volatility. One reason is that log-volatility has similar roughness and scaling properties as fractional Brownian motion (fBM), see Gatheral, Jaisson, and Rosenbaum [37]. Another reason is that fractional volatility models exhibit steep smiles on the short end of the volatility surface similar to those observed in the market, see Bayer, Friz, and Gatheral [6]. Previously, such smiles could be achieved only by adding jumps to the model. Empirical evidence also shows that short-term interest rates exhibit statistical properties, such as long-range dependence, which cannot be modelled by low-dimensional Markov processes, see e.g. Backus and Zin [3].

This thesis addresses (iv) by introducing a class of fractional processes which can be represented as a functional of a Banach-space valued affine process. The affine process is a collection of Ornstein-Uhlenbeck processes indexed by the speed of mean reversion. In particular, such a Markovian representation can be used in the case of fBM. The key idea is to express the fractional integral in the Mandelbrot and Van Ness representation [61] of fBM by a Laplace transform, see Carmona and Coutin [14]. The representation is important for the following reasons. First, due to the history dependence, fBM, seen from a future point in time, is not the same process as the original one. Thus, starting the model at a different time gives a different model. The Markovian structure behind fBM identifies the states of the model, making it comparable across time. Additionally, identifying and understanding model states is necessary for calibration. It is also within reach to solve simple models where the affine structure is preserved. For example, fractional yield curve models in a similar spirit as Biagini, Fink, and Klüppelberg [7] are shown to be tractable in the Markovian framework. Similar investigations are done in a generalisation of the affine stochastic volatility model by Stein and Stein [75] to fractional volatility.

Structure of the thesis

Chapters 2-4 summarise the main contributions of Papers A-E:

- Chapter 2 introduces CRC models and discusses their properties as in Papers A and B. Paper C presents results on the affine transform formula,
which relate real exponential moments of *time-inhomogeneous* affine processes to solutions of Riccati-type differential equations. These results are used in Chapter 2 for the construction of CRC models.

- Chapter 3 presents the computational methodology developed in Paper D to value and hedge long-term zero-coupon bonds trading in short- and medium-term ones.

- Chapter 4 presents a class of fractional processes which can be represented as linear functionals of an infinite-dimensional affine process as described in Paper E. This class is used to construct financial models driven by fractional processes with positively or negatively correlated increments. These processes are closely related to fBM.

The full papers are given in the appendix.
2. Consistent recalibration models

We present a new class of yield curve models which combines the advantages of factor and HJM models. We call this class of models by its most distinguished property consistent recalibration (CRC) models. In Section 2.1 we introduce the class of Hull-White extended affine short rate models, which are the building blocks of CRC models introduced in Section 2.2. The most important properties of CRC models are discussed in Section 2.3. The full theory for a continuous path version of CRC models is developed in Paper A. CRC models within the one-factor Vasicek and Cox-Ingersoll-Ross frameworks are presented in full detail in Paper A. A multifactor example in discrete time is discussed in detail in Paper B. Paper C extends some of the results on affine processes of Keller-Ressel and Mayerhofer [56] to the time-inhomogeneous case, and its main result is used in Paper A to calculate zero-coupon bond prices in the class of Hull-White extended affine short rate models. The proofs of the results presented in this chapter are found in Paper A.

2.1 Hull-White extended affine short rate models

Let \((\Omega, \mathcal{F}, (\mathcal{F}(t))_{t \geq 0}, Q)\) be a filtered probability space satisfying the usual conditions. The measure \(Q\) plays the role of a risk-neutral measure. All continuous-time stochastic processes are defined on \(\Omega\), adapted to \((\mathcal{F}(t))_{t \geq 0}\) and càdlàg. \(W = (W(t))_{t \geq 0}\) is \(d\)-dimensional \((\mathcal{F}(t))_{t \geq 0}\)-Brownian motion. Let \((t_n)_{n \in \mathbb{N}_0}\) be a strictly increasing sequence of deterministic times \(t_n \in [0, \infty)\). All discrete-time processes on this grid are defined on \(\Omega\) and adapted to the filtration \((\mathcal{F}(t_n))_{n \in \mathbb{N}_0}\).

The short rate process \(r = (r(t))_{t \geq 0}\) is determined by a \(d\)-dimensional factor process \(X = (X(t))_{t \geq 0}\) taking values in a state space \(X \subset \mathbb{R}^d\). For essentials on short rate models we refer to [34, Chapter 5]. The evolution of the factor process depends on a \(p\)-dimensional parameter process \(Z\) taking values in \(Z \subset \mathbb{R}^p\). In this section the parameter process \(Z = z\) is assumed to be constant and fixed, whereas it is allowed to vary in Sections 2.2 and 2.3, below. The canonical basis vectors in \(\mathbb{R}^d\) are denoted by \(e^1, \ldots, e^d\).
The factor process $X$ is assumed to be a continuous, $\mathbb{X}$-valued solution of the SDE
\[
dX(t) = \left(\theta(t)e^1 + b_z(X(t))\right)dt + \sqrt{a_z(X(t))}dW(t), \quad X(0) = x \in \mathbb{X}, \tag{2.1}
\]
where $\theta \in C(\mathbb{R}_+)$ and for each $(x, z) \in \mathbb{X} \times \mathbb{Z}$
\[
a_z(x) = a_z + \sum_{i=1}^{d} \alpha^i_x z^i, \quad b_z(x) = b_z + \sum_{i=1}^{d} \beta^i_x z^i,
\]
for symmetric positive semidefinite matrices $a_z, a^1_z, \ldots, a^d_z \in \mathbb{R}^{d \times d}$ and vectors $b_z, \beta^1_z, \ldots, \beta^d_z \in \mathbb{R}^d$. We denote by $\sqrt{a_z(x)}$ the symmetric positive semidefinite square root of $a_z(x)$. We assume that the parameters $(\theta, z)$ are such that SDE (2.1) has a unique continuous, $\mathbb{X}$-valued solution $X$, for each initial condition $X(s) = x$, where $(s, x) \in \mathbb{R}_+ \times \mathbb{X}$. In this case, the parameters $(\theta, z)$ are called admissible. The process $X$, or rather the family of processes obtained by varying the initial conditions in SDE (2.1), is time-inhomogeneous affine, see Filipović [33]. All coefficients in SDE (2.1) except for $\theta$ are time-independent. The extension to a time-dependent drift coefficient $\theta$ is due to Hull and White [50], and $X$ is called Hull-White extended affine. The function $\theta$ is called Hull-White extension and is used to calibrate the model to a given initial yield curve. Without loss of generality, we choose the first component for this extension. The short rate is given in terms of the factor process by
\[
r(t) = \ell + \langle \lambda, X(t) \rangle, \quad t \geq 0, \tag{2.2}
\]
where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product on $\mathbb{R}^d$, $\ell \in \mathbb{R}$ and $\lambda \in \mathbb{R}^d$ satisfies $\langle \lambda, e^1 \rangle \neq 0$. This class of short rate models is called Hull-White extended affine. In order to obtain well-defined bond prices we assume that the factor process $X$ satisfies the moment condition
\[
\mathbb{E}^Q \left[ e^{-\int_0^T (\ell + \langle \lambda, X(s) \rangle)} ds \right] < \infty, \quad \forall \ t \geq 0. \tag{2.3}
\]
Then no-arbitrage arguments provide the zero-coupon bond (ZCB) prices and forward rates.

**Theorem 2.1 (ZCB prices and forward rates)** Let $X$ satisfy condition (2.3). Then for $0 \leq t \leq T$ the ZCB prices in short rate model (2.1)-(2.2) satisfy
\[
P(t, T) = \mathbb{E}^Q \left[ e^{-\int_t^T r(s) ds} \mathcal{F}(T) \right]
\]
\[
= e^{-\ell(T-t) + \int_t^T \theta(s) \langle \Psi_z(T-s), e^1 \rangle ds + \Phi_z(T-t) + \langle \Psi_z(T-t), X(t) \rangle}, \tag{2.4}
\]
\[
= e^{-\ell(T-t) + \int_t^T \theta(s) \langle \Psi_z(T-s), e^1 \rangle ds + \Phi_z(T-t) + \langle \Psi_z(T-t), X(t) \rangle},
\]
\[
20
\]
and the forward rates are given by

\[ h(t)(\tau) = \left[ e^{-\int_0^T \theta(t+s)\langle \psi'(\tau-s), e_1 \rangle ds - \Phi' z(\tau) - \langle \psi' z(\tau), X(t) \rangle} \right] \]

where \((\Phi z, \Psi z) \in C^\infty(\mathbb{R}_+) \times C^\infty(\mathbb{R}_+, \mathbb{R}^d)\) is the unique solution of the system of Riccati ODE's

\[ \partial_\tau \Phi z = \frac{1}{2} \langle \Psi z, a_z \Psi z \rangle + \langle \Psi z, b_z \rangle, \]

\[ \langle \partial_\tau \Psi z, e^i \rangle = \frac{1}{2} \langle \Psi z, a_z^i \Psi z \rangle + \langle \Psi z, b_z^i \rangle - \lambda_i, \quad i \in \{1, \ldots, d\}, \]

with initial condition \((\Phi z(0), \Psi z(0)) = 0\).

Formula (2.4) follows from the relationship between exponential moments and Riccati ODE's for Hull-White extended affine models. We consider the \(X^2\)-valued stochastic process \(\tilde{X} = (\tilde{X}(t))_{t \geq 0}\) defined by

\[ \tilde{X}(t) = \left( X(t), \int_0^t X(s) ds \right). \]

It can be verified that the family of stochastic processes obtained by varying the initial condition of \(\tilde{X}\) is strongly regular affine, see Filipović [33, Theorem 2.14]. By the main result of Paper C we see that the moment condition (2.3) is equivalent to the existence of a solution \((\Phi z, \Psi z)\) of system (2.6). Uniqueness for these equations follows because the coefficients are locally Lipschitz.

At this point it is important to say that \(\Psi z\) depends only on the coefficients \(a^1_z, \ldots, a^d_z\) and \(b^1_z, \ldots, b^d_z\), whereas \(\Phi z\) generally depends on all the coefficients in SDE (2.1). This is important for estimation and calibration as we explain in Section 2.3.1, below.

Relation (2.5) between the forward rate curve \(h\) and the Hull-White extension \(\theta\) is used for calibrating the model, and is going to play a crucial role in the construction of CRC models. For this reason we introduce the following concise notation for equation (2.5)

\[ h(t) = \mathcal{H}_z(X(t), \mathcal{I}(t)\theta), \quad t \geq 0, \]

where for each \((t, x) \in \mathbb{R}_+ \times X\) and given parameter vector \(z \in Z\)

\[ \mathcal{I}(t)\theta = \theta(t+\cdot) \in C(\mathbb{R}_+), \]

\[ \mathcal{H}_z(x, \theta) = \ell - \int_0^t \theta(s) \langle \psi'_z(\cdot-s), e_1 \rangle ds - \Phi'_z(\cdot) - \langle \psi'_z(\cdot), x \rangle \in C^1(\mathbb{R}_+). \]
The map $\mathcal{S}(t) : C(\mathbb{R}_+) \rightarrow C(\mathbb{R}_+)$ is called shift operator. We also introduce the following notation for the inverse operation of (2.5)

$$\mathcal{S}(t)\theta = \mathcal{C}_z(h(t), X(t)), \quad t \geq 0.$$  

The operator $\mathcal{C}_z$ performs calibration of the Hull-White extension to a given forward rate curve. The inversion given by $\mathcal{C}_z$ involves solving a Volterra integral equation. In Paper A we show that $\mathcal{C}_z$ is well-defined for all $(x, h) \in \mathcal{X} \times C^1(\mathbb{R}_+)$ satisfying $h(0) = \ell + \langle \lambda, x \rangle$.

The evolution of the forward rate curve is described by the Heath-Jarrow-Morton (HJM) equation. In the following we formulate the HJM equation for the class of Hull-White extended affine short rate models. The derivation can be found in Paper A. For each $(x, z) \in \mathcal{X} \times \mathcal{Z}$ let $\mu_{z}^{HJM}(x) \in C^\infty(\mathbb{R}_+)$ and $\sigma_{z}^{HJM}(x) \in C^\infty(\mathbb{R}_+; \mathbb{R}^d)$ be given by

$$\mu_{z}^{HJM}(x) = \langle \Psi_z, a_z(x) \Psi'_z \rangle, \quad \sigma_{z}^{HJM}(x) = -\sqrt{a_z(x)} \Psi'_z.$$  

Note that the familiar HJM drift condition holds:

$$\mu_{z}^{HJM}(x)(t) = \int_0^T \sigma_{z}^{HJM}(x) dW(s), \quad \forall t \geq 0. \quad (2.7)$$  

Let $\mathcal{H} \subset C(\mathbb{R}_+)$ be a Hilbert space destined to contain the forward rate curve of the model. By abuse of notation the symbol $h$ is used interchangeably to denote an element of $\mathcal{H}$ and the forward rate curve process. We assume that $\mathcal{H}$ satisfies the following properties simultaneously for all $z \in \mathcal{Z}$.

(H1) The evaluation map $\text{eval}_t : h \mapsto h(t)$ is continuous on $\mathcal{H}$, for each $t \in \mathbb{R}_+$.

(H2) For each $(x, z, w) \in \mathcal{X} \times \mathcal{Z} \times \mathbb{R}^d$, the functions $\mu_{z}^{HJM}(x)$ and $\langle \sigma_{z}^{HJM}(x), w \rangle$ are elements of $\mathcal{H}$.

(H3) The right shifts $(\mathcal{S}(t))_{t \geq 0}$ mapping $h$ to $h(t + \cdot)$ define a strongly continuous semigroup on $\mathcal{H}$ with infinitesimal generator $\mathcal{A}$.

Hilbert spaces of forward rate curves which comply with (H1)-(H3) are constructed in Filipović [32, Sections 7.4.1 and 7.4.2] for the Vasiček and CIR models. In the domain $\mathcal{D}(\mathcal{A}) \subset \mathcal{H} \cap C^1(\mathbb{R}_+)$ (c.f. Filipović [32, Lemma 4.2.2]) of the infinitesimal generator $\mathcal{A}$ we can characterise the process $(h, X)$ as follows.

**Theorem 2.2 (HJM equation)** Let $(h, X)$ be given by Theorem 2.1 and assume that $h(t) \in \mathcal{D}(\mathcal{A})$, for each $t \geq 0$. Then the process $(h, X)$ is a strong solution of the following SPDE on $\mathcal{H} \times \mathcal{X}$:

$$dh(t) = \left[ t \mathcal{A} h(t) + \mu_{z}^{HJM}(X(t)) \right] dt + \sigma_{z}^{HJM}(X(t))dW(t),$$  

$$dX(t) = \left[ \theta(t) e^1 + b_z(X(t)) \right] dt + \sqrt{a_z(X(t))}dW(t).$$
For the concept of strong solution in Theorem 2.2 we refer to Filipović [32], and Da Prato and Zabczyk [24]. As we are going to see in the next section the HJM equation in Theorem 2.2 plays a crucial role in the construction of CRC models.

### 2.2 Construction of consistent recalibration models

The basic idea behind CRC models can be explained as follows: one concatenates forward rate increments of Hull-White extended affine factor models while allowing for random parameter changes. In other words CRC models are tangent to affine models, see Carmona and Nadtochiy [16], Kallsen and Krühner [55], and Richter and Teichmann [72]. The construction leads to tractable Markovian HJM models.

Considering tangent affine models in the context of yield curve modelling is very natural because affine models admit a finite-dimensional realisation for the yield curve process, see Filipović and Teichmann [35]. The fact that the yield curve process takes values in a finite-dimensional subspace of $\mathbb{H}$ is exactly what makes affine models tractable, but it also means that there is no hope for irreducibility on any reasonably large set of yield curves.

CRC models provide an extension of affine models which allows for stochastic parameters without losing analytic tractability. At least we want to simulate yield curve increments accurately and efficiently. This is feasible for CRC models since yield curve increments are well-behaved and understood. Typically, noise in the parameter process leads to irredducible models.

The constant parameter vector $z$ of Section 2.1 is replaced by a stochastic process $Z$ in this section. The situation is particularly simple when $Z$ is piece-wise constant.

A discrete-time stochastic process $(h(t_n), X(t_n), Z(t_n))_{n \in \mathbb{N}_0}$ with values in $\mathbb{H} \times X \times Z$ is called a discrete-time CRC model if there exists a discrete-time process $(\theta(t_n)(\cdot))_{n \in \mathbb{N}_0}$ with values in $C(\mathbb{R}_+)$ such that conditions (CRC1)-(CRC3) below are satisfied, for each $n \in \mathbb{N}_0$.

**(CRC1)** The Hull-White extension $\theta(t_n)$ is obtained by calibration to $h(t_n)$:

$$h(t_n)(0) = \ell + \langle \lambda, X(t_n) \rangle \quad \text{and} \quad \theta(t_n)(\cdot) = \mathcal{E}_{Z(t_n)}(h(t_n), X(t_n))(\cdot),$$

where the bracket $(\cdot)$ indicates the dependence on the time to maturity.

**(CRC2)** The evolution of $X$ on $[t_n, t_{n+1}]$ corresponds to the Hull-White extended affine model determined by the parameters $(\theta(t_n), Z(t_n))$:

$$X(t_{n+1}) = X_{\theta(t_n), Z(t_n)}^{t_n, X(t_n)}(t_{n+1}),$$
where for any \((s, x) \in \mathbb{R}_+ \times X\) and for any admissible parameters \((\theta, z) \in C(\mathbb{R}_+) \times Z\) the process \(X_{\theta, z}^{s, x}\) denotes the unique solution on \([s, \infty)\) of SDE (2.1) with \(\theta(t)\) replaced by \(\theta(t - s)\) and initial condition \(X(s) = x\). Here, we assume that \((\theta(t), Z(t))\) are admissible and that \(X_{\theta(t), Z(t)}^{t_n, X(t_n)}\) satisfies moment condition (2.3).

(CRC3) The forward rate \(h(t_{n+1})\) is determined by \(X(t_{n+1})\) according to the prevailing Hull-White extended affine model:

\[
h(t_{n+1}) = \mathcal{H}_Z(t_n)(X(t_{n+1}), \mathcal{F}(t_{n+1} - t_n) \theta(t_n)).
\]

We construct continuous-time CRC models in two steps. First, we extend the discrete-time CRC model between the grid points \((t_n)_{n \in \mathbb{N}_0}\) assuming a piecewise constant parameter process \(Z\). This allows one to see the connection to HJM models. Once the HJM equation for the extended model is derived then we can define CRC models directly as continuous-time HJM models for a general continuous-time parameter process \(Z\).

We extend the process \((h, X, Z, \theta)\) given by conditions (CRC1)-(CRC3) to \(\mathbb{R}_+\) by defining for each \(t \in [t_n, t_{n+1})\) and \(n \in \mathbb{N}_0\):

\[
\begin{align*}
Z(t) &= Z(t_n), & \theta(t) &= \mathcal{F}(t - t_n) \theta(t_n), \\
X(t) &= X_{\theta(t), Z(t_n)}^{t_n, X(t_n)}(t), & h(t) &= \mathcal{H}_Z(t_n)(X(t), \mathcal{F}(t - t_n) \theta(t_n)).
\end{align*}
\]

(2.8)

Additionally, we impose that \((h, X)\) is continuous at every grid point \(t_n\), and we assume that \((\theta(t), Z(t))\) is admissible for each \(t \in \mathbb{R}_+\). This construction uniquely defines the process \((h, X, Z, \theta)\) on \(\mathbb{R}_+\). Observe that the recalibration still happens on a discrete-time grid since \(Z\) is constant on \([t_n, t_{n+1})\).

In the next theorem we formulate the relation between ZCB prices and forward rates for CRC models. The theorem also shows that the construction leads to yield curve models which are free of arbitrage.

**Theorem 2.3 (ZCB prices and forward rates)** Let \((h, X, Z)\) be a CRC model as defined by (CRC1)-(CRC3) and (2.8) with corresponding process \(\theta\). Define

\[
P(t, T) = e^{-\int_t^T h(s) \, ds}, \quad r(t) = h(t)(0), \quad B(t) = e^{\int_0^t r(s) \, ds}, \quad 0 \leq t \leq T.
\]

Then the discounted ZCB price process \(t \mapsto B^{-1}(t)P(t, T)\) is a \(\mathbb{Q}\)-martingale, for each \(T \geq 0\). In this sense, the bond market is free of arbitrage. Moreover, bond prices and short rates are related by

\[
P(t, T) = \mathbb{E}^\mathbb{Q}\left[ e^{-\int_t^T r(s) \, ds} \mathbb{F}(t) \right],
\]

24
and the following affine bond pricing formulas hold:

\[
P(t, T) = e^{-\ell(T-t) + \int_0^{T-t} \vartheta(s) (\psi_{Z(t)}(T-t-s), e^1) ds + \Phi_{Z(t)}(T-t) + (\psi_{Z(t)}(T-t), X(t))},
\]

\[
h(t, \tau) = \ell - \int_0^{\tau} \vartheta(s) (\psi'_{Z(t)}(\tau-s), e^1) ds - \Phi'_{Z(t)}(\tau) - \langle \psi'_{Z(t)}(\tau), X(t) \rangle.
\]

CRC models have the remarkable property that, even if the parameters change stochastically, the ZCB price \( P(t, T) \) and the forward rate curve \( h(t) \) can be expressed explicitly in terms of \( (X(t), \Phi_{Z(t)}, \psi_{Z(t)}, \vartheta(t)) \). Moreover, it turns out that the HJM equation of CRC models has a nice form. The next theorem shows that the forward rate evolution is obtained replacing the fixed parameter vector \( z \) in the HJM equation of Theorem 2.2 by the stochastic process \( Z \).

**Theorem 2.4 (HJM equation)** Let \( (h, X, Z) \) be a CRC model as defined by (CRC1)-(CRC3) and (2.8) with corresponding process \( \vartheta \), and assume that \( h(t) \in D(\mathcal{A}) \), for each \( t \geq 0 \). Then the following properties hold:

1. **(HJM1)** the expression \( C_{Z(t)}(h(t), X(t)) \) is well-defined and equal to \( \vartheta(t) \), for all \( t \geq 0 \);

2. **(HJM2)** the parameters \( (\vartheta(t), Z(t)) \) are admissible, for all \( t \geq 0 \); and

3. **(HJM3)** the process \( (h, X) \) is a strong solution of the following SPDE on \( \mathbb{H} \times \mathbb{X} \):

\[
dh(t) = \left( \mathcal{A} h(t) + \mu_{Z(t)}(X(t)) \right) dt + \sigma_{Z(t)}(X(t)) dW(t),
\]

\[
dx(t) = \left( C_{Z(t)}(h(t), X(t)) e^1 + b_{Z(t)}(X(t)) \right) dt + \sqrt{a_{Z(t)}(X(t))} dW(t).
\]

We use the HJM equation derived above in order to define CRC models for a general continuous-time parameter process \( Z \), i.e., we do not require \( Z \) to be piecewise constant. A continuous-time CRC model is a continuous process \( (h, X, Z) \) with values in \( \mathbb{H} \times \mathbb{X} \times \mathbb{Z} \) satisfying (HJM1)-(HJM3) of Theorem 2.4. These models are free of arbitrage since the HJM drift condition (2.7) is satisfied.

We could define CRC models directly in the HJM framework assuming a stochastic dynamics for the parameters. However, solely from the HJM representation, one cannot see that the yield curve dynamics is obtained by combining well-understood Hull-White extended affine short rate models, see conditions (CRC1)-(CRC3). As we show in Section 2.3.3, below, the fact that yield curve increments are well-understood is very useful to simulate continuous-time CRC models accurately and efficiently.
2.3 Properties of consistent recalibration models

In this section we show that CRC models combine the flexibility of HJM models with the analytic tractability of affine factor models.

2.3.1 Robust calibration

Model selection should not be based purely on statistical methods; rather, it should be a combination of estimation and calibration. We call this model selection paradigm robust calibration. Accordingly, the model is selected simultaneously from time series and prevailing market prices in the spirit of Cuchiero and Teichmann [22]. Model parameters which are invariant under equivalent measure changes should be estimated by a statistical procedure from time series data. The remaining parameters are calibrated by solving an inverse problem with respect to prevailing market prices. All model parameters should be constant during the life time of the model; only state variables may change. Requiring parameters to remain constant throughout the life time of the model is less restrictive in CRC models than in the underlying affine factor models. The reason is that the parameters of the underlying affine model are turned into state variables of the CRC model. This provides additional modelling flexibility which can lead to a better fit of the yield curve dynamics.

In order to fit CRC models we need to estimate a time series for the parameter process $Z$ from market data, and fit a model for this time series. By Theorem 2.4 quadratic covariations of forward rates for CRC models satisfy

$$d[h(\cdot)(\tau_1), h(\cdot)(\tau_2)](t) = \Psi'_{Z(t)}(\tau_1)^T \left(a_{Z(t)} + \sum_{i=1}^{d} \alpha^i_{Z(t)} X_i(t) \right) \Psi'_{Z(t)}(\tau_2) dt, \quad (2.9)$$

for any times to maturity $\tau_1, \tau_2 \in \mathbb{R}_+$. The right-hand side of this equation depends on the coefficients $a_{Z(t)}$, $\alpha^i_{Z(t)}$ and $\beta^i_{Z(t)}$. Observe the implicit dependence of $\Psi_{Z(t)}$ on $\beta^i_{Z(t)}$. We assume that time series of estimated quadratic covariances are available. Therefore, for any selection of times to maturity $\tau_1$ and $\tau_2$, estimators for the coefficients can be obtained by solving for those coefficients which achieve the best fit in equation (2.9), see Figures 2.3.1-2.3.4. These are model parameters under a risk-neutral measure, but they can be estimated from real-world observations since (2.9) is obtained solely from the volatility of the forward rate process.

For given $a_{Z(t)}$, $\alpha^i_{Z(t)}$, and $\beta^i_{Z(t)}$ the remaining parameter $b_{Z(t)}$ may be calibrated to the prevailing market yield curve by regression methods. Alternatively, it can be estimated by econometric methods. However, this typically requires a market price of risk specification, whose parameters have to be estimated at the same time. Note that for one-factor models $b_{Z(t)}$ is redundant and can be
Figure 2.3.1: Parameter $a$ estimated from market volatilities in the one-factor Vasilec model. The estimation is based on equation (2.9) for $\tau_1 = \tau_2$ (time to maturity in yearly units). See Paper A for more details.

normalised to zero because of the Hull-White extension (c.f. SDE (2.1)). Finally, given the estimates of $a_{Z(t)}$, $\alpha_{Z(t)}^i$, $\beta_{Z(t)}^i$ and $b_{Z(t)}$ we calibrate the Hull-White extension $\theta$ to the prevailing market yield curve, see Figures 2.3.5-2.3.6 for the calibration results as of 1 April 2014.

Varying the estimation and calibration time $t$ produces time series of estimates for $b_{Z(t)}$, $a_{Z(t)}$, $\alpha_{Z(t)}^i$ and $\beta_{Z(t)}^i$ (c.f. Figures 2.3.1-2.3.4). Selecting CRC models requires the additional task of fitting a model for the estimated time series. This completes the model specification in accordance with the robust calibration paradigm. Whenever possible, the parameters are estimated from volatilities of yields, which allows one to bypass the usual inverse problems in calibration. In Papers A and B the procedure is presented in more details.

2.3.2 Consistent recalibration

A model is called consistent if the stochastic process $(h(t))_{t \geq 0}$ does not leave a pre-specified subset $\mathcal{I} \subset \mathcal{H}$ of possible forward rate curves. Additionally, we consider the following requirement: the process $(h(t))_{t \geq 0}$ should, loosely speaking, be able to reach any neighbourhood of any curve in $\mathcal{I}$ with positive probability, since any newly arriving market configuration is a possible model state. Consequently, the model can be recalibrated to any new market configuration without losing consistency to the model with old parameters. For this reason we say that the model satisfies the consistent recalibration property.
Figure 2.3.2: Parameter $\alpha$ estimated from market volatilities in the one-factor Cox-Ingersoll-Ross model. The estimation is based on equation (2.9) for $\tau_1 = \tau_2$ (time to maturity in yearly units). See Paper A for more details.

Figure 2.3.3: Parameter $\beta$ estimated from market volatilities in the one-factor Vasicek model. The estimation is based on equation (2.9) for $\tau_1 = \tau_2$ (time to maturity in yearly units). See Paper A for more details.
2.3. Properties of consistent recalibration models

Speed of mean-reversion in the CIR model

![Graph](attachment:image.png)

Figure 2.3.4: Parameter $\beta$ estimated from market volatilities in the one-factor Cox-Ingersoll-Ross model. The estimation is based on equation (2.9) for $\tau_1 = \tau_2$ (time to maturity in yearly units). See Paper A for more details.

Zero-coupon yields (%)

![Graph](attachment:image.png)

Figure 2.3.5: Calibration of affine models to the initial yield curve (Market as of 1 April 2014). Vasiček 1-2 and CIR 1-2 are time-homogenous (i.e. $\theta$ is constant) affine models with parameters estimated from market data by using two different estimation procedures (1-2). HWE Vasiček 1-2 and HWE CIR 1-2 are their Hull-White extended counterparts, where $\theta$ is calibrated to the initial yield curve. See Paper A for more details.
The consistent recalibration property can be verified by checking Hörmander’s condition, see Baudoin and Teichmann [5]. For a continuous-time CRC model \((h, X, Z)\) taking values in \(H \times X \times Z\) the condition is satisfied as soon as there is noise in the parameter process. In the following we discuss Hörmander’s condition in the Vasicek framework.

Let \(X = \mathbb{R}\), \(\ell = 0\) and \(\lambda = 1\). We assume that the parameter process \(Z\) takes values in \(Z = \mathbb{R}^p\) and satisfies

\[
dZ(t) = \mu(Z(t)) \, dt + \sigma(Z(t)) \, dW^Z(t),
\]

where \(\mu \in C_\infty^b(\mathbb{R}^p; \mathbb{R}^p), \sigma \in C_\infty^b(\mathbb{R}^p; \mathbb{R}^{p \times q}),\) and \(W^Z\) is \(q\)-dimensional Brownian motion independent of \(W\). We write \(C_\infty^b\) for bounded functions with bounded derivatives of all orders. The components of \(\sigma\) are written as \(\sigma = (\sigma_1, \ldots, \sigma_q),\) where \(\sigma_i : Z \rightarrow Z\). For each \((x, z) \in X \times Z\) the volatility and drift coefficients are assumed to be given by

\[
a_z(x) = a_z, \quad b_z(x) = \beta_z x,
\]

where \(a_z \in [0, \infty), \beta_z \in (-\infty, 0)\). Furthermore, assume that the mappings \(z \mapsto \sqrt{a_z}\) and \(z \mapsto \beta_z\) are of class \(C_\infty\) and that \(\beta_z\) is bounded away from 0, i.e.

\[
\sup_{z \in \mathbb{R}^p} \beta_z < 0.
\]

In the CRC Vasicek model (2.10)-(2.11), the consistent recalibration property holds if the speed of the mean reversion process \(\beta_{Z(t)}\) has strictly positive volatility. This is made precise in the following result.
2.3. Properties of consistent recalibration models

Figure 2.3.7: Simulation of CRC models. Updating $h, X, \theta$ is done using (CRC1)-(CRC3). Updating $Z$ is done using an exogenously given model for $Z$.

**Theorem 2.5** Assume that for all $z \in \mathbb{R}^p$, $\partial_z \beta \sigma_i(z) \neq 0$ holds for some $i \in \{1, \ldots, q\}$. Then the consistent recalibration property is satisfied for the CRC Vasicek model (2.10)-(2.11) with respect to the state space $\mathcal{I} = \mathbb{H}$.

The bottom line is that CRC models with stochastic parameter processes allow us to construct irreducible Markovian yield curve models on reasonably large state spaces. The irreducibility of the yield curve model is precisely the property needed to solve the recalibration problem. However, in general, irreducibility does not go along well with analytic tractability. In the next section we show that in the case of CRC models we can at least provide numerical schemes to sample the increments of the model accurately and efficiently.

2.3.3 Analytic tractability

A key property of CRC models is that their increments look infinitesimally like Hull-White extended affine processes, which means that the infinitesimal increments are well-known. Consequently, we can apply the sampling techniques for affine processes, which allow one to simulate accurately and efficiently yield curve increments.

The simulation scheme for CRC models transfers the problem of sampling state variables increments to a finite-dimensional setting. If we assume that a stochastic model for the evolution of the parameter process $Z$ is specified, then consistent recalibration models can be simulated by applying iteratively steps (CRC1)-(CRC3) in the discrete-time construction of Section 2.2. More precisely:

Given $h(0)$ and a model for the discrete-time process $(Z(t_n))_{n \in \mathbb{N}_0}$, we generate paths of the discrete-time process $(h, X, Z, \theta)$ by iteratively executing steps (CRC1)-(CRC3), for each $n \in \mathbb{N}_0$. The algorithm is illustrated in Figure 2.3.7.
stead of simulating increments from an infinite-dimensional space, it is sufficient to simulate increments of the finite-dimensional processes $X$ and $Z$. This allows one to take advantage of the existing high-order schemes for the simulation of affine processes. These advantages are thanks to the affine structure of the CRC increments and are not available for general HJM models.

In Paper A we show that the algorithm of Figure 2.3.7 is a splitting scheme for continuous-time CRC models defined by the HJM SPDE of Theorem 2.4. This observation allows one to prove convergence of discrete-time CRC models to their continuous-time counterpart as the grid size of the discrete model tends to zero. In Paper A we prove convergence in the Vasiček framework (2.10)-(2.11) using semigroup theory.
3. Hedging in bond markets with reinvestment risk

We present a computational methodology to value and hedge long-term ZCB’s trading in short- and medium-term ones. In Section 3.1 we introduce a discrete-time stochastic yield curve model with limited availability of maturity dates. As explained in Paper D the limitation in the maturity dates naturally leads to a bond market model with reinvestment risk, and there is no perfect hedging strategy available for long-term ZCB’s. For this reason our approach relies on the concept of acceptable hedging, which is introduced in Section 3.2. Acceptable hedging of long-term ZCB’s within our framework amounts to solving an infinite-dimensional portfolio optimisation problem with constraints on the terminal wealth corresponding to the acceptance criterion for hedging strategies. A numerical procedure to obtain approximate solutions is discussed in Section 3.3. The numerical solutions provide an extension of the yield curve at the long end based on hedging strategies and risk tolerances. By looking at the shift between two extensions based on different risk tolerances we also obtain spreads corresponding to shortfall risk, which cannot be completely eliminated because of the unavoidable reinvestment risks.

3.1 Bond market model with reinvestment risk

We consider a discrete-time grid with regular grid size $\delta > 0$ and final time horizon $N\delta$, i.e., according to the notation of Section 2.1 we set

$$t_n = n\delta, \quad n \in \{0, \ldots, N\}.$$  

Let $m \in \mathbb{N}$ and assume that in the market only ZCB’s with times to maturity $\delta, \ldots, m\delta$ are available. This implies that a ZCB which matures at time $(n + m)\delta$ is not available before time $n\delta$. For each $n \in \{0, \ldots, N\}$ we define the zero-coupon yields at time $n\delta$ by

$$Y(n\delta, (n + j)\delta) = -\frac{1}{j\delta} \log P(n\delta, (n + j)\delta), \quad j = 1, \ldots, m, \quad (3.1)$$
where $P(n\delta,(n+j)\delta)$ denotes the ZCB price as in Chapter 2.

Using the same notation as in Section 2.1 let $(\Omega,\mathcal{F},(\mathcal{F}(n\delta))_{n\in\{0,\ldots,N\}})$ be a filtered space. We introduce a probability measure $\mathbb{P}$ on $(\Omega,\mathcal{F})$ which plays the role of the real-world measure. Additionally, we define as in Wüthrich and Merz [82, Section 2.3.2] a risk-neutral measure $\mathbb{Q}\sim\mathbb{P}$ on $(\Omega,\mathcal{F})$ for the bank account numeraire (risk-free rollover)

$$B(n) = e^{\delta \sum_{i=1}^{n} Y((i-1)\delta,i\delta)}.$$ 

Our goal is to construct a model for the zero-coupon yields (3.1) in which newly issued ZCB’s are modelled so that reinvestment risk is introduced, i.e., it should not be possible to perfectly replicate long-term ZCB’s, see Dahl [25, 26].

Initially, we formulate our model under $\mathbb{Q}$ considering only times to maturity $\delta,\ldots,(m-1)\delta$. For each $n \in \{1,\ldots,N\}$ the maturities $(n+1)\delta,\ldots,(n+m-1)\delta$ correspond to ZCB’s which have not been newly issued, i.e., these maturities are also available in the market at time $(n-1)\delta$. We construct an HJM model for these times to maturity in the spirit of Teichmann and Wüthrich [77]. Under $\mathbb{Q}$ set for $n \in \{1,\ldots,N\}$ and $j \in \{1,\ldots,m-1\}$

$$j\delta Y(n\delta,(n+j)\delta) = (j+1)\delta Y((n-1)\delta,(n+j)\delta) - \delta Y((n-1)\delta,n\delta) + \sqrt{\delta} \sum_{k=1}^{m-1} \theta(Y((n-1)\delta,(n+j)\delta)) \Sigma_{k,j}^{1/2} \epsilon_k(n),$$

where

(i) $(\epsilon_k(n))_{k\in\{1,\ldots,m-1\}, n\in\{1,\ldots,N\}}$ are independent standard Gaussian random variables under $\mathbb{Q}$;

(ii) $\theta(\cdot)$ is a real function which determines the volatility scaling in the yields;

(iii) $(\Sigma_{k,j})_{k,j\in\{1,\ldots,m-1\}}$ are constant parameters called return directions; and

(iv) $\gamma_j(\cdot) = \delta \sum_{k=1}^{m-1} \theta(\cdot)^2 \Sigma_{k,j}$.

The terms (1)-(3) in equation (3.2) are interpreted as follows.

(1) is the no-arbitrage condition for a deterministic yield curve;

(2) is the HJM no-arbitrage drift term which ensures that discounted ZCB prices are $\mathbb{Q}$-martingales; and

(3) is the volatility part, which depends on the level of yields at the previous time point.
For each $n \in \{1, \ldots, N\}$ the bond which matures at time $(n + m)\delta$ is not traded at time $(n - 1)\delta$, and therefore the dynamics of the newly issued securities is not restricted by the HJM drift condition under $\mathbb{Q}$. For this reason we model the process $(Y(n\delta, (n + m)\delta))_{n \in \{0, \ldots, N\}}$ directly under the real-world measure $\mathbb{P}$. We couple equation (3.2) with the following equation for $n \in \{1, \ldots, N\}$

\[
(3.3) \quad m\delta Y(n\delta, (n + m)\delta) = (m - 1)\delta Y(n\delta, (n + m - 1)\delta) + \delta \varphi(Y((n - 1)\delta, \cdot)) \\
+ \sqrt{\delta} \theta(Y((n - 1)\delta, (n - 1 + m)\delta)) \kappa \eta(n),
\]

where

(i) $(\eta(n))_{n \in \{1, \ldots, N\}}$ are independent standard Gaussian random variables under $\mathbb{P}$;

(ii) $\varphi : \mathbb{R}^m \to \mathbb{R}$ is a function which describes the slope at the long end of the yield curve; and

(iii) $\kappa \in \mathbb{R}$ is a constant parameter.

The terms (3)-(4) in equation (3.3) are interpreted as follows.

(3) is a linear continuation of the yield curve at the long end; and

(4) is the volatility part, which depends on the level of yields at the previous time point.

In Paper D the properties of model (3.2)-(3.3) are discussed in details. In particular, we show that the model is incomplete, and that deterministic cash flows with time to maturity larger than $m\delta$ are not attainable. The unattainability of such cash flows reflects reinvestments risk. At this point it is important to mention that the functions $\theta$ and $\varphi$ are chosen exogenously. The model parameters $(\Sigma_{kj})_{i, j \in \{1, \ldots, m - 1\}}$ and $\kappa$ are calibrated to market data using historical simulation as explained in Paper D. Numerical results suggest that the calibrated parameters are appropriate over a long period of time. This is an essential feature for the long-term hedging problem under consideration. Moreover, the model allows to simulate yield curves efficiently, which is also very important since numerical solutions of the hedging problem will heavily rely on simulations.

### 3.2 Acceptable hedging

A trading strategy is an $m$-dimensional predictable stochastic process

\[
\xi = (\xi_j(n))_{j \in \{1, \ldots, m\}, n \in \{1, \ldots, N\}},
\]
where \( \xi_j(n) \) denotes the amount of money invested from time \((n-1)\delta \) to \( n\delta \) in the ZCB with time to maturity \( j\delta \). A self-financing trading strategy \( \xi \) provides a value process

\[
V_\xi = \{ V_\xi(n) \}_{n\in\{0,\ldots,N\}},
\]

where \( V_\xi(n) \) denotes the value of the investments in \( \xi \) at time \( n\delta \) (trading strategies are called self-financing if there are no outflows or inflows of cash over time).

A claim with maturity \( N\delta \) is an \( \mathcal{F}(N\delta) \)-measurable random variable \( c(N) \). We think of \( c(N) \) as a future (insurance) liability which has to be covered at time \( N\delta \). A claim \( c(N) \) is called attainable if there exists a self-financing trading strategy \( \xi^{c(N)} \) such that \( V_{\xi^{c(N)}}(N) = c(N) \), \( \mathbb{P} \)-a.s. Any attainable claim can be perfectly hedged by investing \( V_{\xi^{c(N)}}(0) \) at time 0 and adjusting the portfolio holdings at times \( t \in \{1,\ldots,N\} \) according to the trading strategy \( \xi^{c(N)} \).

In Section 3.1 we have introduced an interest rate model where ZCB’s with longer times to maturity become available only at future time points. In Paper D we show that this feature immediately leads to an incomplete market model. In particular, we show that the deterministic claim \( c(N) = 1 \) (ZCB with maturity \( N\delta \)) is not attainable for \( m < N \). The unattainability of a ZCB with maturity \( N\delta \) is a desired feature of the model. For instance, a life-time annuity with payments having more than 30 years to maturity cannot be hedged because, in general, there are no bonds available at the financial market with such long times to maturity. Bonds of shorter times to maturity need to be rolled over. This involves reinvestment risk because of the uncertainty in future yields. In practice, we may also have constraints on the trading strategies which can be employed to hedge the liability. We denote by \( \mathcal{D} \) the set of feasible trading strategies.

In the literature there are several approaches to hedging in incomplete market models. For instance, superhedging requires a self-finance strategy \( \xi \) to satisfy \( V_\xi(N) \geq c(N) \), \( \mathbb{P} \)-a.s. For more details on this approach in discrete time see e.g. Föllmer and Schied [36, Chapters 7 and 8]. Such a superhedging strategy is often very expensive and unreasonable in practice.

In Paper D we construct hedging strategies using a more pragmatic approach based on Artzner, Delbaen, and Koch-Medina [2], and Hilli, Koivu, and Pennanen [47]. We define acceptable hedging strategies as those which cover the claim at an acceptable level of risk. Let \( \mathcal{A}^{c(N)} \) be a set of \( \mathcal{F}(N\delta) \)-measurable random variables. \( \mathcal{A}^{c(N)} \) is called acceptance set and contains all payoffs which are within our risk tolerance in hedging \( c(N) \). Examples of acceptance sets based on Value-At-Risk or Expected-Shortfall risk measures are given in [47]. A trading strategy \( \xi \in \mathcal{D} \) is said to hedge \( c(N) \) at acceptance level \( \mathcal{A}^{c(N)} \) if \( V_\xi(N) \in \mathcal{A}^{c(N)} \), \( \mathbb{P} \)-a.s. The
initial value of $c(N)$ at acceptance level $A^{c(N)}_c$ is defined by

$$\inf_{\xi \in \mathbb{D}} \mathbb{V}_\xi(0).$$

In Paper D we present a method to solve an approximation of (3.4) numerically for the claim $c(N) = 1$, where yield curves evolve according to the model of Section 3.1. We restrict ourselves to long-only trading strategies, i.e., we consider the following convex cone of feasible strategies

$$\mathbb{D} = \{\xi \text{ is a self-financing trading strategy} \mid \xi_j(n) \geq 0, \mathbb{P}\text{-a.s., for all } n \text{ and } j\}.$$

In the following section we consider a finite-dimensional restriction of the optimisation problem (3.4), which can be solved numerically. This provides a computational methodology to price and hedge non-tradable ZCB’s with tradable ones under a certain risk tolerance and no short positions, where risk is caused by the unavoidable future reinvestments whose yields are unknown at time 0, and described by acceptance set $A^{c(N)}_c$.

### 3.3 Galerkin approximation of the hedging strategy

In general, the optimisation problem (3.4) cannot be solved analytically or using standard algorithms for convex optimisation, since the set of feasible strategies $\mathbb{D}$ is an infinite-dimensional convex cone. Numerical methods to approximate optimisation problems such as (3.4) by finite-dimensional ones have been discussed in the literature.

A possible approach is scenario discretisation of $Y(n\delta, (n+j)\delta)$, see e.g. Pflug [70], and Hilli and Pennanen [49]. The main idea is to construct a finite set of scenarios for the yield curve. Such a discretisation leads to a finite- (typically high-) dimensional set of feasible strategies and an approximation of (3.4) by a standard convex optimisation problem which can be solved with conventional algorithms. Discretisation of yield curves presents, however, serious computational drawbacks. The dimensionality of the approximation increases exponentially in the number of trading periods $N$, and the problem of hedging long-term cash flows requires approximation for relatively large values of $N$. For this reason we do not consider scenario discretisation.

Instead, we consider the **Galerkin method**, see Koivu and Pennanen [57], and Hilli, Koivu, and Pennanen [48]. This is easier to implement and leads to efficient algorithms. The idea is to search for optimal solutions not over the entire set $\mathbb{D}$, but over a subset of $\mathbb{D}$ consisting of finite linear combinations of feasible trading strategies, called **basis strategies**. A selection of finitely many basis
strategies in $\mathcal{D}$ is fixed. Then optimisation problem (3.4) is solved over the coefficients of linear combinations of the basis strategies. It is important to point out that the Galerkin approach is not only very convenient to obtain numerical solutions, but that it also reflects more realistically hedging strategies for long-term liabilities used in practice. For example, from the point of view of an insurance company, we are interested in finding an optimal asset allocation among pre-specified investment strategies to hedge long-term liabilities, see [47] and [48]. In the Galerkin approach we are doing precisely this for the chosen basis strategies.

The computational advantages of the Galerkin method allow one to find optimal allocations among the basis strategies very efficiently, see Paper D. In contrast to scenario discretisation, increasing $N$ does not increase the complexity of the optimisation directly. The dimension of the optimisation problem depends solely on the number of basis strategies. However, the complexity of evaluating the objective function typically depends on $N$. This is because in our model we are not able to characterise the distribution of $V_\xi(N)$ analytically, and therefore we have to rely on simulations, which are computationally more expensive if $N$ is large. In Paper D we present some typical choices for the basis strategies. We also provide all details on how to estimate the value function by simulation and compute the optimal allocation.

Our valuation methodology can be applied to compute the prices of non-tradable ZCB with times to maturity greater than $m\delta$. These prices are based on optimal allocation among pre-specified basis strategies which hedge the ZCB’s at a pre-specified level of risk. From these prices the corresponding ZCB yields can be calculated. In this way we extend the yield curve for times to maturity greater than $m\delta$, see Figures 3.3.1-3.3.2. Moreover, by looking at the shift between two extensions with different risk tolerances, we also obtain spreads which are related to the shortfall risk of the hedging strategies. In Paper D we discuss in detail the computation of the optimal allocation and the yield curve extension.

Note that our yield curve extension is not based on heuristic extrapolations from parametric curve fitting families such as Svensson [76]. Instead, our extension is based on acceptable hedging with liquid financial instruments, which is in accordance with generally accepted solvency principles.
3.3. **Galerkin approximation of the hedging strategy**

![Graph of yield curve](image1)

**Figure 3.3.1:** Extension of the yield curve at the long end computed from acceptable hedging strategies for $m\delta = 10$ years, $N\delta = 11, \ldots, 20$ years, and different values of the risk tolerance parameter $u_{11}$. The lower the value of $u_{11}$, the lower the risk tolerance in hedging the ZCB’s. See Paper D for more details.

![Graph of optimal weights](image2)

**Figure 3.3.2:** Optimal weights assigned to 10 basis strategies for $m\delta = 10$ years, $N\delta = 20$ years and different values of the risk tolerance parameter $u$. The basis strategies are buy & hold (BH), fixed time to maturity (FTM), fixed proportions (FP) and target date fund (TDF) trading rules. See Paper D for more details.
4. Affine models with fractional features

In Section 4.1 we consider a class of Banach-space valued Ornstein-Uhlenbeck (OU) processes. In general, linear real-valued functionals of these processes are not semimartingales. Instead they are fractional processes which are closely related to fractional Brownian motion (fBM) as explained in Section 4.2. Conditional exponential moments of these processes satisfy an infinite-dimensional version of the affine transform formula. The affine structure allows one to solve selected fractional models as shown in Sections 4.3 and 4.4. This chapter summarises the results of Paper E related to financial models driven by fractional processes.

4.1 \( L^1 \)-valued Ornstein-Uhlenbeck process

We use the notation and setup of Section 2.1 with \( d = 1 \). In particular, \( W \) is one-dimensional \( (\mathcal{F}(t))_{t \geq 0} \)-Brownian motion. Given a collection of \( \mathcal{F}(0) \)-measurable \( \mathbb{R} \)-valued random variables \( X^\beta_1(0) \) and \( X^\beta_2(0) \) indexed by \( \beta \in (0, \infty) \), we define the factor process \( X = (X(t))_{t \geq 0} \) by

\[
X(t) = (X_1(t), X_2(t)) = (X^\beta_1(t), X^\beta_2(t))_{\beta \in (0, \infty)},
\]  

where for each \( \beta \in (0, \infty) \) the process \( (X^\beta_1(t), X^\beta_2(t))_{t \geq 0} \) satisfies the SDE

\[
dX^\beta_1(t) = -\beta X^\beta_1(t) \, dt + dW(t), \\
dX^\beta_2(t) = \left( -\beta X^\beta_2(t) + X^\beta_1(t) \right) \, dt.
\]

The factor process \( X \) and the variable \( \beta \) play the same role as in the finite-dimensional setting of Section 2.1, and therefore by an abuse of notation we use the same symbols.

For each \( \beta \in (0, \infty) \) the process \( (X^\beta_1(t), X^\beta_2(t))_{t \geq 0} \) is a bivariate OU process, and the variable \( \beta \) is related to the speed of mean reversion of the process. In other words the factor process is assumed to be a collection of infinitely many bi-variate OU processes indexed by variable \( \beta \).
In Paper E we show that $X$ takes values in $L^1(\mu) \times L^1(\nu)$, where $\mu$ and $\nu$ are sigma-finite measures on $(0,\infty)$ such that
\[
\int_0^\infty (1 \wedge \beta^{-1}) \mu(d\beta) < \infty, \quad \int_0^\infty (1 \wedge \beta^{-1}) \nu(d\beta) < \infty, \tag{4.3}
\]
and such that $\nu$ has a density $p$ with respect to $\mu$ satisfying for each $t > 0$
\[
\sup_{\beta \in (0,\infty)} p(\beta) e^{-t\beta} < \infty. \tag{4.4}
\]
Additionally, an infinite-dimensional version of the affine transform formula for the conditional exponential moments of $X$ can be derived. After introducing the necessary notation, we formulate these results precisely in the theorem below.

Let $\langle \cdot, \cdot \rangle_\mu : L^1(\mu) \times L^\infty(\mu) \to \mathbb{R}$ and $\langle \cdot, \cdot \rangle_\nu : L^1(\nu) \times L^\infty(\nu) \to \mathbb{R}$ be defined by
\[
\langle f, u \rangle_\mu = \int_0^\infty f(\beta) u(\beta) \mu(d\beta), \quad (f, u) \in L^1(\mu) \times L^\infty(\mu),
\]
\[
\langle g, v \rangle_\nu = \int_0^\infty g(\beta) v(\beta) \nu(d\beta), \quad (g, v) \in L^1(\nu) \times L^\infty(\nu).
\]
The complexification of the spaces are denoted by $L^1(\mu; \mathbb{C})$, etc.

**Theorem 4.1 (Affine structure in $L^1$)** Let $\mu$ and $\nu$ satisfy conditions (4.3)-(4.4) and let $X(0) \in L^1(\mu) \times L^1(\nu)$. Then the process $X$ has a predictable $L^1(\mu) \times L^1(\nu)$-valued version and is Gaussian. Moreover, the process $X$ is affine in the sense that for each $0 \leq t \leq T$ and $(u, v) \in L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})$, the relation
\[
\mathbb{E}^Q \left[ e^{\langle X(t), u \rangle_\mu + \langle X(T), v \rangle_\nu} \mathcal{F}(t) \right] = e^{\phi(T-t, u, v) + \sum_{1}^{2} \langle X(t), \psi_j(T-t, u, v) \rangle_{\mu} + \langle X(T), \psi_j(T-t, u, v) \rangle_{\nu}},
\]
holds with probability one, where the functions
\[
(\phi, \psi_1, \psi_2) : [0,\infty) \times L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C}) \to \mathbb{C} \times L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})
\]
are given by
\[
\phi(t, u, v) = \frac{1}{2} \int_0^t \left( \int_0^\infty \psi_1(s, u, v)(\beta) \mu(d\beta) \right)^2 ds,
\]
\[
\psi_1(t, u, v)(\beta) = e^{-t\beta} (u(\beta) + \tau v(\beta) p(\beta)), \tag{4.5}
\]
\[
\psi_2(t, u, v)(\beta) = e^{-t\beta} v(\beta).
\]
In Paper E we also show that the coefficient functions $(\phi, \psi_1, \psi_2)$ can be characterised as solutions of infinite-dimensional Riccati equations. The theorem above extends some of the results in Carmona and Coutin [14] and is useful for
several reasons. As we explain below, thanks to the affine structure we are able to construct tractable financial models with fractional features. Moreover, these models are driven by a Markov process on an infinite-dimensional state space. A Markovian representation allows one to identify the states of the model. This makes it easier to speak about calibration in a sensible way and to compare the model across time.

In general, for \((u, v) \in L^\infty(\mu) \times L^\infty(\nu)\) the \(\mathbb{R}\)-valued processes \(\langle X_1, u \rangle_\mu\) and \(\langle X_2, v \rangle_\nu\) are not semimartingales. The next result provides sufficient conditions on \(u\) and \(v\) to obtain semimartingales. As this will be needed later in applications, we consider time-dependent functions \(u = u(\beta, t)\) and \(v = v(\beta, t)\).

**Theorem 4.2 (Semimartingale property)** Let \(\mu\) and \(\nu\) satisfy conditions (4.3)-(4.4) and let \(X(0) \in L^1(\mu) \times L^1(\nu)\). Let \(u(\beta, t)\) and \(v(\beta, t)\) be real-valued, deterministic, jointly measurable in \((\beta, t) \in (0, \infty) \times [0, \infty),\) differentiable in \(t\) and satisfying

\[
\forall t \geq 0: \quad \|u(\cdot, t)\|_{L^\infty(\mu)} < \infty \quad \text{and} \quad \|v(\cdot, t)\|_{L^\infty(\nu)} < \infty.
\]

Assume \(X(0) \in L^1(\mu) \times L^1(\nu),\) a.s., and for each \(t \geq 0\)

\[
\begin{align*}
\int_0^\infty \int_0^t \left| \partial_s u(\beta, s) - \beta u(\beta, s) \right| (1 + \beta^{-1/2}) d s \mu(d\beta) &< \infty, \quad (4.6) \\
\int_0^\infty \sqrt{\int_0^t u(\beta, s)^2 d s} \mu(d\beta) &< \infty, \quad (4.7) \\
\int_0^\infty \int_0^t \left| \partial_s v(\beta, s) - \beta v(\beta, s) \right| (1 + \beta^{-1/2}) d s \nu(d\beta) &< \infty, \quad (4.8) \\
\int_0^\infty \int_0^t |v(\beta, s)| (1 + \beta^{-1/2}) d s \nu(d\beta) &< \infty. \quad (4.9)
\end{align*}
\]

Then the processes \(\langle X_1(t), u(\cdot, t) \rangle_\mu\) \(\forall t \geq 0\) and \(\langle X_2(t), v(\cdot, t) \rangle_\nu\) \(\forall t \geq 0\) are semimartingales with decompositions

\[
\begin{align*}
\langle X_1(t), u(\cdot, t) \rangle_\mu &= \langle X_1(0), u(\cdot, 0) \rangle_\mu + \int_0^t \int_0^\infty \left( \partial_s u(\beta, s) - \beta u(x, s) \right) X_1^\beta(s) \mu(d\beta) d s \\
&\quad + \int_0^t \int_0^\infty u(\beta, s) \mu(d\beta) d W(s), \\
\langle X_2(t), v(\cdot, t) \rangle_\nu &= \langle X_2(0), v(\cdot, 0) \rangle_\nu + \int_0^t \int_0^\infty \left( \partial_s v(\beta, s) - \beta v(x, s) \right) X_2^\beta(s) \nu(d\beta) d s \\
&\quad + \int_0^t \int_0^\infty v(\beta, s) X_2^\beta(s) \nu(d\beta) d s.
\end{align*}
\]

As we stated in Theorem 4.1 one can verify that \(X\) is a Gaussian process in \(L^1(\mu) \times L^1(\nu)\). However, the stationary distribution of \(X\) is in general not a Gaussian distribution on \(L^1(\mu) \times L^1(\nu)\), but only on a larger space \(L^1(\mu_\infty) \times L^1(\nu_\infty)\)
corresponding to sigma-finite measures \( \mu_\infty \) and \( \nu_\infty \) on \((0, \infty)\) satisfying the following stronger integrability conditions

\[
\int_0^\infty \beta^{-1/2} \mu_\infty(d\beta) < \infty, \quad \int_0^\infty \beta^{-3/2} \nu_\infty(d\beta) < \infty, \tag{4.11}
\]

and for each \( t > 0 \)

\[
\sup_{\beta \in (0, \infty)} p_\infty(\beta) e^{-t\beta} < \infty, \tag{4.12}
\]

where \( \mu_\infty \) has density \( p_\infty \) with respect to \( \nu_\infty \). For any initial value \( X(0) \in L^1(\mu_\infty) \times L^1(\nu_\infty) \) and any \( t \geq 0 \), we consider \( X(t) \) as a random variable with values in the space \( L^1(\mu_\infty) \times L^1(\nu_\infty) \), which we endow with the weak topology.

In the following it is convenient to define \( W(t) \) for \( t \in \mathbb{R} \) by a two-sided Brownian motion, i.e.

\[
W(t) = W_1(t)1_{t \geq 0} + W_2(-t)1_{t < 0}, \tag{4.13}
\]

where \( (W_1(t))_{t \geq 0} \) and \( (W_2(t))_{t \geq 0} \) are independent Brownian motions.

**Theorem 4.3 (Stationary distribution)** Let \( X(\infty) = (X_1^\beta(\infty), X_2^\beta(\infty))_{\beta > 0} \) be the random variable defined by

\[
X_1^\beta(\infty) = \int_{-\infty}^0 e^{s\beta} dW(s), \tag{4.14}
\]

\[
X_2^\beta(\infty) = -\int_{-\infty}^0 se^{s\beta} dW(s).
\]

Then the random variable \( X(\infty) \) has a Gaussian distribution on \( L^1(\mu_\infty) \times L^1(\nu_\infty) \) and its distribution is stationary in the sense that \( X(t) \) is equal in distribution to \( X(\infty) \) if \( X(0) \) is equal in distribution to \( X(\infty) \). Moreover, \( X(t) \) converges in distribution to \( X(\infty) \) as \( t \to \infty \).

In the following the stationary distribution is used as randomised initial value in a Markovian representation of fBM.

### 4.2 Connection to fractional Brownian motion

In this section we present a Markovian representation of fBM in terms of the process \( X \). We define fBM \( W^H = (W^H(t))_{t \geq 0} \) with initial value \( w^H \in \mathbb{R} \) by the representation of Mandelbrot and Van Ness [61], that is,

\[
W^H(t) = w^H + \frac{1}{\Gamma(H + 1/2)} \int_{-\infty}^t (t-s)^{H-1/2} (-s)^{H-1/2} dW(s), \tag{4.15}
\]
where $H \in (0, 1)$, $W$ is two-sided Brownian motion as defined in (4.13) and $\Gamma$ denotes the Gamma function. For $H \in (0, 1/2)$ and $H \in (1/2, 1)$ the process $W^H$ exhibits short- and long-range dependence, respectively. For $H = 1/2$ the process $W^H$ is in fact a Brownian motion. The parameter $H$ is called Hurst index.

In the following theorem we represent fBM (4.15) as a functional of the process $X$. The representation is obtained replacing the integrand in equation (4.15) by its Laplace transform and applying the stochastic Fubini’s theorem, see e.g. Veraar [80]. This idea goes back to Carmona and Coutin [14], where a Markovian representation of the fractional integral

$$\int_0^t (t-s)^{H-\frac{1}{2}} dW(s)$$

is derived in the case $H \in (0, 1/2)$. Extensions and numerical approximations of this representation can be found in Carmona, Coutin, and Montseny [15], and Muravlev [64].

**Theorem 4.4 (Markovian representation of fBM)** Let $X$ be the process defined in (4.1)-(4.2) with initial value $X(0)$ equal to the random variable $X(\infty)$ defined in (4.14). Furthermore, let $\mu$ and $\nu$ be measures on $(0, \infty)$ given by

$$\mu(d\beta) = \frac{d\beta}{\beta^{1/2} \Gamma(1/2 + H) \Gamma(1/2 - H)}; \quad \nu(d\beta) = \frac{d\beta}{\beta^{-1/2} \Gamma(1/2 + H) \Gamma(1/2 - H)}.$$

Then fBM (4.15) has the following representation

$$W^H_t = \begin{cases} w^H_0 + \int_0^\infty \left( X^\beta_1(t) - X^\beta_1(0) \right) \mu(d\beta), & \text{if } H \in (0, 1/2), \\
 w^H_0 + \int_0^\infty \left( X^\beta_2(t) - X^\beta_2(0) \right) \nu(d\beta), & \text{if } H \in (1/2, 1), \end{cases}$$

where $X - X(0)$ is a continuous process in $L^1(\mu) \times L^1(\nu)$.

The measures $\mu$ and $\nu$ in Theorem 4.4 satisfy conditions (4.3)-(4.4) but not conditions (4.11)-(4.12). It follows from Theorem 4.3 that $X$ takes values in $L^1(\mu_\infty) \times L^1(\nu_\infty)$ where $\mu_\infty$ and $\nu_\infty$ are given by

$$\mu_\infty(d\beta) = (1 \wedge \beta^{1/2}) \mu(d\beta), \quad \nu_\infty(d\beta) = (1 \wedge \beta^{-1/2}) \nu(d\beta).$$

Nevertheless, we show in Paper E that $X - X(0)$ takes values in $L^1(\mu) \times L^1(\nu)$.

The representation in Theorem 4.4 lends itself to numerical implementation. Indeed, the integrals can be approximated by finite sums as described in Carmona, Coutin, and Montseny [15].
4.3 Fractional short rate model

In this section we construct a short rate model where the short rate process has statistical properties similar to fBM. Such a model is interesting for applications since short-term interest rates exhibit long-range dependence, see Backus and Zin [3].

Let $\mu$ and $\nu$ be sigma-finite measures on $(0, \infty)$ satisfying a slightly strengthened version of conditions (4.3)-(4.4) as explained in Paper E. We define the short rate and bank account processes in an analogous way as in the finite-dimensional setting of Section 2.1

\[
r(t) = \ell + \langle X_1(t), \lambda_1 \rangle \mu + \langle X_2(t), \lambda_2 \rangle \nu, \quad B(t) = e^{\int_0^t r(s)ds},
\]

where $\ell \in \mathbb{R}$ is a shift parameter and $(\lambda_1, \lambda_2) \in L^\infty(\mu) \times L^\infty(\nu)$ are scaling functions. Observe that $\ell$, $\lambda_1$, $\lambda_2$, $X$, $\langle \cdot, \cdot \rangle_\mu$ and $\langle \cdot, \cdot \rangle_\nu$ play exactly the same role as in Section 2.1, and therefore by an abuse of notation we use the same symbols.

Since we are interested in models for the short rate process with long-range dependence we typically choose $\lambda_1 = 0$. For example, if we set $\mu(dx) \propto x^{-H}dx$, $\nu(dx) \propto x^{\frac{1}{2}-H}dx$, $H \in (\frac{1}{2}, 1)$ and $(\lambda_1, \lambda_2) = (0, 1)$, then $r$ is not a semimartingale and has the same roughness as fBM with Hurst index $H \in (\frac{1}{2}, 1)$. While the short rate process may take negative values, the probability of yields becoming negative can be reduced by shifting the parameter $\ell$ and scaling the parameters $\lambda_1$ and $\lambda_2$.

As in Section 2.1 we define the ZCB price by

\[
P(t, T) = \mathbb{E}^Q \left[ \frac{B(t)}{B(T)} \bigg| \mathcal{F}(t) \right], \quad 0 \leq t \leq T,
\]

and the (instantaneous) forward rate by

\[
h(t)(\tau) = -\partial_\tau \log \left( P(t, t+\tau) \right), \quad t, \tau \geq 0.
\]

For each $T > 0$ the process $B(\cdot)^{-1}P(\cdot, T)$ is by definition a $Q$-martingale. This means that $Q$ is an equivalent martingale measure by construction, and that the model is free of arbitrage. Our construction leads to the following affine structure for ZCB prices.

**Theorem 4.5 (ZCB prices and forward rates)** In the fractional short rate model (4.16), ZCB prices and forward rates are given by

\[
P(t, T) = e^{-\ell(T-t)+\Phi(T-t,\lambda_1,\lambda_2)+\langle X_1(t), \Psi_1(T-t,\lambda_1,\lambda_2) \rangle_\mu+\langle X_2(t), \Psi_2(T-t,\lambda_1,\lambda_2) \rangle_\nu},
\]

\[
h(t)(\tau) = \ell - \partial_\tau \Phi(\tau, \lambda_1, \lambda_2) - \langle X_1(t), \partial_\tau \Psi_1(\tau, \lambda_1, \lambda_2) \rangle_\mu - \langle X_2(t), \partial_\tau \Psi_2(\tau, \lambda_1, \lambda_2) \rangle_\nu,
\]

46
where for each $\tau \geq 0$ and $\beta \in (0, \infty)$

$$
\Phi(\tau, \lambda_1, \lambda_2) = \frac{1}{2} \int_0^\tau \langle \Psi_1(s, \lambda_1, \lambda_2), 1 \rangle \mu ds,
$$

$$
\Psi_1(\tau, \lambda_1, \lambda_2)(\beta) = e^{-\tau \beta} - 1 + \left( \frac{e^{-\tau \beta} - 1}{\beta^2} + \frac{\tau}{\beta} e^{-\tau \beta} \right) p(\beta) \lambda_2(\beta),
$$

$$
\Psi_2(\tau, \lambda_1, \lambda_2)(\beta) = \frac{e^{-\tau \beta} - 1}{\beta} \lambda_2(\beta).
$$

The coefficient functions $(\Phi, \Psi_1, \Psi_2)$ play the same role as in Theorem 2.1, and therefore we use the same symbols. In analogy to the finite-dimensional setting of Section 2.1 the coefficient functions $(\Phi, \Psi_1, \Psi_2)$ can be characterised as solutions of infinite-dimensional Riccati equations, see Paper E for details.

In the fractional short rate model (4.16), it turns out that ZCB prices and forward rates are semimartingales. This can be verified checking conditions (4.6)-(4.9) in Theorem 4.2. Since forward rates are semimartingales, we can provide the corresponding HJM equation.

**Theorem 4.6 (HJM equation)** In the fractional short rate model (4.16) ZCB prices $(P(t, T))_{0 \leq t \leq T}$ and forward rates $(h(t)(\tau))_{t \geq 0}$ are semimartingales for each fixed $T, \tau > 0$. The forward rate process $h = (h(t)(\tau))_{t \geq 0}$ is a solution of the HJM equation

$$
dh(t) = \left( A h(t) + \mu^{HJM} \right) dt + \sigma^{HJM} dW_t, \quad (4.17)
$$

where $A$ denotes differentiation with respect to time to maturity $\tau$ and $\mu^{HJM}, \sigma^{HJM}$ are measurable functions on $(0, \infty)$ given by

$$
\mu^{HJM}(\tau) = \partial_\tau^2 \Phi(\tau, \lambda_1, \lambda_2), \quad \sigma^{HJM}(\tau) = -\langle \partial_\tau \Psi_1(\tau, \lambda_1, \lambda_2), 1 \rangle \mu.
$$

The coefficients $\mu^{HJM}$ and $\sigma^{HJM}$ play the same role as in Theorem 2.4 and satisfy the HJM drift condition (2.7). From HJM equation (4.17) it follows for each $\tau_1, \tau_2 > 0$

$$
d[h(\tau_1), h(\tau_2)](t) = \langle \partial_t \Phi_1(\tau_1, \lambda_1, \lambda_2), 1 \rangle \mu \langle \partial_t \Phi_1(\tau_2, \lambda_1, \lambda_2), 1 \rangle \mu dt.
$$

Moreover, it follows that for each $T \geq 0$ the stochastic exponential of the process $
\int_0^T \langle \Psi_1(T - s, \lambda_1, \lambda_2), 1 \rangle \mu dW(s)$ defines a $T$-forward density process, see Filipović [34, Section 7.1]. This allows one to work out explicit formulas for interest rates caps and floors as in the standard HJM framework.

In this model the short rate process is a superposition of infinitely many OU processes. The construction allows one to consider short rate processes of the same roughness as fBM. The model has a Markovian representation which
allows to clearly identify the states of the model. Even though the short rate process is not a semimartingale, ZCB prices and forward rates are. Explicit formulas for ZCB prices, forward rates, and interest rate caps and floors can be derived. Our considerations provide two ways of identifying model parameters: either, they could be calibrated to interest rate caps and floors, or they could be estimated from realised covariations of forward rates (c.f. Section 2.3.1).

4.4 Fractional Stein & Stein model

We consider the affine stochastic volatility model introduced by Stein and Stein [75] and extend it to fractional volatility. This extension is natural in the sense that the single OU volatility process in [75] is replaced by a superposition of infinitely many OU processes. We are interested in constructing a volatility process having the same roughness as fBM with Hurst index $H \in (0, \frac{1}{2})$ in accordance with the empirical analysis in Gatheral, Jaisson, and Rosenbaum [37]. In Paper E we show that the construction leads to a fractional volatility model which preserves the affine structure. In the following we summarise this result.

Let $\tilde{W}$ be one-dimensional $(\mathcal{F}(t))_{t \geq 0}$-Brownian motion with correlation coefficient $\langle W, \tilde{W} \rangle(t) = \rho d t$ for given $\rho \in (-1, 1)$. We fix a sigma-finite measure $\mu$ on $(0, \infty)$ satisfying condition (4.3), a function $\lambda \in L^\infty(\mu)$ and an initial value $X_1(0) \in L^1(\mu)$ for the process $X_1$ defined in Section 4.1. Given these model parameters, the price process $S = (S(t))_{t \geq 0}$ is defined by the SDE

$$dS(t) = S(t) \langle X_1(t), \lambda \rangle \mu d\tilde{W}_t.$$ 

In other words the volatility process in the original model [75] is replaced by the process $\langle X_1, \lambda \rangle \mu$. For example, if we set $\mu \propto x^{-H}$, $H \in (0, \frac{1}{2})$ and $\lambda = 1$, then the volatility process has the same roughness as fBM with Hurst index $H \in (0, \frac{1}{2})$.

We introduce the following spaces of simple symmetric tensors, let

$$L^1(\mu) \otimes_s L^1(\mu) = \{y \otimes_2 : y \in L^1(\mu)\}, \quad L^\infty(\mu) \otimes_s L^\infty(\mu) = \{v \otimes_2 : v \in L^\infty(\mu)\},$$

where for any $f : (0, \infty) \rightarrow \mathbb{R}$ the function $f \otimes_2 : (0, \infty)^2 \rightarrow \mathbb{R}$ is given by $(\beta_1, \beta_2) \mapsto f(\beta_1)f(\beta_2)$. Algebraic tensor products are very convenient to bring the affine structure of the model to light. For each $t \geq 0$ we set

$$\Pi(t) = X_1(t) \otimes_2 \in L^1(\mu) \otimes_s L^1(\mu).$$

Then we have the relation

$$\langle X_1(t), \lambda \rangle^2_\mu = \langle X_1(t) \otimes_2, \lambda \otimes_2 \rangle_{\mu \otimes_2}.$$
where $\mu^{\otimes_2}$ denotes the product measure on $(0, \infty)^2$. It follows that the log-price process $L = \log(S)$ satisfies

$$\,dL(t) = -\frac{1}{2} \langle \Pi(t), \lambda^{\otimes_2} \rangle_{\mu^{\otimes_2}} \,dt + \sqrt{\langle \Pi(t), \lambda^{\otimes_2} \rangle_{\mu^{\otimes_2}}} \,d\tilde{W}_t.$$ 

The process $\Pi = (\Pi(t))_{t \geq 0}$ can be characterised as an affine process on $L^1(\mu) \otimes_s L^1(\mu)$, and an affine transform formula can be calculated explicitly.

**Theorem 4.7 (Affine structure of $\Pi$)** Let $\nu^{\otimes_2} \in iL^\infty(\mu) \otimes_s L^\infty(\mu)$. Then with probability one

$$\mathbb{E}^Q \left[ e^{\langle \Pi(T), \nu^{\otimes_2} \rangle_{\mu^{\otimes_2}}} \bigg| \mathcal{F}_t \right] = e^{\chi(T-t, \nu^{\otimes_2}) + \langle \Pi(t), \zeta(T-t, \nu^{\otimes_2}) \rangle_{\mu^{\otimes_2}}}, \quad 0 \leq t \leq T,$$

where $\chi(t, \nu^{\otimes_2}) \in \mathbb{C}$ and $\zeta(t, \nu^{\otimes_2}) \in L^\infty(\mu; \mathbb{C}) \otimes_s L^\infty(\mu; \mathbb{C})$ are given in terms of the coefficient functions $(4.5)$ by

$$\chi(t, \nu^{\otimes_2}) = -\frac{1}{2} \log \left( 1 - 4\phi(t, \nu, 0) \right),$$

$$\zeta(t, \nu^{\otimes_2}) = \psi_1(t, \nu, 0)^{\otimes_2} \frac{1 - 4\phi(t, \nu, 0)}{1 - 4\phi(t, \nu, 0)}.$$

In Paper E we also show that the coefficient functions $(\chi, \zeta)$ can be characterised as solutions of infinite-dimensional Riccati equations.

For applications we are interested in understanding the process $L$. In particular, a characterisation of its distribution and numerical schemes for simulation are needed. Our main contribution in this direction is a characterisation of $(L, \Pi)$ as an affine process taking values in $\mathbb{R} \times L^1(\mu) \otimes_s L^1(\mu)$. However, in this case, no explicit affine transform formula is available.

**Theorem 4.8 (Affine structure of $(L, \Pi)$)** Let $(L(0), \Pi(0)) \in \mathbb{R} \times L^1(\mu) \otimes_s L^1(\mu)$. Then the process $(L, \Pi)$ is an affine process in the sense that for each $0 \leq t \leq T$, $u \in \mathbb{C}$, and $\nu^{\otimes_2} \in iL^\infty(\mu) \otimes_s L^\infty(\mu)$, the logarithmic conditional characteristic function

$$\log \mathbb{E}^Q \left[ e^{L(T)u + \langle \Pi(T), \nu^{\otimes_2} \rangle_{\mu^{\otimes_2}}} \bigg| \mathcal{F}_t \right],$$

is affine in $(L(t), \Pi(t))$.

The proof of the theorem above is based on an approximation of $\langle X_1, \nu \rangle_{\mu}$ which goes back to Carmona, Coutin, and Montseny [15]. This leads to an approximation of $(L, \Pi)$ by finite-dimensional affine models, which provides numerical schemes for simulation of the fractional Stein and Stein model.
5. Conclusions

In Chapter 2 we have introduced a tractable and flexible extension of affine factor models which allows parameters to follow stochastic processes. The additional flexibility of stochastic parameters leads to better fits of the yield curve dynamics to real data. Moreover, turning the (otherwise) static parameters into stochastic processes allows one to recalibrate the model consistently over time to new market information. For this reason the extended models are called consistent recalibration (CRC) models. Infinitesimal increments of CRC models are affine, and therefore the models remain tractable. In particular, yield curves can be simulated efficiently, and it is feasible to compute several model quantities of interest in applications. CRC models are selected from data in accordance with the robust calibration principle, i.e., historical parameters, market prices of risk and Hull-White extensions are inferred from time series data and prevailing market prices using a combination of statistical methods and calibration.

In Chapter 3 we have developed an efficient algorithm to extrapolate the yield curve at the long end by explicitly considering hedging strategies. The method is based on a yield curve model with reinvestment risk which is calibrated to market data. The hedging strategy is constructed by computing the optimal allocation among pre-specified trading strategies according to a certain risk tolerance. The lower the risk tolerance the lower the spread between the extrapolated yields and market yields. Thus, we obtain a natural extrapolation of the yield curve beyond the last liquid time to maturity date. Moreover, we also obtain explicitly the hedging strategies which correspond to the extrapolation.

In Chapter 4 we have constructed financial models driven by fractional processes which can be represented as linear functionals of an infinite-dimensional affine process. We have shown that these processes are closely related to fractional Brownian motion. The Markovian representation of the model clearly defines model states and leads to numerical schemes for simulation. The affine structure is exploited to obtain explicit formulas for several model quantities.
Philipp Harms, David Stefanovits, Josef Teichmann, Mario V. Wüthrich.

Consistent recalibration of yield curve models.

CONSISTENT RECALIBRATION OF YIELD CURVE MODELS

PHILIPP HARMS, DAVID STEFANOVITS, JOSEF TEICHMANN, AND MARIO WÜTHRICH

Abstract. The analytical tractability of affine (short rate) models, such as the Vasiček and the Cox-Ingersoll-Ross models, has made them a popular choice for modelling the dynamics of interest rates. However, in order to account properly for the dynamics of real data, these models need to exhibit time-dependent, or even stochastic, parameters. This in turn breaks their tractability, and modelling and simulating becomes an arduous task. We introduce a new class of Heath-Jarrow-Morton (HJM) models, that both fit the dynamics of real market data and remain tractable. We call these models consistent recalibration (CRC) models. These CRC models appear as limits of concatenations of forward rate increments, each belonging to a Hull-White extended affine factor model with possibly different parameters. That is, we construct HJM models from “tangent” affine models. We develop a theory for a continuous path version of such models, and discuss their numerical implementations within the Vasiček and Cox-Ingersoll-Ross frameworks.

1. Introduction

Modelling the stochastic evolution of yield curves is an important task in risk management, forecasting, decision making, pricing and hedging. We emphasise here three principles of modelling yield curves (or any other traded instrument in finance): we certainly require all models for traded assets’ prices to be free of arbitrage, therefore we do not state this as a principal requirement.

- **Robust calibration:** the model is selected simultaneously from time series and prevailing market prices, as explained in [22]. Model parameters which are invariant under equivalent measure changes should be estimated by a statistical procedure from time series data. The remaining parameters are calibrated by solving an inverse problem with respect to the prevailing market prices. All model parameters should be constant during the life time of the model; only state variables may change.

- **Consistency:** an interest rate model is called consistent if the stochastic process of yield curves does not leave a pre-specified set $I$ of possible market observables (in [32] the set $I$ is assumed to be a finite dimensional submanifold of curves corresponding to a curve fitting method). Here, we add the following requirement: the yield curve process should – loosely speaking – be able to reach any neighbourhood of any yield curve in $I$ with positive probability because any newly arriving market configuration is a possible model state. Consequently, the model can be recalibrated to a new market

2010 Mathematics Subject Classification. 91G30, 60J25, 60J60,
Supported in part by SNF grant 149879.
We gratefully acknowledge support by ETH Foundation.
configuration without losing consistency to the model with old parameters; we say that the model satisfies the consistent recalibration property.

- **Analytic tractability:** relevant quantities of a model can be calculated quickly and accurately. In particular, one should be able to simulate state variable increments efficiently. This can be a delicate problem in the presence of boundary conditions.

**Remark.**

- By a model for the term structure of interest rates, we understand a fully specified stochastic process taking values in the pre-specified set of yield curves $I$. We shall always consider a parametrised class of models consisting of one fully specified model for each initial state in $I$ and each parameter value.

- In practice, interest rate models are recalibrated on a regular basis (e.g. daily) to market data. Suppose that the consistent recalibration property does not hold for today’s model. Then tomorrow’s market yield curve might lie outside of the set of possible realisations. If this happens, then tomorrow’s recalibration necessarily implicates a rejection of today’s model. On the other hand, no inconsistencies occur if recalibration is an update of the state variables of the model and does not involve a change of model parameters.

- We are able to fully calibrate our models under the pricing measure without specifying the market price of risk. The robust calibration principle separates the easier task of estimating volatilities from the more difficult task of estimating drifts. Moreover, it tells exactly which parameters may be estimated from time series (namely those which are invariant under equivalent changes of measure). In future work, our results will form the basis for modelling and filtering the market price of risk.

Usually it is difficult to find models for yield curve evolutions satisfying all three principal requirements with respect to sets $I$ which are sufficiently large to be of practical use (think of open subsets of a Hilbert space of curves). Mathematically speaking, we look for irreducible Markovian models, which are in addition analytically tractable. Irreducibility is understood in the sense that any strict subset of $I$ is not invariant under the model. Often irreducibility does not accord with analytic tractability beyond elliptic models, which are too restrictive in infinite dimension.

Several in part overlapping approaches to yield curve modelling have been developed. The models can be roughly categorised as factor, HJM, principal component analysis (PCA), and filtered historical simulation models. We briefly analyse these models with respect to our requirements and then present our new class of models, which we call consistent recalibration models (CRC models).

1.1. **Factor models.** Factor models are based on a factor process, which usually describes certain market factors, from which – by means of basic principles – the entire yield curve can be derived (see [33] for an overview). Let $X = (X_t)_{t \geq 0}$ be a factor Markov process acting on a finite dimensional state space and depending on a parameter vector $y$, and let $B := g(X)$ be the bank account process, for some positive functional $g$. Then one obtains – with respect to the pricing measure $\mathbb{P}$ – the relation

$$P(t, T) = \mathbb{E} \left[ \frac{B_T}{B_t} \bigg| \mathcal{F}_t \right] = G(t, T, X_t)$$
for some function $G$ also depending on the parameter vector $y$.

Market data arrive in the form of daily yield curves. By means of calibration the initial state $X_0$ and a parameter vector $y_0$ are chosen to explain today’s market data. By choosing the parameter vector rich enough, one receives good fits to today’s market data. Apparently a recalibration at time $t = 1$ can (and will) lead to another state $X_1$ and another parameter vector $y_1$. As states may vary stochastically, the change of $X_0$ to $X_1$ is not a problem, but the change of parameters is. This means that one has to decide at time $t = 1$ whether to continue with the model specified by $y_0$ or whether to switch to the model specified by $y_1$. This problem can be alleviated to some extent by using a combination of filtering and calibration techniques to stabilise the choice of $y$, as described in, e.g., [38, 4]. Nevertheless, robust calibration remains an unresolved issue.

The consistent recalibration property never holds unless the set $I$ is very small. (If $I$ is a sub-manifold, its dimension cannot be larger than that of the factor process.) However, on the positive side, factor models are often analytically tractable, for instance within the affine class (see e.g. [33, 30]).

### 1.2. HJM-models.

Markovian HJM models are an extreme version of factor models: the yield curve itself is taken as state variable (possibly together with some hidden state variables). By this choice the state space is no longer finite dimensional. A potentially difficult drift condition is necessary to guarantee absence of arbitrage (see [44]). Calibration to daily arriving yield curves is now a matter of statistical estimation from the time series of market data. An appropriate parametrisation of instantaneous co-variance, jump structure and drifts will lead to a statistical inference problem, an infinite dimensional though. Hence the paradigm of robust calibration, including the requirement of calibration through estimation, is fulfilled in the optimal sense. If the Markov process acting on yield curves is “irreducible”, i.e. every neighbourhood of a state can be reached with positive probability, then even consistent recalibration is possible (see [5]). However, one usually encounters severe lack of analytic tractability within this model class: even if one can simulate the process (which often requires fairly strong assumptions on the vector fields involved, see [29]), there is not much hope to quickly solve more complicated pricing and hedging questions in general.

### 1.3. PCA- or local PCA models.

Principal component analysis (PCA) or local PCA consider yield curves as outcomes of a statistical model, which is estimated by standard PCA techniques (see e.g. [60, 9, 19]). When the statistical model is too simplistic, often arbitrage enters the field, which is an undesirable feature. A more refined version is actually equivalent to an HJM model with constant vector fields (as e.g. in [45]). Here preserving positive interest rates, which is desired in some situations, is not possible. PCA inspired models, correctly implemented, allow for robust calibration and consistent recalibration, but are usually not very tractable from an analytic point of view.

### 1.4. Filtered (historical) simulation.

Historical simulation is a standard industry technique to simulate distributions of yield curves by considering the relative returns as independent samples of an unknown distribution, see [38, 4]. Certainly this assumption can lead to difficulties with absence of arbitrage, but this can be solved as in [68, 77]. The most important problem is the state-independence of the distribution. Again also (filtered) historical simulation can be embedded into the
realm of HJM models. These models then allow for robust calibration and consistent recalibration, but are usually not very tractable from an analytic point of view.

1.5. Consistent recalibration models. The goal of this work is to present a new model class, which combines the advantages of factor and HJM models: consistent recalibration with respect to a rich set of yield curves $I$, robust calibration, and analytic tractability. We call such a model class by its most distinguished property consistent recalibration models (CRC models).

The basic idea can be explained easily: there are well-known Markovian HJM models which can be built from affine factor models for the short rate process. They are Hull-White extensions (in the sense of interest rate theory) of the given factor models. From [35] it is known that these are the only model classes where each single model within the class admits a finite-dimensional realisation. (This is proven in [35, Theorem 5.3] under some mild regularity assumptions.) The existence of finite dimensional realisations is exactly what makes these models tractable, but it also means that there is no hope for irreducibility on any reasonably large set of yield curves $I$.

We may ask ourselves how a (minimal) enlargement of such model classes could look like to allow for consistent recalibration without losing analytic tractability. Necessarily the increments of the yield curve process should be tractable. This leads to the principle of concatenating yield curve increments of factor models while allowing for random parameter changes. This construction is depicted in Figure 3.2.

In the language of term structure equations, see [72, 16, 55], this amounts to considering models tangent to affine models. Making the parameters of the affine process stochastic (in an independent or dependent way) will lead generically towards consistent recalibration, since the conditions for finite dimensional realisations are not fulfilled anymore. Robust calibration is preserved since we are still working with HJM models. What is less clear is what happens to the analytic tractability: here the observation helps that these newly introduced HJM models look infinitesimally like Hull-White extended affine processes, which means that the Fourier transform of the infinitesimal increments of this HJM model is known and that all sampling techniques of affine processes apply. That is, we can precisely sample the increments of such SPDE’s (see e.g. [1]). This is what we shall call analytically tractable at this level.

The principle behind CRC models applies also to the modelling of more general term structure dynamics. For example, it could be applied to multi-curve interest rate models and to models of the term structure of option prices.

1.6. Example. We illustrate this with an example. Consider the Hull-White extended Cox-Ingersoll Ross (CIR) model. In this model the short rate is given by the SDE

$$dr(t) = \left(\theta(t) + \beta r(t)\right)dt + \sqrt{\alpha r(t)}dW(t),$$

where $\theta(t) \geq 0$ determines the time-dependent level of mean reversion, $\beta < 0$ the speed of mean reversion, and $\alpha > 0$ the level of volatility.

The model can be calibrated to any initial yield curve from a large subset $I$ of curves by choosing an appropriate Hull-White extension $\theta$. For any fixed initial yield curve, the distribution of yield curves at some future $t > 0$ is concentrated at a one-dimensional affine subspace of curves. Therefore, market observations are generally not in the support of the model and the consistent recalibration property
PAPER A

does not hold with respect to $I$. The low dimensionality of the model is also apparent at the level of realised covariations of yields for different times to maturity. In the CIR model, the matrix of covariations has rank one, which is in stark contrast to observations from the market (see Figure 6.16). Finally, calibrated model parameters vary significantly over time as shown in Figures 6.5 and 6.7, which contradicts the requirement of robust calibration.

As a remedy, one could make some model parameters stochastic and include them as state variables. For example, one could make $\alpha = \alpha_y$ and $\beta = \beta_y$ depend on a parameter $y$ and write dynamics of the form

$$dr(t) = \left(\theta(t) + \beta Y(t) r(t)\right)dt + \sqrt{\alpha Y(t)r(t)}dW(t),$$
$$dY(t) = \mu(Y(t))dt + \sigma(Y(t))d\tilde{W}(t).$$

Unfortunately, this usually breaks the analytic tractability of the model.

The key idea of CRC models is to lift the short rate model to a HJM model and to introduce stochastic parameters on that level. Let $h(t)$ denote the forward rate curve at time $t$ in Musiela parametrisation (i.e., as a function of time to maturity). Then CRC models are defined by the joint dynamics

$$dh(t) = \left(A h(t) + \mu^{\text{HJM}} Y(t) r(t)\right)dt + \sigma^{\text{HJM}} Y(t) r(t)dW(t),$$
$$dY(t) = \mu Y(t)dt + \sigma Y(t)d\tilde{W}(t),$$

where $r(t) = h(t)(0)$, $A$ is the generator of the shift semigroup, and $\mu^{\text{HJM}}$, $\sigma^{\text{HJM}}$ are the HJM drift and volatility of the short rate model with constant parameter $y$.

This is a full-fledged HJM model bringing with it the benefits of robust calibration. Indeed, all model parameters can be estimated from realised covariation of yields rather than calibrated by solving high-dimensional inverse problems. Thus, the parameters are estimated from yield curve dynamics instead of calibrated to static yield curves, while an exact match to the current yield curve is guaranteed by the Hull-White extension $\theta$. This will be highlighted below.

The consistent recalibration property holds if the coefficients $\alpha$ and $\beta$ are sufficiently noisy to disperse the forward rate curves all over the set $I$. The irreducibility of the model is also reflected in the much higher ranks of the covariation matrices of yields with different times to maturity. Indeed, our empirical analysis shows that they are closer to those observed in the market than in the corresponding models without CRC extension (see Figure 6.16).

Despite all this flexibility, the model remains tractable. Indeed, if $W$ and $\tilde{W}$ are independent, then the model can be simulated by a splitting scheme which takes advantage of the affine nature of the underlying short rate model. In particular, efficient high-order positivity-preserving simulation schemes for the CIR process such as [1] can be used. No similar schemes are available for general HJM equations with non-Lipschitz vector fields. In our numerical implementation, we achieve first-order convergence of the splitting scheme. We also give a theoretical proof of first-order convergence in the Vasiček case.

1.7. Organisation of the paper. In Section 2, we introduce Hull-White extended affine short rate models, which are the building blocks of CRC models introduced in Section 3. In Sections 4 and 5, the one-factor Vasiček and CIR case is developed in

59
2. Hull-White extended affine short rate models

2.1. Overview. We set the stage for CRC models by describing Hull-White extended affine short rate models, focusing on the correspondence between Hull-White extensions and initial forward rate curves. The one-dimensional short rate model of the introduction is replaced by more general multi-dimensional factor models for the short rate. The parameter \( y \), which becomes stochastic in the CRC setting, is kept constant and fixed.

2.2. Setup and notation. \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) is a filtered probability space. The filtration satisfies the usual conditions. The measure \( \mathbb{P} \) plays the role of a risk-neutral measure. All processes are defined on \( \Omega \), adapted to \((\mathcal{F}_t)_{t \geq 0}\), and càdlàg, \( W = (W_1)_{t \geq 0} \) is \( d \)-dimensional \((\mathcal{F}_t)_{t \geq 0}\)-Brownian motion.

The short rate process \( r = (r(t))_{t \geq 0} \) is determined by a factor process \( X = (X(t))_{t \geq 0} \) with values in a state space \( \mathbb{X} \). The evolution of the factor process depends on a parameter process \( Y = (Y(t))_{t \geq 0} \) with values in a space \( \mathbb{Y} \). In all of Section 2, the parameter process \( Y(t) \equiv y \) is assumed to be constant and fixed, whereas it is allowed to vary in Section 3 below.

The spaces \( \mathbb{X} \) and \( \mathbb{Y} \) are both subsets of some finite dimensional vector space. \( \mathbb{X} \) is, up to permutation of coordinates, of the canonical form \( \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \) with \( d_1 + d_2 = d \geq 1 \). The canonical basis vectors in \( \mathbb{R}^d \) are denoted by \( e_1, \ldots, e_d \).

For each \((x, y) \in \mathbb{X} \times \mathbb{Y}\), there is a symmetric positive semidefinite matrix \( a_y(x) \in \mathbb{R}^{d \times d} \) and a vector \( b_y(x) \in \mathbb{R}^d \), destined to determine the volatility and drift of \( X \). The expressions \( a_y(x) \) and \( b_y(x) \) are affine in \( x \), i.e.,

\[
 a_y(x) = a_y + \sum_{i=1}^d \alpha_y^i x^i, \quad b_y(x) = b_y + \sum_{i=1}^d \beta_y^i x^i, \quad \forall (x, y) \in \mathbb{X} \times \mathbb{Y},
\]

for symmetric positive semidefinite matrices \( a_y, \alpha_y^1, \ldots, \alpha_y^d \in \mathbb{R}^{d \times d} \) and vectors \( b_y, \beta_y^1, \ldots, \beta_y^d \in \mathbb{R}^d \). Moreover, a function \( \theta \in C(\mathbb{R}_+) \) is given, which is used to make the drift of \( X \) time-inhomogeneous.

2.3. Factor process and short rate. The factor process \( X \) is a continuous, \( \mathbb{X} \)-valued solution of the SDE

\[
 dX(t) = \sqrt{a_y(X(t))}dW(t) + \left( \theta(t)e_1 + b_y(X(t)) \right) dt
\]

with initial condition \( X(0) = x \in \mathbb{X} \). The short rate is given by

\[
 r(t) = \ell + \langle \lambda, X(t) \rangle, \quad \forall t \geq 0,
\]

for some fixed \( \ell \in \mathbb{R} \) and \( \lambda \in \mathbb{R}^d \) satisfying \( \langle \lambda, e_1 \rangle \neq 0 \).

Assumption 2.1. It is assumed that SDE (2.1) has a unique continuous, \( \mathbb{X} \)-valued solution \( X \), for each initial condition \( X(s) = x \), where \((s, x) \in \mathbb{R}_+ \times \mathbb{X} \). In this case, the parameters \((y, \theta)\) are called admissible. Moreover, it is assumed that \( X \) satisfies the moment condition

\[
 \mathbb{E} \left[ e^{-\int_0^t (\ell + \langle \lambda, X(s) \rangle) ds} \right] < \infty, \quad \forall t \geq 0.
\]
2.4. \textbf{Exponential moments and Riccati equations}. The process $X$, or rather the family of processes obtained by varying the initial conditions in SDE (2.1), is time-inhomogeneous affine. All coefficients in SDE (2.1) are independent of time, except for the drift $\theta$; we call $X$ \textit{Hull-White extended affine}. This we are going to highlight in detail below. Our main reference for time-inhomogeneous affine processes is [34].

Functions $([\Phi_y, \Psi_y]) \in C^\infty(\mathbb{R}_+) \times C^\infty(\mathbb{R}_+; \mathbb{R}^d)$ are called solutions of the \textit{Riccati equations} (with parameter $y \in \mathcal{Y}$) if

\begin{align}
\tag{2.4a} \Phi'_y &= F_y \circ \Psi_y, & \Phi_y(0) &= 0, \\
\tag{2.4b} \Psi'_y &= R_y \circ \Psi_y - \lambda, & \Psi_y(0) &= 0
\end{align}

holds, where $(F_y, R_y) \in C(\mathbb{R}^d) \times C(\mathbb{R}^d; \mathbb{R}^d)$ are given by

$F_y(u) = \frac{1}{2} \langle u, a_y u \rangle + \langle u, b_y \rangle, \quad R^i_y(u) = \frac{1}{2} \langle u, a^i_y u \rangle + \langle u, \beta^i_y \rangle,$

for all $u \in \mathbb{R}^d$ and $i \in \{1, \ldots, d\}$.

\textbf{Lemma 2.2.} $X$ satisfies moment condition (2.3) if and only if there exists a solution $([\Phi_y, \Psi_y])$ of the Riccati equations (2.4). Moreover, if there exists a solution of the Riccati equations, it is unique.

Note that Lemma 2.2 implies that the moment condition depends only on $y$, not on the choice of Hull-White extension $\theta$.

\textbf{Proof.} Let $Z$ be the $\mathcal{X}^2$-valued process $Z(t) = (X(t), \int_0^t X(s) ds)$. Then $Z$ is an It\^o diffusion whose drift and volatility at time $t \geq 0$ are given by

$$(\theta(t)e_1 + b_y(Z_1(t)), Z_1(t)) \in \mathbb{R}^{2d}, \quad \begin{pmatrix} a_y(Z_1(t)) & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2d \times 2d},$$

respectively. Clearly, these expressions are affine in $Z(t)$. Moreover, the time-inhomogeneity $\theta(t)e_1$ is (by definition) continuous in $t$. Therefore, the process $Z$, or rather the family of processes obtained by varying the initial condition of $Z$, is strongly regular affine, see [34, Theorem 2.14]. For each $(t, u_1, u_2) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$, the functional characteristics of $Z$ are given by

$F(t, u_1, u_2) = \theta(t) (u_1, e_1) + F_y(u_1) \in \mathbb{R}, \quad R(t, u_1, u_2) = (R_y(u_1) + u_2, 0) \in \mathbb{R}^d \times \mathbb{R}^d.$

Moment condition (2.3), expressed in terms of $Z$, reads as follows:

$$\mathbb{E} \left[ e^{-(\lambda, Z_2(T))} \right] < \infty, \quad \forall T \geq 0.$$

By [42], the moment condition is equivalent to the existence of a solution $(\phi, \psi_1, \psi_2)$ of the following Riccati system associated to $Z$:

$$-\partial_t \phi(t, T) = F(t, \psi_1(t, T), \psi_2(t, T)), \quad \phi(T, T) = 0,$$

$$-\partial_t \psi_1(t, T) = R(t, \psi_1(t, T), \psi_2(t, T)), \quad \psi_1(T, T) = 0,$$

$$-\partial_t \psi_2(t, T) = 0, \quad \psi_2(T, T) = -\lambda.$$

Equivalently, the relations $\psi_2(t, T) = -\lambda$ and

$$\phi(t, T) = \int_t^T \theta(s) \langle \Psi_y(T - s), e_1 \rangle ds + \Phi_y(T - t), \quad \psi_1(t, T) = \Psi_y(T - t)$$

61
hold identically, where \((\Phi_y, \Psi_y)\) is a solution of the Riccati equations (2.4). Uniqueness holds for these equations because the vector fields are locally Lipschitz.

2.5. Bond prices and forward rates. By the affine nature of the factor process, bond prices can be obtained by solving the Riccati system of ODE’s.

**Theorem 2.3** (Zero-coupon bond price and forward rate). Let \(X\) satisfy moment condition (2.3) and let \((\Phi_y, \Psi_y)\) be the unique solution of the Riccati equations with parameter \(y\), given by Lemma 2.2. Then the bond prices in the short rate model (2.1)–(2.2) satisfy

\[
P(t, T) = \mathbb{E} \left[ e^{-\int_t^T \tau(s) ds} \middle| \mathcal{F}_t \right] = e^{-\int_0^T \theta(s)(\Psi_y(T-s), e_1) ds + \Phi_y(T-t) + \langle \Psi_y(t), X(t) \rangle},
\]

for all \(T \geq t \geq 0\), and the forward rates are given by

\[
h(t, \tau) = h(t)(\tau) = -\partial_x \log \left( P(t, t + \tau) \right)
\]

\[
= -\int_0^{\tau} \theta(t + s)(\Psi_y'(\tau - s), e_1) ds - \Psi_y'(\tau) - \langle \Psi_y'(\tau), X(t) \rangle,
\]

for all \(t, \tau \geq 0\).

For essentials on short rate models we refer to [33, Chapter 5]. The parametrisation of the forward rate as a function of \(t\) and \(\tau\) is called Musiela parametrisation. It is particularly useful in this paper since \((h(t))_{t \geq 0}\) will be interpreted as a stochastic process taking values in a suitable space of functions on \(\mathbb{R}_+\).

**Proof.** We borrow from the proof of Lemma 2.2, where moment condition (2.3) was shown to be equivalent to the existence of solutions \((\phi, \psi_1, \psi_2)\) of the Riccati system associated to the process \(Z = (X, \int X)\). Moreover, \((\phi, \psi_1, \psi_2)\) are closely related to the solutions \((\Phi_y, \Psi_y)\) of Riccati system (2.4). By the main theorem in [42] and its conditional version, the affine transform formula

\[
\mathbb{E} \left[ e^{-\langle \lambda, Z(T) \rangle} \middle| \mathcal{F}_t \right] = e^{\phi(t,T) + \langle \psi_1(t,T), Z_t(t) \rangle + \langle \psi_2(t,T), Z_t(t) \rangle}, \quad \forall T \geq t \geq 0,
\]

holds. A direct calculation shows this formula to be equivalent to formula (2.5) for bond prices. Formula (2.6) for forward rates is obtained by taking the logarithm and differentiating with respect to \(\tau\).

2.6. Heath-Jarrow-Morton equation. The evolution of forward rate curves is described by the *HJM equation*. For each \((x, y) \in X \times Y\), let \(\mu_y^{\text{HJM}}(x)\) and \(\sigma_y^{\text{HJM}}(x)\) be given by

\[
\mu_y^{\text{HJM}}(x) = \langle \Psi_y, a_y(x) \Psi_y' \rangle \in C^\infty(\mathbb{R}_+), \quad \sigma_y^{\text{HJM}}(x) = -\sqrt{a_y(x)} \Psi_y' \in C^\infty(\mathbb{R}_+; \mathbb{R}^d).
\]

Note that the familiar HJM drift condition holds:

\[
(2.7) \quad \mu_y^{\text{HJM}}(x)(\tau) = \mathbb{E} \left[ \sigma_y^{\text{HJM}}(x)(\tau), \int_0^{\tau} \sigma_y^{\text{HJM}}(x)(s) ds \right], \quad \forall \tau \geq 0.
\]

Let \(\mathbb{H}\) be a Hilbert space, destined to contain the forward rate curves of the model.

**Assumption 2.4.** \(\mathbb{H}\) is a Hilbert space with the following properties:

(i) \(\mathbb{H} \subset C(\mathbb{R}_+)\) and the evaluation map \(\text{eval} : h \mapsto h(\tau)\) is continuous on \(\mathbb{H}\), for each \(\tau \in \mathbb{R}_+\);

(ii) for each \((x, y, z) \in X \times Y \times \mathbb{R}^d\), \(\mu_y^{\text{HJM}}(x)\) and \(\langle \sigma_y^{\text{HJM}}(x), z \rangle\) are elements of \(\mathbb{H}\);
(iii) the right shifts \((S_t)_{t \geq 0}\) mapping \(h\) to \(h(t + \cdot)\) define a strongly continuous semigroup on \(H\) with infinitesimal generator \(A\).

**Theorem 2.5** (HJM equation). Let \((h, X)\) be given by Theorem 2.3 and assume that \(h(t) \in H\), for each \(t \geq 0\). Then the process \((h, X)\) is a strong solution of the following SPDE on \(H \times X\):

\[
\begin{align*}
 dh(t) &= \left( Ah(t) + \mu^\text{HJM}_Y(X(t)) \right) dt + \sigma^\text{HJM}_Y(X(t)) dW(t), \\
 dX(t) &= \sqrt{a_y(X(t))} dW(t) + \left( \theta(t) e_1 + b_y(X(t)) \right) dt.
\end{align*}
\]

(2.8)

For the concepts of mild, weak, and strong solutions of SPDE’s, we refer to [24, Section 6.1]. In the one-factor case, the factor process \(X(t)\) is a simple functional of the forward rate. Then equation (2.8) can be rewritten as an evolution equation for the forward rate process alone (c.f. equations (4.4) and (5.4)). This is also possible in the multi-factor case, but the corresponding functional is generally quite complicated, which is why we chose to present the HJM equation in the form (2.8).

**Proof.** Differentiating formula (2.6) for forward rates with respect to \(t\) and \(\tau\) and using \(\Psi'_y(0) = -\lambda\), one obtains for each \(\tau \geq 0\)

\[
\begin{align*}
 dh(t, \tau) &= -\int_t^{t+\tau} \theta(s) \langle \Psi'_y(t + \tau - s), e_1 \rangle ds + \theta(t + \tau) \langle \lambda, e_1 \rangle \\
&\quad + \theta(t) \langle \Psi'_y(\tau), e_1 \rangle dt - \langle \Psi'_y(\tau), dX(t) \rangle, \\
Ah(t, \tau) &= -\int_t^{t+\tau} \theta(s) \langle \Psi''_y(t + \tau - s), e_1 \rangle ds + \theta(t + \tau) \langle \lambda, e_1 \rangle \\
&\quad - \Phi''_y(\tau) - \langle \Psi''_y(\tau), X(t) \rangle.
\end{align*}
\]

Subtracting the equations and cancelling out the integral as well as the term next to it yields

\[
\begin{align*}
 dh(t, \tau) &= \left( Ah(t, \tau) + \Phi''_y(\tau) + \langle \Psi''_y(\tau), X(t) \rangle + \theta(t) \langle \Psi'_y(\tau), e_1 \rangle \right) dt - \langle \Psi'_y(\tau), dX(t) \rangle. \\
\end{align*}
\]

When \(dX(t)\) is replaced by the right-hand side of SDE (2.1), the \(\theta(t)\)-term cancels out and one obtains for each \(\tau \geq 0\)

\[
\begin{align*}
 dh(t, \tau) &= \left( Ah(t, \tau) + \Phi''_y(\tau) + \langle \Psi''_y(\tau), X(t) \rangle - \langle \Psi'_y(\tau), b_y(X(t)) \rangle \right) dt \\
&\quad - \left\langle \Psi'_y(\tau), \sqrt{a_y(X(t))} dW(t) \right\rangle.
\end{align*}
\]

The symmetric matrix \(\sqrt{a_y(X(t))}\) can be moved to the other side of the scalar product, and one immediately recognises the volatility \(\sigma^\text{HJM}_y(X(t))\). A direct calculation shows that the drift is equal to \(\mu^\text{HJM}_y(X(t))\). Indeed,

\[
\begin{align*}
\Phi''_y + \langle \Psi'_y, x \rangle - \langle \Psi'_y, b_y(x) \rangle &= F'_y \circ \Psi_y \cdot \Psi'_y + \langle R'_y \circ \Psi_y, \Psi'_y, x \rangle - \langle \Psi'_y, b_y(x) \rangle \\
&= \langle \Psi_y, a_y(x) \Psi'_y \rangle = \mu^\text{HJM}_y(x),
\end{align*}
\]

which follows from the relations

\[
F'_y(u) \cdot v = \langle u, a_y v \rangle + \langle v, b_y \rangle, \quad (R'_y)^t(u) \cdot v = \langle u, a'_y v \rangle + \langle v, \beta'_y \rangle.
\]

□
2.7. Forward rates and Hull-White extensions. Relation (2.6) between the forward rate \( h(t) \) and the Hull-White extension \( \theta \) plays a key role in calibration (and recalibration) of the model. It can be expressed concisely as

\[
h(t) = \mathcal{H}_y(X(t), S_t \theta), \quad S_t \theta = C_y(h(t), X(t)), \quad \forall t \geq 0,
\]

where \( S_t \theta \) is the shift operator, \( \mathcal{H}_y \) calculates the initial forward rate curve from the Hull-White extension given parameter \( y \), and \( C_y \) performs the inverse operation of calibrating a Hull-White extension to an initial forward rate curve. Formally, for each \((t, x, \theta) \in \mathbb{R}_+ \times \mathbb{X} \times C(\mathbb{R}_+)\), these operators are given by

\[
S_t \theta = \theta(t + \cdot) \in C(\mathbb{R}_+), \\
\mathcal{H}_y(x, \theta) = \ell - \int_0^x \theta(s) \langle \Psi_y'(s), e_1 \rangle ds \in C^1(\mathbb{R}_+), \\
\mathcal{I}_y(\theta) = \frac{1}{2} \langle \Psi_y'(0), e_1 \rangle \theta(0) + \frac{1}{2} \langle \Psi_y(0), e_1 \rangle \theta(0) \in C^1(\mathbb{R}_+).
\]

Note that \( \mathcal{H}_y \) involves the Volterra integral operator \( \mathcal{I}_y \). The operator \( C_y \) (the letter \( C \) standing for calibration) is defined as the partial inverse of \( \mathcal{H}_y \) given by the following theorem.

**Theorem 2.6 (Calibration to initial forward rate curves).** Let \((h, x) \in C^1(\mathbb{R}_+) \times \mathbb{X} \) satisfy \( h(0) = \ell + \langle \lambda, x \rangle \). Then the Volterra integral equation \( h = \mathcal{H}_y(x, \theta) \) has a unique solution \( \theta \in C(\mathbb{R}_+) \), which we denote by \( C_y(h, x) \).

The theorem is a direct consequence of the following lemma.

**Lemma 2.7.** For each \( y \in \mathcal{Y} \), the Volterra integral operator

\[
\mathcal{I}_y : C(\mathbb{R}_+) \to \{ h \in C^1(\mathbb{R}_+) : h(0) = 0 \}
\]

is bijective.

**Proof.** This follows from [12, Theorem 2.1.8], noting that the integral kernel

\[
K_y(s, t) = \langle \Psi_y'(t - s), e_1 \rangle, \quad \forall t \geq s \geq 0,
\]

satisfies \(|K_y(t, t)| = |\langle \lambda, e_1 \rangle| > 0\) and both \( K_y \) and \( \partial_t K_y \) are continuous. \( \Box \)

Note that calibration of a Hull-White extension \( \theta \) requires the inversion of the Volterra integral operator \( \mathcal{I}_y \). The assumption \( \langle \lambda, e_1 \rangle \neq 0 \) is needed, here.

2.8. Numerical solution of the Volterra equation. In the absence of analytical formulas, the Volterra equation has to be solved numerically. We are aiming at a second order approximation to keep the global error of the simulation scheme of order one. Thus, we approximate the Volterra integral operator \( \mathcal{I}_y \) by the trapezoid rule, which yields an operator \( \hat{\mathcal{I}}_y \) given by

\[
\hat{\mathcal{I}}_y(\theta)(\tau_n) = \delta \left( \frac{1}{2} \langle \Psi_y'(\tau_n), e_1 \rangle \theta(0) + \frac{1}{2} \langle \Psi_y'(0), e_1 \rangle \theta(0) \right) + \frac{1}{2} \sum_{i=1}^{n-1} \langle \Psi_y'(\tau_n - \tau_i), e_1 \rangle \theta(\tau_i) + \frac{1}{2} \langle \Psi_y'(\tau_n), e_1 \rangle \theta(\tau_n),
\]

64
for each $n \in \mathbb{N}^+$, where $\tau_n = n\delta$ constitutes a uniform grid of step size $\delta > 0$. Approximate solutions $\hat{\theta}$ can be constructed by solving for continuous piecewise linear (i.e. linear on each interval $[\tau_n, \tau_{n+1}]$) functions $\hat{\theta}$ satisfying

\begin{equation}
\hat{\theta}(0) = \frac{g'(0)}{(\Psi'_y(0), e_1)}, \quad \hat{I}_y(\hat{\theta})(\tau_n) = g(\tau_n), \quad \forall n \in \mathbb{N}^+.
\end{equation}

As $\hat{I}_y$ is a second order approximation of $I_y$, it is not surprising that $\hat{\theta}$ is a second order approximation of $\theta$.

**Lemma 2.8.** Let $(x, y) \in \mathbb{R} \times \mathcal{Y}$ and $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ piecewise $C^4$ with continuous second derivatives. If $g(0) = 0$, then there is a unique piecewise linear function $\theta \in C(\mathbb{R}_+)$ satisfying (2.10). Moreover, $\hat{\theta}$ is a second order approximation of the exact solution $\theta$ of the Volterra equation $I_y(\theta) = g$ in the sense that for each $T \in \mathbb{R}_+$,

$$
\sup_{t \in [0,T]} |\hat{\theta}(t) - \theta(t)| \leq C \delta^2,
$$

where $C$ is a constant depending only on $T$ and $g$.

The smoothness assumption on the right-hand side $g$ of the Volterra equation is satisfied if the yields are interpolated sufficiently smoothly.

**Proof.** This follows from [59, Theorem 3] and [12, Example 2.4.5 and Theorems 2.4.5, 2.4.8], noting that the integral kernel (2.9) is $C^4$ and strictly bounded away from zero along the diagonal by our assumption $(\Psi'_y(0), e_1) = (\lambda, e_1) \neq 0$. □

Solving for $\hat{\theta}$ can be performed efficiently because (2.10) is a linear system for $(\hat{\theta}(\tau_n), \ldots, \hat{\theta}(\tau_m))$ of lower triangular form.

**2.9. Estimation of the affine coefficients.** The coefficients $a_y, \alpha_y, \beta_y$ can be estimated directly from the empirical covariation of yields. Let $r(t, \tau)$ denote the yield of a zero-coupon bond held from $t$ to $t + \tau$, i.e.

\begin{equation}
(2.11)
\quad r(t, \tau) = -\frac{1}{\tau} \log P(t, t + \tau), \quad \forall t, \tau \in \mathbb{R}_+.
\end{equation}

Then, by equation (2.5) the quadratic covariation of yields with maturity $\tau_i$ and $\tau_j$ satisfies

$$
[r(\cdot, \tau_i), r(\cdot, \tau_j)](t) = \frac{1}{\tau_i \tau_j} \left( t \Psi_y(\tau_i) \Psi_y(\tau_j) + \sum_{k=1}^{d} \Psi_y(\tau_i) \alpha_y^k \Psi_y(\tau_j) \int_0^t X^k(s)ds \right),
$$

for all $t, \tau_i, \tau_j \in \mathbb{R}_+$. If yields $\hat{r}(t_n, \tau_i)$ are given by the market, for times $t_n$ and times to maturity $\tau_i$, then the quadratic covariation can be estimated by the realised covariation, which is defined as

$$
[r(\cdot, \tau_i), r(\cdot, \tau_j)](t_n) = \sum_{k=1}^{n} (\hat{r}(t_k, \tau_i) - \hat{r}(t_{k-1}, \tau_i)) (\hat{r}(t_k, \tau_j) - \hat{r}(t_{k-1}, \tau_j)).
$$

Assuming we also observe the factors $\hat{X}(t_n)$, for times $t_n = n\delta$, then the integrals over the factors are approximated by

$$
\int_0^{t_n} X^i(s)ds \approx \delta \sum_{m=1}^{n} \hat{X}_i(t_m), \quad \text{for } i = 1, \ldots, d.
$$
Thus, fixing a time window of length \( M \) and a time \( t_n \), one has
\[
\frac{\hat{\sigma}(\tau_i, \tau_j)}{(t_n - t_{n-M})} \approx \frac{1}{\tau_i \tau_j} \psi_{y}(\tau_i)^\top a_y \psi_{y}(\tau_j)
\]
(2.12)
\[
\frac{\delta}{\tau_i \tau_j(t_n - t_{n-M})} \sum_{k=1}^{d} \psi_{y}(\tau_i)^\top a_k^y \psi_{y}(\tau_j) \sum_{m=n-M+1}^{n} \hat{X}^y(t_m)
\]
The right-hand side of this equation depends on the coefficients \( a_y, a_y, \beta_y \). (Note the implicit dependence of \( \psi_{y} \) on \( \beta_y \) given by equation (2.4b).) Therefore, for any selection of times to maturity \( \tau_i, \tau_j \), estimators for these coefficients can be obtained by solving for those \( \hat{a}_y, \hat{a}_y, \hat{\beta}_y \) which achieve the best fit in equation (2.12). Note that (2.12) is derived solely from the diffusion coefficient of the yield dynamics and therefore invariant under Girsanov’s change of measure. Thus, the coefficients \( a_y, a_y, \beta_y \) can be estimated from real world observations without specifying the market price of risk. The estimates should not depend on the choice of \( \tau_i, \tau_j \), under the model hypothesis, which provides a means to reject ill-suited models. Varying the calibration time \( t_n \) creates time series of coefficients \( \hat{a}(t_n), \hat{a}(t_n), \hat{\beta}(t_n) \).

We now discuss how the coefficient \( b_y \) can be estimated. First, note that for one-factor models \( b_y \) is redundant and can be normalised to zero because of the Hull-White extension which is calibrated to the prevailing market yield curve. In the multi-factor case only the first component of the vector \( b_y \) is redundant. The other components may also be calibrated to the prevailing market yield curve by regression methods. Alternatively, these can be estimated by econometric methods. However, these require a market price of risk specification. We do not discuss this topic, here.

3. Consistent recalibration of affine short rate models

3.1. Overview. The constant parameter process \( y \) of the previous section is now replaced by a stochastic process \( Y = (Y(t))_{t \geq 0} \). The situation is particularly simple when \( Y \) is piecewise constant. In this case, the Hull-White extension is recalibrated to the prevailing yield curve (i.e., the yield curve given by the model with old parameters) each time the parameter process changes. Later on, the concepts are generalised to arbitrary parameter processes \( Y \), resulting in our definition of general CRC models. Geometrically, these models “locally look like” Hull-White extended affine short rate models with fixed parameter \( y \). This is made precise in Section 3.9. A semigroup point of view is taken in Section 3.11, leading to an interpretation of CRC models with piecewise constant \( Y \) as splitting schemes for more general CRC models.

3.2. Setup and notation. We recall the notion of admissible parameters from Assumption 2.1. For all \((s, x) \in \mathbb{R}_+ \times \mathbb{X}\) and all admissible parameters \((y, \theta) \in \mathcal{Y} \times C(\mathbb{R}_+)\), we let \( X = X^{y, x}_{\theta} \) denote the unique solution on \([s, \infty)\) of the SDE (2.1) with \( \theta(t) \) replaced by \( \theta(t - s) \) and initial condition \( X(s) = x \).

We fix a Hilbert space \( \mathcal{H} \) of forward rate curves satisfying Assumption 2.4 simultaneously for all \( y \in \mathcal{Y} \). Let \((t_n)_{n \in \mathbb{N}_0}\) be a strictly increasing sequence of deterministic times \( t_n \in [0, \infty) \). We assume that all discrete-time processes are defined on this grid and adapted to the discrete-time filtration \((\mathcal{F}_{t_n})_{n \in \mathbb{N}_0}\).
3.3. Discrete-time CRC models.

Definition 3.1 (Discrete-time CRC models). A discrete-time stochastic process \((h, X, Y)\) with values in \(\mathbb{H} \times X \times Y\) is called a discrete-time CRC model if there exists a discrete-time stochastic process \(\theta\) with values in \(C(\mathbb{R}_+)\) such that the following conditions are satisfied, for each \(n \in \mathbb{N}_0\):

(i) The Hull-White extension \(\theta(t_n)\) is obtained by calibration to \(h(t_n)\):

\[
h(t_n)(0) = \ell + \langle \lambda, X(t_n) \rangle \quad \text{and} \quad \theta(t_n) = C_{Y(t_n)}(h(t_n), X(t_n)).
\]

(ii) The evolution of \(X\) on \([t_n, t_{n+1}]\) corresponds to the Hull-White extended affine model determined by the parameters \((Y(t_n), \theta(t_n))\):

\[
X(t_{n+1}) = X_{Y(t_n),\theta(t_n)}^{t_n}(t_{n+1}).
\]

Here, Assumption 2.1 is assumed to hold for the parameters \((Y(t_n), \theta(t_n)) \in Y \times C(\mathbb{R}_+)\).

(iii) The forward rate \(h(t_{n+1})\) is determined by \(X(t_{n+1})\) according to the prevailing Hull-White extended affine model:

\[
h(t_{n+1}) = H_{Y(t_n)}(X(t_{n+1}), S_{t_{n+1}-t_n}) \theta(t_n).
\]

3.4. Simulation. If we assume that a stochastic model for the evolution of the parameter process \(Y\) is specified, then consistent recalibration models can be simulated by applying iteratively steps (i)–(iii) of Definition 3.1.

Algorithm 3.2 (Simulation). Given \(h(0)\) and the discrete-time process \(Y\), calculate a discrete-time process \((h, X, Y, \theta)\) by iteratively executing steps (i)–(iii) of Definition 3.1, for each \(n \in \mathbb{N}_0\). Abort with an error if the assumption in step (ii) is not satisfied, for any \(n \in \mathbb{N}_0\).

The algorithm is illustrated in Figure 3.1. Note that the forward rate increments are calculated from increments of the affine factor process \(X\), which can typically be simulated with high orders of accuracy and proper treatment of boundary conditions. These advantages are thanks to the affine structure of the CRC increments and are not available for general HJM models.

![Figure 3.1. Simulation of CRC models. Updating \(\theta\), \(X\), \(h\) is done using (i), (ii), (iii) of Definition 3.1, respectively. Updating \(Y\) is done using the exogenously given model for \(Y\).](image-url)
3.5. Efficient updating of forward rate curves. Updating the curve of forward rates as prescribed by Definition 3.1(iii) involves calculating integrals on time intervals \([0, \tau]\), for large values of \(\tau\) (see Section 2.7 for the formulas). A significant speed-up can be obtained when this update is done using the alternative formula provided by the following lemma, which involves only integrals over time intervals of length \(\delta = t_{n+1} - t_n\).

**Lemma 3.3** (Efficient updating of forward rate curves). Definition 3.1(iii) can be rewritten as

\[
(3.1) \quad h(t_{n+1}) = S_\delta h(t_n) + S_\delta \Psi_Y^{(t_n)} - \Phi_Y^{(t_n)} + \left\langle S_\delta \Psi_Y^{(t_n)}, X(t_n) \right\rangle \\
- \left\langle \Phi_Y^{(t_n)}, X(t_{n+1}) \right\rangle + \int_0^\delta \theta(t_n)(s) \left\langle S_{\delta-s} \Psi_Y^{(t_n)}, e_1 \right\rangle ds,
\]

where \(\delta = t_{n+1} - t_n\).

**Proof.** By conditions (i) and (iii) of Definition 3.1,
\[
h(t_{n+1}) - S_\delta h(t_n) = H_{Y(t_n)}(X(t_{n+1}), S_\delta \theta(t_n)) - S_\delta H_{Y(t_n)}(X(t_n), \theta(t_n)) \\
= \ell - I_{Y(t_n)}(S_\delta \theta(t_n)) - \Phi_Y^{(t_n)} - \left\langle \Psi_Y^{(t_n)}, X(t_{n+1}) \right\rangle \\
- \ell + S_\delta I_{Y(t_n)}(\theta(t_n)) + S_\delta \Phi_Y^{(t_n)} + \left\langle S_\delta \Psi_Y^{(t_n)}, X(t_n) \right\rangle.
\]

Now the assertion of the lemma follows from the relation

\[
S_\delta I_{\theta}(\theta) - I_{\theta}(S_\delta \theta) = \int_0^\delta \theta(s) \left\langle S_{\delta-s} \Psi_Y e_1 \right\rangle ds, \quad \forall (\delta, \theta) \in \mathbb{R}_+ \times C(\mathbb{R}_+),
\]

which can be easily verified from the definition. \(\square\)

3.6. Discrete-time CRC models, equivalent definition. While the formulation of Definition 3.1 in discrete time is useful for simulation, the connection to HJM models is more easily seen in continuous time. Therefore, we extend the processes \((h, X, Y, \theta)\) to \(\mathbb{R}_+\) by defining for each \(t \in \{t_n, t_{n+1}\}\),

\[
Y(t) = Y(t_n), \quad X(t) = X_{Y(t_n), \theta(t_n)}(t), \quad h(t) = H_{Y(t_n)}(X(t), S_{t-t_n} \theta(t_n)).
\]

Imposing additionally that the processes \(X\) and \(h\) are continuous at every grid point \(t_n\) uniquely characterises \((h, X, Y, \theta)\). Thus, we lead to the following definition, which is essentially equivalent to Definition 3.1 by the previous considerations.

**Definition 3.4** (Discrete-time CRC models). A continuous-time stochastic process \((h, X, Y)\) with values in \(H \times X \times \mathbb{Y}\) is called a discrete-time CRC model if there exists a continuous-time stochastic process \(\theta\) with values in \(C(\mathbb{R}_+)\) such that the following properties hold:

(i) equation (3.2) is satisfied, for each \(t \in \{t_n, t_{n+1}\}\) and \(n \in \mathbb{N}_0;\)
(ii) the process \((h, X)\) is continuous at \(t_n\), for each \(n \in \mathbb{N}_0;\) and
(iii) the parameters \((Y(t), \theta(t))\) are admissible, for each \(t \in \mathbb{R}_+\).

We emphasise that recalibration still happens on a discrete time scale because the parameter process \(Y\) is piecewise constant on \([t_n, t_{n+1}]\).
3.7. Bond prices and forward rates.

**Theorem 3.5 (Zero-coupon bond price and forward rate).** Let \((h,X,Y)\) be a consistent recalibration model as in Definition 3.4 with corresponding process \(\theta\). Define
\[
P(t,T) = e^{-\int_0^T h(t,s-t)ds}, \quad r(t) = h(t,0), \quad B(t) = e^{\int_0^t r(s)ds}.
\]
Then the discounted price process \(t \mapsto P(t,T)/B(t)\) is a \(\mathbb{F}\)-martingale, for each \(T \geq 0\). In this sense, the bond market is free of arbitrage. Moreover, bond prices and short rates are related by
\[
P(t,T) = \mathbb{E} \left[ e^{-\int_t^T r(s)ds} \bigg| \mathcal{F}_t \right]
\]
and the following affine bond pricing formulas hold:
\[
P(t,T) = e^{-(T-t)\ell} \int_0^{T-t} \theta(t,s)(\Psi_{Y(t)}(T-t-s), e_1) ds + \Phi_{Y(t)}(T-t) + \langle \Psi_{Y(t)}(T-t), X(t) \rangle,
\]
\[
h(t,\tau) = \ell - \int_0^\tau \theta(t,s)(\Psi_{Y(t)}(\tau-s), e_1) ds - \Phi_{Y(t)}(\tau) - \langle \Psi_{Y(t)}(\tau), X(t) \rangle.
\]

**Proof.** On each interval \([t_n, t_{n+1}]\), the evolution of forward rate curves \(h(t)\) stems from a Hull-White extended affine short rate model. Therefore, for each \(t \geq 0\), the discounted price process \(t \mapsto P(t,T)/B(t)\) is a martingale on each interval \([t_n, t_{n+1}]\). Moreover, the process is continuously stitched together at the boundaries \(t_n\) of the intervals. It follows that the process is a martingale on \([0, \infty)\). Therefore, the model is free of arbitrage and \(P(t,T) = \mathbb{E}[e^{-\int_t^T r(s)ds}]\) holds, for all \(0 \leq t \leq T\). The affine bond pricing formulas are equivalent to \(h(t) = \mathcal{H}_{Y(t)}(X(t), \theta(t))\), which follows from equation (3.2). 

\[\square\]


**Theorem 3.6 (HJM equation).** Let \((h,X,Y)\) be a consistent recalibration model as in Definition 3.4 with corresponding process \(\theta\) and assume that \(h(t) \in \mathbb{H}\), for each \(t \geq 0\). Then the following properties hold:

(i) the expression \(\mathcal{C}_{Y(t)}(h(t),X(t))\) is well-defined, for all \(t \geq 0\);

(ii) the parameters \((Y(t), \theta(t))\) are admissible, for all \(t \geq 0\); and

(iii) the process \((h, X)\) is a strong solution of the following SPDE on \(\mathbb{H} \times \mathbb{X}\):
\[
dh(t) = \left( Ah(t) + \mu^{HJM}_{Y(t)}(X(t)) \right) dt + \sigma^{HJM}_{Y(t)}(X(t)) dW(t),
\]
\[
dx(t) = \sqrt{a_{Y(t)}(X(t))} dW(t) + \left( \mathcal{C}_{Y(t)}(h(t), X(t))(0) e_1 + b_{Y(t)}(X(t)) \right) dt.
\]

3.9. Geometric interpretation. The consistent recalibration scheme has a nice geometric interpretation. Forward rate curves of a Hull-White extended affine short rate model remain within the finite dimensional manifold with boundary given by
\[
\left\{ -\int_0^T \theta(t+s)(\Psi_{Y}(\tau-s), e_1) ds + \ell - \Phi_{Y}(\tau) - \langle \Psi_{Y}(\tau), x \rangle \bigg| (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \right\},
\]
as can be seen from Theorem 2.3. These submanifolds foliate the space of forward rate curves or large portions thereof. Let \(h\) be a forward rate curve. Then, for every choice of functional characteristics \((F_y, R_y)\), there is at most one leaf through \(h\). However, if \((F_y, R_y)\) is allowed to vary, there are in general many leaves through \(h\). A choice of leaf corresponds to a choice of foliation and thus to a choice functional
characteristics \((F_y, R_y)\). For the characterisation of finite dimensional forward rate models we refer to [35].

A consistent recalibration model is constructed by concatenating forward rate evolutions on leaves belonging to different foliations. This allows the otherwise constant coefficients \((F_y, R_y)\) to change over time. The result is an HJM model which is “tangent” to Hull-White extended affine short rate models. This is illustrated in Figure 3.2.

3.10. Continuous-time CRC models. By continuous-time CRC models, we mean models where the parameter process \(Y\) is not required to be piecewise constant as in the last sections, but is allowed to vary over time. To characterise such models, we use the SPDE derived in Theorem 3.6.

**Definition 3.7** (Continuous-time CRC models). A continuous-time CRC model is a continuous process \((h, X, Y)\) with values in \(H \times X \times Y\) satisfying conditions (i)–(iii) of Theorem 3.6.

Note that these models are free of arbitrage because the HJM drift condition (2.7) is satisfied.

3.11. Semigroup interpretation. If the parameter process \(Y\) is a Markov process on \(Y\) and the SPDE (3.3) is well-posed, then the continuous-time CRC model \((h, X, Y)\) in Definition 3.7 is Markov on \(H \times X \times Y\). Let \(P\) denote the corresponding semigroup on the Banach space \(C_b(H \times X \times Y)\) of bounded continuous functions, i.e.,

\[
P_t f(h_0, x_0, y_0) = \mathbb{E} \left[ f(h(t), X(t), Y(t)) \middle| (h(0), X(0), Y(0)) = (h_0, x_0, y_0) \right].
\]

Moreover, let \(Q\) denote the semigroup on \(C_b(H \times X \times Y)\) obtained by holding the parameter process \(Y(t) \equiv y\) fixed, i.e.,

\[
Q_t f(h_0, x_0, y_0) = \mathbb{E} \left[ f(h(t), X(t), y_0) \middle| (h(0), X(0)) = (h_0, x_0) \right],
\]

where \((h, X)\) are as in Theorem 2.5 with \(y = y_0\). Finally, let \(R\) denote the semigroup on \(C_b(H \times X \times Y)\) describing the evolution of \(Y\), i.e.,

\[
R_t f(h_0, x_0, y_0) = \mathbb{E} \left[ f(h_0, x_0, Y(t)) \middle| Y(0) = y_0 \right].
\]
Then, the concatenation \((R_\delta Q_\delta)^n\) of semigroups describes discrete-time CRC models. Indeed,

\[(R_\delta Q_\delta)^n f(h_0, x_0, y_0) = \mathbb{E}[h(t_n), X(t_n), Y(t_n)], \quad \forall n \in \mathbb{N}_0,
\]

where \(\delta = t_{n+1} - t_n\) is the step size, \((h(t_n), X(t_n))\) is obtained by executing the simulation scheme in Section 3.4 and \(Y(t_n)\) by sampling the Markov process \(Y\) at the discrete time grid.

3.12. **Discrete-time CRC models as splitting schemes.** The semigroup interpretation of Section 3.11 allows one to view discrete-time CRC models as exponential Euler splitting schemes for continuous-time CRC models. To see this, let \(f : \mathcal{H} \times X \times Y \to \mathbb{R}\) be twice differentiable with derivatives uniformly continuous on bounded sets and assume that \(\mathcal{H} \subseteq \text{dom}(A)\). Then Itô’s formula holds for \(f(h(t), X(t), Y(t))\) by [24, Theorem 4.17]. It follows that \(f\) lies in the common domain of the generators \(\mathcal{G}^P, \mathcal{G}^Q, \mathcal{G}^R\) of the semigroups \(\mathcal{P}, \mathcal{Q}, \mathcal{R}\) and

\[\mathcal{G}^P f = \mathcal{G}^Q f + \mathcal{G}^R f.\]

The exponential Euler splitting scheme with respect to this splitting is defined as

\[P_{n\delta} f \approx (\exp(\delta\mathcal{G}^R) \exp(\delta\mathcal{G}^Q))^n f = (R_\delta Q_\delta)^n f, \quad \forall n \in \mathbb{N}_0.
\]

By the considerations in Section 3.11, it coincides with the simulation scheme for discrete-time CRC models.

3.13. **Calibration of CRC models.** In order to calibrate CRC models we need to estimate a time series of the parameter process \(Y\) from market data, and fit a model for this time series. By Theorem 3.6 quadratic covariations of forward rates for CRC models satisfy

\[(3.4) \quad d[h(\cdot, \tau_i), h(\cdot, \tau_j)](t) = \Psi_{Y(t)}'(\tau_i)^\top \left( a_{Y(t)} + \sum_{k=1}^{d} \alpha_{Y(t)}^k X^k(t) \right) \Psi_{Y(t)}'(\tau_j) dt,
\]

for any times to maturity \(\tau_i, \tau_j \in \mathbb{R}_+\). Starting from a time series of estimated quadratic covariations, we can estimate a time series for the parameter process by the same methods as for affine models with constant parameters (see Section 2.7). These are model parameters under a risk neutral measure, but they can be estimated from real world observations since (3.4) is obtained solely from the volatility of the forward rate process.

Calibrating CRC models requires the additional task of selecting and fitting a model for the estimated time series of \(Y\). This completes the model specification under a risk neutral probability measure. We leave the market price of risk specification open in this paper.

3.14. **Robust calibration, consistency, and analytic tractability.** It is time to readdress the question to what extent CRC models satisfy the interest rate modelling principles set forth in the introduction. Arbitrage is excluded because CRC models are HJM models.

- **Robust calibration:** the robust calibration principle is satisfied perfectly. Indeed, the method described in Section 3.13 allows to use the entire present and past market data of yields to select a model. Whenever possible, the parameters are estimated from realised covariations of yields, which allows one to bypass the usual inverse problems in calibration.
Moreover, requiring parameters to remain constant throughout the life time of the model is less restrictive in CRC models than in the underlying affine factor models. The reason is that the parameters of the underlying affine model are turned into state variables of the CRC model.

- **Consistency**: the canonical state space $\mathcal{I}$ of CRC models is the subset of $\mathbb{H} \times X \times Y$ determined by the admissibility condition on the underlying Hull-White extended affine factor model (see Assumption 2.1). Under sensible specifications of the affine factor model, $\mathcal{I} \cap (\mathbb{H} \times \{x\} \times \{y\})$ is large enough to contain all realistic market curves, for each fixed $(x, y) \in X \times Y$. If the Hilbert space $\mathbb{H}$ is continuously embedded in $C^1(\mathbb{R}_+)$, then $\mathcal{I}$ is also large in the topological sense of having non-empty interior.

Consistency holds because the state process $(h, X, Y)$ of CRC models does not leave the set $\mathcal{I}$, by definition. The consistent recalibration property holds if the state process is irreducible on $\mathcal{I}$. This can be verified by checking Hömander’s condition. Generically, Hömander’s condition is satisfied as soon as there is noise in the parameter process $Y$. The exact conditions are worked out for the Vasicek case in Section 4.9.

- **Analytic tractability**: the simulation scheme for CRC models (Algorithm 3.2) transfers the task of sampling state variable increments to a finite-dimensional setting. Namely, instead of simulating forward rate increments from an infinite-dimensional space, it is sufficient to simulate increments of the finite-dimensional processes $X$ and $Y$. This allows one to take advantage of the existing high-order schemes for the simulation of affine processes.

## 4. Consistent Recalibration of Vasicek Models

### 4.1. Overview.

We describe CRC models based on the Hull-White extended Vasicek model in full detail. Moreover, we show using semigroup theory that discrete-time CRC models converge to their continuous-time counterparts.

### 4.2. Setup and notation.

We use the setup of Sections 2.2 and 3.2, setting $X = \mathbb{R}$, $\ell = 0$, $\lambda = 1$. We do not specify the parameter space $Y$, yet, but we assume that for each $(x, y) \in X \times Y$, the volatility and drift coefficients are given by $a_y(x) = a_y \in [0, \infty)$ and $b_y(x) = \beta_y x$ with $\beta_y \in (-\infty, 0)$. For simplicity, we choose equidistant grids of times $t_n = n\delta$ and times to maturity $\tau_n = n\delta$, for all $n \in \mathbb{N}_0$, where $\delta$ is a positive constant.

### 4.3. Hull-White extended Vasicek models.

For each parameter $(y, \theta) \in Y \times C(\mathbb{R}_+)$, the SDE for the short rate process is

\begin{equation}
\begin{aligned}
    dr(t) &= (\theta(t) + \beta_y r(t))\, dt + \sqrt{a_y} \, dW(t),
\end{aligned}
\end{equation}

where $W$ is one-dimensional $(\mathcal{F}_t)_{t \geq 0}$-Brownian motion. Assumption 2.1 is satisfied for each parameter $(y, \theta)$. The functional characteristics $(F, R)$ from Section 2.4 are

\begin{equation}
\begin{aligned}
    F_y(u) &= \frac{a_y}{2} u^2, \\
    R_y(u) &= \beta_y u, \quad \forall u \in \mathbb{R},
\end{aligned}
\end{equation}

and the solutions $(\Phi_y, \Psi_y)$ of the corresponding Riccati equations are

\begin{equation}
\begin{aligned}
    \Phi_y(t) &= \frac{a_y}{4 \beta^3} \left( 2 \beta_y t - 4 e^{\beta_y t} + 3 + e^{2 \beta_y t} \right), \\
    \Psi_y(t) &= \frac{1}{\beta_y} \left( 1 - e^{\beta_y t} \right), \quad \forall t \geq 0.
\end{aligned}
\end{equation}
By Theorem 2.3, the forward rates in the Hull-White extended Vasiček model (4.1) with fixed parameters \((y, \theta)\) are given by \(h(t) = \mathcal{H}_y(r(t), S_t \theta)\), where 
\[
\mathcal{H}_y(x, \theta)(\tau) = \int_0^\tau \theta(s)e^{y_\beta(\tau-s)}ds - \frac{\alpha_y}{2y_\beta} (1 - e^{y_\beta \tau})^2 + e^{y_\beta \tau}x,
\]
for all \((x, \theta, \tau) \in \mathbb{R} \times C(\mathbb{R}_+) \times \mathbb{R}_+\). Due to the simple structure of the integral kernel \(e^{y_\beta(\tau-s)}\), there is a closed-form expression for the calibration operator, 
\[
C_y(h)(\tau) = h(\tau) - \beta_y h(\tau) - \frac{\alpha_y}{2y_\beta} (1 - e^{y_\beta \tau}), \quad \forall(h, \tau) \in C^1(\mathbb{R}_+) \times \mathbb{R}_+.
\]
This can be verified using the definitions. Note that the calibration operator does not depend on \(x\). Therefore, we dropped \(x\) from the notation \(C_y(h, x)\).

The HJM drift and volatility from Section 2.6 are
\[
\mu_y^{\text{HJM}}(\tau) = -\frac{\alpha}{\beta_y} e^{\beta_\gamma \tau} (1 - e^{\beta_\gamma s}) \quad \text{and} \quad \sigma_y^{\text{HJM}}(\tau) = \sqrt{\alpha_y} e^{\beta_\gamma s}, \quad \forall \tau \in \mathbb{R}_+.
\]
Note that these expressions do not depend on \(x\), which is why we dropped \(x\) from the previous notation \(\mu_y^{\text{HJM}}(x)(\tau), \sigma_y^{\text{HJM}}(x)(\tau)\). The HJM equation for forward rates then reads as
\[
dh(t) = \left(\mathcal{A} h(t) + \mu_y^{\text{HJM}}\right)dt + \sigma_y^{\text{HJM}}dW(t).
\]

4.3. Vasiček CRC models. Since the factor process is a function of the forward rate process (i.e., \(X(t) = r(t) = h(t, 0)\)), the corresponding CRC models can be characterised by the process \((h, Y)\) instead of \((h, X, Y)\). Thus, in accordance with Theorem 3.6 and Definition 3.7, a process \((h, Y)\) with values in \(\mathbb{H} \times \mathbb{Y}\) may be called a CRC model if \(h\) satisfies the SPDE
\[
dh(t) = \left(\mathcal{A} h(t) + \mu_{Y(t)}^{\text{HJM}}\right)dt + \sigma_{Y(t)}^{\text{HJM}}dW(t),
\]
with drift \(\mu_{Y(t)}^{\text{HJM}}\) and volatility \(\sigma_{Y(t)}^{\text{HJM}}\) defined in (4.3). Beyond the obvious requirement that these quantities are well-defined, for all \(t \in \mathbb{R}_+\), no further conditions are needed. In other words, the maximally admissible set \(Z\) in the Vasiček case is the entire Hilbert space \(\mathbb{H}\).

4.5. Simulation of Vasiček CRC models. Given a discrete-time process \(Y\) with values in \(\mathbb{Y}\), the CRC model is simulated as described in Algorithm 3.2. The following observations make the algorithm particularly effective. First, the state process \(X\) is a function of the forward rate and can be eliminated as a state variable. Second, the short rate process can be simulated exactly. Indeed, in the model with constant parameter \(y\), \(r(t)\) is normally distributed,
\[
r(t) \sim \mathcal{N}\left(e^{y_\beta \tau_0} + \int_0^t e^{y_\beta (t-s)}\theta(s)ds, \frac{\alpha_y}{2y_\beta} \left(e^{2y_\beta \tau} - 1\right)\right).
\]
Third, inverting the Volterra integral operator can be avoided by using closed-form expression (4.2) of the calibration operator.

Discretisation is done on the uniform grid \(t_n = \tau_n = \delta n\) for a choice of finitely many times to maturity \(\tau_n\). Integrals are approximated to second order by the trapezoid rule, which leads to a global error of order one (see Section 4.6 and Section 6.7). The resulting scheme works as follows.

Algorithm 4.1 (Simulation). Given \((h(0), \mathcal{A} h(0))\) and the discrete-time process \(Y\), execute iteratively the following steps, for each \(n \in \mathbb{N}_0\):
The values of \( \theta(t_n) = CY(t_n)(h(t_n)) \) at times to maturity 0 and \( \delta \) are calculated using (4.2),
\[
\theta(t_n)(0) = Ah(t_n)(0) - \beta_Y(t_n) h(t_n)(0),
\]
\[
\theta(t_n)(\delta) = Ah(t_n)(\delta) - \beta_Y(t_n) h(t_n)(\delta) - \frac{\alpha_Y(t_n)}{2\beta_Y(t_n)} (1 - e^{2\beta_Y(t_n)\delta}),
\]
and \( I_Y(t_n)(\theta(t_n))(\delta) \) is approximated by the trapezoid rule as follows:
\[
\tilde{I}_Y(t_n)(\theta(t_n))(\delta) = -\frac{\delta}{2} \left( e^{2\beta_Y(t_n)\delta} \theta(t_n)(0) + \theta(t_n)(\delta) \right).
\]

(i) A sample \( r(t_{n+1}) \) is drawn such that conditionally on \( F_{t_n} \), \( r(t_{n+1}) \) has normal distribution
\[
r(t_{n+1}) \sim \mathcal{N} \left( \left( e^{\beta_Y(t_n)\delta} h(t_n)(0) - \tilde{I}_Y(t_n)(\theta(t_n))(\delta) \right) \frac{\alpha_Y(t_n)}{2\beta_Y(t_n)} e^{2\beta_Y(t_n)\delta} - 1 \right) .
\]

(ii) A sample \( h(t_{n+1}) \) is calculated from \( h(t_n) \), \( Ah(t_n) \), \( r(t_{n+1}) \) using Lemma 3.3:
\[
h(t_{n+1})(\tau) = h(t_n)(\delta + \tau) + \frac{\alpha_Y(t_n)}{2\beta_Y(t_n)} \left( \left( 1 - e^{\beta_Y(t_n)(\delta + \tau)} \right) \right)^2 \left( 1 - e^{2\beta_Y(t_n)\tau} \right)^2
\]
\[
+ e^{\beta_Y(t_n)\tau} \left( -e^{\beta_Y(t_n)\delta} r(t_n) + r(t_{n+1}) + \tilde{I}_Y(t_n)(\theta(t_n))(\delta) \right),
\]
\[
Ah(t_{n+1})(\tau) = Ah(t_n)(\delta + \tau)
\]
\[
+ \frac{\alpha_Y(t_n)}{2\beta_Y(t_n)} \left( e^{\beta_Y(t_n)\tau} + e^{2\beta_Y(t_n)(\tau + \delta)} - e^{2\beta_Y(t_n)\tau} - e^{\beta_Y(t_n)(\delta + \tau)} \right)
\]
\[
+ \beta_Y(t_n) e^{\beta_Y(t_n)\tau} \left( -e^{\beta_Y(t_n)\delta} r(t_n) + r(t_{n+1}) + \tilde{I}_Y(t_n)(\theta(t_n))(\delta) \right).
\]

Here, \( h(t_{n+1}) \) must be calculated at all times to maturity \( \tau_i \), whereas \( Ah(t_{n+1}) \) is needed only at \( t_0 = 0 \) and \( \tau_1 = \delta \).

### 4.6. Convergence of the simulation scheme.
In this section, we show that discrete-time CRC models converge to continuous-time CRC models as the grid size of the discrete models tends to zero. We are not aiming for the highest generality. Instead, we show how the results follow from standard semigroup theory.

#### Assumption 4.2.
The parameter process \( Y \) takes values in \( Y = \mathbb{R}^p \) and satisfies
\[
dY(t) = \left( AY(t) + \mu(Y(t)) \right) dt + \sigma(Y(t))dW(t)
\]
where \( A : \mathbb{R}^p \rightarrow \mathbb{R}^p \) is a linear mapping generating a semigroup of contractions on \( \mathbb{R}^p \), \( \mu \in C^\infty_b(\mathbb{R}^p; \mathbb{R}^p) \), \( \sigma \in C^\infty_b(\mathbb{R}^p, \mathbb{R}^{p \times q}) \), and \( W \) is \( q \)-dimensional \( \mathcal{F}_t \)-Brownian motion, independent of \( W \). We write \( C^\infty_b \) for bounded functions with bounded derivatives of all orders. The above SDE has a unique solution for any initial condition \( Y(s) = y \), where \( (s, y) \in \mathbb{R}_+ \times Y \).

**Assumption 4.3.** The mappings \( y \mapsto \sqrt{\sigma_y} \) and \( y \mapsto \beta_y \) are of class \( C^\infty_b(\mathbb{R}^p) \) and \( \sup_{y \in Y} \beta_y < 0 \) holds.

As Vasicek CRC models can be characterised in terms of \( (h, Y) \) instead of \( (h, X, Y) \), the semigroups \( (P_t)_{t \geq 0} \), \( (Q_t)_{t \geq 0} \), and \( (R_t)_{t \geq 0} \) from Section 3.11 are now assumed to be defined on \( C_b(\mathbb{R} \times Y) \) instead of \( C_b(\mathbb{R} \times X \times Y) \). Recall that \( (P_t)_{t \geq 0} \) describes the joint evolution of the process \( (h, Y) \), \( (Q_t)_{t \geq 0} \) the evolution of \( h \) with \( Y \) fixed, and \( (R_t)_{t \geq 0} \) the evolution of \( Y \) with \( h \) fixed.
**Theorem 4.4.** There exists a Hilbert space $H$ of continuous functions on $\mathbb{R}_+$ and a Banach space $B$ of continuous functions on $\mathbb{R} \times \mathbb{R}$ such that $(P_t)_{t \geq 0}$, $(Q_t)_{t \geq 0}$, and $(R_t)_{t \geq 0}$ are strongly continuous semigroups on $B$. Moreover, for each $T \in \mathbb{R}_+$ there exists a constant $C$ such that

$$
\|P_t f - (R_{t/n} Q_{t/n})^n f\|_B \leq C n^{-1} \|f\|_B, \quad \forall f \in B', \ t \in [0, T], n \in \mathbb{N}^+,
$$

where $B'$ is a Banach space which is densely and continuously embedded in $B$.

The space $B'$ is large enough to be relevant in applications: any $C^4$ function on $H_0 \times Y$ belongs to $B'$, where $H_0 \supset H$ is defined in the proof below.

**Proof.** We proceed as in [29] and [32]. Let $(\gamma_t)_{t \in \mathbb{N}_0}$ be a strictly increasing sequence of real numbers strictly greater than $3$. For each $i \in \mathbb{N}_0$, define a separable Hilbert space $H_i$ by

$$
H_i = \left\{ h \in L^1_{loc}: h^{(j)} \in L^1_{loc} \text{ and } \int_{(0, \infty)} h^{(j)}(\tau) (1 + \tau) \gamma \ d\tau < \infty, \forall j = 1, \ldots, i \right\},
$$

where $L^1_{loc}$ denotes the space of locally integrable functions on $(0, \infty)$. Every function in $H_0$ is continuous, bounded and has a well-defined limit $h(\infty) = \lim_{\tau \to \infty} h(\tau)$. The scalar product on $H_i$ is defined by

$$
\langle h_1, h_2 \rangle_{H_i} = h_1(\infty) h_2(\infty) + \sum_{j=1}^{i} \int_{(0, \infty)} h_1^{(j)}(\tau) h_2^{(j)}(\tau) (1 + \tau) \gamma \ d\tau.
$$

For each $0 < \zeta > 0$ and $i \in \mathbb{N}_0$, we define the space $H^\zeta_i(\mathbb{H}_i \times \mathbb{Y})$ as the closure of $C_0^\infty(\mathbb{H}_i \times \mathbb{Y})$ under the norm

$$
\|f\|_{H^\zeta_i(\mathbb{H}_i \times \mathbb{Y})} = \sup_{j=0}^{k} \|h\|_{H_i} \langle \zeta \|h\|_{H_i}, \|y\|_{\mathbb{Y}}^2 \rangle^{-1/2} \|D^j f(h, y)\|_{L^1(\mathbb{H}_i \times \mathbb{Y})}.
$$

Together with Assumption 4.2 and 4.3, this implies that the conditions of [29, Sections 3.1.1 and 3.1.2] are satisfied for SPDE (4.5), (4.6) characterising the evolution of $(h, Y)$. (Note that $\beta_y$ needs to be bounded away from zero for $\mu^H_{y}$ and $\mu^{H\text{HM}}_{y}$ to be bounded with bounded derivatives.) Thus, this SPDE admits unique solutions on each space $H_i \times \mathbb{Y}$, given that the initial condition is smooth enough. The same applies to the SPDE for $h$ with fixed $y$ and the SDE for $Y$ with fixed $y$.

Fix $\zeta > 0$ and define $H = H_2$, $B = B^\zeta_0(\mathbb{H}_2 \times \mathbb{Y})$, and $B' = B^\zeta_0(\mathbb{H}_0 \times \mathbb{Y})$. Then $(P_t)_{t \geq 0}$, $(Q_t)_{t \geq 0}$, and $(R_t)_{t \geq 0}$ are strongly continuous semigroups on $B$ by [29, Lemma 13] and quasicontinutive by [29, Lemma 7]. Their generators are denoted by $G^P$, $G^Q$, and $G^R$. By the same lemma, $B'$ is stable under $(P_t)_{t \geq 0}$. Together with [29, Theorem 11] this implies that for each $f \in B'$, the expressions

$$
G^P P_t f, \quad G^Q P_t f, \quad G^R P_t f, \quad G^Q G^Q P_t f, \quad G^Q G^R P_t f, \quad G^R G^Q P_t f
$$

are well-defined with $\mathbb{B}$-norm bounded uniformly in $t \in [0, T]$ and $G^P f = G^Q f + G^R f$.

Thus, the splitting is of formal order one and the result follows from [41, Theorem 2.3 and Section 4.4].
4.7. Consistent recalibration property. If the coefficient $\beta$ in the HJM volatility $\sqrt{\sigma \phi}$ is stochastic, one would expect the forward rate process to reach every point in the Hilbert space with positive probability, i.e., the consistent recalibration property holds. This is made precise here. Assumptions 4.3 and 4.2 remain in place. The components of $\sigma$ are written as $\sigma = (\sigma_1, \ldots, \sigma_q)$, where $\sigma_i : \mathcal{Y} \to \mathbb{Y}$.

**Lemma 4.5.** Assume that for all $y \in \mathcal{Y}$, $\beta'(y)\sigma_i(y) \neq 0$ holds for some $i \in \{1, \ldots, q\}$. Then the consistent recalibration property is satisfied for the Vasiček CRC model (4.3), (4.6) with respect to the state space $\mathcal{I} = \mathbb{H}$.

**Proof.** The mappings $0 \times \sigma_i$ and $\sigma_{\text{HJM}} \times 0$ are vector fields on $\mathbb{H} \times \mathcal{Y}$. The Lie derivative is denoted by $\mathcal{L}$. For any $(h, y) \in \mathbb{H} \times \mathcal{Y}$ and $i \in \{1, \ldots, q\}$ such that $\beta'(y)\sigma_i(y) = \mathcal{L}_{\sigma_i, \beta}(y) \neq 0$, evaluating the vector field

$$\mathcal{L}_{(0 \times \sigma_i)}(\sigma_{\text{HJM}} \times 0) = (d\sigma_{\text{HJM}} \sigma_i) \times 0 = (\tau \mapsto \tau e^{\beta \tau} \mathcal{L}_{\sigma_i, \beta}) \times 0$$

at $(h, y)$ yields a non-zero vector in $V_1 \subset \mathbb{H}$, where

$$V_n = \text{span}_\mathbb{R} \{ (\tau \mapsto \tau^k e^{\beta \tau}) : k = 0, \ldots, n \} \subset \mathbb{H}, \quad \forall n \in \mathbb{N}_0.$$ 

It can be shown by induction that

$$\mathcal{L}_{(0 \times \sigma_i)}^n(\sigma_{\text{HJM}} \times 0) \equiv (\tau \mapsto \tau^n e^{\beta \tau} (\mathcal{L}_{\sigma_i, \beta})^n) \times 0 + \cdots,$$

where the remaining summands are functions on $\mathcal{Y}$ times vector fields $(\tau \mapsto \tau^k e^{\beta \tau}) \times 0$ with $k < n$. It follows from the condition $\mathcal{L}_{\sigma_i, \beta}(y) \neq 0$ that

$$\text{span}_\mathbb{R} \{ \mathcal{L}_{(0 \times \sigma_i)}^k(\sigma_{\text{HJM}} \times 0)(h, y) : k = 0, \ldots, n \} = V_n \times 0 \subset \mathbb{H} \times \mathcal{Y}.$$ 

All iterated Lie brackets with more than one appearance of $\sigma_{\text{HJM}} \times 0$ vanish because $\sigma_{\text{HJM}}(h, y)$ does not depend on $h \in \mathbb{H}$. Therefore, the iterated Lie brackets of the vector fields $\sigma_{\text{HJM}} \times 0 \times \sigma_1, \ldots, 0 \times \sigma_q$, evaluated at $(h, y)$, span the space $\bigcup_n V_n \times \mathcal{Y}$, for some subspace $\mathcal{Y}' \subseteq \mathcal{Y}$. As $\bigcup_n V_n$ is dense in $\mathbb{H}$, Hörmander’s condition is satisfied. By [5, Theorem 1], the forward rate evolution is irreducible on $\mathbb{H}$ in the sense that every finite dimensional distribution of forward rates $h(t)$, $t > 0$, admits a density with respect to the Lebesgue measure. \(\square\)

4.8. **Example.** We present an example of a Vasiček CRC model based on [17]. 

In this model, the volatility is stochastic, but the speed of mean reversion is not. Therefore, the conditions of Lemma 4.5 are not satisfied, and it turns out that the model admits a finite-dimensional realisation. The explicit formula for bond prices in this model will serve as a reference for showing convergence of the numerical simulation scheme for CRC models to the continuous-time limit.

The parameter process $Y$ is a CIR process with values in $\mathcal{Y} = \mathbb{R}_+$ given by SDE (4.6) with coefficients $\mu(y) = m + \mu y$ and $\sigma(y) = \sigma y$ for some $m \geq 0, \mu \leq 0$, and $\sigma \geq 0$. The Vasiček drift and volatility in the HJM equation are given by equation (4.3) with $\beta_y = \beta$ for some constant $\beta < 0$ and $a_y = y$, i.e.,

$$\mu_{\text{HJM}}(\tau) = -\frac{y}{\beta} e^{\beta \tau} (1 - e^{-\beta \tau}), \quad \sigma_{\text{HJM}}^2(\tau) = \sqrt{\beta} e^{\beta \tau}, \quad \forall \tau \in \mathbb{R}_+.$$
If $h(0) \in C^1(\mathbb{R}_+)$, there is a closed-form solution of CRC equation (4.4),

$$h(t, \tau) = h(0, t + \tau) - \int_0^t Y(s) e^{\beta(t+\tau-s)} \left(1 - e^{\beta(t+s)}\right) ds + \int_0^t \sqrt{Y(s)} e^{\beta(t+s)} dW(s).$$

Setting $\xi(t) = \int_0^t Y(s) e^{2\beta(t-s)} ds$ and $r(t) = h(t, 0)$, this can be rewritten as

$$h(t, \tau) = h(0, t + \tau) + e^{\beta\tau} \left(r(t) - h(0, t)\right) + \frac{1}{\beta} \left(e^{2\beta\tau} - e^{\beta\tau}\right) \xi(t).$$

Setting $\tau = 0$ in equation (4.4) and plugging in equation (4.7) yields

$$dr(t) = (Ah(0, t) - \beta h(0, t) + \beta r(t) + \xi(t)) dt + \sqrt{Y(t)} dW(t).$$

Summing up, the process $X = (r, \xi, Y)$ is given by the SDE

$$
\begin{align*}
  dr(t) &= \left(\frac{\partial h}{\partial \tau}(0, t) - \beta h(0, t) + \beta r(t) + \xi(t)\right) dt + \sqrt{Y(t)} dW(t), \\
  d\xi(t) &= (2\beta \xi(t) + Y(t)) dt, \\
  dY(t) &= (m + \mu Y(t)) dt + \sigma \sqrt{Y(t)} d\tilde{W}(t),
\end{align*}
$$

where $h(0) \in C^1(\mathbb{R}_+)$ is a given initial forward rate curve. It follows that $X$ is an affine factor process for the short rate as described in Section 2 with $d = 3, \ell = 0, \lambda = (1, 0, 0)^T$. Thus, the CRC model has a finite-dimensional realisation. If $\sigma = 0$, the affine bond pricing formula is particularly simple: bond prices are given by

$$P(t, T) = e^{\int_0^T (e^{\beta(t-s)} h(0, t) - h(0, s)) ds + \beta^{-1}(1-e^{\beta t}) r(t) - \frac{1}{2} \beta^{-2} (1-e^{\beta t})^2 \xi(t)},$$

where $\xi(t)$ is deterministic and satisfies

$$\xi(t) = \begin{cases} 
  Y(0) e^{2\beta t} - e^{\mu t} + \frac{m(2\beta - \mu - 2\beta e^{\mu t} + \mu e^{2\beta t})}{2\beta \mu (2\beta - \mu)}, & \text{if } \mu < 0, \\
  Y(0) e^{2\beta t} - 1 + \frac{m(e^{2\beta t} - 2\beta t - 1)}{4\beta^2}, & \text{if } \mu = 0,
\end{cases}$$

and $r(t)$ is normally distributed with mean

$$e^{\beta t} r(0) + \int_0^t e^{\beta(t-s)} (Ah(0, s) - \beta h(0, s) + \xi(s)) ds,$$

and variance

$$\frac{Y(0)}{2\beta} (e^{2\beta t} - 1) + \frac{m}{4\beta^2} (-2\beta t + e^{2\beta t} - 1).$$

4.9. Calibration of Vasiček CRC models. As outlined in Section 3.13, we first regard $y$ as fixed and suppress the dependence on $y$ in the notation. For any selection of times to maturity $\tau_i, \tau_j$, estimators for $a, \beta$ can be obtained as described in Section 2.9 by solving for those $\hat{a}, \hat{\beta}$ which achieve the best fit in (2.12), i.e.,

$$\frac{[\hat{r}(\cdot, \tau_i) - \hat{r}(\cdot, \tau_j)](t_n) - [\hat{r}(\cdot, \tau_i) - \hat{r}(\cdot, \tau_j)](t_{n-M})}{t_n - t_{n-M}} \approx \hat{a} \frac{e^{\hat{\beta} \tau_i} - 1}{\beta \tau_i} - \frac{1}{\hat{\beta} \tau_i}.$$
Varying the calibration time $t_n$ creates a time series of coefficients $\tilde{\alpha}(t_n), \tilde{\beta}(t_n)$ for which we need to specify and calibrate a model. Some models are described in Section 6.5, below.

5. Cox-Ingersoll-Ross short rate model

5.1. Overview. We describe CRC models based on CIR short rates, giving a detailed description of the simulation and calibration schemes. For comparison, we briefly digress to the CIR++ model and its CRC version.

5.2. Setup and notation. We use the setup of Sections 2.2 and 3.2, setting $X = \mathbb{R}_+, \ell = 0, \lambda = 1$. We do not specify the parameter space $\mathcal{Y}$, yet, but we assume that for each $(x, y) \in X \times \mathcal{Y}$, the volatility and drift coefficients are given by $a_y(x) = \alpha_y x$ and $b_y(x) = \beta_y x$ for some $\alpha_y \in (0, \infty)$ and $\beta_y \in (-\infty, 0)$. For simplicity, we again choose equidistant grids of times $t_n = n\delta$ and times to maturity $\tau_n = n\delta$, for all $n \in \mathbb{N}_0$, where $\delta$ is a positive constant.

5.3. Hull-White extended Cox-Ingersoll-Ross models. For each fixed set of parameters $(y, \theta) \in \mathcal{Y} \times C(\mathbb{R}_+)$, the SDE for the short rate process is

$$dr(t) = (\theta(t) + \beta_y r(t))dt + \sqrt{\alpha_y r(t)}dW(t),$$

where $W$ is one-dimensional $(\mathcal{F}_t)_{t \geq 0}$-Brownian motion. Thus, $(y, \theta)$ satisfies Assumption 2.1 if and only if $\theta(t) \geq 0$, for all $t \in \mathbb{R}_+$. The functional characteristics $(F, R)$ from Section 2.4 are

$$F_y(u) = 0, \quad R_y(u) = \frac{\alpha_y}{2} u^2 + \beta_y u, \quad \forall u \in \mathbb{R}.$$

Letting $\gamma_y = \sqrt{\beta_y^2 + 2\alpha_y}$, the solutions of the corresponding Riccati equations are

$$\Phi_y(t) = 0, \quad \Psi_y(t) = \frac{-2(e^{\gamma_y t} - 1)}{\gamma_y (e^{\gamma_y t} + 1) - \beta_y (e^{\gamma_y t} - 1)}, \quad \forall t \geq 0.$$

Thus, by Theorem 2.3, the forward rates in the Hull-White extended CIR model (5.1) with fixed parameters $(y, \theta)$ are given by $h(t) = H_y(r(t), \theta)$, where

$$H_y(x, \theta)(\tau) = -\int_0^{\tau} \theta(s)x'(\tau - s)ds - \Psi_y'(\tau)x, \quad \forall (x, \tau) \in \mathbb{R} \times \mathbb{R}_+.$$

In contrast to the Vasiček model, the integral kernel $\Psi_y'$ is more complicated,

$$\Psi_y'(\tau) = \frac{-2\gamma_y e^{\gamma_y \tau}}{\gamma_y (e^{\gamma_y \tau} + 1) - \beta_y (e^{\gamma_y \tau} - 1)} + \frac{2(e^{\gamma_y \tau} - 1)}{(\gamma_y (e^{\gamma_y \tau} + 1) - \beta_y (e^{\gamma_y \tau} - 1))^2},$$

and there does not seem to be a closed-form expression for $\theta = C_y(h, x)$. Instead, it must be calculated numerically as described in Section 2.8.

The HJM drift and volatility are

$$\mu_y^{\text{HJM}}(x)(\tau) = \Psi_y'(\tau)x, \quad \sigma_y^{\text{HJM}}(x)(\tau) = -\sqrt{\alpha_y} x \Psi_y'(\tau),$$

and the HJM equation for forward rates reads as

$$dh(t) = (Ah(t) + \mu_y^{\text{HJM}}(h(t)(0)))dt + \sigma_y^{\text{HJM}}(h(t)(0))dW(t).$$
5.4. Cox-Ingersoll-Ross CRC models. As the factor process is a function of the forward rate process (i.e., \(X(t) = r(t) = h(t, 0)\)), the corresponding CRC models can be characterised by the process \((h, Y)\) instead of \((h, X, Y)\). Thus, in accordance with Theorem 3.6 and Definition 3.7, a process \((h, Y)\) with values in \(\mathbb{H} \times \mathbb{Y}\) may be called a CRC model if \(h\) satisfies the SPDE

\[
\frac{dh(t)}{dt} = \left( A h(t) + \mu_Y^{\text{HJM}}(h(t)(0)) \right) dt + \sigma_Y^{\text{HJM}}(h(t)(0)) dW(t),
\]

with drift \(\mu_Y^{\text{HJM}}\) and volatility \(\sigma_Y^{\text{HJM}}\) as defined in equation (5.3). To ensure that the drift and volatility are well-defined, for all \(t \in \mathbb{R}_+\), it must be assumed that \(h(t)(0) \geq 0\) holds, for all \(t \in \mathbb{R}_+\). The maximally admissible set \(I\) in the CIR model is exactly characterised by this condition.

5.5. Simulation of Cox-Ingersoll-Ross CRC models. The CRC model is simulated as described in Algorithm 3.2. Discretisation in time and time to maturity is done as for the Vasicek model. However, in contrast to the Vasicek model, simulating the short rate process and calibrating Hull-White extensions is done by numerical approximations of order two. The resulting algorithm is presented below.

**Algorithm 5.1** (Simulation). Given an initial curve of forward rates \(h(0)\) and the discrete-time process \(Y\), execute iteratively the following steps, for each \(n \in \mathbb{N}_0\):

(i) The values of \(\theta(t_n) = C_Y(t_n)(h(t_n), r(t_n))\) at times to maturity 0 and \(\delta\) are calculated by applying Lemma 2.8 to \(g = -h(t_n) - \Psi_Y'(t_n) r(t_n)\):

\[
\theta(t_n)(0) = Ah(t_n)(0) - \beta_Y(h(t_n)(0)),
\]

\[
\theta(t_n)(\delta) = \frac{2}{\delta} \left( h(t_n)(\delta) + \Psi_Y'(t_n)(\delta) r(t_n) \right) + \Psi_Y'(t_n)(\delta) \theta(t_n)(0).
\]

(ii) A random draw \(r(t_{n+1}) = X_{Y(t_n), \theta(t_n)}(t_{n+1})\) of the CIR process with parameter \(Y(t_n)\) and time-dependent drift \(\theta(t_n)\) is created using the second-order scheme of [1].

(iii) \((h(t_{n+1}), Ah(t_{n+1}))\) is calculated from \((h(t_n), Ah(t_n), r(t_{n+1}))\) using Lemma 3.3 with integrals approximated by the trapezoid rule:

\[
h(t_{n+1})(\tau) = h(t_n)(\delta + \tau) + \Psi_Y'(t_n)(\delta + \tau) r(t_n) - \Psi_Y'(t_n)(\tau) r(t_{n+1}) + \frac{\delta}{2} \left( \theta(t_n)(0) \Psi_Y'(t_n)(\delta + \tau) + \theta(t_n)(\delta) \Psi_Y'(t_n)(\tau) \right),
\]

\[
Ah(t_{n+1})(\tau) = Ah(t_n)(\delta + \tau) + \Psi_Y'(t_n)(\delta + \tau) r(t_n) - \Psi_Y'(t_n)(\tau) r(t_{n+1}) + \frac{\delta}{2} \left( \theta(t_n)(0) \Psi_Y'(t_n)(\delta + \tau) + \theta(t_n)(\delta) \Psi_Y'(t_n)(\tau) \right).
\]

Here, \(h(t_{n+1})\) must be calculated at all times to maturity \(\tau\), whereas \(Ah(t_{n+1})\) is needed only at \(\tau_0 = 0\).

5.6. Calibration of Cox-Ingersoll-Ross CRC models. We proceed as in the Vasicek case described in Section 4.9, with equation (4.11) replaced by

\[
\frac{\hat{\tau}(\cdot, \tau_i) \cdot \hat{\tau}(\cdot, \tau_j)}{\delta \sum_{m = M - M + 1} \hat{\tau}(m, \tau_k)} \frac{\hat{\tau}(\cdot, \tau_i)}{\tau_i} = \alpha \frac{\Psi(\tau_i)}{\tau_i} \frac{\Psi(\tau_j)}{\tau_j}.
\]

The function \(\Psi\) depends on \(\alpha, \beta\) as shown in equation (5.2).
5.7. **CIR++ models in the CRC framework.** In the CIR++ model [10, Section 3.9], also known as deterministic shift-extended CIR model, the short rate process is defined by \( r(t) = X(t) + \theta(t) \), where \( X \) is a CIR process and \( \theta \) is a deterministic function of time. Note that this is a different time-inhomogeneity than the one described in Section 5.3. In particular, the factor process \( X \) is time-homogeneous and does not coincide with the short rate.

Forward rate curves are given by

\[
h(t) = S_t \theta - b_y \Psi_y - \Psi'_y X(t),
\]

where \( \Psi_y \) is the same as in the CIR case, see equation (5.2). Given the parameter vector \( y \) and the factor \( X \), this equation allows to calibrate \( \theta \) to a given yield curve without having to invert a Volterra integral operator. The HJM equation of the CIR++ model is

\[
\begin{align*}
\frac{dh(t)}{dt} &= (Ah(t) + \mu^\text{HJM}_y(X(t))) dt + \sigma^\text{HJM}_y(X(t)) dW(t), \\
\frac{dX(t)}{dt} &= (b_y + \beta_y X(t)) dt + \sqrt{\alpha_y} X(t) dW(t),
\end{align*}
\]

(5.7)

where \( \mu^\text{HJM}_y \) and \( \sigma^\text{HJM}_y \) are the same as in the CIR case, see equation (5.3).

The CRC extension of the CIR++ model is obtained by replacing the constant parameter vector \( y \) in (5.7) by a stochastic process \( (Y(t))_{t \geq 0} \). The resulting equation is easier to handle than its CIR counterpart for two reasons. First, there are no boundary conditions on \( h \). Indeed, \( \theta \) is allowed to assume negative values and can be calibrated to any forward rate curve. Thus, equation (5.7) is defined on the entire space \( \mathbb{H} \times \mathbb{R}_+ \). Second, the SDE for \( X \) does not depend on \( h \). Therefore, one can first solve for \( X \), and then construct a mild solution \( h \) by stochastic convolution [24, Section 6.1]:

\[
h(t) = S_t h(0) + \int_0^t S_{t-s} \mu^\text{HJM}_{Y(s)}(X(s)) ds + \int_0^t S_{t-s} \sigma^\text{HJM}_{Y(s)}(X(s)) dW(s).
\]

The SDE for \( X \) is finite-dimensional. Therefore, existence and uniqueness of \( X \) can be shown by standard methods. For example, assuming that \( Y \) is independent of \( W \), one can condition on \( Y \) and use results on time-inhomogeneous affine processes [34] to construct \( X \).

Simulation of the CRC model is analogue to Algorithm 3.2. The recalibration step is easier because no Volterra equation is involved.

A disadvantage of the model is the presence of the hidden factor \( X \). In contrast to the CIR version, \( X \) is not a function of the forward rate curve and cannot be directly observed. This is a challenge for calibration. We suggest an analogue approach to Section 3.13. First, \( \beta_{Y(t)}, \sigma_{Y(t)} \), and \( X(t) \) can be identified from the instantaneous covariation

\[
d[r(\cdot, \tau_i), r(\cdot, \tau_j)](t) = \alpha_{Y(t)} \frac{\Psi_{Y(t)}(\tau_i)}{\tau_i} \frac{\Psi_{Y(t)}(\tau_j)}{\tau_j} X(t) dt
\]

of yields with times to maturity \( \tau_i, \tau_j \). Subsequently, \( b_{Y(t)} \) can be calibrated by least squares to the prevailing yield curve. Note that in this approach, \( X(t) \) is identified from the yield curve dynamics instead of extracted from the prevailing yield curve as in the Vasicek and CIR cases. For this reason the calibration is expected to be numerically more difficult.
6. Empirical results

6.1. Overview. CRC models based on Vasicek and CIR short rates are calibrated to Euro area yield curves. Properties of the calibrated models are studied in comparison to market data and models without consistent recalibration. Our empirical findings show that the assumption of constant parameters in the Vasicek and CIR models is too restrictive. Therefore, the additional flexibility provided by CRC models is a useful tool to better capture the market dynamics. This is also reflected in better fits of the covariance matrix of yields.

All figures are shown at the end of the section.

6.2. Description of the data. We consider the zero-coupon yield curves released by the European Central Bank (ECB) on a daily basis. The yields are estimated from AAA-rated (Fitch Ratings) Euro area central government bonds being actively traded on the market. Estimation is done by the ECB using the Svensson family of curves, see [76, 78]. Data is available from September 6, 2004, and we set April 1, 2014 to be the last observation date. In total, this results in 2454 observed yield curves with times to maturity ranging from 3 months up to 30 years. Yields are continuously compounded (c.f. equation (2.11)) and denoted by \( \tilde{r}(t, \tau) \), with \( \tau \) being the time to maturity. A selection of yields is shown in Figures 6.1 and 6.2. The short rate is approximated by the yield with the lowest time to maturity (3 months) continuously compounded (c.f. equation (2.11)) and denoted by \( \tilde{r}(t) \), one can solve (4.11) and (5.6) for \( \hat{\alpha} \) and obtain the estimator

\[
\hat{\alpha}(t) = \frac{[\tilde{r}(t), \tilde{r}(t, \tau_1)](t) - [\tilde{r}(t), \tilde{r}(t, \tau_1)](t - \delta M)}{\delta M},
\]

in the Vasicek case, and

\[
\hat{\alpha}(t) = \frac{[\tilde{r}(t, \tau_1), \tilde{r}(t, \tau_1)](t) - [\tilde{r}(t, \tau_1), \tilde{r}(t, \tau_1)](t - \delta M)}{\delta \sum_{m=0}^{M-1} \tilde{r}(t - \delta m, \tau_1)},
\]

in the CIR case, where the quadratic variation is estimated by (4.11). On the other hand, taking \( \tau_2 \gg 1 \), one can solve (4.11) and (5.6) for \( \beta \) and obtain the estimator

\[
\hat{\beta}(t) = -\frac{1}{\tau_2} \left( \frac{\delta M \hat{\alpha}(t)}{[\tilde{r}(t, \tau_2), \tilde{r}(t, \tau_2)](t) - [\tilde{r}(t, \tau_2), \tilde{r}(t, \tau_2)](t - \delta M)} \right)^{\frac{1}{2}},
\]

in the Vasicek case, and

\[
\hat{\beta}(t) = \frac{\sqrt{\hat{\alpha}(t)}}{\tau_2} \left( \frac{\delta M \hat{\alpha}(t)}{\delta \sum_{m=0}^{M-1} \tilde{r}(t - \delta m, \tau_1)} \right)^{\frac{1}{2}},
\]

\[
\hat{\beta}(t) = \frac{\sqrt{\hat{\alpha}(t)}}{\tau_2} \left( \frac{\delta M \hat{\alpha}(t)}{\delta \sum_{m=0}^{M-1} \tilde{r}(t - \delta m, \tau_1)} \right)^{\frac{1}{2}},
\]

in the CIR case. The resulting trajectories of the estimated Vasicek volatility \( \sqrt{\hat{\alpha}} \) and CIR volatility \( \sqrt{\hat{\alpha}} \) are shown in Figures 6.4 and 6.5. The trajectories of the speed of mean reversion \( -\beta \) for both models are plotted in Figures 6.6 and 6.7. It
turns out that for most parts of the data, \( a \) and \( \alpha \) do not depend much on the times to maturity used in the estimation, whereas \( \beta \) does. Typically, smaller times to maturity result in larger values of \(-\beta\), as shown in Figures 6.6 and 6.7. This means that one-factor Vasiček and CIR models are not flexible enough to reproduce the ECB yield curve movements in full accuracy, and so the choice of times to maturity used in the calibration procedure may have an impact on the results. In particular we set \( \tau_1 = 0.25 \) and \( \tau_2 = 2 \) (i.e. 3 months and 2 years).

The dependence of the estimated parameter \( \beta \) on the choice of times to maturity suggests to use multi-factor models as building blocks for CRC models. In the empirical part of this paper, we aim for a detailed understanding of the one-factor case and leave the extension to multiple factors for future research.

By the theory of Hull-White extensions, an exact match to the initial yield curve is always achieved (Figures 6.8, 6.9, 6.10, and 6.11). However, the corresponding time-homogeneous models often do not match the initial yield curve well (Figures 6.10 and 6.11). This is not surprising as they are calibrated to the yield curve dynamics and not to the initial yield curve. This separation of dynamics and initial calibration is actually one of the strengths of our approach.

To model the dynamics of the Vasiček coefficients \( a, \beta \) and the CIR coefficients \( \alpha, \beta \), we use geometric Brownian motions and/or CIR processes, as laid out in Section 6.5. Note that the assumption of constant coefficients, which is implicit in affine factor models without the CRC extension, is not realistic over long time horizons in view of Figures 6.4 and 6.5.

6.4. Negative levels of mean reversion. A problematic aspect is that the time-dependent drift \( \theta \) obtained by the calibration to the initial yield curve can assume negative values, which are not admissible in the CIR model. The problem occurs mostly in low interest rate scenarios with partially inverted yield curves (Figures 6.10, 6.11, and 6.12). In contrast, negative values of \( \theta \) are allowed in the Vasiček model and might even be desirable for modelling bond markets with negative interest rates.

Since only the short (left) end of \( \theta \) is relevant for CRC models, at each step of the simulation scheme (cf. Algorithms 4.1 and 5.1), it is sufficient to understand how \( \theta(0) \) depends on the prevailing forward rate curve and the coefficients of the affine factor process. The general formula for \( \theta(0) \) is

\[
\theta(0) = C_y(h,x)(0) = \frac{1}{\langle \lambda,e_1 \rangle} \left( Ah(0) - F_y'(0) \cdot \lambda - \langle R_y'(0) \cdot \lambda,x \rangle \right),
\]

which follows by differentiating the relation \( h = H_y(x,\theta) \) with respect to the time to maturity \( \tau \) and evaluating at \( \tau = 0 \). In the Vasiček and CIR case, the formula becomes \( \theta(0) = Ah(0) - \beta y h(0) \). This shows that the problem can be alleviated to some extent by artificially choosing higher levels of \(-\beta\), resulting in higher values of \( \theta(0) \). For this reason we set \( \tau_2 = 2 \) instead of higher values of \( \tau_2 \) in the calibration of \( \beta \) (Figures 6.6 and 6.7).

6.5. Models for parameter evolutions. There are very few restrictions on the choice of parameter process. It can be chosen exogenously for scenario based simulation or calibrated to the market, and it is not restricted by the fit to the initial term structure.

We consider here four models for the evolution of the Vasiček and CIR parameters: a reference model where the parameters are constant, two toy models with constant
mean reversion and time-varying volatility, and one fully stochastic model which is calibrated to the market. In the Vasiček case, the four models are:

(V1) a Hull-White extended Vasiček model with constant coefficients \( \beta_y = \beta_0 \) and \( a_y = a_0 \);

(V2) a Vasiček CRC model with constant mean reversion coefficient \( \beta_y = \beta_0 \) and deterministically increasing volatility given by \( a_y = a_0y, Y(t) = 1 + 3t \);

(V3) a Vasiček CRC model with constant mean reversion coefficient \( \beta_y = \beta_0 \) and stochastically increasing volatility given by \( a_y = y, dY(t) = (4a_0 - Y(t))dt + \sigma \sqrt{Y(t)}d\tilde{W}(t), Y(0) = a_0, \sigma = 3 \cdot 10^{-3} \); and

(V4) a Vasiček CRC model with stochastic coefficients given by geometric Brownian motion: \( \beta_y = y_1, a_y = y_2, dY_1 = \mu_1Y_1(t) + \sigma_1Y_1(t)d\tilde{W}_1(t), dY_2 = \mu_2Y_2(t) + \sigma_2Y_2(t)d\tilde{W}_2(t) \). The coefficients \( \mu_{1,2} \) and \( \sigma_{1,2} \) are deterministic and calibrated to \( M \) observations as described in Section 6.3.

Note that in the model (V2), the volatility coefficient \( a \) at time \( t = 1 \) equals four times its initial value. This is comparable to (V3), where the level of mean reversion of \( a \) is set to four times its initial value.

Models (V2) and (V3) are special cases of Section 4.8. In (V2), which corresponds to \( m = 3a_0, \mu = 0, \) and \( \sigma = 0 \), there is an explicit formula for the moment generating function of the short rate process. By equations (4.9) and (4.10), it is given by (6.5)

\[
E[e^{\eta Y(t)}] = e^{e^{2\eta t}a_0 + \eta \int_0^t e^{2\eta s} (b_0(s) - \beta a_0(s) + \xi(s))ds + \frac{1}{2} \xi(t)}, \quad \forall \eta \in \mathbb{R}, t \in \mathbb{R}_+,
\]

where

\[
\xi(t) = a_0 e^{2\beta_0 t} - 1 + \frac{3a_0 (e^{2\beta_0 t} - 2\beta_0 t - 1)}{4\beta_0^2}.
\]

Model (V3) corresponds to Section 4.8 with \( m = 4a_0, \mu = -1 \) and \( \sigma = 3 \cdot 10^{-3} \).

The semigroup approach of Theorem 4.4 implies convergence of the simulation scheme for (V2) and for (V4) with \( Y_1 \) replaced by \( Y_1 + \epsilon \) for some \( \epsilon > 0 \). In our numerical simulations, we observe convergence for all models (see Section 6.7), including the following CIR counterparts of the models just described:

(CIR1) a Hull-White extended Cox-Ingersoll-Ross model with constant coefficients \( \beta_y = \beta_0 \) and \( a_y = a_0 \);

(CIR2) a Cox-Ingersoll-Ross CRC model with constant mean reversion coefficient \( \beta_y = \beta_0 \) and deterministically increasing volatility given by \( a_y = a_0y, Y(t) = 1 + 3t \);

(CIR3) a Cox-Ingersoll-Ross CRC model with constant mean reversion coefficient \( \beta_y = \beta_0 \) and stochastically increasing volatility given by \( a_y = y, dY(t) = (4a_0 - Y(t))dt + \sigma \sqrt{Y(t)}d\tilde{W}(t), Y(0) = a_0, \sigma = 5 \cdot 10^{-2} \); and

(CIR4) a Cox-Ingersoll-Ross CRC model with stochastic coefficients given by geometric Brownian motion: \( \beta_y = y_1, a_y = y_2, dY_1 = \mu_1Y_1(t) + \sigma_1Y_1(t)d\tilde{W}_1(t), dY_2 = \mu_2Y_2(t) + \sigma_2Y_2(t)d\tilde{W}_2(t) \). The coefficients \( \mu_{1,2} \) and \( \sigma_{1,2} \) are deterministic and calibrated to \( M \) observations as described in Section 6.3.

6.6. Implementation. Simulating CRC models requires iterative sampling of the underlying affine short rate process and recalibrating Hull-White extensions. In
the case of Vasicek and Cox-Ingersoll-Ross CRC models, this is explained in detail in Algorithms 4.1 and 5.1, respectively. The algorithms can be parallelised on a path-by-path level. Parallelisation on lower levels does not pay off because the individual time steps are dependent on each other. In our implementation in R, generating $10^5$ paths with 240 time steps on a cluster of 48 times 2.2GHz processors takes around 10 minutes in the Vasicek case and 20 minutes in the CIR case.

6.7. **Convergence analysis.** Theorem 4.4 predicts first order convergence of the simulation scheme under suitable assumptions on the model. The objective of this section is to demonstrate this convergence in numerical examples for the models described in Section 6.5.

The simulation is started with the initial forward rate curve $h_0$ of September 2, 2013. The parameters $\beta_0, a_0, \alpha_0, \theta_0$ are calibrated as in Section 6.3 with a time window of $M = 100$ observations. Then the moment generating function of the short rate $r(1)$ after one year is calculated by Monte Carlo simulation with $10^5$ sample paths. In the model (V2), the exact value of the moment generating function is known and given by equation (6.5). In the other models, a reference value is calculated by extrapolation from the Monte Carlo estimates. The resulting errors are shown in Figures 6.13, 6.14, and 6.15. As expected from Theorem 4.4 we observe first order convergence for models (V2) and (V4). The errors in Figure 6.15 indicate convergence also in the CIR counterparts.

6.8. **Distributional properties.** Making parameters in the CIR and Vasicek model stochastic in the sense of CRC models has considerable impact on the distribution of short rates and prices. As an example, statistics of the short rate $r(1)$ obtained by simulation are presented in Table 6.1. In the models (V1) and (V2) with deterministic parameters, the short rate process is Gaussian (see Section 4.8). As expected, the simulations show skewness and excess kurtosis values close to zero. In contrast, leptokurtosis appears in the models (V3) and (V4) with stochastic parameters. In the CIR examples, the distribution of $r(1)$ is also affected considerably by the stochasticity of the parameters.

6.9. **Covariation of yields.** A further example where empirical differences between CRC and non-CRC models become apparent is the covariation of yields. In the Hull-White Vasicek and CIR models without the CRC extension, the covariation matrix of yields with different times to maturity has rank 1. This is in stark contrast to the covariations observed in the market. For instance, the $33 \times 33$ covariation matrix of market yields with times to maturity ranging from 3 months to 30 years typically has rank between 5 and 9, as shown in Figure 6.16. The ranks produced by the CRC models (V4) and (CIR4) typically lie between 3 and 5. Thus, they are higher than those of the non-CRC models, but not as high as those of the market.

Numerically, the covariation matrix is calculated as in (4.11) and the ranks are defined as the number of singular values differing significantly from zero. A comparison with Figures 6.18 and 6.17 shows that higher ranks are also related to higher volatility of the parameters.
Figure 6.1. Historical zero-coupon yields estimated by the ECB for various times to maturity.

Figure 6.2. Zero-coupon yield curves estimated by the ECB for various observation dates.
Figure 6.3. Historical short rate approximated by 3 month yields.

Figure 6.4. Volatility parameter of the Vasicek model estimated by (6.1) using a time window of $M = 100$ yields with time to maturity $\tau_1$. 
Figure 6.5. Volatility parameter of the CIR model estimated by (6.2) using a time window of \( M = 100 \) yields with time to maturity \( \tau_1 \).

Figure 6.6. Speed of mean-reversion parameter of the Vasiček model estimated by (6.3) using a time window of \( M = 100 \) yields with times to maturity \( \tau_1 = 0.25 \) and various values of \( \tau_2 \).
Figure 6.7. Speed of mean-reversion parameter of the CIR model estimated by (6.4) using a time window of $M = 100$ yields with times to maturity $\tau_1 = 0.25$ and various values of $\tau_2$.

Figure 6.8. Calibrations of some homogeneous and Hull-White extended models as of September 2, 2013. Vasicek 1 and CIR 1 are homogeneous models calibrated to the yield curve dynamics using (6.1)-(6.4) with $\tau_1 = 0.25$ and $\tau_2 = 2$. Vasiček 2 and CIR 2 are homogeneous models calibrated to the prevailing yield curve by least squares. The Hull-White extended models match the initial yield curve exactly.
Figure 6.9. The Hull-White extensions $\theta(t)$ corresponding to the models of Figure 6.8. In the time-homogeneous models, $\theta(t)$ is a constant function of time to maturity $\tau$.

Figure 6.10. Calibrations of some homogeneous and Hull-White extended models as of 1 April 2014. Vasicek 1 and CIR 1 are homogeneous models calibrated to the yield curve dynamics using (6.1)-(6.4) with $\tau_1 = 0.25$ and $\tau_2 = 2$. Vasicek 2 and CIR 2 are homogeneous models calibrated to the prevailing yield curve by least squares. The Hull-White extended models match the initial yield curve exactly.
Figure 6.11. The Hull-White extensions \( \theta(t) \) corresponding to the models of Figure 6.10. Note that \( \theta(t) \) assumes negative values, which is typical in situations where the yield curve is inverted at the short end.

Figure 6.12. The historical values of \( \theta(t)(0) \) in the CIR model calculated using the estimates of \( \beta \) in Figure 6.7. Negative values occur frequently in 2009 and 2012 to 2014. They are problematic for reasons laid out in Section 6.4.
Figure 6.13. Absolute error (log-log plot) of the Monte Carlo estimate of the moment generating function $E[e^{\eta r(1)}]$ for model (V2). This is calculated as the absolute difference between the estimate and (6.5) for different values of $\delta$. We simulate $10^5$ paths for the estimation.

Figure 6.14. Absolute error (log-log plot) of the Monte Carlo estimate of the moment generating function $E[e^{\eta r(1)}]$ for model (V4) defined in Section 6.5. The true values are estimated by the intercept of the linear extrapolation of the Monte Carlo estimates. The errors are calculated as the absolute difference between the intercept and the estimates. $10^5$ paths were used in the simulation.
**Figure 6.15.** Absolute error (log-log plot) of the Monte Carlo estimate of the moment generating function $E[e^{\eta r(1)}]$ for model (CIR4). The true values are estimated by the intercept of the linear extrapolation of the Monte Carlo estimates. The errors are calculated as the absolute difference between the intercept and the estimates for different values of $\delta$. $10^5$ paths were used in the simulation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean (%)</th>
<th>Median (%)</th>
<th>Volatility (%)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Minimum (%)</th>
<th>Maximum (%)</th>
<th>1st Quartile (%)</th>
<th>3rd Quartile (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V1)</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>-0.02</td>
<td>3.01</td>
<td>-0.19</td>
<td>0.60</td>
<td>0.18</td>
<td>0.31</td>
</tr>
<tr>
<td>(V2)</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>-0.02</td>
<td>3.99</td>
<td>-0.19</td>
<td>0.60</td>
<td>0.18</td>
<td>0.31</td>
</tr>
<tr>
<td>(V3)</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>-0.02</td>
<td>3.01</td>
<td>-0.19</td>
<td>0.60</td>
<td>0.18</td>
<td>0.31</td>
</tr>
<tr>
<td>(V4)</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>-0.02</td>
<td>3.99</td>
<td>-0.19</td>
<td>0.60</td>
<td>0.18</td>
<td>0.31</td>
</tr>
<tr>
<td>(CIR1)</td>
<td>0.25</td>
<td>0.20</td>
<td>0.19</td>
<td>1.79</td>
<td>8.24</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>(CIR2)</td>
<td>0.25</td>
<td>0.12</td>
<td>0.33</td>
<td>2.75</td>
<td>14.47</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>(CIR3)</td>
<td>0.25</td>
<td>0.14</td>
<td>0.30</td>
<td>2.63</td>
<td>13.61</td>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>(CIR4)</td>
<td>0.24</td>
<td>0.19</td>
<td>0.20</td>
<td>1.90</td>
<td>9.06</td>
<td>0.00</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

**Table 6.1.** Statistics of the short rate $r(1)$ at time 1 in the models defined in Section 6.5 obtained by Monte-Carlo simulations. $10^5$ paths and a step size of $\delta = 0.02$ were used in the simulation.
Figure 6.16. Historical rank of the empirical covariation matrix (4.11) based on time windows of $M = 100$ market yields with 33 different times to maturity $\tau_i \in \{0.25, 0.5, 0.75, 1, 2, 3, \ldots, 30\}$. For comparison, the plot also features the average ranks obtained in simulations of the Hull-White extended affine models (V1) and (CIR1) as well as their CRC counterparts (V4) and (CIR4). These models were calibrated using time windows of $M = 100$ observations. The missing values in (CIR1) and (CIR4) are due to non-admissible negative levels of mean reversion at these dates, see Section 6.4 and Figure 6.12. The averages are taken over $10^3$ simulated paths. In the numerical computation of the rank, eigenvalues which are $10^{-6}$ times smaller than the largest eigenvalue are rounded down to zero.

Figure 6.17. Historical values of the parameter $\sigma_1$ in the models (V4) and (CIR4) defined in Section 6.5 estimated using time windows of $M = 100$ observations.
Figure 6.18. Historical values of the parameter $\sigma_2$ in the models (V4) and (CIR4) defined in Section 6.5 estimated using time windows of $M = 100$ observations.
Philipp Harms, David Stefanovits, Josef Teichmann, Mario V. Wüthrich.

Consistent re-calibration of the discrete-time multifactor Vasiček model.

Abstract

The discrete-time multifactor Vasiček model is a tractable Gaussian spot rate model. Typically, two- or three-factor versions allow to capture the dependence structure between yields with different times to maturity in an appropriate way. In practice, re-calibration of the model to the prevailing market conditions leads to model parameters which change over time. Therefore, the model parameters should be understood as being time-dependent, or even stochastic. Following the consistent re-calibration (CRC) approach of [43] we construct models as concatenations of yield curve increments of Hull-White extended multifactor Vasiček models with different parameters. The CRC approach provides attractive tractable models that preserve the no-arbitrage premise. As a numerical example we fit Swiss interest rates using CRC multifactor Vasiček models.

1 Introduction

The tractability of affine models, such as the Vasiček [79] and the Cox-Ingersoll-Ross [20] models, has made them appealing for term structure modelling. Affine term structure models are based on a (multidimensional) factor process which in turn describes the evolution of the spot rate and the bank account processes. No-arbitrage arguments then provide the corresponding zero-coupon bond prices, yield curves and forward rates. Prices in these models are calculated under an equivalent martingale measure for known static model parameters. However, model parameters typically vary over time as financial market conditions change. They may, for instance, be of regime switching nature and need to be permanently re-calibrated to the actual financial market conditions. In practice, this re-calibration is done on a regular basis (as new information becomes available). This implies that model parameters are not static and, henceforth, may also be understood as stochastic processes. The re-calibration should preserve the no-arbitrage condition, which provides side constraints in the re-calibration. The aim of this work is to...
discuss these side constraints with the help of the discrete-time multifactor Vasićek interest rate model, which is a tractable but also flexible model. We show that re-calibration under the side constraints naturally leads to Heath-Jarrow-Morton [44] models with stochastic parameters.

Organisation of the paper. In Section 2 we introduce Hull-White extended discrete-time multifactor Vasićek models, which are the building blocks for consistent re-calibration (CRC) in this work. We define CRC of the Hull-White extended multifactor Vasićek model in Section 3. Section 4 specifies the market price of risk assumptions used to model the factor process under the real world probability measure and the equivalent martingale measure, respectively. In Section 5 we deal with parameter estimation from market data. In Section 6 we fit the model to Swiss interest rate data, and in Section 7 we conclude. All proofs are presented in Appendix A and all figures are in Appendix B.

2 Discrete-time Vasićek model and Hull-White extension

Choose a fixed grid size $\Delta > 0$ and consider the discrete-time grid $\{0, \Delta, 2\Delta, 3\Delta, \ldots\} = N_0\Delta$. For example, a daily grid corresponds to $\Delta = 1/252$ if there are 252 business days per year.

Choose a (sufficiently rich) filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with discrete-time filtration $\mathbb{F} = (\mathcal{F}(t))_{t \in \mathbb{N}_0}$, where $t \in \mathbb{N}_0$ refers to time point $t\Delta$. Assume that $\mathbb{P}$ denotes an equivalent martingale measure for a (strictly positive) bank account numeraire $(B(t))_{t \in \mathbb{N}_0}$. $B(t)$ denotes the value at time $t\Delta$ of an investment of one unit of currency at time 0 into the bank account (i.e., the risk-free rollover relative to $\Delta$).

2.1 Discrete-time multifactor Vasićek model

We choose $n \in \mathbb{N}$ fixed and introduce the discrete-time $n$-factor Vasićek model.

Notation. Subscript indices refer to elements of vectors and matrices. Argument indices refer to time points. We denote the $n \times n$ identity matrix by $\mathbf{I} \in \mathbb{R}^{n \times n}$. We also introduce the vectors $\mathbf{1} = (1, \ldots, 1)\top \in \mathbb{R}^n$ and $\mathbf{e}_1 = (1, 0, \ldots, 0)\top \in \mathbb{R}^n$.

We consider the $n$-dimensional $\mathbb{F}$-adapted factor process

$$X = (X(t))_{t \in \mathbb{N}_0} = (X_1(t), \ldots, X_n(t))_{t \in \mathbb{N}_0},$$

which generates the spot rate and bank account processes as follows

$$r(t) = \mathbf{1}^\top X(t) \quad \text{and} \quad B(t) = \exp \left\{ \Delta \sum_{s=0}^{t-1} r(s) \right\}, \quad (2.1)$$

where $t \in \mathbb{N}_0$; empty sums are set equal to zero. The factor process $X$ is assumed to evolve under $\mathbb{P}$ according to

$$X(t) = b + \beta X(t-1) + \Sigma \mathbf{1} \varepsilon^*(t), \quad t > 0, \quad (2.2)$$
with initial factor $X(0) \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, $\beta \in \mathbb{R}^{n \times n}$, $\Sigma_1 \in \mathbb{R}^{n \times n}$ and $(\varepsilon^\ast(t))_{t \in \mathbb{N}} = (\varepsilon^\ast_1(t), \ldots, \varepsilon^\ast_n(t))^{\top}$ being $\mathbb{P}$-adapted. The following assumptions are in place throughout the paper.

**Assumptions 2.1** We assume that the spectrum of matrix $\beta$ is a subset of $(-1, 1)^n$ and that matrix $\Sigma_1$ is non-singular. Moreover, for each $t \in \mathbb{N}$, we assume that $\varepsilon^\ast(t)$ is independent of $\mathcal{F}(t - 1)$ under $\mathbb{P}^\ast$ with $\varepsilon^\ast(t) \sim N(0, \mathbb{I})$.

**Remark.** In Assumptions 2.1 the condition on matrix $\beta$ ensures that $1 - \beta$ is invertible and that the geometric series generated by $\beta$ converges. The condition on $\Sigma_1$ ensures that $\Sigma = \Sigma_1 \Sigma_1^\top$ is symmetric positive definite. Under Assumptions 2.1 equation (2.2) defines a stationary process, see [11, Section 11.3].

The model defined by equations (2.1)-(2.2) is called discrete-time multifactor Vasiček model. Under the above model assumptions we have for $m > t$

$$
X(m)|\mathcal{F}(t) \sim \mathcal{N} \left( (1 - \beta)^{-1} \left( 1 - \beta^{m-t} \right) b + \beta^{m-t} X(t), \sum_{s=0}^{m-t-1} \beta^s \Sigma (\beta^\top)^s \right).
$$

(2.3)

**Remark.** For $m > t$ the conditional distribution of $X(m)$, given $\mathcal{F}(t)$, depends only on the value $X(t)$ at time $t \Delta$ and on lag $m - t$. In other words, the factor process (2.2) is a time-homogeneous Markov process.

At time $t \Delta$ the price of the zero-coupon bond (ZCB) with maturity date $m \Delta > t \Delta$ with respect to filtration $\mathcal{F}$ and equivalent martingale measure $\mathbb{P}^\ast$ is given by

$$
P(t, m) = \mathbb{E}^\ast \left[ \frac{B(t)}{B(m)} \bigg| \mathcal{F}(t) \right] = \mathbb{E}^\ast \left[ \exp \left( -\Delta \sum_{s=t}^{m-1} 1^\top X(s) \right) \bigg| \mathcal{F}(t) \right].
$$

For the proof of the following result see Appendix A.

**Theorem 2.2** The ZCB prices in the discrete-time multifactor Vasiček model (2.1)-(2.2) with respect to filtration $\mathcal{F}$ and equivalent martingale measure $\mathbb{P}^\ast$ have an affine term structure

$$
P(t, m) = e^{A(t,m) - B(t,m)^\top X(t)}, \quad m > t,
$$

with $A(m - 1, m) = 0$, $B(m - 1, m) = 1 \Delta$ and for $m - 1 > t \geq 0$

$$
A(t, m) = A(t + 1, m) - B(t + 1, m)^\top b + \frac{1}{2} B(t + 1, m)^\top \Sigma B(t + 1, m),
$$

$$
B(t, m) = (1 - \beta^\top)^{-1} \left( 1 - (\beta^\top)^{m-t} \right) 1 \Delta.
$$

In the discrete-time multifactor Vasiček model (2.1)-(2.2) the term structure of interest rates (yield curve) takes the following form at time $t \Delta$ for maturity dates $m \Delta > t \Delta$

$$
Y(t, m) = \frac{1}{(m - t) \Delta} \log P(t, m) = - \frac{A(t, m)}{(m - t) \Delta} + \frac{B(t, m)^\top X(t)}{(m - t) \Delta},
$$

(2.4)

with spot rate at time $t \Delta$ given by $Y(t, t + 1) = 1^\top X(t) = r(t)$. 

99
2.2 Hull-White extended discrete-time multifactor Vasicek model

The possible shapes of the Vasicek yield curve (2.4) are restricted by the choice of the parameters \( b \in \mathbb{R}^n, \beta \in \mathbb{R}^{n \times n} \) and \( \Sigma \in \mathbb{R}^{n \times n} \). These parameters are not sufficiently flexible to exactly calibrate the model to an arbitrary observed initial yield curve. Therefore, we consider the Hull-White extended version (see [50]) of the discrete-time multifactor Vasicek model. We replace the factor process defined in (2.2) as follows. For fixed \( k \in \mathbb{N}_0 \) let \( X^{(k)} \) satisfy

\[
X^{(k)}(t) = b + \theta(t-k)e_1 + \beta X^{(k)}(t-1) + \Sigma^{1/2} \varepsilon^*(t), \quad t > k, \tag{2.5}
\]

with starting factor \( X^{(k)}(k) = \mathbb{E}_k, e_1 = (1,0,\ldots,0)^T \in \mathbb{R}^n \) and function \( \theta : \mathbb{N} \to \mathbb{R} \). Model assumption (2.5) corresponds to (2.2) where the first component of \( b \) is replaced by the time dependent coefficient \( (b_1+\theta(i))_{i \in \mathbb{N}} \) and all other terms ceteris paribus. Without loss of generality we choose the first component for this replacement. Note that parameter \( b_1 \) is redundant in this model specification, but for didactical reasons it is used below. The time dependent coefficient \( \theta \) is called Hull-White extension and it is used to calibrate the model to a given yield curve at a given time point \( k \Delta \). The upper index \( (k) \) denotes that time point and corresponds to the time shift we apply to the Hull-White extension \( \theta \) in model (2.5). The factor process \( X^{(k)} \) generates the spot rate process and the bank account process as in (2.1).

The model defined by assumptions (2.1, 2.5) is called Hull-White extended discrete-time multifactor Vasicek model. Under these model assumptions we have for \( m > t \geq k \)

\[
X^{(k)}(m)|\mathcal{F}(t) \sim \mathcal{N}
\left(
\sum_{s=0}^{m-t-1} \beta^s (b + \theta(m-s-k)e_1) + \beta^{m-t} X^{(k)}(t),
\sum_{s=0}^{m-t-1} \beta^s \Sigma (\beta^\top)^s
\right).
\]

Remark. For \( m > t \geq k \) the conditional distribution of \( X^{(k)}(m) \), given \( \mathcal{F}(t) \), depends only on the factor \( X^{(k)}(t) \) at time \( t \Delta \). In this case, factor process (2.5) is a time-inhomogeneous Markov process. Note that the upper index \( (k) \) in the notation is important since the conditional distribution depends explicitly on the lag \( m-k \).

**Theorem 2.3** The ZCB prices in the Hull-White extended discrete-time multifactor Vasicek model (2.1, 2.5) with respect to filtration \( \mathbb{F} \) and equivalent martingale measure \( \mathbb{P}^* \) have affine term structure

\[
P^{(k)}(t,m) = e^{A^{(k)}(t,m) - B(t,m)\top X^{(k)}(t)}, \quad m > t \geq k,
\]

with \( B(t,m) \) as in Theorem 2.2, \( A(m-1,m) = 0 \) and for \( m-1 > t \geq k \)

\[
A^{(k)}(t,m) = A^{(k)}(t+1,m) - B(t+1,m)\top (b + \theta(t+1-k)e_1)
+ \frac{1}{2} B(t+1,m)\top \Sigma B(t+1,m).
\]

In the Hull-White extended discrete-time multifactor Vasicek model (2.1, 2.5) the yield curve takes the following form at time \( t \Delta \) for maturity dates \( m \Delta > t \Delta \geq k \Delta \)

\[
Y^{(k)}(t,m) = -\frac{1}{(m-t)\Delta} \log P^{(k)}(t,m) = \frac{A^{(k)}(t,m)}{(m-t)\Delta} + \frac{B(t,m)\top X^{(k)}(t)}{(m-t)\Delta}, \tag{2.6}
\]
with spot rate at time $t\Delta$ given by $Y^{(k)}(t, t + 1) = 1^\top X^{(k)}(t)$.

Remark. Note that the coefficient $B(t, m)$ in Theorem 2.3 is not affected by the Hull-White extension $\theta$ and depends solely on $m - t$, whereas the coefficient $A^{(k)}(t, m)$ depends explicitly on Hull-White extension $\theta$.

### 2.3 Calibration of the Hull-White extended model

We consider the term structure model defined by the Hull-White extended factor process $X^{(k)}$ and calibrate the Hull-White extension $\theta \in \mathbb{R}^N$ to a given yield curve at time point $k\Delta$. We explicitly introduce the time index $k$ in model (2.5) because the CRC algorithm is a concatenation of multiple Hull-White extended models which are calibrated at different time points $k\Delta$, see Section 3 below.

Assume that there is a fixed final time to maturity date $M\Delta$ and that we observe at time $k\Delta$ the yield curve $\hat{y}(k) \in \mathbb{R}^M$ for maturity dates $(k + 1)\Delta, \ldots, (k + M)\Delta$. For these maturity dates the Hull-White extended discrete-time multifactor Vasiček yield curve at time $k\Delta$, given by Theorem 2.3, reads as

$$y^{(k)}(k) = \left( -\frac{1}{t\Delta} A^{(k)}(k, k + i) + \frac{1}{t\Delta} B(k, k + i)\right) X^{(k)}(k)\right)_{i=1,\ldots,M} \in \mathbb{R}^M. $$

For given starting factor $X^{(k)}(k) \in \mathbb{R}^n$, and parameters $b \in \mathbb{R}^n$, $\beta \in \mathbb{R}^{n \times n}$ and $\Sigma \in \mathbb{R}^{n \times n}$ our aim is to choose the Hull-White extension $\theta \in \mathbb{R}^N$ such that we get an exact fit at time $k\Delta$ to the yield curve $\hat{y}(k)$, that is,

$$y^{(k)}(k) = \hat{y}(k).$$

The following theorem provides an equivalent condition to (2.7) which allows to calculate the Hull-White extension $\theta \in \mathbb{R}^N$ explicitly.

**Theorem 2.4** Denote by $y^{(k)}(k)$ the yield curve at time $k\Delta$ obtained from the Hull-White extended discrete-time multifactor Vasiček model (2.1, 2.5) for given starting factor $X^{(k)}(k) = \in \mathbb{R}^n$, parameters $b \in \mathbb{R}^n$, $\beta \in \mathbb{R}^{n \times n}$ and $\Sigma \in \mathbb{R}^{n \times n}$, and Hull-White extension $\theta \in \mathbb{R}^N$. For given $y \in \mathbb{R}^M$ identity $y^{(k)}(k) = y$ holds if and only if the Hull-White extension $\theta$ fulfils

$$\theta = C(\beta)^{-1} z(\beta, \Sigma, x, y),$$

(2.8)

where $\theta = (\theta_i)_{i=1,\ldots,M-1} \in \mathbb{R}^{M-1}$, $C(\beta) = (C_{ij}(\beta))_{i,j=1,\ldots,M-1} \in \mathbb{R}^{(M-1) \times (M-1)}$ and $z(\beta, \Sigma, x, y) = (z_i(\beta, \Sigma, x, y))_{i=1,\ldots,M-1} \in \mathbb{R}^{M-1}$ are defined by

$$\theta_i = \theta(i),$$

$$C_{ij}(\beta) = B_1(k + j, k + i + 1) 1_{i,j},$$

$$z_i(\beta, \Sigma, x, y) = \sum_{i=1}^{k+1} \left( \frac{1}{2} B(s, k + i + 1)\Sigma B(s, k + i + 1) - B(s, k + i + 1)\right)$$

$$- 1^\top (1 - \beta^{i+1})(1 - \beta)^{-1} x\Delta + (i + 1)y_{i+1}(k)\Delta,$$

with $i, j = 1, \ldots, M - 1$ and $B(\cdot, \cdot) = (B_1(\cdot, \cdot), \ldots, B_n(\cdot, \cdot))^\top$ given by Theorem 2.2.
Theorem 2.4 shows that the Hull-White extension can be calculated by inverting the \((M - 1) \times (M - 1)\) lower triangular positive definite matrix \(C(\beta)\).

3 Consistent re-calibration

The crucial extension now is the following: we let parameters \(b\), \(\beta\) and \(\Sigma\) vary over time, and we re-calibrate the Hull-White extension in a consistent way at each time point, that is, according to the actual choice of the parameter values using Theorem 2.4. Below we show that this naturally leads to a Heath-Jarrow-Morton [44] (HJM) approach to term structure modelling.

3.1 Consistent re-calibration algorithm

Assume that \((b(k))_{k \in \mathbb{N}_0}, (\beta(k))_{k \in \mathbb{N}_0}\) and \((\Sigma(k))_{k \in \mathbb{N}_0}\) are \(\mathcal{F}\)-adapted parameter processes with \(\beta(k) \text{ and } \Sigma(k)\) satisfying Assumptions 2.1, \(\mathbb{P}^*-\text{a.s.}\), for all \(k \in \mathbb{N}_0\). Based on these parameter processes we define the \(n\)-dimensional \(\mathcal{F}\)-adapted CRC factor process \(X\) which evolves according to steps (i)-(iv) of the CRC algorithm described below. Thus, factor process \(X\) will define a spot rate model similar to (2.1).

The CRC algorithm works as follows.

(i) Initialisation \(k = 0\). Assume that the initial yield curve observation at time 0 is given by \(\hat{y}(0) \in \mathbb{R}^M\). Let \(\theta(0) \in \mathbb{R}^N\) be an \(\mathcal{F}(0)\)-measurable Hull-White extension such that condition (2.7) is satisfied at time 0 for initial factor \(X(0) \in \mathbb{R}^n\), and parameters \(b(0), \beta(0)\) and \(\Sigma(0)\). By Theorem 2.4 the values \(\theta(0) = (\theta(0)(i))_{i=1, \ldots, M-1} \in \mathbb{R}^{M-1}\) are given by

\[
\theta(0) = C(\beta(0))^{-1} z(b(0), \beta(0), \Sigma(0), X(0), \hat{y}(0)) .
\]

This provides Hull-White extended Vasiček yield curve \(y(0)\) identically equal to \(\hat{y}(0)\) for given initial factor \(X(0)\) and parameters \(b(0), \beta(0), \Sigma(0)\).

(ii) Increments of the factor process from \(k \rightarrow k + 1\). Assume factor \(X(k)\), parameters \(b(k), \beta(k)\) and \(\Sigma(k)\), and Hull-White extension \(\theta(k)\) are given. Define the Hull-White extended model \(X(k) = (X(k)(t))_{t \geq k}\) by

\[
X(k)(t) = b(k) + \theta(k)(t - k)e_1 + \beta(k)X(k)(t - 1) + \Sigma(k)e^*(t), \quad t > k,
\]

with starting value \(X(k)(k) = X(k), \mathcal{F}(k)\)-measurable parameters \(b(k), \beta(k)\) and \(\Sigma(k)\), and Hull-White extension \(\theta(k)\). We update the factor process \(X\) at time \((k + 1)\Delta\) according to the \(X(k)\)-dynamics, that is, we set

\[
X(k + 1) = X(k)(k + 1).
\]

This provides \(\mathcal{F}(k+1)\)-measurable yield curve at time \((k+1)\Delta\) for maturity dates \(m\Delta > (k+1)\Delta\)

\[
Y(k)(k+ 1, m) = -\frac{A(k)(k + 1, m)}{(m - (k + 1))\Delta} + \frac{B(k)(k + 1, m)^\top X(k + 1)}{(m - (k + 1))\Delta},
\]

102
with \( A(k)(m-1, m) = 0 \) and \( B(k)(m-1, m) = \Delta_1 \), and recursively for \( m-1 > t \geq k \)
\[
A(k)(t, m) = A(k)(t+1, m) - B(k)(t+1, m) \\
+ \frac{1}{2} B(k)(t+1, m)^\top \Sigma(k) B(k)(t+1, m).
\]
\[
B(k)(t, m) = \left(1 - \beta(k)^\top\right)^{-1} \left(1 - (\beta(k)^\top)^{m-t}\right) 1\Delta.
\]
This is exactly the no-arbitrage price under \( \mathbb{P}^* \) if the parameters \( b(k), \beta(k) \) and \( \Sigma(k) \) and the Hull-White extension \( \theta^{(k)} \) remain constant for all \( t > k \).

(iii) Parameter update and re-calibration at \( k + 1 \). Assume that at time \( (k + 1)\Delta \) the parameters \( (b(k), \beta(k), \Sigma(k)) \) are updated to \( (b(k+1), \beta(k+1), \Sigma(k+1)) \). We may think of this parameter update as a consequence of model selection after we observe a new yield curve at time \( (k + 1)\Delta \). This is discussed in more detail in Section 5 below. The no-arbitrage yield curve at time \( (k + 1)\Delta \) from the model with parameters \( (b(k), \beta(k), \Sigma(k)) \), and Hull-White extension \( \theta^{(k)} \) is given by
\[
y^{(k)}(k+1) = \left(Y^{(k)}(k+1, k+2), \ldots, Y^{(k)}(k+1, k+1+M)\right)^\top \in \mathbb{R}^M.
\]
The parameter update \( (b(k), \beta(k), \Sigma(k)) \rightarrow (b(k+1), \beta(k+1), \Sigma(k+1)) \) requires re-calibration of the Hull-White extension, otherwise arbitrage is introduced into the model. This re-calibration provides \( \mathcal{F}(k+1) \)-measurable Hull-White extension \( \theta^{(k+1)} \in \mathbb{R}^N \) at time \( (k + 1)\Delta \). The values \( \theta^{(k+1)} = (\theta^{(k+1)}(i))_{i=1, \ldots, M-1} \in \mathbb{R}^{M-1} \) are given by, see Theorem 2.4,
\[
\theta^{(k+1)} = C(\beta(k+1))^{-1} \mathcal{Z} \left(b(k+1), \beta(k+1), \Sigma(k+1), \mathcal{X}(k+1), y^{(k)}(k+1)\right),
\]
and the resulting yield curve \( y^{(k+1)}(k+1) \) under the updated parameters is identically equal to \( y^{(k)}(k+1) \). Note that this CRC makes the upper index \( (k) \) in the yield curve superfluous because the Hull-White extension is re-calibrated to the new parameters such that the resulting yield curve remains unchanged. Therefore, we write \( \mathcal{Y}(k, \cdot) \) in the sequel for the CRC yield curve with factor \( \mathcal{X}(k) \), parameters \( b(k), \beta(k), \Sigma(k) \), and Hull-White extension \( \theta^{(k)} \).

(iv) Iteration. Iterate items (ii)-(iii) for \( k \geq 0 \)
\[
\square
\]
Remark. For the implementation of the above algorithm we need to consider the following issue. Assume we start the algorithm at time 0 with initial yield curve \( \mathcal{Y}(0) \in \mathbb{R}^M \). At times \( k\Delta \), for \( k > 0 \), calibration of \( \theta^{(k)} \in \mathbb{R}^{M-1} \) requires yields with times to maturity beyond \( M\Delta \). Either yields for these times to maturity are observable and the length of \( \theta^{(k)} \) is reduced in every step of the CRC algorithm or an appropriate extrapolation method beyond the latest available maturity date is applied in every step.
3.2 Heath-Jarrow-Morton representation

We analyse the yield curve dynamics \((Y(k,·))_{k∈N_0}\) obtained by the CRC algorithm of Section 3.1. Due to re-calibration (3.2) the yield curve fulfils the following identity for \(m > k + 1\)

\[
Y(k + 1, m) = -\frac{A^{(k)}(k + 1, m)}{(m - (k + 1))\Delta} \mathbf{1} + \frac{B^{(k)}(k + 1, m)\mathbf{X}(k + 1)}{(m - (k + 1))\Delta},
\]

\[
= -\frac{A^{(k+1)}(k + 1, m)}{(m - (k + 1))\Delta} \mathbf{1} + \frac{B^{(k+1)}(k + 1, m)\mathbf{X}(k + 1)}{(m - (k + 1))\Delta},
\]

where the first line is based on the \(\mathcal{F}(k)\)-measurable parameters \((b(k), β(k), Σ(k))\) and Hull-White extension \(θ^{(k)}\), and the second line is based on the \(\mathcal{F}(k+1)\)-measurable parameters and Hull-White extension \((b(k + 1), β(k + 1), Σ(k + 1), θ^{(k+1)})\) after CRC step (iii). Note that in the re-calibration only \((b(k + 1), β(k + 1), Σ(k + 1))\) can be chosen exogenously and the Hull-White extension \(θ^{(k+1)}\) is used for consistency property (3.2). Our aim is to express \(Y(k + 1, m)\) as a function of \(\mathbf{X}(k)\) and \(Y(k, m)\). Using equations (3.1) and (3.3) we have for \(m > k + 1\)

\[
Y(k + 1, m) (m - (k + 1))\Delta = -\frac{A^{(k)}(k + 1, m)}{(m - (k + 1))\Delta} \mathbf{1} + \mathbf{X}(k)\mathbf{1}^\top + \frac{1}{2} \mathbf{B}(k + 1, m)^\top Σ(k) \mathbf{B}(k + 1, m)
\]

\[
+ \mathbf{B}(k + 1, m)^\top Σ(k) 1^\top ε^*(k + 1),
\]

This provides the following theorem, see Appendix A for the proof.

**Theorem 3.1** Under equivalent martingale measure \(P^∗\) the yield curve dynamics \((Y(k,·))_{k∈N_0}\) obtained by the CRC algorithm of Section 3.1 has the following HJM representation for \(m > k + 1\)

\[
Y(k + 1, m) (m - (k + 1))\Delta = Y(k, m) (m - k)\Delta - Y(k, k + 1)\Delta
\]

\[
+ \frac{1}{2} \mathbf{B}(k + 1, m)^\top Σ(k) \mathbf{B}(k + 1, m)
\]

\[
+ \mathbf{B}(k + 1, m)^\top Σ(k) 1^\top ε^*(k + 1),
\]

with \(\mathbf{B}(k + 1, m) = (1 - β^\top(k))^{-1} \mathbf{1} - (β(k)^\top)^{m-k-1}) \mathbf{1}\).

**Key observation.** Observe that in Theorem 3.1 a remarkable simplification happens. Simulating CRC algorithm (3.1)-(3.2) to future time points \(kΔ > 0\) does not require the calculation of the Hull-White extensions \((θ^{(k)})_{k∈N_0}\) according to (3.2), but the knowledge of the parameter process \((b(k), β(k), Σ(k))_{k∈N_0}\) is sufficient. The Hull-White extensions are fully encoded in the yield curve process \((Y(k,·))_{k∈N_0}\), and we can avoid inversion of (potentially) high dimensional matrices \(Σ(β(k))_{k∈N_0}\).

**Further remarks.**

- CRC of the multifactor Vasiček spot rate model can be defined directly in the HJM framework assuming a stochastic dynamics for the parameters. However, solely from the HJM representation, one cannot see that the yield curve dynamics is obtained in our case by combining well understood Hull-White extended multifactor Vasiček spot rate models using the CRC algorithm of Section 3, that is, the Hull-White extended multifactor Vasiček model gives an explicit functional form to the HJM representation.
The CRC algorithm of Section 3 does not rely directly on \((ε^*(t))_{t∈\mathbb{N}}\) having independent and Gaussian components. The CRC algorithm is feasible as long as explicit formulas for ZCB prices in the Hull-White extended model are available. Therefore, one may replace the Gaussian innovations by other distributional assumptions such as normal variance mixtures. This replacement is possible provided that conditional exponential moments can be calculated under the new innovation assumption. Under non-Gaussian innovations, it will no longer be the case that the HJM representation does not depend on the Hull-White extension \(θ(k) ∈ \mathbb{R}^N\).

Interpretation of the parameter processes will be given in Section 5, below.

4 Real world dynamics and market-price of risk

All previous derivations were done under an equivalent martingale measure \(P^*\) for the bank account numeraire. In order to statistically estimate parameters from market data we need to specify a Girsanov transformation to the real world measure, which is denoted by \(P\). We present a specific change of measure which provides tractable spot rate dynamics under \(P\). Assume that \((λ(k))_{k∈\mathbb{N}_0}\) and \((Λ(k))_{k∈\mathbb{N}_0}\) are \(\mathbb{R}^n\)- and \(\mathbb{R}^n×n\)-valued \(\mathbb{F}\)-adapted processes, respectively. Let \((X(k))_{k∈\mathbb{N}_0}\) be the factor process obtained by the CRC algorithm of Section 3. Then the \(n\)-dimensional \(\mathbb{F}\)-adapted process \((λ(k) + Λ(k)X(k))_{k∈\mathbb{N}_0}\) describes the market-price of risk dynamics. We define the following \(P^*\)-density process \((ξ(k))_{k∈\mathbb{N}_0}\)

\[
ξ(k) = \exp\left\{-\frac{1}{2} \sum_{s=0}^{k-1} \|λ(s) + Λ(s)X(s)\|^2 + \sum_{s=0}^{k-1} (λ(s) + Λ(s)X(s))^\top ε^*(s+1)\right\}, \quad k ∈ \mathbb{N}_0.
\]

The real world probability measure \(P\) is then defined by the Radon-Nikodym derivative

\[
\frac{dP}{dP^*}\bigg|_{\mathcal{F}(k)} = ξ(k), \quad k ∈ \mathbb{N}_0.
\]  

(4.1)

An immediate consequence is that for \(k ∈ \mathbb{N}_0\)

\[
ε(k+1) = λ(k) + Λ(k)X(k) + ε^*(k+1),
\]

has a standard Gaussian distribution under \(P\), conditionally on \(\mathcal{F}(k)\). This implies that under the real world measure \(P\) the factor process \((X(k))_{k∈\mathbb{N}_0}\) is described by

\[
X(k+1) = a(k) + α(k)X(k) + Σ(k)^{1/2}ε(k+1),
\]

(4.2)

where we define

\[
a(k) = b(k) + θ(k)(1)e_1 - Σ(k)^{1/2}λ(k) \quad \text{and} \quad α(k) = β(k) - Σ(k)^{1/2}Λ(k).
\]

(4.3)

As in Assumptions 2.1 we require \(Λ(k)\) to be such that the spectrum of \(α(k)\) is a subset of \((-1, 1)^n\). Formula (4.2) describes the dynamics of the factor process \((X(k))_{k∈\mathbb{N}_0}\) obtained by the CRC algorithm of Section 3.1 under real world measure \(P\). The following corollary describes the yield curve dynamics obtained by the CRC algorithm under \(P^*\), in analogy to Theorem 3.1.
Corollary 4.1 Under real world measure $\mathbb{P}$ satisfying (4.1) the yield curve dynamics $(Y(k, \cdot))_{k \in \mathbb{N}_0}$ obtained by the CRC algorithm of Section 3.1 has the following HJM representation for $m > k + 1$

$$Y(k + 1, m) (m - (k + 1)) \Delta = Y(k, m)(m - k)\Delta - Y(k, k + 1)\Delta$$

$$+ \frac{1}{2} B^{(k)}(k + 1, m)^\top \Sigma(k) B^{(k)}(k + 1, m)$$

$$- B^{(k)}(k + 1, m)^\top \Sigma(k) \Lambda(k)$$

$$- B^{(k)}(k + 1, m)^\top \Sigma(k) \Lambda(k) \mathbf{X}(k)$$

$$+ B^{(k)}(k + 1, m)^\top \Sigma(k) \varepsilon(k + 1),$$

with $B^{(k)}(k + 1, m) = (1 - \beta(k))^{-1} \left(1 - (\beta(k)^\top)^{m-k-1}\right) \mathbf{1} \Delta$.

Compared to Theorem 3.1 there are additional drift terms $- B^{(k)}(k + 1, m)^\top \Sigma(k) \frac{1}{2} \Lambda(k)$ and $- B^{(k)}(k + 1, m)^\top \Sigma(k) \frac{1}{2} \Lambda(k) \mathbf{X}(k)$ which are characterized by the market-price of risk parameters $\Lambda(k) \in \mathbb{R}^n$ and $\Lambda(k) \in \mathbb{R}^{n \times n}$.

5 Choice of parameter process

The yield curve dynamics obtained by the CRC algorithm of Section 3.1 require exogenous specification of the parameter process of the multifactor Vasicek model (2.1)-(2.2) and the market price of risk process, i.e., we need to model the process

$$(b(t), \beta(t), \Sigma(t), \Lambda(t))_{t \in \mathbb{N}_0}.$$

By equation (3.1) the one-step ahead development of the CRC factor process $\mathbf{X}$ under $\mathbb{P}$ reads as

$$\mathbf{X}(t + 1) = b(t) + \theta^{(t)}(1) \varepsilon_t - \Sigma(t) \frac{1}{2} \Lambda(t) + \left(\beta(t) - \Sigma(t) \frac{1}{2} \Lambda(t)\right) \mathbf{X}(t) + \Sigma(t) \frac{1}{2} \varepsilon(t + 1),$$

with $\mathcal{F}(t)$-measurable parameters $b(t)$, $\beta(t)$ and $\Sigma(t)$, and Hull-White extension $\theta^{(t)}$. Thus, on the one hand, the factor process $(\mathbf{X}(t))_{t \in \mathbb{N}_0}$ evolves according to (5.2), and on the other hand parameters $(b(t), \beta(t), \Sigma(t), \Lambda(t))_{t \in \mathbb{N}_0}$ evolve according to the financial market conditions. Note that the process $(\theta^{(t)})_{t \in \mathbb{N}_0}$ of Hull-White extensions is fully determined through CRC by (3.2). In order to distinguish the evolutions of $(\mathbf{X}(t))_{t \in \mathbb{N}_0}$ and $(b(t), \beta(t), \Sigma(t), \Lambda(t))_{t \in \mathbb{N}_0}$, respectively, we assume that process (5.1) changes at a slower pace than the factor process and, therefore, parameters can be assumed to be constant over a short time window. This assumption motivates the following approach to specifying a model for process (5.1). For each time point $t \Delta$ we fit multifactor Vasicek model (2.1)-(2.2) with fixed parameters $(b, \beta, \Sigma, \Lambda)$ on observations from a time window $\{t - K + 1, \ldots, t\}$ of length $K$. For estimation we assume that we have yield curve observations $(\tilde{Y}(k))_{k = t - K + 1, \ldots, t} = (\tilde{Y}_1(k), \ldots, \tilde{Y}_M(k)))_{k = t - K + 1, \ldots, t}$ for times to maturity $\tau_1 \Delta < \ldots < \tau_M \Delta$. Since yield curves are not necessarily observed on a regular time to maturity grid, we introduce the indices $\tau_1, \ldots, \tau_M \in \mathbb{N}$ to refer to the available times to maturity. Varying the time of estimation $t \Delta$ we obtain time series for the parameters from historical data. Finally,
we fit a stochastic model to these time series. In the following we discuss the interpretation of the parameters and present two different estimation procedures. The two procedures are combined to obtain a full specification of the model parameters.

5.1 Interpretation of parameters

**Level and speed of mean reversion.** By equation (2.3) we have under $\mathbb{P}^*$ for $m > t$

$$E^*[X(m)|\mathcal{F}(t)] = (1 - \beta)^{-1} (1 - \beta^{m-t}) b + \beta^{m-t} X(t),$$

$$E^*[r(m)|\mathcal{F}(t)] = 1^\top (1 - \beta)^{-1} (1 - \beta^{m-t}) b + 1^\top \beta^{m-t} X(t).$$

Thus, $\beta$ determines the speed at which the factor process $(X(t))_{t \in \mathbb{N}_0}$ and the spot rate process $(r(t))_{t \in \mathbb{N}_0}$ return to their long term means

$$\lim_{m \to \infty} E^*[X(m)|\mathcal{F}(t)] = (1 - \beta)^{-1} b \quad \text{and} \quad \lim_{m \to \infty} E^*[r(m)|\mathcal{F}(t)] = 1^\top (1 - \beta)^{-1} b.$$ A sensible choice of $(\beta(t))_{t \in \mathbb{N}_0}$ adapts the speed of mean reversion to the prevailing financial market conditions at each time point $t \Delta$.

**Instantaneous variance.** By equation (2.3) we have under $\mathbb{P}^*$ for $t > 0$

$$\text{Cov}^*[X(t)|\mathcal{F}(t-1)] = \Sigma, \quad \text{and} \quad \text{Var}^*[r(t)|\mathcal{F}(t-1)] = 1^\top \Sigma 1.$$ Thus, matrix $\Sigma$ plays the role of the instantaneous covariance matrix of $X$, and it describes the instantaneous spot rate volatility.

5.2 State space modelling approach

On each time window, we want to use yield curve observations to estimate the parameters of time-homogeneous Vasicek model (2.1)-(2.2). In general, this model is not able to reproduce the yield curve observations exactly. One reason is that the data might be given in the form of parametrised yield curves, and the parametrisation might not be compatible with the Vasicek model. For example, this is the case for the widely used Svensson family [76]. Another reason might be that yield curve observations do not exactly represent risk-free zero-coupon bonds. The discrepancy can be accounted for by adding a noise term to the Vasicek yield curves. This defines a state space model with the factor process as hidden state variable. In this state space model, the parameters of the factor dynamics can be estimated using Kalman filter techniques in conjunction with maximum likelihood estimation, see e.g. [82, Section 3.6.3].

**Transition system.** The evolution of the unobservable process $X$ under $\mathbb{P}$ is assumed to be given on time window $\{t - K + 1, \ldots, t\}$ by

$$X(k) = a + \alpha X(k-1) + \Sigma \xi(k), \quad k \in \{t - K + 1, \ldots, t\},$$

with initial factor $X(t - K) \in \mathbb{R}^n$, and parameters $a = b - \Sigma \lambda$ and $\alpha = \beta - \Sigma \Lambda$. The initial factor $X(t - K)$ is updated according to the output of the Kalman filter for the previous time window $\{t - K, \ldots, t - 1\}$. The initial factor is set to zero for the first time window available.
Remark. Parameters \((b, \beta, \Sigma, \lambda, \Lambda)\) are assumed to be constant over the time window \(\{t - K + 1, \ldots, t\}\). Thus, we drop the index \(k\) compared to equations (4.2)-(4.3). For estimation we assume that the factor process evolves according to the time-homogeneous multifactor Vasiček model (2.1)-(2.2) in that time window. The Hull-White extension is calibrated to the yield curve at time \(t \Delta\) given the estimated parameter values of the time-homogeneous model.

Measurement system. We assume that the observations in the state space model are given by

\[
\hat{Y}(k) = d + DX(k) + S\frac{1}{2}\eta(k), \quad k \in \{t - K, \ldots, t\},
\]

(5.3)

where

\[
\hat{Y}(k) = \left(\hat{Y}(k, k + \tau_1), \ldots, \hat{Y}(k, k + \tau_M)\right)\in \mathbb{R}^M,
\]

\[
d = \left(-(\tau_1\Delta)^{-1}A(k, k + \tau_1), \ldots, -(\tau_M\Delta)^{-1}A(k, k + \tau_M)\right)\in \mathbb{R}^M,
\]

\[
D_{ij} = (\tau_i\Delta)^{-1}B_j(k, k + \tau_i), \quad 1 \leq i \leq M, \quad 1 \leq j \leq n,
\]

with \(A(\cdot, \cdot)\) and \(B(\cdot, \cdot) = (B_1(\cdot, \cdot), \ldots, B_n(\cdot, \cdot))\) given by Theorem 2.2, and \(M\)-dimensional \(\mathcal{F}(k)\)-measurable noise term \(S\frac{1}{2}\eta(k)\) for non-singular \(S\frac{1}{2} \in \mathbb{R}^{M \times M}\). We assume that \(\eta(k)\) is independent of \(\mathcal{F}(k - 1)\) and \(\varepsilon(k)\) under \(\mathbb{P}\), and that \(\eta(k) \sim \mathcal{N}(0, 1)\). The error term \(S\frac{1}{2}\eta\) describes the discrepancy between the yield curve observations and the model. For \(S = 0\) we would obtain a yield curve in (5.3) which corresponds exactly to the multifactor Vasiček one.

Given parameter and market price of risk values \((b, \beta, \Sigma, \lambda, \Lambda)\) we estimate the factor using the following iterative procedure. For \(k \in \{t - K, \ldots, t\}\) and fixed \(t\) we consider the \(\sigma\)-field \(\mathcal{F}(k) = \sigma\left(\hat{Y}(s) \mid t - K \leq s \leq k\right)\subset \mathcal{F}(k)\) and describe the estimation procedure in this state space model.

Anchoring. Fix initial factor \(X(t - K) = x(t - K|t - K - 1)\) and initialise

\[
x(t - K + 1|t - K) = \mathbb{E}\left[X(t - K + 1)\big|\mathcal{F}(t - K)\right] = a + a\mathfrak{x}(t - K|t - K - 1),
\]

\[
\Sigma(t - K + 1|t - K) = \text{Cov} \left[X(t - K + 1)\big|\mathcal{F}(t - K)\right] = \Sigma.
\]

Forecasting the measurement system. At time \(k \in \{t - K + 1, \ldots, t\}\) we have

\[
y(k|k - 1) = \mathbb{E}\left[\hat{Y}(k)\big|\mathcal{F}(k - 1)\right] = d + DX(k|k - 1),
\]

\[
F(k) = \text{Cov}\left[\hat{Y}(k)\big|\mathcal{F}(k - 1)\right] = D\Sigma(k|k - 1)D^\top + S,
\]

\[
\zeta(k) = \tilde{y}(k) - y(k|k - 1).
\]

Bayesian inference in the transition system. The prediction error \(\zeta(k)\) is used to update the unobservable factors.

\[
x(k|k) = \mathbb{E}\left[X(k)\big|\mathcal{F}(k)\right] = x(k|k - 1) + K(k)\zeta(k),
\]

\[
\Sigma(k|k) = \text{Cov}\left[X(k)\big|\mathcal{F}(k)\right] = (I - K(k)D)\Sigma(k|k - 1),
\]

108
where $K(k)$ denotes the Kalman gain matrix given by

$$K(k) = \text{Cov} \left( X(k) \big| \hat{F}Y(k-1) \right) D^\top \text{Cov} \left( \hat{Y}(k) \big| \hat{F}Y(k-1) \right)^{-1} = \Sigma(k|k-1)D^\top F(k)^{-1}.$$ 

**Forecasting the transition system.** For the unobservable factor process we have the following forecast

$$x(k+1|k) = \mathbb{E} \left[ X(k+1) \big| \hat{F}Y(k) \right] = a + \alpha x(k|k),$$

$$\Sigma(k+1|k) = \text{Cov} \left( X(k+1) \big| \hat{F}Y(k) \right) = \alpha \Sigma(k|k)\alpha^\top + \Sigma.$$ 

**Likelihood function.** The Kalman filter procedure above allows one to infer factors $X$ given parameter and market price of risk values. Of course, in this section, we are interested in estimating these values in the first place. For this purpose the procedure above can be used in conjunction with maximum likelihood estimation. For the underlying parameters $\Theta = (b, \beta, \Sigma, a, \alpha)$ we have the following likelihood function given the observations $(\hat{y}(k))_{k=1}^{N}$:

$$L_t(\Theta) = \prod_{k=t-K+1}^{t} \frac{\exp \left( -\frac{1}{2} \zeta(k)\Sigma(k)^{-1}\zeta(k)^\top \right)}{(2\pi)^{\frac{n}{2}} \det F(k)^{\frac{n}{2}}}.$$ (5.4)

The maximum likelihood estimator (MLE) $\hat{\Theta}^{\text{MLE}} = (\hat{b}^{\text{MLE}}, \hat{\beta}^{\text{MLE}}, \hat{\Sigma}^{\text{MLE}}, \hat{a}^{\text{MLE}}, \hat{\alpha}^{\text{MLE}})$ is found by maximizing the likelihood function $L_t(\Theta)$ over $\Theta$, given the data. As in the EM (expectation maximization) algorithm, maximization of the likelihood function is alternated with Kalman filtering until convergence of the estimated parameters $\hat{\Theta}^{\text{MLE}}$ is achieved.

**5.3 Estimation motivated by continuous time modelling**

**Rescaling the time grid.** Assume factor process $(X(t))_{t\in\mathbb{N}_0}$ is given under $\mathbb{P}$ by $X(0) \in \mathbb{R}^n$ and for $t > 0$

$$X(t) = a + \alpha X(t-1) + \Sigma\varepsilon(t),$$

where $a = b - \Sigma\lambda$ and $\alpha = \beta - \Sigma\Lambda$. Furthermore, assume that $\alpha$ is a diagonalisable matrix with $\alpha = TDT^{-1}$ for $T \in \mathbb{R}^{n \times n}$ and diagonal matrix $D \in (-1,1)^{n \times n}$. Then the transformed process $Z = (T^{-1}X(t))_{t\in\mathbb{N}_0}$ evolves according to

$$Z(t) = c + DZ(t-1) + \Psi\varepsilon(t), \quad t > 0,$$

where $c = T^{-1}a$ and $\Psi = T^{-1}\Sigma(T^{-1})^\top$. For $d \in \mathbb{N}_0$ the $d$-step ahead conditional distribution of $Z$ under $\mathbb{P}$ is given by

$$Z(t + d) \big| F(t) \sim \mathcal{N} \left( \mu + \gamma Z(t), \Gamma \right), \quad t \geq 0,$$

where $\mu = (1 - D)^{-1}(1 - D^d)c$, $\gamma = D^d$ and $\Gamma = \sum_{s=0}^{d-1} D^s \Psi D^s$. Suppose we have estimated $\mu \in \mathbb{R}^n$, the diagonal matrix $\gamma \in (-1,1)^n$ and $\Gamma \in \mathbb{R}^{n \times n}$ on the time grid with size $d\Delta$, for instance, using MLE estimation as explained in Section 5.2. We are interested in recovering the parameters $c, D$ and $\Psi$ of the dynamics on the refined time grid with size $\Delta$ from $\mu$, $\gamma$ and $\Gamma$. 

109
The diagonal matrix \( D \) and vector \( c \) are reconstructed from the diagonal matrix \( \gamma \) as follows:

\[
D = \gamma^\frac{1}{2} = I + \frac{1}{d} \log(\gamma) + o \left( \frac{1}{d} \right), \quad \text{as } d \to \infty,
\]

\[
c = (1 - \gamma)^{-\frac{1}{2}} \mu = \frac{1}{d} (1 - \gamma)^{-1} \log(\gamma^{-1}) \mu + o \left( \frac{1}{d} \right), \quad \text{as } d \to \infty,
\]

where logarithmic and power functions applied to diagonal matrices are defined on their diagonal elements. Note that for \( i, j = 1, \ldots, n \) we have

\[
\Gamma_{ij} = \sum_{s=0}^{d-1} \gamma_{ii}^s \Psi_{ij} \gamma_{jj}^s = \Psi_{ij} \sum_{s=0}^{d-1} \left( \frac{1}{\gamma_{ii}^s} \frac{1}{\gamma_{jj}^s} \right)^s = \Psi_{ij} \frac{1 - \gamma_{ij}^s}{1 - (\gamma_{ij})^s}.
\]

Therefore, we recover \( \Psi \) from \( \gamma \) and \( \Gamma \) as follows.

\[
\Psi = \frac{1}{d} \nu + o \left( \frac{1}{d} \right), \quad \text{as } d \to \infty,
\]

where \( \nu = (-\Gamma_{ij} \log(\gamma_{ij})(1 - \gamma_{ij})^{-1})_{i,j=1,\ldots,n} \in \mathbb{R}^{n \times n} \). Consider for \( t > 0 \) the increments \( \mathcal{D}_t Z = Z(t) - Z(t-1) \). From the formulas for \( c, D \) and \( \Psi \) we observe that the \( \mathcal{F}_{t-1} \)-conditional mean of \( \mathcal{D}_t Z \)

\[
c + (D - 1) Z(t-1) = -\frac{1}{d} (1 - \gamma)^{-1} \log(\gamma) \mu + \frac{1}{d} \log(\gamma) Z(t-1) + o \left( \frac{1}{d} \right),
\]

and the \( \mathcal{F}_{t-1} \)-conditional volatility of \( \mathcal{D}_t Z \)

\[
\Psi^T = \sqrt{\frac{T}{d}} \nu + o \left( \sqrt{\frac{T}{d}} \right),
\]

live on different scales as \( d \to \infty \); in fact, volatility dominates for large \( d \). Under \( \mathbb{P} \) for \( t > 0 \) we have

\[
\mathbb{E} \left[ \mathcal{D}_t Z \mathcal{D}_t Z^\top | \mathcal{F}_{t-1} \right] = \text{Cov} [\mathcal{D}_t Z, \mathcal{D}_t Z] | \mathcal{F}_{t-1} + \mathbb{E} [\mathcal{D}_t Z | \mathcal{F}_{t-1}] \mathbb{E} [\mathcal{D}_t Z | \mathcal{F}_{t-1}]^\top
\]

\[
= \text{Cov} [Z(t), Z(t) | \mathcal{F}_{t-1}] + \left( \mathbb{E} [Z(t) | \mathcal{F}_{t-1}] - Z(t-1) \right) \left( \mathbb{E} [Z(t) | \mathcal{F}_{t-1}] - Z(t-1) \right)^\top
\]

\[
= \Psi + (c + (D - 1) Z(t-1)) (c + (D - 1) Z(t-1))^\top.
\]

Therefore, as \( d \to \infty \), we obtain

\[
\mathbb{E} \left[ \mathcal{D}_t X \mathcal{D}_t X^\top | \mathcal{F}_{t-1} \right] = T \mathbb{E} \left[ \mathcal{D}_t Z \mathcal{D}_t Z^\top | \mathcal{F}_{t-1} \right] T^\top
\]

\[
= T \Psi T^\top + T \left( c + (D - 1) Z(t-1) \right) \left( c + (D - 1) Z(t-1) \right)^\top T^\top
\]

\[
= \frac{1}{d} T \nu T^\top + o \left( \frac{1}{d} \right) = T \Psi T^\top + o \left( \frac{1}{d} \right) = \Sigma + o \left( \frac{1}{d} \right),
\]

where \( \mathcal{D}_t X = X(t) - X(t-1) \). This will be used in the next subsection.
Longitudinal realised covariations of yields. We consider the yield curve increments within the discrete-time multifactor Vasiček model (2.1)-(2.2). The increments of the yield process \((Y(t,t+\tau))_{t\in\mathbb{N}_0}\) for fixed time to maturity \(\tau \Delta > 0\) are given by

\[
D_{t,\tau}Y = Y(t,t+\tau) - Y(t-1,t-1+\tau) = \frac{1}{\tau \Delta} B(t,t+\tau)\top (X(t) - X(t-1)) = \frac{1}{\tau \Delta} B(t,t+\tau)\top D_tX,
\]

where \(D_tX|F_{t-1}\overset{P}{\sim} \mathcal{N}(a + (\alpha - 1)X(t-1), \Sigma)\). For times to maturity \(\tau_1 \Delta, \tau_2 \Delta > 0\) we get under \(P\)

\[
\mathbb{E}[D_{t,\tau_1}Y D_{t,\tau_2}Y|F_{t-1}] = \frac{1}{\tau_1 \tau_2 \Delta^2} B(t,t+\tau_1)\top \mathbb{E}\left[D_tX(D_tX)\top|F_{t-1}\right] B(t,t+\tau_2).
\]

By equation (5.5) for small grid size \(\Delta\) we estimate the last expression by

\[
\mathbb{E}[D_{t,\tau_1}Y D_{t,\tau_2}Y|F_{t-1}] \approx \frac{1}{\tau_1 \tau_2} \mathbb{I} (1 - \beta_{\tau_1}) (1 - \beta)^{-1} \Sigma \left(1 - \beta\right)^{-1} \left(1 - \left(\beta\right)^\top\right)^{\tau_2} 1. \tag{5.6}
\]

The latter is interesting for the following reasons.

- Formula (5.6) does not depend on the unobservable factors \(X\).
- Formula (5.6) allows for direct cross-sectional estimation of \(\beta\) and \(\Sigma\). That is, \(\beta\) and \(\Sigma\) can directly be estimated from market observations (without knowing the market-price of risk).
- Formula (5.6) is helpful to determine the number of factors needed to fit the model to market yield curve increments. This can be analysed by principal component analysis.
- Formula (5.6) can also be interpreted as a small-noise approximation for noisy measurement systems of the form (5.3).

Let \(\tilde{y}_1(k)\) and \(\tilde{y}_2(k)\) be market observations for times to maturity \(\tau_1 \Delta\) and \(\tau_2 \Delta\), and at times \(k \in \{t - K + 1, \ldots, t\}\), also specified in Section 5.2. Then the expectation on the left hand side of (5.6) can be estimated by the realised covariance

\[
\widehat{\text{RCov}}(t,\tau_1,\tau_2) = \frac{1}{K} \sum_{k=t-K+1}^{t} (\tilde{y}_1(k) - \tilde{y}_1(k-1))(\tilde{y}_2(k) - \tilde{y}_2(k-1)). \tag{5.7}
\]

The quality of this estimator hinges on two crucial assumptions. First, higher order terms in (5.5) are negligible in comparison to \(\Sigma\). Second, the noise term \(S \hat{\eta}\) in (5.3) leads to a negligible distortion in the sense that observations \(\tilde{Y}\) are reliable indicators for the underlying Vasiček yield curves.

Cross-sectional estimation of \(\beta\) and \(\Sigma\). Realised covariance estimator (5.7) can be used in conjunction with asymptotic relation (5.6) to estimate parameters \(\beta\) and \(\Sigma\) at time \(t \Delta\) in the
following way. For given symmetric weights $w_{ij} = w_{ji} \geq 0$ we solve the least squares problem

\[
\left( \hat{\beta}_{\text{RCov}}, \hat{\Sigma}_{\text{RCov}} \right) = \arg \min_{\beta, \Sigma} \left\{ \sum_{i,j=1}^{M} w_{ij} \left[ \text{Cov}(t, \tau^i, \tau^j) - \frac{1}{\tau^i \tau^j} \right]^{\top} \left( I - \left( \beta \right)^{\top} \right)^{-1} \left( I - \left( \beta \right)^{\top} \right) \right\},
\]

(5.8)

where we optimise over $\beta$ and $\Sigma$ satisfying Assumption 2.1.

5.4 Inference on market-price of risk

Finally we aim at determining parameters $\lambda$ and $\Lambda$ of the change of measure specified in Section 4. For this purpose we combine MLE estimation (Section 5.2) with estimation from realised covariations of yields (Section 5.3). First, we estimate $\beta$ and $\Sigma$ by $\hat{\beta}_{\text{RCov}}$ and $\hat{\Sigma}_{\text{RCov}}$ as in Section 5.3. Second, we estimate $a$, $b$, and $\alpha$ by maximising the log-likelihood

\[
\log L_t(b, \beta, \Sigma, a, \alpha) = \sum_{k=t-K+1}^{t} \log (\det F(k)) - \sum_{k=t-K+1}^{t} \zeta(k) \right)^{\top} F(k)^{-1} \zeta(k) + \text{const.}
\]

for fixed $\beta$ and $\Sigma$ over $b \in \mathbb{R}^n$, $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^{n \times n}$ with spectrum in $(-1,1)^n$, i.e.,

\[
\left( \hat{b}_{\text{MLE}}, \hat{a}_{\text{MLE}}, \hat{\alpha}_{\text{MLE}} \right) = \arg \max_{b,a,\alpha} \log L_t\left( b, \beta_{\text{RCov}}, \Sigma_{\text{RCov}}, a, \alpha \right).
\]

(5.9)

The constraint on the matrix $\alpha$ ensures that the factor process is stationary under the real world measure $P$. From equation (4.3) we have $\lambda = \Sigma^{-\frac{1}{2}} (b - a)$ and $\Lambda = \Sigma^{-\frac{1}{2}} (\beta - \alpha)$. This motivates inference of $\lambda$ by

\[
\hat{\lambda} = \left( \Sigma_{\text{RCov}} \right)^{-\frac{1}{2}} \left( \hat{b}_{\text{MLE}} - \hat{a}_{\text{MLE}} \right),
\]

(5.10)

and inference of $\Lambda$ by

\[
\hat{\Lambda}(k) = \left( \Sigma_{\text{RCov}} \right)^{-\frac{1}{2}} \left( \hat{\beta}_{\text{RCov}} - \hat{\alpha}_{\text{MLE}} \right).
\]

(5.11)

We stress the importance of estimating as many parameters as possible from the realised covariations of yields prior to using maximum likelihood estimation. The MLE procedure of Section 5.2 is computationally intensive and generally does not work well to estimate volatility parameters.

6 Numerical example for Swiss interest rates

6.1 Description and selection of data

We choose $\Delta = 1/252$, which corresponds to a daily time grid (assuming that a financial year has 252 business days). For the Swiss currency (CHF) we consider as yield observations the Swiss Average Rate (SAR), the London InterBank Offered Rate (LIBOR) and the Swiss Confederation Bond (SWCNB). See Figures B.1-B.2.
• **Short times to maturity.** The SAR is an ongoing volume-weighted average rate calculated by the Swiss National Bank (SNB) based on repo transactions between financial institutions. It is used for short times to maturity of at most 3 months. For SAR we have the Over-Night SARON that corresponds to a time to maturity of $\Delta$ (one business day) and the SAR Tomorrow-Next (SARTN) for time to maturity $2\Delta$ (two business days). The latter is not completely correct because SARON is a collateral over-night rate and tomorrow-next is a call money rate for receiving money tomorrow which has to be paid back the next business day. Moreover, we have the SAR for times to maturity of 1 week (SAR1W), 2 weeks (SAR2W), 1 month (SAR1M) and 3 months (SAR3M), see also [54].

• **Short to medium times to maturity.** The LIBOR reflects times to maturity which correspond to 1 month (LIBOR1M), 3 months (LIBOR3M), 6 months (LIBOR6M) and 12 months (LIBOR12M) in the London interbank market.

• **Medium to long times to maturity.** The SWCNB is based on Swiss government bonds and it is used for times to maturity which correspond to 2 years (SWCNB2Y), 3 years (SWCNB3Y), 4 years (SWCNB4Y), 5 years (SWCNB5Y), 7 years (SWCNB7Y), 10 years (SWCNB10Y), 20 years (SWCNB20Y) and 30 years (SWCNB30Y). This data is available from December 8, 1999, and we set September 15, 2014 to be the last observation date. Of course, SAR, LIBOR and SWCNB do not exactly model risk-free zero-coupon bonds and these different classes of instruments are not completely consistent because prices are determined slightly differently for each class. In particular, this can be seen during the 2008–2009 financial crisis. However, this data is in many cases the best approximation to CHF risk-free zero-coupon yields that is available. For the longest times to maturity of SWCNB one may also raise issues about liquidity of these instruments because insurance companies typically run a buy-and-hold strategy for long term bonds.

In Figures B.3-B.6 we compute the realised volatility $\hat{\text{RCov}}(t, \tau, \tau)$ of yield curve observations $(\tilde{y}_T(k))_{k=t-K+1,...,t}$ for different times to maturity $\tau\Delta$ and window length $K$, see equation (5.7). In Figures B.2 and B.6 we observe that SAR fits SWCNB better than LIBOR after the financial crisis of 2008. For this reason we decide to drop LIBOR and build daily yield curves from SAR and SWCNB, only. The mismatch between LIBOR, SAR and SWCNB is attributable to differences in liquidity and credit risk of the underlying instruments.

6.2 Model selection

In this numerical example we restrict ourselves to multifactor Vasicek models with $\beta$ and $\alpha$ of diagonal form

$$\beta = \text{diag}(\beta_{11}, \ldots, \beta_{nn}), \quad \text{and} \quad \alpha = \text{diag}(\alpha_{11}, \ldots, \alpha_{nn}),$$

where $-1 < \beta_{11}, \ldots, \beta_{nn}, \alpha_{11}, \ldots, \alpha_{nn} < 1$. In the following we explain exactly how to perform the delicate task of parameter estimation in the multifactor Vasicek model (2.1)-(2.2) using the procedure explained in Section 5.
Discussion of identification assumptions. We select short times to maturity (SAR) to estimate parameters $b$, $\beta$, $\Sigma$, $a$, and $\alpha$. This is reasonable because these parameters describe the dynamics of the factor process, and thus of the spot rate. As we are working on a small (daily) time grid, asymptotic formulas (5.5) and (5.6) are expected to give good approximations. Additionally, it is reasonable to assume that the noise covariance matrix $S$ in data-generating model (5.3) is negligible compared to (5.6). Therefore, we can estimate the left hand side of (5.6) by the realised covariation of observed yields, see estimator (5.7). Then we determine the Hull-White extension $\theta$ in order to match the prevailing yield curve interpolated from SAR and SWCNB.

Determination of the number of factors. We need to determine the appropriate number of factors $n$. The more factors we use, the better we can fit the model to the data. However, the dimensionality of the estimation problem increases quadratically in the number of factors, and the model may become over-parametrised. Therefore, we look for a trade-off between accuracy of the model and the number of parameters used. In Figure B.7 we determine $\beta_{11}, \ldots, \beta_{nn}$ and $\Sigma$ by solving optimisation (5.8) numerically for three observation dates and $n = 2, 3$. A 3-factor model is able to capture rather accurately the dependence on the time to maturity $\tau$. In Figures B.8-B.10 we compare the realised volatility of the numerical solution of (5.8) to the market realised volatility for all observation dates. We observe that in several periods the 2-factor model is not able to fit the SAR realised volatilities accurately for all times to maturities. The 3-factor model achieves an accurate fit for most observation dates. The model exhibits small mismatches in 2001, 2008-2009, and 2011-2012. These are periods characterised by a sharp reduction in interest rates in response to financial crises. In September 2011, following strong appreciation of the Swiss Franc with respect to the Euro, the SNB pledged to no longer tolerate Euro-Franc exchange rates below the minimum rate of 1.20, effectively enforcing a currency floor for more than three years. As a consequence of the European sovereign debt crisis and the intervention of the SNB starting from 2011 we have a long period of very low (even negative) interest rates.

Determination of Vasiček parameters. Considering the results of Figures B.8-B.10 we restrict ourselves from now on to 3-factor Vasiček models with parameters $a, b \in \mathbb{R}^3$ and

$$
\beta = \text{diag}(\beta_{11}, \beta_{22}, \beta_{33}), \quad \alpha = \text{diag}(\alpha_{22}, \alpha_{22}, \alpha_{33}), \quad \Sigma^2 = \begin{pmatrix}
\Sigma_{11}^2 & 0 & 0 \\
\Sigma_{22}^2 & \Sigma_{22}^2 & 0 \\
\Sigma_{33}^2 & \Sigma_{33}^2 & \Sigma_{33}^2
\end{pmatrix},
$$

where $-1 \leq \beta_{11}, \beta_{22}, \beta_{33}, \alpha_{11}, \alpha_{22}, \alpha_{33} \leq 1$, $\Sigma_{11}^2, \Sigma_{22}^2, \Sigma_{33}^2 > 0$ and $\Sigma_{21}^2, \Sigma_{31}^2, \Sigma_{32}^2 \in \mathbb{R}$.

In Figures B.11-B.13 we plot the numerical solution of optimisations (5.8) and (5.9) for all observation dates. The parameters are reasonable for most of the observation dates. We observe that the estimates of $\beta_{11}$ are close to one for all observation dates. Our values for the speed of mean reversion are reasonable on a daily time grid. Note that $\beta$ scales as $\beta^d$ on a $d$-days time
grid, see Section 5.3. The speeds of mean reversion of $X_2$ and $X_3$ are higher than of $X_1$ for most of the observation dates. We also see that the volatility of $X_1$ is lower than of $X_2$ and $X_3$. In 2011 we observe large spikes in the factor volatilities. Starting from 2011 we have a period with strong correlations among the factors. From these results we conclude that the 3-factor Vasićek model is reasonable for Swiss interest rates. Particularly challenging for the estimation is the period 2011–2014 of low interest rates following the European sovereign debt crisis and the SNB intervention.

In Figure B.11 (rhs) we observe that the difference in the speeds of mean-reversion under the risk-neutral and real world measures is negligible. The difference between $b$ and $a$ is considerable in certain time periods. From the estimation results we conclude that a constant market price of risk assumption is reasonable and set from now on $\Lambda = 0$. In Figure B.14 we compute the objective function of optimisation (5.9) for ($b, \beta, \Sigma, a, \alpha$) = (0, $\hat{\Sigma}^{RCov}$, $\Sigma^{RCov}$, 0, $\hat{\beta}^{RCov}$) and compare it to the numerical solution ($b^{MLE}$, $\hat{\beta}^{RCov}$, $\Sigma^{RCov}$, $a^{MLE}$, $\hat{\beta}^{RCov}$). We observe that in 2003–2005 and 2010–2014 the parameter configuration (0, $\hat{\beta}^{RCov}$, $\Sigma^{RCov}$, 0, $\hat{\beta}^{RCov}$) is nearly optimal. In these periods we have very low interest rates and therefore estimates of $b$ and $a$ close to zero are reasonable. Given the estimated parameters we calibrate the Hull-White extension by equation (3.2) to the full yield curve interpolated from SAR and SWCNB, see Figure B.15. We point out that our fitting method is not a purely statistical procedure; rather, it is a combination of estimation and calibration in accordance with the paradigm of robust calibration as explained in [43].

**Selection of a model for the Vasićek parameters.** In the following we use the CRC approach to construct a modification of the Vasićek model with stochastic volatility. We model the process $(\Sigma(t))_{t \in \mathbb{N}_0}$ by a Heston [46] like approach. We assume deterministic correlations among the factors and stochastic volatility given by

$$
\begin{pmatrix}
\Sigma_{11}(t) \\
\Sigma_{22}(t) \\
\Sigma_{33}(t)
\end{pmatrix} = \varphi + \phi \begin{pmatrix}
\Sigma_{11}(t-1) \\
\Sigma_{22}(t-1) \\
\Sigma_{33}(t-1)
\end{pmatrix} + \begin{pmatrix}
\sqrt{\Sigma_{11}(t-1)} & 0 & 0 \\
0 & \sqrt{\Sigma_{22}(t-1)} & 0 \\
0 & 0 & \sqrt{\Sigma_{33}(t-1)}
\end{pmatrix} \Phi^\dagger \tilde{\varepsilon}(t),
$$

where $\varphi \in \mathbb{R}^3$, $\phi = \text{diag}(\varphi_{11}, \varphi_{22}, \varphi_{33}) \in \mathbb{R}^{3 \times 3}$, $\Phi^\dagger \in \mathbb{R}^{3 \times 3}$ non-singular and, for each $t \in \mathbb{N}$, $\tilde{\varepsilon}(t)$ has a standard Gaussian distribution under $\mathbb{P}$, conditionally given $F(t-1)$. Moreover, we assume that $(\varepsilon(t), \tilde{\varepsilon}(t))$ is multivariate Gaussian under $\mathbb{P}$, conditionally given $F(t-1)$. Note that $\varepsilon(t)$ and $\tilde{\varepsilon}(t)$ are allowed to be correlated. The matrix valued process $(\Sigma(t))_{t \in \mathbb{N}_0}$ is constructed combining this stochastic volatility model with fixed correlation coefficients. This model is able to capture the stylised fact that volatility appears to be more noisy in high volatility periods, see Figure B.12.

We use the volatility time series of Figure B.12 to specify $\varphi$, $\phi$ and $\Phi$. We rewrite the equation for the evolution of the volatility as

$$
\frac{\Sigma_{ii}(t)}{\sqrt{\Sigma_{ii}(t-1)}} = \frac{\varphi_i}{\sqrt{\Sigma_{ii}(t-1)}} + \phi_i \sqrt{\Sigma_{ii}(t-1)} + (\Phi^\dagger \tilde{\varepsilon}(t)), \quad i = 1, 2, 3,
$$

and use least square regression to estimate $\varphi$, $\phi$ and $\Phi$. From the regression residuals we estimate the correlations between $\varepsilon(t)$ and $\tilde{\varepsilon}(t)$. Figures B.16-B.18 show the estimates of $\varphi$, $\phi$ and $\Phi$. 

115
6.3 Simulation and back-testing

Section 6.2 provides a full specification of the Vasiček CRC model under the risk-neutral and real world probability measures. Various model quantities of interest in applications can then be calculated by simulation.

Simulation. The CRC approach has the remarkable property that yield curve increments can be simulated accurately and efficiently using Theorem 3.1 and Corollary 4.1. In contrast, spot rate models with stochastic volatility without CRC have serious computational drawbacks. In such models the calculation of the prevailing yield curve for given state variables requires Monte Carlo simulation. Therefore, the simulation of future yield curves would require nested simulations.

Back-testing. We backtest properties of the monthly returns of a buy & hold portfolio investing equal proportions of wealth in the zero-coupon bonds with times to maturity of 2, 3, 4, 5, 6, and 9 months, and 1, 2, 3, 5, 7, and 10 years. We divide the sample in disjoint monthly periods and calculate the monthly return of this portfolio assuming that at the beginning of each period we invest in the bonds with these times to maturity in equal proportions of wealth. The returns and some summary statistics are shown in Figure B.19. We observe that the returns are positively skewed, leptokurtic, and have heavier tails than the Gaussian distribution. These stylised facts are essential in applications.

For each monthly period we select a 3-factor Vasiček model and its CRC counterpart with stochastic volatility. Then we simulate for each period realisations of the returns of the test portfolio. By construction the Vasiček model generates Gaussian log-returns and is unable to reproduce the stylised facts of the sample, see Tables B.1 and B.2 and Figure B.20. Increasing the number of factors does not help much because the log-returns remain Gaussian. On the other hand, CRC of the Vasiček model with stochastic volatility provides additional modelling flexibility. In particular, we can see from the statistics in Table B.2 and the confidence intervals in Figure B.20 that the model matches the return distribution better than the Vasiček model. The three-factor CRC Vasiček model is a parsimonious and tractable alternative which provides reasonable results.

7 Conclusions

Flexibility and tractability. Consistent recalibration of the multifactor Vasiček model provides a tractable extension which allows parameters to follow stochastic processes. The additional flexibility can lead to better fits of yield curve dynamics and return distributions, as we demonstrated in our numerical example. Nevertheless, the model remains tractable. In particular, yield curves can be simulated efficiently using Theorem 3.1 and Corollary 4.1. This allows one to efficiently calculate model quantities of interest in risk management, forecasting, and pricing.
**Model selection.** CRC models are selected from data in accordance with the robust calibration principle of [43]. First, historical parameters, market-prices of risk, and Hull-White extensions are inferred using a combination of volatility estimation, MLE, and calibration to the prevailing yield curve via formulas (5.8), (5.9), (5.10), (5.11), and (3.2). The only choices in this inference procedure are the number of factors of the Vasiček model and the window length $K$. Then, as a second step, the time series of estimated historical parameters are used to select a model for the parameter evolution. This results in a complete specification of the CRC model under the real world and the pricing measure.

**Application to modelling of Swiss interest rate.** We fitted a 3-factor Vasiček CRC model with stochastic volatility to Swiss interest rate data. The model achieves a reasonably good fit in most time periods. The tractability of CRC allowed us to compute several model quantities by simulation. We looked at the historical performance of a representative buy & hold portfolio of Swiss bonds and concluded that a multifactor Vasiček model is unable to describe the returns of this portfolio accurately. In contrast, the CRC version of the model provides the necessary flexibility for a good fit.
A Proofs

Proof of Theorem 2.2. We prove Theorem 2.2 by induction as in [82, Theorem 3.16] where ZCB prices are derived under the assumption that $\beta$ and $\Delta$ are diagonal matrices. We have $P(m-1, m) = \exp(-\Gamma^T X(m-1) \Delta)$, which proves the claim for $t = m - 1$. Assume that Theorem 2.2 holds for $t + 1 \in \{2, \ldots, m - 1\}$. We verify that it also holds for $t \in \{1, \ldots, m - 2\}$. Under equivalent martingale measure $P^*$ we have using the tower property for conditional expectations and the induction assumption

\[
P(t, m) = \exp \left\{ -\Gamma^T X(t) \Delta \right\} E^*[E^\left[ \exp \left\{ -\Delta \sum_{s=t+1}^{m-1} 1^T X(s) \right\} | F(t+1) \right] | F(t) |]
\]

Finally, note that the recursive formula for $B(\cdot, \cdot)$ implies

\[
B(t, m) = \sum_{i=0}^{m-1} \left( \beta^\top \right)^i \Delta = \left( 1 - \beta^\top \right)^{-1} \left( 1 - \left( \beta^\top \right)^{m-1} \right) \Delta.
\]

This concludes the proof. □

Proof of Theorem 2.3. Proof goes by induction as the proof of Theorem 2.2. □

Proof of Theorem 2.4. First, observe that the condition $y(i)(k) = y$ imposes conditions only on the values $\theta(1), \ldots, \theta(M - 1)$. Secondly, note that the vector $\theta$ such that the condition is satisfied can be calculated recursively in the following way.

(i) First component $\theta_1$. We have $A(k+1, k+2) = 0$, $B(k+1, k+2) = 1\Delta$ and

\[
A(k, k+2) = -\Gamma^T b \Delta - \theta(1) \Delta + \frac{1}{2} \Gamma^T \Sigma \Delta^2,
\]

see Theorem 2.3. Solving the last equation for $\theta_1$ we have

\[
\theta_1 = \frac{1}{2} \Gamma^T \Sigma \Delta - \frac{1}{2} \Gamma^T b - A(k, k+2) \Delta^{-1}.
\]

From (5.6) we obtain

\[
A(k, k+2) = \frac{1}{1 - \beta^2} \left( 1 - \beta^{-1} \right) \Delta - 2\gamma_1 \Delta.
\]

This is equivalent to

\[
\theta_1 = \frac{1}{2} \Gamma^T \Sigma \Delta - \frac{1}{2} \Gamma^T b - A(k, k+2) \Delta^{-1} = \frac{1}{2} \Gamma^T \Sigma \Delta - \frac{1}{2} \Gamma^T b - A(k, k+2) \Delta^{-1} + 2\gamma_1 \Delta.
\]

(ii) Recursion $i \rightarrow i + 1$. Assume we have determined $\theta_1, \ldots, \theta_i$, for $i = 1, \ldots, M - 2$. We want to determine $\theta_{i+1}$. We have $A(k+i+1, k+i+2) = 0$ and iteration implies

\[
A(k, k+i+2) = \frac{1}{2} \sum_{s=k+1}^{k+i+1} B(s, k+i+2) (b + \theta(s-k) e_1) + \frac{1}{2} \sum_{s=k+1}^{k+i+1} B(s, k+i+2) \Sigma B(s, k+i+2).
\]

118
Solving the last equation for $\theta_{i+1}$ and using $B(k+i+1,k+i+2) = 1$ we have
$$\theta_{i+1} = -\frac{1}{\Delta} A(k)(k,k+2) \theta_k - \frac{1}{\Delta} \sum_{s=k+1}^{k+i+1} B(s,k+i+2) (b + \theta(s-k)e_1) - 1^\top b$$
$$+ \frac{1}{2\Delta} \sum_{s=k+1}^{k+i+1} B(s,k+i+2) \Sigma B(s,k+i+2).$$

From (5.6) we obtain
$$A^{(k)}(k,k+i+2) = 1^\top \left(1 - \beta^{i+2}\right) (1 - \beta)^{-1} x \Delta - y_{i+2}(i+2) \Delta.$$ This is equivalent to
$$\theta_{i+1} = (i+2) y_{i+2} - 1^\top \left(1 - \beta^{i+2}\right) (1 - \beta)^{-1} x - 1^\top \frac{1}{\Delta} \sum_{s=k+1}^{k+i+1} B(s,k+i+2) \Sigma B(s,k+i+2)$$
$$- \frac{1}{\Delta} \sum_{s=k+1}^{k+i+1} B(s,k+i+2) \theta_{s-k} + \frac{1}{2\Delta} \sum_{s=k+1}^{k+i+1} B(s,k+i+2) \Sigma B(s,k+i+2).$$ (A.2)

This recursion allows to determine the components of $\theta$. Note that equation (A.2) can be written as
$$(C(\beta)\theta)_{i+1} = z_{i+1}(b,\beta,\Sigma,x,y), \quad i = 1, \ldots, M - 2.$$  

Observe that the lower triangular matrix $C(\beta)$ is invertible since $\text{det}(C(\beta)) = \Delta^{M-1} > 0$. Hence, equations (A.1) and (A.2) prove (2.8).

\begin{proof}[Proof of Theorem 3.1] We add and subtract $-A^{(k)}(k,m) + B^{(k)}(k,m)^\top X(k)$ to equation (3.4) and obtain
$$Y(k+1,m)(m-(k+1)) \Delta = A^{(k)}(k,m) - A^{(k)}(k+1,m) - A^{(k)}(k,m)$$
$$+ B^{(k)}(k,m)^\top X(k) - B^{(k)}(k,m)^\top X(k) + B^{(k)}(k+1,m)^\top \left(b(k) + \theta(k)(k+1)e_1 + \beta(k)x(k) + \Sigma(k)^{1/2} \varepsilon^*(k+1)\right).$$ (A.3)

We have the two identities
$$-A^{(k)}(k,m) + B^{(k)}(k,m)^\top X(k) = Y(k,m)(m-k) \Delta,$$
$$A^{(k)}(k,m) - A^{(k)}(k+1,m) = -B^{(k)}(k,m)^\top \left(b(k) + \theta(k)(k+1)e_1 \right) + \frac{1}{2} B^{(k)}(k+1,m)^\top \Sigma(k) B^{(k)}(k+1,m).$$

Therefore, the right-hand side of equality (A.3) is rewritten as
$$Y(k+1,m)(m-(k+1)) \Delta = Y(k,m)(m-k) \Delta + \left(B(k+1,m)^\top \beta(k) - B(k,m)^\top \right) X(k)$$
$$+ \frac{1}{2} B^{(k+1,m)^\top \Sigma(k) B^{(k+1,m)} + B^{(k+1,m)^\top \Sigma(k)^{1/2} \varepsilon^*(k+1).$$

Observe that
$$B^{(k)}(k+1,m)^\top \beta(k) = \left( \sum_{s=0}^{m-k-2} \left(\beta^s(k)\right)^\top 1 \right) \beta(k) \Delta = 1^\top \sum_{s=1}^{m-k-1} \beta(k)^s \Delta = 1^\top B(k,m)^\top - 1^\top \Delta,$$

and that $Y(k, k+1) = 1^\top X(k)$. This proves the claim. \qed

\end{proof}
B Figures

Figure B.1: Yield rates (lhs) Swiss Average Rate (SAR); (rhs) London InterBank Offered Rate (LIBOR) from December 8, 1999, until September 15, 2014.

Figure B.2: Yield rates (lhs) Swiss Confederation Bond (SWCNB); (rhs) a selection of SAR, LIBOR and SWCNB from December 8, 1999, until September 15, 2014. Note that LIBOR looks rather differently from SAR and SWCNB after the financial crisis of 2008.
Figure B.3: SAR realised volatility $\hat{\text{RCov}}(t, \tau, \tau)^{\frac{1}{2}}$ for $\tau = 1, 2, 5, 10, 21, 63$, window length $K = 21$ (lhs) and $K = 126$ (rhs).

Figure B.4: LIBOR realised volatility $\hat{\text{RCov}}(t, \tau, \tau)^{\frac{1}{2}}$ for $\tau = 21, 63, 126, 252$, window length $K = 21$ (lhs) and $K = 126$ (rhs).
Figure B.5: SWCNB realised volatility $\hat{R}_{\text{Cov}}(t, \tau, \tau)^{1/2}$ for $\tau/252 = 2, 3, 4, 5, 7, 10, 20, 30$, window length $K = 21$ (lhs) and $K = 126$ (rhs).

Figure B.6: A selection of SAR, LIBOR and SWCNB realised volatility $\hat{R}_{\text{Cov}}(t, \tau, \tau)^{1/2}$ for $\tau = 1, 63, 252, 504$, window length $K = 21$ (lhs) and $K = 126$ (rhs). Note that LIBOR looks rather differently from SAR and SWCNB after the financial crisis of 2008.
Figure B.7: SAR realised volatility $\hat{RCov}(t, \tau, \tau)^{\frac{1}{2}}$ for $K = 126$, $\tau = 1, 2, 5, 10, 21, 63$ and three observation dates compared to the realised volatility of the 2- (lhs) and 3-factor (rhs) Vasicek model fitted by optimisation (5.8) for $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$ and $w_{ij} = 1_{\{i=j\}}$. The 3-factor model achieves an accurate fit.

Figure B.8: SAR realised volatility $\hat{RCov}(t, \tau, \tau)^{\frac{1}{2}}$ for $K = 126$, $\tau = 1$ (lhs), $\tau = 2$ (rhs) and all observation dates compared to the realised volatility of the 2- and 3-factor Vasicek models fitted by optimisation (5.8) for $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$ and $w_{ij} = 1_{\{i=j\}}$. 

123
Figure B.9: SAR realised volatility $\hat{\text{RCov}}(t, \tau, \tau)_{12}$ for $K = 126$, $\tau = 5$ (lhs), $\tau = 10$ (rhs) and all observation dates compared to the realised volatility of the 2- and 3-factor Vasiček models fitted by optimisation (5.8) for $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$ and $w_{ij} = 1_{\{i=j\}}$.

Figure B.10: SAR realised volatility $\hat{\text{RCov}}(t, \tau, \tau)_{12}$ for $K = 126$, $\tau = 21$ (lhs), $\tau = 63$ (rhs) and all observation dates compared to the realised volatility of the 2- and 3-factor Vasiček models fitted by optimisation (5.8) for $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$ and $w_{ij} = 1_{\{i=j\}}$. 
Figure B.11: Estimation of $\beta_{11}$, $\beta_{22}$ and $\beta_{33}$ (lhs), and $(\Sigma^2 \Lambda)_{11} = \beta_{11} - \alpha_{11}$, $(\Sigma^2 \Lambda)_{22} = \beta_{22} - \alpha_{22}$ and $(\Sigma^2 \Lambda)_{33} = \beta_{33} - \alpha_{33}$ (rhs) by optimisations (5.8) and (5.9) in the 3-factor model for $K = 126$, $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$, $w_{ij} = 1_{(i=j)}$ and $S = 10^{-5} \cdot I$. The values determine the speed of mean reversion of the factors. Since we are considering a daily time grid, values close to one (slow mean reversion) are reasonable. We observe that the difference in the speed of mean-reversion under the risk-neutral and real world measures is negligible.

Figure B.12: Estimation of $\Sigma_{11}$, $\Sigma_{22}$ and $\Sigma_{33}$ (lhs), and correlations $\rho_{21}$, $\rho_{11}$ and $\rho_{22}$ (rhs) by optimisation (5.8) in the 3-factor model for $K = 126$, $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$ and $w_{ij} = 1_{(i=j)}$. We observe large spikes in the volatilities and strong correlations among the factors during the European sovereign debt crisis and after SNB intervention in 2011.
Figure B.13: Estimation of $b_1$, $b_2$ and $b_3$ (lhs), and $(\Sigma^2 \lambda)_1 = b_1 - a_1$, $(\Sigma^2 \lambda)_2 = b_2 - a_2$ and $(\Sigma^2 \lambda)_3 = b_3 - a_3$ (rhs) by optimisations (5.8) and (5.9) in the 3-factor model for $K = 126$, $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$, $w_{ij} = 1_{\{i=j\}}$ and $S = 10^{-5} \cdot 1$. The difference between $b$ and $a$ is considerable in 2000–2002 and 2006–2009.

Figure B.14: Objective function $\log \mathcal{L}_t$ (lhs), and $(\Sigma^2 \lambda)_1 = b_1 - a_1$, $(\Sigma^2 \lambda)_2 = b_2 - a_2$ and $(\Sigma^2 \lambda)_3 = b_3 - a_3$ (rhs) given by optimisation (5.9) in the 3-factor model for $K = 126$, $M = 6$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 5$, $\tau_4 = 10$, $\tau_5 = 21$, $\tau_6 = 63$, $w_{ij} = 1_{\{i=j\}}$ and $S = 10^{-5} \cdot 1$. We compare the value of the objective function for $(b, \beta, \Sigma, \alpha) = (0, \beta^{RCov}, \Sigma^{RCov}, 0, \alpha^{RCov})$ and the numerical solution of the optimisation. The configuration $(0, \beta^{RCov}, \Sigma^{RCov}, 0, \alpha^{RCov})$ is almost optimal in low interest rate times.
Figure B.15: 3-factor Hull-White extended Vasicek yield curve (lhs) and Hull-White extension $\theta$ (rhs) as of 29 September, 2006. The parameters are estimated as in Figure B.11-B.13. The initial factors are obtained from the Kalman filter for the estimated parameters. The calibration of the Hull-White extension requires yields on a time to maturity grid of size $\Delta$. These are interpolated from SAR and SWCNB using cubic splines.

Figure B.16: Estimation of $\varphi_1$, $\varphi_2$ and $\varphi_3$ by least square regression (two different scales). We use a time window of 252 observations for the regression.
Estimates of $\phi_{11}$, $\phi_{22}$, and $\phi_{33}$.

Figure B.17: Estimation of $\phi_{11}$, $\phi_{22}$, and $\phi_{33}$ (lhs), and $\Phi_{11}$, $\Phi_{22}$ and $\Phi_{33}$ (rhs) by least square regression. We use a time window of 252 observations for the regression.

Estimates of correlations $\rho_{ij} = \Phi_{ij} / (\Phi_{ii} \Phi_{jj})$.

Figure B.18: Estimation of correlations $\tilde{\rho}_{21}$, $\tilde{\rho}_{31}$ and $\tilde{\rho}_{32}$ (lhs), and correlations $\text{Cor} [\varepsilon(t), \tilde{\varepsilon}(t) \mid \mathcal{F}(t - 1)]$ (rhs). We use a time window of 252 observation for the regression. The residuals $\varepsilon$ are calculated using the parameter estimates of Figures B.11-B.13.
Figure B.19: Logarithmic monthly returns of a buy & hold portfolio investing in equal wealth proportions in the zero-coupon bonds with times to maturity of 2, 3, 4, 5, 6 and 9 months, and 1, 2, 3, 5, 7 and 10 years. For each monthly period we calculate the logarithmic return of this portfolio assuming that at the beginning of each period we are invested in the bonds with these times to maturity in equal proportions of wealth.

Table B.1: Statistics computed from simulations of the test portfolio returns for some of the monthly periods in the Vasiček model. For each monthly period we simulate $10^4$ realisations.
Table B.2: Statistics computed from simulations of the test portfolio returns for some of the monthly periods in the CRC counterpart of the Vasiček model with stochastic volatility. For each monthly period we simulate $10^4$ realisations.

<table>
<thead>
<tr>
<th>Time</th>
<th>Mean (%)</th>
<th>Volatility (%)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>1st Quartile (%)</th>
<th>3rd Quartile (%)</th>
<th>5% Quantile (%)</th>
<th>95% Quantile (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002-01-03</td>
<td>0.174</td>
<td>0.569</td>
<td>0.053</td>
<td>3.675</td>
<td>-0.202</td>
<td>0.568</td>
<td>-0.726</td>
<td>1.045</td>
</tr>
<tr>
<td>2003-01-07</td>
<td>0.228</td>
<td>1.227</td>
<td>0.149</td>
<td>3.546</td>
<td>-0.573</td>
<td>0.999</td>
<td>-1.738</td>
<td>2.291</td>
</tr>
<tr>
<td>2004-01-20</td>
<td>0.172</td>
<td>0.227</td>
<td>-0.065</td>
<td>3.195</td>
<td>0.025</td>
<td>0.319</td>
<td>-0.216</td>
<td>0.544</td>
</tr>
<tr>
<td>2005-01-21</td>
<td>0.178</td>
<td>0.848</td>
<td>0.128</td>
<td>3.990</td>
<td>-0.401</td>
<td>0.730</td>
<td>-1.162</td>
<td>1.600</td>
</tr>
<tr>
<td>2006-01-12</td>
<td>0.165</td>
<td>0.190</td>
<td>0.226</td>
<td>3.373</td>
<td>0.042</td>
<td>0.281</td>
<td>-0.149</td>
<td>0.486</td>
</tr>
<tr>
<td>2007-01-25</td>
<td>0.147</td>
<td>0.101</td>
<td>0.127</td>
<td>3.253</td>
<td>0.080</td>
<td>0.213</td>
<td>-0.012</td>
<td>0.310</td>
</tr>
<tr>
<td>2008-01-09</td>
<td>0.215</td>
<td>0.645</td>
<td>-0.029</td>
<td>3.200</td>
<td>-0.192</td>
<td>0.629</td>
<td>-0.809</td>
<td>1.257</td>
</tr>
<tr>
<td>2009-01-20</td>
<td>0.237</td>
<td>1.313</td>
<td>0.906</td>
<td>4.720</td>
<td>-0.517</td>
<td>1.015</td>
<td>-1.745</td>
<td>2.134</td>
</tr>
<tr>
<td>2010-01-06</td>
<td>0.156</td>
<td>0.480</td>
<td>1.045</td>
<td>4.033</td>
<td>-0.101</td>
<td>0.402</td>
<td>-0.550</td>
<td>0.827</td>
</tr>
<tr>
<td>2011-01-17</td>
<td>0.158</td>
<td>0.506</td>
<td>-0.315</td>
<td>3.709</td>
<td>-0.161</td>
<td>0.481</td>
<td>-0.709</td>
<td>0.988</td>
</tr>
<tr>
<td>2012-01-25</td>
<td>0.153</td>
<td>0.512</td>
<td>1.120</td>
<td>4.124</td>
<td>-0.139</td>
<td>0.414</td>
<td>-0.604</td>
<td>0.916</td>
</tr>
<tr>
<td>2013-01-11</td>
<td>0.146</td>
<td>0.710</td>
<td>0.319</td>
<td>5.987</td>
<td>-0.169</td>
<td>0.422</td>
<td>-0.789</td>
<td>1.073</td>
</tr>
<tr>
<td>2014-01-21</td>
<td>0.152</td>
<td>0.586</td>
<td>-0.621</td>
<td>6.521</td>
<td>-0.127</td>
<td>0.454</td>
<td>-0.616</td>
<td>0.895</td>
</tr>
</tbody>
</table>

Figure B.20: Confidence levels computed from simulations of the test portfolio returns in the Vasiček model (lhs), and its CRC counterpart with stochastic volatility (rhs). For each monthly period we simulate $10^4$ realisations.
Philipp Harms, David Stefanovits, Josef Teichmann.

*Exponential moments of time-inhomogeneous affine processes.*
EXPONENTIAL MOMENTS OF TIME-INHOMOGENEOUS AFFINE PROCESSES

PHILIPP HARMS, DAVID STEFANOVITS, AND JOSEF TEICHMANN

Abstract. We show that the affine transform formula, which relates real exponential moments of affine processes to solutions of Riccati-type differential equations, holds for affine semimartingales. This extends some of the results of Keller-Ressel and Mayerhofer [56] to the time-inhomogeneous case.

1. Definitions and main results

We extend Theorem 2.14 of Keller-Ressel and Mayerhofer [56] to the time-inhomogeneous case. We use the same setup and notation and make changes when necessary.

1.1. Affine processes. For some \(d \geq 1\), let \(D\) be a non-empty convex subset of \(\mathbb{R}^d\) with Borel sigma algebra \(\mathcal{B}(D)\). Without loss of generality (see Remark 2.3 in [56]) we may assume that \(D\) contains 0 and that the linear span of \(D\) is the full space \(\mathbb{R}^d\). Under this assumption it follows in particular that the interior \(D^\circ\) of \(D\) is non-empty. Associated to \(D\) is the set

\[
U = \{ u \in \mathbb{C}^d : x \mapsto e^{\langle x, u \rangle} \text{ is bounded on } D \},
\]

where \(\langle , \rangle\) denotes an inner product on \(\mathbb{R}^d\).

Let \((\Omega, \mathcal{F}, \mathbb{F})\) be a filtered space, with \(\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}\) a right-continuous filtration, and let \(X\) be a càdlàg, \(\mathbb{F}\)-adapted process on \(\Omega\) taking values in \(D\). For each \(s \geq 0\) and \(x \in D\), let \(\mathbb{P}^{s,x}\) be a probability measure defined on the filtered space \((\Omega, \mathcal{F}, \mathbb{F})\). We assume that \(\mathbb{F}\) contains the null sets of \(\mathbb{P}^{s,x}\), for each \(s \geq 0\) and \(x \in D\). Moreover, we assume that \(X\) is a time-inhomogeneous conservative Markov process with respect to the family of measures \(\mathbb{P}^{s,x}\). More precisely, the transition kernel of \(X\), which is defined by

\[
p_{s,t}(x, A) = \mathbb{P}^{s,x}(X_t \in A), \quad (s, t \geq 0, x \in D, A \in \mathcal{B}(D)),
\]

satisfies the following properties:

(a) \(x \mapsto p_{s,t}(x, A)\) is \(\mathcal{B}(D)\)-measurable for all \(s, t \geq 0, A \in \mathcal{B}(A)\),

(b) \(p_{s,t}(x, D) = 1\) for all \(s, t \geq 0, x \in D\),

(c) \(p_{s,t}(x, \{x\}) = 1\) for all \(s \geq t \geq 0, x \in D\), and

(d) the Chapman-Kolmogorov equation

\[
p_{r,t}(x, A) = \int_D p_{s,t}(y, A)p_{r,s}(x, dy)
\]

holds for every \(0 \leq r \leq s \leq t, x \in D\), and \(A \in \mathcal{B}(D)\).
Remark 1.1. Note that it is implicit in the definitions that $P^{s,x}(X_t = x, \forall t \leq s) = 1$.

Definition 1.2 (Affine process). The process $X$ is called affine with state space $D$, if its transition kernel $p_{s,t}(x, A)$ satisfies the following:

(i) it is stochastically continuous, i.e. for each $x \in D$, $p_{s,t}(x, \cdot)$ is weakly continuous in $(s,t) \in \mathbb{R}_{\geq 0}^2$, and

(ii) there exist functions $\phi : \mathbb{R}_{\geq 0}^2 \times U \to \mathbb{C}$ and $\psi : \mathbb{R}_{\geq 0} \times U \to \mathbb{C}^d$ such that

$$\int_D e^{(u,\xi)} p_{s,t}(x, d\xi) = e^{\phi(s,t,u)+(x,\psi(s,t,u))}$$

for all $s, t \geq 0, x \in D,$ and $u \in U$.

Assumption 1.3. For each $(t, x) \in \mathbb{R}_{>0} \times D$, the process $X$ is a $\mathbb{P}^{t,x}$-semimartingale. Its semimartingale characteristics with respect to the truncation function $h(\xi) = 1(\|\xi\| \leq 1)\xi$ are given by

\begin{align}
A_s &= \int_t^{t+s} a_r(X_{r-}) dr, \\
B_s &= \int_0^{t+s} b_r(X_{r-}) dr,
\end{align}

where $a_r(x), b_r(x)$ and $K_r(x, d\xi)$ are affine functions of the form

\begin{align}
a_r(x) &= a_r + x_1 \alpha_r^1 + \cdots + x_d \alpha_r^d, \\
b_r(x) &= b_r + x_1 \beta_r^1 + \cdots + x_d \beta_r^d, \\
K_r(x, d\xi) &= m_r(d\xi) + x_1 \mu_r^1(d\xi) + \cdots + x_d \mu_r^d(d\xi)
\end{align}

and for each $(r, x) \in \mathbb{R}_{\geq 0} \times D$ it holds that $a_r(x)$ is a positive semidefinite $d \times d$ matrix, $b_r(x)$ is a $\mathbb{R}^d$-vector, and $K_r(x, d\xi)$ is a Radon measure on $\mathbb{R}^d$, satisfying

$$\int_{\mathbb{R}^d} (\|\xi\|^2 \wedge 1) K_r(x, d\xi) < \infty$$

and $K_r(x, \{0\}) = 0$.

1.2. Real moments of affine processes.

Definition 1.4. Given $a_t(x), b_t(x), K_t(x, d\xi)$ as in Assumption 1.3, define for each $t \geq 0$ and $x \in D$ the function $\mathcal{R}_{t,x} : \mathbb{R}^d \to (-\infty, \infty]$ by

$$\mathcal{R}_{t,x}(y) = \frac{1}{2} \langle y, a_t(x)y \rangle + \langle b_t(x), y \rangle$$

\begin{align}
&+ \int_{\mathbb{R}^d \setminus \{0\}} \left( e^{\langle y, \xi \rangle} - 1 - (h(\xi), y) \right) K_t(x, d\xi).
\end{align}

For each fixed $(t, x)$, the function $\mathcal{R}_{t,x}$ is a convex and lower semicontinuous. Lower semicontinuity follows from Fatou’s Lemma applied to the integral with respect to $K_t(x, d\xi)$. As for any convex function, the effective domain $\mathcal{Y}_{t,x}$ is the set of arguments for which $\mathcal{R}_{t,x}$ takes finite values. Taking the intersection over all $x \in D$ leads to the following definition.
Definition 1.5. Given an affine process $X$ and the associated function $R_{t,x}$ as in Def. 1.4, define for each $t \geq 0$ the set

$$\mathcal{Y}_t = \bigcap_{x \in D} \left\{ y \in \mathbb{R}^d : \int_{\|\xi\| \geq 1} e^{\langle y, \xi \rangle} K_t(x, d\xi) < \infty \right\}.$$  

As an intersection of convex sets, also $\mathcal{Y}_t$ is convex. Moreover, $\mathcal{Y}_t$ contains $0$ and hence is non-empty, because $R_{t,x}(0) = 0$, for all $t \geq 0, x \in D$.

Since the functions $a_1(x), b_1(x)$ and $K_1(x, d\xi)$ are affine in $x$, we can decompose $R_{t,x}$ into $R_{t,x}(y) = F_t(y) + (R_t(y), x)$. For arguments $y \in \mathcal{Y}_t$, the functions $F_t$ and $R_t$ are uniquely specified, since $D$ contains $0$ and $d$ linearly independent points.

Proposition 1.6. Let $X$ be an affine process with state space $D$. Then, for each $t \geq 0$, there exist functions $F_t : \mathcal{Y}_t \rightarrow \mathbb{R}, R_t : \mathcal{Y}_t \rightarrow \mathbb{R}^d$ such that

$$R_{t,x}(y) = F_t(y) + (R_t(y), x)$$

for all $x \in D, y \in \mathcal{Y}_t$. Let $(e_1, \ldots, e_d)$ be a set of $d$ linearly independent vectors in $\mathbb{R}^d$. Then we can write $F_t$ and $R_t(y) := (R_t(y), e_i)$ as

\begin{align*}
(1.8a) \quad F_t(y) &= \frac{1}{2} \left( \langle u, a_t y \rangle + \langle b_t, y \rangle + \int_{\mathbb{R}^d \setminus \{0\}} \left( e^{\langle \xi, y \rangle} - 1 - \langle h(\xi), y \rangle \right) m_t(d\xi) \right), \\
(1.8b) \quad R_t(y) &= \frac{1}{2} \left( \langle y, a_t \rangle + \langle \beta_t, y \rangle + \int_{\mathbb{R}^d \setminus \{0\}} \left( e^{\langle \xi, y \rangle} - 1 - \langle h(\xi), y \rangle \right) \mu_t(d\xi) \right).
\end{align*}

Proof. This follows immediately from Definition 1.4 and Assumption 1.3.

Remark 1.7. Setting $x = 0$ in (1.6) yields that $F_t(y)$ is a convex and lower semi-continuous function of Lévy-Khintchine form, for each $t \geq 0$. The same is not necessarily true for $R^1_t, \ldots, R^d_t$, since the matrices $a_t$ may not be positive semidefinite or the measures $\mu_t$ may be signed measures.

The functions $F_t(y)$ and $R_t(y)$ define a system of ODEs, called Riccati equations.

Definition 1.8 (Riccati system). Let $X$ be an affine process and $F_t, R_t$, and $\mathcal{Y}_t$ be defined as in Definition 1.5 and Proposition 1.6. Let $T \geq S \geq 0, y \in \mathcal{Y}_T$, and let

$$p : t \mapsto p(t, T, y), \quad q : t \mapsto q(t, T, y)$$

be continuous functions mapping $[S, T]$ to $\mathbb{R}$ resp. $\mathbb{R}^d$ that satisfy $q(t, T, y) \in \mathcal{Y}_t$ and

\begin{align*}
(1.9a) \quad p(t, T, y) - p(S, T, y) &= - \int_S^t F_s(q(s, T, y)) ds, \quad p(T, T, y) = 0, \\
(1.9b) \quad q(t, T, y) - q(S, T, y) &= - \int_S^t R_s(q(s, T, y)) ds, \quad q(T, T, y) = y
\end{align*}

for all $t \in [S, T]$. Then we call $(p, q)$ a solution (on the time interval $[S, T]$ and with end point $y$) of the Riccati system associated to $X$.

It is important to note that in general the function $R_t$ is locally Lipschitz continuous only on the interior of $\mathcal{Y}_t$, but may fail to be Lipschitz continuous at the boundary of $\mathcal{Y}_t$. Hence solutions of (1.9) reaching or starting at the boundary of $\mathcal{Y}_t$ may not be unique. For this reason we add the following definition.
Definition 1.9 (Minimal solution). Let $X$ be an affine process, and let $(p, q)$ a solution on $[S, T]$ with end point $y \in \mathcal{Y}_T$ to the associated Riccati system. We call $(p, q)$ a minimal solution, if for any other solution $(\tilde{p}, \tilde{q})$ on $[\tilde{S}, T] \subseteq [S, T]$ with the same end point $q(T, T, y) = \tilde{q}(T, T, y) = y$ it holds that
\begin{equation}
(1.10) \quad p(t, T, y) + \langle q(t, T, y), x \rangle \leq \tilde{p}(t, T, y) + \langle \tilde{q}(t, T, y), x \rangle \quad \forall x \in D, \quad \forall t \in [S, T].
\end{equation}
for all $t \in [\tilde{S}, T]$ and $x \in D$.

Remark 1.10. By setting $q_x(t, T, y) := p(t, T, y) + \langle q(t, T, y), x \rangle$, the Riccati system may be written in condensed form as
\begin{equation}
(1.11) \quad -\frac{\partial_t q_x(t, T, y)}{\partial t} = \mathcal{R}_{t,x}(q(t, T, y)), \quad q_x(T, T, y) = y, \quad \forall x \in D.
\end{equation}
In this notation the minimality property can we written as
\begin{equation}
q_x(t, T, y) \leq \tilde{q}_x(t, T, y), \quad \forall x \in D, \quad t \in [S, T].
\end{equation}

Remark 1.11. The following properties are easy to see: If for a given end value $y \in \mathcal{Y}_T$ there is only one solution to the Riccati system, then it is automatically a minimal solution. Also, if for a given end value a minimal solution $(p, q)$ exists on $[S, T]$, it is automatically the unique minimal solution. Indeed, if there were another minimal solution $(\tilde{p}, \tilde{q})$, then
\begin{equation}
(1.12) \quad p(t, T, y) + \langle q(t, T, y), x \rangle = \tilde{p}(t, T, y) + \langle \tilde{q}(t, T, y), x \rangle \quad \forall x \in D, \quad t \in [S, T].
\end{equation}
for all $t \in [S, T]$, $x \in D$. Since $D$ contains $d$ linearly independent points and 0 it follows that $p = \tilde{p}$ and $q = \tilde{q}$ in this case.

We can now formulate our main results on real exponential moments of affine processes.

Theorem 1.12 (Real moments of affine processes). Let $X$ be an affine process on $D$ satisfying Assumption 1.3 and let $T \geq 0$.

(a) Let $y \in \mathbb{R}^d$ and suppose that $E^{S,x}[e^{(y,X_T)}] < \infty$ for some $t \geq 0$ and $x \in D^2$.
Then there exist a unique minimal solution $(p, q)$ on $[S, T]$ of the Riccati system (1.9), such that
\begin{equation}
(1.12) \quad E^{S,x}[e^{(y,X_T)}] = e^{p(t, T, y) + \langle q(t, T, y), x \rangle}, \quad \forall x \in D, \quad t \in [S, T].
\end{equation}

(b) Suppose that the Riccati system (1.9) has solutions $(\tilde{p}, \tilde{q})$ on $[S, T]$ with end point $y$. Then $E^{S,x}[e^{(y,X_T)}] < \infty$ and there exist unique minimal solutions $(p, q)$ on $[S, T]$ of the Riccati system such that (1.12) holds for all $x \in D$, $t \in [S, T]$.

Remark 1.13. We emphasize that in point (b) of the theorem $p = \tilde{p}$ and $q = \tilde{q}$ does not necessarily hold, i.e. the candidate solutions $(\tilde{p}, \tilde{q})$ have to be replaced by the minimal solutions $(p, q)$ in order for (1.12) to hold true.

The following Corollary is a conditional version of Theorem 1.12:

Corollary 1.14. Suppose that the conditions of either Theorem 1.12a or 1.12b are satisfied, and let $(p, q)$ be the associated minimal solutions of the Riccati system (1.9). Then also $E^{S,x}[e^{\langle q(t, T, y), X_t \rangle}] < \infty$ and
\begin{equation}
E^{S,x}[e^{(y,X_T)}|\mathcal{F}_t] = e^{p(t, T, y) + \langle q(t, T, y), X_t \rangle}
\end{equation}
holds for all $x \in D$, $t \in [S, T]$. 

136
2. Proofs for real moments of affine processes

2.1. Decomposability and dependency on the starting value. Definition 1.2 of an affine process immediately implies a decomposability property of the probability measures \( \mathbb{P}^{s,x} \) (see also Duffie, Filipović, and Schachermayer [30, Thm 2.15]).

To describe this property, we identify each measure \( \mathbb{P}^{s,x} \), \((s,x) \in \mathbb{R}_{\geq 0} \times D \), with the law of \( X \), i.e., with a probability measure on the Skorokhod space \( \mathcal{D}(\mathbb{R}_{\geq 0}, \mathbb{R}^d) \) of càdlàg paths \( \omega : \mathbb{R}_{\geq 0} \to \mathbb{R}^d \). Thus, we assume \( \Omega = \mathcal{D}(\mathbb{R}_{\geq 0}, \mathbb{R}^d) \) and \( X_t(\omega) = \omega_s \). As in Duffie, Filipović, and Schachermayer [30, Def. 2.14], we write \( \mathbb{P} \times \mathbb{P}' \) for the image of \( \mathbb{P} \times \mathbb{P}' \) under the measurable mapping \( (\omega, \omega') \mapsto \omega + \omega' : (\Omega \times \Omega, \mathcal{F} \times \mathcal{F}) \to (\Omega, \mathcal{F}) \).

**Proposition 2.1.** Let \( X \) be an affine process with state space \( D \). Then the probability laws \( \mathbb{P}^{s,x} \) satisfy the following decomposability property: Suppose that \( x, \xi \) and \( x + \xi \) are in \( D \). Then

\[
\mathbb{P}^{t,x} * \mathbb{P}^{s,x+\xi} = \mathbb{P}^{t,0} * \mathbb{P}^{s,x+\xi}
\]

**Proof.** This can be shown as [56, Proposition 4.1]. \( \square \)

In the following, we set

\[
g(s, t, y, x) = \mathbb{E}^{s,x} \left[ e^{(y,X_t)} \right] = \int_D e^{(y,\xi)} p_{s,t}(x, d\xi),
\]

for all \( t \geq s \geq 0, y \in \mathbb{R}^d \), and \( x \in D \). Note that \( g(t, y, x) \) is always strictly positive, but might take the value \( +\infty \). By approximating \( g(t, y, x) \) monotonically from below by bounded functions and using the Chapman-Kolmogorov equation we derive that

\[
g(r, t, y, x) = \int_D g(s, t, y, \xi)p_{r,s}(x, d\xi)
\]

holds for all \( t \geq s \geq r \geq 0, y \in \mathbb{R}^d \), and \( x \in D \), where \( +\infty \) is allowed on both sides and in the integrand. The following Lemma concerns the role of the starting value \( X_0 = x \) of the affine process with regards to finiteness of exponential moments.

**Lemma 2.2.** Let \( X \) be an affine process on \( D \) and let \((T, y) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^d \). Then the following holds:

(a) \( \mathbb{E}^S \left[ e^{(y,X_T)} \right] = \infty \) implies \( \mathbb{E}^{S,x} \left[ e^{(y,X_T)} \right] = \infty \) for all \( x \in D^0 \).

(b) \( \mathbb{E}^{S,x} \left[ e^{(y,X_T)} \right] < \infty \) for some \( x \in D^0 \) implies \( \mathbb{E}^{S,x} \left[ e^{(y,X_T)} \right] < \infty \) for all \( x \in D \).

(c) \( \mathbb{E}^{S,x} \left[ e^{(y,X_T)} \right] < \infty \) for all \( x \in D \) implies \( \mathbb{E}^{t,x} \left[ e^{(y,X_T)} \right] < \infty \) for all \( t \in [S,T] \), \( x \in D \).

**Proof.** Statements (a) and (b) can be shown as in [56, Lemma 4.2]. To prove (c), let

\[
A = \{ t \in [S,T] : \exists x \in D \text{ s.t. } g(t, T, x) = \infty \}.
\]

We proceed by contraposition: assuming \( A \) is nonempty we show that \( S \in A \). Under this assumption, the infimum of \( A \) is finite and lies in the interval \([S,T] \). Fix an arbitrary point \( x \in D^0 \) and choose \( \epsilon > 0 \) such that \( B_\epsilon(x) \subseteq D^0 \). Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a continuous function, which is bounded from above by the indicator function of \( B_\epsilon(x) \) and satisfies \( f(x) = 1 \). By the weak continuity of the semigroup, the expression \( \int_D f(\xi)p_{s,t}(x, d\xi) \) is jointly continuous in \((s,t)\) and assumes the value 1.
on the diagonal $s = t$. Thus, there is $\delta > 0$ such that whenever $s, t$ are nonnegative and satisfy $\inf A - \delta \leq s \leq t \leq \inf A + \delta$, then
\begin{equation}
(2.3) \quad p_{s,t}(x, D^p) \geq p_{s,t}(x, B_0(x)) \geq \int_D f_s(\xi)p_{s,t}(x, d\xi) \geq \frac{1}{2}.
\end{equation}
Choose $t \in A$ satisfying $|t - \inf A| \leq \delta$, which is possible by the definition of the infimum. From part (b), the condition $t \in A$ can be strengthened to $g(t, T, u, x) = \infty$, for all $x \in D^p$. It follows by (2.2) and (2.3) that
\[
g(s, T, y, x) = \int_D g(t, T, y, \xi)p_{s,t}(x, d\xi) = \infty
\]
holds for each $s$ satisfying $0 \leq s \leq t$ and $|s - \inf A| \leq \delta$. This means $s \in A$. It follows that $\inf A = S$ and $S \in A$, which proves (c).

2.2. From moments to Riccati equations. Lemma 2.3 and Proposition 2.4 in this section show that the Riccati system has a solution if the corresponding exponential moment exists. The minimality property of the solution will be shown in Section 2.4.

Lemma 2.3. Let $X$ be an affine process on $D$, and let $T \geq S \geq 0$. Suppose that for some $x \in D^p$ and $y \in \mathbb{R}^d$ it holds that $\mathbb{E}^{S,x}[e^{\langle y, X_T \rangle}] < \infty$. Then $y \in \mathcal{Y}_T$ and the following holds:

(a) There exist functions $t \mapsto p(t, T, y) \in \mathbb{R}$ and $t \mapsto q(t, T, y) \in \mathbb{R}^d$ defined for $t \in [S, T]$ such that (1.12) holds.

(b) For each $t \in [S, T]$, $\mathbb{E}^{S,x}[e^{\langle q(t, T, y), X_t \rangle}] < \infty$ and
\[
(2.4) \quad \mathbb{E}^{S,x}\left[ e^{\langle y, X_T \rangle} \middle| \mathcal{F}_t \right] = e^{p(t, T, y) + \langle q(t, T, y), X_t \rangle} \quad \text{for all } x \in D.
\]

(c) The functions $p(t, T, y), q(t, T, y)$ satisfy the semi-flow equations
\begin{align}
(2.5a) & \quad p(S, T, y) = p(t, T, y) + p(S, t, q(t, T, y)), \quad p(T, T, y) = 0, \\
(2.5b) & \quad q(S, T, y) = q(S, t, q(t, T, y)), \quad q(T, T, y) = y,
\end{align}
for all $t \in [S, T]$.

Proof. This can be shown as in [56, Lemma 4.3].

Proposition 2.4. Let $X$ be an affine process on $D$ and let $T \geq S \geq 0$. Let $y \in \mathbb{R}^d$ and suppose that $\mathbb{E}^{S,x}[e^{\langle y, X_T \rangle}] < \infty$ for some $x \in D^p$. Then $y \in \mathcal{Y}_T$ and there exists a solution $(p, q)$ on $[S, T]$ of the Riccati system (1.9), such that (1.12) holds for all $x \in D$, $t \in [S, T]$.

Proof. As in Cuchiero [21] and Keller-Ressel and Mayerhofer [56], we enlarge the measurable space $(\Omega, \mathcal{F}, \mathbb{P})$ such that it supports $d + 1$ copies of the affine process $X$, which we denote by $X^0, \ldots, X^d$. Without loss of generality it can be assumed that $X = X^0$. In what follows, we will use the convention that upper indices correspond to the different instances of the process $X$, while lower indices correspond to the coordinate projections of a single process. Let $x = (x^0, \ldots, x^d)$ be $d + 1$ affinely independent points in $D$. For each $S \geq 0$, we denote by $\mathbb{P}^{S,x}$ a probability measure such that $X^0, \ldots, X^d$ are independent and $X^t$ has law $\mathbb{P}^{S,x_t}$, for each $i \in \{0, \ldots, d\}$.
For each $t \in [S,T]$, let
\[
\Xi(t) = \begin{pmatrix}
1 & X_0^0(t) & \cdots & X_0^d(t) \\
\vdots & \vdots & \ddots & \vdots \\
1 & X_d^0(t) & \cdots & X_d^d(t)
\end{pmatrix}.
\]
Assume that $\mathbb{P}^{S,x}(\forall t \in [S,T]: \det(\Xi(t)) \neq 0) > 0$. Then it follows by the same arguments as in [56, Proposition 4.5] that the functions $(p,q)$ given by Lemma 2.3 are solutions of the Riccati system (1.9). Equation (1.12) holds by construction.

In general, $\Xi(t)$ might not be invertible for all $t \in [S,T]$, with positive probability. However, since invertibility of the matrix $\Xi(t)$ is an open condition on the values of the affine process $(X^0, \ldots, X^d)$, Lemma A.2 provides some $\delta > 0$ such that
\[
\inf_{t \in [S,T]} \mathbb{P}^{S,x}(\forall s \in [t,t+\delta]: \det(\Xi(s)) > 0) \geq \frac{1}{2}
\]
Let $S_k = S \lor (T - k\delta)$, for each $k \in \mathbb{N}$. By Lemma 2.2 we know that
\[
\mathbb{E}^{S_k,x^t}[e^{\langle y,X_t^k \rangle}] < \infty
\]
and by Lemma 2.3,
\[
\mathbb{E}^{S_k,x^t}[e^{\langle y,X_t^k \rangle}|\mathcal{F}_t] = e^{p(t,T,y)+\langle q(t,T,y),X_t^k \rangle}
\]
for each $i \in \{0,\ldots,d\}, t \in [S_k,T]$. Since $|S_k - T| \leq \delta$, the arguments of [56, Proposition 4.5] show that $(p,q)$ is a solution of the Riccati system (1.9) on the interval $[S_k,T]$.

We now work our way backwards to $S_k \leq S_1$ by induction. Suppose that $(p,q)$ are solutions of the Riccati system on $[S_k,T]$. We show that they can be extended to solutions on $[S_{k+1},T]$. By Lemma 2.2, $\mathbb{E}^{S_{k+1},x}[e^{\langle y,X_T \rangle}] < \infty$ implies that $\mathbb{E}^{S_{k+1},x}[e^{\langle y,X_T \rangle}] < \infty$ and by Lemma 2.3, $\mathbb{E}^{S_{k+1},x}[\langle q(S_k,T,y),X_{S_k} \rangle] < \infty$. Set $y' = q(S_k,T,y)$, $S' = S_{k+1}$, and $T' = S_k$. Then—proceeding exactly as above—we obtain functions $t \mapsto p(t,T',y')$ and $t \mapsto q(t,T',y')$ defined on the interval $[S',T']$ and satisfying (1.9) with $(S,T,y)$ replaced by $(S',T',y')$. By the induction hypothesis, (1.9) already holds for all $t \in [T',T]$. By the flow property, this is equivalent to (1.9) holding on $[S',T]$. Thus, (1.9) holds on all intervals $[S_k,T], k \in \mathbb{N}$, and the proof is complete.

2.3. From Riccati equations to moments. If the Riccati system has a solution, then the corresponding exponential moment is finite, as the following proposition shows.

**Proposition 2.5.** Let $X$ be an affine process taking values in $D$. Let $y \in \mathcal{Y}_T$ and suppose that the Riccati system (1.9) has a solution $(\hat{p},\hat{q})$ that ends at $y$ and exists on the interval $[S,T]$ for some $S \leq T$. Then $\mathbb{E}^{S,y}[e^{\langle y,X_T \rangle}] < \infty$ and the affine transform formula (1.12) holds, where $(p,q)$ is also a solution on $[S,T]$ to (1.9).

**Proof.** This can be shown as in [56, Proposition 4.6].

2.4. Proof of Theorem 1.12. To establish Theorem 1.12 it remains to show the minimality of $(p,q)$ in (1.12).
Lemma 2.6. Let \((p, q)\) and \((\tilde{p}, \tilde{q})\) be given as in Proposition 2.5. Then for all \(t \in [S, T]\) and \(x \in D\),
\[
p(t, T, y) + \langle q(t, T, y), x \rangle \leq \tilde{p}(t, T, y) + \langle \tilde{q}(t, T, y), x \rangle
\]
Proof. We define
\[
M^p_t = e^{p(t, T, y) + \langle q(t, T, y), X_t \rangle}, \quad \tilde{M}^\tilde{p}_t = e^{\tilde{p}(t, T, y) + \langle \tilde{q}(t, T, y), X_t \rangle}.
\]
Then, for each \(x \in D\) the process \(M^p\) is a \(\mathbb{P}^{S,x}\)-martingale (see (2.6)); \(\tilde{M}^\tilde{p}\) is a \(\mathbb{P}^{S,x}\)-supermartingale, and they satisfy \(M^\tilde{p}_T = \tilde{M}^\tilde{p}_T\). Hence,
\[
M^p_t = \mathbb{E}^{S,x}[M^p_T | \mathcal{F}_t] = \mathbb{E}^{S,x}[\tilde{M}^\tilde{p}_T | \mathcal{F}_t] \leq \tilde{M}^\tilde{p}_t,
\]
for all \(t \in [S, T]\). Taking logarithms, the claimed inequality follows. \(\square\)

Proof of Theorem 1.12. This follows by the same arguments as in [56], which we repeat here for the sake of completeness. In view of Remark 1.11, part (a) is established if we can show that the solution \((p, q)\) of the Riccati system established in Proposition 2.4 is minimal. Let \((\tilde{p}, \tilde{q})\) be another solution on \([S', T]\) of the Riccati system, \(S' \geq S\). Then by Proposition 2.5 there exists \((p^*, q^*)\) such that (1.12) holds for all \(y \in D\) and \(t \in [S', T]\), as is the case for \((p, q)\). By taking logarithms of the respective right sides of (1.12) and by applying Lemma 2.6 we see that
\[
p + \langle q, y \rangle = p^* + \langle q^*, y \rangle \leq \tilde{p} + \langle \tilde{q}, x \rangle,
\]
on \([S', T]\) and for all \(y \in D\). Hence by Definition 1.9 \((p, q)\) is the minimal solution of the Riccati system, and we are done with part (a).

Part (b) follows immediately from Lemma 2.6, Definition 1.9 and Remark 1.11. \(\square\)

Appendix A. Auxiliary lemmas

Lemma A.1. Let Assumption 1.3 be in place. Then for each \(x \in D\) the expressions
\[
a_r(x), b_r(x), \int \left(\|\xi\|^2 \wedge 1\right) K_r(x, d\xi), \quad (r \geq 0)
\]
are locally integrable in \(r\).

Proof. The expressions in (A.1) are affine in \(x\). Therefore, it is sufficient to establish local integrability for one tuple of affinely independent points \(x^0, \ldots, x^d \in D^o\), which we fix once and for all. Since affine independence is an open condition, there exists for each \(i \in \{0, \ldots, d\}\) an open set \(U^i \subset D\) containing \(x^i\) such that \(U^0 \times \cdots \times U^d\) consists of affinely independent points, only. For each \(i \in \{0, \ldots, d\}\), there is a continuous, non-negative function \(f^i\) on \(D\) which is dominated by the indicator function of \(U^i\) and satisfies \(f^i(x^i) = 1\). This function allows us to exploit the weak continuity of the transition probabilities \(p_{s,t}(x^i, \cdot)\): indeed, when \(s < t\) is close enough to \(t\), for some fixed \(t > 0\), then
\[
\mathbb{P}^{s,x^i}(X_t \in U^i) = p_{s,t}(x^i, U^i) \geq \int_D f(\xi)p_{s,t}(x^i, d\xi) > 0
\]
holds simultaneously for all \(i \in \{0, \ldots, d\}\). Since \(\mathbb{P}^{s,x^i}\) admits (1.4) as semimartingale characteristics, the functions
\[
a_r(X_{r-}), b_r(X_{r-}), \int \left(\|\xi\|^2 \wedge 1\right) K_r(X_{r-}, d\xi), \quad r \geq s,
\]
are locally integrable in \(r\).
are locally integrable in $r$, $\mathbb{P}^s,x^i$-almost surely. Moreover, $\Delta X_t = 0$ holds, $\mathbb{P}^s,x^i$-almost surely, because the random measure $K_t(X_{r-}, d\xi)$ assigns weight zero to $\mathbb{R}^d \times \{r\}$, for each $r \geq 0$. Taken together, this means that $\mathbb{P}^s,x^i$ assigns positive probability to the event that the functions in (A.2) are locally integrable and that $X_t = X_{t-} \in U^i$. In particular, there exists at least one càdlàg path $\omega^i$ of $X$ with these properties, for each $i \in \{0, \ldots, d\}$.

By the continuity of the path $\omega^i$ at time $t$, there exist $\delta > 0$ such that $\omega^i(r) \in U^i$, for each $r \in [t - \delta, t + \delta]$ and $i \in \{0, \ldots, d\}$. We will now show that for each $i \in \{0, \ldots, d\}$, the function $r \mapsto a_r(x^i)$ is integrable on the interval $[t - \delta, t + \delta]$. Integrability of the remaining functions in (A.1) can be shown by the same arguments. For each $r \in [t - \delta, t + \delta]$, we can write

$$
(A.3) \quad \left( \begin{array}{c} a_r(x^0) \\ \vdots \\ a_r(x^d) \end{array} \right) = \left( \begin{array}{ccc} \omega_0^0(r) & \cdots & \omega_0^d(r) \\ \vdots & \ddots & \vdots \\ \omega_d^0(r) & \cdots & \omega_d^d(r) \end{array} \right)^{-1} \left( \begin{array}{c} a_r(\omega^0(r)) \\ \vdots \\ a_r(\omega^d(r)) \end{array} \right),
$$

because $a_r(x)$ is affine in $x$ and $\omega^0(r), \ldots, \omega^d(r)$ are affinely independent. Integrability of the functions on the left-hand side of (A.3) follows from boundedness of the coefficients of the matrix and integrability of the coefficients of the vector on the right-hand side of (A.3). Thus, $r \mapsto a_r(x^i)$ and, by the same argument, all functions in (A.1), are integrable on $[t - \delta, t + \delta]$, for each $i \in \{0, \ldots, d\}$. Since $t \geq 0$ was chosen arbitrarily, this completes the proof. 

\[ \square \]

**Lemma A.2.** Let $X$ be subject to Assumption 1.3. Let $U \subseteq D$ and $x \in U^\circ$. Then, for each $T > 0$ and $\epsilon > 0$, there exists $\delta > 0$ such that

$$
\inf_{t \in [0,T]} \mathbb{P}^t,x(\forall s \in [t, t + \delta] : X_s \in U) \geq 1 - \epsilon.
$$

**Proof.** Let $(t, x) \in \mathbb{R}_{\geq 0} \times D$ be fixed but arbitrary. Choose $R \in (0, 1/2]$ small enough such that $B_R(x) \subseteq U$. We define the stopping time $\tau = \inf \{s \in [t, \infty) : \|X_s\| > R\}$. Letting $X^\tau$ ($X^\tau_\tau$) denote the process $X$ stopped at (before) $\tau$, we have

$$
\mathbb{P}^t,x \left( \sup_{s \in [t, t + \delta]} \|X_s - x\| < R \right)
\geq \mathbb{P}^t,x \left( \sup_{s \in [t, t + \delta]} \|X^\tau_\tau - x\| < R/2 \text{ and } \forall s \in [t, t + \delta] : \|\Delta X^\tau_\tau\| < R/2 \right)
\geq 1 - \mathbb{P}^t,x \left( \sup_{s \in [t, t + \delta]} \|X^\tau_\tau - x\| \geq R/2 \right) - \mathbb{P}^t,x \left( \exists s \in [t, t + \delta] : \|\Delta X^\tau_\tau\| \geq R/2 \right).
$$

We will show that by choosing $\delta$ small enough, the probabilities $p_1$ and $p_2$ can be made arbitrarily small, locally uniformly in $t$.

We start by estimating the first probability. By Chebyshev’s inequality,

$$
p_1 \leq \frac{4}{R^2} \mathbb{E}^{t,x} \left[ \left( \sup_{s \in [t, t + \delta]} \|X^\tau_\tau - x\| \right)^2 \right].
$$

Recall that $X$ admits the $\mathbb{P}^t,x$-semimartingale characteristics $A, B, \nu$ given by (1.4). Let $J(dt, d\xi)$ denote the Poisson random measure associated to the jumps of $X$ with
predictable compensator $\nu(dt, d\xi)$. Then $X$ can be decomposed as $X = x + M + G$ into a $\mathcal{P}^{t,x}$-local martingale $M = X^e + h(\xi) \ast (J - \nu)$ and a process of finite variation $G = (\xi - h(\xi)) \ast J + B$. By the arithmetic-geometric inequality,

$$p_1 \leq \frac{8}{R^2} \mathbb{E}^{t,x} \left[ \left( \sup_{s \in [t,t+\delta]} \|M^\tau_s\| \right)^2 + \left( \sup_{s \in [t,t+\delta]} \|G^\tau_s\| \right)^2 \right].$$

By the Burkholder-Davis-Gundy inequality,

$$p_1 \leq \frac{32}{R^2} \mathbb{E}^{t,x} \left[ \text{tr} \left( ([M,M])_{t+\delta}^\tau \right) + \frac{8}{R^2} \mathbb{E}^{t,x} \left[ (\text{Var}(G))_{t+\delta}^\tau \right]^2 \right],$$

where $\text{tr}([M,M])$ denotes the trace of the matrix-valued quadratic covariation process and $\text{Var}(G)$ is the variance process of $G$. For each $s < \tau$, $\|\Delta X_s\|$ is bounded by $2R \leq 1$, which implies $\Delta X_s = -h(\Delta X_s) = 0$. Therefore,

$$\text{Var}(G)_{t+\delta}^\tau = (\|\xi - h(\xi)\| \ast J)_{t+\delta}^\tau + \text{Var}(B)_{t+\delta}^\tau = \text{Var}(B)_{t+\delta}^\tau.$$

Thus,

$$p_1 \leq \frac{32}{R^2} \mathbb{E}^{t,x} \left[ \text{tr} \left( (X^e, X^e)^{t+\delta} \right) + (\|h(\xi)\|^2 \ast J)_{t+\delta}^\tau + (\text{Var}(B))_{t+\delta}^\tau \right].$$

The interval $[0,t]$ can be neglected because $X$ is constant and $M = G = 0$, there. Consequently, we obtain

$$p_1 \leq \frac{32}{R^2} \mathbb{E}^{t,x} \left[ \int_t^{t+\delta} \left( \text{tr} (a_s(X_{s-}) + \int_{\mathbb{R}^d} \|h(\xi)\|^2 K_s(X_{s-}, d\xi) + \|b_s(X_{s-})\| ) \right) ds \right].$$

We now plug in the definition of $a_s(x), b_s(x), K_s(x, d\xi)$ as affine functions in $x$ and account for the fact that $\|X_{s-}\| \leq R \leq 1$ holds for each $s \leq \tau$. Moreover, we can estimate $\|h(\xi)\|^2$ by $\|\xi\|^2 \wedge 1$ and obtain

(A.4)  $$p_1 \leq \frac{32}{R^2} \int_t^{t+\delta} \left( \text{tr} (a_s + \alpha_s^1 + \cdots + \alpha_s^d) + \|b_s\| + \|\beta_s^1\| + \cdots + \|\beta_s^d\| \right)
$$

$$+ \int_{\mathbb{R}^d} \left( \|\xi\|^2 \wedge 1 \right) \left( m_s + \mu_s^1 + \cdots + \mu_s^d \right) (d\xi) ds.$$

We now estimate the second probability $p_2$, defined as the probability that $X^\tau$ has a jump of size at least $R/2$ within the interval $[t,t+\delta]$. Note that the integer-valued process $\mathbb{1}_{\|\xi\| \geq R/2} \ast J$ counts the number of jumps of size at least $R/2$. By Markov’s inequality

$$p_2 = \mathbb{P}^{t,x} \left( \left( \mathbb{1}_{\|\xi\| \geq R/2} \ast J \right)_{t+\delta}^\tau \geq 1 \right) \leq \mathbb{E}^{t,x} \left[ \mathbb{1}_{\|\xi\| \geq R/2} \ast J \right]_{t+\delta}^\tau.$$

Since the indicator function is non-negative, $J$ can be replaced by its compensator. Moreover, the indicator function can be estimated by $4/R^2(\|\xi\|^2 \wedge 1)$ and one obtains

$$p_2 \leq \frac{4}{R^2} \mathbb{E}^{t,x} \left[ \int_t^{t+\delta} \int_{\mathbb{R}^d} \left( \|\xi\|^2 \wedge 1 \right) K_s(X_{s-}, d\xi) ds \right].$$

Plugging in the definition of $K_s(x, d\xi)$ and using the fact that $\|X_{s-}\| \leq R \leq 1$ holds for each $s \leq \tau$, one obtains

(A.5)  $$p_2 \leq \frac{4}{R^2} \mathbb{E}^{t,x} \left[ \int_t^{t+\delta} \int_{\mathbb{R}^d} \left( \|\xi\|^2 \wedge 1 \right) \left( m_s + \mu_s^1 + \cdots + \mu_s^d \right) (d\xi) ds \right].$$
By Lemma A.1, the integrals on the right-hand side of (A.4) and (A.5) are finite and depend continuously on $t$ and $\delta$. Thus, the assertion of the lemma holds true locally in $t$. A compactness argument concludes the proof.

Department of Mathematics, ETH Zürich
E-mail address: philipp.harms@math.ethz.ch
E-mail address: david.stefanovits@math.ethz.ch
E-mail address: josef.teichmann@math.ethz.ch
David Stefanovits, Mario V. Wüthrich.
Hedging of long term zero-coupon bonds in a market model with reinvestment risk.
Hedging of long term zero-coupon bonds in a market model with reinvestment risk

David Stefanovits* and Mario V. Wüthrich*

Abstract
We present a computational methodology to value and hedge long term zero-coupon bonds trading in short and medium term ones. For this purpose we develop a discrete time stochastic yield curve model with limited availability of maturity dates at a fixed time point and newly issued bonds at future time points. This involves reinvestment risk and there is no perfect hedging strategy available for long term liabilities. We calibrate the model to market data and describe optimal hedging strategies under a given risk tolerance. These considerations provide a natural extrapolation of the yield curve beyond the last liquid maturity date, and a framework which allows to value long term insurance liabilities, for instance, under Solvency 2. Moreover, we determine the optimal trading strategy replicating the liabilities under the given risk tolerance.

1 Introduction
This paper is motivated by the problem of valuing and hedging long term insurance liabilities. Solvency directives require to value liabilities in a market-consistent way. That is, insurance cash-flows should be optimally replicated by an appropriate investment strategy of liquid financial instruments. This replication might prove difficult when considering long term liability cash-flows such as life insurance contracts or pension liabilities. This is because the value of these cash-flows may depend on factors, such as mortality rates, which cannot be traded in a liquid market. Therefore the insurer is facing underwriting risk. Moreover, the value of these cash-flows may also depend on yields beyond the maturities available in the liquid bond market. In that case, securities with shorter times to maturity need to be rolled over and the insurer is also subject to reinvestment risk. Because of the absence of liquid markets for these risks, the insurer cannot hedge the liabilities completely, or complete hedging may be highly capital inefficient. A possible approach is to value long term contracts as the minimal capital which must be invested in liquid financial instruments to cover the contract at an acceptable level of risk, see Hilli et al. [47]. In the literature this type of hedging technique is known as acceptable hedging. A mathematical presentation can be found in Artzner et al. [2].
In this paper we apply this valuation approach to a specific sub-problem of the one outlined above. We aim to value long term deterministic cash-flows with short and medium
term zero-coupon bonds. This means that we only consider reinvestment risk. Our valuation methodology is not based on simple yield curve extrapolation using families of parametrized curves, such as the Svensson family (see e.g. [63]), but rather on replication with actual investment strategies. We also assume that we can trade short and medium term zero-coupon bonds in a liquid market. In practice, this is not entirely the case, since most of the medium term securities in the liquid government and corporate bond market are coupon bearing. Because coupon bonds are simply linear combinations of zero-coupon bonds with different maturities, our methodology can be modified for that case. However, the valuation algorithm becomes more complex from the computational point of view. The first step of our work is to model the yield curve process appropriately. Classical stochastic term structure models, which assume that bonds with unlimited times to maturity are traded at each point in time, do not provide the right framework for our problem and do not seem appropriate for real world solvency modeling. In our set-up we can only trade bonds up to a certain time to maturity, and at each point in time a new bond is issued into the market with maturity date unavailable at the previous time points. This involves roll over and reinvestment risk in a natural way. To our knowledge there are only a few papers in the literature dealing with the problem of modeling bonds up to a finite time to maturity with newly issued bonds as time passes. See Sommer [74], and Dahl [26, 25]. In Sommer [74] a continuous time model with new bonds being issued continuously in time is considered, whereas in the continuous time setting of Dahl [25] new bonds are issued at a fixed set of time points. In Dahl [26] reinvestment risk is introduced in a discrete time non-recombining binomial model.

We work in a discrete time set-up and present a Heath-Jarrow-Morton type model with time-dependent volatility structure based on ideas developed in Ortega et al. [68], and Teichmann and Wüthrich [77]. In particular, we modify the model presented in [77] introducing the restriction of finitely many available times to maturity, and modeling newly issued zero-coupon bonds using an additional stochastic structure. In this way we obtain a market model with reinvestment risk. Except for the market price of risk, the calibration of the model to market data is straightforward. The model has two essential features for the problem of valuing long term liabilities: its calibration is consistent over a long period of time and it does not require nested simulations to generate yield curve scenarios. The model is presented in Section 2 and the calibration procedure in Section 5. The rest of the paper is organized as follows. In Section 3 we introduce trading strategies in this market model and prove formally that deterministic long term cash-flows are unattainable. Then, we define the concept of acceptable hedging in the context of our model and formulate the corresponding dynamic stochastic optimization problem. In Section 4.1 we apply the Galerkin approximation technique which allows to efficiently solve the optimization problem numerically, see e.g. Koivu and Pennanen [57]. To apply this technique, classes of parametric trading rules need to be considered. In Section 4.2 we consider four classes and derive their properties. Finally, in Section 6 we apply the methodology to zero-coupon bond price time series derived from Swiss government bonds.
2 Bond market model with reinvestment risk

2.1 Bond market model with finitely many instruments

We propose a discrete-time bond market model. Choose a fixed grid size \( \delta > 0 \) and \( d, n \in \mathbb{N} \) with \( d < n \). Let \( J = \{0, \delta, \ldots, n\delta\} \) be the set of discrete time points with final time horizon \( T = n\delta \). Let \( D = \{1, \ldots, d\} \) and assume that at each time point \( t \in J \) zero-coupon bonds with times to maturity \( j\delta \) for \( j \in D \) are available at the financial market. Define the sets \( J' = J \setminus \{0\} = \{\delta, \ldots, n\delta\} \) and \( D^- = D \setminus \{d\} = \{1, \ldots, d-1\} \). Our aim is to model the zero-coupon bond prices stochastically at each \( t \in J' \) taking into account that the zero-coupon bonds with maturities \( t + j\delta \) for \( j \in D^- \) are available at the market at time \( t - \delta \), whereas the one with maturity \( t + d\delta \) is not.

Let \( (\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}) \) be a filtered probability space with real-world probability measure \( \mathbb{P} \) and filtration \( \mathbb{F} = (\mathcal{F}_t)_{t \in J'} \). We aim to model the \( d \)-dimensional \( \mathbb{P}^\ast \)-adapted yield curve process

\[
\left( Y_t \right)_{t \in J} = \left( (Y_{t,1}, \ldots, Y_{t,d}) \right)_{t \in J}.
\]

Note that for \( t \in J \) close to \( T \) the yield curve \( Y_t \) refers to securities which may expire after the final time horizon \( T \). The zero-coupon bond price process is defined by

\[
(P_t)_{t \in J} = \left( (P_{t,1}, \ldots, P_{t,d}) \right)_{t \in J},
\]

where \( P_{t,j} = \exp(-j\delta Y_{t,j}) \) for \( j \in D \). This price process is by assumption \( \mathbb{F} \)-adapted and observable at the market. Let \( (\varepsilon_t)_{t \in J'} = \left( (\varepsilon_{t,1}, \ldots, \varepsilon_{t,d}) \right)_{t \in J'} \) be \( \mathbb{F} \)-adapted, \( d \)-dimensional independent standard Gaussian innovations under \( \mathbb{P} \), i.e. \( \varepsilon_t \) is \( \mathcal{F}_t \)-measurable, independent of \( \mathcal{F}_{t-\delta} \) and \( \varepsilon_t \sim \mathcal{N}(0, I_d) \) under \( \mathbb{P} \) for all \( t \in J' \).

Define a probability measure \( \mathbb{P}^\ast \) equivalent to \( \mathbb{P} \) via the density process \( (\xi_t)_{t \in J'} \) given by

\[
\xi_t = \prod_{s=1}^{t/\delta} \exp \left(-\frac{1}{2} ||\mu_{s,\delta}||^2 + \mu'_{s,\delta} \varepsilon_{s,\delta} \right),
\]

where \( (\mu_t)_{t \in J'} \) is a \( d \)-dimensional previsible process called the market price of risk process. Let \( (\varepsilon_t)_{t \in J'} \) be \( \mathbb{F} \)-adapted, \( d \)-dimensional independent standard Gaussian innovations under \( \mathbb{P}^\ast \). For \( t \in J' \) we have under \( \mathbb{P}^\ast \)

\[
\varepsilon_t - \mu_t \sim \mathcal{N}(0, I_d), \quad \text{given } \mathcal{F}_{t-\delta}.
\]  

(1)

For more details concerning \( \mathbb{P}^\ast \) we refer to Wüthrich and Merz [82], Sections 2.3 and 2.4.

The bank account process is defined by \( B_0 = 1 \) and for \( t \in J' \)

\[
B_t = \exp \left( \sum_{s=1}^{t/\delta} \delta Y_{(s-1)\delta,1} \right) = \prod_{s=1}^{t/\delta} P_{(s-1)\delta,1}^{-1}.
\]

This process is previsible and describes the one period risk-free roll over, also called locally riskless investment (see Example 5.5 in Föllmer and Schied [36]). In order to obtain a model which is free of arbitrage we require under \( \mathbb{P}^\ast \) for all \( t \in J' \) and \( j \in D^- \)

\[
B_{t-\delta}^{-1} P_{t-\delta,j+1} = \mathbb{E}^\ast[B_{t}^{-1} P_{t,j} | \mathcal{F}_{t-\delta}].
\]  

(2)
The existence of an equivalent probability measure \( P^* \sim P \) such that the price process fulfills (2) rules out arbitrage in the resulting pricing model, we refer to Delbaen and Schachermayer [28] for the general version of the no-arbitrage condition. In our setting condition (2) is formulated for \( j \in D^- \) because the bond with time to maturity \( d\delta \) is not available at previous time points. As we discuss below, the extension of the model to bonds that are only available at later time points immediately leads to an incomplete market model with infinitely many equivalent measures which satisfy (2). Our aim is to price these bonds using dynamic hedging arguments with instruments available at the market.

### 2.2 Stochastic yield curve modeling

First, we formulate our model under \( P^* \) considering only maturities in \( D^- \). We construct a semi-parametric model based on ideas presented in Teichmann and Wüthrich [77]. Choose \( t \in J' \). For given yield curve \( Y_{t-d} \) we make the following model assumption for \( j \in D^- \)

\[
j\delta Y_{t,j} = (j + 1)\delta Y_{t-\delta,j+1} - \delta Y_{t-\delta,1} + \alpha_j(Y_{t-\delta,j+1} + \sqrt{\delta} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1}) \lambda_{ij} \varepsilon_{t,i}^*),
\]

where \( h : \mathbb{R} \rightarrow \mathbb{R} \) is a real function which describes how yield volatilities scale with respect to yield levels. Set \( \Lambda^- = (\Lambda^-_{ij})_{i,j \in D^-} \in \mathbb{R}^{(d-1) \times (d-1)} \).

Our explicit choice of \( h \) is discussed later on. The function \( h \) is called volatility scaling factor and \( \Lambda^- \) is called matrix of return directions. If we drop the last two terms on the right-hand side of (3), we obtain the no-arbitrage condition in case of a deterministic yield curve. The last term specifies uncertainty in the future yield curve. The drift term \( \alpha_j \) is determined by the no-arbitrage condition (2). From [77], Lemma 3.1, we have for all \( t \in J' \) and \( j \in D^- \)

\[
\alpha_j(Y_{t-\delta,j+1}) = \frac{\delta}{2} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1})^2 \lambda_{ij}^2.
\]

We introduce some notation to simplify (3) and (4). Let \( \varepsilon^-_t = (\varepsilon^-_{t,1}, \ldots, \varepsilon^-_{t,d-1})' \) and \( \mu^-_t = (\mu^-_{t,1}, \ldots, \mu^-_{t,d-1})' \). Define the functions \( \zeta^- : \mathbb{R}^d \rightarrow \mathbb{R}^{(d-1) \times (d-1)} \) and \( \Sigma^- : \mathbb{R}^d \rightarrow \mathbb{R}^{(d-1) \times (d-1)} \) by

\[
y \mapsto \zeta^- (y) = \text{diag}(h(y_2), \ldots, h(y_d)), \quad \text{and} \quad y \mapsto \Sigma^- (y) = \zeta^- (y) \Lambda^- \zeta^- (y).
\]

We consider the process \( (\tilde{Y}^-_t)_{t \in J'} = ((\tilde{Y}^-_{t,1}, \ldots, \tilde{Y}^-_{t,d-1}))_{t \in J'} \) given by

\[
\tilde{Y}^-_{t,j} = j\delta Y_{t,j} - (j + 1)\delta Y_{t-\delta,j+1} = -\log \frac{P_{t,j}}{P_{t-\delta,j+1}}.
\]

These describe the one period log-returns on the bonds. Under the real-world measure \( \mathbb{P} \) we obtain the following representation (see Section A.1 in the appendix).
Lemma 2.1. Let $t \in J'$. Under (1), model equation (3) and condition (4) can be formulated as
\[ \tilde{Y}_t^- = \delta \left[ -Y_{t-\delta,1} 1^- + \frac{1}{2} \text{sp}(\Sigma(Y_{t-\delta})) \right] + \sqrt{\delta} \zeta(Y_{t-\delta}) \Lambda^- (\epsilon_t - \mu_t), \] (5)
where $1^- = (1, \ldots, 1') \in \mathbb{R}^{d-1}$ and $\text{sp}(\Sigma(-)) = (\Sigma(-)_{11}, \ldots, \Sigma(-)_{d-1,d-1})'$. Thus, $\tilde{Y}_t ^-$ is multivariate Gaussian distributed under $P$, given $F_{t-\delta}$, with conditional mean
\[ \mathbb{E}[\tilde{Y}_t^- | F_{t-\delta}] = \delta \left[ -Y_{t-\delta,1} 1^- + \frac{1}{2} \text{sp}(\Sigma(Y_{t-\delta})) \right] - \sqrt{\delta} \zeta(Y_{t-\delta}) \Lambda^- \mu_t, \]
and conditional covariance matrix
\[ \text{Cov}[\tilde{Y}_t^- | F_{t-\delta}] = \delta \Sigma^-(Y_{t-\delta}). \]

The lemma provides a vectorial representation of (3) and (4), and characterizes the multivariate distribution of the one period log-returns on the bonds.

For $t \in J'$ the bond which matures at time $t + d\delta$ is not traded at time $t - \delta$ and therefore the choice of the model for this security is not restricted by the no-arbitrage condition (2) under $P^*$. For this reason we formulate a stochastic model for the process $(Y_t, t \in J)$ directly under $P$. We propose to couple equation (3) with the following stochastic representation for $t \in J'$ and a given yield curve $Y_{t-\delta}$
\[ d\delta Y_{t,d} = (d-1)\delta Y_{t,d-1} + \delta \beta(Y_{t-\delta}) + \sqrt{\delta} h(Y_{t-\delta}, \kappa_{t,d}), \] (6)
where $\kappa \in \mathbb{R}$ is a constant and $\beta : \mathbb{R}^d \rightarrow \mathbb{R}$ is a real function which describes the slope at the long end of the yield curve. We call this function long end slope factor and choose it explicitly later on. The first two terms in (6) describe a linear continuation of the yield curve at the long end. The last term adds a stochastic part. This completes the formulation of our model assumptions.

To condense the model assumptions in one equation we introduce additional notation. Define the functions $\zeta : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $\Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by
\[ y \mapsto \zeta(y) = \text{diag}(h(y_2), \ldots, h(y_d), h(y_d)), \quad \text{and} \quad y \mapsto \Sigma(y) = \zeta(y) \Lambda \Lambda' \zeta(y), \] (7)
where
\[ \Lambda = \begin{bmatrix} \Lambda^- & 0 \\ \vdots & \vdots \\ \Lambda_{d-1,1} & \cdots & \Lambda_{d-1,d-1} & \kappa \end{bmatrix} \in \mathbb{R}^{d \times d}. \]

We also define the transformed yield curve process $\tilde{Y}_t = (\tilde{Y}_t^- , d\delta Y_{t,d} - d\delta Y_{t-\delta,d})'$ and $\tilde{\mu}_t = (\mu_t', 0)'$. We formulate our bond market model under the real-world measure $P$ as follows (see Section A.2 in the appendix).

Lemma 2.2. Let $t \in J'$. Under (1), model equations (3), (6) and condition (4) can be formulated as
\[ \tilde{Y}_t = \delta \left[ -Y_{t-\delta,1} 1 + \frac{1}{2} \text{sp}(\Sigma(Y_{t-\delta})) + \left( \beta(Y_{t-\delta}) - \frac{1}{2} \zeta(Y_{t-\delta}) \Lambda \right) \right] _{d,d} + \sqrt{\delta} \zeta(Y_{t-\delta}) \Lambda (\epsilon_t - \tilde{\mu}_t), \] (8)
where \( \mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^d \) and \( \mathbf{e}_d = (0, \ldots, 0, 1)' \in \mathbb{R}^d \). Thus, \( \tilde{\mathbf{Y}}_t \) is multivariate Gaussian distributed under \( \mathbb{P} \), given \( \mathcal{F}_{t-\delta} \), having initial value \( V_0(X) = 0 \), terminal value \( VT(X) \geq 0 \), \( \mathbb{P} \)-a.s., and \( \mathbb{P}[VT(X) > 0] > 0 \). Such trading strategies are distributed under \( \mathbb{P} \).

Note that only the conditional distribution given the information in the previous period is Gaussian. For general \( h \), conditioning over multiple periods leads to more complicated distributions and there are no closed form expressions for the moments available.

### 3 Trading strategies and hedging

A trading strategy is a \( d \)-dimensional previsible stochastic process

\[
\mathbf{X} = (X_t)_{t \in \mathcal{J}} = ((X_{t,1}, \ldots, X_{t,d}))_{t \in \mathcal{J}}.
\]

where \( X_{t,j} \) denotes the amount of money invested from time \( t - \delta \) to \( t \) in the zero-coupon bond with time to maturity \( j\delta \). We define the \( \mathbb{F} \)-adapted return process \( (R_t)_{t \in \mathcal{J}} = ((R_{t,1}, \ldots, R_{t,d}))_{t \in \mathcal{J}} \) by

\[
R_{t,j} = \begin{cases} 
R_{t,j-1}F_{t-\delta,j}^{-1} = \exp\{j\delta Y_{t-\delta,j} - (j - 1)\delta Y_{t,j-1}\}, & \text{for } j \in \mathcal{D} \setminus \{1\}, \\
F_{t-\delta,j}^{-1} = \exp\{\delta Y_{t-\delta,1}\}, & \text{for } j = 1.
\end{cases}
\]

This return denotes the value at time \( t \in \mathcal{J}' \) of one unit of cash invested at time \( t - \delta \) in the security with time to maturity \( j\delta \). We also introduce the \( \mathbb{F} \)-adapted process \( (V_t^{-}(X))_{t \in \mathcal{J}'} \) defined by the value of \( \mathbf{X} \) at time \( t \in \mathcal{J}' \) before rebalancing, i.e. \( V_t^{-}(X) = X_t'R_t \). The \( \mathbb{F} \)-adapted value process \( (V_t(\mathbf{X}))_{t \in \mathcal{J}} \) for a trading strategy \( \mathbf{X} \) is defined by

\[
V_t(\mathbf{X}) = \begin{cases} 
X_{t+\delta}^{-1}' & \text{for } t < T, \\
VT^{-}(\mathbf{X}) & \text{for } t = T.
\end{cases}
\]

Trading strategies with no outflows or inflows of cash over time are called self-financing. Formally, \( \mathbf{X} \) is self-financing if for all \( t \in \mathcal{J}' \)

\[
V_t(\mathbf{X}) = V_t^{-}(\mathbf{X}). \tag{9}
\]

For \( \mathbf{X} \) self-financing we identify \( V_t^{-}(\mathbf{X}) \) with \( V_t(\mathbf{X}) \) and we only use the latter notation. Note that linear combinations of self-financing trading strategies are self-financing.

An arbitrage opportunity is a self-financing trading strategy \( \mathbf{X} \) having initial value \( V_0(\mathbf{X}) = 0 \), terminal value \( VT(\mathbf{X}) \geq 0 \), \( \mathbb{P} \)-a.s., and \( \mathbb{P}[VT(\mathbf{X}) > 0] > 0 \). Such trading strategies are...
ruled out in our market model by the existence of an equivalent probability measure satisfying (2). A claim with maturity $T$ is an $\mathcal{F}_T$-measurable random variable $c_T$. We think of $c_T$ as a future liability which has to be covered at time $T$. A claim $c_T$ is called attainable if there exists a self-financing trading strategy $X_{c_T}$ such that $V_T(X_{c_T}) = c_T$, $\mathbb{P}$-a.s. An attainable claim can perfectly be replicated by investing $V_0(X_{c_T})$ at time 0 and adjusting the portfolio holdings at times $t \in J'$ according to the trading strategy $X_{c_T}$. In our market model we have the following.

**Theorem 3.1.** The market model (8) is incomplete, i.e. there are claims with maturity $T$ which are not attainable. In particular, the claim $c_T = 1$, which corresponds to a zero-coupon bond with time to maturity $T > d\delta$, is not attainable.

The first statement follows from Föllmer and Schied [36], Theorem 5.38. The second statement is intuitively clear given our model assumptions. A formal proof is given in Appendix A.3. The unattainability of $c_T = 1$ for $T > d\delta$ is a desired property of our model. For instance, a life-time annuity with expected payments having more than 30 years to maturity cannot be hedged because there are no bonds available at the financial market with such long times to maturity. Bonds of shorter maturities need to be rolled over. This involves reinvestment risk because of the uncertainty in future yields.

In the literature there are several approaches to tackle the problem of hedging claims in incomplete financial markets. Super-replication is one example. A super-replicating strategy $X$ is a self-financing trading strategy such that $V_T(X) \geq c_T$, $\mathbb{P}$-a.s. Such $X$ is often very expensive (see Example 3.4 below) and in some cases it does not even exist. For more details see Föllmer and Schied [36], Sections 7 and 8.

In this paper we follow a more practical approach based on Artzner et al. [2], and Hilli et al. [47]. We define acceptable hedging strategies as those which cover the claim at maturity at an acceptable level of risk. Let $\mathcal{A}$ denote a set of $\mathcal{F}_T$-measurable random variables. This set is called acceptance set and is interpreted as the set of all payoffs at maturity which are within our risk tolerance in hedging the claim $c_T$. For example, in Hilli et al. [47] acceptance sets of the form $\mathcal{A} = \{V \text{ is } \mathcal{F}_T\text{-measurable} \mid \rho(V - c_T) \leq 0\}$ for given risk measures $\rho$ are considered. A self-financing trading strategy $X$ is said to hedge $c_T$ at acceptance level $\mathcal{A}$ if $V_T(X) \in \mathcal{A}$, $\mathbb{P}$-a.s. The initial value of $c_T$ at acceptance level $\mathcal{A}$ is defined by

$$V_0(\mathcal{A}) = \inf_{X \text{ self-financing} \atop V_T(X) \in \mathcal{A}, \mathbb{P} \text{-a.s.}} V_0(X). \quad (10)$$

**Remark 3.2.** In [2] and [47] $\mathcal{A}$ is defined with respect to the net terminal payoff after the claim has been paid, i.e. $V_T(X) - c_T$. We define $\mathcal{A}$ with respect to the terminal payoff $V_T(X)$. Note that $\mathcal{A}$ depends on the claim $c_T$ because it should be chosen reasonably relative to $c_T$. In this way we obtain a more concise representation since the dependence from $c_T$ is implicit in $\mathcal{A}$. For clarity we should write $\mathcal{A} = \mathcal{A}(c_T)$. We drop this explicit dependence in the notation, since later on we fix $c_T = 1$.

**Remark 3.3.** We only consider constraints for the value at the final time horizon $T$. In practice, we might be interested in additional value constraints at previous time points (e.g. at the end of each calendar year). Our methodology can also be applied for these constraints. For computational simplicity we do not consider them in this paper.
The rest of the paper exposes a method to solve (10) numerically for the claim $c_T = 1$ in the market model of Section 2. We define the set of feasible trading strategies by

$$D = \{ X \text{ is a self-financing trading strategy } | X_t \geq 0, \text{ P-a.s., for all } t \in J' \},$$

i.e. we restrict ourselves to long-only trading strategies. Note that $D$ is a convex cone. In this work we consider acceptance sets of the following form

$$A = \{ V \text{ is } F_T \text{-measurable } | E[f(V)] \leq 0 \},$$

(11)

where $f : \mathbb{R} \to \mathbb{R}^m$ is a continuous function with convex components. We consider such functions only on $\mathbb{R}_{\geq 0}$ because $V_T(X) \geq 0$, P-a.s., for all $X \in D$.

**Example 3.4.** Let $m = 1$, $f_0 > 0$ and define $f(v) = \max\{f_0 - v, 0\}$. This corresponds to the acceptance set $A = \{ V | V \geq f_0, \text{ P-a.s.} \}$. In particular, for $f_0 = 1$, this corresponds to super-replication of $c_T = 1$. Note that for all $t \in J'$ we have $R_{t+1} \geq 1$, P-a.s. Then, the risk-free roll over with sufficiently large initial capital provides a super-replicating strategy. Typically, such a super-replication is rather capital inefficient.

**Example 3.5.** Let $m = 1$, $f_0 > 0$, $u > 0$, $k \in \mathbb{N}$ and define $f(v) = \max\{f_0 - v, 0\}^k - u^k$. In this way we require the upper bound $u^k$ to the lower $k$-th partial moment with target level $f_0$. We interpret $u > 0$ as a risk tolerance parameter. Higher lower partial moments indicate higher shortfall risk. We use these acceptance sets in the numerical example of Section 6.

Summarizing we study the following optimization problem

$$\begin{align*}
\text{minimize} & \quad V_0(X) \quad \text{over } \quad X \in D \\
\text{subject to} & \quad E[f(V_T(X))] \leq 0.
\end{align*}$$

(12)

This corresponds to (10) for the acceptance set (11). Solving (12) we achieve the following: a methodology to price and hedge non-tradable long term zero-coupon bonds with tradable ones under a certain risk tolerance and no short positions, where risk is caused by uncertainty in the interest rates on the unavoidable reinvestments.

### 4 Acceptable hedging and stochastic optimization

#### 4.1 Galerkin approximation

Since the distribution of $(Y_t)_{t \in J}$ is continuous, $D$ is typically infinite dimensional and (12) cannot be solved analytically or using standard algorithms for convex optimization. For this reason we consider finite dimensional approximations. Instead of (12) it is more convenient for the application of certain numerical methods to consider the problem

$$\begin{align*}
\text{minimize} & \quad V_0(X) + \theta E[f(V_T(X))] \quad \text{over } \quad X \in D,
\end{align*}$$

(13)

1Throughout this paper for a vector $v$ we write $v \leq 0$ meaning that all components of $v$ are lower or equal to zero. In the same sense we also write $v \geq 0$. 

154
where \( \theta : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \) denotes the convex and increasing penalty function

\[
v \mapsto \theta(v) = \sum_{i=1}^{m} \theta_{0i} \max\{ v_i, 0 \}^2,
\]

for \( \theta_{01}, \ldots, \theta_{0m} > 0 \). This is because defining parametric classes of trading strategies in \( \mathcal{D} \) is easier compared to \( \mathcal{D} \cap \{ X | V_T(X) \in \mathcal{A}, \text{P-a.s.} \} \). The second term in (13) is positive whenever the inequality constraints in (12) are violated and zero otherwise. The solution of (13) approximates the one of (12) for large \( \theta_{01}, \ldots, \theta_{0m} \). The choice of smooth \( \theta \) is convenient for the application of numerical optimization algorithms. The problem (13) is convex (see Appendix A.4).

**Remark 4.1.** An alternative approach to obtain numerical solutions for (12) is to follow Hilli et al. [47] and minimize \( E[f(V_T(X))] \) for fixed values of the initial cost \( V_0(X) \). Then, the optimal initial wealth is the value for which the minimum is zero.

Numerical methods to approximate infinite dimensional optimization problems such as (13) to finite dimensional ones have been discussed in the literature. A possible approach is the discretization of \( (Y_t)_{t \in J} \), see e.g. Hilli and Pennanen [49] or Pflug [70]. This technique is referred to as scenario discretization. The main idea is to construct a finite set of scenarios for the yield curve. Such a discretization leads to a finite dimensional \( \mathcal{D} \) and a standard convex optimization problem that can be solved with conventional algorithms.

A well known method to construct such scenarios is conditional sampling, where these are obtained by recursively sampling from the conditional distribution \( Y_t | \mathcal{F}_{t-\Delta} \), see Shapiro [73]. However, scenario discretization presents some serious computational drawbacks. The dimensionality of the resulting approximation typically increases exponentially as the number of trading periods \( n \) rises. We do not want to be constrained in the number of trading periods and therefore we do not consider scenario discretization.

Instead, we turn our attention to the Galerkin method, see Koivu and Pennanen [57], and Hilli et al. [48]. This is computationally more attractive and easier to implement. The idea is to look for optimal solutions not over the entire set \( \mathcal{D} \), but over a subset of \( \mathcal{D} \) consisting of finite linear combinations of feasible trading strategies, called basis strategies. Let \( \chi = (X(1), \ldots, X(N))' \in \mathcal{D}^N \) be \( N \) such basis strategies. The set of conical combinations \( \{w'\chi | w \in \mathbb{R}^N_{\geq 0} \} \) is a finite dimensional subset of \( \mathcal{D} \). Let \( V_t(\chi) = (V_t(X(1)), \ldots, V_t(X(N)))' \) for \( t \in J \). The Galerkin approximation of (13) is given by

\[
\text{minimize } w'V_0(\chi) + \theta(E[f(w'V_T(\chi))]) \text{ over } w \in \mathbb{R}^N_{\geq 0}.
\]

The convexity is preserved in (14), because it is simply the convex optimization problem (13) restricted to a finite dimensional convex subset of \( \mathcal{D} \). This approximation leads to a standard finite dimensional convex minimization problem for which standard numerical algorithms can be used.

Observe that (14) is in some sense a one-period optimization problem. It depends directly only on \( V_0(\chi) \) and \( V_T(\chi) \). In contrast to scenario discretization, increasing \( n \) does not increase the complexity of the optimization directly. The dimension of (14) depends solely on \( N \). The complexity of evaluating the term \( E[f(w'V_T(\chi))] \) typically depends on the number of trading periods \( n \). This is because \( n \) influences the valuation of \( V_T(\chi) \). Since in
our market model we are not able to find closed form solutions for such expectations (except for very simple special cases) we need to evaluate them through Monte Carlo simulations, which are computationally more expensive if $n$ is large. This is nevertheless easier and typically less time consuming compared to the evaluation in tree-based discretization. In particular, for the model of Section 2, we do not have to work with nested simulations. We only need to generate $nd$ standard Gaussian random numbers in order to obtain one observation of $V_T(\chi)$. Because of the computational advantages, we take (14) together with (8) as our basic model to partially hedge and price long term zero-coupon bonds.

### 4.2 Parametric basis strategies

In this section we consider four parametric families of basis strategies which are in $\mathcal{D}$ for every possible parameter choice. These are buy & hold (BH), fixed times to maturity (FTM), fixed proportions (FP) and target date fund (TDF) trading strategies. Let $V_0 \geq 0$ and $\pi = (\pi_1, \ldots, \pi_d) \in \mathbb{R}_{\geq 0}^d$ so that $\pi'1 = 1$. The BH, FTM and FP families are defined by:

\[ \begin{align*}
X_t^{BH} &= (X_t^{BH}R_{t-\delta,2}, \ldots, X_t^{BH}R_{t-\delta,d}, X_t^{BH}R_{t-\delta,1})', \\
X_t^{FTM} &= (X_t^{BH}R_{t-\delta,1}, \ldots, X_t^{BH}R_{t-\delta,d})', \\
X_t^{FP} &= (R_{t-\delta}X_t^{FP})\pi.
\end{align*} \]

At time 0 we invest $V_0$ in the $d$ zero-coupon bonds according to the proportions $\pi$. For BH trading strategies we do not rebalance at future time points. At each point in time we invest the cash obtained from the security that expires in the newly available zero-coupon bond with time to maturity $\delta d$. For FTM trading strategies we rebalance the portfolio at each point in time so that the total portfolio value is invested in each bond in the security with the same time to maturity. For FP trading strategies we rebalance the portfolio at each point in time so that the total portfolio value is invested in each bond according to the proportions $\pi$.

We define TDF trading strategies similarly to Bodie and Treuillard [8]. Fix $d' \in \mathbb{D}^-$ and divide the set of times to maturity into two disjoint subsets $\{1, \ldots, d'\}$ and $\{d'+1, \ldots, d\}$. The bonds with shorter maturities $1, \ldots, d'$ are referred to as lower duration and those with longer maturities as higher duration. Let $\eta = (\eta_1, \ldots, \eta_d) \in \mathbb{R}_{\geq 0}^d$ so that

\[ \sum_{j=1}^{d'} \eta_j = \sum_{j=d'+1}^{d} \eta_j = 1. \]

Let $0 \leq p_\delta < p_T \leq 1$ be the initial and terminal proportions invested in the lower duration bonds. We increase this proportion linearly from $p_\delta$ to $p_T$ to reflect the fact that the duration of the claim decreases as its expire date approaches. That is,

\[ p_t = p_\delta + \frac{p_T - p_\delta}{T - \delta}(t - \delta). \]

The trading strategy is then defined by $X_t^{TDF} = (p_\delta \eta_1 1_{1 \leq j \leq d'} + (1 - p_\delta) \eta_1 1_{d' < j \leq d})V_0$ and for $t > \delta$:

\[ X_t^{TDF} = \left( p_\delta \eta_1 1_{1 \leq j \leq d'} + (1 - p_\delta) \eta_1 1_{d' < j \leq d} \right) R_{t-\delta}X_t^{TDF}, \quad j \in \mathcal{D}. \]
This means that at each point in time the portfolio is rebalanced so that the proportion invested in low duration bonds is \( p_l \), the one invested in high duration bonds is \( 1 - p_l \) and those invested within the low and high duration ones are given by \( \eta \). Note that these four families are, in general, not conical. Therefore, using trading strategies in one of these as basis for (14), we optimize over strategies that do not belong to the family.

**Remark 4.2.** Let \( X \) be a BH, FTM, FP or TDF trading strategy, \( X_{\text{norm}} \) have \( V_0 = 1 \) and otherwise be identical to \( X \). Then, \( X_t = V_0 X_{t_{\text{norm}}} \) for \( t \in J' \).

For \( j_1, \ldots, j_n \in D \) we define the \( \mathbb{F} \)-adapted stochastic process \((B_t(j_1, \ldots, j_n))_{t \in J'}\) by \( B_0(j_1, \ldots, j_n) = 1 \) and for \( t \in J' \)

\[
B_t(j_1, \ldots, j_n) = \prod_{s=1}^{\gamma t} R_{s,j_s}. \tag{15}
\]

This process corresponds to the value at time \( t \) of investing one unit of cash in the zero-coupon bond with time to maturity \( j_1 \) at time 0, rolling over at time \( \delta \) in the one with time to maturity \( j_2 \), and so on for \( j_3, \ldots, j_n \). In particular we have that \( B_t(1, \ldots, 1) = B_t \) is the classical bank account process. Note that we have \( B_t(j_1, \ldots, j_n) = B_t(j_1, \ldots, j_2, 1, \ldots, 1) \) for all \( j_1, \ldots, j_n \in D \) and \( t \in J' \). Applying the definitions above recursively we have the following result (see Appendix A.5):

**Lemma 4.3.** Let \( X \) be a FTM, FP or TDF strategy. Then, we have for \( t \in J' \)

\[
V_t(X) = \sum_{j_1, \ldots, j_{\gamma t} = 1}^d a_t(j_1, \ldots, j_{\gamma t}, 1, \ldots, 1) B_t(j_1, \ldots, j_{\gamma t}, 1, \ldots, 1),
\]

where

\[
a_t(j_1, \ldots, j_n) = \begin{cases} 
V_0 \pi_{j_1} & \text{if } X \text{ is FTM}, \\
V_0 \prod_{s=1}^{\gamma t} \pi_{j_s} & \text{if } X \text{ is FP}, \\
V_0 \prod_{s=1}^{\gamma t} (p_{s,j} \delta)_{1_{j_s} < j_s} (1 - p_{s,j})_{\eta_j} & \text{if } X \text{ is TDF}.
\end{cases}
\]

The above lemma states that the value processes for FTM, FP and TDF trading strategies are linear combinations of the random variables \( B_t(\ldots) \) with deterministic coefficients. These random variables depend only on the market model whereas the coefficients \( a_t(\ldots, \cdot) \) depend only on the strategy parameters. Thus, we obtain a split in two terms, the market model (random) and the deterministic coefficients from the strategy which allows to evaluate the objective function in (14) more efficiently. A closed form expression in terms of \( B_t(\ldots) \) can also be worked out for the value process of BH trading strategies, but it does not have a simple form as for the other three families. On the distribution of \( B_t(\ldots) \) in our model we know the following (see Appendix A.6).

**Lemma 4.4.** Let \( t \in J' \), \( s = \frac{t}{\delta} \in \{1, \ldots, n\} \) and \( j_1, \ldots, j_n \in D \). For \( j_s > 1 \), the random variable \( \log B_t(j_1, \ldots, j_n) \) is normally distributed under \( \mathbb{P} \), given \( \mathcal{F}_{t-s} \), with conditional mean

\[
\log P_{t-s,1} + \log B_{t-s}(j_1, \ldots, j_n) + \sqrt{\delta} (\zeta(Y_{t-s}) \Delta \mu_t)_{j_s-1} - \frac{\delta}{2} \text{sp}(\Sigma(Y_{t-s}))_{j_s-1}.
\]
and variance $\delta \text{sp}(\Sigma(Y_{t-\delta}))_{j_s-1}$. For $j_s = 1$, $B_t(j_1, \ldots, j_n)$ is previsible and given by $B_t(j_1, \ldots, j_n) = B_{t-\delta}(j_1, \ldots, j_n)P_{t-\delta,1}$.

The conditional expectation and variance of $B_t(\ldots)$, assuming $j_s > 1$, are given by

\[
\begin{align*}
E[B_t(j_1, \ldots, j_n)|F_{t-\delta}] &= B_{t-\delta}(j_1, \ldots, j_n)P_{t-\delta,1} \exp \left\{ \sqrt{\delta} (\zeta(Y_{t-\delta}) \Lambda \mu_t)_{j_s-1} \right\}, \\
\text{Var}[B_t(j_1, \ldots, j_n)|F_{t-\delta}] &= E[B_t(j_1, \ldots, j_n)|F_{t-\delta}]^2 \left( \exp \left\{ \delta \text{sp}(\Sigma(Y_{t-\delta}))_{j_s-1} \right\} - 1 \right).
\end{align*}
\]

Observe that the conditional distribution of $B_t(\ldots)$, given $F_{t-\delta}$, does not depend on $\beta$. If we condition over multiple periods we obtain distributions which are not log-normal and depend on $\beta$.

5 Calibration procedure

Our aim is to solve (14) numerically under the model of Section 2. This requires the calibration to market data. Once $h$ and $\beta$ have been chosen, it remains to calibrate $\Lambda$ to real market data. Our explicit choice of $h$ and $\beta$ is discussed in Section 6. Assume that $\mu \equiv 0$, i.e., $\mu_t = 0$, $F$-a.s., for all $t \in J'$. This implies that the real world measure $P$ satisfies condition (2). This assumption is discussed in more details in Section 6. Under this assumption, using Lemma 2.2 and equation (26) in the appendix, we have that $Y_t$ is multivariate Gaussian distributed, given $F_{t-\delta}$, with conditional mean

\[
\delta \left[ -Y_{t-\delta,1}1 + \frac{1}{2} \text{sp}(\Sigma(Y_{t-\delta})) + \left( \beta(Y_{t-\delta}) + \frac{1}{2} (\text{sp}(\Sigma(Y_{t-\delta}))_{d} - \text{sp}(\Sigma(Y_{t-\delta}))_{d-1}) \right) \right],
\]

and conditional covariance matrix $\delta \Sigma(Y_{t-\delta})$ for $t \in J'$. Therefore, we do not need to directly calibrate $\Lambda$ to market data but only $\Sigma(\cdot)$ as defined in (7). In particular, we need to construct estimators for the elements of $\Lambda'$. In this section we show that the calibration procedure presented in Teichmann and Wütrich [77] can also be applied with slight modifications to our market model specified by the above conditional distribution.

Let $(Y_{k\delta})_{k=0, \ldots, K}$ be observations of the yield curve and $(Y_{k\delta})_{k=1, \ldots, K}$ be the corresponding transformed observations. We introduce some additional notation to simplify the expressions below. For $k = 0, \ldots, K$ let $b_k = Y_{k\delta}11_{k}\delta$, $\beta = \beta(Y_{k\delta})$, $\gamma_k = \beta - b_k$, $\zeta = \zeta(Y_{k\delta})$, $\zeta_{k,j} = \text{sp}(\zeta_{k})_j$ for $j \in D$ and $\Sigma_k = \Sigma(Y_{k\delta})$. Define the following matrix based on the observations,

\[
C(K) = \frac{1}{\sqrt{\text{det}(\Sigma)}} \left( \zeta_{k-1} Y_{k\delta} \right)_{k=1, \ldots, K} \in \mathbb{R}^{K \times K},
\]

and the function $S(K) : \mathbb{R}^{d} \rightarrow \mathbb{R}^{d \times d}$ by $y \mapsto \zeta(y)C(K)C'(K)\zeta(y)$. From the the proof of Theorem 3.4 in [77] we have for $y \in \mathbb{R}^{d}$

\[
\mathbb{E}[S(K)(y)] = \delta \Sigma(y) + \frac{\delta^2}{K} \sum_{k=0}^{K-1} \zeta(y) \mathbb{E}[G_{h,\beta,\Sigma}(b_k1, Y_{k\delta})] \zeta(y),
\]

where $G_{h,\beta,\Sigma} : \mathbb{R}^{d} \times \mathbb{R}^{d} \rightarrow \mathbb{R}^{d \times d}$ has the form

\[
(z, y) \mapsto \zeta(y)^{-1}g_{\beta,\Sigma}(z, y)g_{\beta,\Sigma}(z, y)^{-1} \zeta(y)
\]

where $g_{\beta,\Sigma}$ is a function of $\beta$ and $\Sigma$.
with \( g_{\beta,\Sigma} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) given by
\[
(z, y) \mapsto -z + \frac{1}{2} \text{sp}(\Sigma(y)) + \left( \beta(y) - \frac{1}{2} \left( \text{sp}(\Sigma(y))_d - \text{sp}(\Sigma(y))_{d-1} \right) \right) e_d.
\]

Similar to (3.6) in [77] we get the crucial observation in (16) that \( S_{(K)}(y) \) may serve as an estimator for \( \delta \Sigma(y) \) with a bias going to zero as \( \delta \rightarrow 0 \).

We use (16) to derive estimators for \( \Lambda \). We set \( y = 1 \). For \( i, j \in D^- \) we obtain the same estimators for \( s_{ij} \) as in [77] formulas (3.12)-(3.13). For the last diagonal element \( s_{dd} \) we have
\[
\frac{\mathbb{E}[S_{(K)}(1)]}{\delta \zeta(1)} = s_{dd} + \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[ \left( \gamma_k + \frac{1}{2} \zeta_k^2_{d-1} \right) \frac{1}{2} \zeta_k^{d-2} \right]
\]
\[
= s_{dd} + \frac{\delta}{K} \sum_{k=0}^{K-1} \left( A_k^{(1)} + s_{d-1,d-1} A_k^{(2)} + \frac{1}{4} s_{d-1,d-1} A_k^{(3)} \right),
\]
where we have used that \( \zeta_k, d-1 = \zeta_k, d \) for all \( k = 0, \ldots, K \) and set
\[
A_k^{(1)} = \mathbb{E}[\gamma_k^2 \zeta_k^{d-2}], \quad A_k^{(2)} = \mathbb{E}[\gamma_k] \quad \text{and} \quad A_k^{(3)} = \mathbb{E}[\zeta_k^2].
\]

For the elements \( (s_{id})_{i=1, \ldots, d-1} \) we have
\[
\frac{\mathbb{E}[S_{(K)}(1)]}{\delta \zeta(1)_{ii} \zeta(1)_{dd}} = s_{id} - \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[ b_k \left( \gamma_k + \frac{1}{2} \zeta_k^2_{d-1} s_{d-1,d-1} \right) \zeta_k^{d-1} \right]
\]
\[
+ \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[ \frac{1}{2} \zeta_k^2 \zeta_k^{d-1} \left( \gamma_k + \frac{1}{2} \zeta_k^2_{d-1} s_{d-1,d-1} \right) \zeta_k^{d-1} \right]
\]
\[
= s_{id} + \frac{\delta}{K} \sum_{k=0}^{K-1} \left( B_{k,i}^{(1)} + \frac{1}{2} s_{d-1,d-1} \left( B_{k,i}^{(2)} + B_{k,i}^{(3)} \right) + \frac{1}{2} s_{ii} B_{k,i}^{(4)} \right),
\]
where we have used again that \( \zeta_k, d-1 = \zeta_k, d \) for all \( k = 0, \ldots, K \) and set
\[
B_{k,i}^{(1)} = \mathbb{E}[b_k \gamma_k \zeta_k^{d-1} \zeta_k^{d-1}], \quad B_{k,i}^{(2)} = \mathbb{E}[b_k \zeta_k^{d-1} \zeta_k^{d-1}],
\]
\[
B_{k,i}^{(3)} = \mathbb{E}[\gamma_k \zeta_k \zeta_k^{d-1} \zeta_k^{d-1}] \quad \text{and} \quad B_{k,i}^{(4)} = \mathbb{E}[\zeta_k, d \zeta_k, d].
\]

Note that \( s_{ii} = s_{dd} \) for all \( i \in D \). We can derive estimates for \( s_{id} \) in the following way. We replace the expectations (18) and (20) by the observations. For the diagonal elements \( (s_{ii})_{i \in D^-} \) we can use the estimates derived in [77] formula (3.12). By solving the linear equations (17) and (19) for \( s_{id} \) we obtain the estimates. This is the approach presented in [77] and completes our estimation procedure.

\section{Numerical example: Swiss government bonds}

\subsection{Choice of data, \( h(\cdot) \) and \( \beta(\cdot) \)}

In this section we consider CHF zero-coupon bond yields derived from Swiss government bonds. We work on a monthly grid, i.e. \( \delta = 1/12 \). The Swiss government issues bonds
with times to maturity in
\[ \{1/12, 1/6, 1/4, 1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30\}. \]

However, securities with times to maturity of 15 years or more are not liquid. For this reason we consider only bonds with times to maturity up to 10 years and set \( d = 120 \).

Note that medium and long term Swiss government bonds (i.e. those issued with times to maturity larger than one year) are coupon bearing securities. Therefore, the yield curve needs to be extrapolated from their market prices. For this purpose the Swiss National Bank uses the Svensson method, see Müller [63]. Bloomberg L.P. uses a similar procedure and provides these extrapolated rates for a history of more than 18 years and times to maturity in \( \{j\delta | j \in D_{\text{obs}}\} \), where
\[ D_{\text{obs}} = \{3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120\}, \]
see Bloomberg tickers: F25603M Index, F25606M Index, F25601Y Index and so on. We choose these time series from 31.01.1995 to 31.12.2012 (i.e. \( K = 215 \)) as our monthly yield curve observations. Therefore, we have observations
\[ (Y_{k\delta,j})_{k=0,...,K}, j \in D_{\text{obs}}. \]

To obtain the yield curve for all times to maturity in \( D \) we set for \( k \in \{0, \ldots, K\}, j \in D \setminus D_{\text{obs}} \) and \( j > 3 \)
\[ Y_{k\delta,j} = \frac{j - d_l(j)}{d_u(j) - d_l(j)} Y_{k\delta,d_l(j)} + \frac{d_u(j) - j}{d_u(j) - d_l(j)} Y_{k\delta,d_u(j)}, \]
where \( d_l(j) = \max\{k \in D_{\text{obs}} | k \leq j\} \) and \( d_u(j) = \min\{k \in D_{\text{obs}} | k \geq j\} \). This means we interpolate linearly between the two nearest known values. We continue the curve linearly at the short end to extrapolate data for \( \delta j = 1/12 \) and \( 1/6 \). In this way we obtain observations \( (Y_{k\delta})_{k=0,...,K} \) for the full yield curve. Some of these are shown in Figure 6.1 below.

For the volatility scaling factor we choose the square root scaling \( h(y) = \sqrt{y} \), which has been proved reasonable for yield curve modeling, see Guillaume et al. [27]. For the long end slope factor, which describes the steepness of the yield curve at the long end, we make the following intuitive choice
\[ \beta(y) = \frac{1}{\delta} (d \delta y_d - (d - 1) \delta y_{d-1}) = dy_d - (d - 1) y_d - 1. \]

6.2 Calibration results

In order to test if our model is reasonable and consistent, we split the sample of Section 6.1 in the middle and calibrate \( \Lambda' \) to the first half. We use the procedure described in Section 5 under the assumption \( \mu \equiv 0 \). This assumption is discussed below. For the calibration we use Teichmann and Wüthrich [77] formulas (3.12)-(3.13), and equations (17) and (19). For these estimators we consider first and second order terms in \( \delta \). We only use the first half of the data for the calibration since \( \Lambda' \) is constant by model assumption, and we
want to test if the model provides reasonable and consistent out-of-sample results. For this purpose we define the following process for \( \vartheta \in \mathbb{R}^d \)

\[
(\phi_t(\vartheta))_{t \in J'} = \left( \frac{1}{\sqrt{\delta}} \vartheta' \zeta(Y_{t-\delta})^{-1}(\hat{Y}_t - \psi_t) \right)_{t \in J'},
\]

(21)

where

\[
\psi_t = \delta \left[ -Y_{t-\delta} + 1 + \frac{1}{2}\text{sp}(\zeta(Y_{t-\delta}) \Lambda \zeta(Y_{t-\delta})) + \beta(Y_{t-\delta})e_d \right.
\]

\[
+ \frac{1}{2}\text{sp}(\zeta(Y_{t-\delta}) \Lambda \zeta(Y_{t-\delta}))e_d - \frac{1}{2}\text{sp}(\zeta(Y_{t-\delta}) \Lambda \zeta(Y_{t-\delta}))_{d-1}e_d \right],
\]

which depends on \( \Lambda \Lambda' \). A direct consequence of Lemma 2.2 under the assumption \( \mu^- \equiv 0 \) is that \( \psi_t = \mathbb{E}[Y_t|\mathcal{F}_{t-\delta}] \) for all \( t \in J' \) and for any \( \vartheta \in \mathbb{R}^d \)

\[
(\phi_t(\vartheta))_{t \in J'} \overset{iid}{\sim} \mathcal{N}(0, \vartheta' \Lambda \Lambda' \vartheta).
\]

We compute (21) for different values of \( \vartheta \) using the observations \( (Y_{k\delta})_{k=0,...,K} \) and the matrix \( \Lambda \Lambda' \) calibrated to the first half of the sample. We do this for

\[
\vartheta_1 = \frac{1}{[\frac{d}{4}]^2} \sum_{j=1}^{[\frac{d}{4}]} e_j, \quad \vartheta_2 = \frac{1}{[\frac{d}{4}]^2} \sum_{j=\lceil \frac{d}{4} \rceil + 1}^{d-\lceil \frac{d}{4} \rceil} e_j, \quad \text{and} \quad \vartheta_3 = \frac{1}{[\frac{d}{4}]^2} \sum_{j=d-\lceil \frac{d}{4} \rceil + 1}^{d} e_j,
\]

(23)

which correspond to short, medium and long term parts of the yield curve. We obtain observations

\[
(\phi_{k\delta}(\vartheta_i))_{k=0,...,K}, i=1,2,3.
\]

If the model is reasonable then these sequences should look like i.i.d. normal random variables including the out-of-sample part, see (22). Figure 6.2 shows that the serial correlation is low also for small lags in all three cases as expected from the model. From Figure 6.3 we can see that the observations of the sequences appear to be reasonable given the model over the full period. There are no particular changes visible in the second half of the sample which has not been used to calibrate the model. Table 6.1 provides some statistics. The observed standard deviations are close to the ones given by the model. Application of the Jarque-Bera test (see [53]) shows that for the medium and long term parts the normality of the series cannot be rejected at 5% significance level, whereas for the short end normality is rejected at 1% level. The series also appear to be slightly negatively drifted. Given these results, we conclude that our model does a good job in explaining the yield curve development in the medium and long term parts, whereas for the short term part the model does not appear to work optimally. The assumption \( \mu^- \equiv 0 \) does not seem optimal either.

There is no clear procedure to estimate the market price of risk process \( \mu \). A heuristic approach is to extrapolate it from the sample applying linear filtering techniques. Lemma 2.2 implies for any \( \vartheta \in \mathbb{R}^d \) and \( t \in J' \) that \( \vartheta' \Lambda \mu_t = \mathbb{E}[\phi_t(\vartheta)|\mathcal{F}_{t-\delta}] \). As discussed above the series \( (\phi_{k\delta}(\vartheta_i))_{k=0,...,K} \) for \( i = 1,2,3 \) show very low serial correlation and we can see
from Figure 6.4 that a constant drift assumption is reasonable for these series. Hence, we make the following approximation for \( t \in J' \) and \( i = 1, 2, 3 \)

\[ \vartheta_i \Lambda t \approx \frac{1}{K+1} \sum_{k=0}^{K} \phi_{k\delta}(\vartheta_i). \]

We can use these approximations of the projections on the short, medium and long term parts to obtain an approximation for \( \Lambda \mu \), i.e. we set for \( t \in J' \)

\[ \Lambda t \approx \lfloor \frac{d}{4} \rfloor + 1 \sum_{k=0}^{K} \left( \phi_{k\delta}(\vartheta_1)\vartheta_1 + \phi_{k\delta}(\vartheta_2)\vartheta_2 + \phi_{k\delta}(\vartheta_3)\vartheta_3 \right), \quad (24) \]

In this way we approximate \( \mu \) by a time-independent deterministic vector.

### 6.3 Pricing and replication for \( T = 20 \)

Our aim is to solve (14) numerically for \( T = 20 \) (i.e. \( n = 240 \)) and the stochastic model of Section 2 calibrated on the sample of Section 6.1. The first step is to simulate yield curve developments. We set \( \mu \equiv 0 \). Alternatively the heuristic approach (24) might be used. The complexity of the pricing algorithm is not affected by the presence of \( \mu \). We calibrate the model on the full sample using the procedure of Section 6.2. Our market model provides two essential features for this application. First, as we verified in the previous section, the calibrated parameters are appropriate over a long period of time. This is important since the calibrated model is being used to simulate yield paths for several years ahead. Secondly, generating yield curves in our model does not require nested simulations. This is crucial since we need to simulate with reasonable computational time monthly yield curve observations for a 20 years period several thousand times.

We choose the yield curve at 31.12.2012 as the starting point for our simulation. Observe that financial market conditions in 2012 resulted in extraordinary low yields for Swiss government bonds with flat and near zero short term rates. For the number of simulated developments (each one having 20 years of monthly yield curve simulations) we consider \( M = 125, 250, 500, 1000, 2000 \) and 4000.

We consider the acceptance set given in Example 3.5. We set \( m = 1, f_0 = 1 \) and \( k = 2 \). This means \( f(v) = \max\{1-v, 0\}^2 - u^2 \), where we consider the values \( u = 0.2, 0.15, 0.1, 0.05 \) and 0.025 for the risk tolerance parameter. This corresponds to the acceptance set

\[ \mathcal{A} = \{ V \text{ is } \mathcal{F}_T - \text{measurable} \mid E[\max\{1 - V, 0\}^2] \leq u^2 \}, \]

i.e. terminal values are acceptable if the shortfall risk is not greater than \( u^2 \), where this is measured by the lower partial second moment of the loss. Hence, a lower value of \( u \) corresponds to a lower risk tolerance.

We define 10 basis strategies, i.e. \( N = 10 \). We set \( V_0 = 1 \) and \( \pi_i = \vartheta_i \) for \( i = 1, 2, 3 \), where \( \vartheta_i \) denote the weight vector defined in Section 6.2. Using these parameters we define 3 BH, FTM and FP strategies. Let \( p_0 = 0, p_T = 1, d' = d - \lfloor \frac{d}{4} \rfloor \) and \( \eta = \frac{d'}{2} \pi_1 + \frac{d}{2} \pi_2 + \pi_3 \). From this we define one TDF strategy. We introduce the abbreviations BH1, BH2, BH3, FTM1, FTM2, FTM3, FP1, FP2, FP3 and TDF to reference these 10 basis strategies, where
the numbers correspond to \( \pi_1, \pi_2, \pi_3 \) respectively. We evaluate the terminal values of these strategies for each simulated yield curve development. This can be done very efficiently for FTM, FP and TDF strategies using Lemma 4.3.

**Remark 6.1.** The choice \( V_0 = 1 \) is irrelevant to (14). To see this, let \( \chi = (X^{(1)}, \ldots, X^{(N)})' \) be a \( N \)-dimensional vector of BH, FTM, FP or TDF trading strategies, \( \chi^\text{norm} \) have \( V_0(\chi) = 1 \) and otherwise be identical to \( \chi \). Then, Remark 4.1 implies

\[
\{w'\chi \mid w \in \mathbb{R}_+^N \} = \left\{ \sum_{i=1}^N w_i V_0(X^{(i)})X^{(i)} \mid w \in \mathbb{R}_+^N \right\} = \{w'\chi^\text{norm} \mid w \in \mathbb{R}_+^N \}.
\]

This implies that the solution of (14) for BH, FTM, FP or TDF basis strategies does not depend on \( V_0(\chi) \).

Finally, we solve (14) using a quasi-Newton method. The expectations in (14) are evaluated using the terminal values obtained from the simulations. In this example we consider only 10 basis strategies and so the time needed to solve the optimization accounts only for a minimal part of the total computational time, which is almost entirely determined by the simulations. If we consider hundreds of basis strategies then the situation might be different.

Our numerical results for \( T = 20 \) are presented in Tables 6.2, 6.3 and Figure 6.5. We have solved (14) for different values of \( M \) and \( u \). First, we see that the resulting prices in Table 6.2 are stable as \( M \) increases for fixed values of \( u \). This indicates convergence. Secondly, we observe that prices increase considerably with decreasing \( u \). This makes sense since we require lower shortfall risk for the optimal replicating strategy. Reasonably the risk tolerance parameter plays a crucial role in determining the value of the long term zero-coupon bond. In Table 6.3 we report for fixed \( M \) some risk measures related to the probability distribution of the loss. We observe that all the risk measures considered decrease as \( u \) decreases. This is sound and allows shortfall risk to be managed with respect to several risk measures at the same time by selecting \( u \). We also observe that the lower partial second moments are nearly equal to the values of \( u \). Of course, this must be the case since we are using this measure for the optimization. From Figure 6.5 we observe that the solution trend to be concentrated in 3-4 basis strategies and the concentration increases as \( u \) decreases. These are BH1, FTM1, FP1, which invest in short term securities, and TDF, which invests across all maturities and increases the proportion invested in short term securities as time passes. This concentration makes intuitively sense. In our example BH1, FTM1 and FP1 have low terminal value volatility which appears to be an attractive property for the optimization algorithm. The attractiveness of TDF might be explained as follows. Among our basis strategies, TDF most closely matches the duration over time of the long term zero-coupon bond.

Next we apply our pricing algorithm for \( T = 11, 12, \ldots, 20 \) and derive an extension of the CHF yield curve for these times to maturity. For this purpose we cannot set one single value for \( u \) across all times to maturity since we are considering losses over different periods and this would reflect different risk tolerances across maturities. Therefore we set \( u_{11} > 0 \) and consider the square root of time scaling

\[
u_T = u_{11} \sqrt{T - 10},
\]
for $T = 11, 12, \ldots, 20$. This rule is well known in the literature to scale volatility estimates and relies on the rationale of one-period returns being i.i.d. over time. In non-i.i.d. environments, such as in our model, this rule produces estimates which are reasonable on average but show too large volatility fluctuations, see Christoffersen et al. [18] for more details. In Figure 6.6 we show the extension. This extension has a positive spread compared to the market yield curve which becomes smaller as $u_{11}$ decreases. This makes intuitively sense since the derived yields are not risk-free. Because of reinvestment risk we can only replicate those securities up to a certain risk tolerance specified by $u_{11} > 0$. We also observe that the steepness and concavity of the extension matches the market yield curve better as $u_{11}$ decreases. Thus, we obtain a natural extension of the yield curve beyond the last liquidity traded time to maturity. This extrapolation is obtained by searching explicitly for replicating strategies up to the risk tolerance.

Figure 6.1: CHF yield curve (values in percent) at the end of December for some of the years in the sample. We can observe that the yields at the end of 2012 are at historical lows with short term rates being near zero.
Figure 6.2: Autocorrelation of the observed series \( (\phi_k \delta(\cdot))_{k=0,\ldots,K} \) for different lags. The three lines correspond to \( \vartheta_1 \) (short), \( \vartheta_2 \) (medium) and \( \vartheta_3 \) (long). See also (21), (22) and (23).

Figure 6.3: Values of the observed series \( (\phi_k \delta(\cdot))_{k=0,\ldots,K} \) for \( \vartheta_1 \) (short), \( \vartheta_2 \) (medium) and \( \vartheta_3 \) (long). See also (21), (22) and (23).
Table 6.1: This table presents the sample mean and standard deviation of the observed series \((\phi(\cdot) k \delta)_{k=0,\ldots,K}\) in the three cases. It also presents the standard deviation implied by the model, i.e. \(\sqrt{\vartheta' \Lambda \vartheta}\), and the Jarque-Bera test statistic for normality. The 5% and 1% critical values of the test are 5.99 and 9.21, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Sample Mean</th>
<th>Sample Std</th>
<th>Model Std</th>
<th>Jarque-Bera Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>short</td>
<td>-0.02</td>
<td>0.09</td>
<td>0.10</td>
<td>14.40</td>
</tr>
<tr>
<td>medium</td>
<td>-0.06</td>
<td>0.22</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td>long</td>
<td>-0.09</td>
<td>0.31</td>
<td>0.30</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Figure 6.4: Cumulated values over time of the observed series \((\phi(\cdot) k \delta)_{k=0,\ldots,K}\) for \(\vartheta_1\) (short), \(\vartheta_2\) (medium) and \(\vartheta_3\) (long). See also (21), (22) and (23). We also plot the 24 and 48 months symmetric moving averages to smooth the series out. We can see that the drift is quite stable over time.

<table>
<thead>
<tr>
<th></th>
<th>(u = 20%)</th>
<th>(u = 15%)</th>
<th>(u = 10%)</th>
<th>(u = 5%)</th>
<th>(u = 2.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M = 125)</td>
<td>0.6490</td>
<td>0.7070</td>
<td>0.7703</td>
<td>0.8471</td>
<td>0.8940</td>
</tr>
<tr>
<td>(M = 250)</td>
<td>0.6872</td>
<td>0.7155</td>
<td>0.7788</td>
<td>0.8508</td>
<td>0.8920</td>
</tr>
<tr>
<td>(M = 500)</td>
<td>0.6575</td>
<td>0.7134</td>
<td>0.7811</td>
<td>0.8523</td>
<td>0.8975</td>
</tr>
<tr>
<td>(M = 1000)</td>
<td>0.6564</td>
<td>0.7156</td>
<td>0.7796</td>
<td>0.8533</td>
<td>0.8990</td>
</tr>
<tr>
<td>(M = 2000)</td>
<td>0.6554</td>
<td>0.7143</td>
<td>0.7789</td>
<td>0.8535</td>
<td>0.8992</td>
</tr>
<tr>
<td>(M = 4000)</td>
<td>0.6558</td>
<td>0.7121</td>
<td>0.7773</td>
<td>0.8521</td>
<td>0.8964</td>
</tr>
</tbody>
</table>

Table 6.2: Prices of a 20 years to maturity zero-coupon bond given by the approximation algorithm for different values of \(M, u\).
Table 6.3: $T = 20$ and $M = 4000$. Shortfall risk measures of the optimal portfolio for different values of $u$. LSM1 and LSM2 are the lower standardized moments given by $E[\max\{1 - V(0)^k\}]$ for $k = 1, 2$ respectively. VaR and ES are the Value-At-Risk and Expected Shortfall measures for 95% and 99% confidence levels. All four measures are computed applying sample estimators on the simulated terminal values of the optimal portfolio.

<table>
<thead>
<tr>
<th></th>
<th>LSM1</th>
<th>LSM2</th>
<th>VaR95</th>
<th>VaR99</th>
<th>ES95</th>
<th>ES99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u = 20%$</td>
<td>0.1541</td>
<td>0.2004</td>
<td>0.3254</td>
<td>0.3338</td>
<td>0.3314</td>
<td>0.3393</td>
</tr>
<tr>
<td>$u = 15%$</td>
<td>0.1087</td>
<td>0.1508</td>
<td>0.2655</td>
<td>0.2678</td>
<td>0.2656</td>
<td>0.2738</td>
</tr>
<tr>
<td>$u = 10%$</td>
<td>0.0675</td>
<td>0.1015</td>
<td>0.1913</td>
<td>0.2011</td>
<td>0.1978</td>
<td>0.2066</td>
</tr>
<tr>
<td>$u = 5%$</td>
<td>0.0295</td>
<td>0.0514</td>
<td>0.1123</td>
<td>0.1241</td>
<td>0.1201</td>
<td>0.1304</td>
</tr>
<tr>
<td>$u = 2.5%$</td>
<td>0.0127</td>
<td>0.0262</td>
<td>0.0653</td>
<td>0.0778</td>
<td>0.0739</td>
<td>0.0853</td>
</tr>
</tbody>
</table>

Figure 6.5: Optimal portfolio weights assigned to the 10 basis strategies for $M = 4000$ and different values of $u$. 

167
Conclusions

In this paper we have developed a computationally efficient algorithm to extrapolate the yield curve at the long end by explicitly considering replication strategies. Our procedure is based on a stochastic yield curve model with reinvestment risk which can be effectively calibrated to market data. The hedging strategy is constructed by taking an optimal linear combination of trading rules specified in advance according to a certain risk tolerance. This involves reinvestment risk, which cannot be completely eliminated, and the lower this risk the smaller the spread between the extrapolated yields and the market yields. Thus, we obtain a natural extension of the yield curve beyond the last liquid time to maturity and moreover, we also obtain the hedging strategy which replicates this value.

A Proofs

A.1 Proof of Lemma 2.1

Let $t \in J'$ and $j \in D^-$. We set $\mathbf{y} = \mathbf{Y}_{t-\delta} \in \mathbb{R}^d$ for this proof. For the drift term (4) we have

$$
\alpha_j(y_{j+1}) = \frac{\delta}{2} h(y_{j+1})^2 \sum_{i=1}^{d-1} \lambda^2_{ij} = \frac{\delta}{2} \zeta^2_j \text{sp}(\Lambda^- (\Lambda^-)'_j) = \frac{\delta}{2} \text{sp}(\Sigma^- (\mathbf{y}))_j.
$$
Using (1) we rewrite the model assumption (3) under the real world measure \( P \) and obtain

\[
\tilde{Y}_{t,j} = -\delta Y_{t-\delta,1} + \alpha_j(Y_{t-\delta,j+1}) + \sqrt{\delta} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1}) \lambda_{ij}(\varepsilon_{t,i} - \mu_i,1)
\]

\[
= -\delta Y_{t-\delta,1} + \frac{\delta}{2} \text{sp}(\Sigma^-)(y))_j + \sqrt{\delta}(\zeta^-)(y)\Lambda^-((\xi^- - \mu^-)_j).
\]

The statement on the conditional normality of \( \tilde{Y}_t \) follows directly from (5). \( \Box \)

### A.2 Proof of Lemma 2.2

Let \( t \in J' \) and set \( y = Y_{t-\delta} \) for this proof. For the components \( (\tilde{Y}_{t,j})_{j \in D^-} \) equation (8) is equivalent to (5). Therefore, we only need to consider the random variable \( \tilde{Y}_{t,d} \). From the definitions of \( \zeta(.) \) and \( \Lambda \) observe that \( \zeta(\cdot)_{dd} = \zeta(\cdot)_{d-1,d-1} \), \( \Lambda_{dd} = \kappa \) and \( \Lambda_{d,j} = \Lambda_{d-1,j} = \lambda_{j,d-1} \) for \( j \in D^- \). Hence, for \( \tilde{\mu}^- = (\mu^-, 0) \) we have

\[
(\zeta(y)\Lambda(\varepsilon_t - \mu))_{d} = \zeta(y)\Lambda^-((\xi^- - \mu^-)_{d-1} + h(y_t)\kappa_{\varepsilon t,d}.
\]

and

\[
\text{sp}(\Sigma(y))_{d} = \zeta(y)^2 \sum_{i=1}^{d-1} \lambda^2_{i,d-1} + \zeta(y)^2 \kappa^2 = \text{sp}(\Sigma(y))_{d-1} + (\zeta(y)\Lambda)_{dd}^2.
\]

Using model assumptions (3) and (6) we obtain

\[
\tilde{Y}_{t,d} = d\delta Y_{t,d} - d\delta y_d
\]

\[
= \delta(-y_1 + \frac{1}{2} \text{sp}(\Sigma^-)(y))_{d-1} + \beta(y) + \sqrt{\delta}(\zeta^-)(y)\Lambda^-((\xi^- - \mu^-))_{d-1} + h(y_t)\kappa_{\varepsilon t,d})
\]

\[
= \delta(-y_1 + \frac{1}{2} \text{sp}(\Sigma(y))_{d} + \beta(y) - \frac{1}{2} (\zeta(y)\Lambda)^2_{dd}) + \sqrt{\delta}(\zeta(y)\Lambda((\xi^- - \mu^-))_{d})
\]

This proves the stochastic representation (8) for the last component under the real-world measure \( P \). The statements on the conditional distribution follow directly from (8). \( \Box \)

### A.3 Proof of Theorem 3.1

We only have to prove the second part. The statement is intuitively clear because \( \varepsilon_{t,d} \) and \( \varepsilon_{t,j} \) are independent, given \( F_{t-\delta} \), for all \( t \in J' \) and \( j \in D^- \). We prove the statement indirectly. Assume that the claim \( c_T = 1 \) is attainable. Let \( X \) be a replicating strategy for \( c_T \). This means that \( X \) is self-financing and \( V_T(X) = 1, P \)-a.s. The no-arbitrage condition
(2) and the self-financing condition (9) imply for all $t \in J'$
\[
E^*[B^{-1}_t V_t(X) | F_{t-\delta}] = B^{-1}_t E^*[V^{-1}_t(X) | F_{t-\delta}] = B^{-1}_t E^*[X_t | R_t | F_{t-\delta}]
\]
\[
= B^{-1}_t X_{t,1} P^{-1}_{t-\delta,1} + B^{-1}_t \sum_{j=2}^d X_{t,j} P^{-1}_{t-\delta,j} E^*[P_{t,j-1} | F_{t-\delta}]
\]
\[
= B^{-1}_{t-\delta} P^{-1}_{t-\delta,1} X_{t,1} P^{-1}_{t-\delta,1} + B^{-1}_t \sum_{j=2}^d X_{t,j} P^{-1}_{t-\delta,j} \frac{B^{-1}_t}{B_{t-\delta}} P_{t-\delta,j}
\]
\[
= B^{-1}_{t-\delta} \sum_{j=1}^d X_{t,j} = B^{-1}_{t-\delta} V_{t-\delta}(X),
\]
i.e. the discounted value process $(B^{-1}_t V_t(X))_{t \in J}$ is a $(\mathbb{P}^*, \mathbb{F})$-martingale. The value of the replicating strategy at time $T - d\delta$ is then given by (see explanation below)
\[
B^{-1}_{T-d\delta} V_{T-d\delta}(X) = E^*[B^{-1}_T V_T(X)|F_{T-d\delta}] = E^*[B^{-1}_T P_{T-d\delta}|F_{T-d\delta}]
\]
\[
= E^*[E^*[B^{-1}_{T-d\delta} P_{T-d\delta,1}|F_{T-d\delta}]] = E^*[E^*[B^{-1}_{T-d\delta} P_{T-d\delta,2}|F_{T-d\delta}]] = \ldots = B^{-1}_{T-d\delta} P_{T-d\delta,d}.
\]
In the first step we use the martingale property. The second step follows from $\mathbb{P}^* \sim \mathbb{P}$ and $V_T(X) = 1$, $\mathbb{P}$-a.s. We then use the tower property of conditional expectations and the no-arbitrage condition (2) iteratively. This result is intuitively clear: at time $T - d\delta$ holding the portfolio $X$ or investing $P(T - d\delta, T)$ in the zero-coupon bond with time to maturity $d\delta$ are both going to generate the same payoff of one unit of cash at maturity. Because of no-arbitrage the value of both strategies at time $T - d\delta$ must be the same. In this proof we set $y = Y_{T-(d\delta+1)d}, \mu = (\mu^-, 0) = (\mu_{T-d\delta, 0})$ and $x = X_{T-d\delta}$. Note that these random variables are $\mathcal{F}_{T-(d\delta+1)d}$-measurable. Using (8), (25) and (26) we have
\[
V_{T-d\delta}(X) = P_{T-d\delta} \exp \{-d\delta Y_{T-d\delta} \} = \exp \{\delta(dy_d - \beta(y))\} k_{d-1}^Y E_{d-1}^Y
\]
where for $j \in D^-$ we set
\[
k_j^- = \exp \left\{ \delta \left( y_j - \frac{1}{2} sp(\Sigma^-)(y)_j \right) + \sqrt{\delta}(\zeta^-(y)A^- \mu^-)_j \right\},
\]
\[
E_j^- = \exp \left\{ -\sqrt{\delta}(\zeta^-(y)A^- \epsilon^-)_j \right\}, E_d = \exp \{-\sqrt{\delta}h(y_d)\}.
\]
Define $k^- = (k_1^-, \ldots, k_j^-)^\prime$ and $E^- = (E_1^-, \ldots, E_d^-)^\prime$. Note that $k^-$ is $\mathcal{F}_{T-(d\delta+1)d}$-measurable, and $E^-$ and $E_d$ are $\mathcal{F}_{T-d\delta}$-measurable. Note that $k^-, E^-, E_d$ have strictly positive components, $\mathbb{P}$-a.s. For all $j \in D^-$ the conditional distribution of $\log E_j^-$ is normal with zero mean and variance $\delta sp(\Sigma^-)(y)_j$. The conditional distribution of $\log E_d$ is also normal with zero mean and variance $\delta h(y_d)^2 \kappa^2$. On the other hand, using the self-financing condition (9) and the model equation (5) we obtain
\[
V_{T-d\delta}(X) = x' R_{T-d\delta} = x_1 \exp\{\delta y_1\} + \sum_{j=1}^{d-1} x_{j+1} \exp\{(j+1)\delta y_{j+1}\} \exp\{-j\delta Y_{T-d\delta, j}\}
\]
\[
= x_1 \exp\{\delta y_1\} + \sum_{j=1}^{d-1} x_{j+1} k_j^- E_j^-.
\]
From the model assumption we know that the components of $\varepsilon_t$ are independent and therefore $E_d$ is independent of $E^{-}$. This independence and the two expressions derived above for $V_{T-d|d}(X)$ imply

$$0 = \exp\{-\delta(dy|d - \beta(y))\} \sum_{j=1}^{d} x_j k^j \Cov[E_j, E_d|\mathcal{F}_{T-1}]$$

$$= \exp\{-\delta(dy|d - \beta(y))\} \Cov[V_{T-d|d}(X) - x_1 \exp\{\delta y_1\}, E_d|\mathcal{F}_{T-1}]$$

$$= \exp\{-\delta(dy|d - \beta(y))\} \Cov[V_{T-d|d}(X), E_d|\mathcal{F}_{T-1}]$$

$$= k^{-1}_d \Cov[E_{d-1} E_d|\mathcal{F}_{T-1}]$$

$$= k^{-1}_d \left( E[E_{d-1} E_d|\mathcal{F}_{T-1}] - \mathbb{E}[E_{d-1} E_d|\mathcal{F}_{T-1}] E[E_{d-1}|\mathcal{F}_{T-1}] \right)$$

$$= k^{-1}_d \left( E[E_{d-1}|\mathcal{F}_{T-1}] (E[E_{d-1}|\mathcal{F}_{T-1}] - E[E_d|\mathcal{F}_{T-1}]^2) \right)$$

$$= k^{-1}_d \left( E[E_{d-1}|\mathcal{F}_{T-1}] \Var[E_d|\mathcal{F}_{T-1}] \right)$$

$$= k^{-1}_d \exp \left\{ \frac{1}{2} \delta \sum(y_{j-1}) \left( \exp\{\delta h(y_{j-1})^2 \kappa^2\} - 1 \right) \exp\{\delta h(y_{j-1})^2 \kappa^2\} \right\} > 0,$$

where the right-hand side of the above equation is strictly larger than zero, given $\mathcal{F}_{T-1}$ (assuming $\kappa \neq 0$ and $h(y_{j-1}) \neq 0$). This is a contradiction and proves the claim.

A.4 Convexity of (13)

The convexity of $\mathcal{D}$ is clear. We consider the objective function in (13). Let $X, Y \in \mathcal{D}$ and $s \in [0, 1]$. We estimate (see explanation below)

$$V_0(sX + (1-s)Y) + \theta(\mathbb{E}[f(V_T(sX + (1-s)Y))])$$

$$= sV_0(X) + (1-s)V_0(Y) + \theta(\mathbb{E}[f(sV_T(X) + (1-s)V_T(Y))])$$

$$\leq sV_0(X) + (1-s)V_0(Y) + \theta(\mathbb{E}[f(V_T(X)) + (1-s)f(V_T(Y))])$$

$$= sV_0(X) + (1-s)V_0(Y) + \theta(s\mathbb{E}[f(V_T(X))] + (1-s)\mathbb{E}[f(V_T(Y))])$$

$$\leq s(V_0(X) + \theta(\mathbb{E}[f(V_T(X))]) + (1-s)(V_0(Y) + \theta(\mathbb{E}[f(V_T(Y))]),$$

where in the first equality we use the linearity of $V_0(\cdot)$ and $V_T(\cdot)$. In the second step we use the convexity of $f$, the monotonicity of the expected value and the fact that $\theta$ is increasing. Finally in the fourth step the convexity of $\theta$ is used. This proves the convexity of the objective function.

A.5 Proof of Lemma 4.3

The given identity follows by straightforward application of the definitions. We show it only for TDF strategies and for the other strategies proceed similarly. Let $t \in J'$ and set $X = X_{\text{TDF}}$. Since the strategies considered are all self-financing we have $V_t(X) = X_{\text{TDF}}$. In the next step we use the linearity of $V_0(\cdot)$ and $V_T(\cdot)$. In the second step we use the convexity of $f$, the monotonicity of the expected value and the fact that $\theta$ is increasing. Finally in the fourth step the convexity of $\theta$ is used. This proves the convexity of the objective function.

\[\square\]
$V_t^-(X) = X_t R_t$ for all $t \in J'$. Iterating the definition of TDF strategies we obtain

$$V_t(X) = X_t R_t = \sum_{j,k=1}^d [p_t \pi_k 1_{1 \leq k \leq d'} + (1 - p_t) \pi_k 1_{d' < k \leq d}] X_t \delta R_{t-d',k} R_{t,j} = \ldots$$

$$= V_0 \sum_{j_1,\ldots,j_d} B_t(j_1,\ldots,j_n) = B_{t-d}(j_1,\ldots,j_n) R_{t,j} = B_{t-d}(j_1,\ldots,j_n) \exp \{-\tilde{Y}_{s,j_s-1}\}.$$

A.6 Proof of Lemma 4.4

Let $t \in J'$, $s = \frac{t}{\delta}$ and $j_1, \ldots, j_n \in D$. For this proof we set $y = Y_{(s-1)\delta} = Y_{t-\delta}$. The statement for $j_s = 1$ follows directly from (15). Assuming $j_s > 1$ and using (15) we have

$$B_t(j_1,\ldots,j_n) = B_{t-\delta}(j_1,\ldots,j_n) R_{t,j} = B_{t-\delta}(j_1,\ldots,j_n) \exp\{-\tilde{Y}_{s,j_s-1}\}.$$

Then, using Lemma 2.2, we have

$$\frac{B_t(j_1,\ldots,j_n)}{B_{t-\delta}(j_1,\ldots,j_n)} = P_{t-\delta,1} \exp \left\{ -\frac{\delta}{2} \mathcal{S}(y))_{j_s-1} + \sqrt{\delta}(\zeta(y) \Lambda(\tilde{\mu}_t - \varepsilon_t))_{j_s-1} \right\}.$$

Taking the logarithm of the right-hand side proves the statement.
Philipp Harms, David Stefanovits.

**Affine representations of fractional processes with applications in mathematical finance.**

AFFINE REPRESENTATIONS OF FRACTIONAL PROCESSES
WITH APPLICATIONS IN MATHEMATICAL FINANCE

PHILIPP HARMS AND DAVID STEFANOVITS

Abstract. Fractional processes have gained popularity in financial modeling due to the dependence structure of their increments and the roughness of their sample paths. The non-Markovianity of these processes gives, however, rise to conceptual and practical difficulties in computation and calibration. To address these issues, we show that a certain class of fractional processes can be represented as linear functionals of an infinite dimensional affine process. We demonstrate by means of several examples that the affine structure allows one to construct tractable financial models with fractional features.

1. Introduction

Empirical evidence suggests that certain financial time series may not be captured well by low-dimensional Markovian models. In particular, this applies to short-term interest rates, which tend to have long-range dependence [3], and to volatilities of stock prices, which have rough sample paths and behave essentially as fractional Brownian motion with small Hurst index [37]. Dependent increments and rough sample paths are, however, characteristic features of fractional processes.

The wide-spread adoption of fractional processes in financial modeling was impeded by several difficulties. Conceptually, one of the major challenges is the lack of the Markov property. In the absence of the Markov property, it is unclear what the states of the model are. This makes it difficult to talk about calibration in a sensible way and to compare the model across time. Moreover, PDE methods for option pricing cannot be used.

In this paper we introduce a class of fractional processes which can be represented as linear functionals of an infinite-dimensional affine process. The key idea, which goes back to Carmona and Coutin [14], is to express the fractional integral in the Mandelbrot-Van Ness representation of fractional Brownian motion by a Laplace transform: for each $H < 1/2$, by the stochastic Fubini’s theorem,

$$
\int_0^t (t-s)^{H-\frac{1}{2}} dW_s \propto \int_0^t \int_0^\infty e^{-x(t-s)} \frac{dx}{x^{H+\frac{1}{2}}} dW_s = \int_0^\infty \int_0^t e^{-x(t-s)} dW_s \frac{dx}{x^{H+\frac{1}{2}}}.
$$

The right-hand side is a superposition of infinitely many Ornstein-Uhlenbeck (OU) processes with varying speed of mean reversion. Extensions and numerical approximations of this representation can be found in Carmona, Coutin, and Montseny [15] and Muravlev [64]. We show that the collection of OU processes, indexed by the speed of mean reversion, is a Banach-space valued affine process. Linear functionals

\textit{2010 Mathematics Subject Classification.} 60G22, 60G15, 60J25, 91G30.

Supported in part by SNF grant 149879 and the ETH Foundation.

We are grateful to Josef Teichmann and Mario Wüthrich for helpful comments and suggestions.
of this process are in general not semimartingales. Instead, they are fractional processes with positively or negatively correlated increments and are closely related to fractional Brownian motion. More precisely, fractional Brownian motion is obtained by randomizing the initial condition of the OU processes according to the stationary distribution.

Our result is relevant in mathematical finance for the following reasons. First, it is within reach to solve some simple fractional models where the affine structure is preserved. We demonstrate this by means of several examples in this paper. In particular, we construct interest rate models where either the short rate or the bank account process is modeled by a fractional process. In contrast to [67] and [7], we build the model such that discounted zero-coupon bond prices are martingales. This implies absence of arbitrage by construction, while certain quantities of the model such as the short rate may very well be non-semimartingales. We also build a fractional version of the stochastic volatility model by Stein and Stein [75].

Second, there is recently a high interest in non-affine fractional volatility models such as the fractional Bergomi and SABR models [62, 37]. It is a major challenge to derive short-time, large-time, and wing asymptotics for these models, as well as to develop numerical schemes for pricing and calibration. Hopefully, the Markovian point of view and the affine structure will be helpful for achieving these goals.

Third, the Markovian structure is useful for characterizing the behavior of fractional Brownian motion after a stopping time. This is crucial for characterizing arbitrage opportunities in models with fractional price processes (c.f. the stickiness property in [40, 23] and the notion of arbitrage times in [69]).

The paper is structured as follows. In Section 2 we prove that the collection of OU processes is indeed a Banach-space valued affine process. In Section 3 we deduce the affine representation of fractional Brownian motion. Section 4 is dedicated to applications in interest rate modeling and Section 5 to a fractional version of the stochastic volatility model of Stein and Stein [75].

2. Infinite-dimensional Ornstein-Uhlenbeck process

2.1. Setup and notation. Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}}, \mathbb{Q})\) be a filtered probability space satisfying the usual conditions and let \(W\) be two-sided \((\mathcal{F}_t)\)-Brownian motion on \(\Omega\). The probability measure \(\mathbb{Q}\) plays the role of a risk-neutral measure.

**Definition 2.1 (OU processes).** Given a collection of \(\mathcal{F}_0\)-measurable \(\mathbb{R}\)-valued random variables \(Y_0^x, Z_0^x\) indexed by \(x \in (0, \infty)\), let for each \(t \geq 0\)

\[
Y_t^x = Y_0^x e^{-tx} + \int_0^t e^{-(t-s)x} dW_s,
\]

\[
Z_t^x = Z_0^x e^{-tx} + \int_0^t e^{-(t-s)x} Y_s^x ds.
\]

**Remark 2.2.** For each \(x \in (0, \infty)\), the process \((Y_t^x, Z_t^x)_{t \geq 0}\) solves the SDE

\[
dY_t^x = -xY_t^x dt + dW_t, \quad dZ_t^x = (-xZ_t^x + Y_t^x)dt.
\]

Therefore, it is a bi-variate OU process, and the variable \(x\) is related to the speed of mean reversion of the process (see Lemma D.1 in the Appendix for details).
2.2. Ornstein-Uhlenbeck process in \( L^1 \). Let \( Y_t = (Y^x_t)_{x > 0} \) and \( Z_t = (Z^x_t)_{x > 0} \) denote the collection of OU processes indexed by the speed of mean reversion \( x \). We show in this section that the process \((Y_t, Z_t)_{t \geq 0}\) takes values in \( L^1(\mu) \times L^1(\nu) \), where the measures \( \mu \) and \( \nu \) are subject to the following conditions.

**Assumption 2.3** (Integrability condition). \( \mu \) and \( \nu \) are sigma-finite measures on \((0, \infty)\) such that \( \nu \) has a density \( p \) with respect to \( \mu \) and for each \( t > 0 \),

\[
\int_0^\infty (1 + x^{-\frac{3}{2}})\mu(dx) < \infty, \quad \int_0^\infty (1 + x^{-\frac{3}{2}})\nu(dx) < \infty, \quad \sup_{x \in (0, \infty)} p(x)e^{-tx} < \infty.
\]

The pairing between \( L^1(\mu) \) and \( L^\infty(\mu) \) is denoted by \( \langle \cdot, \cdot \rangle_\mu \), and similarly for \( L^1(\nu) \) and \( L^\infty(\nu) \). The complexification of these spaces is denoted by \( L^1(\mu; \mathbb{C}) \), etc.

**Theorem 2.4** (OU process in \( L^1 \)). Let \( \mu, \nu \) satisfy Assumption 2.3 and let \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu) \). Then the process \((Y_t, Z_t)_{t \geq 0}\) has a predictable \( L^1(\mu) \times L^1(\nu) \)-valued version and is Gaussian.

**Remark 2.5.** Carmona and Coutin [14] show the weaker statement that for each fixed \( t \geq 0 \), the random variable \( Y_t \) lies a.s. in \( L^1(\mu) \).

**Proof.** It is shown in Lemma D.1 that for each \( x \in (0, \infty) \) the process \((Y^x_t, Z^x_t)_{t \geq 0}\) can be represented as

\[
Y^x_t = Y^x_0e^{-tx} + \int_0^t e^{-(t-s)x}dW_s,
\]

\[
Z^x_t = Z^x_0e^{-tx} + Y^x_0te^{-tx} + \int_0^t (t-s)e^{-(t-s)x}dW_s.
\]

By Assumption 2.3, the deterministic parts in the above representation are \( L^1(\mu) \)- and \( L^1(\nu) \)-valued functions, respectively. Therefore, we can assume without loss of generality that \( Y_0 \) and \( Z_0 \) are zero.

In Lemma D.2 it is shown that for each fixed \( t \geq 0 \), \((Y_t, Z_t) \in L^1(\mu) \times L^1(\nu) \) holds almost surely. Moreover, for any \((u, v) \in L^\infty(\mu) \times L^\infty(\nu) \), the random variables \((Y_t, u)_\mu \) and \((Z_t, v)_\mu \) are centered Gaussian, as shown in Lemma D.3. Let \( P_t : L^\infty(\mu) \rightarrow L^1(\mu) \) and \( Q_t : L^\infty(\mu) \rightarrow L^1(\nu) \) be the associated covariance operators, which are calculated explicitly in Lemma D.4.

We now show that \( Y_t \) is a version of an \( L^1(\mu) \)-valued stochastic convolution. To this aim, let \( H_t \subseteq L^1(\mu) \) be the reproducing kernel Hilbert space of \( P_t \) (see Appendix B). The inclusion of \( H_t \) in \( L^1(\mu) \) is \( \gamma \)-radonifying because \( Y_t \) provides an instance of a Gaussian random variable with covariance operator \( P_t \) [65, Theorem 7.4]. For each \( s > 0 \) define \( \Theta_1(s) \in L^1(\mu) \) and \( \Theta^*_1(s) : L^\infty(\mu) \rightarrow \mathbb{R} \) by

\[
\Theta_1(s)(x) = e^{-sx}, \quad \Theta^*_1(s)(u) = \langle \Theta_1(s), u \rangle_\mu.
\]

Then \( \Theta^*_1 \) satisfies for each \( t \geq 0 \) and any \( u \in L^\infty(\mu) \)

\[
\int_0^t (\Theta^*_1(t-s)(u))^2ds = \int_0^t \left( \int_0^\infty e^{-s(t-s)}u(x)\mu(dx) \right)^2ds = \langle Pu, u \rangle_\mu < \infty,
\]

where the order of integration can be exchanged because condition (A.1) is satisfied by Equation (C.22). By [13, Theorem 3.3], the bound on \( \Theta^*_1 \) and the \( \gamma \)-radonifying property of the inclusion of \( H_t \) in \( L^1(\mu) \) imply that the stochastic convolution of \( \Theta_1 \)
with $W$ exists as an $L^1(\mu)$-valued $(\mathcal{F}_t)_{t \geq 0}$-predictable process $\tilde{Y}$ such that for each $t \geq 0$ and any $u \in L^\infty(\mu)$, \[
abla \theta_1, u\|_{\mu} = \int_0^t \theta_1(t-s)(u)dW_s \]
holds almost surely. The same equation is also satisfied by $Y_t$. As stochastic convolutions are unique up to modifications [13, Theorem 3.3], $\mathbb{Q}(Y_t = \tilde{Y}_t) = 1$ holds for each $t \geq 0$. This proves that $Y$ has a predictable, $L^1(\mu)$-valued version.

We use the same argument to show that $Z$ has a predictable, $L^1(\nu)$-valued version. For each $s > 0$ define $\Theta_2(s) \in L^1(\nu)$ and $\tilde{\Theta}_2(s) : L^\infty(\nu) \to \mathbb{R}$ by \[
abla \theta_2(s)(x) = se^{-sx}, \quad \tilde{\Theta}_2(s)(v) = (\Theta_2(s), v)_\nu. \]

Then $\Theta_2$ satisfies for each $t \geq 0$ and any $v \in L^\infty(\nu)$ \[
abla \int_0^t \left(\tilde{\Theta}_2(t-s)(v)\right)^2 ds = \int_0^t \left(\int_0^{\infty} (t-s)e^{-x(t-s)}v(x)dx\right)^2 ds = (Q_tv, v)_\nu < \infty, \]
where the order of integration can be exchanged because condition (A.1) is satisfied by Equation (C.23). By the same argument as above there exists an $L^1(\nu)$-valued $(\mathcal{F}_t)_{t \geq 0}$-predictable process $\tilde{Z}$ such that for each $t \geq 0$ and any $v \in L^\infty(\nu)$, \[
abla \langle \tilde{Z}_t, v\rangle_\nu = \int_0^t \tilde{\Theta}_2(t-s)(u)dW_s, \]
holds almost surely. As $Z$ satisfies the same equation and stochastic convolutions are unique up to modifications [13, Theorem 3.3], $\tilde{Z}$ is a version of $Z$. \]

2.3. Affine structure. We derive an infinite-dimensional affine transformation formula for the conditional exponential moments of $(Y, u)_\mu$ and $(Z, v)_\nu$ for test functions $u \in L^\infty(\mu; \mathbb{C})$ and $v \in L^\infty(\nu; \mathbb{C})$.

**Theorem 2.6** (Affine structure). Let $\mu, \nu$ satisfy Assumption 2.3 and let $(Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)$. Then the process $(Y, Z)$ is affine in the sense that for each $0 \leq t \leq T$ and $(u, v) \in L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})$, the relation \[
abla \mathbb{E} \left[ e^{(Y_T, u)_\mu + (Z_T, v)_\nu} | \mathcal{F}_t \right] = e^{\phi_0(T-t, u, v) + (Y_t, \phi_1(T-t, u, v))_\mu + (Z_t, \phi_2(T-t, u, v))_\nu} \]
holds with probability one, where the functions \[
abla (\phi_0, \phi_1, \phi_2) : [0, \infty) \times L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C}) \to \mathbb{C} \times L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C}) \]
are given by \[
abla \phi_0(t, u, v) = \frac{1}{2} \int_0^T \left( \int_0^\infty \phi_1(s, u, v)(x)\mu(dx) \right)^2 ds, \]
\[
abla (\phi_1, \phi_2)(t, u, v)(x) = e^{-tx}(u(x) + \tau v(x)p(x)), \]
\[
abla (\phi_2)(t, u, v)(x) = e^{-tx}v(x). \]

**Proof.** Lemma D.3 states that for each $0 \leq t \leq T$, the random variable $(Y_T, u)_\mu + (Z_T, v)_\nu$ is Gaussian, given $\mathcal{F}_t$, with mean \[
abla \int_0^\infty Y_t^x e^{-(T-t)x}u(x)d\mu(dx) + \int_0^\infty \left( Z_t^x e^{-x(T-t)} + Y_t^x(T-t)e^{-x(T-t)} \right) \tau v(x)v(dx) \]

\[
abla = (Y_t, \phi_1(T-t, u, v))_\mu + (Z_t, \phi_2(T-t, u, v))_\nu. \]
By Itô’s isometry, the conditional variance of $(Y_T, u)_\mu + (Z_T, v)_\nu$ given $\mathcal{F}_t$ is

$$\int_t^T \left( \int_0^\infty e^{-(T-s)x} u(x) \mu(dx) + \int_0^\infty (T-s)e^{-x(T-s)}v(x)\nu(dx) \right)^2 \, ds,$$

which equals $2\phi_0(T-t, u, v)$. Thus,

$$\mathbb{E}\left[ e^{(Y_T, u)_\mu + (Z_T, v)_\nu} \big| \mathcal{F}_t \right] = e^{\frac{1}{2} \text{Var}(Y_T, u)_\mu + (Z_T, v)_\nu} \mathbb{E}[e^{(Y_T, u)_\mu + (Z_T, v)_\nu}].$$

The coefficient functions $(\phi_0, \phi_1, \phi_2)$ are solutions of an infinite-dimensional system of Riccati equations. To formulate the equations, we need to introduce some topology. We endow the spaces $L^\infty(\mu; \mathbb{C})$ and $L^\infty(\nu; \mathbb{C})$ with the weak-star topology. Then they are locally convex separable Hausdorff vector spaces. In particular, differentiability of curves with values in these spaces is well-defined.

**Definition 2.7 (Riccati equations).** Mappings $\phi_0, \phi_1, \phi_2$ as in (2.6) are called solutions of the Riccati equations if they are continuous in $t$ on the interval $[0, \infty)$, differentiable in $t$ on the interval $(0, \infty)$, and satisfy

$$\begin{align*}
\partial_t (\phi_0, \phi_1, \phi_2)(\tau, u, v) &= (R_0, R_1, R_2)(\phi_1(\tau, u, v), \phi_2(\tau, u, v)), \\
(\phi_0, \phi_1, \phi_2)(0, u, v) &= (0, u, v),
\end{align*}$$

(2.8)

where the mappings

$$(R_0, R_1, R_2): L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C}) \to \mathbb{C} \times L^0(\mu; \mathbb{C}) \times L^0(\nu; \mathbb{C})$$

are given by

$$\begin{align*}
R_0(u, v) &= \frac{1}{2} \left( \int_0^\infty u(x) \mu(dx) \right)^2, \\
R_1(u, v)(x) &= -ux(x) + p(x)v(x), \\
R_2(u, v)(x) &= -xv(x).
\end{align*}$$

**Lemma 2.8 (Riccati equations).** The functions $(\phi_0, \phi_1, \phi_2)$ defined in Equation (2.7) are the unique solution of the Riccati equations (2.8).

**Proof.** It is straightforward to verify that the functions $(\phi_0, \phi_1, \phi_2)$ given by Equation (2.7) solve the Riccati equations in the sense of Definition 2.7. Let $(\overline{\phi_0}, \overline{\phi_1}, \overline{\phi_2})$ be any other solution. Then $e^{\varepsilon t}(\phi_2 - \overline{\phi_2})$ has vanishing derivative and initial condition, implying that it is constant and $\phi_2 = \overline{\phi_2}$. The same applies to $e^{\varepsilon t}(\phi_1 - \overline{\phi_1})$, showing that $\phi_1 = \overline{\phi_1}$, and to $\phi_0 - \overline{\phi_0}$, showing that $\phi_0 = \overline{\phi_0}$. □

**2.4. Continuity of sample paths.** Under the following conditions on the measures $\mu$ and $\nu$, the process $(Y, Z)$ has continuous sample paths in $L^1(\mu) \times L^1(\nu)$ with respect to the norm topology.

**Assumption 2.9 (Integrability condition).** $\mu$ and $\nu$ are sigma-finite measures on $(0, \infty)$ satisfying

$$\int_0^\infty \log(1 + tx)x^{-\frac{1}{2}} \mu(dx) < \infty, \quad \int_0^\infty \log(1 + tx)x^{-\frac{1}{2}} \nu(dx) < \infty.$$  

Moreover, $\nu$ has a density $p$ with respect to $\mu$, such that for each $t > 0$

$$\sup_{x \in (0, \infty)} p(x)e^{-tx} < \infty.$$
Remark 2.10. Compared to Assumption 2.3, Assumption 2.9 is weaker near zero and stronger near infinity, as can be seen from the limits

\[ \forall t > 0: \lim_{x \to 0^+} \frac{\log(1 + tx)x^{-\frac{3}{2}}}{1 \wedge x^{-\frac{3}{2}}} = 0, \quad \lim_{x \to \infty} \frac{\log(1 + tx)x^{-\frac{3}{2}}}{1 \wedge x^{-\frac{3}{2}}} = \infty. \]

Theorem 2.11 (Continuity of sample paths). Under Assumption 2.9, the process \((Y, Z)\) has continuous sample paths in \(L^1(\mu) \times L^1(\nu)\) if the initial condition \((Y_0, Z_0)\) lies in this space.

Remark 2.12. Note that Theorem 2.11 does not guarantee that \((Y, Z)\) is a Gaussian process in \(L^1(\mu) \times L^1(\nu)\); this follows from Theorem 2.4 under Assumption 2.3.

Proof. The expressions \(Y_t^x e^{-tx}\) and \(Z_t^x e^{-tx} + Y_t^x te^{-tx}\) define continuous \(L^1(\mu)\)- and \(L^1(\nu)\)-valued functions, respectively. Thus, it follows from the representation of \((Y, Z)\) in Equation (2.3) that we may assume \((Y_0, Z_0) = 0\) without loss of generality. By Lemma D.5, and Assumption 2.9 on \(\mu\), integration with respect to \(\mu\) yields

\[ \mathbb{E} \left[ \int_0^\infty \sup_{s \in [0,t]} |Y_t^x| \mu(dx) \right] \leq C \int_0^\infty \log(1 + tx)x^{-\frac{1}{2}} \mu(dx) < \infty, \]

where we are allowed to exchange the order of integration since the integrand is positive. This implies that \(\mathbb{Q}[\forall t: Y_t \in L^1(\mu)] = 1\). Moreover, by the dominated convergence theorem with the sup process of \(Y\) as majorant, \(\mathbb{Q}[Y \in C([0, \infty); L^1(\mu))] = 1\). For the process \(Z\), the estimate of Lemma D.5 and Assumption 2.9 on \(\nu\) show that \(\mathbb{Q}[\forall t: Z_t^x \in L^1(\nu)] = 1\). As before, the dominated convergence theorem with the sup process of \(Z\) as majorant implies \(\mathbb{Q}[Z \in C([0, \infty); L^1(\nu))] = 1\). \(\square\)

2.5. Semimartingale property. In this section we investigate under which conditions linear functionals of the process \((Y, Z)\) are semimartingales. We consider time-dependent linear functionals as this will be needed later in applications.

Theorem 2.13 (Semimartingale property). Let Assumption 2.3 be in place. Let \(f_t^x\) and \(g_t^x\) be real-valued, deterministic, jointly measurable in \((x, t) \in (0, \infty) \times [0, \infty)\), differentiable in \(t\) and satisfy

\[ \forall t \geq 0: \|f_t\|_{L^\infty(\mu)} < \infty \text{ and } \|g_t\|_{L^\infty(\nu)} < \infty. \]

Assume \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)\), a.s., and for each \(t \geq 0\)

\begin{align*}
(2.9) \quad & \int_0^\infty \int_0^t |\partial_s f_s^x - xf_s^x|(1 \wedge x^{-\frac{3}{2}})ds \mu(dx) < \infty, \\
(2.10) \quad & \int_0^\infty \sqrt{\int_0^t (f_s^x)^2} ds \mu(dx) < \infty, \\
(2.11) \quad & \int_0^\infty \int_0^t |\partial_s g_s^x - xg_s^x|(1 \wedge x^{-\frac{3}{2}})dv \mu(dx) < \infty, \\
(2.12) \quad & \int_0^\infty \int_0^t |g_s^x|(1 \wedge x^{-\frac{3}{2}})dv \mu(dx) < \infty. 
\end{align*}
Then \((\langle Y_t, f_t \rangle_\mu)_{t \geq 0}\) and \((\langle Z_t, g_t \rangle_\nu)_{t \geq 0}\) are semimartingales with decompositions
\begin{align}
\langle Y_t, f_t \rangle_\mu &= \langle Y_0, f_0 \rangle_\mu + \int_0^t \int_0^\infty (\partial_s f_s - x f_s) Y_s^x \mu(dx) ds \\
+ & \int_0^t \int_0^\infty f_s^x \mu(dx) dW_s,
\end{align}
(2.13)
\begin{align}
\langle Z_t, g_t \rangle_\nu &= \langle Z_0, g_0 \rangle_\nu + \int_0^t \int_0^\infty (\partial_s g_s - x g_s) Z_s^x \nu(dx) ds \\
+ & \int_0^t \int_0^\infty g_s^x \nu(dx) ds.
\end{align}
Proof. First observe that
\begin{align}
\langle Y_t, f_t \rangle_\mu &= \langle Y_t - Y_0^t e^{-xt}, f_t \rangle_\mu + \langle Y_0^t e^{-xt}, f_t \rangle_\mu, \\
\langle Z_t, g_t \rangle_\nu &= \langle Z_t - Z_0^t e^{-xt} - Y_0^t e^{-xt}, g_t \rangle_\nu + \langle Z_0^t e^{-xt}, g_t \rangle_\nu + \langle Y_0^t e^{-xt}, g_t \rangle_\nu.
\end{align}
Since \(\langle Y_0^t e^{-xt}, f_t \rangle_\mu, \langle Z_0^t e^{-xt}, g_t \rangle_\nu\) are finite variation processes we assume without loss of generality that \(Y_0 = Z_0 = 0\). By SDE (2.2) for \((Y, Z)\) and Itô’s formula, the semimartingale decomposition of the process \((f_t^x Y_t^x, g_t^x Z_t^x)\) is given by
\begin{align}
f_t^x Y_t^x &= \int_0^t (\partial_s f_s^x - x f_s^x) Y_s^x ds + \int_0^t f_s^x dW_s, \\
g_t^x Z_t^x &= \int_0^t (\partial_s g_s^x - x g_s^x) Z_s^x ds + \int_0^t g_s^x dY_s.
\end{align}
Therefore,
\begin{align}
\langle Y_t, f_t \rangle_\mu &= \int_0^\infty \int_0^\infty (\partial_s f_s^x - x f_s^x) Y_s^x ds \mu(dx) + \int_0^\infty \int_0^t f_s^x dW_s \mu(dx), \\
\langle Z_t, g_t \rangle_\nu &= \int_0^\infty \int_0^\infty (\partial_s g_s^x - x g_s^x) Z_s^x ds \nu(dx) + \int_0^\infty \int_0^t g_s^x dY_s ds \nu(dx).
\end{align}
By Theorem A.1 one obtains the semimartingale decompositions of \(\langle Y_t, f_t \rangle_\mu\) and \(\langle Z_t, g_t \rangle_\nu\). By Lemma D.6 and Equations (2.9)-(2.12) conditions (A.1) and (A.2) are satisfied.

2.6. Stationary distribution. We show that the stationary distribution of \((Y, Z)\) in general not a Gaussian distribution on \(L^1(\mu) \times L^1(\nu)\), but only on a larger space \(L^1(\mu_\infty) \times L^1(\nu_\infty)\) corresponding to stronger integrability conditions on the measures \(\mu_\infty\) and \(\nu_\infty\).

Assumption 2.14 (Integrability condition). \(\mu_\infty, \nu_\infty\) are sigma-finite measures on (0, \(\infty\)) such that \(\nu_\infty\) has a density \(p_\infty\) with respect to \(\mu_\infty\) and
\begin{align}
\int_0^\infty x^{-1/2} \mu_\infty(dx) < \infty, \\
\int_0^\infty x^{-3/2} \nu_\infty(dx) < \infty, \\
\sup_{x \in (0, \infty)} p_\infty(x)e^{-tx} < \infty.
\end{align}

Remark 2.15. Assumption 2.14 is more stringent than Assumption 2.3. The difference is the decay of the measures near zero: \(\mu, \nu\) satisfy Assumption 2.3 if and only if the measures
\begin{align}
\mu_\infty(dx) &= (1 \land x^{1/2}) \mu(dx), \\
\nu_\infty(dx) &= (1 \land x^{1/2}) \nu(dx)
\end{align}
satisfy Assumption 2.14. In this case, \(L^1(\mu) \times L^1(\nu) \subset L^1(\mu_\infty) \times L^1(\nu_\infty)\).
Theorem 2.16 (Stationary distribution). The random variables $Y_{\infty} = (Y_x)_{x \geq 0}$ and $Z_{\infty} = (Z_x)_{x > 0}$ defined by
\begin{align}
Y_x^\infty &= \int_{-\infty}^{0} e^{sx} dW_s, \\
Z_x^\infty &= -\int_{-\infty}^{0} se^{sx} dW_s
\end{align}
are normally distributed on $L^1(\mu_{\infty}) \times L^1(\nu_{\infty})$. Their distribution is stationary in the sense that $(Y_t, Z_t)$ is equal in distribution to $(Y_{\infty}, Z_{\infty})$ if $(Y_0, Z_0)$ is equal in distribution to $(Y_{\infty}, Z_{\infty})$.

Proof. $(Y_{\infty}, Z_{\infty}) \in L^1(\mu_{\infty}) \times L^1(\nu_{\infty})$ holds almost surely because
\begin{align}
E \left[ \|Y_\infty\|_{L^1(\mu_{\infty})} \right] &= \int_0^{\infty} E \left[ \int_{-\infty}^{0} e^{sx} dW_s \right] \mu_{\infty}(dx) = \int_0^{\infty} \sqrt{\frac{1}{2\pi x}} \mu_{\infty}(dx) < \infty, \\
E \left[ \|Z_\infty\|_{L^1(\nu_{\infty})} \right] &= \int_0^{\infty} E \left[ \int_{-\infty}^{0} se^{sx} dW_s \right] \nu_{\infty}(dx) = \int_0^{\infty} \sqrt{\frac{1}{4\pi x^3}} \nu_{\infty}(dx) < \infty.
\end{align}
For each $u, v \in L^\infty(\mu_{\infty}) \times L^\infty(\nu_{\infty})$, the random variable $\langle Y_{\infty}, u \rangle_{\mu_{\infty}} + \langle Z_{\infty}, v \rangle_{\nu_{\infty}}$ can be expressed by Fubini (Theorem A.1) as
\begin{align}
\langle Y_{\infty}, u \rangle_{\mu_{\infty}} + \langle Z_{\infty}, v \rangle_{\nu_{\infty}} &= \int_0^{\infty} \int_{-\infty}^{\infty} e^{sx} u(x) \mu_{\infty}(dx) dW_s \\
&\quad + \int_{-\infty}^{\infty} e^{sx} v(x) \nu_{\infty}(dx) dW_s.
\end{align}
Condition (A.2) of Fubini’s theorem is satisfied because
\begin{align}
\int_0^{\infty} \sqrt{\int_{-\infty}^{0} e^{sx} u(x)^2 ds} \mu_{\infty}(dx) &\leq \|u\|_{L^\infty(\mu_{\infty})} \int_0^{\infty} \sqrt{\frac{1}{2\pi x}} \mu_{\infty}(dx) < \infty, \\
\int_0^{\infty} \sqrt{\int_{-\infty}^{0} s e^{sx} v(x)^2 ds} \nu_{\infty}(dx) &\leq \|v\|_{L^\infty(\nu_{\infty})} \int_0^{\infty} \sqrt{\frac{1}{4\pi x^3}} \nu_{\infty}(dx) < \infty.
\end{align}
Therefore, $(Y_{\infty}, u)_{\mu_{\infty}} + (Z_{\infty}, v)_{\nu_{\infty}}$ is a centered Gaussian random variable on $L^1(\mu_{\infty}) \times L^1(\nu_{\infty})$. To show that the distribution of $(Y_{\infty}, Z_{\infty})$ is stationary, let us assume that $(Y_0, Z_0) = (Y_{\infty}, Z_{\infty})$. Then Lemma D.1 implies
\begin{align}
Y_t^\infty &= \int_{-\infty}^{t} e^{-(t-s)^2} dW_s, \\
Z_t^\infty &= \int_{-\infty}^{t} (t-s) e^{-(t-s)^2} dW_s,
\end{align}
which is equal in distribution to $Y_{\infty}$ and $Z_{\infty}$, respectively.

Theorem 2.17 (Convergence to the stationary distribution). For any initial condition $(Y_0, Z_0) \in L^1(\mu_{\infty}) \times L^1(\nu_{\infty})$ and any $t \geq 0$, we consider $(Y_t, Z_t)$ as a random variable with values in the space $L^1(\mu_{\infty}) \times L^1(\nu_{\infty})$, which we endow with the weak topology. Then $(Y_t, Z_t)$ converges in distribution to $(Y_{\infty}, Z_{\infty})$ as $t \to \infty$.

Proof. Let $(u, v) \in L^\infty(\mu_{\infty}) \times L^\infty(\nu_{\infty})$. By Equation (2.16) and Itô’s isometry the variance of the centered Gaussian random variable $\langle Y_{\infty}, u \rangle_{\mu_{\infty}} + \langle Z_{\infty}, v \rangle_{\nu_{\infty}}$ is
\begin{align}
E \left[ (\langle Y_{\infty}, u \rangle_{\mu_{\infty}} + \langle Z_{\infty}, v \rangle_{\nu_{\infty}})^2 \right] &= \int_{-\infty}^{\infty} \left( \int_0^{\infty} e^{sx} (u(x) + sv(x)p_{\infty}(x)) \mu_{\infty}(dx) \right)^2 ds.
\end{align}
Assume for a moment that \((Y_0, Z_0) = 0\). As the measures \(\mu_\infty\) and \(\nu_\infty\) satisfy the conditions of Theorem 2.6,
\[
\lim_{t \to \infty} \mathbb{E} \left[ e^{(Y_t, u)_\mu+ (Z_t, v)_\nu} \right] = \lim_{t \to \infty} e^{\int_0^\infty \left( \int e^{-t\gamma} (u(x) + sv(x)p_\infty(x)) \mu_\infty(dx) \right)^2 \, ds}
\]
\[
= e^{\frac{1}{2} \int_0^\infty (\int e^{-\gamma} (u(x) + sv(x)p_\infty(x)) \mu_\infty(dx)^2 \, ds}
\]
\[
= e^{\frac{1}{2} \text{Var}((Y_\infty, u), (Z_\infty, v))}
\]
\[
= \mathbb{E} \left[ e^{(Y_\infty, u)_\mu+ (Z_\infty, v)_\nu} \right].
\]
This shows point-wise convergence of the characteristic functions of \((Y_t, Z_t)\) to the characteristic functions of \((Y_\infty, Z_\infty)\). By Lemma D.7 the laws of the random variables \((Y_t, Z_t)\) are tight on the space \(\mathcal{L}^1(\mu_\infty) \times \mathcal{L}^1(\nu_\infty)\) with the weak topology. It follows that \((Y_t, Z_t)\) converges in distribution on \(\mathcal{L}^1(\mu_\infty) \times \mathcal{L}^1(\nu_\infty)\) to \((Y_\infty, Z_\infty)\) (see e.g. [31, Theorem 9]).

To account for arbitrary initial conditions \((Y_0, Z_0) \in \mathcal{L}^1(\mu_\infty) \times \mathcal{L}^1(\nu_\infty)\), we need to add the deterministic functions \(Y_0^z e^{-tx}\) and \(Z_0^z e^{-tx}\) to the processes \(Y_t^z\) and \(Z_t^z\) considered above (see Lemma D.1). For \(t \to \infty\), these functions converge to zero in the corresponding \(\mathcal{L}^1\) spaces. It follows that convergence in distribution to \((Y_\infty, Z_\infty)\) holds regardless of the initial condition.

\[\square\]

2.7. Ornstein-Uhlenbeck process with values in \(L^2\). We defined \((Y, Z)\) as an \(L^1\)-valued process because the construction of fractional Brownian motion in Section 3 involves a pairing of \((Y, Z)\) with the constant function 1. Nevertheless, it is good to know that \((Y, Z)\) can also be understood as an \(L^2\)-valued process.

**Assumption 2.18** (Integrability condition). \(\mu\) and \(\nu\) are sigma-finite measures on \((0, \infty)\) such that \(\nu\) has a density \(p\) with respect to \(\mu\) and for each \(t > 0\),
\[
\int_0^\infty (1 \wedge x^{-1}) \mu(dx) < \infty, \quad \int_0^\infty (1 \wedge x^{-3}) \nu(dx) < \infty, \quad \sup_{x \in (0, \infty)} e^{-tx} p(x) < \infty.
\]

**Theorem 2.19** (OU process in \(L^2\)). Let \(\mu, \nu\) satisfy Assumption 2.18 and let \((Y_0, Z_0) \in \mathcal{L}^2(\mu) \times \mathcal{L}^2(\nu)\). Then the process \((Y_t, Z_t)\) has a predictable \(L^2(\mu) \times L^2(\nu)\)-valued version and is a Gaussian affine process on \(L^2(\mu) \times L^2(\nu)\).

The theorem can be proven along the lines of Theorems 2.4 and 2.6. Here we present an alternative proof, which uses the theory of Hilbert-space valued stochastic convolutions.

**Proof.** We want to construct the stochastic convolutions in Equation (2.3) as \(L^2(\mu)\)- and \(L^2(\nu)\)-valued processes, respectively. The setting of [24, Sections 5.1.1–5.1.2] does not apply directly because the integrand in the expression \(\int_0^t e^{-t\gamma} x dW_x\) cannot be written as a strongly continuous semigroup acting on \(L^2(\mu)\). Nevertheless, we can adapt the arguments of [24, Theorem 5.2 and Proposition 3.6] to our setting. Let
\[
S_t : \mathbb{R} \to L^2(\mu), \quad \lambda \mapsto (x \mapsto e^{-t\gamma} \lambda).
\]

183
Then the $L^2(\mu)$-valued convolution $\int_0^t S_{t-s}dW_s$ exists by [24, Theorem 5.2] because
\[
\int_0^t \|S_t\|^2_{HS(\mathbb{R},L^2(\mu))} ds = \int_0^t \|S_t1\|^2_{L^2(\mu)} ds = \int_0^t \int_0^\infty e^{-2sx} \mu(dx) ds = \int_0^\infty \left(1 - e^{-2xt}\right) \mu(dx) < \infty
\]
by Equation (C.4) and Assumption 2.18. It is mean-square continuous by the same arguments as in the proof of [24, Theorem 5.2]. Therefore, it is predictable [24, Proposition 3.6]. Similarly, it can be shown that $Z$ has a predictable, $L^2(\nu)$-valued version. The affine structure can be derived as in Section 2.3. □

**Assumption 2.20** (Integrability condition). $\mu$ and $\nu$ are sigma-finite measures on $(0, \infty)$ such that $\nu$ has a density $p$ with respect to $\mu$. There is $\epsilon \in (0,1)$ such that for each $t > 0$,
\[
\int_0^\infty (1 \wedge x^{-1+\epsilon})\mu(dx) < \infty, \quad \int_0^\infty (1 \wedge x^{-3+\epsilon})\nu(dx) < \infty, \quad \sup_{x \in (0,\infty)} e^{-tx} p(x) < \infty.
\]

**Theorem 2.21** (Continuity of sample paths). Under Assumption 2.20, the process $(Y, Z)$ has continuous sample paths in $L^2(\mu) \times L^2(\nu)$ if the initial condition $(Y_0, Z_0)$ lies in this space.

**Proof.** Let $S$ be as in the proof of Theorem 2.19. Then the estimate
\[
\int_0^t s^{-\epsilon} \|S_t\|^2_{HS(\mathbb{R},L^2(\mu))} ds = \int_0^t \int_0^\infty s^{-\epsilon} e^{-2sx} \mu(dx) ds 
\leq \int_0^\infty \left(2^{\epsilon-1} \Gamma(1-\epsilon) \vee \frac{1^{1-\epsilon}}{1-\epsilon}\right) (1 \wedge x^{-1-\epsilon}) \mu(dx) < \infty
\]
holds by Equation (C.7) for $\epsilon \in (0,1)$ as in Assumption 2.20. Therefore, [24, Theorem 5.11] may be applied, showing that $Y$ has continuous sample paths in $L^2(\mu)$. (While the stochastic convolution $Y$ is not covered by the setting of [24, Section 5.1.1–5.1.2], the same arguments as in the proof of Theorem 2.19 show that [24, Theorem 5.11] holds.) Similarly, it may be shown that the process $Z$ given by Equation (2.3) has continuous sample paths in $L^2(\nu)$. □

**2.8. Smoothness in the spatial dimension.** We show in the following theorem that $(Y^x_t, Z^x_t)$ varies smoothly in $x$. To this aim, we extend Definition 2.1 of $(Y^x_t, Z^x_t)$ to $x \leq 0$ in the obvious way. The space $C^k(\mathbb{R})$, $k \in \mathbb{N} \cup \{\infty\}$, is the Fréchet space with the topology of uniform convergence of derivatives up to order $k$ on compact sets.

**Theorem 2.22** (Smoothness in the spacial dimension). For each $k \in \mathbb{N} \cup \{\infty\}$ and initial value $(Y_0^x, Z_0^x) \in C^k(\mathbb{R}) \times C^k(\mathbb{R})$, the process $(Y, Z)$ is a Gaussian process on $C^k(\mathbb{R})^2$ with continuous sample paths.

**Proof.** The deterministic parts in Equation (2.3) are smooth in $t$ and $x$. We set them to zero by assuming without loss of generality that $(Y_0, Z_0) = 0$. By partial integration, the stochastic integrals in Equation (2.3) can be transformed into
Lebesgue integrals:

\[ Y_t^x = W_t - \int_0^t W_s x e^{-(t-s)x} ds, \]

\[ Z_t^x = \int_0^t W_s \left( e^{-(t-s)x} - (t-s) x e^{-(t-s)x} \right) ds. \]

The integrands, seen as functions of \((s, t)\), are continuous with values in \(C^\infty(\mathbb{R})\). This shows that \((Y_t, Z_t)_{t \geq 0}\) has continuous sample paths in \(C^\infty(\mathbb{R})^2\). The \(k\)-th spatial derivative, expressed as a stochastic integral, is given by

\[ \partial^k_x Y_t^x = \int_0^t (s-t)^k e^{-(t-s)x} dW_s, \quad \partial^k_x Z_t^x = -\int_0^t (s-t)^{k+1} e^{-(t-s)x} dW_s. \]

To show that \((Y, Z)\) is a Gaussian process, it suffices to test with linear functionals on \(C^k([-K, K])\) for \(K \in \mathbb{N}\). By the Riesz representation theorem, the dual of \(C^k([-K, K])\) is \(\mathbb{R}^k \times \mathcal{M}([-K, K])\), where \(\mathcal{M}\) stands for the space of signed regular Borel measures endowed with the total variation norm [31, IV.13.36]. The pairing of \(Y_t \in C^k([-K, K])\) with an element \((m, \mu)\) of the dual space \(\mathbb{R}^k \times \mathcal{M}([-K, K])\) reads as

\[ \langle Y_t, (m, \mu) \rangle = \sum_{j=0}^{k-1} m_j \partial_{x^j}|_{x=0} Y_t^x + \int_{-K}^{K} \partial^j_x Y_t^x \mu(dx) \]

\[ = \sum_{j=0}^{k-1} m_j \int_0^t (s-t)^j dW_s + \int_{-K}^{K} \int_0^t (s-t)^j e^{-(t-s)x} dW_s \mu(dx). \]

By the stochastic Fubini theorem (Theorem A.1), the order of the integrals in the last expression can be exchanged. The assumptions of Theorem A.1 are satisfied because \(\mu\) is a finite measure and the integrand is bounded. This shows that \((Y_t, (m, \mu))\) is Gaussian. As \((m, \mu)\) was arbitrary, \(Y_t\) is Gaussian on \(C^k([-K, K])\), for each fixed \(t\). A similar argument shows that \(Z_t\) is Gaussian on the same space. \(\square\)

3. Fractional Brownian motion as a functional of a Markov process

The goal in this section is to obtain a Markovian representation of fractional Brownian motion (fBM) in terms of \((Y, Z)\). We use the representation of Mandelbrot and Van Ness [61] to define fBM.

**Definition 3.1** (fBM). Fractional Brownian motion \(W^H\) with initial value \(w_0^H \in \mathbb{R}\) and Hurst index \(H \in (0, 1)\) is defined for each \(t \geq 0\) as

\[ W_t^H = w_0^H + \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^{t} \left( (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) dW_s, \]

where \(W = (W_t)_{t \in \mathbb{R}}\) is two-sided Brownian motion as defined in Section 2.1.

3.1. Markovian representation of fBM on \(L^1\)-spaces.

**Definition 3.2** (Markovian representation). Let \((Y, Z)\) be the process in Definition 2.1 with initial value \((Y_0, Z_0)\) equal to the random variable \((Y_\infty, Z_\infty)\) defined in Equation (2.15). Furthermore, let \(\mu, \nu\) be measures on \((0, \infty)\) given by

\[ \mu(dx) = \frac{dx}{x^{\frac{1}{2}+H} \Gamma(H + \frac{1}{2}) \Gamma(H - \frac{1}{2})}, \quad \nu(dx) = \frac{dx}{x^{\frac{H}{2}-\frac{1}{2}} \Gamma(\frac{1}{2} + H) \Gamma(\frac{3}{2} - H)}. \]
Remark 3.3. The measures $\mu, \nu$ in the definition above satisfy Assumption 2.3, but not Assumption 2.14. It follows by Theorem 2.11 that $(Y, Z)$ has continuous paths in $L^1(\mu_\infty) \times L^1(\nu_\infty)$ with $(\mu_\infty, \nu_\infty)$ as in Equation (2.14), but not necessarily in $L^1(\mu) \times L^1(\nu)$. Nevertheless, $(Y - Y_0, Z - Z_0)$ has continuous paths in $L^1(\mu) \times L^1(\nu)$, as shown in the proof of Theorem 3.4.

Theorem 3.4 (Markovian representation). Under the specification of Definition 3.2, fBM has representation

\[
W_t^H = \begin{cases} 
  w_0^H + \int_0^\infty (Y_t^x - Y_0^x) \mu(dx), & \text{if } H < \frac{1}{2}, \\
  w_0^H + \int_0^\infty (Z_t^x - Z_0^x) \nu(dx), & \text{if } H > \frac{1}{2}, 
\end{cases}
\]

where $(Y - Y_0, Z - Z_0)$ is a continuous process in $L^1(\mu) \times L^1(\nu)$.

Remark 3.5. The fractional integral in Definition 3.1 can be decomposed as

\[
W_t^H = w_0^H + \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^0 \left( (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right) dW_s + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s.
\]

Markovian representations of the integral appearing in the definition of $W_t^H$ were found by Carmona and Coutin [14] for $H < 1/2$ and by Carmona, Coutin, and Montseny [15] for $H > 1/2$. Muravlev [64] incorporated also the integral $\int_0^t$ in his representation and interpreted it as a random initial value. Moreover, in contrast to [15], his representation is time-homogeneous also in the case $H > 1/2$. Our representation can be seen as a modification of [64] which allows us to identify an infinite-dimensional state space for the Markov process (c.f. Section 2).

Proof of Theorem 3.4 for $H < \frac{1}{2}$. The function $\tau \mapsto \frac{\tau^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})}$ on $(0, \infty)$ appearing in the definition of $W_t^H$ is the Laplace transform of $\mu$, i.e., for each $\tau > 0$ and $H < \frac{1}{2}$

\[
\mathcal{L}(\mu)(\tau) = \int_0^\infty e^{-\tau x} \mu(dx) = \frac{\tau^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})}.
\]

Therefore,

\[
W_t^H = w_0^H + \int_0^t \int_{-\infty}^\infty \left( e^{-x(t-s)} - e^{-x(-s)} \right) \mu(dx) dW_s + \int_0^t \int_0^\infty e^{-x(t-s)} \mu(dx) dW_s.
\]

By the stochastic Fubini's theorem A.1,

\[
W_t^H = w_0^H + \int_0^\infty \int_{-\infty}^0 \left( e^{-x(t-s)} - e^{-x(-s)} \right) dW_s \mu(dx) + \int_0^\infty \int_0^t e^{-x(t-s)} dW_s \mu(dx).
\]
Condition (A.2) of Fubini’s theorem is satisfied because
\[
\int_0^\infty \sqrt{\int_{\mathbb{R}} (e^{-x(t-s)} - e^{-x(-s)})^2 \, ds} \mu(dx) = \int_0^\infty \frac{1 - e^{-tx}}{\sqrt{2x}} \mu(dx) \\
\leq \int_0^\infty \sqrt{\frac{1 - e^{-tx}}{x}} \mu(dx) < \infty,
\]
where we use \(1 - e^{-tx} \leq \sqrt{1 - e^{-tx}}\) and Equation (C.12). By the definition of \(Y^x_t\),

\[
W^H_t = w^H_0 + \int_0^t \left( e^{-xt} - 1 \right) Y^x_0 \mu(dx) + \int_0^t \int_0^\infty e^{-x(t-s)} dW_s \mu(dx)
\]

\[= w^H_0 + \int_0^t (Y^x_t - Y^x_s) \mu(dx) .\]

The expressions

\[
(e^{-xt} - 1) Y^x_0 \quad \text{and} \quad \int_0^t e^{-x(t-s)} dW_s
\]

define continuous \(L^1(\mu)\)-valued processes: the first expression has majorant \((1 \lor t)(1 \wedge x) Y^x_0\) in \(L^1(\mu)\), which allows one to apply the dominated convergence theorem, and the second expression is treated in Theorem 2.11. \(\square\)

**Proof of Theorem 3.4 for** \(H > \frac{1}{2}\). As the function \(\tau^{H-\frac{1}{2}} / \Gamma(H + \frac{1}{2})\) is the Laplace transform of the measure \(\nu\), the relation

\[
\tau \mathcal{L}(\nu)(\tau) = \tau \int_0^\infty e^{-xt} \nu(dx) = \frac{\tau^{H-\frac{1}{2}}}{\Gamma(H + \frac{1}{2})},
\]

holds for each \(\tau > 0\) and \(H \in (\frac{1}{2}, 1)\). Therefore,

\[
W^H_t = w^H_0 + \int_0^t \int_0^\infty \left( (t-s)e^{-x(t-s)} + se^{xs} \right) \nu(dx) dW_s + \int_0^t \int_0^\infty (t-s)e^{-x(t-s)} \nu(dx) dW_s.
\]

By the stochastic Fubini’s theorem A.1,

\[
W^H_t = w^H_0 + \int_0^\infty \int_0^t \left( (t-s)e^{-x(t-s)} + se^{xs} \right) dW_s \nu(dx) \]

\[+ \int_0^\infty \int_0^t (t-s)e^{-x(t-s)} dW_s \nu(dx).
\]
Condition (A.2) of Fubini’s theorem is satisfied because
\[
\int_0^\infty \sqrt{\int_0^\infty ((t-s)e^{-x(t-s)} + se^{xs})^2} \, ds \, d\nu(dx)
\]
\[
= \int_0^\infty \frac{1-2e^{-tx}(tx+1) + 2txe^{-2tx}(tx+1) + e^{-2tx}}{4x^3} \, d\nu(dx)
\]
\[
\leq \int_0^{1/t} \sqrt{\frac{2}{6x}} \, d\nu(dx) + \int_{1/t}^\infty \sqrt{\frac{2}{3x}} \, d\nu(dx)
\]
\[
\leq \sqrt{2}(t \wedge 1) \int_0^\infty (x^{-\frac{1}{2}} \wedge x^{-\frac{3}{2}}) \, d\nu(dx) < \infty,
\]
where we used Equations (C.13) and (C.14). Using the definition of \((Y^x, Z^x)\) in Equation (2.1), Equation (3.2) can be expressed as
\[
W^H_t = w^H_0 + \int_0^t \int_{-\infty}^0 e^{xs} \left(te^{-xt} + s(1 - e^{-xt})\right) \, dW_s \rho(dx)
\]
\[
+ \int_0^t \int_{-\infty}^0 (t - s)e^{-x(t-s)} \, dW_s \rho(dx)
\]
\[
= w^H_0 + \int_0^t \left(te^{-xt} \int_{-\infty}^0 e^{sx} \, dW_s + (1 - e^{-xt}) \int_{-\infty}^0 se^{sx} \, dW_s\right) \, d\rho(dx)
\]
\[
+ \int_0^t \left(Z^x_t - Z^x_0 e^{-xt} - Y^x_0 te^{-xt}\right) \, d\rho(dx)
\]
\[
= w^H_0 + \int_0^t \left(Z^x_t - Z^x_0\right) \, d\rho(dx).
\]
By Lemma D.1, \(Z^x_t - Z^x_0\) can be written as the sum of the following expressions:
\[
Z^x_0(e^{-tx} - 1), \quad Y^x_0 te^{-tx}, \quad \int_0^t (t-s)e^{-(t-s)x} \, dW_s.
\]
All three expressions define continuous \(L^1(\rho)\)-valued processes: the first and second expression have \(|Z^x_0(1\vee t)(1\wedge x)|\) and \(|Y^x_0(1\vee t)(1\wedge x)|\) as majorants in \(L^1(\rho)\), which allows one to apply the dominated convergence theorem, and the third expression is treated in Theorem 2.11.

**Remark 3.6.** The representation in Theorem 3.4 lends itself to numerical implementation. Indeed, the integrals can be approximated by finite sums as described in [15]. Alternatively, aiming for a more parsimonious representation, one has in the case \(H > 1/2\)
\[
W^H_t = w^H_0 - \int_0^t \partial_x (Y^x_t - Y^x_0) \, d\rho(dx).
\]
This follows from the following deterministic relationship between \(Y\) and \(Z\) (c.f. Theorem 2.22)
\[
Z^x_t = -\partial_x Y^x_t + (\partial_x Y^x_0 + Z^x_0)e^{-tx}, \quad t \geq 0.
\]
Remark 3.7. The case $H = 1/2$ fits into the framework of Theorem 3.4 with $\mu$ equal to the Dirac measure. Indeed, the process $(Y_t^0 - Y_0^0)_{t \geq 0}$ is Brownian motion, as can be seen from the definition of $Y$. Moreover, the choice of $\mu$ as a Dirac measure is in line with the proof of Theorem 3.4 where $\mu$ is defined as the inverse Laplace transform of the integrand in Definition 3.1. Note that the representing Markov process $Y_t \in L^1(\mu)$ is one-dimensional and can be identified with Brownian motion.

3.2. Filtrations. The filtration generated by $W^H$ is essentially the same as the one generated by $(Y, Z)$, as shown in the following lemmas. Therefore, the law of fractional Brownian motion after a stopping time can be characterized using the strong Markov property of $(Y, Z)$. This is important for understanding the existence of arbitrage opportunities in models with fractional price processes (see, e.g. the stickiness property in [40, 23] and the notion of arbitrage times in [69]).

Lemma 3.8 (Filtrations). Let $H < 1/2$. Then the completed filtrations generated by the processes $W - W_0$, $W^H - W_0^H$, and $Y - Y_0$ are equal. The same statement holds for $H > 1/2$ with $Y$ replaced by $Z$.

Proof. Let $\mathcal{N}$ denote the $\mathbb{Q}$-null sets. Then the following sigma algebras are equal for each $T \geq 0$:

$$
\sigma(W_t - W_0, 0 \leq t \leq T) \vee \mathcal{N} = \sigma(W_t^H - W_0^H, 0 \leq t \leq T) \vee \mathcal{N}
$$

$$
\subseteq \sigma(Y_t - Y_0, 0 \leq t \leq T) \vee \mathcal{N}
$$

$$
\subseteq \sigma(W_t - W_0, 0 \leq t \leq T) \vee \mathcal{N}.
$$

The first equality above follows from [71, Proposition 1]. The proof for $H > 1/2$ is similar. \qed

From a Markovian point of view, the canonical definition of fractional Brownian motion is $V_t^H = (Y_t, 1)_\mu$ or $V_t^H = (Z_t, 1)_\nu$, depending on whether $H$ is smaller or greater than $1/2$. Here the initial value $(Y_0, Z_0)$ is fixed and deterministic. Moreover, the initial value $W_0$ can be normalized to zero. Then the following lemma holds.

Lemma 3.9 (Filtrations). If $H > 1/2$, then the completed filtrations generated by the processes $W$, $V^H$, and $Y$ are equal. The same statement holds for $H > 1/2$ with $Y$ replaced by $Z$.

Proof. As before, $\mathcal{N}$ denotes the $\mathbb{Q}$-null sets. Let us assume for a moment that the initial value $(Y_0, Z_0)$ is zero. Then one has for each $T \geq 0$

$$
\sigma(W_t, 0 \leq t \leq T) \vee \mathcal{N} = \sigma(V_t^H, 0 \leq t \leq T) \vee \mathcal{N}
$$

$$
\subseteq \sigma(Y_t, 0 \leq t \leq T) \vee \mathcal{N}
$$

$$
\subseteq \sigma(W_t, 0 \leq t \leq T) \vee \mathcal{N}.
$$

The first equality above follows from [71, Proposition 1] applied to a Brownian path which is set to zero for all $t \leq 0$, noting that the relevant integrals are defined pathwise. To get rid of the assumption on $(Y_0, Z_0)$, note that the process $(Y, Z)$ depends on the initial condition $(Y_0, Z_0)$ only via a deterministic function, which is $\mathcal{N}$-measurable. The proof for $H > 1/2$ is similar. \qed

3.3. Markovian representation of fBM on $L^2$-spaces. There is also an $L^2$-version of the results of Section 3.1.
Definition 3.10 (Markovian representation). Let \((Y, Z)\) be the process in Definition 2.1 with initial value \((Y_0, Z_0)\) equal to the random variable \((Y_{\infty}, Z_{\infty})\) defined in Equation (2.15). Furthermore, let \(f(x) = 1 \wedge x^{-1/2}\) and
\[
\mu(dx) = \frac{dx}{f(x)x^{1+H}\Gamma(H + \frac{1}{2})\Gamma(\frac{1}{2} - H)}, \quad \nu(dx) = \frac{dx}{f(x)x^{H-\frac{1}{2}}\Gamma(H)\Gamma(\frac{3}{2} - H)}.
\]

Remark 3.11. The measures \(\mu\) and \(\nu\) in the definition above satisfy Assumptions 2.18 and 2.20, but \((Y_{\infty}, Z_{\infty})\) does not take values in \(L^2(\mu) \times L^2(\nu)\) (c.f. Remark 3.3). Nevertheless, the process \((Y - Y_0, Z - Z_0)\) does, as the following theorem shows.

Theorem 3.12 (Markovian representation). Under the specification of Definition 3.10, fBM has representation
\[
W_t^H = \begin{cases} 
   u_t^H + \int_0^\infty (Y_t - Y_0) f(x) \mu(dx), & \text{if } H < \frac{1}{2}; \\
   u_t^H + \int_0^\infty (Z_t - Z_0) f(x) \nu(dx), & \text{if } H > \frac{1}{2}; 
\end{cases}
\]
where \((Y - Y_0, Z - Z_0)\) is a continuous \(L^2(\mu) \times L^2(\nu)\)-valued process and \(f \in L^2(\mu) \cap L^2(\nu)\).

This can be shown along the lines of the proof of Theorem 3.4.

4. Applications to interest rate modeling

In this section we construct two interest rate models: one with fractional short rate and another one with fractional bank account process. In both models, the affine structure gives rise to explicit formulas for zero-coupon bond (ZCB) prices, forward rates, and calls and puts on ZCB’s.

4.1. Essentials of interest rate modeling. We refer to [33] for further details.

Definition 4.1 (Interest rates). The bank account is given by a positive process \(B = (B_t)_{t \geq 0}\) such that \(B_{t}^{-1}\) is integrable for all \(t \geq 0\). Zero-coupon bond (ZCB) prices are given by
\[
P(t, T) = \mathbb{E}\left[ \frac{B_T}{B_t} \middle| \mathcal{F}_t \right], \quad T \geq t \geq 0,
\]
and the (instantaneous) forward rates are given by
\[
h(t)(\tau) = -\frac{\partial}{\partial \tau}|_{\tau=t} \log P(t, T), \quad t, \tau \geq 0.
\]

Remark 4.2. Note that for each \(T > 0\) the process \(B^{-1}P(\cdot, T)\) is by definition a martingale. This means that \(\mathbb{Q}\) is a risk-neutral measure by construction, and that the model is free of arbitrage.

Definition 4.3 (Call and put options). The prices at time \(t \geq 0\) of European call and put options with expiry date \(T < S\) and strike \(K\) on the ZCB with maturity date \(S > T\) are given by
\[
\pi_t^{\text{Call}}(T, S, K) = \mathbb{E}\left[ \frac{B_0}{B_T} (P(T, S) - K)^+ \middle| \mathcal{F}_t \right], \quad \pi_t^{\text{Put}}(T, S, K) = \mathbb{E}\left[ \frac{B_0}{B_T} (K - P(T, S))^+ \middle| \mathcal{F}_t \right].
\]
Definition 4.4 (Caps and floors). Consider interest rate cap and floor with maturity $T_n$, strike rate $\kappa$ and payment dates $0 < T_0 < T_1 < \ldots < T_n$ where $T_k - T_{k-1} = \Delta$, $k \in \mathbb{N}$. At time $t < T_0$ the cap and floor prices are given

$$\text{Cp}_t = \sum_{k=1}^{n} \text{Cpl}(T_{k-1}, T_k, \kappa), \quad \text{and} \quad \text{Fl}_t = \sum_{k=1}^{n} \text{Fll}(T_{k-1}, T_k, \kappa),$$

where

$$\text{Cpl}(T_{k-1}, T_k, \kappa) = (1 + \Delta \kappa)\pi^\text{Put}_t (T_{k-1}, T_k, (1 + \Delta \kappa)^{-1}).$$

$$\text{Fll}(T_{k-1}, T_k, \kappa) = (1 + \Delta \kappa)\pi^\text{Call}_t (T_{k-1}, T_k, (1 + \Delta \kappa)^{-1}).$$

In order to calculate prices of call and put options on ZCB’s it is convenient to consider forward measure changes.

Definition 4.5 (Forward measure). For $0 \leq t \leq T$ define the $T$-forward measure $Q^T$ by the following Radon-Nikodym derivative

$$\xi(t, T) = \frac{dQ^T}{dQ} \bigg|_{\mathcal{F}_t} = \mathbb{E} \left[ \frac{B_0}{P(0, T)B_T} \bigg|_{\mathcal{F}_t} \right] = \frac{B_t^{-1}P(t, T)}{B_0^{-1}P(0, T)},$$

where the last equality follows directly from Definition 4.1.

The following property is useful for computations. The symbol $\mathcal{E}$ denotes the stochastic exponential, see e.g. [33, Section 4.1].

Theorem 4.6 (Black-Scholes formula). Assume that there is a process $(v(t, T))_{t \geq 0}$ such that for each $0 \leq t \leq T$,

$$\xi(t, T) = \mathcal{E} \left( \int_0^t v(s, T)ds \right) \bigg|_{\mathcal{F}_t}.$$

Then, for any $S, T > 0$ the process $W^T = W - \int_0^T v(s, T)ds$ is $Q^T$-Brownian motion and the price process of the ZCB with maturity date $S$ discounted by the ZCB with maturity $T$

$$\frac{P(t, S)}{P(t, T)} = \frac{P(0, S)}{P(0, T)} \mathcal{E} \left( \int_0^S (v(s, S) - v(s, T))dW^T_s \right), \quad t \in [0, S \wedge T],$$

is a $Q^T$-martingale. Moreover, assuming that $v(\cdot, T)$ is deterministic, call and put option prices are given by the following version of the Black-Scholes formula

$$\pi^\text{Call}_t = P(t, S)\Phi_0^\text{Gauss}(d_1) - KP(t, T)\Phi_0^\text{Gauss}(d_2),$$

$$\pi^\text{Put}_t = KP(t, T)\Phi_0^\text{Gauss}(-d_2) - P(t, S)\Phi_0^\text{Gauss}(-d_1),$$

where $\Phi_0^\text{Gauss}$ is the standard Gaussian cumulative distribution function and

$$d_{1,2} = \frac{\log \left( \frac{P(t, S)}{KP(t, T)} \right) + \frac{1}{2} \int_t^T (v(\cdot, S) - v(\cdot, T))^2 ds}{\sqrt{\int_t^T (v(\cdot, S) - v(\cdot, T))^2 ds}}.$$

Proof. The derivation of [33, Section 7] can also be used in this setting because the discounted ZCB price process $B^{-1}P(\cdot, T)$ is a martingale by construction. \qed
4.2. Fractional short rate process. In this section, we construct an interest rate model with a fractional short rate. To this aim, we fix measures $\mu, \nu$ on $(0, \infty)$ satisfying the following slightly strengthened version of Assumption 2.3.

**Assumption 4.7.** $\mu$ and $\nu$ are sigma-finite measures on $(0, \infty)$. The measure $\nu$ has a density $p$ with respect to $\mu$, and there exists $\beta \in (0, 2)$ such that for each $t > 0$,
\[
\int_0^\infty (1 \wedge x^{-\frac{1}{2}}) \mu(dx) < \infty, \quad \int_0^\infty (1 \wedge x^{-\frac{1}{2}}) \nu(dx) < \infty, \quad \sup_{x \in (0, \infty)} p(x)(1 \wedge x^{-\beta}) < \infty.
\]

Moreover, we fix $(u, v) \in L^\infty(\mu) \times L^\infty(\nu)$, $\ell \in \mathbb{R}$, and an initial value $(Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)$ for the process $(Y, Z)$ defined in Section 2. Given these model parameters, we define the short rate and bank account as

\[
(4.3) \quad r_t = \ell + \langle Y_t, u \rangle_\mu + \langle Z_t, v \rangle_\nu, \quad B_t = \exp \left( \int_0^t r_s ds \right).
\]

**Example 4.8.** Set $u = v = 1$ and consider measures of the form $\mu(dx) \propto x^{-\alpha} dx$ and $\nu(dx) \propto x^2 \nu dx$ for $\alpha \in (\frac{1}{2}, 1)$. Then Assumption 4.7 is satisfied. The process $(Y, Z)$ takes values in $L^1(\mu) \times L^1(\nu)$ and has continuous sample paths by Theorem 2.11. Therefore, it can be used to construct fractional processes as in Section 3. In particular, $(Y, u)_\mu$ is a fractional process of the same roughness as fBM with Hurst index $H = \alpha - \frac{1}{4} \in (0, \frac{1}{2})$, and $(Z, v)_\nu$ has the same roughness as fBM with $H = \alpha \in (\frac{1}{2}, 1)$. None of the two processes are semimartingales.

**Remark 4.9.** While the short rate may take negative values, the probability of yields becoming negative can be reduced by shifting the parameter $\ell$ and scaling the parameters $u, v$. Often times, either $u$ or $v$ will be set to zero, unless one is interested in mixing processes with long- and short-range dependence.

**Theorem 4.10** (Bond prices and forward rates). In the fractional short rate model (4.3), ZCB prices and forward rates are given by
\[
P(t, T) = e^{-\ell(T-t) + \Phi_0(T-t, u, v) + \langle Y_t, \Phi_1(T-t, u, v) \rangle_\mu + \langle Z_t, \Phi_2(T-t, u, v) \rangle_\nu}, \quad 0 \leq t \leq T,
\]
\[
h(t)(\tau) = \ell - \partial_\tau \Phi_0(\tau, u, v) - \langle Y_t, \partial_\tau \Phi_1(\tau, u, v) \rangle_\mu - \langle Z_t, \partial_\tau \Phi_2(\tau, u, v) \rangle_\nu, \quad t, \tau \geq 0,
\]
where for each $\tau \geq 0$ and $x \in (0, \infty)$
\[
\Phi_0(\tau, u, v) = \frac{1}{2} \int_0^\tau \langle \Phi_1(s, u, v), 1 \rangle_\mu^2 ds,
\]
\[
\Phi_1(\tau, u, v)(x) = e^{-\tau x} - 1 - u(x) + \left( e^{-\tau x} - 1 - \frac{\tau}{x} e^{-\tau x} \right) p(x) v(x),
\]
\[
\Phi_2(\tau, u, v)(x) = \frac{e^{-\tau x} - 1}{x} v(x).
\]

**Proof.** Lemma E.2 implies that the random variable $\int_t^T \left( \langle Y_s, u \rangle_\mu + \langle Z_s, v \rangle_\nu \right) ds$ is Gaussian, given $\mathcal{F}_t$, with mean
\[
- \langle Y_t, \Phi_1(T-t, u, v) \rangle_\mu - \langle Z_t, \Phi_2(T-t, u, v) \rangle_\nu
\]
and variance $2\Phi_0(T-t, u, v)$. Thus, the formula for ZCB prices follows from the formula of the moment generating function of the normal distribution. The expression for the forward rates follows by differentiation with respect to the time to maturity. \hfill \Box
Remark 4.11. The functions $\Phi_0, \Phi_1, \Phi_2$ are the unique solution of the Riccati equations
\begin{align*}
\partial_\tau \Phi_0(\tau, u, v) &= R_0(\Phi_1(\tau, u, v), \Phi_2(\tau, u, v)), \quad \Phi_0(0, u, v) = 0, \\
\partial_\tau \Phi_1(\tau, u, v) &= R_1(\Phi_1(\tau, u, v), \Phi_2(\tau, u, v)) - u, \quad \Phi_1(0, u, v) = 0, \\
\partial_\tau \Phi_2(\tau, u, v) &= R_2(\Phi_1(\tau, u, v), \Phi_2(\tau, u, v)) - v, \quad \Phi_2(0, u, v) = 0,
\end{align*}
with $R_0, R_1, R_2$ as in Lemma 2.8. Here, solutions are defined in analogy to Definition 2.7 and Lemma 2.8.

Theorem 4.12 (HJM equation). In the fractional short rate model (4.3) bond prices $(P(t,T))_{0 \leq t \leq T}$ and forward rates $(h(t)(\tau))_{\tau > 0}$ are semimartingales for each fixed $T, \tau > 0$. The forward rate process $h = (h(t)(\tau))_{\tau > 0}$ is a solution of the HJM equation
\begin{equation}
dh(t) = (Ah(t) + \mu_{\text{HJM}}) \, dt + \sigma_{\text{HJM}} \, dW_t,
\end{equation}
where $A$ denotes differentiation with respect to time to maturity $\tau$ and $\mu_{\text{HJM}}, \sigma_{\text{HJM}}$ are measurable functions on $(0, \infty)$ given by
\begin{align*}
\mu_{\text{HJM}}(\tau) &= \partial^2_\tau \Phi_0(\tau, u, v), \quad \sigma_{\text{HJM}}(\tau) = -\langle \partial_\tau \Phi_1(\tau, u, v), 1 \rangle_{\mu}.
\end{align*}

Proof. The semimartingale property of prices and forward rates follows from Lemmas E.3 and E.4, which are based on Theorem 2.13. The semimartingale decomposition of $h(\cdot)(\tau)$ is obtained by collecting the terms in Equation (2.13):
\begin{align*}
dh(t)(\tau) &= -d(Y_t, \partial_\tau \Phi_1(\tau, u, v))_{\mu} - d\langle Z_t, \partial_\tau \Phi_2(\tau, u, v) \rangle_{\nu} \\
&= (\langle Y_t, x \partial_\tau \Phi_1(\tau, u, v) - \partial_\tau \Phi_2(\tau, u, v)p \rangle_{\mu} + \langle Z_t, x \partial_\tau \Phi_2(\tau, u, v) \rangle_{\nu}) \, dt \\
&\quad - \langle \partial_\tau \Phi_1(\tau, u, v), 1 \rangle_{\mu} \, dW_t.
\end{align*}

Note that by abuse of notation, we wrote $x \partial_\tau \Psi_i(\tau, u, v)$ to designate the function $x \mapsto \partial_\tau \Psi_i(\tau, u, v)(x)$ for $i = 1, 2$. The second derivatives of $\Psi_i$ are
\begin{align*}
\partial^2_\tau \Phi_1(\tau, u, v) &= -x \partial_\tau \Phi_1(\tau, u, v) + \partial_\tau \Phi_2(\tau, u, v)p, \\
\partial^2_\tau \Phi_2(\tau, u, v) &= -x \partial_\tau \Phi_2(\tau, u, v).
\end{align*}

Therefore, we have for all $t \geq 0$ and $\tau > 0$
\begin{align*}
Ah(t)(\tau) &= -\partial^2_\tau \Phi_0(\tau, u, v) + \langle Y_t, x \partial_\tau \Phi_1(\tau, u, v) - \partial_\tau \Phi_2(\tau, u, v)p \rangle_{\mu} \\
&\quad + \langle Z_t, x \partial_\tau \Phi_2(\tau, u, v) \rangle_{\nu}.
\end{align*}

It follows that
\begin{equation}
dh(t)(\tau) = (Ah(t)(\tau) + \partial^2_\tau \Phi_0(\tau, u, v)) \, dt - \langle \partial_\tau \Phi_1(\tau, u, v), 1 \rangle_{\mu} \, dW_t,
\end{equation}
which allows one to identify $\mu_{\text{HJM}}$ and $\sigma_{\text{HJM}}$. \qed

Remark 4.13. The HJM drift condition is satisfied because
\begin{equation}
\mu_{\text{HJM}} = \partial^2_\tau \Phi_0(\tau, u, v) = \langle \partial_\tau \Phi_1(\tau, u, v), 1 \rangle_{\mu} \langle \Phi_1(\tau, u, v), 1 \rangle_{\mu} \\
= \sigma_{\text{HJM}}(\tau) \int_0^\tau \sigma_{\text{HJM}}(s) \, ds.
\end{equation}

Corollary 4.14 (Covariations). For each $\tau_1, \tau_2 > 0$ the following relation holds:
\begin{equation}
d[h(\cdot)(\tau_1), h(\cdot)(\tau_2)]_t = \langle \partial_\tau \Phi_1(\tau_1, u, v), 1 \rangle_{\mu} \langle \partial_\tau \Phi_1(\tau_2, u, v), 1 \rangle_{\mu} \, dt.
\end{equation}
To show that the Black-Scholes formula of Theorem 4.6 holds, we verify that the \( T \)-forward density process is the stochastic exponential of \( \int_0^T v(s, T) dW_s \) for a deterministic function \( v(\cdot, T) \).

**Corollary 4.15** (Forward measure). For \( 0 \leq t \leq T \) the density process \( \xi(t, T) \) takes the form (4.1) with deterministic \( v(t, T) = \langle \Phi_1(T - t, u, v), 1 \rangle_\mu \).

**Proof.** In Lemma E.3 we verified that the expressions \( \langle Y, \Phi_1(T - \cdot, u, v) \rangle_\mu \) and \( \langle Z, \Phi_2(T - \cdot, u, v) \rangle_\nu \) are semimartingales. Their semimartingale decompositions are given by Equation (2.13):

\[
d\langle Y_t, \Phi_1(T - t, u, v) \rangle_\mu = \int_0^t \left( u(x) - e^{-(T-t)x} \frac{1 - p(x)}{x} \right) Y_s^\tau \mu(dx)dt + \langle \Phi_1(T - t, u, v), 1 \rangle_\mu dW_t,
\]

\[
d\langle Z_t, \Phi_2(T - t, u, v) \rangle_\nu = \left( \langle v, Z_t \rangle_\nu + \langle \Phi_2(T - t, u, v), Y_t \rangle_\nu \right) dt.
\]

By the formula for bond prices in Theorem 4.10, \( \log(\xi(t, T)) \) satisfies

\[
d(\log(\xi(t, T))) = \left( -\langle Y_t, u \rangle_\mu - \langle Z_t, v \rangle_\nu - \partial_t \Phi_0(T - t, u, v) \right) dt + d\langle Y_t, \Phi_1(T - t, u, v) \rangle_\mu + d\langle Z_t, \Phi_2(T - t, u, v) \rangle_\nu.
\]

Applying Itô’s formula and canceling out terms yields

\[
d\xi(t, T) = \xi(t, T) \left( d(\log(\xi(t, T))) + \frac{1}{2} d(\log(\xi(\cdot, T) ))_I \right) = \xi(t, T) \langle \Phi_1(T - t, u, v), 1 \rangle_\mu dW_t,
\]

which implies that \( \xi \) is a stochastic exponential of the form (4.1) with \( v(t, T) = \langle \Phi_1(T - t, u, v), 1 \rangle_\mu \).

**Remark 4.16.** We summarize the results of Section 4.2. We considered a model with fractional short rate process constructed as a superposition of infinitely many OU processes. We derived closed-form expressions for ZCB prices and forward rates, and HJM equation (4.5) holds. It follows that prices of interest rate derivatives can be calculated as in the standard HJM framework (see e.g. [33, Section 6 and 7]), even though the short rate is not a semimartingale. Our results provide two ways of identifying model parameters: either, they could be calibrated to interest rate caps and floors using Black-Scholes formula (4.2) (c.f. Corollary 4.15), or they could be estimated from realized covariances of forward rates (c.f. Corollary 4.14).

4.3 Fractional bank account process. In this section, we construct an interest rate model with a fractional bank account process. To this aim, we fix measures \( \mu, \nu \) on \((0, \infty)\) satisfying Assumption 2.3. Moreover, we fix \( (u, v) \in L^\infty(\mu) \times L^\infty(\nu) \), \( \ell \in \mathbb{R} \), and an initial value \( (Y_0, Z_0) \in L^1(\mu) \times L^1(\nu) \) for the process \((Y, Z)\) defined in Section 2. Given these model parameters, we define the bank account process as

\[
B_t = e^{\ell t + \langle Y_t, u \rangle_\mu + \langle Z_t, v \rangle_\nu}.
\]

**Theorem 4.17** (Bond prices and forward rates). In the fractional bank account model (4.7), ZCB prices and forward rates are given by

\[
P(t, T) = e^{-\ell(T-t) + \langle Y_t, \phi_0(T-t, u, v) \rangle_\mu + \langle Z_t, \phi_2(T-t, u, v) \rangle_\nu}.
\]

\[
h(t)(\tau) = \ell - \partial_\tau \phi_0(\tau, -u, -v) - \langle Y_t, \partial_\tau \phi_1(\tau, -u, -v) \rangle_\mu - \langle Z_t, \partial_\tau \phi_2(\tau, -u, -v) \rangle_\nu.
\]
for each $0 \leq t \leq T$ and $\tau > 0$, where $\phi_0, \phi_1,$ and $\phi_2$ are given by Theorem 2.6.

Proof. The formula for the ZCB prices follows directly from Theorem 2.6 and Equation (4.7), and the formula for the forward rates follows by definition. □

**Theorem 4.18 (HJM equation).** Discounted bond prices $(B_t^{-1}P(t,T))_{t \geq 0}$ and forward rates $(h(t)(\tau))_{t \geq 0}$ are semimartingales for each $T, \tau > 0$. The forward rates solve HJM equation (4.5) with $\mu^{HJM}$ and $\sigma^{HJM}$ given by

\[ \mu^{HJM}(\tau) = \partial_t^2 \phi_0(\tau, -u, -v), \quad \sigma^{HJM}(\tau) = -\langle \partial_t \phi_1(\tau, -u, -v), 1 \rangle_\mu. \]

**Remark 4.19.** In contrast to discounted bond prices and forward rates, undiscounted bond prices $(P(t,T))_{0 \leq t \leq T}$ are not semimartingales, in general. For example, they are not semimartingales because $\langle \phi_1(\tau, -u, -v), 1 \rangle$ and $\langle \phi_2(\tau, -u, -v), 1 \rangle$ are not semimartingales, in general.

Proof. Discounted bond prices are martingales by definition. Forward rates are martingales by Lemma F.1. The semimartingale decomposition of the forward rate process is given by Equation (2.13) and reads as

\[ dh(t)(\tau) = -d \langle Y_t, \partial_t \phi_1(\tau, -u, -v) \rangle_\mu - d \langle Z_t, \partial_t \phi_2(\tau, -u, -v) \rangle_\nu \]

\[ = \langle Y_t, x \partial_x \phi_1(\tau, -u, -v) - \partial_t \phi_2(\tau, -u, -v) \rangle_\mu dt + \langle Z_t, x \partial_x \phi_2(\tau, -u, -v) \rangle_\nu dt - \langle \partial_t \phi_1(\tau, -u, -v), 1 \rangle_\mu dW_t, \]

where by abuse of notation we wrote $x \partial_x \phi_1(\tau, -u, -v)$ to designate the function $x \rightarrow x \partial_x \phi_1(\tau, -u, -v)$. The second derivatives of $\phi_1, \phi_2$ are

\[ \partial_t^2 \phi_1(\tau, -u, -v) = -x \partial_x \phi_1(\tau, -u, -v) + \partial_t \phi_2(\tau, -u, -v), \]

\[ \partial_t^2 \phi_2(\tau, -u, -v) = -x \partial_x \phi_2(\tau, -u, -v). \]

Hence, for all $t \geq 0$ and for all $\tau > 0$ we have

\[ Ah(t)(\tau) = -\partial_t^2 \phi_0(\tau, -u, -v) + \langle Y_t, x \partial_x \phi_1(\tau, -u, -v) \rangle_\mu + \langle Z_t, x \partial_x \phi_2(\tau, -u, -v) \rangle_\nu \]

\[ = \langle Y_t, \partial_t \phi_1(\tau, -u, -v) \rangle_\mu - \langle \partial_t \phi_1(\tau, -u, -v), 1 \rangle_\mu dW_t. \]

Therefore, the semimartingale decomposition of $h(\cdot)(\tau)$ can be written as

\[ dh(t)(\tau) = (Ah(t)(\tau) + \partial_t^2 \phi_0(\tau, -u, -v)) dt - \langle \partial_t \phi_1(\tau, -u, -v), 1 \rangle_\mu dW_t, \]

which allows one to identify $\mu^{HJM}$ and $\sigma^{HJM}$.

**Remark 4.20.** The HJM drift condition is satisfied:

\[ \mu^{HJM}(\tau) = \partial_t^2 \phi_0(\tau, -u, -v) = \langle \partial_t \phi_1(\tau, -u, -v), 1 \rangle_\mu \]

\[ = \sigma^{HJM}(\tau) \int_0^\tau \sigma^{HJM}(s)ds. \]

**Corollary 4.21 (Covariations).** For fixed $\tau_1, \tau_2 > 0$

\[ d[h(\cdot)(\tau_1), h(\cdot)(\tau_2)]_t = \langle \partial_t \phi_1(\tau_1, -u, -v), 1 \rangle_\mu \langle \partial_t \phi_1(\tau_2, -u, -v), 1 \rangle_\mu dt. \]

To show that the Black-Scholes formula of Theorem 4.6 holds, we verify that the $T$-forward density process is the stochastic exponential of $\int_0^T v(s,T) dW_s$ for a deterministic function $v(\cdot, T)$.

**Corollary 4.22 (Forward measure).** For $0 \leq t \leq T$ the density process $\xi(t,T)$ takes the form (4.1) with deterministic $v(t,T) = \langle \phi_1(\tau, -u, -v), 1 \rangle_\mu$.  

195
Proof. By Theorem 4.17, the density process $\xi(t,T)$ can be expressed equivalently as

$$d(\log(\xi(t,T))) = -\partial_t\phi_0(T-t, -u, -v) + d(Y_t, \phi_1(T-t, -u, -v))_\mu + d(Z_t, \phi_2(T-t, -u, -v))_\nu.$$ 

The processes $(\langle Y_t, \phi_1(T-t, -u, -v) \rangle_\mu)_{t \geq 0}$ and $(\langle Z_t, \phi_2(T-t, -u, -v) \rangle_\nu)_{t \geq 0}$ are semimartingales with decompositions given by

$$d(\langle Y_t, \phi_1(T-t, -1, -1) \rangle_\mu) = -\langle Y_t, \phi_2(T-t, -u, -v) \rangle_\mu dt + \langle \phi_1(T-t, -u, -v), 1 \rangle_\mu dW_t,$$

$$d(\langle Z_t, \phi_2(T-t, -u, -v) \rangle_\nu) = \langle Y_t, \phi_2(T-t, -u, -v) \rangle_\nu dt.$$ 

By Itō's formula, using ODE (2.8) for $\phi_0$, one obtains

$$d\xi(t,T) = \xi(t,T) \left( d(\log(\xi(t,T))) + \frac{1}{2} d[\log(\xi(t,T))]_t \right) = \xi(t,T) \langle \phi_1(T-t, -u, -v), 1 \rangle_\mu dW_t.$$ 

$\square$

**Remark 4.23.** We summarize the results of Section 4.3. We defined an interest rate model where the logarithmic bank account is a fractional process constructed as a superposition of infinitely many OU process. We derived closed-form expressions for ZCB prices and forward rates. While ZCB prices are typically not semimartingales, discounted ZCB prices and forward rates are. In the same way as in the fractional short rate model, the model parameters can be identified by calibration to caps and floors using Black-Scholes formula (4.2) (c.f. Corollary 4.22), or alternatively by estimation from forward rate realized covariations (c.f. Corollary 4.21).

5. **Fractional Stein & Stein model**

In this section we generalize an affine stochastic volatility model by Stein and Stein [75] to fractional volatility. In the original model, the volatility process is a single OU process. In our model, it is a fractional process constructed as a superposition of infinitely many OU processes. In accordance with empirical facts about realized volatility [37] we restrict ourselves to fractional processes with roughness and dependence structure similar to fBM of Hurst index $H < 1/2$.

5.1. **Setup and notation.** Let $\tilde{W}$ be $(\mathcal{F}_t)_{t \geq 0}$-Brownian motion with correlation $d\langle W, \tilde{W} \rangle_t = \rho dt$ for some $\rho \in (-1, 1)$. We fix a measure $\mu$ on $(0, \infty)$ satisfying Assumption 2.3, a function $v \in L^\infty(\mu)$, and an initial value $Y_0 \in L^1(\mu)$ for the process $Y$ defined in Section 2. Given these model parameters, the process $S = (S_t)_{t \geq 0}$ is defined by the SDE

$$dS_t = S_t(Y_t, v)_\mu d\tilde{W}_t.$$

To bring the SDE for the process $S$ into an affine form, we introduce the following spaces of simple symmetric tensors:

$$L^1(\mu) \otimes_s L^1(\mu) = \{y \otimes^2: y \in L^1(\mu)\} \subset L^1(\mu) \otimes^2 \subset L^1(\mu \otimes^2),$$

$$L^\infty(\mu) \otimes_s L^\infty(\mu) = \{v \otimes^2: v \in L^\infty(\mu)\} \subset L^\infty(\mu) \otimes^2 \subset L^\infty(\mu \otimes^2).$$

1All tensor products are algebraic; we do not complete the tensor products.
For each \( t \geq 0 \) we set \( \Pi_t = Y \otimes^2 t \in L^1(\mu) \otimes L^1(\mu) \). Then the relation \( \langle Y_t, v \rangle^2 = \langle Y_t \otimes^2, v \otimes^2 \rangle_{\mu \otimes^2} \) holds. Therefore, the log-price process \( X = \log(S) \) satisfies
\[
(5.1) \quad dX_t = -\frac{1}{2} \langle \Pi_t, v \otimes^2 \rangle_{\mu \otimes^2} dt + \sqrt{\langle \Pi_t, v \otimes^2 \rangle_{\mu \otimes^2}} dW_t.
\]

5.2. Affine structure of \( \Pi \). The following theorem characterizes \( \Pi \) as an affine process with values in \( L^1(\mu) \otimes L^1(\mu) \).

**Theorem 5.1** (Affine structure). Let \( v \otimes^2 \in iL^\infty(\mu) \otimes L^\infty(\mu) \). Then, with probability one,
\[
E \left[ e^{(\Pi_{t-v} \otimes^2)_{\mu \otimes^2}} \bigg| \mathcal{F}_t \right] = e^{\psi_0(t-v) + (\Pi_t, \psi_1(t-v))_{\mu \otimes^2}}, \quad 0 \leq t \leq T,
\]
where \( \psi_0(t, v) \in \mathbb{C} \) and \( \psi_1(t, v) \in L^\infty(\mu; \mathbb{C}) \otimes L^\infty(\mu; \mathbb{C}) \) are given by
\[
\psi_0(t, v) = -\frac{1}{2} \log(1 - 4\phi_0(t, v, 0)),
\]
\[
\psi_1(t, v) = \frac{\phi_1(t, v, 0) \otimes^2}{1 - 4\phi_0(t, v, 0)}.
\]

**Remark 5.2.** An immediate observation is that for each \((x, y) \in (0, \infty)^2\), the tuple \((\Pi^x,\Pi^y,\Pi^{y\otimes^2})\) is an affine process. This can be seen from the following SDE for \( \Pi_t^x = Y_t^x Y_t^y \), which follows from Itô’s rule:
\[
d\Pi_t^y = (1 - (x + y)\Pi_t^{x\otimes^2}) dt + \sqrt{\Pi_t^{x\otimes^2} + 2\Pi_t^{2y} + \Pi_t^{y\otimes^2}} dW_t.
\]

More generally, for any finite set of points \( x_i \), the process \((\Pi^{x_i},\Pi^{x_j})_{i,j}\) is affine. Theorem 5.1 generalizes this observation to infinitely many points \( x^i, x^j \in (0, \infty) \). A version of Theorem 5.1 with \( v \otimes^2 \) replaced by arbitrary symmetric test functions is given in Lemma G.3.

**Proof.** By Lemma D.3 the random variable \( \frac{1}{\sqrt{2\phi_0(T-t, v, 0)}} \langle Y_T, v \rangle_{\mu} \) is Gaussian, given \( \mathcal{F}_t \), with mean
\[
\langle Y_t, \phi_1(T - t, v, 0) \rangle_{\mu} / \sqrt{2\phi_0(T - t, v, 0)},
\]
and unit variance. Hence, the random variable
\[
\frac{\langle \Pi_T, v \otimes^2 \rangle_{\mu \otimes^2}}{2\phi_0(T - t, v, 0)} = \left( \frac{\langle Y_T, v \rangle_{\mu}}{\sqrt{2\phi_0(T - t, v, 0)}} \right)^2,
\]
is non central \( \chi^2 \)-distributed, given \( \mathcal{F}_t \), with one degree of freedom and non centrality parameter
\[
\frac{\langle Y_t, \phi_1(T - t, v, 0) \rangle_{\mu}^2}{2\phi_0(T - t, v, 0)} = \frac{\langle \Pi_t, \phi_1(T - t, v, 0) \otimes^2 \rangle_{\mu \otimes^2}}{2\phi_0(T - t, v, 0)}.
\]
The statement follows from the formula for the characteristic function of the non central \( \chi^2 \) distribution. \( \square \)

The coefficient functions \((\psi_0, \psi_1)\) of Theorem 5.1 are solutions of an infinite dimensional version of the Riccati ODE’s in the sense of Definition 2.7.
Lemma 5.3 (Riccati equations). For any \(v \otimes 2 \in iL^\infty(\mu) \otimes_s L^\infty(\mu)\), the functions \(\psi_0(\cdot, v \otimes 2)\) and \(\psi_1(\cdot, v \otimes 2)\) given by Theorem 5.1 solve the following system of differential equations

\[
\begin{align*}
\partial_\tau \psi_0 (\tau, v \otimes 2) &= F_0 (\psi_1 (\tau, v \otimes 2), \psi_0 (0, v \otimes 2)) = 0, \\
\partial_\tau \psi_1 (\tau, v \otimes 2) &= F_1 (\psi_1 (\tau, v \otimes 2), \psi_1 (0, v \otimes 2)) = v \otimes 2,
\end{align*}
\]

where for any \(w \in L^\infty(\mu; \mathbb{C})^\otimes 2\), \(F_0(w)\) is a complex number given by

\[
F_0(w) = \int_0^\infty \int_0^\infty w(x, y) \mu(dx) \mu(dy),
\]

and \(F_1(w)\) is a measurable function on \((0, \infty)^2\) given by

\[
F_1(w)(x, y) = -(x + y)w(x, y) + 2 \int_0^\infty \int_0^\infty w(x, x')w(y, y') \mu(dx') \mu(dy').
\]

Proof. The initial conditions are satisfied by Lemma 2.8. We differentiate with respect to \(\tau\) and use Lemma 2.8:

\[
\begin{align*}
\partial_\tau \psi_0 (\tau, v \otimes 2) &= \frac{2}{1 - 4\phi_0(\tau, v, 0)} \partial_\tau \phi_0(\tau, v, 0) = F_0 (\psi_1 (\tau, v \otimes 2)), \\
\partial_\tau \psi_1 (\tau, v \otimes 2)(x, y) &= -x\phi_1(\tau, v, 0)(x)\phi_1(\tau, v, 0)(y) - y\phi_1(\tau, v, 0)(x)\phi_1(\tau, v, 0)(y) + 2\phi_1(\tau, v, 0)(x)\phi_1(\tau, v, 0)(y) \\
&\quad + \frac{2\phi_1(\tau, v, 0)(x)\phi_1(\tau, v, 0)(y)}{(1 - 4\phi_0(\tau, v, 0))^2} \left( \int_0^\infty \phi_1(\tau, v, 0)(z) \mu(dz) \right)^2 \\
&= F_1 (\psi_1 (\tau, v \otimes 2))(x, y).
\end{align*}
\]

\[\square\]

5.3. Affine structure of \((X, \Pi)\). The following theorem shows that \((X, \Pi)\) is an affine process with values in \(\mathbb{R} \times L^1(\mu) \otimes_s L^1(\mu)\). The proof is based on an approximation of \((Y, w)_\mu\) going back to Carmona, Coutin, and Montseny [15]. This approximation also provides a mean for simulating the fractional Stein and Stein model.

Theorem 5.4 (Affine structure). Let \(\mu\) satisfy Assumption 2.3 and \((X_0, \Pi_0) \in \mathbb{R} \times L^1(\mu) \otimes_s L^1(\mu)\). Then \((X, \Pi)\) is an affine process in the sense that for each \(0 \leq t \leq T, u \in \mathbb{R}\), and \(v \otimes 2 \in iL^\infty(\mu) \otimes_s L^\infty(\mu)\), the logarithmic conditional characteristic function

\[
\log \mathbb{E} \left[ e^{X_{Tt}} (\Pi, v \otimes 2)_{\mu} | F_t \right],
\]

is affine in \((X_t, \Pi_t)\).

Proof. We approximate the measure \(\mu\) by a sequence \(\mu^n\) of atomic measures. If \(\mu^n\) are suitably chosen, it follows from [15] that \((Y, v)_\mu^n\) converges uniformly on compacts in probability (ucp) to \((Y, v)_\mu\). It follows that \((\Pi, v \otimes 2)_{(\mu^n)^{\otimes 2}} = (Y, v)_\mu^n\) converges ucp to \((\Pi, v \otimes 2)_{\mu^{\otimes 2}} = (Y, v)_\mu^{\otimes 2}\). Let \(X^n\) be the corresponding process solving Equation (5.1) with \(\mu\) replaced by \(\mu^n\). As stochastic integrals are continuous in the ucp topology, it follows that \(X^n_T\) converges in probability to \(X_T\). This implies convergence of the logarithmic characteristic function in Theorem 5.4. For each \(n\), the logarithm characteristic function is affine by Remark 5.2 and the affine nature of Equation (5.1). The result follows. \[\square\]
5.4. The uncorrelated case. By “uncorrelated” we mean $d(W, \tilde{W})_t = \rho dt = 0$. In the uncorrelated case, the distribution of $X_T$ depends immediately on the integrated variance, which is defined as

$$IV(t, T) = \frac{1}{T-t} \int_t^T (Y_s, v)_\mu^2 ds = \frac{1}{T-t} \int_t^T \langle \Pi_s, \tilde{v} \otimes 2 \rangle_{\mu \otimes 2} ds.$$ 

This dependence is made precise in the following lemma.

Lemma 5.5 (Conditional CDF). In the uncorrelated case $\rho = 0$, the $\mathcal{F}_t$-conditional cumulative distribution function of $X_T$ is

$$\mathbb{Q}[X_T \leq x | \mathcal{F}_t] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \mathbb{E} \left[ \exp \left( -\frac{(y - X_t + \frac{T-t}{2} IV(t, T))^2}{2(T-t) IV(t, T)} \right) \right] dy,$$

and the $\mathcal{F}_t$-conditional characteristic function is

$$\mathbb{E} \left[ e^{X_T u + \langle \Pi_T, \tilde{v} \otimes 2 \rangle_{\mu \otimes 2}} | \mathcal{F}_t \right] = e^{X_t u + \langle \Pi_t, \tilde{v} \otimes 2 \rangle_{\mu \otimes 2} \mathbb{E} \left[ e^{\frac{T-t}{2} (u^2 - u) IV(t, T)} | \mathcal{F}_t \right],$$

where $0 \leq t \leq T$, $u \in \mathbb{R}$, $\tilde{v} \otimes 2 \in iL^\infty(\mu) \otimes_s L^\infty(\mu)$.

Proof. This can be seen as in [75] by conditioning on the sigma algebra generated by $(Y_t, v)_{0 \leq t \leq T}$ and by using the independence of $W$ and $\tilde{W}$. \qed

The Fourier transform of the integrated variance process can be calculated explicitly using the affine structure of the process $\Pi$. Thus, in theory, it is possible to characterize the conditional distribution of the integrated variance. An example is given in the next corollary.

Corollary 5.6 (Conditional moments). For each $0 \leq t \leq T$, the first and second $\mathcal{F}_t$-conditional moments of the integrated variance $IV(t, T)$ are given by

$$\mathbb{E}[IV(t, T) | \mathcal{F}_t] = \int_t^T \left( 2\phi_0(s - t, v, 0) + \langle \Pi_s, \phi_1(s - t, v, 0) \tilde{v} \otimes 2 \rangle_{\mu \otimes 2} \right) ds,$$

$$\mathbb{E}[ IV^2(t, T) | \mathcal{F}_t ] = 4 \int_t^T \int_s^T \left( \phi_0(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) \phi_0(s_1 \wedge s_2 - t, v, 0) 
+ 2\phi_0(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) \langle \Pi_s, \phi_1(s_1 \wedge s_2 - t, v, 0) \tilde{v} \otimes 2 \rangle_{\mu \otimes 2} 
+ \frac{1}{4} \mathbb{E} \left[ (\Pi_s \wedge s_2, \tilde{w}(s_1, s_2))_{\mu \otimes 2}^2 | \mathcal{F}_t \right] \right) ds_1 ds_2,$$

where $\tilde{w}(s_1, s_2) = v \otimes \phi_1(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) + \phi_1(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) \otimes v$ is symmetric two tensor and the last expectation is given by Lemma G.5.

Proof. We obtain the formula for the conditional mean using Lemma G.4. Note that we are allowed to exchange the conditional expectation and integration because the integrand is positive. For the second moment we use the tower property of
conditional expectations and Lemma G.4 for the conditional mean:

\[
\mathbb{E} \left[ V(t,T)^2 \mid \mathcal{F}_t \right] = \int_t^T \int_t^T \mathbb{E} \left[ \langle \Pi_{s_1,1}^{\otimes 2}, \Pi_{s_2,1}^{\otimes 2} \rangle_{\mu^{\otimes 2}} \bigg| \mathcal{F}_t \right] ds_2 ds_1
\]

\[
= \int_t^T \int_t^T \mathbb{E} \left[ \langle \Pi_{s_1,1}^{\otimes 2}, \Pi_{s_2,1}^{\otimes 2} \rangle_{\mu^{\otimes 2}} \bigg| \mathcal{F}_s \right] ds_2 ds_1
\]

\[
= 4 \int_t^T \int_t^T \phi_0(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) \phi_0(s_1 \wedge s_2 - t, v, 0)
\]

\[
+ 2 \phi_0(s_1 \vee s_2 - s_1 \wedge s_2, v, 0) \langle \Pi_t, \phi_1(s_1 \wedge s_2 - t, v, 0) \rangle_{\mu^{\otimes 2}}
\]

\[
+ \mathbb{E} \left[ \langle \Pi_{s_1,1}^{\otimes 2}, \Pi_{s_2,1}^{\otimes 2} \rangle_{\mu^{\otimes 2}} \bigg| \mathcal{F}_t \right] ds_2 ds_1.
\]

We set \( s = s_1 \wedge s_2 \) and \( \tau = s_1 \vee s_2 - s_1 \wedge s_2 \). Observe that

\[
\langle \Pi_s, v^{\otimes 2} \rangle_{\mu^{\otimes 2}} \langle \Pi_s, \phi_1(\tau, v, 0) \rangle_{\mu^{\otimes 2}} = \langle Y_s^{\otimes 2}, v^{\otimes 2} \rangle_{\mu^{\otimes 2}} \langle Y_s^{\otimes 2}, \phi_1(\tau, v, 0) \rangle_{\mu^{\otimes 2}}
\]

\[
= \left( \langle Y_s, v \rangle_{\mu} \langle Y_s, \phi_1(\tau, v, 0) \rangle_{\mu} \right)^2 = \langle \Pi_s, v \otimes \phi_1(\tau, v, 0) \rangle_{\mu^{\otimes 2}}^2 = 1/4 \langle \Pi_s, w \rangle_{\mu^{\otimes 2}}^2,
\]

where \( w = v \otimes \phi_1(\tau, v, 0) + \phi_1(\tau, v, 0) \otimes v \) is a symmetric two tensor. The result follows from Lemma G.5.

\( \square \)

**Remark 5.7.** We summarize the results of Section 5. We generalized the stochastic volatility model by Stein and Stein [75] to fractional volatility. We introduced an affine framework for formulating the model. Namely, we showed that \( \sigma \) measurable with respect to the product \( \mathcal{B} \) and \( \Pi = Y^{\otimes 2} \), the model can be approximated by finite-dimensional affine models as shown in the proof of Theorem 5.4.

**Appendix A. Stochastic Fubini’s Theorem**

We refer to the version of the theorem proved in [80]. Let \( \mu \) be a \( \sigma \)-finite measure on \((0, \infty)\). Fix \( T \geq 0 \) and denote by \( \mathcal{P} \) the \( \sigma \)-algebra on \([0, T] \times \Omega \) generated by all progressively measurable processes.

**Theorem A.1** (Stochastic Fubini Theorem). Let \( G : (0, \infty) \times [0, T] \times \Omega \to \mathbb{R} \) be measurable with respect to the product \( \sigma \)-algebra \( \mathcal{B}(0, \infty) \otimes \mathcal{P} \). Define processes \( \zeta_{1,2} : (0, \infty) \times [0, T] \times \Omega \to \mathbb{R} \) and \( \eta : [0, T] \times \Omega \to \mathbb{R} \) by

\[
\zeta_1(x, t, \omega) = \int_0^t G(x, s, \omega) ds,
\]

\[
\zeta_2(x, t, \omega) = \left( \int_0^t G(x, s, \cdot) dW_s \right)(\omega),
\]

\[
\eta(t, \omega) = \int_0^\infty G(x, t, \omega) \mu(dx).
\]

(i) Assume \( G \) satisfies for almost all \( \omega \in \Omega \)

\[
\int_0^\infty \int_0^T |G(x, s, \omega)| ds \mu(dx) < \infty.
\]

Then, for almost all \( \omega \in \Omega \) and for all \( t \in [0, T] \) we have \( \zeta_1(\cdot, t, \omega) \in L^1(\mu) \) and

\[
\int_0^\infty \zeta_1(x, t, \omega) \mu(dx) = \int_0^t \eta(s, \omega) ds.
\]

200
(ii) Assume $G$ satisfies for almost all $\omega \in \Omega$

\[
\int_{0}^{\infty} \sqrt{\int_{0}^{T} G(x,s,\omega)^2 ds} \mu(dx) < \infty.
\]

Then, for almost all $\omega \in \Omega$ and for all $t \in [0,T]$ we have $\zeta_{2}(.,t,\omega) \in L^{1}(\mu)$ and

\[
\int_{0}^{\infty} \zeta_{2}(x,t,\omega) \mu(dx) = \left( \int_{0}^{t} \eta(s,\cdot) dW_s \right)(\omega).
\]

Remark A.2. Note that

\[
\int_{0}^{\infty} \int_{0}^{T} E |G(x,s)| \, ds \mu(dx) < \infty \quad \text{and} \quad \int_{0}^{\infty} E \left[ \sqrt{\int_{0}^{T} G(x,s)^2 ds} \right] \mu(dx) < \infty
\]

imply that conditions (A.1) and (A.2) hold with probability one.

**Appendix B. Reproducing kernel Hilbert spaces**

We adapt the exposition of [65, Section 8] to our setting and refer to this reference for further details. Let $P: L^{\infty}(\mu; \mathbb{C}) \to L^{1}(\mu; \mathbb{C})$ be a positive and symmetric bounded linear operator, i.e., $\langle Pu, u \rangle_{\mu} \geq 0$ and $\langle Pu, v \rangle_{\mu} = \langle Pv, u \rangle_{\mu}$ for all $u, v \in L^{\infty}(\mu; \mathbb{C})$.

The bilinear form $(Pu, Pv) \mapsto \langle Pu, v \rangle_{\mu}$ defines an inner product on the image of $P$. The completion of the image of $P$ with respect to this inner product is a Hilbert space, which we denote by $\text{im}(P)$. The inclusion of the image of $P$ in $L^{1}(\mu; \mathbb{C})$ extends to a bounded injective operator $i: \text{im}(P) \to L^{1}(\mu; \mathbb{C})$. The space $H = \text{im}(i) \subseteq L^{1}(\mu; \mathbb{C})$ with the Hilbert structure induced by the bijection $i: \text{im}(P) \to H$ is called the reproducing kernel Hilbert space of $P$. If $u, v \in L^{\infty}(\mu; \mathbb{C})$, then $Pu, Pv \in H$ and $\langle Pu, Pv \rangle_{H} = \langle Pu, v \rangle_{\mu}$, where the inclusion $i$ is dropped from our notation.

**Appendix C. Basic estimates**

We collect some inequalities and estimates which are used throughout the paper.

**Lemma C.1** (Elementary inequalities). The following inequalities hold true for all $x, y > 0$

\[
\begin{align*}
\text{(C.1)} & \quad 1 \wedge xy \leq (1 \vee x) (1 \wedge y), \\
\text{(C.2)} & \quad y \wedge x^{-1} \leq (1 \vee y) (1 \wedge x^{-1}),
\end{align*}
\]

and for all $\alpha, \tau > 0$ and $0 < \epsilon < 1$,

\[
\begin{align*}
\text{(C.3)} & \quad e^{-x \tau} \leq \left( 1 \vee \left( \frac{\tau}{\alpha} \right)^{-\alpha} \right) (1 \wedge x^{-\alpha}), \\
\text{(C.4)} & \quad \frac{1 - e^{-\tau x}}{x} \leq (1 \vee \tau) (1 \wedge x^{-1}), \\
\text{(C.5)} & \quad \frac{1 - e^{-\tau x} (1 + \tau x)}{x^2} \leq (1 \vee \tau^2) (1 \wedge x^{-2}), \\
\text{(C.6)} & \quad \frac{1 - e^{-\tau x} (1 + \tau x + \frac{1}{2} \tau^2 x^2)}{x^3} \leq (1 \vee \tau^3) (1 \wedge x^{-3}), \\
\text{(C.7)} & \quad \int_{0}^{t} s^{-\epsilon} e^{-2sx} ds \leq \left( 2^{\epsilon-1} \Gamma(1 - \epsilon) \vee \frac{t^{1-\epsilon}}{1-\epsilon} \right) (1 \wedge x^{\epsilon-1}).
\end{align*}
\]

\[\text{In [65] the space } \text{im}(P) \text{ is called reproducing kernel Hilbert space of } P.\]
Proof. For the inequalities (C.1)-(C.2) consider the following four cases separately.

(1) If $0 < x, y \leq 1$. Then, $1 \land xy = xy \leq y = (1 \lor x) (1 \land y)$ and $y \land x^{-1} = y \leq \frac{1}{1 \lor y} \left(1 \land x^{-1}\right)$.

(2) If $0 < x \leq 1 \leq y$. Then, $1 \land xy \leq 1 = (1 \lor x) (1 \land y)$ and $y \land x^{-1} \leq y = (1 \lor y) \left(1 \land x^{-1}\right)$.

(3) If $0 < y \leq 1 \leq x$. Then, $1 \land xy \leq xy = (1 \lor x) (1 \land y)$ and $y \land x^{-1} \leq x^{-1} = (1 \lor y) \left(1 \land x^{-1}\right)$.

(4) If $1 \leq x, y$. Then, $1 \land xy = 1 \leq (1 \lor x) (1 \land y)$ and $y \land x^{-1} = x^{-1} \leq \frac{1}{y} (1 \land x^{-1})$.

Consider the functions $f(x, \tau) = e^{-x\tau}$ and $g(x, \tau, \alpha) = x^\alpha f(x, \tau)$. Obviously, $f(x, \tau) \leq 1$ for all $x, \tau > 0$. Note that $\partial_x g(x, \tau, \alpha) = x^{\alpha-1} e^{-x\tau} (\alpha - \tau x)$ and $g$ attains its maximum in $x$ at $\frac{\tau}{\alpha}$. Hence, Equation (C.3) follows from

$$f(x, \tau) = \frac{g(x, \tau, \alpha)}{x^\alpha} \leq \frac{g \left( \frac{\tau}{\alpha}, \tau, \alpha \right)}{x^\alpha} = \left( \frac{\tau}{\alpha} x \right)^{-\alpha} e^{-\alpha} \leq \left( \frac{\tau}{\alpha} x \right)^{-\alpha},$$

and Equation (C.1).

Define $k_1(x, \tau) = \frac{1-e^{-x \tau}}{x}$, $k_2(x, \tau) = \frac{1-e^{-x(1+\tau \alpha)}}{x^2}$ and $k_3(x, \tau) = \frac{1-e^{-x(1+\tau x+\frac{1}{2} \tau^2 x^2)}}{x^3}$.

Computing the derivatives with respect to $x$ shows that $k_{1,2,3}(x, \tau)$ are decreasing functions in $x$ for all $\tau > 0$. The inequalities (C.4)-(C.6) follow from

$$\lim_{x \to -\infty} k_{1,2,3}(x, \tau) = 0, \quad \lim_{x \to 0^+} k_i(x, \tau) = \begin{cases} \tau, & i = 1, \\ \frac{\tau^2}{2}, & i = 2, \\ \frac{\tau^3}{6}, & i = 3, \end{cases}$$

and Equation (C.2). Equation (C.7) follows from the relation

$$\int_0^\infty s^{-\epsilon} e^{-2sx} ds = \int_0^{2tx} (2x)^{\epsilon-1} s^{-\epsilon} e^{-s} ds,$$

and from the following two estimates:

$$\int_0^{2tx} (2x)^{\epsilon-1} s^{-\epsilon} e^{-s} ds \leq \int_0^\infty (2x)^{\epsilon-1} s^{-\epsilon} e^{-s} ds = (2x)^{\epsilon-1} \Gamma(1-\epsilon),$$

$$\int_0^{2tx} (2x)^{\epsilon-1} s^{-\epsilon} e^{-s} ds \leq \int_0^{2tx} (2x)^{\epsilon-1} s^{-\epsilon} e^{-s} ds = \frac{t^{1-\epsilon}}{1-\epsilon}.$$ \hfill \Box

Lemma C.2 (Integrability of elementary expressions). Let Assumption 2.3 be in place and let $\tau, \alpha > 0$. Then

(C.8) $\int_0^\infty e^{-x\tau} \mu(dx) < \infty$,

(C.9) $\int_0^\infty e^{-x\tau} \nu(dx) < \infty$,

(C.10) $\int_0^\infty x^\alpha e^{-x\tau} \mu(dx) < \infty$,

(C.11) $\int_0^\infty x^\alpha e^{-x\tau} \nu(dx) < \infty$,

(C.12) $\int_0^\infty \sqrt{\frac{1-e^{-2tx}}{x}} \mu(dx) < \infty$. 

202
Further, for each $0 \leq t < T$ we have

\begin{align}
(C.15) & \quad \int_0^\infty \sqrt{\int_t^T e^{-2s(T-s)} ds} \mu(dx) < \infty, \\
(C.16) & \quad \int_0^\infty \sqrt{\int_t^T (T-s)^2 e^{-2s(T-s)} ds} \nu(dx) < \infty, \\
(C.17) & \quad \int_0^\infty \int_t^T e^{-x(T-s)} ds \mu(dx) < \infty, \\
(C.18) & \quad \int_0^\infty \int_t^T (T-s) e^{-x(T-s)} ds \nu(dx) < \infty, \\
(C.19) & \quad \int_0^\infty \int_t^T \frac{1 - e^{-x(T-s)}}{x} (1 \wedge x^{-1}) ds \nu(dx) < \infty, \\
(C.20) & \quad \int_0^\infty \sqrt{\int_t^T \left( \frac{1 - e^{-x(T-s)}}{x} \right)^2 ds} \mu(dx) < \infty, \\
(C.21) & \quad \int_0^\infty \sqrt{\int_t^T \left( \frac{1 - e^{-x(T-s)}}{x} \right)^2 ds} \mu(dx) < \infty, \\
(C.22) & \quad \int_0^\infty \int_0^\infty \int_t^T e^{-(x+y)(T-s)} ds \mu(dx) \nu(dy) < \infty, \\
(C.23) & \quad \int_0^\infty \int_0^\infty \int_t^T (T-s)^2 e^{-(x+y)(T-s)} ds \nu(dx) \nu(dy) < \infty.
\end{align}

Proof. Equations (C.8) and (C.9) follow directly from (C.3) for $\alpha = \frac{1}{2}$ and $\alpha = \frac{3}{2}$, respectively. Applying Equation (C.3) for $\beta \geq \alpha$ we obtain

\begin{align*}
\int_0^\infty x^\alpha e^{-x \tau} \mu(dx) & \leq \int_0^1 e^{-x \tau} \mu(dx) + \int_1^\infty x^\alpha e^{-x \tau} \mu(dx) \\
& \leq \int_0^1 \left( 1 \vee \left( \frac{x}{\beta} \right)^{\beta} \right) \mu(dx) + \int_1^\infty x^{\alpha - \beta} \left( 1 \vee \left( \frac{x}{\beta} \right)^{\beta} \right) \mu(dx) \\
& = \left( 1 \vee \left( \frac{x}{\beta} \right)^{\beta} \right) \int_0^\infty \left( 1 \wedge x^{\alpha - \beta} \right) \mu(dx),
\end{align*}

and in the same way $\int_0^\infty x^\alpha e^{-x \tau} \nu(dx) \leq \int_0^1 (1 \vee \left( \frac{x}{\beta} \right)^{\beta} \nu(dx) + \int_1^\infty x^{\alpha - \beta} (1 \vee \left( \frac{x}{\beta} \right)^{\beta} \nu(dx)$. Setting $\beta = \alpha + \frac{1}{2}$ and $\beta = \alpha + \frac{3}{2}$ one proves (C.10) and (C.11), respectively. By Equation (C.4) we obtain Equation (C.12)

\begin{align}
(C.24) & \quad \int_0^\infty \sqrt{\frac{1 - e^{-2x \tau}}{x}} \mu(dx) \leq \left( 1 \vee (2\tau)^{\frac{1}{2}} \right) \int_0^\infty \left( 1 \wedge x^{-\frac{1}{2}} \right) \mu(dx) < \infty.
\end{align}
By Equation (C.12) we obtain Equation (C.15)
\[
\int_0^\infty \sqrt{\int_t^T e^{-2x(T-s)}ds} \mu(dx) = \int_0^\infty \sqrt{\frac{1 - e^{-2(T-t)x}}{2x}} \mu(dx) < \infty.
\]

By Equation (C.6) we obtain Equation (C.13)
\[
\int_0^\infty \sqrt{1 - e^{-2\tau x}(1 + 2\tau x + 2\tau^2 x^2)} \nu(dx) \leq (1 \lor (2\tau)^2) \int_0^\infty (1 \land x^{-\frac{1}{2}}) \nu(dx) < \infty.
\]

Equation (C.14) follows from
\[
\int_0^\infty \sqrt{1 - 2e^{-\tau x}(\tau x + 1) + 2\tau xe^{-2\tau x}(\tau x + 1) + e^{-2\tau x}} \frac{1}{4x^3} \nu(dx) \leq \int_0^{1/\tau} \sqrt{\frac{\tau^2}{6x}} \nu(dx) + \int_{1/\tau}^\infty \sqrt{\frac{2}{x^2}} \nu(dx) \leq \sqrt{2(\tau \lor 1) \int_0^\infty (x^{-\frac{1}{2}} \land x^{-\frac{1}{2}}) \nu(dx) < \infty}.
\]

Equation (C.13) implies Equation (C.16)
\[
\int_0^\infty \sqrt{\int_t^T (T-s)^2 e^{-2x(T-s)}ds} \nu(dx) = \int_0^\infty \sqrt{\frac{1 - e^{-2(T-t)x}}{2x}} \left(1 + 2(T-t)x + 2(T-t)^2 x^2\right) \frac{1}{4x^3} \nu(dx) < \infty.
\]

Equation (C.17) is obtained using (C.3) for $\alpha = \frac{1}{2}$
\[
\int_0^\infty \int_t^T e^{-x(T-s)} ds \mu(dx) \leq \int_t^T (1 \lor (T-s)^{-\frac{1}{2}}) ds \int_0^\infty (1 \land x^{-\frac{1}{2}}) \mu(dx) = (t \lor (T - 1) - t + 2\sqrt{T - (t \lor (T - 1))} \int_0^\infty (1 \land x^{-\frac{1}{2}}) \mu(dx) < \infty.
\]

Equation (C.18) is obtained using (C.3) for $\alpha = \frac{3}{2}$
\[
\int_0^\infty \int_t^T (T-s)e^{-x(T-s)} ds \nu(dx) \leq \int_t^T (T-s) \left(1 \lor (T-s)^{-\frac{3}{2}}\right) ds \int_0^\infty (1 \land x^{-\frac{3}{2}}) \mu(dx) \leq \left(\int_t^T (T-s) ds \lor \int_t^T (T-s)^{-\frac{3}{2}} ds\right) \int_0^\infty (1 \land x^{-\frac{3}{2}}) \mu(dx) = \left(\frac{(T-t)^2}{2} \lor 2\sqrt{T-t}\right) \int_0^\infty (1 \land x^{-\frac{3}{2}}) \mu(dx) < \infty.
\]
Equation (C.4) immediately implies Equation (C.19)
\[
\int_0^\infty \int_t^T \frac{1 - e^{-x(T-s)}}{x} (1 \wedge x^{-\frac{1}{2}}) ds \nu(dx) \\
\leq \int_0^\infty \left(1 \wedge x^{-\frac{1}{2}}\right) \nu(dx) \int_t^T (1 \vee (T-s)) ds < \infty,
\]

and Equation (C.20)
\[
\int_0^\infty \sqrt{\int_t^T \left(1 - e^{-x(T-s)}(1 \wedge x^{-1}) - \frac{1}{x}\right) ds \mu(dx)} \\
\leq \sqrt{\int_t^T (1 \vee (T-s)^2) ds} \int_0^\infty \left(1 \wedge \frac{1}{x}\right) \mu(dx) < \infty.
\]

Equation (C.5) immediately implies Equation (C.21)
\[
\int_0^\infty \sqrt{\int_t^T \left(e^{-x(T-s)}(1 + x(T-s)) - 1\right)^2 ds \mu(dx)} \\
\leq \sqrt{\int_t^T (1 \vee (T-s)^4) ds} \int_0^\infty \left(1 \wedge x^{-2}\right) \mu(dx) < \infty.
\]

Equation (C.22) follows from Equation (C.15) applying Cauchy-Schwarz inequality
\[
\int_0^\infty \int_0^\infty \int_t^T e^{-(x+y)(T-s)} ds \mu(dx) \mu(dy) \\
\leq \int_0^\infty \int_0^\infty \sqrt{\int_t^T e^{-2x(T-s)} ds} \sqrt{\int_t^T e^{-2y(T-s)} ds} \mu(dx) \mu(dy) \\
= \left(\int_0^\infty \sqrt{\int_t^T e^{-2x(T-s)} ds} \mu(dx) \right)^2 < \infty.
\]

In the same way Equation (C.23) follows from Equation (C.16)
\[
\int_0^\infty \int_0^\infty \int_t^T (T-s)^2 e^{-(x+y)(T-s)} ds \nu(dx) \nu(dy) \\
\leq \left(\int_0^\infty \sqrt{\int_t^T (T-s)^2 e^{-2y(T-s)} ds} \right)^2 < \infty.
\]

**Appendix D. Auxiliary results for Section 2**

**Lemma D.1** (Conditional moments of \((Y,Z)\)). For each \(x \in (0, \infty)\) and \(0 \leq t \leq T\), the process \((Y^x, Z^z)\) can be represented as
\[
\begin{align*}
Y_t^x &= Y_t^x e^{-(T-t)x} + \int_t^T e^{-(T-s)x} dW_s, \\
Z_t^x &= Z_t^x e^{-(T-t)x} + Y_t^x (T-t) e^{-(T-t)x} + \int_t^T e^{-(T-s)x} (T-s) dW_s.
\end{align*}
\]
The random variables \( Y^x_0 \) and \( Z^x_0 \) have conditional means given by

\[
E[Y^x_1|F_t] = Y^x_t e^{-(T-t)x}, \quad E[Z^x_1|F_t] = Z^x_t e^{-(T-t)x} + Y^x_t(T-t)e^{-(T-t)x}.
\]

Moreover, for \( x_1, x_2 \in (0, \infty) \) we have conditional covariances

\[
\text{Cov}(Y^x_1, Y^{x_2}_1|F_t) = \frac{1 - e^{-(T-t)(x_1+x_2)}}{x_1 + x_2},
\]
\[
\text{Cov}(Y^x_1, Z^{x_2}_1|F_t) = \frac{1 - e^{-(T-t)(x_1+x_2)}(1 + (T-t)(x_1 + x_2))}{(x_1 + x_2)^2},
\]
\[
\text{Cov}(Z^x_1, Z^{x_2}_1|F_t) = \frac{2 - e^{-(T-t)(x_1+x_2)}(2 + 2(T-t)(x_1 + x_2) + (T-t)^2(x_1 + x_2)^2)}{(x_1 + x_2)^3}.
\]

**Proof.** The representation in Equation (D.1) can be deduced from the SDE (2.2) for \((Y^x, Z^x)\) using Theorem A.1(ii)

\[
Z^x_t = Z^x_t e^{-(T-t)x} + \int_t^T e^{-(T-s)x} \left( Y^x_s e^{-(s-t)x} + \int_t^s e^{-(s-u)x} dW_u \right) ds
\]
\[
= Z^x_t e^{-(T-t)x} + Y^x_t(T-t)e^{-(T-t)x} + \int_t^T \int_s^T e^{-(T-s)u} dW_u ds
\]
\[
= Z^x_t e^{-(T-t)x} + Y^x_t(T-t)e^{-(T-t)x} + \int_t^T \int_u^T e^{-(T-s)u} ds dW_u
\]
\[
= Z^x_t e^{-(T-t)x} + Y^x_t(T-t)e^{-(T-t)x} + \int_t^T (T-u)e^{-(T-u)x} dW_u.
\]

The condition (A.2) is satisfied because \( \int_t^T \sqrt{\int_s^T e^{-2(T-t)u} duds} < \infty \). The conditional means can be read off directly from the representation of \((Y^x, Z^x)\). The formulas for the conditional covariances are obtained using Itô's isometry by calculating the following integrals

\[
\text{Cov}(Y^x_1, Y^{x_2}_1|F_t) = \int_t^T e^{-(T-s)(x_1+x_2)} ds,
\]
\[
\text{Cov}(Y^x_1, Z^{x_2}_1|F_t) = \int_t^T (T-s)e^{-(T-s)(x_1+x_2)} ds,
\]
\[
\text{Cov}(Z^x_1, Z^{x_2}_1|F_t) = \int_t^T (T-s)^2e^{-(T-s)(x_1+x_2)} ds.
\]

**Lemma D.2** (Integrability of \((Y, Z)\)). Let Assumption 2.3 be in place and assume \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)\) a.s. Then, for each \( t \geq 0, Y_t \in L^1(\mu) \) and \( Z_t \in L^1(\nu) \) holds with probability one.

**Proof.** By Lemma D.1 we have for \((Y^x, Z^x)\)

\[
Y^x_t = Y^x_0 e^{-tx} + \int_0^t e^{-(t-s)x} dW_s,
\]
\[
Z^x_t = Z^x_0 e^{-tx} + Y^x_0 te^{-tx} + \int_0^t (t-s)e^{-(t-s)x} dW_s.
\]
The deterministic parts are integrable because
\[
\int_0^\infty |Y_t^x|e^{-tx}\mu(dx) \leq \|Y_0\|_{L^1(\mu)} < \infty, \\
\int_0^\infty |Z_t^x|e^{-tx}\nu(dx) \leq \|Z_0\|_{L^1(\nu)} < \infty, \\
\int_0^\infty |Z_t^x|e^{-tx}\nu(dx) \leq \sup_{x \in (0, \infty)} (p(x)e^{-tx}) t \|Y_0\|_{L^1(\mu)} < \infty,
\]
where Assumption 2.3 is used in the last line. Therefore we can assume without loss of generality that \(Y_0, Z_0\) vanish. Then for each \(t \geq 0\),
\[
\mathbb{E} \left[ \|Y_t\|_{L^1(\mu)} \right] = \int_0^\infty \mathbb{E} \left[ \|Y_t^x\| \right] \mu(dx) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty \sqrt{\text{Var}(Y_t^x)} \mu(dx) < \infty,
\]
\[
\mathbb{E} \left[ \|Z_t\|_{L^1(\nu)} \right] = \int_0^\infty \mathbb{E} \left[ \|Z_t^x\| \right] \nu(dx) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^\infty \sqrt{\text{Var}(Z_t^x)} \nu(dx) < \infty,
\]
which follows from Equations (C.12) and (C.14). Therefore, \(Y_t \in L^1(\mu)\) and \(Z_t \in L^1(\nu)\) holds almost surely. □

**Lemma D.3** (Linear functionals of \((Y, Z)\)). Let Assumption 2.3 be in place and assume \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)\) a.s. Then the process \((Y, Z)\) satisfies for each \(0 \leq t \leq T\) and \((u, v) \in L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})\)
\[
\langle Y_t, u \rangle_\mu = \int_0^\infty Y_t^x e^{-(T-t)x} u(x) \mu(dx) + \int_t^T \int_0^\infty e^{-(T-t)x} u(x) \mu(dx) dW_s, \\
\langle Z_t, v \rangle_\nu = \int_0^\infty \left( Z_t^x e^{-(T-t)x} + Y_t^x (T-t) e^{-(T-t)x} \right) v(x) \nu(dx) \\
+ \int_t^T \int_0^\infty (T-s) e^{-(T-s)x} v(x) \nu(dx) dW_s.
\]
In particular, the random variable \(\langle Y_T, u \rangle_\mu + \langle Z_T, v \rangle_\nu\) is Gaussian, given \(\mathcal{F}_t\).

**Proof.** The statement follows from Lemmas D.1 and D.2 and from Theorem A.1. Condition (A.2) of the stochastic Fubini theorem are satisfied by Equations (C.15) and (C.16). □

**Lemma D.4** (Covariance operators). Let Assumption 2.3 be in place and \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu)\) a.s. Then for all \((u_1, u_2) \in L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C}), (v_1, v_2) \in L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})\) and for all \(0 \leq t \leq T\)
\[
\text{Cov} \left( \langle Y_T, u_1 \rangle_\mu, \langle Y_T, u_2 \rangle_\mu \bigg| \mathcal{F}_t \right) = \langle P_{T-t} u_1, u_2 \rangle_\mu, \\
\text{Cov} \left( \langle Z_T, v_1 \rangle_\mu, \langle Z_T, v_2 \rangle_\mu \bigg| \mathcal{F}_t \right) = \langle Q_{T-t} v_1, v_2 \rangle_\nu.
\]

207
where \( P_\tau : L^\infty(\mu; \mathbb{C}) \to L^1(\mu; \mathbb{C}) \) and \( Q_\tau : L^\infty(\nu; \mathbb{C}) \to L^1(\nu; \mathbb{C}) \) are bounded linear operators given by

\[
P_\tau u(x) = \int_0^\infty \frac{1 - e^{-\tau(x+y)}}{x+y} u(y) \mu(dy),
\]

\[
Q_\tau v(x) = \int_0^\infty \frac{2 - e^{-\tau(x+y)}(2 + 2\tau(x+y) + \tau^2(x+y)^2)}{(x+y)^3} v(y) \nu(dy),
\]

for \( u \in L^\infty(\mu; \mathbb{C}), v \in L^\infty(\nu; \mathbb{C}) \) and \( \tau \geq 0 \). In particular, \( Y_T \) and \( Z_T \) are Gaussian random variables, given \( \mathcal{F}_t \), with covariance operators \( P_{T-t} \) and \( Q_{T-t} \), respectively.

**Proof.** For each \( t \geq 0 \) and any \( u_{1,2} \in L^\infty(\mu) \) and \( v_{1,2} \in L^\infty(\nu) \) we have using the representation of Lemma D.3

\[
\text{Cov} \left( \left( Y_T, u_1 \right)_{\mu}, \left( Y_T, u_2 \right)_{\mu} \right) \mid \mathcal{F}_t \right) = \int_0^\infty \int_0^\infty \text{Cov} \left( Y^x_T, Y^y_T \mid \mathcal{F}_t \right) u_1(x) u_2(y) \mu(dy) \mu(dx) \leq \|u\|_{L^\infty(\mu)} \|v\|_{L^\infty(\mu)} < \infty,
\]

\[
\text{Cov} \left( \left( Z_T, v_1 \right)_{\nu}, \left( Z_T, v_2 \right)_{\nu} \right) \mid \mathcal{F}_t \right) = \int_0^\infty \int_0^\infty \text{Cov} \left( Z^x_T, Z^y_T \mid \mathcal{F}_t \right) v_1(x) v_2(y) \nu(dy) \nu(dx) \leq \|u\|_{L^\infty(\nu)} \|v\|_{L^\infty(\nu)} < \infty.
\]

By Equations (C.22) and (C.23) we have

\[
\|P_\tau u\|_{L^1(\mu)} = \int_0^\infty \int_0^\infty \int_0^x e^{-s(x+y)} |u(x)| ds \mu(dx) \mu(dy) \leq C \|u\|_{L^\infty(\mu)} < \infty,
\]

\[
\|Q_\tau v\|_{L^1(\nu)} = \int_0^\infty \int_0^\infty \int_0^x s^2 e^{-s(x+y)} |v(x)| ds \nu(dx) \nu(dy) \leq C \|v\|_{L^\infty(\nu)} < \infty,
\]

for some constant \( C \). The last two inequalities imply that \( P_\tau : L^\infty(\mu; \mathbb{C}) \to L^1(\mu; \mathbb{C}) \) and \( Q_\tau : L^\infty(\nu; \mathbb{C}) \to L^1(\nu; \mathbb{C}) \) are bounded linear operators. \( \square \)

**Lemma D.5 (Maximum inequality for OU processes).** There exists a constant \( C > 0 \) such that for each \( t \geq 0 \) and \( x > 0 \)

\[
\mathbb{E} \left[ \sup_{s \in [0,t]} |Y^x_s|^2 \right] \leq C \log(1 + tx)x^{-\frac{1}{2}},
\]

\[
\mathbb{E} \left[ \sup_{s \in [0,t]} |Z^x_s|^2 \right] \leq C \log(1 + tx)x^{-\frac{1}{2}}.
\]

**Proof.** The inequality for \( Y^x \) follows from the maximal inequalities for OU processes developed by Graversen and Peskir [39]. For the process \( Z^x \), we estimate for each
$t \geq 0$ and $x > 0$

$$\mathbb{E} \left[ \sup_{s \in [0,t]} |Z_s^x| \right] \leq \mathbb{E} \left[ \int_0^t e^{-((t-s)x^\frac{1}{2})} |Y_s^x| \, ds \right] \leq C \int_0^t e^{-((t-s)x^\frac{1}{2})} \log(1 + sx)x^{-\frac{1}{2}} \, ds$$

$$= C \left[ e^{-((t-s)x^\frac{1}{2})} \log(1 + sx)x^{-\frac{1}{2}} \right]_0^t - C \int_0^t e^{-((t-s)x^\frac{1}{2})} (1 + sx)^{-\frac{1}{2}} \, dt$$

$$= C \log(1 + tx)x^{-\frac{1}{2}} - C \int_0^t e^{-((t-s)x^\frac{1}{2})} (1 + sx)^{-\frac{1}{2}} \, ds$$

$$\leq C \log(1 + tx)x^{-\frac{1}{2}}.$$  

\[ \square \]

**Lemma D.6 (Auxiliary estimates for semimartingale decomposition).** Let $G(x, t)$ be deterministic and jointly measurable in $(x, t) \in (0, \infty) \times [0, \infty)$. Assume $Y_0 = Z_0 = 0$. Then, with probability one,

$$\text{Then, with probability one,} 
\int_0^\infty \int_0^t |G(x, t)Y_s^x(\omega)| \, dsm(\mu)(dx) \leq (1 \vee t^\frac{3}{2}) \int_0^\infty \int_0^t |G(x, s)Y_s^x(\omega)| \, dsm(\mu)(dx),$$

$$\int_0^\infty \int_0^t |G(x, s)Z_s^x(\omega)| \, dsv(\nu)(dx) \leq (1 \vee t^\frac{3}{2}) \int_0^\infty \int_0^t |G(x, s)Z_s^x(\omega)| \, dsv(\nu)(dx).$$

**Proof.** Note that for each $s \geq 0$ the random variables $|Y_s^x|$ and $|Z_s^x|$ are half-normal distributed with mean

$$\mathbb{E}[|Y_s^x|] = \frac{1 - e^{-2sx}}{\pi x}, \text{ and } \mathbb{E}[|Z_s^x|] = \sqrt{\frac{1 - e^{-2sx} (1 + 2sx + 2^2x^2)}{2\pi x^3}}.$$  

By (C.4) we have

$$\int_0^\infty \int_0^t \mathbb{E} \left[ |G(x, s)Y_s^x| \right] \, dsm(\mu)(dx) = \int_0^\infty \int_0^t |G(x, s)| \sqrt{\frac{1 - e^{-2sx}}{\pi x}} \, dsm(\mu)(dx)$$

$$\leq (1 \vee t^\frac{3}{2}) \int_0^\infty \int_0^t |G(x, s)(1 \vee x^{-\frac{1}{2}}) \, dsm(\mu)(dx).$$

By (C.6) we have

$$\int_0^\infty \int_0^t \mathbb{E} \left[ |G(x, s)Z_s^x| \right] \, dsv(\nu)(dx)$$

$$= \int_0^\infty \int_0^t |G(x, s)| \sqrt{\frac{1 - e^{-2sx} (1 + 2sx + 2^2x^2)}{2\pi x^3}} \, dsv(\nu)(dx)$$

$$\leq (1 \vee t^\frac{3}{2}) \int_0^\infty \int_0^t |G(x, s)(1 \vee x^{-\frac{1}{2}}) \, dsv(\nu)(dx).$$

Then the inequalities hold true with probability one.  

\[ \square \]

**Lemma D.7 (Tightness).** Let $\mu_\infty, \nu_\infty$ satisfy Assumption 2.14. Then the laws of the random variables $(Y_t, Z_t)_{t \geq 0}$ are tight on the space $L^1(\mu_\infty) \times L^1(\nu_\infty)$ with the weak topology.

**Proof.** We generalize the proof of [14, Proposition 2] to our setting. We endow $L^1(\mu_\infty) \times L^1(\nu_\infty)$ with the weak topology and assume that $(Y_0, Z_0) = 0$. We will show using [31, Theorem IV.8.9] that for any $M \geq 0$, the set

$$K_M = \left\{ (y, z) \in L^1(\mu_\infty) \times L^1(\nu_\infty) : \|y\|_{L^2(\mu_\infty)}^2 + \|z\|_{L^2(\nu_\infty)}^2 \leq M \right\}$$

209
is pre-compact in \( L^1(\mu_\infty) \times L^1(\nu_\infty) \). For any measurable set \( E \subseteq [0, \infty) \) and \((y,z) \in K_M\), the Cauchy-Schwartz inequality implies
\[
\|1_Ey\|_{L^1(\mu_\infty)} \leq \|y\|_{L^2(\mu_\infty^{1/2})} \|1_E\|_{L^2(\mu_\infty^{1/2})},
\]
\[
\|1_Ez\|_{L^1(\mu_\infty)} \leq \|z\|_{L^2(\mu_\infty^{1/2})} \|1_E\|_{L^2(\mu_\infty^{1/2})}.
\]
Setting \( E = [0, \infty) \) shows that \( K_M \) is bounded in \( L^1(\mu_\infty) \times L^1(\nu_\infty) \). Moreover, if \( E_n \subseteq [0, \infty) \) is a sequence of measurable sets which decreases to the empty set, then the above estimate shows that
\[
\lim_{n \to \infty} \sup_{(y,z) \in K} \|1_{E_n}y\|_{L^1(\mu_\infty)} + \|1_{E_n}z\|_{L^1(\mu_\infty)} = 0.
\]
Therefore, the conditions of [31, Theorem IV.8.9] are satisfied and \( K_M \) is pre-compact.

By Prokhorov’s theorem, the laws of \((Y_t, Z_t)_{t \geq 0}\) are tight if
\[
\lim_{M \to \infty} \sup_{t \geq 0} Q[(Y_t, Z_t) \notin K_M] = 0.
\]
This follows from the estimate
\[
Q[(Y_t, Z_t) \notin K_M] \leq \frac{1}{M} E \left[ \|Y_t\|^2_{L^2(\mu_\infty^{1/2})} + \|Z_t\|^2_{L^2(\nu_\infty^{1/2})} \right] = \frac{1}{M} \left( \int_0^\infty \text{Cov}(Y_t^x) \sqrt{\mu_\infty}(dx) + \text{Cov}(Z_t^x) \sqrt{\nu_\infty}(dx) \right) = \frac{1}{M} \left( \int_0^\infty \sqrt{\mu_\infty}(dx) + \int_0^\infty \frac{1}{4x^2} \sqrt{\nu_\infty}(dx) \right),
\]
where the right-hand side is finite by Assumption 2.14.

\section*{Appendix E. Auxiliary results for Section 4.2}

\textbf{Lemma E.1} (Integrability condition). Under Assumption 4.7, the following condition is satisfied:
\[
\sup_{x \in (0, \infty)} p(x) \int_0^t e^{-sx} ds < \infty.
\]

\textit{Proof.} By assumption, there is \( \beta \in (0, 2) \) such that \( p(x)(1 \wedge x^{-\beta}) \) is bounded in \( x \).

Then the lemma follows from the estimate
\[
\int_0^t e^{-sx} ds \leq \int_0^t s \left( 1 \vee \left( \frac{s}{3} \right)^{-\beta} \right) (1 \wedge x^{-\beta}) ds \leq \int_0^t \left( s + \left( \frac{s}{3} \right)^{1-\beta} \right) (1 \wedge x^{-\beta}) ds = \left( \frac{t^2}{2} + \frac{1}{2-\beta} \left( \frac{t}{3} \right)^{2-\beta} \right) (1 \wedge x^{-\beta}).
\]

\textbf{Lemma E.2} (Time-integrals of \((Y, Z)\)). Let Assumption 4.7 be in place and assume \((Y_0, Z_0) \in L^1(\mu) \times L^1(\nu) \) a.s. Then, for each \( 0 \leq t \leq T \) and for all \((u, v) \in \)
$L^\infty(\mu; \mathbb{C}) \times L^\infty(\nu; \mathbb{C})$ one has

$$\int_t^T \left( \langle Y_s, u \rangle_\mu + \langle Z_s, v \rangle_\nu \right) ds = -\langle Y_t, \Phi_1(T-t, u, v) \rangle_\mu - \langle Z_t, \Phi_2(T-t, u, v) \rangle_\nu$$

with $\Phi_1, \Phi_2$ as in Theorem 4.10. In particular, the random variable $\int_t^T \left( \langle Y_s, u \rangle_\mu + \langle Z_s, v \rangle_\nu \right) ds$ is Gaussian, given $\mathcal{F}_t$.

**Proof.** The time-derivatives of $\Phi_1, \Phi_2$ are given by

$$\partial_\tau \Phi_1(\tau, u, v)(x) = -e^{-\tau x}(u(x) + \tau p(x)v(x)), \quad \partial_\tau \Phi_2(\tau, u, v)(x) = -e^{-\tau x}v(x).$$

It follows from Lemma D.3 that for any $0 \leq t \leq s$,

$$\langle Y_s, u \rangle + \langle Z_s, v \rangle = -\langle Y_t, \partial_\tau \Phi_1(s-t, u, v) \rangle_\mu - \langle Z_t, \partial_\tau \Phi_2(s-t, u, v) \rangle_\mu$$

$$- \int_t^s \langle \partial_\tau \Phi_1(s-r, u, v), 1 \rangle_\mu dW_r.$$

The result follows by integrating over $s \in [t, T]$ and applying Fubini’s theorem (Theorem A.1) to each of the three summands above. For the first summand, Condition (A.1) of Theorem A.1 is satisfied by Lemma E.1 and the estimate

$$\int_0^\infty \int_t^T |Y^2_t \partial_\tau \Phi_1(s-t, u, v)| ds \mu(dx)$$

$$\leq ||u||_{L^\infty(\mu)} ||Y^2_t||_{L^1(\mu)} + ||v||_{L^\infty(\nu)} \int_0^\infty |Y^2_t| \int_t^T (s-t)e^{-(s-t)\tau} ds \mu(dx)$$

$$= ||u||_{L^\infty(\mu)} ||Y_t||_{L^1(\mu)} + ||v||_{L^\infty(\nu)} \int_0^\infty |Y^2_t| \int_0^{T-t} se^{-sx} ds \mu(dx) < \infty.$$

For the second summand, Condition (A.1) reads as

$$\int_0^\infty \int_t^T |Z^x_t e^{-(s-t)x}v(x)| ds \nu(dx) \leq (T-t)||v||_{L^\infty(\nu)} ||Z_t||_{L^1(\nu)} < \infty.$$

For the third summand, we first use Fubini’s theorem to exchange the order of integration with respect to $\mu(dx)$ and $dW_r$:

$$\int_s^t \langle \partial_\tau \Phi_1(s-r, u, v), 1 \rangle_\mu dW_r = -\int_0^\infty \int_t^s e^{-(s-r)x}(u(x)+(s-r)p(x)v(x)) dW_r \mu(dx).$$

This is allowed because Equation (A.2) is satisfied by Equations (C.12) and (C.16):

$$\int_0^\infty \sqrt{\int_t^s e^{-2(s-r)x} |u(x)| \mu(dx)} < \infty,$$

$$\int_0^\infty \sqrt{\int_t^s (s-r)^2 e^{-2(s-r)x} |v(x)| \nu(dx)} < \infty.$$
Then we interchange the order of integration with respect to \(dW_r\) and the product measure \(\mu(dx)ds\), which brings the third summand into the form

\[
- \int_t^T \int_0^\infty e^{-(s-r)x} (u(x) + (s-r)p(x)v(x)) dW_r \mu(dx) ds \\
= - \int_t^T \int_0^T \int_0^\infty e^{-(s-r)x} (u(x) + (s-r)p(x)v(x)) \mu(dx) ds dW_r.
\]

This is allowed because Condition (A.2) is satisfied by Equations (C.24) and (C.25):

\[
\int_t^T \int_0^\infty \sqrt{\int_t^s e^{-(s-r)x} u(x) |\mu(dx)| ds} < \infty,
\]

\[
\int_t^T \int_0^\infty \sqrt{\int_t^s (s-r)^2 e^{-(s-r)x} v(x) |\nu(dx)| ds} < \infty.
\]

Finally, we exchange the innermost integrals \(\mu(dx)\) and \(ds\), which is justified by Condition (A.1) and Equations (C.17) and (C.18). Then the third summand is given by

\[
- \int_t^T \int_r^\infty \int_r^T e^{-(s-r)x} (u(x) + (s-r)p(x)v(x)) ds \mu(dx) dW_r \\
= - \int_t^T \langle \Phi_1(T - r, u, v), 1 \rangle \mu dW_r. \quad \Box
\]

**Lemma E.3** (Semimartingale property). Under Assumption 4.7, the expressions \(\langle Y_t, \Phi_1(T - t, u, v) \rangle_{\mu}\) and \(\langle Z_t, \Phi_2(T - t, u, v) \rangle_{\nu}\) are continuous semimartingales in \(t \in [0, T]\), for each fixed \(T > 0\) and \((u, v) \in L^\infty(\mu) \times L^\infty(\nu)\).

**Proof.** We verify the conditions of Theorem 2.13. In the following estimates it can be assumed without loss of generality that the functions \(u\) and \(v\) are equal to 1 because they are bounded. Conditions (2.9) and (2.10) for \(f^*_x = \Phi_1(T - t, u, v)(x)\) are satisfied by Equations (C.19), (C.20) and (C.21):

\[
\int_0^\infty \int_0^t |\partial_s f^*_x - x f^*_x| (1 \wedge x^{-\frac{1}{2}}) d\mu(dx) \\
= \int_0^\infty \int_0^t \left(1 + \frac{1 - e^{-(t-s)x}}{x} p(x)\right) (1 \wedge x^{-\frac{1}{2}}) d\mu(dx) < \infty,
\]

\[
\int_0^\infty \sqrt{\int_0^t (f^*_x)^2 d\mu(dx)} \leq \int_0^\infty \sqrt{\int_0^t \frac{e^{-(t-s)x} - 1}{x^2} ds} d\mu(dx) \\
+ \int_0^\infty \sqrt{\int_0^t \frac{e^{-(t-s)x} - 1}{x^2} + \frac{r e^{-(t-s)x}}{x^2}} ds d\nu(dx) < \infty.
\]

Conditions (2.11) and (2.12) are satisfied for \(g^*_x = \Phi_2(T - t, u, v)(x)\) by Equation (C.19):

\[
\int_0^\infty \int_0^t |\partial_s g^*_x - x g^*_x| (1 \wedge x^{-\frac{1}{2}}) d\nu(dx) = \int_0^\infty \int_0^t (1 \wedge x^{-\frac{1}{2}}) d\nu(dx) < \infty,
\]

\[
\int_0^\infty |g^*_x| (1 \wedge x^{-\frac{1}{2}}) d\nu(dx) = \int_0^\infty \int_0^t \frac{1 - e^{-(t-s)x}}{x} (1 \wedge x^{-\frac{1}{2}}) d\nu(dx) < \infty.
\]
Thus, we have verified the conditions of Theorem 2.13 and the statement of the lemma follows. □

Lemma E.4 (Semimartingale property). Under Assumption 4.7, the expressions 
\( \langle Y_t, \partial_r \Phi_1(\tau, u, v) \rangle \), and \( \langle Z_t, \partial_r \Phi_2(\tau, u, v) \rangle \), are continuous semimartingales in \( t \in [0, T] \), for each fixed \( \tau > 0 \) and \( (u, v) \in L^\infty(\mu) \times L^\infty(\nu) \).

Proof. We calculate
\[
\partial_r \Phi_1(\tau, u, v)(x) = -e^{-\tau x}(u(x) + \tau p(x)v(x)), \quad \partial_r \Phi_2(\tau, u, v)(x) = -e^{-\tau x}v(x).
\]

We show the semimartingale property by verifying the conditions of Theorem 2.13. In the following estimates it can be assumed without loss of generality that the functions \( u \) and \( v \) are equal to 1 because they are bounded. Conditions (2.9) and (2.10) for \( f^*_t = \partial_r \Phi_1(\tau, u, v)(x) \) are satisfied by Equations (C.8)–(C.11):
\[
\int_0^\infty \int_0^t |\partial_s f^*_s - x f^*_s| (1 \wedge x^{-\frac{1}{2}})ds\mu(dx) \\
= \int_0^\infty \int_0^t xe^{-\tau s}(1 + \tau p(x))(1 \wedge x^{-\frac{1}{2}})ds\mu(dx) < \infty,
\]
\[
\int_0^\infty \sqrt{\int_0^t (f^*_s)^2 ds\mu(dx)} \leq \int_0^\infty \sqrt{\int_0^t 2e^{-2\tau s}ds\mu(dx)} \\
+ \int_0^\infty \sqrt{\int_0^t 2\tau^2 e^{-2\tau s}ds\nu(dx)} < \infty.
\]

Conditions (2.11) and (2.12) for \( g^*_t = \partial_r \Phi_2(\tau, u, v)(x) \) are satisfied by Equation (C.11):
\[
\int_0^\infty \int_0^t |\partial_s g^*_s - x g^*_s| (1 \wedge x^{-\frac{1}{2}})ds\nu(dx) = \int_0^\infty \int_0^t xe^{-\tau s}(1 \wedge x^{-\frac{1}{2}})ds\nu(dx) < \infty,
\]
\[
\int_0^\infty \int_0^t |g^*_s| (1 \wedge x^{-\frac{1}{2}})ds\nu(dx) = \int_0^\infty \int_0^t xe^{-\tau s}(1 \wedge x^{-\frac{1}{2}})ds\nu(dx) < \infty.
\]

Thus, we have verified the conditions of Theorem 2.13 and the statement of the lemma follows. □

Appendix F. Auxiliary results for Section 4.3

Lemma F.1 (Semimartingale property). Under Assumption 2.3, the expressions 
\( \langle Y_t, \partial_r \phi_1(\tau, -u, -v) \rangle \), and \( \langle Z_t, \partial_r \phi_2(\tau, -u, -v) \rangle \), are continuous semimartingales in \( t \in [0, \infty) \) for each fixed \( \tau > 0 \) and \( (u, v) \in L^\infty(\mu) \times L^\infty(\nu) \).

Proof. We verify the conditions of Theorem 2.13. As \( u \) and \( v \) are bounded we may assume without loss of generality in the following estimates that \( u = v = 1 \). Conditions (2.9)–(2.12) for \( f^*_t = \partial_r \phi_1(\tau, -u, -v)(x) \) and \( g^*_t = \partial_r \phi_2(\tau, -u, -v)(x) \)
are satisfied by Equations (C.8)–(C.11):
\[
\int_0^\infty \int_0^t |\partial_s f_s x - x f_s x| (1 \wedge x^{-\frac{1}{2}}) ds \mu(dx) = t \int_0^\infty |x \partial_x \phi_1(\tau, -u, -v)| (1 \wedge x^{-\frac{1}{2}}) \mu(dx)
\leq t \int_0^\infty x^2 e^{-x^2 \tau} \mu(dx) + t \int_0^\infty x e^{-x^2 \tau} \nu(dx) + t t \int_0^\infty x^2 e^{-x^2 \tau} \nu(dx) < \infty,
\]
\[
\int_0^\infty \sqrt{t} (f_s^2) ds \mu(dx) = \sqrt{t} \int_0^\infty |\partial_x \phi_2(\tau, -u, -v)| \mu(dx)
\leq \sqrt{t} \int_0^\infty x e^{-x^2 \tau} \mu(dx) + \sqrt{t} \int_0^t e^{-x^2 \tau} \nu(dx) + \sqrt{t} \int_0^t x e^{-x^2 \tau} \nu(dx) < \infty,
\]
\[
\int_0^\infty \int_0^t |\partial_s g_s x - x g_s x| (1 \wedge x^{-\frac{1}{2}}) ds \nu(dx) = t \int_0^\infty |x \partial_x \phi_2(\tau, -u, -v)| (1 \wedge x^{-\frac{1}{2}}) \nu(dx)
= t \int_0^\infty x^2 e^{-x^2 \tau} \nu(dx) < \infty,
\]
\[
\int_0^\infty \int_0^t |g_s^2| (1 \wedge x^{-\frac{1}{2}}) ds \nu(dx) = t \int_0^\infty |\partial_x \phi_2(\tau, -u, -v)| \nu(dx)
\leq t \int_0^\infty e^{-x^2 \tau} \nu(dx) < \infty.
\]

\[\square\]

**Appendix G. Auxiliary results for Section 5**

**Lemma G.1** (Injectivity of the covariance operator). For any \( \tau > 0 \), the mapping \( P_\tau \) is an injective linear operator from \( L^\infty(\mu; \mathbb{C}) \) to the complexification of the Hilbert space \( H_\tau \).

**Proof.** For simplicity, we write \( H_\tau \) for the complexified space \( H_\tau \otimes \mathbb{R} \mathbb{C} \) (see Appendix B). If \( P_\tau v = 0 \) for some \( v \in L^\infty(\mu; \mathbb{C}) \), then
\[
0 = (P_\tau v, P_\tau v)_{H_\tau} = (P_\tau v, v)_{\mu} = \int_0^\tau \left( \int_0^\infty v(x) e^{-x^2 \tau} \mu(dx) \right) ds.
\]
Therefore, the Laplace transform \( \mathcal{L}(v\mu)(s) \) of the complex measure \( v\mu \) vanishes at almost all \( s \in [0, \tau] \). As \( \mathcal{L}(v\mu)(s) \) is analytic in \( s \), it vanishes identically. By the injectivity of the Laplace transform [52, Section 3.8], the complex measure \( v\mu \) vanishes, which is equivalent to \( v = 0 \) in \( L^\infty(\mu; \mathbb{C}) \).

**Lemma G.2** (Diagonalization of symmetric two-tensors). For each \( \tau \geq 0 \) any symmetric two-tensor \( w \in L^\infty(\mu; \mathbb{C}) \otimes^2 \) has a representation as a sum of squares
\[
w = \sum_{k=1}^n \partial_k v_k \otimes v_k, \quad \text{with } \partial_k \in \mathbb{C} \text{ and } v_k \in L^\infty(\mu; \mathbb{C}),
\]
such that the functions \( v_k \) are orthonormal with respect to the covariance operator \( P_\tau \) defined in Lemma D.4, i.e., \( (P_\tau v_k, v_l)_{\mu} = \delta_{kl} \).

**Proof.** For simplicity, we write \( H_\tau \) for the complexified space \( H_\tau \otimes \mathbb{R} \mathbb{C} \). Let \( w = \sum_{k=1}^n w_k \otimes w_k \in L^\infty(\mu; \mathbb{C}) \otimes^2 \) be any symmetric two-tensor and set \( V = \text{span}_\mathbb{C}\{w_1, \ldots, w_n\} \). By Lemma G.1, the bilinear form \( (P_\tau, \cdot) \) is a scalar product on the finite-dimensional vector space \( V \). The desired representation of \( w \) is obtained by diagonalizing \( w \in V \otimes^2 \) with respect to this scalar product.

\[\square\]
Lemma G.3 (Affine structure). Let $\mu$ satisfy Assumption 2.3 and $Y_0 \in L^1(\mu)$. Let $w = \sum_{k=1}^{n} \vartheta_k v^2_k \in iL^\infty(\mu) \otimes 2$ be a symmetric tensor with decomposition into sums of squares in the sense of Lemma G.2, and $0 \leq t \leq T$.

$$E \left[ e^{(T, w)_{\mu \otimes 2}} | F_t \right] = e^{\psi_0(T-t, w) + \langle \Pi_t, \psi_1(T-t, w) \rangle_{\mu \otimes 2}},$$

where $(\psi_0, \psi_1) : [0, \infty) \times L^\infty(\mu; \mathbb{C}) \otimes 2 \to \mathbb{C} \times L^\infty(\mu; \mathbb{C}) \otimes 2$ are given by

$$\psi_0(\tau, w) = \frac{1}{2} \sum_{k=1}^{n} \log(1 - 2\vartheta_k),$$

$$\psi_1(\tau, w)(x, y) = \sum_{k=1}^{n} \frac{\vartheta_k}{1 - 2\vartheta_k} v_k(x)v_k(y)e^{-(T-t)(x+y)}.$$

Proof. Let $0 \leq t \leq T$ be fixed and let $w = \sum_{k=1}^{n} \vartheta_k v^2_k$ be a decomposition of $w$ into sums of squares in the sense of Lemma G.2. By Lemmas D.3, D.4, and G.2 the random variables $\langle Y_1, v_1 \rangle_{\mu}, \ldots, \langle Y_r, v_r \rangle_{\mu}$ are independent Gaussian, given $F_t$, with conditional means

$$E \left[ \langle Y, v \rangle_{\mu} | F_t \right] = \langle Y, \phi_1(T-t, v, \bot) \rangle_{\mu}, \quad k \in \{1, \ldots, n\},$$

and unit variances. Hence, the random variables $\langle Y, v \rangle_{\mu}^2, \ldots, \langle Y, v \rangle_{\mu}^2$ are independent non-central $\chi^2$, given $F_t$, with non-centrality parameters

$$\langle Y, \phi_1(T-t, v, \bot) \rangle_{\mu \otimes 2}^2 = \langle \Pi_t, \phi_1(T-t, v, \bot) \rangle_{\mu \otimes 2}^2, \quad k \in \{1, \ldots, n\}.$$

We obtain the affine transformation formula using independence and the characteristic function of the non-central $\chi^2$ distribution

$$E \left[ e^{(T, w)_{\mu \otimes 2}} | F_t \right] = \prod_{k=1}^{n} E \left[ e^{\vartheta_k \langle Y, v \rangle_{\mu}^2} | F_t \right] = \exp \left( \frac{1}{2} \sum_{k=1}^{n} \log(1 - 2\vartheta_k) \right) \times e^{\sum_{k=1}^{n} \frac{\vartheta_k}{1 - 2\vartheta_k} \langle \Pi_t, \phi_1(T-t, v, \bot) \rangle_{\mu \otimes 2}^2}.$$

We recognize the functions $\psi_0$ and $\psi_1$ on the right-hand side above. \hfill \Box

Lemma G.4 (Conditional mean). For each $v_{\otimes 2} \in L^\infty(\mu) \otimes_2 L^\infty(\mu)$ and $0 \leq t \leq T$, the $F_t$-conditional mean of $\langle \Pi_T, v_{\otimes 2} \rangle_{\mu \otimes 2}$ is given by

$$E \left[ \langle \Pi_T, v_{\otimes 2} \rangle_{\mu \otimes 2} | F_t \right] = 2\phi_0(\tau, v, \bot) + \langle \Pi_t, \phi_1(\tau, v, \bot) \rangle_{\mu \otimes 2}.$$

Proof. We use Theorem 5.1 to calculate

$$E \left[ \langle \Pi_T, v_{\otimes 2} \rangle_{\mu \otimes 2} | F_t \right] = \frac{1}{t} \partial_{|q|=0} E \left[ e^{\vartheta_0(T-t, v_{\otimes 2} q) + \langle \Pi_t, \psi_1(T-t, v_{\otimes 2} q) \rangle_{\mu \otimes 2}} | F_t \right]$$

$$= \frac{1}{t} \partial_{|q|=0} e^{\vartheta_0(T-t, v_{\otimes 2} q) + \langle \Pi_t, \phi_1(T-t, v_{\otimes 2} q) \rangle_{\mu \otimes 2}}$$

$$= \frac{1}{t} \partial_{|q|=0} e^{-\frac{1}{2} \log(1 - 4\phi_0(\tau, v, \bot)) + \langle \Pi_t, \phi_1(\tau, v, \bot) \rangle_{\mu \otimes 2}}$$

$$= \frac{1}{t} \partial_{|q|=0} e^{-\frac{1}{2} \log(1 - 4\phi_0(\tau, v, \bot)) + \langle \Pi_t, \phi_1(\tau, v, \bot) \rangle_{\mu \otimes 2}}$$

$$= 2\phi_0(\tau, v, \bot) + \langle \Pi_t, \phi_1(\tau, v, \bot) \rangle_{\mu \otimes 2}. \hfill \Box$$

215
Lemma G.5 (Conditional second moment). Let \( w \in L^\infty(\mu) \otimes L^\infty(\mu) \) be a symmetric tensor with sum-of-squares representation

\[
w = \sum_{k=1}^{n} \theta_k v_k^2.
\]
as in Lemma G.2. Then for each \( 0 \leq t \leq T \), the \( \mathcal{F}_t \)-conditional second moment of \( \langle \Pi_t, w \rangle_{\mu^\otimes 2} \) is given by

\[
\begin{align*}
\mathbb{E} \left[ \langle \Pi_t, w \rangle_{\mu^\otimes 2}^2 \right] &= 2 \sum_{k=1}^{n} \theta_k^2 + 2 \sum_{k=1}^{n} \theta_k^2 \langle \Pi_t, \phi_1(t, v_k, 0) \rangle^2_{\mu^\otimes 2} \\
&\quad + \left( \sum_{k=1}^{n} \theta_k + \sum_{k=1}^{n} \theta_k \langle \Pi_t, \phi_1(t, v_k, 0) \rangle_{\mu^\otimes 2} \right)^2.
\end{align*}
\]

Proof. As in the proof of Lemma G.3, we have

\[
\psi_0(T - t, iqw, 0) = -\frac{1}{2} \sum_{k=1}^{n} \log (1 - 2i q \theta_k),
\]

\[
\psi_1(T - t, iqw, 0)(x, y) = e^{-(T-t)(x+y)} \sum_{k=1}^{n} \frac{i q \theta_k}{1 - 2i q \theta_k} v_k(x) v_k(y).
\]

For the derivatives of \( \psi_0(T - t, iqw, 0) \) with respect to \( q \) we have

\[
\partial_q \psi_0(T - t, iqw, 0) = i \sum_{k=1}^{n} \frac{\theta_k}{1 - 2i q \theta_k},
\]

\[
\partial_q^2 \psi_0(T - t, iqw, 0) = -2 \sum_{k=1}^{n} \frac{\theta_k^2}{(1 - 2i q \theta_k)^2},
\]

and for the derivatives of \( \psi_1(T - t, iqw, 0) \) with respect to \( q \) we have

\[
\partial_q \psi_1(T - t, iqw, 0)(x, y) = i e^{-(T-t)(x+y)} \sum_{k=1}^{n} \frac{\theta_k}{1 - 2i q \theta_k} v_k(x) v_k(y),
\]

\[
\partial_q^2 \psi_1(T - t, iqw, 0)(x, y) = -2 e^{-(T-t)(x+y)} \sum_{k=1}^{n} \frac{\theta_k^2}{(1 - 2i q \theta_k)^2} v_k(x) v_k(y).
\]

Using the characteristic function we obtain

\[
\begin{align*}
\mathbb{E} \left[ \langle \Pi_t, w \rangle_{\mu^\otimes 2}^2 \right] &= -\partial_q^2 \psi_0(0) = 0 \mathbb{E} \left[ e^{iq \langle \Pi_t, w \rangle_{\mu^\otimes 2}} \right] \\
&= 2 \sum_{k=1}^{n} \theta_k^2 + 2 \sum_{k=1}^{n} \theta_k^2 \langle \Pi_t, \phi_1(t, v_k, 0) \rangle^2_{\mu^\otimes 2} \\
&\quad + \left( \sum_{k=1}^{n} \theta_k + \sum_{k=1}^{n} \theta_k \langle \Pi_t, \phi_1(t, v_k, 0) \rangle_{\mu^\otimes 2} \right)^2. \quad \Box
\end{align*}
\]
Cumulative Bibliography


<table>
<thead>
<tr>
<th>Number</th>
<th>Author(s)</th>
<th>Title</th>
<th>Year</th>
<th>Reference</th>
</tr>
</thead>
</table>


Curriculum Vitae

PERSONAL DATA

Name: David Stefanovits
Date of Birth: January 1, 1986
Nationality: Swiss

EDUCATION

Ph.D. studies in mathematics, ETH Zurich.
Supervisor: Prof. Dr. Mario V. Wüthrich.

Undergraduate Studies: 10/2005 – 06/2010
Master of Science ETH in Applied Mathematics.

High School: 09/2001 – 06/2005
Matura at Liceo Cantonale Lugano I.

EMPLOYMENT

ETH Zurich: 06/2014 – 04/2016
Research and teaching assistant.

1741 Asset Management Ltd, Zurich: 01/2012 – 06/2014
Quantitative research.

Quantitative research in the asset management division.

Internship.

Teaching assistant.