UiO : Department of Mathematics
University of Oslo

A functional approach to forward rate modelling

Kim André Arntsen

Master's Thesis, Spring 2020

This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Finance, Insurance and Risk*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Acknowledgements

I wish to express my most profound appreciation to my supervisor Fred Espen Benth for being an inexhaustible source of knowledge, in addition to his encouraging and positive disposition. I would also like to extend my thanks to my family and friends for their support.

Abstract

Benth and Süss recently published a paper where they proposed a continuous-time cointegration paradigm in a finite and infinite-dimensional framework. Motivated by their article, we present a Hilbert-valued multi-market forward rate model satisfying the Heath-Jarrow-Morton equation. Also, we give some insight into the no-arbitrage condition in terms of the covariance operator in the so-called Filipovic space. Moreover, we perform an empirical study of the Norwegian yield curve as Nelson-Siegel smoothed government bond observations in a functional data analysis setting. In particular, we carry out a functional Kwiatkowski—Phillips—Schmidt-–Shin test of stationarity following the lines of Horváth, Kokoszka and Rice [24].

Contents

1 Introduction

The topic of this thesis is to model the term structure of interest rates as a curve taking values in a separable Hilbert space. The aim of this thesis is twofold. Firstly, we construct a framework for studying forward interest rates and cointegration in the Heath-Jarrow-Morton framework, loosely based on the paper by Benth and Süss [4]. Secondly, we introduce the functional data analysis (FDA) and perform an empirical study in terms of functional data analysis. In particular, we carry out a functional Kwiatkowski–Phillips–Schmidt–Shin (fKPSS) test of stationarity on Norwegian government bond data as Nelson-Siegel smoothed curves in a functional time-series framework, following the lines of Hórvath, Kokoszka, and Rice [24].

At the intersection of stochastic analysis and functional inference, we are concerned with the inherent infinite dimensionality of continuous functions, signifying the necessity of functional analysis in both stochastic analysis, probability, and even statistics. As such, there has been a tremendous development in the field over last century, from the groundbreaking early works of Volterra, Gateaux, Frechet, and Lévy [1], to the introduction of the axiomatic probability theory by Kolmogorov in the 1930s [30]. Subsequently, Itô introduced the stochastic integral and formulated the so-called Itô calculus, which is considered the foundation of stochastic analysis. To this end, we provide the reader with a swift yet comprehensive introduction to the infinite-dimensional stochastic analysis.

HJM-modelling and forward curves

Our main theoretical concern in the modelling part, is the function-valued stochastic differential equation

$$\mathrm{d}X(t) = (AX(t) + f(X(t)))\mathrm{d}t + B(X(t))\mathrm{d}Z(t), \quad X(t_0) = X(0) \tag{1.1}$$

where A is a densely defined unbounded operator which generates a C_0-semigroup on a Hilbert space H, $f(X(t))$ an H-valued function, $B(X(t))$ a linear operator, and a $Z(t)$ an U-valued noise process, where $X(0)$ is a $\mathcal{F}_0$-measurable random variable.

[1] See the preface of Huang and Yan [26] for a rigorous introduction to the history of function valued stochastic processes

Such an equation is thoroughly studied in Peszat and Zabczyk [37], for $Z(t) = L(t)$ being a square integrable Lévy process, Carmona and Tehranchi [13] for $Z(t) = W(t)$, a cylindrical defined Wiener process, and in Prato and Zabczyk [14] with $Z(t) = W(t)$ a Q-Wiener process.

Let H and U denote two different separable Hilbert spaces. We will build a forward rate model which evolves according to a Heath-Jarrow-Morton model,

$$\mathrm{d}f(t,T) = \alpha(t,T)\mathrm{d}t + \langle \sigma(t,T), \mathrm{d}Z(t)\rangle_{\mathsf{U}}.$$

Here, α is H-valued and σ is $B(\mathsf{U},\mathsf{H})$-valued, such that the forward curves are H-valued. In order to obtain arbitrage-free dynamics of the forward rates, we must ensure that the drift of the process, satisfies the so-called *HJM-condition*, which is not a straightforward task. By arbitrage-free dynamics, we mean that the discounted bond price process

$$\hat{P}(t,T) = \exp\left(-\int_0^t r(s)\mathrm{d}s\right) P(t,T), \tag{1.2}$$

must be a local martingale under a risk neutral probability measure $\mathbb{Q}$. Here $r(s)$ denotes the *short rate process.* We will, to some extent avoid the problem of changing measures with the aid of well-established results on the no-arbitrage condition. A surprising link between stochastic partial differential equations and the HJM equation emerged when Musiela [33] parametrized the curves in time to maturity instead of the maturity time, i.e., we set $x = T - t$, such that $T = t + x$. The forward rate dynamics is then,

$$\mathrm{d}f(t,t+\cdot) = (\frac{\partial}{\partial x} f(t,t+\cdot) + \alpha(t,t+\cdot))\mathrm{d}t + \sigma(t,t+\cdot)\mathrm{d}Z(t), \tag{1.3}$$

which puts us in the same situation as (1.1). Here $\frac{\partial}{\partial x}$ denotes the partial derivative, which generates a C_0-semigroup under certain conditions on H.

Now for the purpose of modeling forward rates, our state-space of choice will be the Filipovic space H_w (see Filipovic [17]), a weighted Sobolev-type of space, defined by functions satisfying

$$\|f\|_w^2 = f^2(0) + \int_0^\infty w(x)(f'(x))^2\mathrm{d}x < \infty, \tag{1.4}$$

where f' is the weak derivative of f, $w(0) = 1$ and $\int_0^\infty w(x)^{-1}\mathrm{d}x < \infty$. This space has a plenitude of pleasant properties with regards to forward rate modelling in Hilbert spaces, and is therefore the preferred state-space, and noise space of choice later on.

We end the first part of the thesis by proposing a multi-market forward rate model where the noise in the forward curves originates from a sum of an idiosyncratic terms, $Y(t), Z(t)$ and a shared term $X(t)$. That is, we let f_1 and f_2 denote two forward rates from different markets and define

$$f_1(t,t+\cdot) = X(t,t+\cdot) + Y(t,t+\cdot) \tag{1.5}$$

$$f_2(t,t+\cdot) = X(t,t+\cdot) + Z(t,t+\cdot), \tag{1.6}$$

where X, Y, Z are three possibly correlated processes which evolves according to (1.1), i.e.,

$$\begin{aligned}
\mathrm{d}X(t) &= (A_1X(t) + F_1(X(t)))\mathrm{d}t + \Sigma_1\mathrm{d}W_1(t)\\
\mathrm{d}Y(t) &= (A_2Y(t) + F_2(Y(t)))\mathrm{d}t + \Sigma_2\mathrm{d}W_2(t)\\
\mathrm{d}Z(t) &= (A_3Z(t) + F_3(Z(t)))\mathrm{d}t + \Sigma_3\mathrm{d}W_3(t).
\end{aligned}$$

The next step is to study the properties of the multi-dimensional two-factor model under the HJM framework. Let $A_1 = A_2 = A_3 = \frac{\partial}{\partial \xi}$, and consider the H^2-valued process,

$$\mathrm{d}\begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} = (\begin{bmatrix} \frac{\partial}{\partial \xi} & 0 \\ 0 & \frac{\partial}{\partial \xi} \end{bmatrix}\begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} + \begin{bmatrix} F_{1,2}(f_1(t)) \\ F_{1,3}(f_2(t)) \end{bmatrix})\mathrm{d}t + \begin{bmatrix} \Sigma_1 & \Sigma_2 & 0 \\ \Sigma_1 & 0 & \Sigma_3 \end{bmatrix}\begin{bmatrix} \mathrm{d}W_1(t) \\ \mathrm{d}W_2(t) \\ \mathrm{d}W_3(t), \end{bmatrix}$$

where $F_{i,j}(\cdot) = F_i(\cdot) + F_j(\cdot)$, for $i = 1$ and $j = 2, 3$. Then, we give a criteria for obtaining arbitrage-free dynamics of $\mathrm{d}\boldsymbol{f}(t) = \mathrm{d}(f_1(t), f_2(t))^T$, according to the HJM-condition. In particular, we show that the no-arbitrage condition for $\mathsf{H} = \mathsf{H}_w$-valued forward curves and $\mathsf{U} = \mathsf{H}_w$-valued noise, vanishes for a time-independent non-random noise operator Σ, when the covariance operator Q, of the Wiener-process in H_w is an integral operator.

The purpose of the last section in the study of forward curves and cointegration, is to provide a connection between stationarity and invariant measures in terms of the solutions of $\boldsymbol{f}(t)$, as presented in Tehranchi [47].

Functional Data Analysis and the functional KPSS test of stationarity

Functional data analysis is concerned with statistical methods applied to observations, which we regard as continuous functions. The term *functional data*, first introduced by Ramsay [40] in 1982, extends the classical notion of statistics to the more general framework in terms of functional analysis. The idéa of applying statistical concepts and methods on continuous mathematical objects is found already in 1950 Grenander [18], which studied inference on stochastic processes. A walk-through of the history of functional data analysis is found in the summary of Wang et al. [48].

We will introduce the basics of functional data analysis, including estimation, principal component analysis, and functional data series. In particular, we prepare the reader with the necessary background in order to perform the test of functional stationarity. To that end, we also include some theory regarding functional dependency, specifically m-dependency and $L^p - m$ approximability.

We will study forward curves, which we often will call yield curves. They are theoretically continuous in time and is therefore of interest to examine in a functional data framework. We are using the Nelson-Siegel family of parametrized functions to smooth discrete government bond observations to obtain the yield curves. We apply the functional test of stationarity developed by Hórvath et al. [24] on the obtained Norwegian yield curve. The null hypothesis is formulated under the assumption of stationarity motivated by the works of Kwiatkowski et al. [31], in contrast to Dickey and Fuller [15] type of tests, which assumes the existence of a unit-root.

Given yield curves as functional data $X_1(t), X_2(t), \dots, X_N(t) \in L^2([0,1])$, the null-hypothesis is then,

$$H_0 : X_j(t) = \mu(t) + \eta_j(t), \quad \text{for } 1 \leq j \leq N \text{ and } 0 \leq t \leq 1 \tag{1.7}$$

where $\mu(t)$ is the mean, and $\eta_j(t)$ is a stationary process in terms of Bernoulli shifts. The test statistic is defined as,

$$T_N = \int_0^1 \int_0^1 (S_N(x,t) - xS_N(1,t))^2 \mathrm{d}t\mathrm{d}x,$$

where

$$S_N(x,t) = \frac{1}{N^{1/2}} \sum_{i=1}^{\lfloor Nx \rfloor} X_i(t), \quad 0 \leq x, t \leq 1.$$

Later on, we will see that the test statistic depends on the long-run covariance of the error terms of (1.7). Consequently, we find that we cannot reject the null-hypothesis of the fKPSS test applied to segments of the Nelson-Siegel smoothed Norwegian yield curves. As far as we know, there has never been done a similar functional test on the Norwegian yield curve. Besides, our findings are contrary to Kokoszka and Young's [29] suggestion of non-stationarity found in the American yield curve. However, they smooth the bond observations using splines directly, unlike our indirect Nelson-Siegel approach.

Regarding inference for functional data, the monograph by Bosq [11] is a great unification of the theory of function-valued stochastic process and estimation, although rather technical. Other thorough introductions to functional data analysis are e.g., Ramsay [39], Horváth et al. [22], Hsing and Eubank [25], which all aims at a frequentist and tractable approach to the subject. Given the plenitude of possible approaches to the matter, we try to respectfully introduce FDA without being too technical, in line with Horváth et al. [22].

The importance of interest rate modelling in the actuarial sciences

This thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk*, with a specialization in insurance. Therefore, we give some motivation for why our contributions regarding the study of yield curves, may be of interest in the actuarial sciences and insurance industry. In the study of mathematical finance, we first encounter interest rates (either deterministic or stochastic) in the money market account as part of the risk-free asset in investment strategies. Hence, a financial engineer needs to handle the interest rate, when for instance, pricing options.

Also, a practitioner in life insurance must deal with the pricing of products, which depends on stretches over extended periods, for which it is crucial to handle the time value of money. In practice, the companies must use the yield curve constructed by the European Insurance and Occupational Pensions Authority as part of Solvency 2 for discounting purposes. This scheme applies to all countries in the European Union. Nevertheless, there is still a need to assess the risk. Since we are not capable of foreseeing what the interest rate might be decades forward in time, we are forced to model the uncertainty of the interest rates mathematically, often done by SDE's in terms of the short rate or the forward rate, as mentioned before.

The writing process

After completing a course on interest rate modelling offered at the university, we discussed the option of writing a thesis in which interest rate modelling was a vital part. My supervisor handed me the paper by himself and Süss [4], which proposed a continuous-time paradigm for studying cointegration in terms of finite and infinite-dimensional valued stochastic processes. The starting point of this thesis was, therefore, to investigate the cointegration properties of forward curves from different markets, under the Heath-Jarrow-Morton framework driven by Lévy noise.

Multiple choices needed to be taken at this point. What should be the state space for the forward curves? Should the forward curves be finite or infinite-valued? Lévy noise, Wiener noise? Cylindrical Wiener or Q-Wiener? After some time, the choice of framework fell on the infinite-dimensional Hilbert-valued configuration with Q-Wiener noise. This choice was motivated by the troublesome hedging strategies emerging in finite-rank models, as shown in section 2.4 in Carmona and Tehranchi [13]. In short, the finite rank models do not incorporate maturity-specific risk in hedging contingent claims, and there is not necessarily a unique hedging portfolio of zero-coupon bonds.

The Filipovic space H_w, is tailor-made for the function-valued HJM model, and is, therefore, a natural choice of state space for the forward curves. A drawback of the infinite-dimensional framework is that of the increased complexity compared to the finite models. As a consequence, a big part of this thesis consisted of learning the necessary functional analysis.

Motivated by the complex nature of interest rate products, I wanted to get some hands-on experience by performing an empirical study of Norwegian interest rates, in particular, yield curves. Since yield curves are intrinsically infinite-dimensional, my supervisor introduced me to the field of "functional statistics", called functional data analysis. In consultation with my supervisor, we settled on performing a test of stationarity for functional data, proposed by Hórvath et al. [24].

As time went by, the thesis turned into a two-part project, consisting of HJM modelling and functional data analysis of the Norwegian yield curve, with stationarity as the common denominator. In what follows, it will become apparent that we conclude with the possibility of further study in several parts of the thesis.

List of Symbols

$\mathcal{A}$	σ-algebra of subsets of some set
$\mathbb{1}_A$	Indicator function
B	Banach space
$B_{\mathrm{HS}}(\mathsf{H}, \mathsf{E})$	Space of Hilbert–Schmidt operators from H to$\mathsf{E}$
$\mathcal{B}(\cdot)$	Borel σ-algebra of some set
$\delta(\cdot)$	Evaluation operator
$B(\mathsf{H}, \mathsf{E})$	Space of bounded operators from H to E
$\mathcal{C}$	Hilbert–Schmidt operator on H_w
$\mathcal{E}$	σ-algebra on S^∞
$Df(\cdot)$	Fréchet derivative of f
F	Hilbert space
$\mathcal{F}$	σ-algebra of subsets in probability space
$\mathcal{F}_t$	Filtration with respect to $\mathcal{F}$
$g(t, \xi)$	Musiela notation forward curve, $g(t, t + \cdot)$
H	Hilbert space
H^n	Product Hilbert space
H_w	Filipovic space
$L(\mathsf{H}, \mathsf{E})$	Space of linear operators from H to E
$L_2^0 = L_2(\mathsf{U}_0, \mathsf{H})$	Space of Hilbert–Schmidt operators from U_0 to H
L_2^p	Space of L^2 functions satisfying $v_p(X) < \infty$
$L^2([a, b])$	square integrable functions defined on $[a, b]$
$\mathbb{N}$	The set of natural numbers
Ω	Nonempty measurable set
Ω_T	The product space $[0, T] \times \Omega$
$\mathbb{P}$	Probability measure
$\mathbb{P}_T$	Product Lebesgue measure
$P_X(H)$	The law of X
$\mathcal{P}_T$	σ-algebra
$\mathcal{P}_\infty$	σ-algebra
$\phi_X(y)$	Characteristic functional of X

$\mathbb{Q}$	Risk-neutral probability measure
Q	Covariance operator
Q_0	Integral operator form of covariance operator in H_w
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of real numbers restricted to $[0, \infty)$
$\mathbb{R}^n$	Product space of real numbers
$\mathbb{R}^{n \times m}$	Set of real-valued $m \times n$ matrices
$\sigma(\cdot)$	Generator of σ-algebra
$S(t)$	C_0-semigroup on some space
$\mathcal{S}^2(0, T; L_2^0)$	Space of stochastically integrable elements
$\mathcal{S}_W(0, T; L_2^0)$	Space of stochastically integrable elements (with weakened conditions)
S^∞	Infinite product sequence space
$\hat{S}^2_{\mathrm{PCA}}$	Projection sum in PCA
Σ	Noise operator in HJMM equation
U	Hilbert space
$\mathsf{U}_0 = Q^{1/2}(\mathsf{U})$	reproducing kernel of $W(t)$
$V(t)$	Portfolio strategy at time t
$W(t)$	Wiener process
$\mathbb{Z}$	Set of positive natural numbers
X	Nonempty measurable set

2 Preliminary Functional Analysis and Probability Theory

The theory of probability and stochastics in function spaces requires an understanding of general function and operator analysis. Since this is a thesis considering interest rate modeling, it is not obvious for the reader to be familiar with such theory. To that end, we put forth a generous preliminary chapter, introducing some of the essential concepts in the theory of stochastic analysis in function spaces. We give an introduction to operators in Hilbert spaces, followed by integration theory, probability and stochastic differential equations. Finally, we present some of the key tools from mathematical finance in the functional setting.

2.1 Operators on Hilbert spaces

We briefly introduce some key concepts from function spaces, in particular that of *Hilbert* and *Banach* spaces, which are complete inner product- and normed spaces respectively. We will for the most part work in the Hilbert framework, however, some of the results presented will be for Banach spaces, which in turn, also applies for Hilbert spaces. This section is influenced by Horváth and Kokoszka [22], and the appendix of Peszat and Zabczyk [37].

For mathematical convenience we will throughout this paper denote by B and H, *separable* Banach- and Hilbert spaces respectively. Separability in terms of Hilbert spaces implies the existence of a countable orthonormal basis. Let E be a separable Hilbert space, then we denote by $L(\mathsf{H}, \mathsf{E})$, the space of linear operators from H to E, which is also a separable Hilbert space. When the codomain of Ψ is a field such as $\mathbb{R}$, we call Ψ a *linear functional*, or just a *functional*. We will in this thesis always assume that the inner products are defined over the field $\mathbb{R}$.

If Ψ satisfies the Lipschitz map $\|\Psi(x)\|_{\mathsf{E}} \leq k\|x\|_{\mathsf{H}}$ for $k \in \mathbb{R}$ and $h \in \mathsf{H}$, we say that Ψ is *bounded*, and we denote by $B(\mathsf{H}, \mathsf{E})$ the space of all bounded linear operators, endowed with the *operator norm*

$$||\Psi||_{B(\mathsf{H},\mathsf{E})} = \sup_{h\in\mathsf{H},h\neq 0} \frac{||\Psi h||_{\mathsf{E}}}{||h||_{\mathsf{H}}}, \quad \Psi \in B(\mathsf{H},\mathsf{E}).$$

For a *continuous* linear operator Ψ, the connection between bounded linear operators and continuous linear operators is shown by the equivalence of the following statements

(i) The linear operator Ψ is bounded.
(ii) The linear operator Ψ is continuous.
(iii) The linear operator Ψ is continuous at 0,

thus boundedness in this framework implies continuity of the linear operator Ψ. By continuity, we mean the existence of a $\delta > 0$ for all $\varepsilon > 0$ such that $\|x - y\| < \delta$ implies $\|\Psi x - \Psi y\| < \varepsilon$, for $x, y \in \mathsf{H}$.

We define the *dual space* of H, denoted H^* as the space of all bounded functionals from H to $\mathbb{R}$,, which is also a Hilbert space. For any orthonormal basis $\{e_n\}_n > 0$ in H, we have that $\|e_n\|_{\mathsf{H}} = 1$ and $\langle e_n, e_m\rangle_{\mathsf{H}} = 0$ for $n \neq m$, analogous to that of bases in $\mathbb{R}^n$. Also recall that any element $h \in \mathsf{H}$ has the representation

$$h = \sum_{n=1}^{\infty} \langle h, e_n\rangle_{\mathsf{H}} e_n, \quad h \in \mathsf{H}, \tag{2.1}$$

for any basis $\{e_n\}_{n\leq 1}$ in H.

We say an operator $\Psi : \mathsf{H} \to \mathsf{E}$ is *compact* if the image under Ψ of any bounded subset of H is a *relatively compact* subset, i.e. the closure is compact. Denote by $K(\mathsf{H},\mathsf{E})$ space of all compact operators from H to E.

Let $\Psi \in L(\mathsf{E},\mathsf{H})$. We denote by $\Psi^* \in L(\mathsf{E},\mathsf{H})$ the *adjoint* of Ψ, uniquely defined by the relation

$$\langle \Psi^* h, x\rangle_{\mathsf{E}} = \langle h, \Psi x\rangle_{\mathsf{H}}, \quad \text{for all } h \in \mathsf{H}, x \in \mathsf{E}. \tag{2.2}$$

An operator $\Psi \in B(\mathsf{H})$ is *self-adjoint* if $\Psi^* = \Psi$, and we say that Ψ is *positive-definite* if it is self-adjoint and $\langle \Psi x, x\rangle_{\mathsf{H}} \geq 0$ for all $x \in \mathsf{H}$.

Recall that for finite dimensional operators defined by matrices, we mean by *trace*, the sum of the diagonal elements. We then define the *trace* for infinite dimensional operators.

Definition 2.1.1. Suppose $\Psi \in B(\mathsf{H},\mathsf{E})$, then

$$\mathrm{Tr}(\Psi) = \sum_{i=0}^{\infty} \langle e_n, \Psi e_n\rangle, \tag{2.3}$$

is the trace of the operator Ψ.

We can now present the *Hilbert–Schmidt* operator, which is related to the trace.

Definition 2.1.2. Let $\Psi \in B(\mathsf{H}, \mathsf{E})$. We say that Ψ is *Hilbert–Schmidt* if it has finite Hilbert-Schmidt norm $||\cdot||_{\text{HS}}$, that is

$$||\Psi||_{\text{HS}}^2 := \text{Tr}(\Psi^*\Psi) = \sum_{k=1}^{\infty} ||\Psi e_k||_{\mathsf{H}}^2 < \infty, \tag{2.4}$$

for any orthonormal basis $\{e_k\}_{k\geq 1}$ in H.

The space of all Hilbert–Schmidt operators $B_{\text{HS}}(\mathsf{H}, \mathsf{E})$ from H to E, is again a separable Hilbert space with inner product

$$\langle \Psi_1, \Psi_2 \rangle_{B_{\text{HS}}(\mathsf{H},\mathsf{E})} := \sum_{k=1}^{\infty} \langle \Psi_1 e_k, \Psi_2 e_k \rangle_{\mathsf{H}}. \tag{2.5}$$

And if Ψ is a symmetric positive-definite Hilbert–Schmidt operator, it admits the spectral type of representation

$$\Psi(h) = \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle v_j, \quad h \in \mathsf{H}, \tag{2.6}$$

for orthonormal eigenfunctions v_j of T, that is,

$$\Psi(v_j) = \lambda_j v_j. \tag{2.7}$$

We conclude this section with an example of a Hilbert–Schmidt operator. First recall that a *measure space* is defined by the triple $(\Omega, \mathcal{A}, \mu)$, where Ω is a non-empty set, $\mathcal{A}$ a σ-algebra of subsets of Ω, and μ is a measure.

A classic example of a Hilbert-Schmidt operator is the *integral operator*. Consider the $L^2(\Omega, \mathcal{A}, \mu; \mathbb{R})$ space of measurable functions $f : \Omega \to \mathbb{R}$ satisfying $\int_\Omega f^2(t)\mu(\text{d}t) < \infty$. For simplicity we let $\Omega = [a, b]$ be a compact set in $\mathbb{R}$, and μ the Lebesgue measure. By $L^2([a, b])$ we mean the L^2 space of measurable functions from Ω the compact subset $[a, b]$ of $\mathbb{R}$ with the Lebesgue measure μ.

Example 2.1.3. We define an integral operator as the map,

$$\Psi(f)(t) = \int_a^b K(s, t) f(t) \text{d}t, \tag{2.8}$$

where $K(\cdot, \cdot)\colon [a, b] \times [a, b] \to \mathbb{R}$ is called the kernel of the integral operator. The kernel in this case is understood as real-valued function, and is not to be mixed up with the term kernel or nullspace from linear algebra. If $\int_a^b \int_a^b K^2(s, t)\text{d}t < \infty$, then $\Psi(f)(\cdot)$ is a Hilbert–Schmidt operator, and it can be shown that the Hilbert–Schmidt norm of Ψ is

$$\|\Psi\|_{\text{HS}}^2 = \int_a^b \int_a^b K^2(s, t)\text{d}t. \tag{2.9}$$

2.2 Integration theory

The object of this section is to make sense of integration where the integrand is a Banach-valued variable. We will introduce the *Bochner integral*, developed by Bochner [10], which extends the Lebesgue integral to account for Banach valued functions. For that reason, we will assume that the reader is familiar with some general measure theory. See for instance McDonald and Weiss [32] for a thorough introduction.

Consider the measure space $(\mathsf{X}, \mathcal{A}, \mu)$. We denote by $\int_{\mathsf{X}} f \mathrm{d}\mu$, the Bochner integral of $f : \mathsf{X} \to \mathsf{B}$ with respect to the measure μ, from which we want to define formally. Recall that for functions $f : \mathsf{X} \to \mathbb{R}$, the above integral is the Lebesgue integral.

Despite our interest in Hilbert spaces, we present the *Bochner integral* for Banach-valued functions following chap. 2.6 of Hsing and Eubank [25]. We will see that the definition of the Bochner integral is similar to the Lebesgue integral. For that reason, we include some of the steps which make up the construction of the Bochner integral. We define the Bochner integral for simple functions, and extend the integral to hold for measurable functions $f : \mathsf{X} \to \mathsf{B}$.

We conclude this section by providing some of the essential properties of the Bochner integral. We will see that many of the features we know of from Lebesgue integration, carry over to the functional-valued case. We start by defining the Bochner integral of simple functions.

Definition 2.2.1. The function $f : \mathsf{X} \to \mathsf{B}$ is called *simple*, if it can be represented as

$$f(\omega) = \sum_{i=1}^{n} \mathbb{1}_{A_i}(\omega) f_i, \tag{2.10}$$

for some finite $n \in \mathbb{N}, A_i \in \mathcal{A}$ and $f_i \in \mathsf{B}$. If in addition $\mu(A_i) < \infty$ for all $A_1, \ldots, A_n$, the *Bochner integral* of f is defined to be

$$\int_{\mathsf{X}} f \mathrm{d}\mu = \sum_{i=1}^{n} \mu(A_i) f_i. \tag{2.11}$$

Moreover, we extend the Bochner integral to account for measurable functions on Ω through the following theorem.

Theorem 2.2.2 ([10, Theorem 2.6.4])**.** *If $f : \mathsf{X} \to \mathsf{B}$ is a measurable function, satisfying*

$$\int_{\mathsf{X}} ||f|| \mathrm{d}\mu < \infty,$$

suppose there exists a finite-dimensional subspace $\mathsf{B}_n \subseteq \mathsf{B}$, such that

$$\lim_{n\to\infty} \int_{\mathsf{X}} ||f - g_n|| \mathrm{d}\mu = 0,$$

for B_n*-valued measurable functions* $\{g_n\}_{n\geq 1}$*. Then there exists a sequence of simple integrable functions* $\{f_n\}_{n\geq 1}$ *such that*

$$\lim_{n\to\infty}\int_{\mathsf{X}} ||f - f_n||\mathrm{d}\mu = 0.$$

The Bochner integral of f *is then*

$$\int_{\mathsf{X}} f\mathrm{d}\mu = \lim_{n\to\infty}\int_{\mathsf{X}} f_n\mathrm{d}\mu.$$

Note that the Bochner-integral is a linear operator,

$$I(f)(\cdot) = \int_{\mathsf{X}} f\mathrm{d}\mu. \tag{2.12}$$

We provide a result on linear operators applied on the Bochner integral.

Theorem 2.2.3 ([25, Theorem 3.1.7])**.** *Let* B_1 *and* B_2 *be Banach spaces and* $f:\Omega\to\mathsf{B}_1$ *a Bochner integrable function. If* $\Psi\in B(\mathsf{B}_1,\mathsf{B}_2)$*, then* $\Psi(f)(\cdot)$ *is Bochner integrable, and*

$$\Psi(\int_{\mathsf{X}} f\mathrm{d}\mu) = \int_{\mathsf{X}} \Psi f\mathrm{d}\mu. \tag{2.13}$$

An example of the above theorem is that the functional $\langle\cdot, b\rangle : \mathsf{B}\times\mathsf{B}\to\mathbb{R}$, satisfies

$$\langle\int_{\mathsf{X}} f\mathrm{d}\mu, b\rangle = \int_{\mathsf{X}}\langle f, b\rangle\mathrm{d}\mu \tag{2.14}$$

Since we will be working mostly with functions taking values in a separable Hilbert spaces, we state the following theorem to ease the requirements for being Bochner integrable.

Theorem 2.2.4 ([25, Theorem 2.6.5])**.** *For any separable Hilbert space* H*, the measurable function* $f:\mathsf{X}\to\mathsf{H}$*, satisfying* $\int_{\mathsf{X}}\|f\|_{\mathsf{H}}\,\mathrm{d}\mu<\infty$*, is Bochner integrable.*

When we later on specify Bochner-integrability of an H-valued f, we mean that the Lebesgue integral of the norm of such a function is integrable. Moreover, we have that the property of *dominated convergence* also holds for the Bochner integral.

Theorem 2.2.5 ([25, Theorem 2.6.6])**.** *Let* $\{f_n\}_{n\geq 0}$ *be a sequence of Bochner integrable functions, converging to some* $f\in\mathsf{B}$*. If there exists a nonnegative Lebesgue measurable function* g *satisfying the bound* $\|f_n\|\leq g$ *for all* n *almost everywhere. Then, we have that* f *is Bochner integrable and*

$$\int_{\mathsf{X}} f\mathrm{d}\mu = \lim_{n\to\infty}\int_{\mathsf{X}} f_n\mathrm{d}\mu. \tag{2.15}$$

Recall that by *almost everywhere* we mean that a property holds everywhere except for a set of measure zero.

There also exist a bound for the norm of a Bochner integral, similar to the monotonicity of the Lebesgue integral.

Theorem 2.2.6 ([25, Theorem 2.6.7]). *For Bochner-integrable $f : \mathsf{X} \to \mathsf{B}$, the following inequality holds,*

$$\|\int_{\mathsf{X}} f \mathrm{d}\mu\| \leq \int_{\mathsf{X}} \|f\| \mathrm{d}\mu. \tag{2.16}$$

The main purpose of introducing the Bochner integral is for the sake of computing *expetations* of function-valued random variables, which we will discuss in the next section. In conclusion, we have for functions $f : \mathsf{X} \to \mathsf{H}$ satisfying

$$\int_{\mathsf{X}} \|f\|_{\mathsf{X}} \mathrm{d}\mu < \infty,$$

the existence of a sequence of simple integrable functions $\{f_n\}_{n\geq 1}$ such that

$$\int_{\mathsf{X}} f \mathrm{d}\mu = \lim_{n\to\infty} \int_{\mathsf{X}} f_n \mathrm{d}\mu = \lim_{n\to\infty} \sum_{i=1}^{k_n} \mu(A_i) f_{i,n}, \tag{2.17}$$

where k_n is finite for all n.

2.3 Probability in Hilbert spaces

In this section we extend the notion of probability on $\mathbb{R}^n$, to a possibly infinite-dimensional separable Hilbert space H, following the axiomatic measure-theoretic approach carried out by Kolmogorov in the 1930s [1]. The state-space of choice will be the separable Hilbert space H defined on reals. However we will consider the *complete* probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ where $\mathcal{F}_t$ denotes the *filtration*, i.e. $\{\mathcal{F}_t\}_t$ is an increasing familiy of $\sigma-$algebras defined on the σ-algebra $\mathcal{F}$ of subsets of Ω. By a *complete* probability space we mean that for all $F \in \mathcal{F}$ satisfying $P(F) = 0$ we must have $F_{\text{sub}} \in \mathcal{F}$, for all $F_{\text{sub}} \subset F$.

We start this section by introducing the H-random variable, before we introduce the expectation and covariance of H-random variables. Furthermore, we present the stochastic processes, conditional expectation and different notions of convergence. This section is influenced by the introductions of Da Prato and Zabczyk [14] and Peszat and Zabczyk [37]. For a thorough presentation of stochastic processes in Banach- and Hilbert spaces, see the monologue of Bosq [11].

We denote by $(\mathsf{H}, \mathcal{H})$ a *measurable space*, which is a measure space without specifying the explicit measure. Suppose H is a Let $\mathcal{H} = \mathcal{B}(\mathsf{H})$ be the *Borel*

[1] See Shafer and Vovk [42].

σ-algebra, which is the smallest σ-algebra generated by the open sets of H. Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(\mathsf{H}, \mathcal{H})$, and recall that a *measurable function* is a function $f : \Omega \to \mathsf{H}$ if,

$$f^{-1}(H) = \{\omega \in \Omega | f(\omega) \in H\} \in \mathcal{F}, \text{for all } H \in \mathcal{H}.$$

We provide the definition of an H-valued random variable.

Definition 2.3.1. Suppose $(\mathsf{H}, \mathcal{H})$ is a measurable space. An H-*valued* random variable, which we will call an H-random variable, is any measurable function

$$X : \Omega \to \mathsf{H}. \tag{2.18}$$

Furthermore the *law of X* is denoted by

$$P_X(H) = \mathbb{P}\left(X^{-1}(H)\right), \quad H \in \mathcal{H}. \tag{2.19}$$

The expectation of a random variable is defined in terms of an integral over the sample space Ω, with respect to the probability measure $\mathbb{P}$. Since we assume that X is H-valued, we must resort to the Bochner integral in giving a meaningful definition of the expectation.

Definition 2.3.2. Let X be a H-random variable. If X is Bochner-integrable i.e., $\int_\Omega \|X\|_{\mathsf{H}} \mathbb{P}(\mathrm{d}\omega) < \infty$, the expectation of X is given by the Bochner integral

$$E[X] = \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega).$$

Equivalently, we can define the expected value identifying the unique element $\mu = E[X]$ by the relation

$$E[\langle x, X\rangle] = \langle x, \mu\rangle, \quad \text{for all } x \in \mathsf{H}. \tag{2.20}$$

It is also necessary to introduce the concept of conditional probability.

Proposition 2.3.3. *[37, Proposition 3.13] Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be a H-valued integrable random variable. Then there is a unique integrable G-measurable H-valued random variable $E[X|\mathcal{G}]$ such that*

$$\int_G X(\omega)\mathbb{P}(\mathrm{d}\omega) = \int_G E[X|\mathcal{G}]\, \mathbb{P}(\mathrm{d}\omega), \quad \textit{for all } G \in \mathcal{G}, \textit{almost surely.} \tag{2.21}$$

Almost surely is the probabilistic notion of the having a property hold almost everywhere. For the readers convenience we will for the most part omit the ω-dependency of the random variables. We define the *covariance operator* of a random variable.

Definition 2.3.4. Let X be an H-random variable satisfying $E\left[\|X\|^2_{\mathsf{H}}\right] < \infty$. The covariance operator of X is then given by

$$C_X(x) = E\left[\langle X - E[X], x\rangle (X - E[X])\right], \quad \text{for } x \in \mathsf{H}. \tag{2.22}$$

The covariance is inherently defined in terms of the linear *tensor* operator. That is, let $a, b \in \mathsf{H}$, then we define $a \otimes b(\cdot) = a\langle b, \cdot\rangle$, which in turn means that,

$$C_X(x) = E\left[(X - E\left[X\right]) \otimes (X - E\left[X\right])\right]. \tag{2.23}$$

Moreover, we can now describe the variation between two random variables as the *correlation*.

Definition 2.3.5. Let X and Y be two Bochner integrable H-random variables. The *correlation* is then given by

$$\mathrm{Cor}(X, Y) = E\left[(X - E\left[X\right]) \otimes (Y - E\left[Y\right])\right] \tag{2.24}$$

To each H-random variable X, there exists an unique *characteristic functional* which completely determines the law of X, P_X, just as in the finite-dimensional case.

Definition 2.3.6. Let i denote imaginary unit. The *characteristic functional* of X is then given by,

$$\phi_X(h) = \int_{\mathsf{H}} e^{i\langle h, X\rangle}\mathbb{P}(\mathrm{d}x), \quad h \in \mathsf{H}. \tag{2.25}$$

Another constituent element of modern probability theory is a *stochastic processes*.

Definition 2.3.7. Let I denote any time interval on $\mathbb{R}_+$, and $\mathcal{B}(I)$ the *Borel* σ-algebra, which is the smallest σ-algebra containing all open sets of I. We say that a *stochastic process* $X(t) = \{X(t)\}_{t\in I}$, is any family of random elements taking values in H.

The stochastic processes we encounter in this thesis will be assumed continuous, meaning that we notation wise will write $\{X(t)\}_{t\in[a,b]}$ for $t \in [a, b]$. If $t \in [0, \infty)$ we just write $X(t)$.

We say that $X(t)$ is *adapted* (to the filtration $\mathcal{F}_t$) if it is $\mathcal{F}_t$-measurable. We present some definitions related to stochastic processes.

Definition 2.3.8. Let $X(t)$ be a H-valued stochastic process. If there exists a H-valued stochastic process $Y(t)$ such that

$$\mathbb{P}(X(t) = Y(t)), \tag{2.26}$$

for all t, we call $Y(t)$ a *modification* of $X(t)$.

Definition 2.3.9. We say that an H-valued stochastic process $X(t)$ is *stochastically continuous* if

$$\lim_{s\to t} E\left(\|X(t) - X(s)\|_{\mathsf{H}} > \varepsilon\right) = 0,$$

for all $\varepsilon > 0$.

Moreover we define the *martingale* property for H-valued stochastic processes. This property is of great importance in mathematical finance.

Definition 2.3.10. If $E[\|X(t)\|_{\mathsf{H}}] < \infty$ for all t, then the H-valued stochastic process $X(t)$ is a *martingale*, if

$$E[X(t)|\mathcal{F}_s] = X(s), \quad \text{a.s.}, \tag{2.27}$$

for any $t, s \in [0, \infty)$ with $s \leq t$.

We also introduce different notions of stochastic convergence of random variables, namely *almost surely convergene* and *convergence in probability.*

Definition 2.3.11. Let X and $\{X_n\}_{n\geq 1}$ be H-valued random variables. Then, if

$$\mathbb{P}\big(\|X_n(\omega) - X(\omega)\|_{\mathsf{H}} \underset{n\to\infty}{\to} 0\big) = 1,$$

we say that X_n converges to X almost surely, which we denote by $X_n \underset{a.s}{\to} X$. Moreover, we have that X_n converges in *probability* denoted $X_n \underset{p}{\to} X$ if for each $\varepsilon > 0$,

$$\mathbb{P}\big(\|X_n - X\|_{\mathsf{H}} > \varepsilon\big) \underset{n\to\infty}{\to} 0.$$

Notice that the the expectation and Bochner integral commute.

Lemma 2.3.12. *Let $X(s)$ be an $\mathcal{F}_s$-measurable random process which is Bochner-integrable for all $s \geq 1$. Then*

$$E\left[\int_{\mathsf{H}} X(s)\mathrm{d}s\right] = \int_{\mathsf{H}} E[X(s)]\,\mathrm{d}s. \tag{2.28}$$

Proof. Let g by an element of H. We have,

$$\begin{aligned}\langle E\left[\int_{\mathsf{H}} X(s)\mu(\mathrm{d}s)\right], g\rangle &= E\left[\langle \int_{\mathsf{H}} X(s)\mu(\mathrm{d}s), g\rangle\right] = E\left[\int_{\mathsf{H}} \langle X(s), g\rangle \mu(\mathrm{d}s)\right] \\ &= \int_{\mathsf{H}} E[\langle X(s), g\rangle]\,\mu(\mathrm{d}s) = \langle \int_{\mathsf{H}} E[X(s)], g\rangle \mu(\mathrm{d}s),\end{aligned}$$

where we made use of the linearity of the Bochner integral with respect to functionals, whereas the last equation follows by Fubini–Tonelli theorem for Lebesgue integrals. ∎

Example 2.3.13. Consider the separable Hilbert space $L^2([a, b])$, which has inner product $\langle x, y\rangle_2 = \int_a^b x(t)y(t)\mathrm{d}t$. We compute the covariance operator of the $L^2([a, b])$-valued mean zero Bochner integrable random variable X,

$$\begin{aligned}C_X(x) &= E[\langle x, X\rangle_2 X] \\ &= E\left[\int_a^b x(t)X(t)\mathrm{d}tX\right] \\ &= E\left[\int_a^b x(t)X(t)X(s)\mathrm{d}t\right].\end{aligned}$$

Moreover we interchange the expectation and integral by Fubini-Tonelli, for which we obtain

$$E\left[\int_a^b x(t)X(t)X(s)\mathrm{d}t\right] = \int_a^b E\left[X(t)X(s)\right]x(t)\mathrm{d}t \tag{2.29}$$

$$= \int_a^b K(s,t)x(t)\mathrm{d}t. \tag{2.30}$$

Notice that the covariance operator is an integral operator with kernel $K(\cdot,\cdot)$, which is just one of many pleasant properties of the $L^2([a,b])$ space of functions.

We end this section by discussing normally distributed random variables in infinite-dimensional spaces, in particular the *Gaussian measure*. We define an H-random variable X to be *centered Gaussian* if for all $x \in \mathsf{H}$, $\langle X, x\rangle$ is a Gaussan random variable in $\mathbb{R}$. Recall that we are familiar with Gaussian measures on $\mathbb{R}$ as the probability density function of a (mean-zero) normal random variable X,

$$f_X(A) = \frac{1}{\sqrt{2\pi}}\int_A e^{-y^2/2}\mathrm{d}y, \quad A \in \mathcal{B}(\mathbb{R}). \tag{2.31}$$

This motivates the definition of Gaussian measures on Hilbert spaces.

Definition 2.3.14. The probability measure μ is called Gaussian if for any $h \in \mathsf{H}$, there exists $m \in \mathbb{R}$ and $\sigma \geq 0$ such that the quantity

$$\mu(x \in \mathsf{H} : \langle h, x\rangle \in H) \quad \text{for all } H \in \mathcal{B}(\mathbb{R}), \tag{2.32}$$

has the law $\mathcal{N}(m,\sigma)$.

2.4 Fréchet Derivative

We extend the notion of differentiability for functions in finite-dimensional spaces to possibly infinite-dimensional normed spaces, as presented in Hsing and Eubank [25]. From elementary calculus we define the differential operator of a function $f : \mathbb{R} \to \mathbb{R}$ with $x, h \in \mathbb{R}$, as the limit

$$f'(x) = \lim_{h\to 0}\frac{f(x+h) - f(x)}{h}, \tag{2.33}$$

for all x.

Now, if $f \in L(\mathsf{H},\mathsf{E})$, and H,E are normed spaces over $\mathbb{R}$ with $x, h \in \mathsf{H}$ and $f(\cdot) \in \mathsf{E}$, what can we say about the operator $f'(x)$? Clearly, $f'(\cdot)$ evaluated in any point is an element of a normed space, not an operator. In addition, we have that a fraction consisting of elements of a normed space is nonsensical, which motivates a definition based on norms of the elements instead. However, in order to obtain an operator acting on elements in H, simple computations shows that (2.33) is equivalent to,

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0.$$

The above can be lifted to account for elements in normed spaces, and functions on normed spaces,

$$\lim_{\|h\|_{\mathsf{H}}\to 0} \frac{\|f(x+h) - f(x) - f'(x)h\|_{\mathsf{E}}}{\|h\|_{\mathsf{H}}} = 0.$$

Thus our candidate for a differential operator on normed spaces is the term $f'(x)h$. We summarize the above in the following definition.

Definition 2.4.1. Let H and E be two normed spaces, and $f : \mathsf{H} \to \mathsf{E}$ a function defined on some open subset U of H. If f satisfies,

$$\lim_{\|h\|_{\mathsf{H}}\to 0} \frac{\|f(x+h) - f(x) - f'(x)h\|_{\mathsf{E}}}{\|h\|_{\mathsf{H}}} = 0, \tag{2.34}$$

then we call $f'(x)$ the Fréchet derivative of $f(x)$, and denote by the Frechet derivative $Df(\cdot)$.

2.5 Hilbert-valued Wiener Process and Stochastic Integrals

The goal of this section is to give meaning to integrals on the form

$$\int_0^t \Phi(s)\mathrm{d}W(s), \tag{2.35}$$

where $W(t)$ is a Wiener process taking values in Hilbert space H. We establish the notion of such an integral following the formulation proposed in Da Prato and Zabczyk [14].

The approach is similar to the construction of the stochastic integral for $\mathbb{R}^n$-valued Wiener process, the so-called *Itô-integral.* There are mainly two ways of defining a Wiener process in a Hilbert space, namely the *Q-Wiener process* and the *cylindrical Wiener process*. In this thesis, we will assume that $W(t)$ is a Q-Wiener process, i.e., a stochastic process taking values in a possibly infinite-dimensional Hilbert space H with covariance of positive trace class. Intuitively we want for all $h \in \mathsf{H}$

$$\langle W(t), h\rangle, \quad t \geq 0, \tag{2.36}$$

that the projection of $W(t)$ is a real-valued Wiener process. Hence the law $W(t)$ should be a mean zero Gaussian measure. Moreover, we have for $s \leq t$

$$\begin{aligned} E\left[\langle W(t), h\rangle\langle W(s), h\rangle\right] &= E\left[\langle (W(t)-W(s)) + W(s), h\rangle\langle W(s), h\rangle\right] \\ &= E\left[(\langle (W(t)-W(s)), h\rangle + \langle W(s), h\rangle)\langle W(s), h\rangle\right] \\ &= E\left[(\langle (W(t)-W(s)), h\rangle\langle W(s), h\rangle\right] + E\left[\langle W(s), h\rangle^2\right] \\ &= sE\left[\langle W(1), h\rangle^2\right] = s\langle Qh, h\rangle, \end{aligned}$$

where we used the property of independent increments. Using the same method of reasoning one can show that

$$E\left[\langle W(t), h\rangle\langle W(s), k\rangle\right] = sE\left[\langle W(1), h\rangle\langle W(1), k\rangle\right] = \langle Qh, k\rangle, \tag{2.37}$$

where Q is the associated covariance operator. The definition of the H-valued Wiener process is as follows.

Definition 2.5.1. An H-valued process $W(t)$, is called a *Q-Wiener process* if it satisfies

(i) $W(0) = 0$

(ii) $W(t)$ has continuous sample paths

(iii) $W(t)$ has independent increments

(iv) The law of $(W(t) - W(s))$ is $\mathcal{N}(0, (t-s)Q)$,

where the covariance Q, is a positive trace class operator on H.

As with the Bochner integral, we outline the construction of the stochastic integral, omitting some of the proofs and technicalities. For $T < \infty$, we denote by $\Phi(t)$, $t \in [0, T]$ an $L(\mathsf{U}, \mathsf{H})$ valued process. We say that $\phi(t)$ is an *elementary process* if there exists a finite sequence $0 = t_0 < t_1 < \cdots < t_k = T$ and $\phi_0, \phi_1, \ldots, \phi_{n-1}$ $L(\mathsf{U}, \mathsf{H})$ valued random variables also taking a finite number of values so that each ϕ_k is $\mathcal{F}_{t_k}$-measurable and

$$\Phi(t) = \phi_k, \quad \text{for } t \in (t_k, t_{k+1}], k = 1, \ldots, n-1. \tag{2.38}$$

Thus we define the stochastic integral of an elementary process, as

$$\int_0^t \Phi(s)\mathrm{d}W(s) = \sum_{k=0}^{n-1} \phi_k(W(t_{k+1} \wedge t) - W(t_k \wedge t)), \tag{2.39}$$

where $a \wedge b = \min\{a, b\}$.

In the extension of $\Phi(t)$ from an elementary process to a general $L(\mathsf{U}, \mathsf{H})$ valued stochastic process, we introduce the subspace $\mathsf{U}_0 = Q^{1/2}(\mathsf{U}) \subset \mathsf{U}$ being the *reproducing kernel* of $W(t)$, with inner product

$$\langle u, v\rangle_0 = \sum_{k=1}^{\infty} \frac{1}{\lambda_k}\langle u, e_k\rangle\langle v, e_k\rangle = \langle Q^{-1/2}u, Q^{-1/2}v\rangle, \quad \text{for } u, v \in \mathsf{U}_0,$$

which defines a Hilbert space.

Let $L_2^0 = L_2(\mathsf{U}_0, \mathsf{H})$ be the space of all Hilbert–Schmidt operators from U_0 to H equipped with norm

$$\begin{aligned}\|\Psi\|_{L_2^0}^2 &= \sum_{j,k=1}^{\infty} |\langle \Psi g_j, f_k\rangle|^2 = \sum_{j,k=1}^{\infty} |\langle \Psi \lambda_j e_j, f_k\rangle|^2 = \sum_{j,k=1}^{\infty} \lambda_j^2 |\langle e_j, f_k\rangle|^2 \\ &= \|\Psi Q^{1/2}\|_{\mathrm{HS}}^2 = \mathrm{Tr}\left[(\Psi Q^{1/2})(\Psi Q^{1/2})^*\right],\end{aligned}$$

where we let $\{g_j\}_{j\geq 1} = \{\lambda_j e_j\}_{j\geq 1}$ with $\{e_j\}_{j\geq 1}$ and $\{f_j\}_{j\geq 1}$ being complete orthonormal bases in U_0. Moreover, if $\Psi(t)$, $t \in [0, T]$ is a measurable L_2^0-valued measurable process, define the norm

$$\|\!|\Psi|\!\|_t = \left(E\left[\int_0^t \|\Psi(s)\|_{L_2^0}^2 \,\mathrm{d}s\right]\right)^{1/2} \tag{2.40}$$

$$= \left(E\left[\int_0^t \mathrm{Tr}\left[(\Psi(s)Q^{1/2})(\Psi(s)Q^{1/2})^*\right] \mathrm{d}s\right]\right)^{1/2}. \tag{2.41}$$

Proposition 2.5.2 ([14, Proposition 4.20])**.** *If Φ is an elementary process and $\|\!|\Psi|\!\|_T < \infty$, then the process $M(t) = \int_0^t \Phi(s)\mathrm{d}W(s)$ is a continuous, square integrable H-valued martingale on $[0, T]$. We also have that*

$$E\left[M(t)\right]^2 = \|\!|\Phi|\!\|_t^2, \quad \textit{for } t \in [0, T]. \tag{2.42}$$

The above result can be extended to valid for any L_2^0-predictable process.

Proposition 2.5.3 ([14, Proposition 4.22])**.** *All elementary processes Φ are L_2^0-predictable. In addition, if Φ is an L_2^0-predictable process such that $\|\!|\Phi|\!\|_T < \infty$, then there exists a sequence of elementary processes $\{\Phi_n\}_{n\geq 0}$ such that $\|\!|\Phi - \Phi_n|\!\|_T \to 0$ as $n \to \infty$.*

Hence, we have showed that the set of all L_2^0-predictable processes Φ with $\|\!|\Phi|\!\|_T < \infty$ forms a Hilbert space which we denote $\mathcal{S}^2(0, T; L_2^0)$. The elementary processes form a dense set in $\mathcal{S}^2(0, T; L_2^0)$, such that we can justify the generalization of the stochastic integral $M(t)$, to all elements of $\mathcal{S}^2(0, T; L_2^0)$. Consequently, the identity (2.42) holds, and $M(t)$ is a continuous, square integrable H-valued martingale.

A natural starting point in order to study the integrands is to regard them as random variables defined on the product space $\Omega_T = [0, T) \times \Omega$, equipped with the product σ-algebra $\mathcal{B}([0, T]) \times \mathcal{F}$. Denote by $\mathbb{P}_T$ the product of the Lebesgue measure on $[0, T]$, and the probability measure $\mathbb{P}$. Note that if $T = \infty$, the above sets can be thought of as $[0, \infty)$.

To capture the adaptability of the stochastic process,let us denote by $\mathcal{P}_\infty$, the σ-algebra generated by the sets

$$(s, t] \times F, \quad 0 \leq s < t < \infty,$$

for $F \in \mathcal{F}_s$, including $\{0\} \times F$ for $F \in \mathcal{F}_0$. Hence, the class of integrable stochastic processes is then the measurable maps from $(\Omega_\infty, \mathcal{P}_T)$ into $(L_2^0, \mathcal{B}(L_2^0))$.

Finally, we weaken the condition for stochastic processes to be integrable by introducing the $\mathcal{S}_W(0,T;L_2^0)$ -space, which is a linear space defined by Φ's satisfying

$$\mathbb{P}(\int_0^T \|\Phi(s)\|_{L_2^0}^2 \mathrm{d}s < \infty) = 1. \tag{2.43}$$

All the elements of $\mathcal{S}_W(0,T;L_2^0)$ are stochastically integrable [2].

Theorem 2.5.4. *[14, Proposition 4.28] Given $\Phi_1, \Phi_2 \in \mathcal{S}_W^2(0,T;L_2^0)$, then*

$$E\left[\int_\Omega \Phi_i(s)\mathrm{dW(s)}\right] = 0, \tag{2.44}$$

$$E\left[\|\int_\Omega \Phi_i(s)\mathrm{d(s)}\|^2\right] < \infty, \tag{2.45}$$

with $t \in [0,T]$ for $i = 1,2$. In addition, for $t, s \in [0,T]$, the correlation operator is given by

$$\mathrm{Cor}(\int_\Omega \Phi_1(s)\mathrm{d}W(s), \int_\Omega \Phi_2(s)\mathrm{d}W(s)) = E\left[\int_0^{t\wedge s} (\Phi_2(r)Q^{1/2})(\Phi_1(r)Q^{1/2})^*\mathrm{d}r\right].$$

The stochastic Fubini also hold.

Theorem 2.5.5 ([14, Theorem 4.33])**.** *Let $(\mathsf{H}, \mathcal{H})$ be a measurable space. Let $\Phi(t,\omega,x)$ be a measurable map from $(\Omega_T \times H, \mathcal{P}_T \times \mathcal{B}(H))$ into $(L_2^0, \mathcal{B}(L_2^0))$. If*

$$\int_{\mathsf{H}} \|\|\Phi(\cdot,\cdot,x)\|\|_T \mu(\mathrm{d}x) < \infty, \tag{2.46}$$

then the following equation holds,

$$\int_{\mathsf{H}} (\int_0^T \Phi(t,x)\mathrm{dW(t)})\mu(\mathrm{dx}) = \int_0^T (\int_{\mathsf{H}} \Phi(t,x)\mu(\mathrm{dx}))\mathrm{dW(t)}. \tag{2.47}$$

We have now constructed the stochastic integral with respect to the Q-Wiener process. However, it is possible to make sense of stochastic integrals when the Wiener processes $W(t)$ is not necessarily is of finite trace. The stochastic integral with respect to a *cylindrical* Wiener process can be approximated by a limit of stochastic integrals with respect to finite-dimensional Wiener processes.

2.6 Hilbert-valued stochastic differential equations and semigroups

In this section, we introduce the theory of function-valued stochastic equations (SDE's), in particular, Hilbert-valued SDE's. The framework of infinite-dimensional stochastic differential equations differs substantially from the

[2] See section 4.2 in Da Prato and Zabczyk [14]

finite-dimensional SDE theory, so we will carefully assemble the necessary preliminaries.

Motivated by the finite-dimensional setting, we extend the framework to account for Hilbert-valued state-space, forcing the drift and volatility terms to be linear operators. Moreover, the solutions of such equations will depend on a particular group of operations called *semigroups*.

From the Itô calculus, we are familiar with multidimensional stochastic differential equations as *Itô-processes*, being on the form

$$\mathrm{d}X(t) = \alpha(t, X(t))\mathrm{d}t + \sigma(t, X(t))\mathrm{d}W(t), \tag{2.48}$$

or equivalently

$$X(t) = X(0) + \int_0^t \alpha(s, X(s))\mathrm{d}s + \int_0^t \sigma(s, X(s))\mathrm{d}W(s). \tag{2.49}$$

Here we assume that for $T > 0$, $a(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $b(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are measurable functions. If α and σ satisfy some Lipschitz bound, we can prove existence and uniqueness of (2.48), see chapter 5 of Oksendal [35]. Moreover, recall that there are different notions of a solution to such an equation, namely, *weak* and *strong* solutions. Simply put, if $X(t)$ is a solution depending on the Wiener process $W(t)$ given in (2.48) we have a *strong* solution. A *weak* solution is the pair $(X(t), \tilde{W}(t))$ of processes where $X(t)$ is adapted to the filtration $\tilde{\mathcal{F}}_t$, and $\tilde{W}(t)$ is a $\tilde{\mathcal{F}}_t$-Wiener process. Now, in the case of infinite-dimensional stochastic equations, the functions α and σ are now linear operators. This section is influenced by Da Prato and Zabczyk [14].

Before embarking on the introduction to the infinite-dimensional SDE's we illustrate some of the technicalities we may encounter through the deterministic *Cauchy problem* or equation:

$$\begin{cases} f'(t) = A_0 f(t), & t \geq 0 \\ f(0) = x \in \mathsf{H}, \end{cases} \tag{2.50}$$

where A_0 is a possibly unbounded linear operator defined on a dense linear subspace $D(A_0)$ of H, whereas $f'(t)$ denotes the *strong derivative*. By strong derivative we mean the existence of limit,

$$f'(t) = \lim_{h\to 0} \frac{f(t+h) - f(t)}{h}.$$

Definition 2.6.1. The Cauchy problem is said to be *well-posed* if it satisfies the following properties:

1. For any $x \in D(A_0)$ there exists a unique strongly differentiable function $f(t, x)$ satisfying (2.50) for all $t \in [0, \infty)$.

2. For any sequence $\{x_n\}_{n\geq 0} \in D(A)$, with $\lim_{n\to\infty} x_n = 0$, we have

$$\lim_{n\to\infty} f(t, x_n) = 0, \tag{2.51}$$

for all $t \in [0, \infty)$.

Assuming the Cauchy problem is well-posed we define the operator $S(t) : D(A) \to \mathsf{H}$ by

$$S(t)x = f(t, x), \quad \text{for all } x \in D(A), t \geq 0.$$

Notice that $S(0)x = f(0, x) = x$, thus we have that $S(0) = I$. Also $S(t)S(s)x = S(t)f(s, x) = f(t, f(s, x))$, which by the assumption of uniqueness of $f(t, x)$ we have that $f(t, f(s, x)) = f(t + s, x)$. This motives the fact that operator S should satisfy $S(t + s) = S(t)S(s)$. Lastly, we have that $S(\cdot)x$ is continuous. We shall name such a bounded operator $S(\cdot)$ a *semigroup.*

Definition 2.6.2. A family of bounded operators $\{S(t)\}_{t\geq 0}$ on a Banach space B is called a C_0-*semigroup* if the following holds

(i) $S(0) = I$

(ii) $S(s)S(t) = S(s + t)$, for all $s, t \geq 0$.

(iii) For $t \geq 0$, the mapping $t \mapsto S(t)x \in \mathsf{B}$ is continuous for each $x \in \mathsf{B}$, that is

$$\lim_{t\downarrow 0} \|S(t)x - x\| = 0,$$

for all $x \in \mathsf{B}$.

Definition 2.6.3. Let $S(t)$ be a C_0-semigroup, and $b \in \mathsf{B}$. The *generator* of $S(t)$ denoted by A is the limit

$$\lim_{t\downarrow 0} \frac{S(t)b - b}{t}, \tag{2.52}$$

if it exists. We denote by $D(A)$, the set of all $b \in \mathsf{B}$ satisfying the above limit.

We give an example of a generator.

Example 2.6.4. Define the operation $S(t)h(\cdot) = h(\cdot + t)$, for a function defined on some seperable Hilbert space where evaluation makes sense. Notice that $S(0)h(\cdot) = h(\cdot)$, and $S(t + s)h(\cdot) = h(\cdot + t + s) = S(t)h(\cdot + s) = S(t)S(s)h(\cdot)$, hence $S(t)$ defines a semigroup. If $\lim_{t\downarrow 0}\|S(t)h - h\| = 0$, we have that the generator A of $S(t)$, defined on $D(A)$ is given by,

$$A = \lim_{t\downarrow 0} \frac{S(t)h(\cdot) - h(\cdot)}{t} = \lim_{t\downarrow 0} \frac{h(\cdot + t) - h(\cdot)}{t} = \frac{\partial}{\partial x} h(x). \tag{2.53}$$

The application of the C_0-semigroup $S(t)$ is understood by studying the abstract Cauchy problem,

$$\frac{d}{dt}u(t) = Au(t), \quad u(0) = u_0,$$

for which, has the solution $t \mapsto S(t)u_0$ if $u_0 \in D(A)$. See theorem 9.2 in Peszat and Zabczyk [37], for the proof.

We specify the setting above to account for H-valued functions, and add an additional term $\phi(t)$ which alters the above equation,

$$\frac{d}{dt}u(t) = Au(t) + \phi(t), \quad u(0) = u_0 \in \mathsf{H}. \tag{2.54}$$

Now if $\phi(t)$ is continuously differentiable and $u_0 \in D(A)$, the *variation-of-constants formula* gives the *mild solution* to (2.54),

$$u(t) = S(t)u_0 + \int_0^t S(t-s)\phi(s)\mathrm{d}s, \quad t \geq 0. \tag{2.55}$$

However, the equation of main interest in this paper, is the linear affine equation,

$$\mathrm{d}X = (AX + \beta(X))\,\mathrm{d}t + \sigma(X)\mathrm{d}Z(t). \tag{2.56}$$

Here X is taking values in H, the drift is split in two parts, one depending on some possibly unbounded operator A with domain $D(A)$ as the generator of a semigroup $S(t)$ on H, and a function $\beta : D(\beta) \to \mathsf{H}$. Moreover, the noise $\sigma : D(\sigma) \to L(H, \mathsf{H})$. is modeled by some square integrable stochastic process $Z(t)$. In order to discuss the solution of equations on the form (2.56), we must establish the notion of a solution to such a system.

We assume that β and σ satisfies the Lipschitz conditions when acted on by S.

Definition 2.6.5. Suppose $D(\beta)$ and $D(\sigma)$ are dense in H, and there exists functions $a, b : (0,\infty) \to (0,\infty)$ satisfying $\int_0^T a(t)\mathrm{d}t < \infty$ and $\int_0^T b^2(t)\mathrm{d}t < \infty$ for all $T < \infty$. Then we call β and σ respectively for *semigroup Lipschitz* if for all $t > 0$, the following holds,

(β) For $x, y \in D(\beta)$

$$\|S(t)\beta(x)\|_{\mathsf{H}} \leq a(t)\left(1 + \|x\|_{\mathsf{H}}\right), \quad \text{and}$$
$$\|S(t)\left(\beta(x) - \beta(y)\right)\|_{\mathsf{H}} \leq a(t)\|x - y\|_{\mathsf{H}}.$$

(σ) For $x, y \in D(\sigma)$

$$\|S(t)\sigma(x)\|_{L_0^2} \leq b(t)\left(1 + \|x\|_{\mathsf{H}}\right), \quad \text{and}$$
$$\|S(t)\left(\sigma(x) - \sigma(y)\right)\|_{L_0^2} \leq b(t)\|x - y\|_{\mathsf{H}}.$$

Furthermore, we provide a concise definition of a solution to (2.56).

Definition 2.6.6. Suppose X_0 is a square integrable $\mathcal{F}_{t_0}$-measurable random variable in H. Then we say that $X : [t_0, \infty) \times \Omega \to \mathsf{H}$ is a *mild solution* to (2.56), starting at t_0, if

$$\sup_{t\in[t_0,T]} \left[\|X(t)\|_{\mathsf{H}}^2\right] < \infty, \quad \text{for all } \ T \in (t_0, \infty),$$

and

$$X(t) = S(t-t_0)X_0 + \int_{t_0}^{t} S(t-s)\beta(X(s))ds + \int_{t_0}^{t} S(t-s)\sigma(X(s))dZ(s),$$

for all $t \geq t_0$.

We restate a result concerning the existence of a solution from Peszat and Zybczyk [37].

Theorem 2.6.7. *Given $t_0 \geq 0$, with X_0 being a $\mathcal{F}_{t_0}$-measurable square integrable random variable in H, and β and σ semigroup Lipschitz, then there exist a unique solution $X(\cdot, t_0, X_0)$ of (2.56).*

If restrict the noise to be driven by a Q-Wiener process and disregard the X dependence in the volatility and put $\sigma(X) = \Sigma \in B(\mathsf{U}, \mathsf{H})$, Da Prato and Zabczyk [14], gives the requirements for (2.56) to have a unique *weak* solution in this weaker situation. A weak solution is a predictable H-valued process $X(t)$ which is Bochner integrable for all $t \in [0, T]$ and,

$$\langle X(t), y\rangle = \langle X(t), y\rangle + \int_0^t \left[\langle X(s), A^*y\rangle + \langle \beta(t), y\rangle\right]\mathrm{d}s + \langle \Sigma W(t), y\rangle, \quad \mathbb{P} - a.s. \tag{2.57}$$

The equation is then

$$\mathrm{d}X(t) = (AX(t) + \beta(t))\mathrm{d}t + \Sigma \mathrm{d}W(t) \tag{2.58}$$

Again we assume that A must generate a C_0-semigroup in H and $\Sigma \in B(\mathsf{U}, \mathsf{H})$.

First, we provide some properties of the *stochastic convolution* with respect to the Q-Wiener process,

$$W_A(t) = \int_0^t S(t-s)B\mathrm{d}W(s), t \geq 0. \tag{2.59}$$

For (2.59) be integrable we must have that

$$\int_0^T \|S(u)B\|_{L_2^0}^2 \mathrm{d}u = \int_0^T \mathrm{Tr}\left(S(u)BQB^*S(u)^*\right)\mathrm{d}u < \infty. \tag{2.60}$$

Hence, if $W_A(t)$ is integrable then

(i) The stochastic convolution $W_A(\cdot))$, is Gaussian.

(ii) The covariance of $W_A(t)$, is given by

$$\mathrm{Cov}(W_A(t)) = \int_0^t S(u)BQB^*S^*(u)\mathrm{d}u, \quad t \in [0,T]. \tag{2.61}$$

Theorem 2.6.8. *[14, Theorem 5.4] If $\beta(\cdot)$ is predictable and Bochner integrable on $[0,T]$, and $X(0)$ is $\mathcal{F}_0$-measurable, then*

$$X(t) = S(t)X(0) + \int_0^t S(t-s)\beta(s)\mathrm{d}s + \int_0^t S(t-s)\Sigma\mathrm{d}W(s), \tag{2.62}$$

is a unique weak solution.

We have now shown the requirements for systems of linear affine type to possess a solution in the case of a square-integrable martingale $Z(t)$ and in the particular case of $Z(t) = W(t)$. Remark that we have only presented the bare necessities to study solutions of the mentioned equation. For an in-depth study of e.g., continuity and regularity of the solutions to (2.56), we refer the reader to Peszat and Zabczyk [37] or Da Prato and Zabczyk [14].

2.7 Financial mathematics

One of the main goals of financial mathematics is to provide fair pricing of financial objects such as bonds and options. A fair price is a price for which neither the buyer nor seller admits an immediate opportunity to earn money without taking any risk. We will describe fair pricing in terms of *arbitrage*, and we will state the fundamental theorem of asset pricing. In this section we follow section 2.5 from Carmona and Tehranchi [13].

Let $\{\psi(t), \phi_1(t), \dots \phi_d(t)\}$ be a $d+1$ dimensional stochastic process which we call a *trading strategy*. Moreover, let $B(t)$ denote a risk-free asset such as the bank account, and define the *financial market* as the $d+1$ dimensional stochastic process $\{B(t), P_1(t), \dots P_d(t)\}$, where the P_i's for $i = 1, \dots, d$ are the underlyings. We will write for short the vectors $\phi(t) = \{\phi_1(t), \dots \phi_d(t)\}$ and $P(t) = \{P_1(t), \dots P_d(t)\}$.

The *portfolio strategy* is then given by the formula

$$V(t) = \psi(t)B(t) + \langle \phi(t), P(t) \rangle, \tag{2.63}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{R}^d$.

Let $0 < t_1 < \dots < t_d$, be d deterministic times such that

$$\phi_j(t) = \sum_{i=1}^{d} \phi_j(t_i) \mathbb{1}_{(t_j, t_{j+1}]}(t),$$

and require that each $\phi_j(t_i)$ is $\mathcal{F}_{t_i}$-measurable for all $i = 1, \dots, d$.

Furthermore, the portfolio strategy must be *self-financing*, which means that we do not allow for any external income or expense to influence the wealth $V(t)$. Therefore, we must have that

$$V(t_{i+1}) - V(t_i) = \psi(t_i)\left(B(t_{i+1}) - B(t_i)\right) + \langle \phi(t), P(t_{i+1}) - P(t_i))\rangle.$$

Moreover, we have by (2.63) for all t_i, $i = 1, \ldots, d$, that

$$\psi(t_i) = \frac{1}{B(t_i)}\left(V(t) - \langle \phi(t), P(t_i)\rangle\right),$$

which lets us establish,

$$\begin{aligned} V(t_{i+1}) - V(t_i) &= \frac{1}{B(t_i)}\left(V(t_i) - \langle \phi(t), P(t)\rangle\right)\left(B(t_{i+1}) - B(t_i)\right) \\ &\quad + \langle \phi(t), P(t_{i+1}) - P(t_i))\rangle \\ &= \frac{B(t_{i+1})}{B(t_i)}\left(V(t_i) - \langle \phi(t), P(t)\rangle\right) - V(t_i) + \langle \phi(t), P(t)\rangle \\ &\quad + \langle \phi(t), P(t_{i+1}) - P(t_i))\rangle. \end{aligned}$$

now we divide by $B(t_{i+1})$ and define $\hat{V}(t) = \frac{V(t)}{B(t)}$ and $\hat{P}(t) = \frac{P(t)}{B(t)}$ to be the discounted wealth and asset prices respectively, such that

$$\begin{aligned} \hat{V}(t_{i+1}) =& \hat{V}(t) - \langle \phi(t), \hat{P}(t)\rangle + \langle \phi(t), \frac{P(t_i)}{B(t_{i+1})}\rangle \\ &+ \langle \phi(t), \hat{P}(t_{i+1}) - \frac{P(t_i)}{B(t_{i+1})}\rangle, \end{aligned}$$

hence

$$\hat{V}(t_{i+1}) - \hat{V}(t_i) = \langle \phi, \hat{P}(t_{i+1}) - \hat{P}(t_i))\rangle. \tag{2.64}$$

If we let $d \to \infty$, we obtain

$$\hat{V}(t) = \hat{V}(0) + \int_0^t \langle \phi(s), d\hat{P}(s)\rangle. \tag{2.65}$$

For the stochastic integral to be well-defined, we must assume that the discounted trading strategy process $\phi_j(t)$ is predictable and integrable for all j, in addition to discounted price process $\hat{P}(t)$ must be a semi-martingale.

If an investor has unlimited financial resources, this may allow for martingale type betting strategies to always yield profit, since the debt amount is not bounded. To avoid such paradoxes, we must provide a technical condition called *admissability*.

Definition 2.7.1. The trading strategy $\{\phi(t)\}_{t\geq 0}$ is said to be *admissible* if the stochastic integral,

$$\int_0^t \langle \phi(s), \mathrm{d}\hat{P}(s)\rangle,$$

is uniformly bounded from below for $t \geq 0$ and $\omega \in \Omega$.

Furthermore, we can now define *arbitrage*.

Definition 2.7.2. An admissible trading strategy $\{\phi(t)\}_{t\geq 0}$ satisfying

$$\mathbb{P}(\int_0^t \langle \phi(s), \mathrm{d}\hat{P}(s)\rangle \geq 0) = 1$$
$$\mathbb{P}(\int_0^t \langle \phi(s), \mathrm{d}\hat{P}(s)\rangle > 0) > 0,$$

is said to admit an *arbitrage* or to have *arbitrage opportunities*.

We conclude this section by stating the *fundamental theorem of asset pricing*.

Theorem 2.7.3. *A trading strategy $\phi(t)$ does not admit any arbitrage opportunities if and only if there exists an equivalent probability measure $\mathbb{Q}$ such that $\hat{P}$ under $\mathbb{Q}$ is a local martingale.*

By equivalence of the measures $\mathbb{P}$ and $\mathbb{Q}$ we mean that $\mathbb{Q}(F) = 0$ if and only if $\mathbb{P}(F) = 0$ for all $F \in \mathcal{F}$, that is, the set of null sets in $\mathbb{Q}$ are equal to that of $\mathbb{P}$. The equivalent probability measure $\mathbb{Q}$, is often called a *risk-neutral* measure.

In practice, we are often in the setting of the historical probability measure $\mathbb{P}$. If we want to price options through the Black and Scholes [9] paradigm, we must derive the explicit risk-neutral probability measure $\mathbb{Q}$. In obtaining the risk-neutral measure from $\mathbb{P}$, we apply techniques of measure change such as the *Girsanov transform* or the *Esscher transform*. We will not, however, spend much time on that matter, but an H-valued version of Girsanov's theorem is stated in the appendix, for the sake of completeness.

3 Interest Rates Models and Cointegration

The object of this thesis is the study of the term structure of interest rates. Providing models for the term structure is imperative for any financial institution by means of, for example, quantifying financial risk, pricing bond securities such as swaps, caps, and floors. Besides, interest rates must also be considered in the monetary policy as part of controlling inflation as well as the issuance of municipal bonds. Moreover, we give an introduction to the Heath-Jarrow-Morton (HJM) framework for modelling forward rates, as well as the Filipovic space H_w, which has proven to be suitable a state-space in the HJM paradigm.

We conclude this chapter by presenting a multi-factor model consisting of linear affine processes taking values in some separable Hilbert space, from which we fit the HJM framework. Finally, we provide some zero-coupon bond prices across the different factors given the identity operator as a coefficient, before we briefly say something about stationarity and HJM models.

We will be working in the continuous-time financial framework, and we will be dealing with continuously compounding interest rates.

3.1 Zero-Coupon Bonds and Interest rates

The main contract of interest in this thesis is the so-called *default-free zero coupon bond*, denoted $\{P(t,T)\}_{t\in[0,T]}$, which describes the time t value of a bond paying the holder one unit of cash at maturity T. By default-free, we mean that we disregard the possibility for the borrower to not be able to pay at maturity time. The bond prices can be expressed by the so-called *instantaneous forward rate* functions $\{f(t,T)\}_{t\in[0,T]}$, which in turn, are derived from *forward rate agreements*.

Given three time points t, T and S, where $t < T < S$, we say the *continuously compounding forward interest rate* at time t for the time interval between S and T, denoted $f(t;T,S)$, is the investment of $P(t,T)$ at time T producing the cash flow $P(t,S)$ at time S. The quantity $f(t;T,S)$ is also known as a *forward rate agreement*, and is accordingly defined as,

$$e^{f(t;T,S)(T-S)}P(t,S) = P(t,T), \tag{3.1}$$

or equivalently,

$$f(t;T,S) = \frac{\log P(t,S) - \log P(t,T)}{S-T}. \tag{3.2}$$

If we let S tend towards T we obtain what is called the *instantaneous forward rate* with maturity T, prevailing at time t,

$$f(t,T) = \lim_{S\downarrow T} f(t;T,S) = -\frac{\partial}{\partial T}\log P(t,T). \tag{3.3}$$

Motivated by (3.3), we define the zero-coupon bond price in terms of the instantaneous forward rates.

Definition 3.1.1. The zero coupon bond prices, denoted $P(t,T)$, is given by

$$P(t,T) = \exp\left(-\int_t^T f(t,s)\mathrm{d}s\right), \quad \text{for } t \in [0,T]. \tag{3.4}$$

The map $T \mapsto P(t,T)$ is called the *discount* curve, which describes the time value of cash, whereas $t \mapsto P(t,T)$ is considered a stochastic process for a fixed maturity.

Moreover, we define the *short rate* and two examples of short rate models. The *instantaneous short rate* $r(t)$ is given in terms of the forward rate at maturity time t,

$$r(t) = f(t,t). \tag{3.5}$$

The short rate appears as the integrand in the money-market account and can be both deterministic and stochastic, the latter being a more realistic description of the evolution of the interest rate. We can model the short rates by stochastic differential equations, for example, the simple Vasicek model which is of *Ornstein-Uhlenbeck* type, given by

$$\mathrm{d}r(t) = \alpha(\beta - r(t))\mathrm{d}t + \sigma\mathrm{d}W(t), \tag{3.6}$$

where α, β, σ are constants, and $W(t)$ is the Wiener process. It can be shown that $\lim_{t\to\infty} E[r(t)] = \beta$, meaning that the model holds the *mean reversion* or property, and β being the mean level. A known drawback of this model is the possibility of obtaining negative interest rates. The *Cox-Ingersoll-Ross* model corrects the possibility of negative interest rates by requiring that $2\alpha\beta \geq \sigma$, in the model

$$r(t) = \alpha(\beta + r(t))\mathrm{d}t + \sigma\sqrt{r(t)}\mathrm{d}W(t). \tag{3.7}$$

But, in the light of recent monetary actions, it is not obvious why we always should neglect the possibility negative interest rates. In fact, the aftermath of the 2008-2009 global financial crisis, lead to several central banks in Europe in addition to Japan, upheld the implementation of a negative interest rate policy(NIRP) [1]. Nevertheless, this debate is beyond the scope of this thesis.

A disadvantage of modelling yield curves using short rate models, is that they only admit one source of noise. Naturally, one cannot incorporate maturity-specific noise in the yield curves in the case of a single source of noise. Hence, there has been proposed models for the *forward rate* instead. Heath, Jarrow and Morton [20] proposed the following model for the forward rate

$$\mathrm{d}f(t,T) = \alpha(t,T)\mathrm{d}t + \sum_{j=1}^{\mathrm{d}} \sigma_j(t,T)\mathrm{d}W_j(t), \tag{3.8}$$

Here, the dynamics of $T \mapsto f(t,T)$ are assumed to be Itô processes, where the drift $\alpha(t,T)$ takes values in $\mathbb{R}$, and $\sigma_j(t,T)$ being an adapted stochastic process on $\mathbb{R}$ and $W_j(t)$ a d-dimensional Wiener process with $d \times d$ covariance matrix Q.

We refer to (3.8), as the HJM framework. For pricing purposes, we want the dynamics of the HJM equation to satisfy the so-called *HJM drift condition*, which ensure that the discounted bond price process is a local martingale under risk-neutral measure $\mathbb{Q}$. In Heath, Jarrow and Morton [20], they found that in order for the discounted bond proces process to be a local martingale under $\mathbb{Q}$, the drift of (3.8) must be on the form

$$\alpha(t,T) = \langle \sigma(t,T), Q \int_t^T \sigma(t,u)\mathrm{d}u \rangle, \tag{3.9}$$

where the inner product is just the scalar product, with $\int_t^T \sigma(t,u)\mathrm{d}u$ being a d-dimensional vector integral.

For the sake yield curve modelling, the forward rate models provides us with d sources of noise, which is an improvement of the aforementioned short rate models. Theoretically, we would like $d \to \infty$, in order to model yield curve as correctly as possible with respect to noise. This motivates the need for a function space valued forward rate paradigm consisting of infinite-dimensional noise, hence we chose to study the HJM equation for Hilbert-valued forward rates. To this end, we introduce a suitable function space for studying forward curves before we turn back to the HJM equation.

3.2 The Forward Curve Space H_w

A natural choice of state-space for modeling curves is the separable Hilbert space $L^2(\mathbb{R}_+)$. However, Filipovic [17] introduced the so-called *Filipovic space* providing a framework for studying forward curves. A shortcoming of the

[1]See Boungou [12] for a discussion of the economical motivation and implication of an NIRP.

classical L^2-spaces, is that the elements are equivalence classes of functions, making evaluation troublesome.

Let $f(x)$, for $x \in \mathbb{R}_+$ denote any forward curve. It is natural to assume some regularity in $f'(x)$, thus we let,

$$\int_{\mathbb{R}_+} |f'(x)|^2 \mathrm{d}x < \infty,$$

moreover, in forcing a decaying or flattening structure of $f(x)$, a weight function $w(x)$ is introduced, such that

$$\int_{\mathbb{R}_+} |f'(x)|^2 w(x) \mathrm{d}x < \infty.$$

In order to define a norm without turning to equivalence classes, the constant term of the squared short rate is added. We now give a formal definition of the *Filipovic space* denoted H_w.

Definition 3.2.1. Let $w : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonely increasing measurable function satisfying $w(0) = 1$ and $\int_0^\infty w(x)^{-1} \mathrm{d}x < \infty$. We define

$$||f||_\omega^2 = f^2(0) + \int_0^\infty w(x)(f'(x))^2 \; \mathrm{d}x < \infty, \tag{3.10}$$

where f' is the weak derivative of f, with associated inner product

$$\langle f, g \rangle_w = f(0)g(0) + \int_0^\infty \omega(x) f'(x) g'(x) \; \mathrm{d}x. \tag{3.11}$$

We define

$$\mathsf{H}_\omega = \{ f \in L^1(\mathbb{R}_+) | \text{There exists } f' \in L^1(\mathbb{R}_+) \text{ and } ||f||_\omega < \infty \}$$

For the rest of this section when we write $\langle \cdot, \cdot \rangle$ we imply the Filipovic inner product as in (3.11).

We list some of the properties of H_ω, due to Filipovic [17].

(i) H_ω equipped with $|| \cdot ||_\omega$ defines a separable Hilbert space.
(ii) The shift operator $S(t) : f \mapsto f(t + \xi)$ for $t \geq 0$, defines a C_0-semigroup on H_ω.
(iii) The evaluation map $\delta : f \mapsto f(x)$ defines a linear functional on H_ω.
(iv) Given the operator $\mathcal{W} f = \sqrt{w} f'$, the map $(\delta_0, \mathcal{W}) : \mathsf{H}_w \to \mathbb{R} \times L^2(\mathbb{R}_+), f \mapsto (f(0), \sqrt{(w)} f')$ defines an isometric isomorphism of H_w and $\mathbb{R} \times L^2(\mathbb{R}_+)$.

In conclusion we have that H_w equipped with the norm $|| \cdot ||_w$ defines is a separable Hilbert space which is isometric to $\mathbb{R} \times L^2(\mathbb{R}_+)$. The fact that the shift operator defines a C_0-semigroup on H_w is necessary for the existence of solutions in the function valued HJM model we will present in the next section. Moreover, since the evaluation operator δ is a linear functional on H_w, it is valid to evaluate functions $h \in \mathsf{H}_w$, in contrast to the equivalence classes in L^2-spaces.

Moreover we want to characterize covariance operators on H_w. As we have seen by property (iv) above, the space H_w resembles in some sense $L^2(\mathbb{R}_+)$. Recall covariance operators in $L^2(\mathbb{R}_+)$ can be defined as an integral operator

$$Qf(x) = \int_0^\infty q(x,y)f(y)\mathrm{d}y, \tag{3.12}$$

for $f \in L^2(\mathbb{R}_+)$ and square-integrable kernel $q(x,y)$. We should therefore expect a similar covariance structure in H_w. The following result regarding Hilbert–Schmidt operators in H_w is due to Benth and Krühner [2]. The idea is that given a complete specification of the Hilbert–Scmidt operators in H_w, we can use the fact that any positive semidefinite trace class operator is the square of a symmetric Hilbert–Schmidt operator [2].

Theorem 3.2.2. *Given a Hilbert–Scmidt operator $\mathcal{C}$ on H_w. Then there are $c \in \mathbb{R}$,and $g, h \in \mathsf{H}_w$, where $h(0) = g(0) = 0$, and a square integrable function $b : \mathbb{R}_+^2 \to \mathbb{R}$, such that*

$$\mathcal{C}f(x) = cf(0) + \langle g, f\rangle + f(0)h(x) + \int_0^\infty q(x,z)f'(z)\mathrm{d}z, \tag{3.13}$$

where the kernel q is given by

$$q(x,z) = \int_0^x \sqrt{\frac{w(z)}{w(y)}}b(y,z)\mathrm{d}z$$

In addition we have that $\mathcal{C}$ is symmetric if b is symmetric and $g = h$.

We now state a corollary which describes all covariance operators on H_w.

Corollary 3.2.3. *[2, Corollary 4.2] Assume the same specification on c, g, h, b, q from Theorem 3.2.2. Let $\ell(x,z) = h'(z)w(z) + q(x,z),$. If $\ell(0,\cdot)/\sqrt{w} \in L^2(\mathbb{R})$ and the map $(x,z) \mapsto \frac{w(z)}{w(x)}(\partial_1 \ell(x,z))^2$ is symmetric and integrable, then class of positive trace class operators on H_w are given by*

$$Qf(x) = \big(f(0)c + \int_0^\infty \ell(0,z)f'(z)\mathrm{d}z\big)\big(c + \int_0^x \frac{\ell(0,z)}{w(z)}\mathrm{d}z\big) \tag{3.14}$$

$$+ f(0)\int_0^\infty \frac{\ell(x, z\ell(0,z))}{w(z)}\mathrm{d}z \tag{3.15}$$

$$+ \int_0^\infty \int_0^\infty \ell(x,z)\partial_1\ell(z,y)\mathrm{d}z f'(y)\mathrm{d}y, \tag{3.16}$$

Thus Q above gives a complete characterization of covariance operators on H_w. We conclude with a technical result regarding the inner product - covariance operator representation of $\langle f(x), Q\int_0^x f(y)\mathrm{d}y\rangle$, which we later on will recognize as the no-arbitrage condition in the HJM model.

[2] See Benth and Krühner [2]

Proposition 3.2.4. *Let Q denote the covariance operator in H_w. If $f \in \mathsf{H}_w$, together with the specifications from Theorem 3.2.2, then*

$$\langle f(x), Q\int_0^x f(y)\mathrm{d}y\rangle = \int_0^\infty w(t)f(x)(Qf)'(x)\mathrm{d}x, \tag{3.17}$$

with

$$(Qf)'(x) = h'(x)(\langle h, f\rangle + c) + f(0)\int_0^\infty h'(z)\partial_1\ell(x,z)\mathrm{d}z + \int_0^\infty k(x,y)f'(y)\mathrm{d}y. \tag{3.18}$$

Proof. Let $F(x) = \int_0^x f(y)\mathrm{d}y$. Note that $\langle QF(x), f(x)\rangle = \langle F(x), Qf(x)\rangle$, by the self-adjoint property of Q. Furthermore we compute the derivative of the Q applied to f,

$$\begin{aligned}(Qf)'(x) &= \frac{\ell(0,x)}{w(x)}(\int_0^\infty \ell(0,z)f'(z)\mathrm{d}z + c) + f(0)\int_0^\infty \frac{\ell(0,z)\partial_1\ell(x,z)}{w(z)}\mathrm{d}z \\ &\quad + \int_0^\infty\int_0^\infty \partial_1\ell(x,z)\partial_1\ell(z,y)\mathrm{d}z f'(y)\mathrm{d}y \\ &= h'(x)(\langle h,f\rangle + c) + f(0)\int_0^\infty h'(z)\partial_1\ell(x,z)\mathrm{d}z \\ &\quad + \int_0^\infty\int_0^\infty \partial_1\ell(x,z)\partial_1\ell(z,y)\mathrm{d}z f'(y)\mathrm{d}y.\end{aligned}$$

Define $k(x,y) = \partial_1\ell(x,z)\partial_1\ell(z,y)$, hence

$$(Qf)'(x) = h'(x)(\langle h, f\rangle + c) + f(0)\int_0^\infty h'(z)\partial_1\ell(x,z)\mathrm{d}z + \int_0^\infty k(x,y)f'(y)\mathrm{d}y.$$

Then we have

$$\langle F(x), Qf(x)\rangle = \int_0^\infty w(x)f(x)(Qf)'(x)\mathrm{d}x$$

■

We will in this thesis work with the sub-class of the covariance derivative of the operator Q in H_w, where we assume the h in Theorem 3.2.2 is such that $h'(x) = 0$ for all x. Such an operator we denote by Q_0. This in turn, transforms the no-arbitrage condition to an integral operator depending on w, f and Q. Moreover

Corollary 3.2.5. *If $k(x,y) = \partial_1\ell(x,z)\partial_1\ell(z,y)$, so that*

$$(Q_0 f)'(x) = \int_0^\infty k(x,y)f'(y)\mathrm{d}y, \tag{3.19}$$

then for $g(x) = k \in \mathsf{H}_w$ for all $x \in \mathbb{R}_+$ we have $(Q_0 g)'(\cdot) = 0$, which implies

$$\langle F(x), Qf(x)\rangle = \int_0^\infty w(x)f(x)(Qf)'(x)\mathrm{d}x = 0. \tag{3.20}$$

The above corollary states that the no-arbitrage condition defined by the inner product is just zero, in H_w.

3.3 Heath-Jarrow-Morton-Musiela Methodology

We briefly introduced the Heath-Jarrow-Morton framework in 2.1, in this section, however, we will consider the forward curves and Wiener noise as Hilbert-valued functions, and therefore we will state a drift condition analogous to (3.9).

We follow the notation from Peszat and Zabchyk [37] as well as their treatise of the Lévy noise generalization of (3.8). We will, however, assume that the state space of the forward curves is a separable Hilbert space H, and let $W(t)$ be a Q-Wiener process defined on the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ taking values in a separable Hilbert space U. This alters the dynamics of the equation studied in Peszat and Zabchyk [37], to that of

$$\mathrm{d}f(t,T) = \alpha(t,T)\mathrm{d}t + \langle \sigma(t,T), \mathrm{d}W(t)\rangle_{\mathsf{U}}, \quad t \leq T, \tag{3.21}$$

or equivalently,

$$f(t,T) = f(0,T) + \int_0^t \alpha(s,T)\mathrm{d}t + \int_0^t \langle \sigma(s,T), \mathrm{d}W(s)\rangle_{\mathsf{U}}. \tag{3.22}$$

We suppose that $\alpha(t,T)$ is H-valued, and $\sigma(t,T) \in B(\mathsf{U},\mathsf{H})$ are predictable processes, with $\alpha(t,T) = \sigma(t,T) = 0$ for $t \geq T$ and $T \geq 0$. Furthermore, we chose to study (3.22) with the second variable in terms of time to maturity, hence we let $x = T - t$, the so-called *Musiela parametrization*, see Musiela [33]. This reveals a surprising link between HJM modeling and stochastic partial differential equations. Following Peszat and Zabchyk [37], we present the derivation of the *Heath–Jarrow–Morton–Musiela* equation. Define

$$\begin{aligned} f(t)(x) &= f(t,t+x) \\ \alpha(t)(x) &= \alpha(t,t+x) \\ (b(t)u)(x) &= \langle \sigma(t,t+x), u\rangle_{\mathsf{U}}, \end{aligned}$$

and let $S(t)g(x) = g(t+x)$ denote the shift semigroup. Then we transform (3.22) to,

$$\begin{aligned} f(t,x) &= f(0,t+x) + \int_0^t \alpha(s,t+x)\mathrm{d}s + \int_0^t \langle \sigma(s,t+x), \mathrm{d}W(s)\rangle_{\mathsf{U}} \\ &= f(0)(t+x) + \int_0^t \alpha(s)(t-s+x)\mathrm{d}s + \int_0^t b(s)(t-s+x)\mathrm{d}W(s) \\ &= S(t)f(0)(x) + \int_0^t S(t-s)\alpha(s)(x)\mathrm{d}s + \int_0^t S(t-s)b(s)(x)\mathrm{d}W(s), \end{aligned}$$

hence we write,

$$f(t) = S(t)f(0) + \int_0^t S(t-s)\alpha(s)\mathrm{d}s + \int_0^t S(t-s)b(s)\mathrm{d}W(s). \tag{3.23}$$

Now, since the infinitesimal generator of the shift semigroup is given by $A = \frac{\partial}{\partial x}$, we write (3.23) as the mild solution to

$$\mathrm{d}f(t) = (\frac{\partial}{\partial x} f(t) + \alpha(t))\mathrm{d}t + b(t)\mathrm{d}W(s). \tag{3.24}$$

For the discounted zero-coupon price dynamics to be local martingales, we present the following theorem due to Peszat and Zabczyk [37], which provides us with the arbitrage-free dynamics of (3.23).

Theorem 3.3.1. *[37, Theorem 20.3] The discounted bond price process of* (3.21) *are local martingales under* $\mathbb{Q}$ *if and only if*

$$\int_0^\theta \alpha(t,u)\mathrm{d}u = J(\int_0^\theta \sigma(t,u)\mathrm{d}u), \tag{3.25}$$

where

$$J(y) = \frac{1}{2}\langle Qy, y\rangle_{\mathsf{U}}$$

for all $\theta < T$ *, and* P *a.s. for* $t \in [0, \theta]$.

The main part of their proof is applying Itô's formula for Hilbert-valued semimartingales to $\Psi(\langle 1_{[0,T]}, f(t)\rangle_{\mathsf{H}})$, for $\Psi \in C^2(\mathbb{R})$.

Before we turn to the HJM condition for (**??**), we propose a result on the Fréchet derivative of $J(u)$ for the Q-Wiener process.

Proposition 3.3.2. *Consider the* Q*-Wiener process* $W(t)$ *with Laplace transform given by* $J(u) = \langle Qu, u\rangle$*. Then the Fréchet derivative of* $J(u)$ *is given by*

$$DJ(u)(\cdot) = \langle Qu, \cdot\rangle. \tag{3.26}$$

Proof. Direct computation gives that

$$\begin{aligned} J(z+h) - J(z) &= \frac{1}{2}\langle Q(z+h), z+h\rangle - \frac{1}{2}\langle Qz, z\rangle \\ &= \frac{1}{2}\langle Qz, z+h\rangle + \frac{1}{2}\langle Qh, z+h\rangle - \frac{1}{2}\langle Qz, z\rangle \\ &= \frac{1}{2}\langle Qz, h\rangle + \frac{1}{2}\langle Qh, z+h\rangle \\ &= \langle Qz, h\rangle + \frac{1}{2}\langle Qh, h\rangle. \end{aligned}$$

Thus we claim that $DJ(z)(\cdot) = \langle Qz, \cdot\rangle$, and by definition we obtain

$$\lim_{\|h\|\to 0} \frac{\|J(z+h) - J(z) - DJ(z)(h)\|_{\mathbb{R}}}{\|h\|_{\mathsf{U}}} = \lim_{\|h\|\to 0} \frac{|\langle Qh, h\rangle_{\mathsf{U}}|}{\|h\|_{\mathsf{U}}}. \tag{3.27}$$

which by Cauchy-Schwarz ensures that $\frac{1}{2}\langle Qh, h\rangle_{\mathsf{U}} \leq \frac{1}{2}\|Qh\|\|h\|$, proving the assertion. ■

The derivation of the HJMM equation in Peszat and Zabchyk [37] is somewhat vague, when it comes to the no-arbitrage condition. To that end, we carefully assemble the no arbitrage condition. With Q-Wiener noise, the no-arbitrage condition simplifies to $\langle b(t)(\xi), DJ\left(\int_0^\xi b(t)(u)\mathrm{d}u\right)\rangle_{\mathsf{H}}$, and $J(x) = \frac{1}{2}\langle Qx, x\rangle$ where Q is the covariance operator of $W(t)$. Hence

$$\langle b(t)(\xi), DJ\left(\int_0^\xi b(t)(u)\mathrm{d}u\right)\rangle = \langle b(t)(\xi), \frac{\partial}{\partial\xi}\left(\frac{1}{2}\langle Q\int_0^\xi b(t)(u)\mathrm{d}u, \int_0^\xi b(t)(u)\mathrm{d}u\rangle\right\rangle.$$

Moreover, we illustrate differentiation on inner products, which is quite similar to that of the product rule in single variable calculus. Let $f, g \in \mathsf{H}$, hence

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle f(t), g(t)\rangle = \lim_{h\to 0}\left(\frac{\langle f(t+h), g(t+h)\rangle - \langle f(t), g(t)\rangle}{h}\right)\\
&= \lim_{h\to 0}\left(\frac{\langle f(t+h), g(t+h)\rangle + \langle f(t+h), g(t)\rangle - \langle f(t+h), g(t)\rangle - \langle f(t), g(t)\rangle}{h}\right)\\
&= \lim_{h\to 0}\left(\frac{\langle f(t+h) - f(t), g(t+h)\rangle - \langle f(t+h), g(t+h) - g(t)\rangle}{h}\right)\\
&= \langle f'(t), g(t)\rangle + \langle f(t), g'(t)\rangle.
\end{aligned}$$

Which for (3.28) implies,

$$\begin{aligned}
&\frac{\partial}{\partial\xi}\left(\frac{1}{2}\langle Q\int_0^\xi b(t)(u)du, \int_0^\xi b(t)(u)du\rangle\right) =\\
&\frac{1}{2}(\langle\frac{\partial}{\partial\xi}Q\int_0^\xi b(t)(u)du, \int_0^\xi b(t)(u)du\rangle + \langle Q\int_0^\xi b(t)(u)du, \frac{\partial}{\partial\xi}\int_0^\xi b(t)(u)du\rangle),
\end{aligned}$$

which with the assumption that the covariance- and differential operator commute, we obtain by the fundamental theorem of analysis and the fact that Q is self-adjoint,

$$\begin{aligned}
\frac{1}{2}(\langle Qb(t)(\xi), \int_0^\xi b(t)(u)\mathrm{d}u\rangle + \langle Q\int_0^\xi b(t)(u)\mathrm{d}u, b(t)(\xi)\rangle)&\\
= \langle b(t)(\xi), \int_0^\xi Qb(t)(u)\mathrm{d}u\rangle.&
\end{aligned}$$

This is verified by another approach to the no-arbitrage theorem, where we instead consider the Fréchet derivative of $J(\cdot)$. We identify the functional $b(t)(\xi)u = \langle\sigma(t, t+\xi), u\rangle$ by the element $\tilde{\sigma}(t, t+\xi)$, and identify the the Fréchet derivative $DJ(z)(\cdot) = \langle Qz, \cdot\rangle$ by the element $\tilde{D}J(z) = Qz$, hence

$$\begin{aligned}
\langle b(t)(\xi), DJ\left(\int_0^\xi b(t)(u)\mathrm{d}u\right)\rangle &= \langle\tilde{\sigma}(t, t+\xi), DJ\left(\int_0^\xi \tilde{\sigma}(t, t+\xi)\mathrm{d}u\right)\rangle\\
&= \langle\tilde{\sigma}(t, t+\xi), Q\int_0^\xi \tilde{\sigma}(t, t+\eta)\mathrm{d}\eta\rangle.
\end{aligned}$$

Consequently, we write $\eta = t + \eta$ for short in the second argument, and then we state the Heath-Jarrow-Morton-Musiela equation

$$\mathrm{d}X(t,\xi) = \left(\frac{\partial}{\partial \xi} X(t,\xi) + \langle \tilde{\sigma}(t,\xi), Q \int_0^{\xi} \tilde{\sigma}(t,\eta)\mathrm{d}\eta \rangle \right) \mathrm{d}t \qquad (3.28)$$
$$+ \tilde{\sigma}(t,\xi)\mathrm{d}W(t),$$

If we choose $\mathsf{U} = \mathsf{H} = \mathsf{H}_w$, we must specify that the Wiener process $W(t) = W(t,\xi)$ has two arguments after applying a semigroup operation.

3.4 Towards a cointegrated HJMM model

Motivated by Benth and Süss [4], we lay the grounds for further work within a cointegrated interest rate model. Given two stochastic processes $g_1(t)$ and $g_2(t)$ they may not alone be stationary, however, if there exist a, b such that the linear combination $ag_1(t) + bg_2(t)$ is stationary, we say that $g_1(t)$ and $g_2(t)$ are *cointegrated*. We can think of the $g(t)$'s as a stochastic process describing the forward curve of e.g., Norway, US, and EU, where t denotes the date. For that reason, we present a multi-dimensional forward curve process $\boldsymbol{g}(t) = (g_1(t), g_2(t), \dots, g_n(t))$, which we model according to the HJM framework. Such a model can be fit to describe an international interest rate marked, from which one can study the complex relationships between the different forward curves.

Moreover, we let each forward rate process consist of two sources of risk, defined as linear affine processes taking values in a Hilbert space H. For the reader's convenience, we first present a single-market two-factor model and extend it to a multi-market two-factor model. We let the curves in section evolve according to Musiela's parametrization, i.e., by $g(t,\xi)$, we mean $g(t, t+\xi)$, where the second argument is bounded by T.

Single market two-factor forward rate process

We model the forward curves as a sum of two H-valued linear affine processes. We define the forward curve as

$$g(t,\xi) = X(t,\xi) + Y(t,\xi), \qquad (3.29)$$

where the dynamics of X and Y is given by

$$\mathrm{d}X(t,\xi) = (A_X(t,\xi)X(t,\xi) + F_X(t))\,\mathrm{d}t + \Sigma_X \mathrm{d}W_1(t) \qquad (3.30)$$
$$\mathrm{d}Y(t,\xi) = (A_Y(t,\xi)Y(t,\xi) + F_Y(t))\,\mathrm{d}t + \Sigma_Y \mathrm{d}W_2(t). \qquad (3.31)$$

For the sake of generality, we do not want to be too strict when defining the above operators. We assume that A_X, A_Y are two possibly unbounded operators. The second drift terms F_X and F_Y are assumed to depend on only t

initially, and taking values in H. The noise $W_1(t), W_2(t)$ are possibly correlated U-valued Q-Wiener processes, and $\Sigma_X, \Sigma_Y \in L(\mathsf{U}, \mathsf{H})$.

It follows that

$$\begin{aligned} \mathrm{d}g(t,\xi) = &(A_X X(t,\xi) + A_Y Y(t,\xi) + F_Y(t) + F_X(t))\,\mathrm{d}t \\ &+ \Sigma_X dW_1(t) + \Sigma_Y \mathrm{d}W_2(t). \end{aligned} \tag{3.32}$$

We define the product Hilbert space $V = \mathsf{U}\times\mathsf{U}$, and identify $\Sigma \in V$ as the linear map $\Sigma(f,h) = \Sigma_X f + \Sigma_Y h$. Let the noise be given by the two-dimensional noise process $Z(t) = (W_1(t), W_2(t))$ on V. We rewrite (3.32),

$$\mathrm{d}g(t,\xi) = (A_X X(t,\xi) + A_Y Y(t,\xi) + F_Y(t) + F_X(t))\,\mathrm{d}t + \Sigma \mathrm{d}Z(t).$$

The no-arbitrage condition from the HJMM equation states that our drift term $F_{XY} := F_X + F_y$ must satisfy,

$$\begin{aligned} F_{XY} &= \langle \Sigma, Q\int_0^\xi \Sigma(\eta)\mathrm{d}\eta\rangle_V \\ &= \langle \Sigma_X, Q\int_0^\xi \Sigma_X(\eta)\mathrm{d}\eta\rangle_U + \langle \Sigma_Y, Q\int_0^\xi \Sigma_Y(\eta)\mathrm{d}\eta\rangle_U, \end{aligned}$$

where the inner product is induced by the Hilbertianity of the product space V.

If we also let $A_X = A_Y = \frac{\partial}{\partial\xi}$, we can write (3.32) as

$$\begin{aligned} \mathrm{d}g(t,\xi) = &\left(\frac{\partial}{\partial\xi} r(t,\xi) + \langle \Sigma_X, Q\int_0^\xi \Sigma_X(\eta)\mathrm{d}\eta\rangle_U + \langle \Sigma_Y, Q\int_0^\xi \Sigma_Y(\eta)\mathrm{d}\eta\rangle_U\right)\mathrm{d}t \\ &+ \Sigma \mathrm{d}Z(t). \end{aligned}$$

Let now H_w be the state space of $g(t,\xi)$ and the noise. Let $\Sigma_X(f) = \sigma_X(\cdot)f(\cdot)$ and $\Sigma_Y(f) = \sigma_Y(\cdot)f(\cdot)$ with $\sigma_X(\cdot) = \sigma_Y(\cdot) = 1$ and $\sigma_X(\cdot), \sigma_Y(\cdot) \in \mathsf{H}_w$. Suppose the covariance of the Wiener processes is $Q = Q_0$. By identification we have that

$$F_{XY} = \xi\langle 1, Q1\rangle_{\mathsf{H}_w} + \xi\langle 1, Q1\rangle_{\mathsf{H}_w} = 2\xi\langle 1, Q1\rangle_{\mathsf{H}_w}.$$

which yields the following dynamics for the model,

$$\begin{aligned} \mathrm{d}X(t,\xi) &= (\partial_\xi X(t,\xi) + \xi\langle 1, Q(1)\rangle)\,\mathrm{d}t + \mathrm{d}W_1(t,\xi) \\ \mathrm{d}Y(t,\xi) &= (\partial_\xi Y(t,\xi) + \xi\langle 1, Q(1)\rangle)\,\mathrm{d}t + \mathrm{d}W_2(t,\xi), \end{aligned}$$

where we simplify the derivative operator $\frac{\partial}{\partial\xi} = \partial_\xi$. Recall that the covariance operator in H_w applied to constants is just zero, simplifying the above to

$$\begin{aligned} \mathrm{d}X(t,\xi) &= (\partial_\xi X(t,\xi))\,\mathrm{d}t + \mathrm{d}W_1(t,\xi) \\ \mathrm{d}Y(t,\xi) &= (\partial_\xi Y(t,\xi))\,\mathrm{d}t + \mathrm{d}W_2(t,\xi). \end{aligned}$$

The mild solutions of X and Y are given by

$$X(t) = S(t)X(0) + \int_0^t S(t-u)\mathrm{d}W_1(u)$$
$$Y(t) = S(t)Y(0) + \int_0^t S(t-u)\mathrm{d}W_2(u),$$

where $S(t)$ is the semigroup generated by ∂_ξ. The mild solution of (3.29) is then given by,

$$g(t) = S(t)g(0) + \int_0^t S(t-u)\mathrm{d}W_1(u) + \int_0^t S(t-u)\mathrm{d}W_2(u). \tag{3.33}$$

We conclude this section by relating the forward rate pricing to the implied short rates. We have that,

$$f(t,T) = \delta_{T-t}X(t, t+\cdot) \tag{3.34}$$

which brings us back to the price of a contract which delivers at maturity time T. Moreover the implied short rates are given by

$$r(t) = \delta_0 X(t, t+\cdot). \tag{3.35}$$

Notice that

$$\delta_x S(t)g(\cdot) = \delta_x g(\cdot + t) = g(x+t), \tag{3.36}$$

which implies that $\delta_x S(t) = \delta_{x+t}$, thus,

$$\begin{aligned}\delta_{T-t}g(t,t+\cdot) &= \delta_{T-t}S(t)g(0,t+\cdot) + \int_0^t \delta_{T-t}S(t-u)\mathrm{d}W_1(u) \\ &\quad + \int_0^t \delta_{T-t}S(t-u)\mathrm{d}W_2(u) \\ &= g(0,T) + \int_0^t \delta_{T-u}\mathrm{d}W_1(u) + \int_0^t \delta_{T-u}\mathrm{d}W_2(u)\end{aligned}$$

Also, the implied short rate is straightforwardly computed,

$$\delta_0 g(t, t+\cdot) = g(0,t) + \int_0^t \delta_{t-u}\mathrm{d}W_1(u) + \int_0^t \delta_{t-u}\mathrm{d}W_2(u).$$

If we now let the volatility be constant, that is, $\sigma_X(\cdot) = \sigma_Y(\cdot) = \sigma \in \mathsf{H}_w$, we obtain for the no-arbitrage condition

$$F_{XY} = \xi\sigma^2\langle 1, Q(1)\rangle_w + \xi\sigma^2\langle 1, Q(1)\rangle_w = 2\xi\sigma^2\langle 1, Q(1)\rangle_w,$$

which again is just zero. The only difference in this case is the volatility-term in the stochastic integral, meaning we have the following mild solutions,

$$X(t) = X(0) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_1(u)$$
$$Y(t) = Y(0) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_2(u).$$

Multi-market two-factor forward rate model

We extend the former model to account for two forward rate processes, $g_1(t,\xi)$ and $g_2(t,\xi)$ given by

$$g_1(t,\xi) = X(t,\xi) + Y(t,\xi)$$
$$g_2(t,\xi) = X(t,\xi) + Z(t,\xi),$$

where we use the same X and Y as in (3.31), and define Z accordingly,

$$\mathrm{d}Z(t,\xi) = (A_Z(t,\xi)Z(t,\xi) + F_Z(t))\,\mathrm{d}t + \Sigma_Z \mathrm{d}W_3(t).$$

Following the same procedure as before forcing $A_X = A_Z = A_Y = \frac{\partial}{\partial\xi}$, our model has the dynamics,

$$\begin{aligned}\mathrm{d}g_1(t,\xi) &= \Big(\frac{\partial}{\partial\xi}X(t,\xi) + \frac{\partial}{\partial\xi}Y(t,\xi) + F_X(t) + F_Y(t)\Big)\,\mathrm{d}t \\ &\quad + \Sigma_X \mathrm{d}W_1(t) + \Sigma_Y \mathrm{d}W_2(t), \\ \mathrm{d}g_2(t,\xi) &= \Big(\frac{\partial}{\partial\xi}X(t,\xi) + \frac{\partial}{\partial\xi}Z(t,\xi) + F_X(t) + F_Z(t)\Big)\,\mathrm{d}t \\ &\quad + \Sigma_X \mathrm{d}W_1(t) + \Sigma_Z \mathrm{d}W_3(t).\end{aligned}$$

Thus we can write the equation on matrix operator form

$$\mathrm{d}\boldsymbol{g} = (\begin{bmatrix}\frac{\partial}{\partial\xi} & 0 \\ 0 & \frac{\partial}{\partial\xi}\end{bmatrix}\begin{bmatrix}g_1(t,\xi) \\ g_2(t,\xi)\end{bmatrix} + \begin{bmatrix}F_1(t) \\ F_2(t)\end{bmatrix})\mathrm{d}t + \begin{bmatrix}\Sigma_X & \Sigma_Y & 0 \\ \Sigma_X & 0 & \Sigma_Z\end{bmatrix}\begin{bmatrix}\mathrm{d}W_1(t) \\ \mathrm{d}W_2(t) \\ \mathrm{d}W_3(t)\end{bmatrix} \tag{3.37}$$

We define $\Sigma : \mathsf{U}^{2\times 3} \to \mathsf{H}^2$ and $Z(t) = (W_1(t), W_2(t), W_3(t))^T$, and let $\Sigma_{XY}, \Sigma_{XZ} \in L(V,H)$ where $\Sigma_{XY}(f,h) = \Sigma_X f + \Sigma_Y h$ and $F_{XY} = F_X + F_Y$.

$$\mathrm{d}\boldsymbol{g}(t,\xi) = (\partial_\xi \boldsymbol{g}(t,\xi) + F_{XY})\mathrm{d}t + \Sigma \mathrm{dZ(t)} \tag{3.38}$$

The operators Σ_{XZ} and F_{XZ} is defined analogously. To satisfy the no-arbitrage condition in this model, the drift terms must be on the form

$$F_{XY} = \langle \Sigma_X, Q\int_0^\xi \Sigma_X(\eta)\mathrm{d}\eta\rangle_U + \langle \Sigma_Y, Q\int_0^\xi \Sigma_Y(\eta)\mathrm{d}\eta\rangle_U$$
$$F_{XZ} = \langle \Sigma_X, Q\int_0^\xi \Sigma_X(\eta)\mathrm{d}\eta\rangle_U + \langle \Sigma_Z, Q\int_0^\xi \Sigma_Z(\eta)\mathrm{d}\eta\rangle_U.$$

Let $\Sigma_X = \Sigma_Y = \Sigma_Z = \sigma$ like in the previous forward rate model with $\mathsf{U} = \mathsf{H} = \mathsf{H}_w$. Thus

$$g_1(t) = S(t)g_1(0) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_1(u) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_2(u),$$
$$g_2(t) = S(t)g_2(0) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_1(u) + \sigma^2\int_0^t S(t-u)\mathrm{d}W_3(u).$$

This approach may be generalized to account for a finite number of curves, say $g_1, \dots, g_n$ with two sources of noise. In theory, there is no restrictions in extending such a model to include additional sources of noise. Standard methods of approximating yield curves, such as the Nelson-Siegel or Nelson-Siegel-Svensson family, however, consists of three or four parameters, which may suggest some redundancy in choosing too many noise components. We will define the Nelson-Siegel functions later in this thesis.

Moreover, the multi-market two-factor forward curve model is defined as follows. Let $X(t,\xi)$ denote the common noise source, and $Y_i(t,\xi)$, the specific risk corresponding to each curve for $i = 1, \dots, n$. Let $X(t,\xi)$ be defined as in (3.30) and suppose,

$$Y_i(t,\xi) = (\frac{\partial}{\partial \xi} Y_i(t,\xi) + F_{Y_i}(t))\mathrm{d}t + \Sigma_{Y_i} \mathrm{d}W(t), \quad \text{for } 1 \le i \le n.$$

Define $F_i(t) = F_X(t) + F_{Y_i}(t)$ for $i = 1, \dots, n$ so that we may put $F(t) = (F_1(t), \dots, F_n(t))^T : \mathbb{R}_+ \to \mathsf{H}_w^n$. Furthermore, we define the matrix operator $\Sigma : (\mathbb{R}_+ \to L(\mathsf{H}^{n\times(n+1)}, \mathsf{H}^n)$ by,

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{Y_1} & 0 & 0 & \cdots & 0 \\ \Sigma_X & 0 & \Sigma_{Y_2} & 0 & \cdots & 0 \\ \Sigma_X & 0 & 0 & \Sigma_{Y_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \Sigma_X & 0 & 0 & \cdots & 0 & \Sigma_{Y_n} \end{bmatrix}. \tag{3.39}$$

whereas $Z(t) = (W_1(t), \dots, W_{n+1}(t))^T$. We write the differential equation similar to (3.38), which gives,

$$\mathrm{d}\boldsymbol{g}(t) = (\partial_\xi \boldsymbol{g}(t) + F(t))\mathrm{d}t + \Sigma \mathrm{d}Z(t), \tag{3.40}$$

where ∂_ξ denotes the $n \times n$ operator matrix with the partial derivatives along diagonal. Let $\boldsymbol{g}(t) = (g_1(t), \dots, g_n(t))^T$, then if $F(t)$ and Σ satisfies the semigroup Lipschitzianity and $\boldsymbol{g}(0) \in \mathsf{H}_w^n$, then there exists unique process $\boldsymbol{g}(t) \in \mathsf{H}_w^n$, being the mild solution to (3.40) given by,

$$\boldsymbol{g}(t) = S(t)\boldsymbol{g}(0) + \int_0^t S(t-u)F(u)\mathrm{d}u + \int_0^t S(t-u)\Sigma \mathrm{d}Z(u). \tag{3.41}$$

The no-arbitrage condition forces $F(t) = \langle \Sigma, Q_0 \int_0^\xi \Sigma(\eta)\mathrm{d}\eta \rangle_{w^n}$, which by w^n we mean the induced n-product Hilbert space on H_w^n, which implies that we can show by identifying the constant element σ as the noise, we obtain the mild solution for the each curve,

$$g_i(t) = S(t)g_i(0) + \sigma^2 \int_0^t S(t-u)\mathrm{d}W_i(u) + \sigma^2 \int_0^t S(t-u)\mathrm{d}W_{i+1}(u) \tag{3.42}$$

for $i = 1, \dots, n$, which can be used to compute e.g. covariances and correlations between different curves, and also pricing of zero-coupon bonds.

3.5 Heath-Jarrow-Morton and Stationarity

A central topic of this thesis is the study of stationarity in yield curves. We, therefore present a brief section regarding stationarity in the HJM model. It is not straightforward to give a canonical definition of stationarity. Still, there is a consensus that we, by strong stationarity for stochastic processes, mean shift or translation invariance of the finite-dimensional distributions. Furthermore, we link the property of stationarity to curves in the HJM framework, in terms of invariant measures, as presented in Tehranchi [47]. For this section we consider the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, unless otherwise mentioned.

Let X be a an H-valued random variable whose distribution is defined by $P_X(A) = \mathbb{P}(X \in A)$ for $A \in \mathcal{A}$. We give the standard definition of strong stationarity in terms of distributions.

Definition 3.5.1. An H-valued stochastic process $\{X(t)\}_{t\in\mathbb{Z}}$, is said to be *stationary* if its finite dimensional distributions are shift invariant, i.e.,

$$P_{X(t+\tau)}(A) = P_{X(t)}(A), \quad \text{for all } A \in \mathcal{A},\ t \in \mathbb{Z}, \tag{3.43}$$

and all τ being in the set of all finite subsets of $\mathbb{Z}$.

Motivated by Benth and Süss [4], we also define stationarity in terms of convergence of measures.

Definition 3.5.2. Let $\{X(t)\}_{t\in\mathbb{Z}}$ be an H-valued stochastic process. We say that $X(t)$ is *stationary* if there exists a probability measure μ on $\mathcal{B}(\mathsf{H})$ such that $P_{X(t)} \to \mu$, when $t \to \infty$. The convergence is in terms of measures, i.e.,

$$\int_{\mathsf{H}} f(x) P_{X(t)}(\mathrm{d}x) \to \int_{\mathsf{H}} f(x)\mu(\mathrm{d}x), \quad \text{as } t \to \infty \tag{3.44}$$

where $f : \mathsf{H} \to \mathbb{R}$ is any bounded measurable function.

In the HJM model, we only know of the dynamics of the process, so it might be challenging to study stationarity in terms of the solution. It is also possible for the HJM to not have any solutions, for bad choices of parameters. Another approach to stationarity is that of studying invariant measures. The main attribute of an invariant measure, is related to the law of the forward curves, or any stochastic process for that matter. Let $f(0,\cdot)$ be an H_w-valued $\mathcal{F}_0$-measurable random variable with law μ. If the law of $f(t,\cdot)$ is unchanged for all $t \geq 0$, we say that μ is an *invariant measure.*

In Shalizi and Kontorovich [43], they provide a unification of stationarity and shift-invariance, where they give a proof of the following equivalence

$$\text{stationarity in terms of } Definition\ 3.5.1 \Leftrightarrow \text{shift invariance}, \tag{3.45}$$

which states that stationarity shift invariance are equivalent statements.

Recall the HJM model in H_w from the last section. We specify an HJM model as in (3.23), where we omit the spatial variable,

$$f(t) = S(t)f(0) + \int_0^t S(t-s)\alpha(s)\mathrm{d}s + \int_0^t S(t-s)b(s)\mathrm{d}W(s), \qquad (3.46)$$

for $f(0) \in \mathsf{H}_w$ being a $\mathcal{F}_0$-measurable random variable. Now for *time-homogeneous* HJM models on H_w, Tehranchi [47] showed that there exists an infinite family of invariant measures on H_w with the following properties

(i) Given a specified HJM model which ensures the existence of a continuous solutions in H_w. For every curve $f(0,\cdot)$ with marginal distribution of the initial long rate $f(0,\infty)$, given by ν, we there exists an unique measure μ^ν such that the law of $f(t,\cdot)$ converges to μ^ν.

(ii) For every bounded $\psi : \mathsf{H}_w \to \mathbb{R}$ such that $|\psi(f) - \psi(g)| \le \|f-g\|_w$, we have that

$$|E[\psi(f)] - \int_{\mathsf{H}_w} \psi(f)\mu^\nu(\mathrm{d}f)| \le (1 + E[\|f(0,\cdot)\|_{\mathsf{H}_w}])e^{-\beta t/2}, \qquad (3.47)$$

for some finite β which acts as an upper bound in the technical specification of the model which we have omitted.

3.6 Closing remarks on HJM and cointegration

In the first part of this thesis, we have introduced the basics of functional analysis and probability in Hilbert spaces. We put forth the necessary notions for understanding Hilbert-valued stochastic differential equations, and applied the theory on forward curves. We studied the forward curves as H_w-valued mild solutions of the linear affine equations introduced in section 2.6, and established a so-called no-arbitrage condition for the Musiela-parametrized forward curves in the Heath-Jarrow-Morton framework. In conclusion, we have in this chapter, laid the groundworks for further study in terms of cointegration. In that manner, we propose several paths from which one can extend the study of forward curves.

- The study carried out in this section could be extended to Lévy noise, or cylindrically defined Wiener noise, although the no-arbitrage condition becomes cumbersome.
- One could compute the covariance and correlation between the different forward curves $g_i(t), g_j(t)$ for i and j, and investigate the how the variation depends on different time to maturities.
- Pricing options and zero-coupon bonds.
- Define an explicit function $w(t)$ for the H_w-valued forward curves.
- There is plenitude of different other model-choices one can explore for the forward curves. One could for instance propose a multi-market model consisting of continuous-time autoregressive moving-average (CARMA) process as defined in Benth and Süss [3], or as a functional autoregressive process (FAR(1)) [3].

[3] We will introduce such a process in section 4.4

4 Introduction to Functional Data Analysis

In recent years, technological advancements have led to an enormous increase in the amount of data being collected, whereas the complexity of modern hardware has provided us with advanced computational power. This evolution has given rise to increased attention to new methods of data analysis, including that of considering statistical observations as continuous objects, which eventually came to be the subfield of statistics called *Functional data analysis*. Typical objects of study are weather data, growth functions, MRI images, and financial data.

Dealing with continuous data, we are in a situation where the number of measurements or data-points m, can be vastly greater than the number of observations N. The large number of data-points poses a challenge in studying the data, however, if we approximate each observation by using a building block of smooth functions described by $p < m$ coefficients, we can obtain a substantial decrease in the complexity of the mentioned task.

The general idea is that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and fix a function space F. We consider observations as realizations of a random variable X from the probability space taking values in F, that is

$$X_1, X_2, \ldots, X_N \in \mathsf{F}. \tag{4.1}$$

For our practical purpose, and in the literature for the most part, F is chosen to be a separable Hilbert space. In particular, it is common to choose $\mathsf{F} = L^2([a,b])$, more specifically, $\mathsf{F} = L^2([0,1])$, which is the function space of choice in the seminal monograph of Ramsay [39], but also in Hórvath and Kokoszka [22] and Hsing and Eubank [25]. When we write F in the remaining chapters, we mean a separable Hilbert space.

We begin by presenting the challenge of transforming discrete data to continuous and some canonical results from finite-dimensional statistics, which in some sense carries over to the infinite-dimensional framework. We also present some preliminary results concerning the Filipovic space, H_w, which we studied earlier on. After that, we give an introduction to inference in function spaces and conclude with some theory on functional time series. For the functional time series section, we emphasize the concept of m-dependency in the functional sense, introduced by Hörmann and Kokoszka [21].

4.1 Smoothing

In real life, we do not encounter intrinsically continuous observations. To that end, we must carefully smooth the discrete set of points using some family of functions. By smoothing, we mean fitting the data either by a function which passes through all the data-points (interpolation), or regression-type of methods which purpose is to minimize the error between the data-points and family of functions. In the best case, we have a plenitude of discrete observations, and thus there exists an underlying continuity. But more than often, the collected data can be sparse, complicating the task of smoothing.

Smoothing can be done either by a parametric- or non-parametric method. Typical parametric methods include *basis function* systems, which aims at smoothing the observations $y_i(t)$ by a linear combination of K basis functions,

$$y(t) = \sum_{j=1}^{K} \alpha_j \phi_j(t), \tag{4.2}$$

where $\alpha_j \in \mathbb{R}$ and $\{\phi_j(t)\}_{j=1,\dots,K}$ is a family of basis functions. Classical examples of basis functions are, the Fourier basis $\{\cos(j\omega t), \sin(j\omega t)\}_{j=1,\dots,K}$ for periodic data, and polynomial basis $\{t^{j-1}\}_{j=1,\dots,K}$ for possibly erratic data.

A subset of the polynomial smoothing methods is the so-called *splines*. In short, splines are piecewise polynomials defined on some set $[a, b]$ taking values in $\mathbb{R}$. We will only encounter B-splines in this thesis. Thus we provide a definition, following Joy [27].

Definition 4.1.1. A *B-spline*, is a polynomial spline function defined by

$$y(t) = \sum_{i=0}^{n} \alpha_j N_{i,j}(t), \tag{4.3}$$

where $\{\alpha_i\}_{i=0}^{n}$ is sequence of control points, k the order of the B-spline, and $N_{i,j}$ are the basis functions defined in terms of a nondecreasing sequence t_i for $0, \dots, n+k$ of *knots*, such that

$$N_{i,1}(t) = \begin{cases} 1 & \text{if } t \in [t_i, t_{i+1}] \\ 0 & \text{otherwise,} \end{cases}$$

and

$$N_{i,j}(t) = \frac{t - t_i}{t_{i+j} - t_i} N_{i,j-1}(t) + \frac{t_{i+j+1} - t}{t_{i+j+1} - t_{i+1}} N_{i+1,j-1}(t).$$

For an in-depth study of smoothing in terms of functional data, see for instance Ramsay [39].

4.2 Estimation

Statistical analysis in the scalar sense relies on canonical asymptotic results such as the *law of large numbers* (LLN) and the *central limit theorem* (CLT). An important question to ask, is then, does there exist analog results for the LLN and CLT in the case of F-valued random variables? We will briefly present that, in fact, yes, there does exist F-valued versions of the major theorems in statistics. In addition, we present what is meant by *mean* and *covariance* for F-valued random variables, followed by their sample counterparts.

From introductory statistics we know that the central limit theorem connects the asymptotic average of any sequence of iid random variables to the normal distribution. The below theorem is the equivalent of the CLT for random variables taking values in a separable Hilbert space. The proof can be found in Bosq [11].

Theorem 4.2.1. *Let $\{X_i\}_{i\geq 1}$ be a sequence of iid mean zero random variables taking values in some separable Hilbert space. If $E\left[\|X_1\|^2\right] < \infty$, then*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i \to Z,$$

with Z being a Gaussian with covariance operator $C(x) = E\left[\langle X_1, x\rangle X_1\right]$.

The LLN states that the average of an iid sequence of random variables, converges in probability to the mean function $\mu(t)$. The proof of which can also be found in Bosq [11].

Theorem 4.2.2. *Let $\{X_i\}_{i\geq 0}$ be sequence of iid random variables taking values in a separable Hilbert space. If $E\left[\|X_1\|^2\right] < \infty$, then the mean $\mu = E\left[X_1\right]$ is uniquely defined by $\langle \mu, x\rangle = E\left[\langle X, x\rangle\right]$ and*

$$\frac{1}{N}\sum_{i=1}^{N} X_i \to \mu$$

Now that we have established the motivation for doing inference in separable Hilbert spaces, we define the canonical parameters and their estimators.

Definition 4.2.3. Let $X(t)$ be an F-valued random variable. We define *mean*, *covariance function* and *covariance operator* as follows,

$$\begin{aligned}
\mu(t) &= E\left[X(t)\right] \\
c(t,s) &= E\left[(X(t)-\mu(t))(X(s)-\mu(s))\right] \\
C(\cdot) &= E\left[\langle X-\mu, \cdot\rangle (X-\mu)\right].
\end{aligned}$$

Moreover, we consider the observations $X_1(t), \ldots, X_N(t)$ as realization of an F-valued random variable $X(t)$. The naive estimators of the above parameters are then given by,

$$\hat{\mu}(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t)$$
$$\hat{c}(t,s) = \frac{1}{N}\sum_{i=1}^{N}(X_i(t)-\hat{\mu}(t))(X_i(s)-\hat{\mu}(s))$$
$$\hat{C}(\cdot) = \frac{1}{N}\sum_{i=1}^{N}\langle X_i-\hat{\mu},\cdot\rangle(X_i-\hat{\mu})$$

There is a striking resemblance between ordinary statistics and the functional versions, the only apparent difference being the temporal dependence in t. Note that, however, the sample covariance operator $\hat{C}(\cdot)$ maps a possibly infinite-dimensional element to a finite-dimensional subspace spanned by $X_1,\ldots,X_N$. This illustrates the limitations of estimation, when dealing with function-valued objects.

A useful property to establish is the *consistency* of the estimators, that being, we want the estimators to converge to the parameters in some sense. Before showing that $\mu(t)$ is consistent in mean square, we provide an auxiliary result,

Lemma 4.2.4. *If $X_1, X_2 \in L^2([a,b])$ are independent and $E[X_1]=0$, then*

$$E[\langle X_1, X_2\rangle] = 0.$$

Proof. This can be computed directly by interchanging integral and expectation,

$$E[\langle X_1, X_2\rangle] = E\left[\int_a^b X_1(t)X_2(t)\mathrm{d}t\right] = \int_a^b E[X_1(t)]\,E[X_2(t)]\,\mathrm{d}t,$$

which by independence shows the assertion. ■

The below theorem which states that $\hat{\mu}(t)$ is a consistent mean square estimator, can be found in Horváth and Kokoszka [22]. We provide a proof where we fill some of the computational steps.

Theorem 4.2.5. *Suppose $\{X_i\}_{i\geq 1}$ is a sequence of iid random variables in $L^2([a,b])$. If $E[\|X\|^2] < \infty$, then $\hat{\mu}$ is an unbiased consistent estimator of μ in L^2 norm, i.e $E[\hat{\mu}] = \mu$.*

Proof. Let $\{X_i\}_{i\geq 1}$ be sequence of realized random variables of X.First we have that $E[X_i] = \mu(t)$ for $i = 1,\ldots$ and almost all $t \in [0,1]$. Hence

$$E[\hat{\mu}] = E\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] = \frac{1}{N}\sum_{i=1}^{N} E[X_i] = \mu.$$

Furthermore

$$E\left[\|\hat{\mu}-\mu\|^2\right] = E\left[\langle \frac{1}{N}\sum_{i=1}^{N} X_i - \mu, \frac{1}{N}\sum_{i=1}^{N} X_j - \mu\rangle\right]$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\langle X_i-\mu, X_j-\mu\rangle\right]$$
$$= \frac{1}{N^2}\sum_{i=1}^{N} E\left[\|X_i-\mu\|^2\right],$$

using Lemma 4.2.4. Since the X_i's are independent realizations of X, we get that

$$\frac{1}{N^2}\sum_{i=1}^{N} E\left[\|X_i-\mu\|^2\right] = \frac{1}{N}E\left[\|X-\mu\|^2\right] \to 0$$

■

Since there has been done extensive research on the statistical properties of the $L^2([a,b])$ space, we put forth an example of what might happen in a general function space.

Example 4.2.6. Let us consider observations in the Filipovic space H_w, which we studied in the last chapter. Recall that

$$\|f\|_w^2 = f(0)^2 + \int_0^\infty w(x)f'(x)^2 \mathrm{d}x. \tag{4.4}$$

We assume that $X_1, X_2, \dots X_N$ are iid realization of a random variable X of finite expectation taking values in H_w with $E\left[\|X\|_w\right] < \infty$. Again, we consider the canonical estimator of the mean function,

$$\hat{\mu}(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t), \tag{4.5}$$

where $t \in \mathbb{R}_+$. Notice that the above estimator is, indeed, unbiased for any choice of function space where $E\left[X_i\right] = \mu(t)$ exists for all t. Following the same reasoning from the L^2-proof, we obtain for the H_w observations,

$$E\left[\|\hat{\mu}-\mu\|_w^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\langle X_i-\mu, X_j-\mu\rangle_w\right]. \tag{4.6}$$

For $i \neq j$ we have that

$$E\left[\langle X_i(t), X_j(t)\rangle_w\right] = E\left[X_i(0)X_j(0)\right] + E\left[\int_0^\infty w(x)\partial_x X_i(x)\partial_x X_j(x)\mathrm{d}x\right]$$
$$= E\left[X_i(0)\right]E\left[X_j(0)\right] + \int_0^\infty w(x)E\left[\partial_x X_i(x)\right]E\left[\partial_x X_j(x)\right]\mathrm{d}x$$
$$= \mu(0)^2 + \int_0^\infty w(x)\mu'(x)^2\mathrm{d}x = \|\mu\|_w^2.$$

Above, we interchanged expectation and integral by Fubini-Tonelli in the second term. Secondly, since the generator ∂_x is closed in H_w, we interchange expectation and differentiation. Thus,

$$\begin{aligned} E\left[\|\hat{\mu}-\mu\|_w^2\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\mu\|_w^2 + \frac{1}{N^2}\sum_{i=1}^{N} E\left[\|X_i-\mu\|_w^2\right] \\ &= \|\mu\|_w^2 + \frac{1}{N}E\left[\|X-\mu\|_w^2\right], \end{aligned}$$

which shows that the naive estimator of the mean in H_w, is not consistent in square norm.

The problem above is that $E\left[\langle X_j-\mu, X_i-\mu\rangle\right] \neq 0$ for $i \neq j$. However if we assume $X_1,\ldots,X_N \in \mathsf{F}$, such that for independent X_1 and X_2 with the condition $E[X_1]=0$, we have $E\left[\langle X_1, X_2\rangle_{\mathsf{F}}\right]=0$. If so, many of the properties shown to hold for observations in L^2 as in Horváth and Kokoszka [22], may also hold for observations in any general separable Hilbert space F. In particular, we have for $E\left[\|X\|_{\mathsf{F}}^4\right] < \infty$, that the following bounds for the covariance operator holds

$$\|\hat{C}\|_{\mathrm{HS}}^2 = \sum_{i=1}^{\infty}\lambda_i^2 \leq E\left[\|X\|_{\mathsf{F}}^4\right] \tag{4.7}$$

$$\|\hat{C}-C\|_{\mathrm{HS}}^2 \leq \frac{1}{N}E\left[\|X\|_{\mathsf{F}}^4\right] \tag{4.8}$$

See Chapter 2 in Horváth and Kokoszka [22] for the proofs. The bounds above asserts that the Hilbert–Schmidt norm of the estimator of the covariance is bounded when $E\left[\|X\|_{\mathsf{F}}^4\right]$ is bounded. Moreover, the distance between $\hat{C}$ and C in Hilbert–Schmidt norm gets smaller for large N. An implication of these bounds is that for operators which are close, the eigenvalues are also close [1], which we will use in the next section when discussing principal component analysis.

4.3 Principal Component Analysis

In this section, we attempt to explain some of the basics of *principal component analysis* (PCA) and how one can generalize the matrix framework to that of Hilbert-valued functions. *Functional principal component analysis* is an inferential tool, and depicts one of the central idéas of FDA, namely, dimension reduction. We obtain a finite-dimensional approximation of the infinite-dimensional observations $X_1,\ldots,X_N$, by a so-called *optimal empirical orthonormal basis*. A byproduct of this approach are the generated *principal components*, which represents the functions which are most correlated with the variability of the data.

[1] See Bosq [11] for the asymptotic bounds of the eigenvalues

Basics of PCA

Given a symmetric $p \times p$ matrix C, we know that there exist an orthonormal matrix U consisting of the eigenvectors of C in the sense that

$$U = [u_1 \cdots u_p], \tag{4.9}$$

where the u_i's are column vectors. Therefore, we have $U^T U = I$, and $Cu_j = \lambda_j u_j$, where λ_j is an eigenvector of C. By T, we mean the transpose of a finite-dimensional vector or matrix. Moreover, we have that

$$U^T C U = \Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p), \tag{4.10}$$

where Λ denotes the $p \times p$ matrix consisting only of the diagonal eigenvalue elements with $\lambda_1 > \lambda_2 > \cdots > \lambda_p$.

The representation (4.10), implies that we can write $C = U\Lambda U^T$. We can use this to find the vector x, which maximizes the quantity $x^T C x$. Notice that

$$x^T C x = x^T U \Lambda U^T x = y^T \Lambda y, \tag{4.11}$$

where $y = U^T x$. By the orthonormality of U we can conclude that $\|x\| = \|y\|$, which shows that it is sufficient to find the y which maximizes $y^t \Lambda y$. Moreover,

$$y \Lambda y = \sum_{j=1}^{p} \lambda_j y_j^2, \tag{4.12}$$

thus for $y = [1\ 0\ \cdots\ 0]^T$ and $x = u_1$ we end up with λ_1 as maximum.

The above derivation can be lifted to the function setting. Recall that spectral decomposition of an Hilbert–Schmidt operator (2.6), which is

$$\Psi(h) = \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle v_j, \quad h \in \mathsf{H}.$$

We also know that the covariance operator is a Hilbert–Schmidt operator. Thus we want to maximize $\langle T(h), h \rangle$ subject to $\|x\| = 1$, but we have that

$$\langle \Psi(h), h \rangle = \langle \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle v_j, h \rangle = \sum_{j=1}^{\infty} \langle \lambda_j \langle h, v_j \rangle v_j, h \rangle \tag{4.13}$$

$$= \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle \langle v_j, h \rangle = \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle^2. \tag{4.14}$$

By Parseval, we may write $\|x\| = \sum_{j=1}^{\infty} \langle x, v_j \rangle^2$, which alters the optimization to the following problem,

$$\max \sum_{j=1}^{\infty} \lambda_j \langle h, v_j \rangle^2 \ \text{ subject to } \ \sum_{j=1}^{\infty} \langle x, v_j \rangle^2 = 1. \tag{4.15}$$

Pick $\langle x, v_1\rangle^2 = 1$ and $\langle x, v_j\rangle = 0$ for all $j \neq 1$, thus $\langle \Psi(x), x\rangle$ is maximized at $\pm v_1$ with maximum λ_1.

Furthermore, we maximize $\langle \Psi(x), x\rangle$ subject to $\|x\| = 1$ and $\langle x, v_1\rangle = 0$, which is maximized at $x = v_2$ with maximum λ_2. This procedure may be repeated and as summarized by Theorem 3.2 in Horváth, and Kokoszka [22], we have that for any symmetric, positive definite Hilbert–Schmidt operator satisfying $\lambda_1 > \cdots > \lambda_{p+1}$, that

$$\sup_{\|x\|=1, \langle x, v_j\rangle = 0} \langle \Psi(x), x\rangle = \lambda_i, \quad \text{for } 1 \leq j \leq i-1,\ i < p \tag{4.16}$$

where the supremum is obtained if $x = v_i$. The element v_j is unique up to a sign.

Functional principal components

Consider the observations $y_1, y_2 \ldots, y_N$ on some separable Hilbert space, and define the sum

$$\hat{S}^2_{\mathrm{PCA}} = \sum_{i=1}^{N} \|y_i - \sum_{j=1}^{p} \langle y_j, u_j\rangle u_j\|^2. \tag{4.17}$$

If we can find an orthonormal basis $u_1, \ldots, u_p$ which minimizes the above term, we have a finite dimensional set of vectors which in some sense approximates the set of observations y_i for $i = 1 \ldots, N$.

If we minimize $\hat{S}_{\mathrm{PCA}}$ subject to $\|u_j\| = 1$ we get that,

$$\begin{aligned}
\hat{S}^2_{\mathrm{PCA}} &= \sum_{i=1}^{N} \|y_i - \sum_{j=1}^{p} \langle y_j, u_j\rangle u_j\|^2 \\
&= \sum_{i=1}^{N} \|y_i\|^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{p} \langle y_i, u_j\rangle^2 + \sum_{i=1}^{N}\sum_{j=1}^{p} \langle y_i, u_j\rangle^2 \|u_j\|^2 \\
&= \sum_{i=1}^{N} \|y_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{p} \langle y_i, u_j\rangle^2,
\end{aligned}$$

by the assumption of orthonormality. The minima of $\hat{S}^2_{\mathrm{PCA}}$ is reached by maximizing $\sum_{i=1}^{N}\sum_{j=1}^{p} \langle y_i, u_j\rangle^2$. Recall that,

$$\langle \hat{C}(u), u\rangle = \langle \frac{1}{N}\sum_{i=1}^{N} \langle y_i, u\rangle y_i, u\rangle = \frac{1}{N}\sum_{i=1}^{N} \langle y_i, u\rangle^2. \tag{4.18}$$

Thus minimizing $\hat{S}^2_{\mathrm{PCA}}$ is equivalent to maximizing

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{p} \langle y_i, u_j\rangle^2 &= \sum_{j=1}^{p} \langle \hat{C}(u), u\rangle \\
&= \sum_{j=1}^{p}\sum_{k=1}^{\infty} \hat{\lambda}_k \langle u_k, \hat{v}_k\rangle^2,
\end{aligned} \tag{4.19}$$

which by (4.16) gives that maximum of (4.19) is $\sum_{k=1}^{p} = \lambda_k$, and is obtained when $u_1 = \hat{v}_1, u_2 = \hat{v}_2, \ldots u_p = \hat{v}_p$.

Consequently, we define the *empirical functional principal components* (EFPC's) of given observations $X_1, \ldots X_N$ as the eigenfunctions $\hat{v}_j$ of the sample covariance operator $\hat{C}$. In contrast, the *functional principal components* are defined as the eigenfunctions v_j of the covariance operator C. Let $\hat{v}_i$ for $i = 1, \ldots, N$ be a basis in $\mathbb{R}^N$, so that

$$\frac{1}{N}\langle X_i, x\rangle^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\langle X_i, \hat{v}_j\rangle^2 = \sum_{j=1}^{N}\hat{\lambda}_j, \tag{4.20}$$

hence we say that the variance in the direction of $\hat{v}_j$ is $\hat{\lambda}$, in addition to the amount of variance explained by $\hat{v}_j$ is the fraction $\hat{\lambda}_j / \sum_{k=1}^{N} \hat{\lambda}_k$.

4.4 Functional Time Series

In this section, we give a short introduction to time series. Following the lines of Shumway and Stoffer [45], we define expectation and autocovariance, before we introduce the autoregressive models for time-series. Finally we give an informal introduction to functional time-series, in particular the autoregressive functional model of order one. We also provide a criteria for when a functional autoregressive model admits a causal stationary solution.

Introduction to Time Series

First we consider some classical time-series theory. We let $\{x_t\}_{t\in\mathbb{Z}}$ be a $\mathbb{R}$-valued stochastic process. Such an object is called a *time series* due to the temporal dependence in t. If $f_t(x)$ denotes the density of x_t we define the *mean function* of x_t to be

$$\mu_t = E\left[x_t\right] = \int_{-\infty}^{\infty} x f_t(x)\mathrm{d}x.$$

A measure of dependency is the covariance between times of itself, which in the realm of time series is called *autocovariance*. The definition is analogue to the covariance parameter from before,

$$\mathrm{cov}(x_s, x_t) = E\left[(x_s - \mu_s)(x_t - \mu_t)\right]. \tag{4.21}$$

Under the assumption of weakly stationarity, time series are often modeled by *autoregressive-moving-average* (ARMA) models. The ARMA models are a combination of autoregressive and moving average type of models. Autoregressive models depends linearly on its past including an error term, while moving average models depend linearly on the past of a white noise process. Thus we denote *white noise* by w_t, and let c be a constant, and ϕ_i, θ_i are real-valued parameters. Then, we say that for finite $p, q \in \mathbb{N}$,

$$x_t = c + w_t + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_i w_{t-i}, \tag{4.22}$$

is a model of order p,q, denoted ARMA(p, q). By white noise, we mean an iid process with zero mean and variance σ^2. A typical choice of a distribution for the white noise is $w_t \sim \mathcal{N}(0, \sigma^2)$.

However, we look into the autoregressive models, in particular, which is when $q = 0$. In the case of an AR(1) time series, we have

$$x_t = \phi x_{t-1} + w_t = \cdots = \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}, \tag{4.23}$$

iterating back in time, assuming finite variance we get

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}, \tag{4.24}$$

If $|\phi| < 1$, the representation (4.24) exists, and is called the *causal stationary solution* of the model.

Moreover, we may identify the order of an AR(p) or MA(p) model, through its autocovariance function (ACF), or *partial autocovariance function* (PACF). The order of the MA model is determined where the ACF displays a sharp cut-off, whereas the AR model is determined where the PACF cuts-off sharply [2]. By cutting-off, we mean an abrupt decrease such that the value is below some threshold.

We conclude this section by introducing the Functional Autoregressive Model of order one FAR(1). Let $\Phi \in B(\mathsf{H}, \mathsf{H})$. The functional analogue of the autoregressive process FAR(1), is defined by

$$X_n = \Phi(X_{n-1}) + \varepsilon_n, \tag{4.25}$$

where $\{X_n\}_{n\in\mathbb{Z}}$ is a mean zero sequence of elements in $L^2([a, b])$. The error process $\{\varepsilon_n\}_{n\in\mathbb{Z}}$ is assumed to be an iid mean zero sequence taking values in $L^2([a, b])$ with $E\left[\|\varepsilon_n\|^2\right] < \infty$.

Horváth and Kokoszka[22] shows that for the FAR(1) model, we can obtain a stationary causal solution similar to (4.25), given by

$$X_n = \sum_{j=0}^{\infty} \Phi^j(\varepsilon_{n-j}), \tag{4.26}$$

[2]See Shumway and Stoffer [45] for details.

if there exists $j_0 \in \mathbb{Z}$, such that

$$\|\Phi^{j_0}\| < 1. \tag{4.27}$$

The series converges a.s. in L^2-norm as well.

In Horváth and Kokoszka[22], they show how one can estimate Φ, as well as performing methods of forecasting, using so-called *predictive-factors*. The FAR(1) model is also used for *change point detection*, which aims at finding abrupt changes in the observations. By means of change point detection using FAR(1) models, one tests for changes in the linear operator through the recursive scheme

$$X_{n+1} = \Psi_n(X_n) + \varepsilon_{n+1}, \tag{4.28}$$

for finite n, and observations $X_n \in L^2([a,b])$ with zero mean $\varepsilon_{n-1} \in L^2([a,b])$.

4.5 Functional Dependency

In extending the notion of dependence to that of functional time-series, we first define what is known as m-dependent time series. Moreover we define $L^p - m$ *approximable functions*, introduced by Hörmann and Kokoszka [21]. The section is concluded an example of approximability of the functional autoregressive model.

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a sequence taking values in a function space F, denote by $\mathcal{F}_h^- = \sigma(\dots, X_{h-2}, X_{h-1}, X_h)$ and $\mathcal{F}_h^+ = \sigma(X_h, X_{h+1}, X_{h+2}, \dots)$ the σ-algebras generated by the observations of $\{X_n\}_{n\in\mathbb{Z}}$, which is said to be m-*dependent* if for any h, $\mathcal{F}_h^-$ and $\mathcal{F}_{h+m}^+$ are independent. While few sequences in practice exhibit this property, methods of approximating m-dependent series, dates all the way back to Billingsley [7]. Even so, we will follow the approach by Hórvath and Kokoszka [22], which is based on the formulation in Hörmann and Kokoszka [21].

Definition 4.5.1. Let $L_2^p([a,b])$, be the space of all $L^2([a,b])$-valued functions, which for all X satisfy,

$$v_p(X) = E\left(\left[\int_a^b X^2(t)\mathrm{d}t\right]^{2/p}\right)^{1/p} < \infty \tag{4.29}$$

For the sake of defining $L^p - m-$approximability, let $S = \{1, \dots, N\}$ for $N \geq 2$, and denote by S^∞ the infinite Cartesian product, consisting of all infinite sequences on the form,

$$\omega = (z_1(\omega), z_2(\omega), \dots), \tag{4.30}$$

so that $z_j(\omega) \in S$ for all $\omega \in S^\infty$, and $j \geq 1$. In fact, it is not trivial to construct a probability space on S^∞, therefore, we briefly discuss the procedure as given in Billingsley [8]. Let $S^n = S \times \cdots \times S$ be the n-dimensional Cartesian product of S, consisting of sequences on the form $(y_1, \ldots, y_n)$. The set

$$\{\omega : (z_1(\omega), z_2(\omega), \ldots, z_n(\omega) = (y_1, \ldots, y_n)\},$$

depicts the event that the first n repetitions of S gives the outcome sequence $(y_1, \ldots, y_n)$. Consider the *cylinder* set,

$$E = \{\omega : (z_1(\omega), z_2(\omega), \ldots, z_n(\omega)) \in A\}, \tag{4.31}$$

where $A \subset S^n$. Let $\mathcal{E}_0$, be the class of all such finite sets, and define the probability measure,

$$\mathbb{P}_\omega(E) = \sum_A p_{y_1} p_{y_2} \cdots p_{y_n}, \tag{4.32}$$

which consists of the product of all possible outcomes $(y_1, \ldots, y_n)$ of A. In Billingsley [8], it is shown that $\mathcal{E}_0$ is a σ-algebra on S^n, which by an extension result can generate a σ-algebra, $\mathcal{E}$ on S^∞. A similar extension can also be performed for $\mathbb{P}_\omega$, such that $(S^\infty, \mathcal{E}, \mathbb{P}_\omega)$ defines a probability space. Moreover, we can then define an $L^p - m-$*approximable* sequence.

Definition 4.5.2. Let S denote a measurable space. A process $\{X_n\} \in L_2^p$ is L^p-*m*-*approximable* if each X_n can be represented as

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots),$$

where $f : S^\infty \to \mathsf{F}$.

Let now $\{\varepsilon_i^{'}\}_{i\in\mathbb{Z}}$ be an independent copy of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$, and define

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{'}, \varepsilon_{n-m-1}^{'}, \ldots), \tag{4.33}$$

which implies

$$\sum_{m=1}^{\infty} v_p(X_n - X_n^{(m)}) < \infty. \tag{4.34}$$

The $X_N^{(m)}$'s from (4.33) is not m-dependent, to this end Hörmann and Kokoszka [21] instead use a so-called *coupling construction*. For each n, define an independent copy $\{\varepsilon_k^{(n)}\}$ of $\{\varepsilon_k\}$, which we instead use in the construction of (4.33), thus

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{m}, \varepsilon_{n-m-1}^{m}, \ldots). \tag{4.35}$$

Then for each $m \geq 1$, the sequence $\{X_n^{(m)}\}_{n\in\mathbb{Z}}$ satisfy (4.5.3), are strictly stationary and m-dependent.

Consider a FAR(1) model, defined as in the last section,

$$X_n = \Phi(X_{n-1}) + \varepsilon_n(t).$$

Suppose that the model admits a stationary causal solution $X_n(t) = \sum_{j=1}^{\infty} \Phi^j(\varepsilon_{n-1})$, and define the sequence

$$X_n^{(m)} = \sum_{j=0}^{m} \Phi^j(\varepsilon_{n-1}) + \sum_{j=m+1}^{\infty} \Phi^j(\varepsilon_{n-1}).$$

We want to show that $\sum_{m=1}^{\infty} v_p(X_m - X_m^{(m)}) < \infty$, and in doing so we need an auxilary result on the $v_p(\cdot)$ function.

Lemma 4.5.3. *For $\Phi \in B(\mathsf{H})$ and $Y \in L^2([a,b])$, we have the bound*

$$v_p(\Phi(Y)) \leq \|\Phi\| v_p(Y). \tag{4.36}$$

Proof. Recall that for bounded Φ, we have $\|\Phi(Y)\| \leq \|\Phi\|\|Y\|$. Thus,

$$v_p(\Phi(Y)) = E\left[\|\Phi(Y)\|^p\right]^{1/p} \tag{4.37}$$
$$\leq E\left[\|\Phi\|^p \|Y\|^p\right]^{1/p} = \|\Phi\| v_p(Y). \tag{4.38}$$

■

Now, using Lemma 4.5.3, Horváth and Kokoszka [22] computes the following bound,

$$v_p\left(X_m - X_m^{(m)}\right) \leq \sum_{j=m}^{\infty} \|\Psi\|^j v_p(\varepsilon_0), \tag{4.39}$$

which means that $\sum_{m=1}^{\infty} v_2(X_m - X_m^{(m)}) < \infty$, since $v_2(\varepsilon_0) < \infty$ by definition.

We define by a *Linear process*, any sequence $\{X_n\}_{n\geq 1} \in L^2([0,1])$ such that

$$X_n = \sum_{i=1}^{\infty} \Psi_i(\varepsilon_{n-i}), \tag{4.40}$$

with iid zero mean errors $\varepsilon_n \in L^2([0,1])$.

Proposition 4.5.4. *[22, Proposition 16.1] Let $\{X_i\}_{i\geq 1} \in L^2([0,1])$ with $v_p(\varepsilon_0) < \infty$ for $p \geq 2$ be a FAR(1) process with operator Ψ. If the operator satisfies*

$$\sum_{m=1}^{\infty} \sum_{j=m}^{\infty} \|\Psi_j\| < \infty, \tag{4.41}$$

then the sequence $\{X_i\}_{i\geq 1}$, is L^p-m-approximable.

The property of L^p-m-approximability relates to the *long-run-covariance* of time series. Recall that for a weakly stationary scalar time series $\{x_t\}_{t\in\mathbb{Z}}$, the long-run-covariance is defined as

$$\sigma^2 = \sum_{i=-\infty}^{\infty} \operatorname{Cov}(x_0, x_j) = \sum_{j=-\infty}^{\infty} \gamma_j \tag{4.42}$$

The long-run-covariance is related to the following central-limit theorem, which proof can be found in Hamilton [19] .

Theorem 4.5.5. *Let* $\{x_t\}_{t\in\mathbb{Z}}$ *be a time series given by,*

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_j, \tag{4.43}$$

where the errors are iid random variables with $E\left[\varepsilon_j^2\right] < \infty$ *for all* $j \in \mathbb{Z}$*. If* $\sum_{j=0}^{\infty}|\psi| < \infty$*, then*

$$\sqrt{T}(\mu_T - \mu) \underset{d}{\rightarrow} \mathcal{N}(0, \sum_{j=-\infty}^{\infty} \gamma_j). \tag{4.44}$$

We conclude this chapter with a result regarding the existence of a central limit theorem for L^p-m-approximable sequences.

Theorem 4.5.6. *[22, Theorem 16.3] Let* $\{X_n\}_{n\in\mathbb{Z}}$ *be a mean zero* L^p*-*m*-approximable sequence. Then*

$$\frac{1}{N}\sum_{i=1}^{N} X_i \underset{d}{\rightarrow} Y, \tag{4.45}$$

where Y *is a gaussian process with*

$$E\left[Y(t)\right] = 0,$$
$$E\left[Y(t)Y(s)\right] = E\left[X_0(t)X_0(s)\right] + \sum_{i=1}^{\infty} E\left[X_0(t)X_i(s)\right] + \sum_{i=1}^{\infty} E\left[X_0(s)X_i(t)\right].$$

5 Empirical study of the Norwegian Yield Curve

5.1 Explorative Data Analysis

We present the data which is to be studied under the assumption of stationarity. We want to test whether the Norwegian yield curve is stationary. However, the choice of such underlying is not uniquely defined. The European Insurance and Occupational Pensions Authority produce the risk-free interest rate, which is used for modeling purposes under Solvency 2. On the other hand, it is useful for financial institutions to measure the risk of the interest rate curve. To that end, we choose to construct the yield curves independently, by smoothing government bond observations using the Nelson-Siegel approach. Moreover, the bonds issued by the Norwegian central bank is limited to having a time to maturity of 10 years as their most lengthy product. Which is far from capturing the asymptotic behavior of the yield curve. Therefore we will use the smoothing method of Nelson-Siegel also to extrapolate the interest rate curve to reach 60 years to maturity.

Data description

The data from Norges Bank [36] comprises 4082 trading days, from 08-01-2003 to 07-05-2019. The central bank of Norway, Norges Bank issues daily bonds as securities with maturity of $3, 6, 9, 12$ months and $3, 5, 10$ years. The bond prices in NOK are reported as annual effective yields, given in percent.

Analysis of the Bond data

We first perform some naive exploratory data analysis of the bond observations. In Figure 5.1, we plot the 3-month bond price together with ACF and PACF-plots. Recall that the ACF is used to determine if the time series is admits a moving average property, while the PACF-plot describes the autoregressive order of the time series. We see from Figure 5.1, the volatility of the 3-month bond price around the time of the financial crisis differs from the rest of the plot. The ACF plot is not interesting, meaning the 3-month bond price time

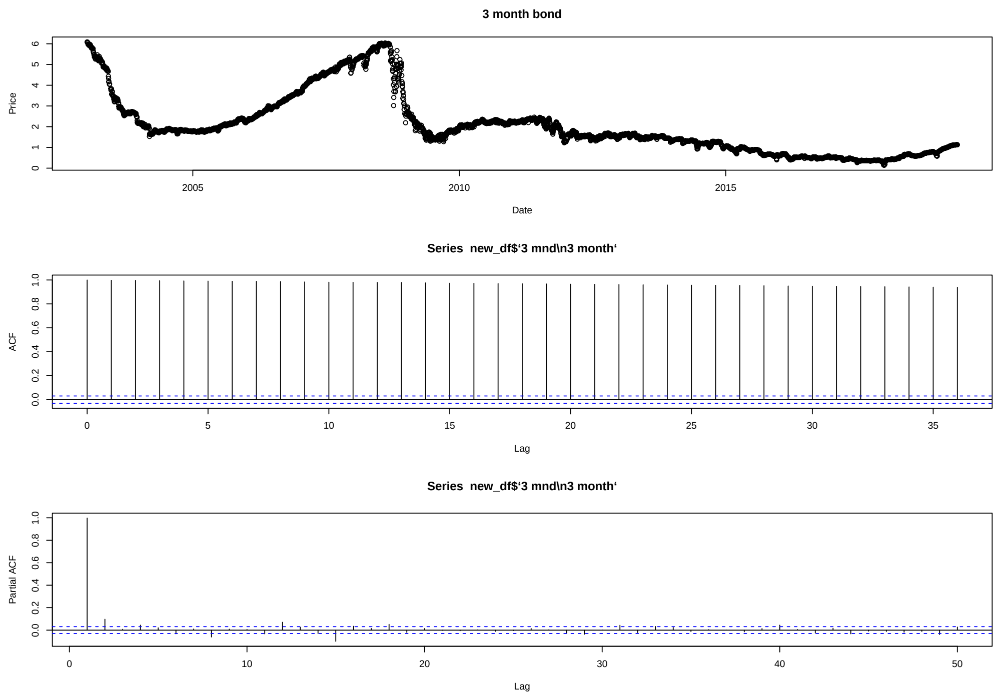


Figure 5.1: Historical 3-Month bond price, together with its ACF and PACF

series is not of moving average type. However, the PACF-plot abruptly cuts off at lag 2. A lag 2-cut-off may indicate that the time series may be modeled as an autoregressive time series of order 2. We should also pay attention to the sporadically significant lags at other times within the lag 50 window.

In Figure 5.2, we see the plots of the 6-month bond time series. The plot resembles the 3-month bond time series to a great extent but differs slightly around the time of the financial crisis by being marginally less volatile. This is expected since the 6-month bond has a longer time to maturity. Again, we also notice that there is no cut-off in the ACF-plot. The PACF-plot, however, cuts off at lag 2, which indicates an autoregressive property of order 2 in the case of a 6 -month-bond price time series. We omit the remaining ACF-plots of the of the bond observations, as they show the same sluggish-decaying behavior as the 3 and 6-month bonds..

The 9-month and 12-month bond observations seen in Figure 5.3 depicts much of the same behavior as the other plots, but with significant values up until lag-3. Notice that there are occurring some sporadically significant values. For the 3-year bond shown in Figure 5.4, we notice a considerable change in the dynamics of the time series. The overall shape is preserved, but the volatility is significantly reduced, showing a more smooth plot in contrast to the previous ones. The PACF-plot of the 3-year bond time series shows a cut-off at lag 2, but the second lag being more subtle than the others.

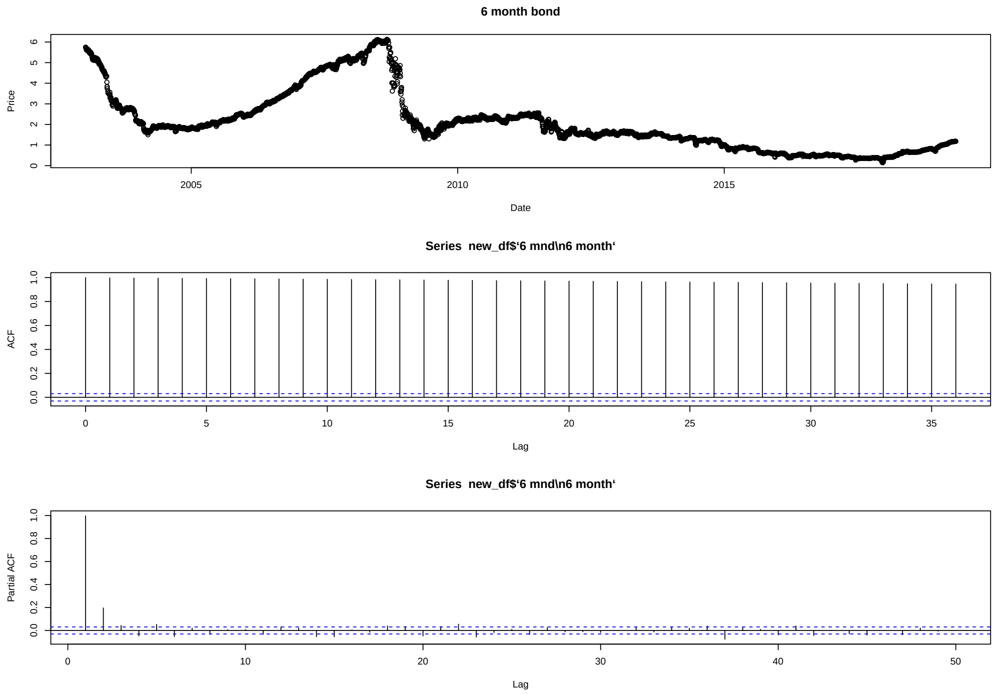


Figure 5.2: Historical 6-Month bond price, together with its ACF and PACF

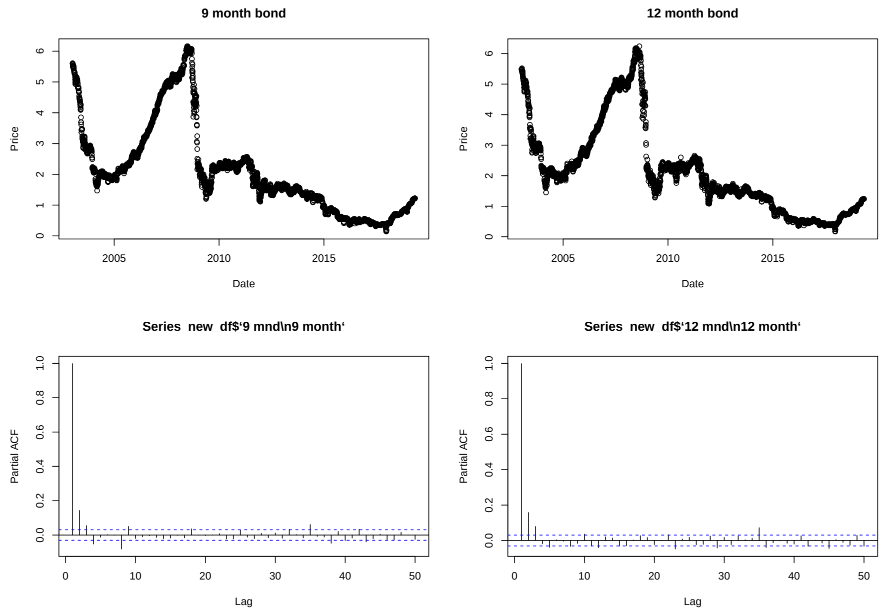


Figure 5.3: 9-Month and 12-month bonds

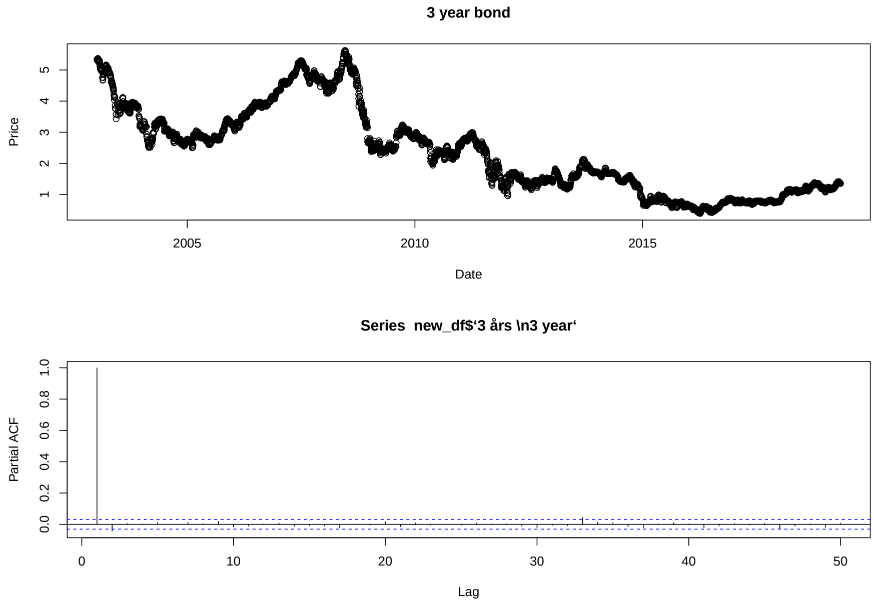


Figure 5.4: 3-year bond

Finally, we have from Figure 5.5 the plot of both the 5 year and 10 year-bond price time series. The time series plot of 3 year and 5 year bonds looks similar, while the 10 year bonds, to some extent, seem less affected by the financial crisis locally. This is not a surprise for an underlying with such a long time to maturity. The PACF-plot of both the 5 year and 10 year bonds cuts of at lag 2. Consequently, we provide in Table 1 a summary of the estimated coefficients for order two autoregressive models for each time to maturity. Recall that such a model can be represented as the time series

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2}. \tag{5.1}$$

From table 1, we have listed the estimated coefficients for an AR(2) model, as a compromise between order one, two or three. The σ^2 term denotes the variance of the error term. We notice that the bonds with a short time to maturity have similar coefficient values, so do the products with a longer time to maturity. This poses a challenge in modelling the entire yield curve as one autoregressive model due to the temporal dependence.

5.2 Nelson-Siegel fitting the of the curves

In the paper of Nelson and Siegel [34], they present a parametrically parsimonious model of the yield curve. Their motivation was to describe the three shapes generally found in yield curves in practice: the level, hump, and

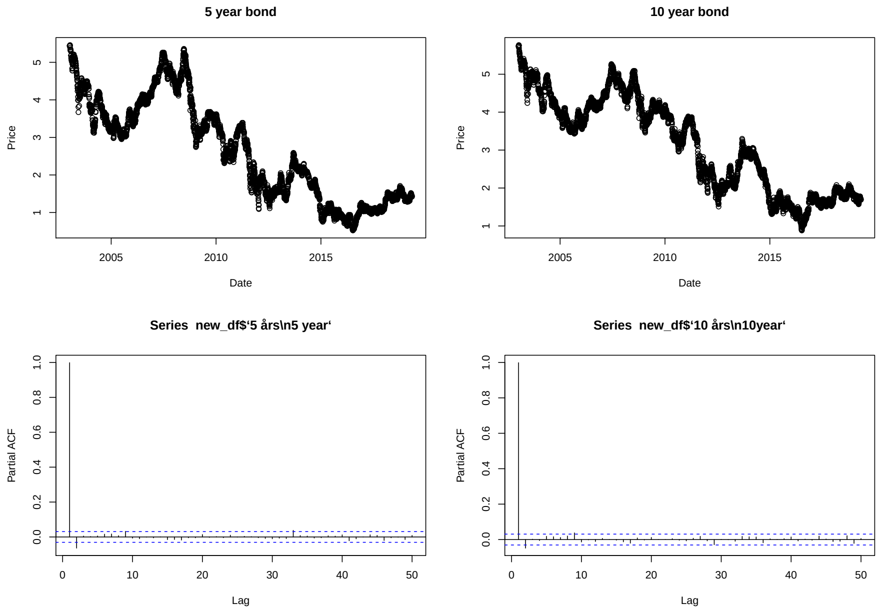


Figure 5.5: 5 -year and 10-year bond

Estimated Autoregressive Models			
Bond	Coefficient 1	Coefficient 2	σ^2
3 month	0.901	0.097	0.008984
6 month	0.8020	0.1963	0.00844
9 month	0.8556	0.1430	0.006643
12 month	0.8406	0.1579	0.00749
3 year	1.0426	-0.0438	0.004359
5 year	1.0628	-0.0641	0.004316
10 year	1.0471	-0.0485	0.004073

Table 5.1: Estimated AR(2) model for the bond observations.

slope. We think of the level being the asymptote and the hump describing the curvature - which is either convex or concave. Lastly, we have the slope, which describes how fast the function decays. The Nelson-Siegel functions, given in terms of time to maturity T, are

$$\begin{aligned} f_{NS}(T) = \beta_0 + \beta_1 \frac{1 - \exp(-\lambda T)}{\lambda T} \\ + \beta_2 (\frac{1 - \exp(-\lambda T)}{\lambda T} - \exp(-\lambda T)). \end{aligned} \tag{5.2}$$

The asymptotic level of the yield curves are given by the constant term β_0, since

$$\lim_{T \to \infty} = \beta_0,$$

hence β_0 is called a *long-term factor*. Notice that the loading term on β_1, is 1 at $T = 0$, as

$$\lim_{T \to 0} = \frac{1 - \exp(-\lambda T)}{\lambda T} = 1.$$

Thereafter, it decays quickly to 0, which motivates the fact that β_1 is considered a *short-term factor*, which represent the slope. The loading on coefficient β_2, can be shown to only affect the middle part of the functions, which coined the name *medium-term* factor, or curvature. For a detailed discussion of the Nelson-Siegel family of functions see Diebold and Rudebusch [16].

We now give some examples of fitted curves. See from Figure 5.6 a plot of the bond observations together with the Nelson-Siegel curve. Notice that the location of the bonds with a short time to maturity produced a hump early on in the yield curve. In Figure 5.7, there is no such hump, and the growth of the function slowly decays, which may indicate a healthy development of economy. In contrast, we see that Figure 5.8 depicts the yield curve under the financial crisis, which indicates an insecure money market a long time ahead. Also notice that the model has some trouble with the fit, in comparison to the prior fits.

It is known that the Nelson-Siegel model may have some trouble producing highly accurate yield curves for certain types of observations. We will not dive into the realm of error study, but instead, assume that the curves represent meaningful yield curves. Our goal is merely to study the property of stationarity, given the structure of the curves. If we were to predict future yield curves, we would have had to take a closer look at the errors of fitting and how they interfere with the predicted curves.

In Figure 5.9, we produced a surface plot of all the generated yield curves, up until our last observation on 07-05-2019. By "Rate" in the z-axis we mean annualized yield in percent. We notice the big spikes between 2007 and 2008, which clearly illustrates the impact the financial crisis had on the Norwegian yield curve. The most significant spike seems somewhat disproportionate to

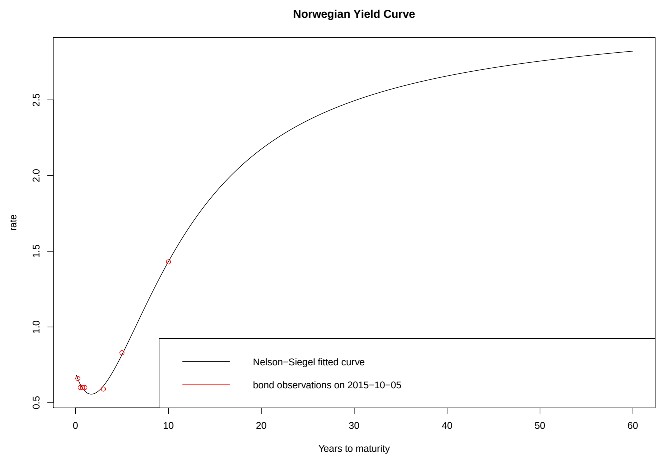


Figure 5.6: Example of Nelson-Siegel fitted data

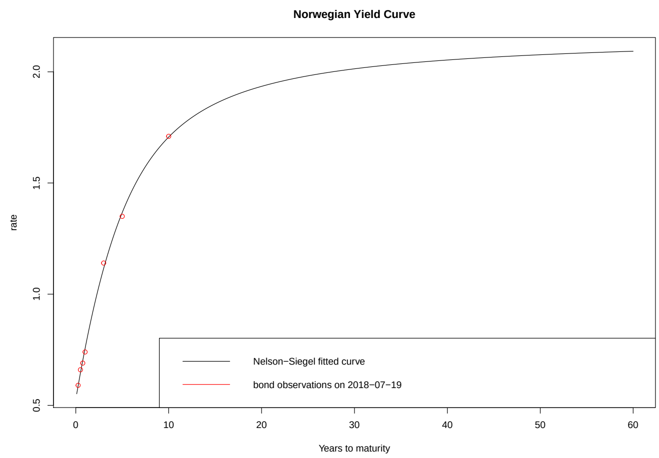


Figure 5.7: Example of Nelson-Siegel fitted data

the rest of the observations, which may suggest an imperfect fit. In the years following, we see that the interest rates slowly decays.

In 1994, Svensson [46] extended the Nelson-Siegel model, which is now called the Nelson-Siegel-Svensson model. This model includes a fourth term, which can be thought of as a second hump, and is together with the Nelson-Siegel model widely used by central banks to model the yield curve [1].

[1] See Aljinović et al. [1], for a review of methods used by several central banks across the world.

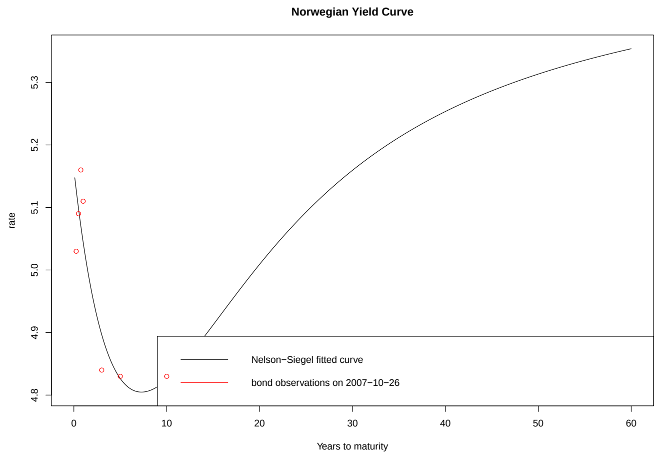


Figure 5.8: Example of Nelson-Siegel fitted data

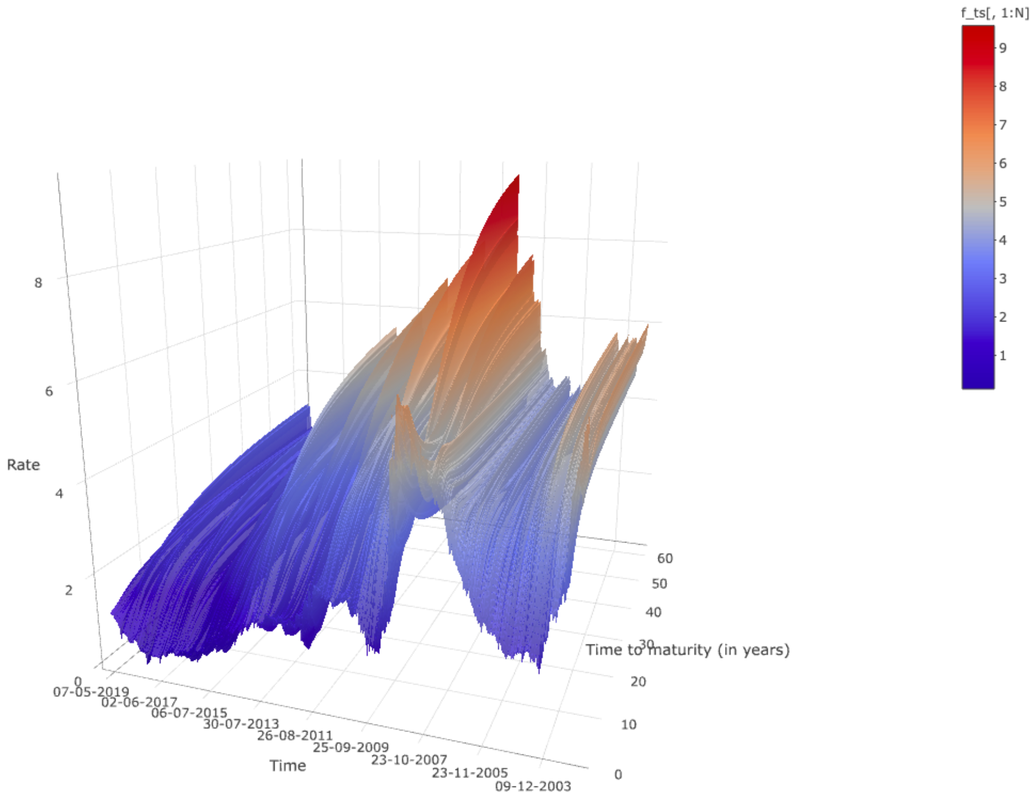


Figure 5.9: The entire historic yield curve

5.3 Mean, Variance and PCA of the Norwegian Yield curves

Now that we have investigated the raw data, as well as the transition to smooth curves, we are now applying the methods of estimation to the yield curves. We continue our empirical pursuits by studying the mean and variance of the Norwegian yield curves. Also, we put forth a brief functional principal component analysis. Most of the plots from this section is due to the comprehensive R package **FDA** by Ramsay et al. [38], which includes a plethora of vital tools for performing functional data analysis.

Mean and Correlation

Starting with the mean function, recall that from Definition 4.2.3, the natural estimator for the functional mean is given by

$$\hat{\mu}(x) = \frac{1}{N}\sum_{i=1}^{N} X_i(x), \quad \text{for } 0 \leq x \leq 60.$$

In Figure 5.10, we see that there are four plots of the mean function of the observations. The blue dashed line denotes a confidence band consisting of the 5th and 95th percentile of the observations at a fixed time to maturity. Going backward in time, we see that the confidence around the mean decreases, but interestingly the most substantial gap, which is seen in the lower right plot - is only at about 5 points. Notice also that the confidence bands for the recent years seem relatively small, even back to 2016.

The empirical covariance and correlation with respect to matuarity times s and t, are given by the equations

$$\begin{aligned}
\hat{c}(t,s) &= \frac{1}{N}\sum_{i=1}^{N}(X_i(t)-\hat{\mu}(t))(X_i(s)-\hat{\mu}(s)),\\
\hat{\rho}(t,s) &= \frac{\hat{c}(t,s)}{\sqrt{\hat{c}(t,t)}\sqrt{\hat{c}(s,s)}},
\end{aligned}$$

for $0 \leq t, s \leq 60$. As we can see from Figure 5.11, the correlation increases for larger s and t. Nevertheless, as the s and t get further apart, the correlation decays fast, but is never below 0.7. Looking closer at the correlation surface, we notice some small irregularities around zero to three years to maturity, which might be due to the majority of our bond observations are located before three years to maturity.

Principal Component Analysis

We present a plot of the four first principal components, which we recall are the estimated eigenfunctions $\pm\hat{v}_j$ for $j = 1, \dots, 4$ of the empirical covariance

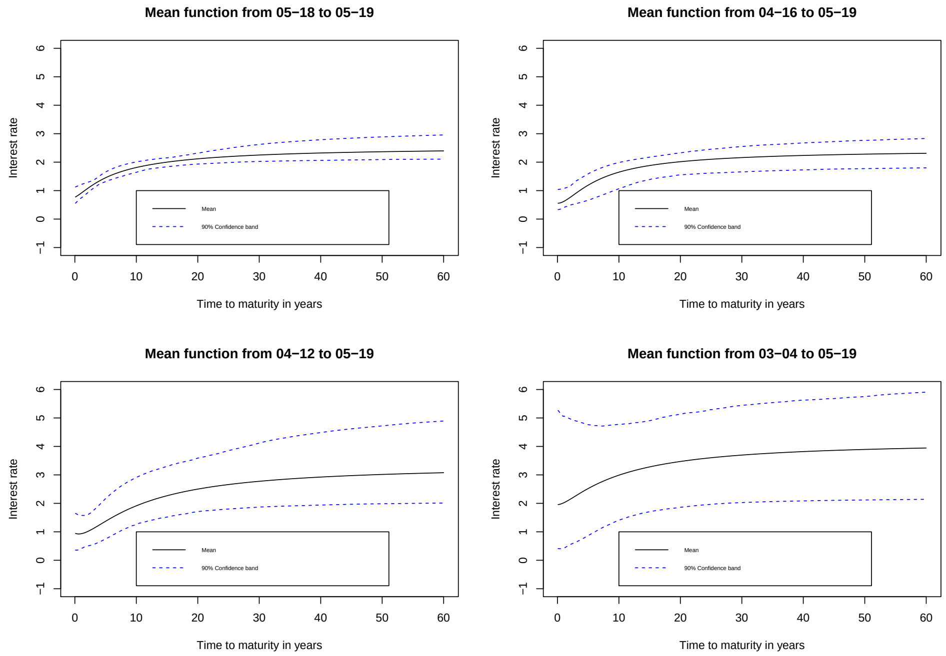
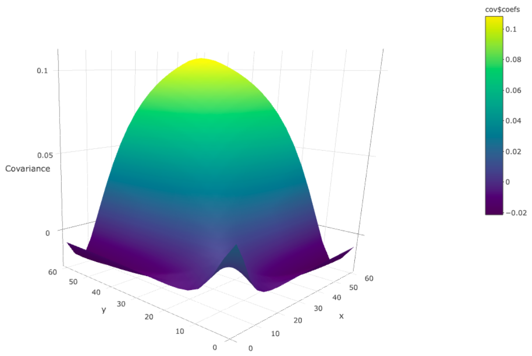
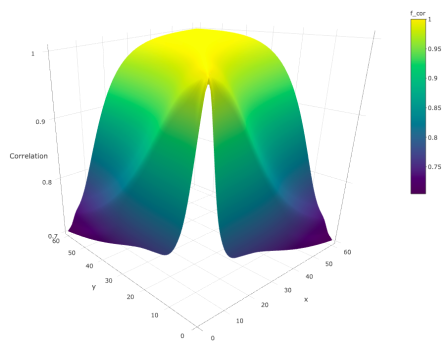


Figure 5.10: Historical estimated mean function of the Norwegian yield curves



(a) Sample covariance surface



(b) Sample correlation surface

Figure 5.11: Dependency surface plots

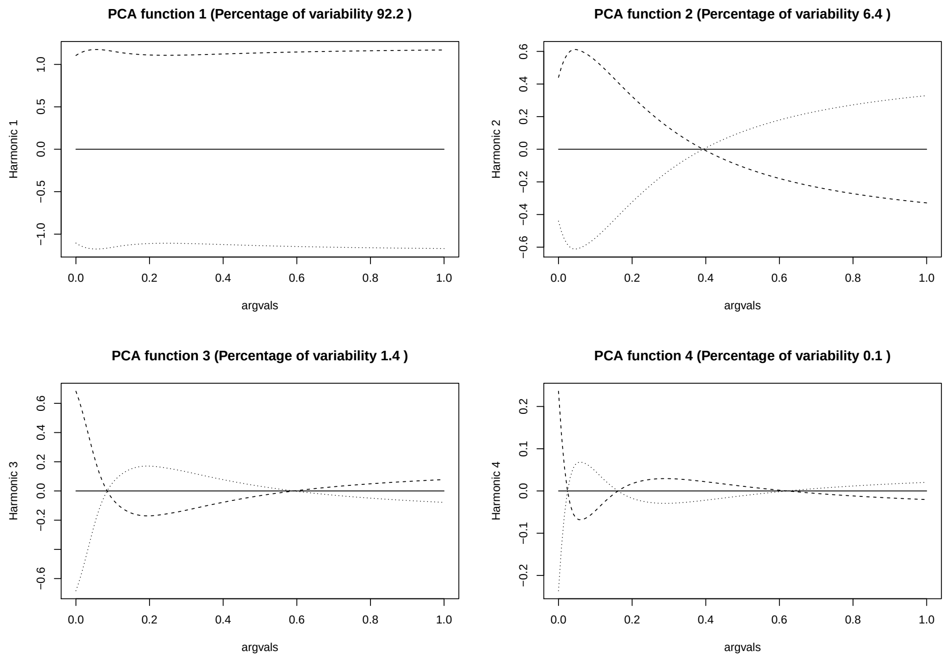


Figure 5.12: The first four principal components

operator $\hat{C}$. The estimated functional principal components describes the direction of noise which correlates the most with the data. In Figure 5.12, we see the first four principal components of the yield curves. Recall that the percentage of variability in the plots corresponds to the estimated eigenvalues $\hat{\lambda}$ of the covariance operator.

The first three components describe nearly all of the variance in the data. This is to be expected since we smoothed the data with a three-parameter family of functions. The first component is almost constant and represents the average yield over the maturities. The second component describes the "hump", whereas the third component, in some sense, captures the decay part of the yield. The fourth component, which also has slight "hump" before reverting to zero, accounts for 0.1% of the total variance and is therefore insignificant.

We notice that the third principal component admits a hump on a smaller scale than in the second principal component. This may indicate that we can smooth the data with the Nelson-Siegel-Svensson family of functions, which admits the possibility of a second hump.

5.4 The Functional KPSS test

Setting the scene

In this section we set the scene for the functional KPSS test, as formulated in Horváth et al. [24]. The test assumes $L^2([0,1])$-valued observations, with the usual norm and inner product.

We put forth some technical assumptions. Let $f : S^\infty \to L^2([0,1])$ be some measurable function, and $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ a sequence of iid functions taking values in S. Furthermore, let $\{\eta_j\}_{j\in\mathbb{Z}}$ be a sequence of *Bernoulli shifts*,

$$\eta_j = f(\varepsilon_j, \varepsilon_{j-1}, \dots), \tag{5.3}$$

which implies that η_j is a well-defined stationary stochastic process. The term in (5.3) is in itself a topic of several publications, see for instance Rosenblatt [41], who first devised the construction.

Furthermore, we let $E[\eta_0(t)] = 0$ for all t, and $E[\|\eta_0\|^{2+\delta}] < \infty$ for $0 < \delta < 1$, and the sequence $\{\eta_n\}_{n=-\infty}^{\infty}$ can be approximated by a so-called ℓ-dependent sequence $\{\eta_{n,\ell}\}_{n=-\infty}^{\infty}$ which satisfies

$$\sum_{\ell=1}^{\infty}(E\left[\|\eta_n - \eta_{n,\ell}\|^{2+\delta}\right])^{1/k}, \tag{5.4}$$

for some $k > 2 + \delta$. The above sequence $\eta_{n,\ell}$ is defined as

$$\eta_{n,\ell} = g(\varepsilon_n, \varepsilon_{n-1}, \dots, \eta_{n-\ell+1}, \varepsilon^*_{n,\ell,n-\ell}, \varepsilon^*_{n,\ell,n-\ell-1}, \dots), \tag{5.5}$$

where the $\varepsilon_{n,\ell,i}$'s are independent copies of ε_0 independent of$\{\varepsilon_i\}_{i\in\mathbb{Z}}$. We will call the above statements, the *stationarity assumptions*.

The functional hypothesis test

Furthermore, we state the hypothesis test,

$$H_0 : X_j(t) = \mu(t) + \eta_j(t), \quad 1 \le j \le N, \tag{5.6}$$

with $\mu(t) \in L^2([0,1])$. We are not testing for stationarity in the strict sense, but for stationary about the mean function $\mu(t)$. The test hypothesis test is an extension of the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, known from econometrics. They constructed a test where the null-hypothesis is stationarity, as opposed to the earlier works of Dickey and Fuller [15] which assumed the existence of a unit root in H_0.

The general alternative hypothesis is that H_0 does not hold. But, in the paper of Hórvath et al. [24], they propose three alternative hypotheses, making this a so-called *Portmanteau* test. Also, they consider two test statistics, one based on the curves themselves, T_N, and one based on the finite-dimensional projections of the curves on the functional principal components, M_N. Here, we will also restrict ourselves only to pursue the first test statistic T_N, as well as simplifying the test to only account for the alternative hypothesis of non-stationarity.

We define,

$$T_N = \int_0^1 \int_0^1 Z_N^2(x,t)\mathrm{d}t\mathrm{d}x, \tag{5.7}$$

with

$$Z_N(x,t) = S_N(x,t) - xS_N(1,t), \quad 0, x, t \leq 1, \tag{5.8}$$

where

$$S_N(x,t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Nx \rfloor} X_i(t), \quad 0 \leq x, t \leq 1, \tag{5.9}$$

where $\lfloor \cdot \rfloor$ is the *floor* function which maps a real number x to the greatest integer less than or equal to x, for example $\lfloor 3.4 \rfloor = 3$. For $X_j(t) = X_j$, i.e., for scalar observations, the test statistic T_N coincide with the numerator of the KPSS test statistic.

Furthermore, the limit distribution of T_N depends on the eigenvalues of the *long run covariance* of the errors,

$$C(t,s) = E[\eta_0(t)\eta_0(s)] + \sum_{\ell=1}^{\infty} E[\eta_0(t)\eta_\ell(s)] + \sum_{\ell=1}^{\infty} E[\eta_0(s)\eta_\ell(t)]. \tag{5.10}$$

The long-run covariance function $C(t,s)$ is positive definite, and thus there exist $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots$ and orthonormal functions $\{\phi_j(t)\}_{j=1}^{\infty}$ with $0 \leq t \leq 1$, such that the following holds,

$$\lambda_j \phi_j(t) = \int_0^1 C(t,s)\phi_j(s)\mathrm{d}s, \tag{5.11}$$

for $1 \leq j < \infty$.

For the purpose of proving the asymptotic distribution of the test statistic, we present the Gaussian process introduced in Berkes et al. [5]

$$\Gamma(x,t) = \sum_{i=1}^{\infty} \lambda^{1/2} W_i(x)\phi(t). \tag{5.12}$$

We can now present the theorem which characterizes the limit distribution of T_N. But before doing so, we must establish a preliminary theorem.

Theorem 5.4.1. *[24, Theorem A.1] Assume that the stationarity assumptions above holds. Then*

$$\sum_{\ell=1}^{\infty} \lambda_\ell < \infty, \tag{5.13}$$

and for all N, we can define a sequence of Gaussian process $\Gamma_N(x,t)$ such that

$$\Gamma_N(x,t) \underset{d}{=} \Gamma(x,t), \quad \text{for } 0 \le x, t \le 1, \tag{5.14}$$

and

$$\sup_{0\le x\le 1} \int_0^1 (V_N(x,t) - \Gamma_N(x,t))^2 \mathrm{d}t = \mathcal{O}(1), \tag{5.15}$$

where

$$V_N(x,t) = N^{-1/2} \sum_{i=1}^{\lfloor Nx \rfloor} \eta_i(t). \tag{5.16}$$

Clearly, we cannot in practice observe all covariances of $C(t,s)$, hence the need for an estimator arises. Note that under H_0, we have that

$$C(t,s) = \mathrm{Cov}(X_0(t), X_0(s)) + \sum_{i=1}^{\infty} \left(\mathrm{Cov}(X_0(t), X_i(s)) + \mathrm{Cov}(X_0(s), X_i(t))\right),$$

for $0 \le t, s \le 1$.

It is shown in Hórvath et al. [23] that we have for the estimator of the long run covariance,

$$\int_0^1 \int_0^1 (\hat{C}_N(t,s) - C(t,s)))^2 \mathrm{d}t\mathrm{d}s \underset{P}{\longrightarrow} 0, \tag{5.17}$$

where,

$$\hat{C}_N(t,s) = \gamma_0(t,s) + \sum_{i=1}^{N-1} K(\frac{i}{h})(\gamma_i(t,s) - \gamma_i(s,t)), \tag{5.18}$$

with

$$\gamma_i(t,s) = \frac{1}{N} \sum_{j=i+1}^{N} (X_j(t) - \bar{X}_N(t))(X_{j-i}(s) - \bar{X}_N(s)). \tag{5.19}$$

Here $\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i(t)$, and the function $K(\cdot)$ is a *kernel function*. Kernel functions are heavily used in econometrics as a tool for discretizing the infinite numbers of autocovariances. The function $K(\cdot)$ must satisfy the properties

(i) $K(0) = 1$
(ii) $K(u) = 0$ if $u > c$ for some $c > 0$
(iii) $K(\cdot)$ is continuous on the compact set $[0, c]$, given the c above.

The parameter h is called the *window*, or the *smoothing bandwidth*, which depends on the number of observations in following way,

(i)
$$h(N) \to \infty$$

(ii)
$$\frac{h(N)}{N} \to 0 \quad \text{as} \quad N \to \infty.$$

Typical choices of kernels are e.g. the *Parzen* kernel, *Bartlett* kernel or *flat top* kernel. Simulating functional autoregressive process, Hórvath et al. [24] obtained satisfactory results using the flat top kernel,

$$K(t) = \begin{cases} 1, & 0 \leq t < 0.1 \\ 1.1 - |t|, & 0.1 \leq t < 1.1 \\ 0, & |t| > 1.1 \end{cases}, \tag{5.20}$$

with $h = N^{1/2}$. However, the study of kernels and bandwidths has been of great importance in the realm of econometric research, and choosing a combination should therefore not be taken lightly. But an in-depth study of such choices is beyond the scope of this thesis.

We can now formulate the relation

$$\hat{\lambda}_j \hat{\phi}_j = \int_0^1 \hat{C}_N(t, s)\hat{\phi}_j(s)\mathrm{d}s, \tag{5.21}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots$ and $\hat{\phi}_1(t), \hat{\phi}_2(t), \ldots$ denotes the empirical eigenvalues and eigenfunctions respectively.

For our application of the functional hypothesis test, we characterize the asymptotic distribution of T_N as presented in Horváth et al. [24]. We give the same proof as they provide, but we include some of the computational gaps.

Theorem 5.4.2. *[24, Theorem 2.1] Assume that H_0 and the technical assumption from before holds, then*

$$T_N \xrightarrow[d]{} \sum_{i=1}^{\infty} \lambda_i \int_0^1 B_i^2(x)\mathrm{d}x, \tag{5.22}$$

where the $B_i(x)$'s denote independent Brownian bridges, i.e.

$$B_i(x) = (W_i(x)|W_i(1) = 0), \tag{5.23}$$

for standard Wiener processes $W_i(x)$ where $x \in [0, 1]$.

Proof. Let

$$V_N^0(x, t) = V_N(x, t) - xV_N(1, t),$$

where

$$V_N(x, t) = N^{-1/2} \sum_{i=1}^{N} \eta_i(t).$$

The partial sum definition of Z_N yields,

$$\begin{aligned} Z_N(x, t) &= S_N(x, t) - xS_N(x, t) \\ &= N^{-1/2} \sum_{i=1}^{\lfloor Nx \rfloor} X_i(t) - xN^{-1/2} \sum_{i=1}^{Nx} X_i(t). \end{aligned}$$

By assumption, we have $X_i(t) = \mu(t) - \eta_i(t)$, which by straightforward computations gives us that

$$Z_N(x,t) = V_N^0(x,t) + \mu(t)\left(\frac{\lfloor Nx \rfloor - Nx}{N^{1/2}}\right). \tag{5.24}$$

Furthermore, we show that Z_N can be approximated by V_N^0 with an error of order $\mathcal{O}(1)$,

$$\begin{aligned}
\sup_{x\in[0,1]} ||Z_N(x,t) - V_N^0(x,t)|| &= \sup_{x\in[0,1]} ||\mu(t)\left(\frac{\lfloor Nx \rfloor - Nx}{N^{1/2}}\right)|| \\
&\leq ||\mu(t)\left(\frac{1}{N^{1/2}}\right)|| \\
&= \frac{1}{N^{1/2}}||\mu||.
\end{aligned}$$

We may then rewrite (5.7) as,

$$\int_0^1\int_0^1 Z_N^2(x,t)\mathrm{d}x\mathrm{d}t = \int_0^1\int_0^1 \left(V_N^0(x,t)\right)^2 \mathrm{d}x\mathrm{d}t + \mathcal{O}(1).$$

Define

$$\Gamma(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x)\varphi_i(t),$$

which by Theorem 5.4.1 implies

$$T_N \xrightarrow[d]{} \int_0^1\int_0^1 (\Gamma^0(x,t))^2\mathrm{d}x\mathrm{d}t,$$

where $\Gamma^0(x,t) = \Gamma(x,t) - x\Gamma(1,t)$.

By the definition of $\Gamma(x,t)$ we get that

$$\Gamma^0(x,t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} B_i(x)\varphi(t),$$

where $B_1, B_2, \dots$ are iid Brownian bridges.

Consequently

$$\begin{aligned}
\int_0^1\int_0^1 (\Gamma^0(x,t))^2\mathrm{d}x\mathrm{d}t &= \int_0^1\int_0^1 \left(\sum_{i=1}^{\infty} \lambda_i^{1/2} B_i(x)\varphi(t)\right)^2 \mathrm{d}x\mathrm{d}t \\
&= \int_0^1 \left(\sum_{i=1}^{\infty} \lambda_i B_i^2(x)\right)\mathrm{d}x,
\end{aligned}$$

since $\phi(t)$ is a family of orthonormal functions. We interchange the sum and integral by the Dominated Convergence Theorem, and the limit is proved. ∎

The test statistic can then be approximated by quantity

$$\hat{T}_d = \sum_{i=1}^{d} \hat{\lambda}_i \int_0^1 B_i(x)\mathrm{d}x. \tag{5.25}$$

It remains to describe the distribution of $\int_0^1 B_i(x)\mathrm{d}x$. In Hórvath et al. [24] they suggest using an expansion from Shorack and Wellner [44], which leads us to define the following approximation

$$\hat{T}_{d,J} = \sum_{i=1}^{d} \hat{\lambda}_i \sum_{j=1}^{J} \frac{Z_j^2}{j^2\pi}, \tag{5.26}$$

where the Z_j's are iid standard normal random variables. In obtaining the test statistic $\hat{T}_{d,J}$ we will Monte Carlo simulate the above expression. Accordingly, the null hypothesis if rejected if T_N is larger than the $95th$ percentile of the simulated distribution of $\hat{T}_{d,J}$.

5.5 Performing the Functional KPSS Test

In this section, we perform the functional KPSS test. We want to test if the full history of the Nelson-Siegel smoothed Norwegian yield curves are stationary. Recall that the data [36] comprises 4082 trading days, from 08-01-2003 to 07-05-2019. Following the lines of Horváth et al. [24] and Kokoszka and Young [29], we separate the data into consecutive segments of length $N = 50, 150$ and 500 days.

When using tools from the R-package **FDA**, we must first transform our multivariate data into a functional time series object. Such an object much be defined through a basis in R. For simplicity, we use the B-spline basis provided by the same R-package since our data is non-periodic. We found it sufficient to pass the argument of 20 basis functions, in approximating the already smooth yield curves, i.e., we approximate the yield functions by cubic splines with 20 basis functions and 18 knots.

We estimate the long-run covariance $\hat{C}(t,s)$ by (5.18), with the flat-top kernel from (5.20) and bandwidth $h = \sqrt{N}$. We obtain the estimated eigenvalues of $\hat{C}(t,s)$, through the relation

$$\hat{\lambda}_j \hat{\phi}_j = \int_0^1 \hat{C}_N(t,s)\hat{\phi}_j \mathrm{d}s.$$

The test statistic (5.7) is naively approximated as a Riemann-sum. We found $x = [0\ 0.01\ 0.02\ \cdots\ 1]$, to be sufficient for our study, i.e. x has 100 entries evenly distributed between 0 and 1. The variable t has 600 evenly

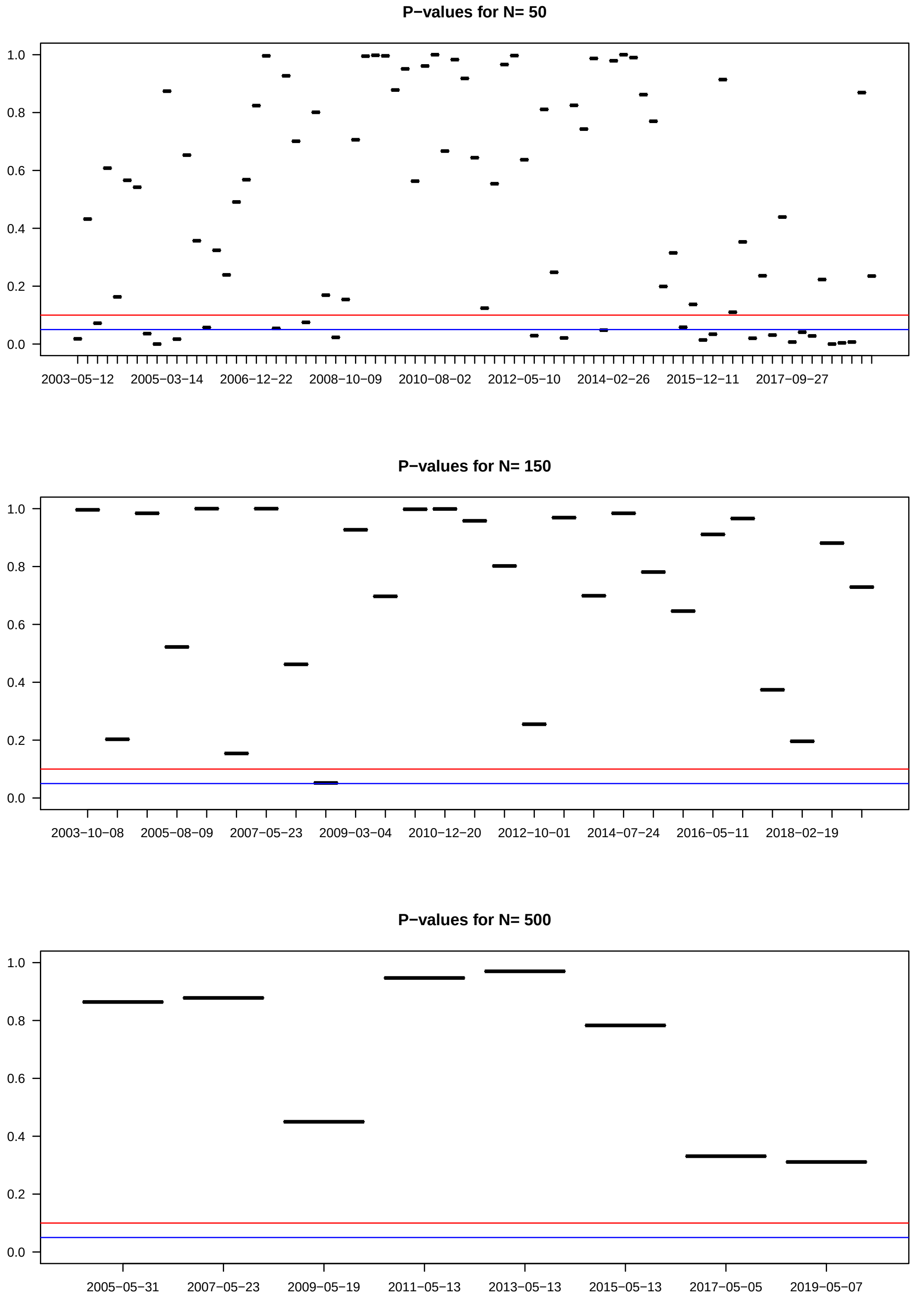


Figure 5.13: P-values for $N = 50, 150, 500$

distributed between 0 and 1, which means

$$T_N = \int_0^1 \int_0^1 Z_N^2(x,t)\mathrm{d}t\mathrm{d}x \tag{5.27}$$

$$\approx \frac{1}{100 \times 600} \sum_{i=1}^{100} \sum_{j=1}^{600} Z_N^2(x,t). \tag{5.28}$$

Then, we compute the distribution of the test statistic

$$\hat{T}_{d,J} = \sum_{i=1}^{d} \hat{\lambda}_i \sum_{j=1}^{J} \frac{Z_j^2}{j^2\pi},$$

under H_0 with $d = 10$, and $J = 100$. Doing so, we obtain an empirical distribution by Monte Carlo simulation of $\hat{T}_{d,J}$ for each segment. Suppose we have simulated observations of $T_1, T_2 \dots T_N$, we, thereafter, establish the *empirical cumulative distribution function* (ECDF) of each $\hat{T}_{d,J}^{(i,N)}$ for $i = 1, 2 \dots 4082/N$ for $N = 50, 150$ and 500. We define the ECFD as

$$F_N^{(i)}(t) = \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{(T_j \leq t)} \tag{5.29}$$

Since this is a one-sided test, and we reject the hypothesis if $95 \leq F_N^{(i)}(T_N)$ we provide an overview of the empirical P-values in Figure 5.13. The P-values are calculated from $P^{(i,N)} = 1 - F_N^{(i)}(T_N)$, where $i = 1, 2 \dots 4082/N$ belongs to each segment for $N = 50, 150, 500$. The red and blue line indicates the 0.10 and 0.05 rejection levels respectively.

In Figure 5.13 we chose $d = 10$ and $J = 100$ in simulating the distribution of $\hat{T}_{d,J}^{(i,N)}$. For $N = 50$, we reject 18/82 segments, and for $N = 150$ and $N = 500$, we do not reject any of the time segments at the 0.05 level. This indicates that for shorter time periods, the test is inclined to reject the hypothesis of stationarity.

In Kokoszka and Young [29], they perform a functional KPSS test on raw bond data from daily United States Federal Reserve yield curve rates defined for maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months, without using any financial smoothing method. Whereas they suggest the yield curves might be non-stationary, we cannot conclude that the same deduction holds for the Nelson-Siegel smoothed Norwegian yield curves. In Kokoszka and Young [29], however, they perform a KPSS type of test, which under H_0, the data are assumed *trend stationary* in contrast to our level type of test. That is, their null hypothesis is formulated as

$$X_j(t) = \mu(t) + n\xi(t) + \eta_j(t),$$

for $1 \leq j \leq N$, where $\mu(t)$ and $n\xi(t)$ denotes the intercept and slope, respectively.

5.6 Concluding remarks

Throughout this chapter, we have presented the framework of functional data analysis, extended the notion of scalar time series to that of functional time series, let alone performed an empirical study of Norwegian yield curves. In total, we could not reject the null-hypothesis of stationarity in the Norwegian yield curves when conducting the functional KPSS test. The rejections of the null-hypothesis in the small-segment test, suggests the existence of some short term non-stationarity in our data. As opposed to the conclusion of non-stationarity in Kokoszka and Young [29], we find no reason to reassess the established notion of stationarity in yield curves, especially for the Norwegian yield curves.

Regarding the untreated government bond observations, we found indications of an autoregressive structure of the bonds. Modelling the bonds separately, we found it constructive to use an order of two or three, although the slight difference between the short and long time to maturity may pose a challenge in doing so. If we can accurately capture an autoregressive structure of the bonds, one can, for instance, implement classical methods for forecasting.

Practitioners in the field of functional data analysis rarely use other spaces than $L^2([a, b])$, when dealing with observation taking values in a separable Hilbert space. Due to the pleasant behavior of the classical L^2-spaces, they are heavily used in practice, often without taking into consideration the innate equivalence-class construction of such spaces.

To that end, we suggest examining other choices for domains for the observations. In the particular case of yield curves, we suggest an investigation of the inferential properties of observations taking values in the Filipovic space, H_w. It is not necessarily an easy task. We showed earlier in Example 4.2.6 that the natural estimator for the mean is not consistent in square norm for H_w-valued observations. We, therefore, might need strict requirements on the function space we are working within for the estimators to behave nicely. A starting point for further analysis could be to construct a function space consisting of all such spaces where, e.g., the mean is consistent, and expand from thereon.

Additionally, there is the option of modeling yield curves as a functional autoregressive model. If we can correctly represent the yield curve as an FAR(1) process $\{X_n\}_{n\geq 1}$ as in (4.25), we can use methods of functional time series forecasting, for instance the *predictive factor method* as proposed in Kargin and Onatski [28], or using so-called *estimated kernels* as in Besse et al. [6]. The latter article also reports that the functional methods outperformed the classical scalar time series prediction methods in their application on geophysical data.

In summary, functional data analysis is still in the early stages of being a branch of statistics. We have presented a few of the immediate links to classical scalar statistics, as well as some justification for why the functional methods in theory succeeds, through the LLN and CLT. Already there exists a sizable toolbox for different types of analysis of functional observations, including testing for stationarity, independence, and change points, not to mention two-sample inference for mean and covariance function, all of which can be found in Horváth and Kokoszka [22]. The need for understanding functional analysis might frighten practitioners in the field of data analysis; however, we anticipate increased use of functional methods in the industry, when standard software becomes accessible.

6 Appendix

6.1 Additional Hilbert Space Properties

We present three fundamental properties of Hilbert spaces. Denote by H a separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$.

Theorem 6.1.1 (Riesz Representation Theorem)**.** *Consider the functional $\mathcal{L} \in B(\mathsf{H}, \mathbb{R})$. There exists an unique element $h \in \mathsf{H}$ called the representer or the identifier of $\mathcal{L}$, defined by the relation*

$$\mathcal{L}(\cdot) = \langle \cdot, h \rangle \tag{6.1}$$

Theorem 6.1.2 (Cauchy–Schwarz inequality)**.** *Given two elements $h_1, h_2 \in \mathsf{H}$. Then the following inequality holds for all two elements h_1 and h_2,*

$$\langle h_1, h_2 \rangle \leq \|h_1\| \|h_2\| \tag{6.2}$$

Theorem 6.1.3 (Parseval's identity)**.** *Given any orthonormal basis $\{e_i\}_{i \geq 1}$, we have that the norm of x squared may be represented as a infinite sum of inner products,*

$$\|x\|^2 = \sum_{i=1}^{\infty} \langle x, e_i \rangle^2. \tag{6.3}$$

6.2 Itô Formula

We present Itô's formula for Hilbert-valued stochastic processes.

Theorem 6.2.1. *[14, theorem 4.32] Let Φ be L_0^2-valued process stochastically integrable in $[0,T]$ and ϕ a H-valued Bochner integrable process on $[0,T]$ $\mathbb{P}$-a.s. If also $X(0)$ is a $\mathcal{F}_0$-measurable H-valued random variable, then the following process,*

$$X(t) = X(0) + \int_0^t \phi(s)\mathrm{d}s + \int_0^t \Phi(s)\mathrm{d}sW(s), \tag{6.4}$$

is well-defined for all $t \in [0,T]$. Moreover, if $F : [0,T] \times \mathsf{H} \to \mathbb{R}$ and its partial derivatives F_t, F_x, F_{xx} are uniformly continuous on bounded subsets of $[0,T] \times \mathsf{H}$, we have

$$\begin{aligned} F(t, X(t)) = F(0, X(0)) &+ \int_0^t \langle F_x(s, X(s)), \phi(s)\mathrm{d}W(s)\rangle \\ &+ \int_0^t \big[F_t(s, X(s)) + F_x(s, X(s), \phi(s)) \\ &+ \frac{1}{2}\mathrm{Tr}(F_{xx}(s, X(s))(\Phi(s)Q^{1/2})(\Phi(s)Q^{1/2}))\big]\mathrm{d}s. \end{aligned}$$

6.3 Girsanov theorem

Girsanov's theorem is used to transform a probability measure $\mathbb{P}$ to an equivalent risk-neutral probability measure $\mathbb{Q}$. The measure $\mathbb{Q}$ plays a crucial role in the derivation of the celebrated Black-Scholes formula [9] for pricing options, and in establishing arbitrage free dynamics in general.

Theorem 6.3.1. *[14, Theorem 10.14] Let $\Psi(\cdot)$ be a U_0-valued $\mathcal{F}_t$-predictable process such that*

$$E\left[\exp(\int_0^T \langle \Psi(s), \mathrm{d}W(s)\rangle_{U_0} - \frac{1}{2}\int_0^T \|\Psi(s)\|^2_{\mathsf{U}_0} \mathrm{d}s)\right] = 1, \tag{6.5}$$

then the Q-Wiener process, $\widehat{W}(t)$ with respect to $\{\mathcal{F}_t\}_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \widehat{\mathbb{P}})$, is defined by the relation

$$\widehat{W}(t) = W(t) - \int_0^t \Psi(s)\mathrm{d}s, \tag{6.6}$$

where

$$\widehat{\mathbb{P}}(\mathrm{d}\omega) = \exp\left(\int_0^T \langle \Psi(s), \mathrm{d}W(s)\rangle_{\mathsf{U}_0} - \frac{1}{2}\int_0^T \|\Psi(s)\|^2_{\mathsf{U}_0} \mathrm{d}s\right) \mathrm{d}\mathbb{P}(\mathrm{d}\omega). \tag{6.7}$$

6.4 R-Code

```
library(YieldCurve)
library(readxl)
library(rmutil)
library(plot3D)
library(fastR)
library(fda)
library(tidyr)
library(plotly)
library(dplyr)
library(fChange)
library(qualityTools)
library(GeneralizedHyperbolic)
final_bond_data <- read_excel("~/MEGA/master_thesis/project/latex/projectv2
    /final_bond_data.xlsx",
                              skip = 5,na = "")
new_df = na.omit(final_bond_data[,c(1,8,9,10,11,5,6,7)]) # changing the
    order of the bonds and removing rows with NA
maturities = c(c(3,6,9,12)/12,3,5,10)
f_maturities = c(c(3,6,9,12)/12,3,5,10,15,30)   # Toy-maturities

# Nelson-Siegel function output
NS = function(NSvector,t){
  beta_0 = NSvector[1]
  beta_1 = NSvector[2]
  beta_2 = NSvector[3]
  lambda = NSvector[4]
  return (beta_0 + beta_1*((1-exp(-lambda*t))/(lambda*t))+beta_2*((1-exp(-
      lambda*t))/(lambda*t)-exp(-lambda*t)))
}

##### --- Plots of bond observations along with acf and pacf  --- #####
par(mfrow=c(2,2))
# 3 Month
plot(new_df$X__1, new_df$‘3 mnd
3 month‘, ylab = "Price",main="3 month bond",xlab = "Date")
acf(new_df$‘3 mnd
3 month‘)
pacf(new_df$‘3 mnd
3 month‘,lag.max = 50)

# 6 Month
plot(new_df$X__1, new_df$‘6 mnd
6 month‘, ylab = "Price",main="6 month bond",xlab = "Date")
acf(new_df$‘6 mnd
6 month‘)
pacf(new_df$‘6 mnd
6 month‘,lag.max = 50)

# 9 Month
plot(new_df$X__1, new_df$‘9 mnd
9 month‘, ylab = "Price",main="9 month bond",xlab = "Date")
```

```
acf(new_df$‘9 mnd
9 month‘,lag.max=100)
pacf(new_df$‘9 mnd
9 month‘,lag.max=50)

# 12 Month
plot(new_df$X__1, new_df$‘12 mnd
12 month‘, ylab = "Price",main="12 month bond",xlab = "Date")
acf(new_df$‘12 mnd
12 month‘)
pacf(new_df$‘12 mnd
12 month‘,lag.max = 50)

# 3 Year
plot(new_df$X__1, new_df$‘3 års
3 year‘, ylab = "Price",main="3 year bond",xlab = "Date")
acf(new_df$‘3 års
3 year‘)
pacf(new_df$‘3 års
3 year‘,lag.max=50)

# 5 Year
plot(new_df$X__1, new_df$‘5 års
5 year‘, ylab = "Price",main="5 year bond",xlab = "Date")
acf(new_df$‘5 års
5 year‘)
pacf(new_df$‘5 års
5 year‘,lag.max = 50)

# 10 Year
plot(new_df$X__1, new_df$‘10 års
10year‘, ylab = "Price",main="10 year bond",xlab = "Date")
acf(new_df$‘10 års
10year‘)
pacf(new_df$‘10 års
10year‘,lag.max = 50)
##### ---------------------------------------------- #####

##### --- AR modeling of the bonds --- #####
ar(new_df$‘3 mnd
3 month‘,order.max = 2,se.fit=FALSE)
ar(new_df$‘6 mnd
6 month‘,order.max = 2)
ar(new_df$‘9 mnd
9 month‘,order.max = 2)
ar(new_df$‘12 mnd
12 month‘,order.max = 2)
ar(new_df$‘3 års
3 year‘,order.max = 2)
ar(new_df$‘5 års
5 year‘,order.max = 2)
ar(new_df$‘10 års
10year‘,order.max=2)
####------------------------------######
```

```
###### --- Constructing Nelson-Siegel plot together with bond data ---
    #####
day = 2890
yieldcurve = new_df[day,c(2,3,4,5,6,7,8)]
NSParameters = Nelson.Siegel(rate = yieldcurve , maturity=maturities)
NSfunc = NS(NSParameters,seq(from = 0, to = 60, by = 0.1))
mylegend = c("Nelson-Siegel fitted curve", paste0("bond observations on ",
    as.Date(new_df[[day,1]])))
plot(seq(from = 0, to = 60, by = 0.1),NSfunc, type="l",main="Norwegian
    Yield Curve",xlab="Years to maturity",ylab="rate")
legend("bottomright",legend=mylegend,col=c(1,2),lty=1)
lines(maturities,yieldcurve,col="2",type="p")
##### ----------------------------------------------------- #####

# Creating functional data object and plotting
nrows = dim(new_df)[1]
ncolumns = 60/0.1
yieldcurve = new_df[1,c(2,3,4,5,6,7,8)]
NSParameters = Nelson.Siegel(rate = yieldcurve , maturity=maturities)
NSfunc = NS(NSParameters,seq(from = 0, to = 60, by = 0.1))

f_ts = data.frame(format(as.Date(new_df[[1,1]]),"%d-%m%-%Y") = NSfunc[-1])
f_ts = data.frame(f_ts)
f_ts = NSfunc[-1]
for(i in 2:nrows) {
  yieldcurve = new_df[i,c(2,3,4,5,6,7,8)]
  NSParameters = Nelson.Siegel(rate = yieldcurve , maturity=maturities)
  NSfunc = NS(NSParameters,seq(from = 0, to = 60, by = 0.1))
  f_ts = cbind(f_ts,NSfunc[-1])
}
df = data.frame(f_ts)

#####---3D Plot of the yield curves---#####
N = 600
axx <- list(
  title = "Time"
)
axy <- list(
  title = "Time to maturity (in years)"
)
axz <- list(
  title = "Rate"
)
y = seq(0.1,60,0.1)
p = plot_ly(x=~format(as.Date(new_df$X__1)[1:N],"%d-%m%-%Y"),y = ~y,
  z = ~f_ts[,1:N],type = "surface",opacity=0.9,colorscale='Picnic') %>%
      layout(scene = list(xaxis=axx,yaxis=axy,zaxis=axz))
p
###### --------------------------- #######

###### --- Plots mean function of interest rate observations --- #####
###### --- together with 90% confidence band defined by the quantiles ---
    #####
```

```
###### --- x here is how far back in time we compute the mean from ---
    #####
mean_function_plot = function(x,y){
  lq = c()
  hq = c()
  for (i in 1:length(f_ts[,1])){
    lq[i] = quantile(f_ts[i,x:y],c(0.05,0.95))[1]
    hq[i] = quantile(f_ts[i,x:y],c(0.05,0.95))[2]
  }
  plot(seq(0.1,60,0.1),rowMeans(f_ts[,x:y]),ylim=c(-1,9),ylab="Interest
      rate",
       xlab="Time to maturity in years",type="l",
       main=paste(format(as.Date(new_df[[x,1]]),"%d-%m%-%y"),"-",format(as.
           Date(new_df[[y,1]]),"%d-%m%-%y"),sep=" "))
  lines(seq(0.1,60,0.1),lq,col="blue",lty=2)
  lines(seq(0.1,60,0.1),hq,col="blue",lty=2)
  #legend(10,1,legend=c("Mean", "90% Confidence band"),
         #col=c("black", "blue"), lty=1:2,cex = 0.5)
}

par(mfrow=c(2,2))
mean_function_plot(3097,2846)
mean_function_plot(2846,2594)
mean_function_plot(2594,2345)
mean_function_plot(2345,2092)

#####--------------------------------#####

##### --- Creating FDA object and plotting estimators --- #####
a = 255
b = 500
basis =  create.bspline.basis(rangeval = c(0, 1), nbasis = 30)
Domain = seq(0, 1, length = nrow(f_ts[,a:b]))
f_data = Data2fd(argvals = Domain , f_ts[,a:b], basisobj = basis)
f_data_c = center.fd(f_data)

plot(f_data_c)
f_data_deriv = deriv.fd(f_data)
f_mean = mean.fd(f_data)
f_std = std.fd(f_data)
f_cov = var.fd(f_data)
f_pca = pca.fd(f_data_c,nharm =4)
f_cor = cor.fd(seq(0,1,0.01),f_data_c)
plot(f_pca$scores[,4])

par(mfrow=c(2,1))
plot(seq(0,60,length = length(f_std$coefs)),f_std$coefs)

axz1 <- list(
  title = "Covariance"
)
axz2 <- list(
  title = "Correlation"
)
```

```
plot_ly(x = seq(0,60,length= nrow(cov$coefs)), y = seq(0,60,length= nrow(
    cov$coefs)),  z=~cov$coefs,type='surface')%>%layout(scene = list(zaxis=
    axz1))

plot_ly(x = seq(0, 1, length = nrow(f_ts[,a:b])),y = seq(0, 1, length =
    nrow(f_ts[,a:b])),z=cov$coefs,type='surface')
plot_ly(x = seq(0, 60, length = nrow(f_cor)),y = seq(0, 60, length = nrow(f
    _cor)), z=~f_cor,type='surface')%>% layout(scene = list(zaxis=axz2))
par(mfrow=c(2,2))
plot.pca.fd(f_pca,pointplot=FALSE)

#####--------------------------------#####


### TESTING FOR STATIONARITY ###
format(as.Date(new_df[[86,1]]),"%d-%m-%y")
format(as.Date(new_df[[335,1]]),"%d-%m-%y")
format(as.Date(new_df[[586,1]]),"%d-%m-%y")
format(as.Date(new_df[[839,1]]),"%d-%m-%y")
format(as.Date(new_df[[1090,1]]),"%d-%m-%y")
format(as.Date(new_df[[1340,1]]),"%d-%m-%y")
format(as.Date(new_df[[1589,1]]),"%d-%m-%y")
format(as.Date(new_df[[1840,1]]),"%d-%m-%y")
format(as.Date(new_df[[2092,1]]),"%d-%m-%y")
format(as.Date(new_df[[2345,1]]),"%d-%m-%y")
format(as.Date(new_df[[2594,1]]),"%d-%m-%y")
format(as.Date(new_df[[2846,1]]),"%d-%m-%y")
format(as.Date(new_df[[3097,1]]),"%d-%m-%y")
format(as.Date(new_df[[3348,1]]),"%d-%m-%y")
format(as.Date(new_df[[3601,1]]),"%d-%m-%y")
format(as.Date(new_df[[3845,1]]),"%d-%m-%y")
yrs = c
    (1,86,335,586,839,1090,1340,1589,1840,2092,2345,2594,2846,3097,3348,3601,3845)


sub = c(50,150,500)
#q = c()
t = c()
pvals = c()
par(mfrow=c(3,1))
for (j in sub){
  yrs = seq(1,4082,by=j)
  t = c()
  for (i in 2:length(yrs) ) {
    a = yrs[i-1]
    b = yrs[i]
    basis =  create.bspline.basis(rangeval = c(0, 1), nbasis = 20)
    Domain = seq(0, 1, length = nrow(f_ts[,a:b]))
    f_data = Data2fd(argvals = Domain , f_ts[,a:b], basisobj = basis)
    f_data_c = center.fd(f_data)
    C = LongRun(f_data, h=(b-a)^{1/2}, kern_type = "FT",is_change=TRUE)
    D= C$e_val
    p = ecdf(sapply(1:1000, function(x) Tn_Distribution(D,10,100)))
```

```
    #q = c(q,quantile(sapply(1:1000, function(x) Tn_Distribution(D,5,100)),
         0.95))
    #t = c(t,test_statistic(f_ts[,a:b],0.1))
    t = c(t,p(test_statistic(f_ts[,a:b],0.01)))
  }
  dyrs=head(yrs,-1)
  d = format(as.Date(new_df$X__1[dyrs]))
  d = (factor(d))
  pvals=c(pvals,t)
  plot(factor(d),1-rev(t),las=1,ylim=c(0,1),main=paste("P-values for N=",j,
      sep=" "))
  abline(h=0.05, col="blue")
  abline(h=0.1, col="red")
}
1-t
dyrs=head(yrs,-1)
d = format(as.Date(new_df$X__1[dyrs]))
d = (factor(d))
plot(factor(d),1-rev(t),las=1,ylim=c(0,1))
abline(h=0.05, col="blue")
abline(h=0.1, col="red")
#####--------------------------###


###### ----- Computing sample covariance of yield curves ----- #####
centered_fts = f_ts[,1:1500] - rowMeans(f_ts[,1:1500])
Cov_fts = (1/1500)*centered_fts%*%t(centered_fts)
matplot(Cov_fts, type="l")
persp3D(seq(0.1,30,0.1),seq(0.1,30,0.1), Cov_fts,zlab="Covariance")
t = 100
X = f_ts[,1:253]
N = dim(X)[2]
plot(E, type ="l")
matplot(E,type="l")
TN_Dist = plot(sapply(1:1000, function(x) Tn_Distribution(D,4,100)))
sum(D[1:4])/sum(D)
##### ----------------------------------------- #####

##### ----- Computing the test statistic and the test statistic
    distribution -----#####
test_statistic = function(X,resolution){
  N = dim(X)[2]
  x = seq(0,1, resolution)
  Z = numeric(dim(X)[1])
  for (j in x) {
    Nx = floor(j*N)
    if (Nx < 2){
      sum1 = X[,1:Nx]
    } else{
      sum1 = rowSums(X[,1:Nx])
    }
    sum2 = j*rowSums(X[,1:N])
    Z = cbind(Z,N^{-0.5}*(sum1-sum2))
  }
```

```
  Tn = (1/dim(Z)[1])*(1/dim(Z)[2])*sum(Z^2)
  return(Tn)
}

Tn_Distribution = function (lambda,d,J) {
  Tn = 0
  Z = rnorm(J, mean = 0, sd = 1)^2
  numerator = (seq(from=1, to = J, by= 1)^2)*pi^2
  for (i in 1:d) {
    Tn = Tn + lambda[i]*sum(Z/numerator)
  }
  return (Tn)
}

##### ----------------------------------------------------------------
    #####
```

Bibliography

[1] Aljinović, Z., Poklepović, T., and Katalinić, K. “Best fit model for yield curve estimation”. *Croatian Operational Research Review* 3.1 (2012), pp. 28–40.

[2] Benth, F. E. and Krühner, P. “Representation of infinite-dimensional forward price models in commodity markets”. *Communications in Mathematics and Statistics* 2.1 (2014), pp. 47–106.

[3] Benth, F. E. and Suess, A. “Continuous-time autoregressive moving-average processes in Hilbert space”. *arXiv preprint arXiv:1701.04618* (2017).

[4] Benth, F. E. and Süss, A. “Cointegration in continuous time for factor models”. *Mathematics and Financial Economics* 13.1 (2019), pp. 87–114.

[5] Berkes, I., Horváth, L., and Rice, G. “Weak invariance principles for sums of dependent random functions”. *Stochastic Processes and their Applications* 123.2 (2013), pp. 385–403.

[6] Besse, P. C., Cardot, H., and Stephenson, D. B. “Autoregressive forecasting of some functional climatic variations”. *Scandinavian Journal of Statistics* 27.4 (2000), pp. 673–687.

[7] Billingsley, P. *Convergence of probability measures*. John Wiley & Sons, 2013.

[8] Billingsley, P. *Probability and measure*. John Wiley & Sons, 2008.

[9] Black, F. and Scholes, M. “The pricing of options and corporate liabilities”. *Journal of political economy* 81.3 (1973), pp. 637–654.

[10] Bochner, S. “Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind”. *Fundamenta Mathematicae* 20.1 (1933), pp. 262–176.

[11] Bosq, D. *Linear processes in function spaces: theory and applications*. Vol. 149. Springer Science & Business Media, 2012.

[12] Boungou, W. “Negative Interest Rates, Bank Profitability and Risk-taking”. *Bank Profitability and Risk-taking (July 8, 2019)* (2019).

[13] Carmona, R. and Tehranchi, M. R. *Interest rate models: an infinite dimensional stochastic analysis perspective*. Springer Science & Business Media, 2007.

[14] Da Prato, G. and Zabczyk, J. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

[15] Dickey, D. A. and Fuller, W. A. "Distribution of the estimators for autoregressive time series with a unit root". *Journal of the American statistical association* 74.366a (1979), pp. 427–431.

[16] Diebold, F. X. and Rudebusch, G. D. *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press, 2013.

[17] Filipovic, D. *Consistency problems for Heath-Jarrow-Morton interest rate models*. Springer Science & Business Media, 2001.

[18] Grenander, U. "Stochastic processes and statistical inference". *Arkiv för matematik* 1.3 (1950), pp. 195–277.

[19] Hamilton, J. D. *Time series analysis*. Vol. 2. Princeton New Jersey, 1994.

[20] Heath, D., Jarrow, R., and Morton, A. "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation". *Econometrica: Journal of the Econometric Society* (1992), pp. 77–105.

[21] Hörmann, S., Kokoszka, P., et al. "Weakly dependent functional data". *The Annals of Statistics* 38.3 (2010), pp. 1845–1884.

[22] Horváth, L. and Kokoszka, P. *Inference for functional data with applications*. Vol. 200. Springer Science & Business Media, 2012.

[23] Horváth, L., Kokoszka, P., and Reeder, R. "Estimation of the mean of functional time series and a two-sample problem". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.1 (2013), pp. 103–122.

[24] Horváth, L., Kokoszka, P., and Rice, G. "Testing stationarity of functional time series". *Journal of Econometrics* 179.1 (2014), pp. 66–82.

[25] Hsing, T. and Eubank, R. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.

[26] Huang, Z.-y. and Yan, J.-a. *Introduction to infinite dimensional stochastic analysis*. Vol. 502. Springer Science & Business Media, 2012.

[27] Joy, K. I. *Definition of a B-spline curve*. URL: https://wwwx.cs.unc.edu/~dm/UNC/COMP258/LECTURES/B-spline.pdf.

[28] Kargin, V. and Onatski, A. "Curve forecasting by functional autoregression". *Journal of Multivariate Analysis* 99.10 (2008), pp. 2508–2526.

[29] Kokoszka, P. and Young, G. "Testing trend stationarity of functional time series with application to yield and daily price curves". *Statistics and Its Interface* 10.1 (2017), pp. 81–92.

[30] Kolmogorov, A. and Bharucha-Reid, A. T. *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications, 2018.

[31] Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics* 54.1-3 (1992), pp. 159–178.

[32] McDonald, J. N. and Weiss, N. A. *A course in real analysis*. Taylor & Francis US, 1999.

[33] Musiela, M. "Stochastic PDEs and term structure models". *Journées Internationales de Finance* (1993).

[34] Nelson, C. R. and Siegel, A. F. "Parsimonious modeling of yield curves". *Journal of business* (1987), pp. 473–489.

[35] Øksendal, B. "Stochastic differential equations". *Stochastic differential equations*. Springer, 2003, pp. 65–84.

[36] *Oslo Stock Exchange 4 pm. Calculations by Norges Bank. Collected 07.05.19 URL*.

[37] Peszat, S. and Zabczyk, J. *Stochastic partial differential equations with Lévy noise: An evolution equation approach*. Vol. 113. Cambridge University Press, 2007.

[38] Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. *fda: Functional Data Analysis*. R package version 2.4.8. 2018. URL: https://CRAN.R-project.org/package=fda.

[39] Ramsay, J. O. "Functional data analysis". *Encyclopedia of Statistical Sciences* 4 (2004).

[40] Ramsay, J. "When the data are functions". *Psychometrika* 47.4 (1982), pp. 379–396.

[41] Rosenblatt, M. "Stationary processes as shifts of functions of independent random variables". *Journal of Mathematics and Mechanics* (1959), pp. 665–681.

[42] Shafer, G. and Vovk, V. "The origins and legacy of Kolmogorov's Grundbegriffe". *arXiv preprint arXiv:1802.06071* (2018).

[43] Shalizi, C. R. and Kontorovich, A. "Almost None of the Theory of Stochastic Processes". *Lecture Notes* (2010).

[44] Shorack, G. R. and Wellner, J. A. *Empirical processes with applications to statistics*. SIAM, 2009.

[45] Shumway, R. H. and Stoffer, D. S. *Time series analysis and its applications: with R examples*. Springer, 2017.

[46] Svensson, L. E. *Estimating and interpreting forward interest rates: Sweden 1992-1994*. Tech. rep. National bureau of economic research, 1994.

[47] Tehranchi, M. "A note on invariant measures for HJM models". *Finance and Stochastics* 9.3 (2005), pp. 389–398.

[48] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. "Functional data analysis". *Annual Review of Statistics and Its Application* 3 (2016), pp. 257–295.