Maximum Likelihood calibration of the Vasicek model to the Swedish interest rate market

MVEX01-18-12

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Axel Gerebrink
Justin Lundgren
Fredrik Malmström
Oscar Thorén

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2018

Maximum Likelihood calibration of the Vasicek model to the Swedish interest rate market

MVEX01-18-12

Kandidatarbete i matematik inom civilingenjörsprogrammet Industriell ekonomi vid Chalmers

Axel Gerebrink Justin Lundgren Fredrik Malmström Oscar Thorén

Handledare: Simone Calogero
Examinator: Marina Axelson-Fisk
Maria Roginskaya

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2018

Insitutionen för Matematiska vetenskaper
Göteborg 2018

Populärvetenskaplig presentation

Att det i nuläget råder ett negativt ränteklimat har troligtvis inte undgått någon då detta regelbundet cirkulerar i nyhetsbevakningen. Ur ett historiskt perspektiv kan nuvarande ränteklimat sägas ha sin grund i 2008 års finanskris, vilket kan anses ha orsakts av en avreglerad finansmarknad i USA som skapade ekonomiska svallvågor över nästan hela värden. Efterdyningarna av finanskrisen har resulterat i att räntorna sänkts i stora delar av världen i syfte att stimulera ekonomin. I Sverige har detta, tillsammans med svårigheter att möta inflationsmålet på 2 %, resulterat i att vi sedan 2015 har negativa räntenivåer från *Riksbanken*.

Investeringar i finansiella instrument som har räntan som underliggande tillgång har sedan länge varit stabila och förhållandevis säkra investeringar. De allra säkraste anses vara obligationer utgivna av stater. På den svenska obligationsmarknaden är svenska staten den största utgivaren och majoriteten av de utgivna instrumenten består av statsobligationer. Den låga räntan resulterar i att avkastningen på dessa obligationer avtar vilket självfallet minskar deras popularitet.

För att förutspå förändringar i ekonomin används matematiska modeller baserade på historisk data. Kritik kan riktas mot dessa modeller då det inte finns garanti för att framtida mönster över huvud taget ska uppvisa samma beteende som de historiska. Utan att använda dessa modeller skulle det istället handla om mer eller mindre kvalificerade gissningar för att avgöra vad som kan tänkas ske i framtiden. Det skiftande ränteklimatet har även påverkat dessa matematiska modeller och attribut som tidigare ansetts vara en svaghet betraktas nu som önskvärda. Det tydligaste exemplet på detta är möjligheten att modellera negativa räntor.

Vasiceks modell presenterades av Oldřich Vašíček 1977 som en möjlighet att modellera framtida räntevärden men har mött kritik på just nämnda ämne kring att modellera negativa räntevärden. Då detta nu i allra högsta grad kan anses aktuellt är det givetvis intressant att undersöka hur Vasiceks modell kan tänkas estimera det framtida svenska ränteklimatet.

Att skatta de ingående parametrarna i Vasiceks modell kan vid första anblick ses som en rätt trivial uppgift, även om matematiska tillvägagångsätt såsom maximum likelihood metoden kan innehålla en hel del algebra. Det är dock viktigt att hela tiden ha komplexiteten kring rådande räntesituation i tankarna och inse att viss modifiering kan tänkas krävas för att komplettera vanliga tillvägagångssätt. Så har även fallet visats vara för Vasiceks modell vilket ytterligare understryker det faktum att matematiska modeller med negativa räntevärden tillviss del är att segla på okänt vatten.

Trots komplexiteten och utmaningarna som tidigare nämnts så har Vasiceks modell ändå visat sig på ett önskvärt vis kunna modellera avkastningskurvan på svenska statsobligationer. Det har även visat sig att via räntesimuleringar så kan Vasiceks modell användas för prissättning av diverse räntederivat. Detta har till viss del gett upprättelse åt den snart 80 år gamle Vasicek, även om han själv vid framtagandet av modellen 1977 säkerligen såg negativa räntenivåer som en utopi.

Blickar vi framåt i tiden så har räntemarknaden visat en tendens att återgå till, vad som generellt sett betraktats vara, normala räntenivåer. Detta är fullt rimligt med tanke på de fluktuationer som räntemarknaden historiskt sett visat upp. Att ett land konstant skulle uppvisa negativa räntenivåer under långa tidsperioder får i skrivande stund betraktas som osannolikt och det kan mycket väl hända att Vasiceks modell återigen möter kritik i framtiden för dess förmåga att modellera negativa räntevärden. Dock kvarstår problematiken med att estimering för framtiden egentligen inte är mer än kvalificerade gissningar, vilket den amerikanske författaren Jonathan Raymond satte fingret på i citatet "You can't know what the future holds, though you might conjecture on it, and if you're psychic, you might venture a guess".

Sammanfattning

Sedan 2015 har räntan i Sverige varit negativ vilket historiskt sett är ett väldigt ovanligt fenomen. Därav kan de flesta matematiska räntemodeller ej modellera negativa räntevärden. Denna rapport ämnar undersöka hur den svenska räntan kan modelleras i dagens ränteklimat.

Detta kommer ske via att applicera Vasiceks modell till svenska statsskuldsväxlar. Vasiceks modell blev fokus för rapporten tack vare det faktum att det är en Gaussisk process som fångar den långsiktiga räntan och kan därför i kvalitativa termer anses modellera svenska statsobligationer. Beräkningarna gentemot modellen gjordes via maximum likelihood-metoden. Via detta upptäcktes att parametern som ämnar mäta hastigheten till den långsiktiga räntan, α, ej var väntevärdesriktig för värden nära noll. När detta togs i beaktande visade sig Vasiceks modell resultera i en relativt bra estimering i jämförelse med den svenska avkastningskurvan. Dock resulterade detta i en modell med negativ α-parameter, vilket överskred modellens tendens att gå mot den långsiktiga räntan. Istället är väntevärdet av räntan en evigt stigande funktion. Från detta kan det argumenteras att modellen inte fångar konceptet kring långsiktiga räntenivåer vilket gör modellen icke-önskvärd för tidsperioder längre än tio år.

Vidare ämnar denna rapport prissätta räntederivat via Vasiceks modell. Specifikt så testas caps och floors prissättas via Blacks modell vilket i sin tur är beroende av framtidsräntan. Dock uppmärksammades att Blacks modell inte tillåter negativa räntevärden då detta ger logaritmen av negativa värden. Istället presenteras en prissättningsmetod av dessa derivat via Monte Carlo simuleringar vilket i sin tur gav önskvärda resultat för priset på caps/floors. Från detta kan en slutsats dras att prissättning kan göras av dessa derivat för den relevanta tidsperioden fem till tio år.

Nyckelord: statsobligationer, avkastningskurva, Vasicek-modellen, maximum likelihood-metoden, korttidsränta, caps och floors.

Abstract

Since 2015 the interest rate in Sweden has been negative, which historically is a very unusual phenomenon. Hence, most mathematical interest rate models do not account for negative interest rates. This paper aims to investigate how to model the Swedish interest rate in today's environment.

This is carried out by applying the Vasicek model to Swedish treasury bills. The Vasicek model was chosen since it is a Gaussian process that incorporates mean reversion and thus in qualitative terms, it was feasible to assume that it can reproduce the yield of Swedish bonds. The calibration of the model to the Swedish treasury bill data was done with the maximum likelihood method. We found that the maximum likelihood estimation of the rate of reversion parameter α was biased for values close to zero. When accounting for this bias, we found that the Vasicek model resulted in a rather good fit in comparison with the listed Swedish yield curve. However, this also resulted in a model with negative rate of reversion, which exceeded the mean reversion. Thus, our model never incorporates mean reversion and is instead ever increasing, which makes the model invalid for longer time spans than ten years.

Furthermore, this thesis also aims to price interest rate derivatives using the Vasicek model as underlying interest rate. In particular, we try to price caps and floors using the Black model, which is reliant on the forward rate. We found however, that the Black model does not allow for negative interest rate since this implies the logarithm of negative values. Instead we present how to price these derivatives using Monte Carlo simulations, which resulted in a satisfactory relation between the cap/floor price and their rate. Thus, we conclude that pricing caps/floors with this method is suitable as long as the model is valid, i.e. five to ten years.

Key words: government bonds, yield curve, Vasicek model, maximum likelihood estimation, spot rate, caps and floors

Contents

Preface

This bachelor thesis is written by four students at Industrial Engineering and Management at Chalmers University of Technology. All four students have within their studies chosen to immerse themselves in Financial Mathematics and the thesis is therefore written in collaboration with the department of Mathematical Sciences.

The thesis has been written during the spring of 2018. The context is characterized by the macroeconomic situation of the period, with negative interest rate in Sweden. Throughout history, this is a rarely seen phenomenon which makes our argumentation, analysis and conclusions deeply dependent on today's condition. The future will decide if the content is relevant just for now, or pertinent for longer time periods ahead. In fact, at the end of this project, the Swedish bond market has changed in such a manner that the spot rate is lower today than it was at the beginning of this project.

Given that the topic of this thesis not only requires knowledge in the purely mathematical field, our thesis will be distinguished from other theses written at the department of Mathematical Sciences. We therefore intend to present the theory needed to familiarize the reader with the interest rate market.

Throughout the course of this project our supervisor, Simone Calogero, has been greatly helpful and dedicated to guide us through all the setbacks that occurred. We would like to thank him for his commitment and at the same time declare his significance in our project.

During the work of this thesis the majority of the work has been conducted together rather than individually. Regarding the writing process, everybody has mainly been in charge over different sections. The group members have different areas of expertise and have therefore been assigned tasks in line with these. Nevertheless, all group members have revised and taken part in all the sections of the thesis and it can therefore be seen as the result of a collective effort rather than a collection of individual contributions. For further sectioning see Appendix.

Considering the topic of ethical aspects, there exists no such elements within the scope of this thesis and has therefore not been taken into account.

1 Introduction

The following section aims to act as an opening into the field of interest rate markets. It will shed light on the prevailing Swedish market conditions and present difficulties that have arisen in the last couples of years.

1.1 Background

Since the beginning of 2015 the Swedish interest rate climate has experienced negative interest rates, changing the conditions of the money and capital markets [26]. These low interest rate levels have raised questions amongst the banking society on how for example credit risk is managed and where the most profit line of business actually is to be found. Furthermore, it has provided a more benevolent attitude for long term investments from society's point of view. The low interest rate levels have also raised questions regarding the mathematical models implemented to manage financial derivatives, something that this thesis aims to further investigate [12].

In discussions around the money and capital markets, the Swedish government plays a major role in obtaining and managing these lines of businesses. In order to finance the public sector, the Swedish government sells various kinds of securities, for example government bonds, government bills and lottery bonds. *Riksgälden* is responsible for managing this trade and in December 2017, *Riksgälden* decided to stop selling lottery bonds because of the low interest rates [31]. Due to this unusual phenomenon with negative interest rates, investors and money managers are today rewarded with a negative yield on bonds with shorter maturity than five years. According to *Riksgälden*, there is a negative trend in the daily traded volume of government bonds since 2010, which is an obvious consequence of the negative yield. Despite this, there are still investors who buy these bonds, often with the intention to sell them on the secondary market. It is therefore important to be able to price these interest-bearing securities with negative interest rates as underlying value when they are traded on the secondary market. That is to say, a model is requested that can provide a negative interest rate and at the same time reproduce a satisfactory yield curve. Several papers such as [11], [36] and [10] highlight valuation problems that occur in negative interest rate environments and delve deeper into which models that can be used.

Furthermore, this brings the discussion regarding the Vasicek model into question. Vasicek introduced his model in 1977 but has since then often been criticized for its ability to create and model negative interest rates. By contrast, this is something that in today's financial climate is considered desirable and therefore is aligned with the challenges accompanying negative interest rates.

1.2 Purpose

The aim of this bachelor thesis is to examine whether or not the Vasicek model is appropriate to model the yield curve of Swedish government bills and bonds. It is from the beginning known that the model succeeds in qualitative terms, i.e., that the Vasicek model is capable of reproducing negative interest rates. It remains for us to prove whether it is also true in quantitative terms. If the model turns out to produce satisfactory yield levels, the project also aims to include pricing of interest rate derivatives, namely interest rates caps and floors.

1.3 Outline

The thesis will begin with a theory chapter where the basics behind the bond market and the Swedish market derivatives in particular are described. We then introduce the fair pricing of bonds and how one can derive the yield from the price. Before we present interest rates caps and floors, we have to introduce the forward rate, which acts as an underlying factor. Towards the end of the chapter, the concepts behind Vasicek model will be described and how to use the particular model to price zero coupon bonds and interest rate caps and floors.

We will subsequently proceed to the chapter regarding the method within the scope of this thesis, in which we will present and delve deeper into the maximum likelihood method in order to provide the reader with a foundation for further analysis. We will then present how the maximum likelihood method will be used to estimate the parameters in the Vasicek model. At the end of the chapter, we will also present an approach of how to price caps and floors using Monte Carlo simulations.

The fourth chapter will present the results, given the methodology from previous chapters. A comparison is here to be found amongst the actual yield curve and estimated yield curve for Swedish government bills and bonds. The result from the Monte Carlo simulations, with respect to pricing of caps and floors, will also be presented here.

The thesis will then end with a conclusion chapter, in which we will elaborate our findings and present our illation whether or not the Vasicek model is appropriate to use. We also include a discussion regarding the parameter values and bias.

Throughout every section in the thesis, references will be done to appendices. This could include proofs, clarification, MATLAB code or a more comprehensive discussion. These appendices are to be found at the end of the thesis.

2 Theory

In order to keep a nation running, including hospitals, welfare and other elements within the public sector, the government is in need of funding. This is commonly done by collecting taxes and borrowing money from the world financial markets. From a government perspective, the most frequently used financial instruments for this purpose are called government bills and bonds. The actors trading within these derivatives ranges from financial institutions and other countries to private investors [31]. The scope of the first part of this chapter will primarily be to describe the bond market in detail with a particular focus on the Swedish situation. Furthermore, the chapter intends to provide the reader with the basic foundations of the relevant concepts within financial mathematics and how to apply these in the bond market context.

2.1 The bond market

The characteristics of a government bond can vary significantly based on its factors. The two most notable factors when attempting to categorize government bonds are 1) time to maturity and 2) the possibility of paying coupons. Hence, there exist different names for different government bonds in order to provide sectioning according to the bond specifics. Different countries most commonly have different names for their individual bonds. In Sweden there are two main types of bonds, referred to as *Statsskuldväxlar* (treasury bills), with a maturity of one year or less, and *Statsobligationer* (government bonds), with a maturity of more than one year. The state institution, *Riksgälden*, has the authority to issue and manage these securities. Looking abroad, the bond market structure varies. Whilst Sweden has divided the interest rate securities in two segments, the US have divided their bonds into treasury bills, with shorter maturity than one year, treasury notes, with maturity between one to ten years, and treasury bonds, with more than ten years to maturity [17]. For simplicity, the thesis will from now on refer to all government interest rate securities with a maturity of one year or more as government bonds.

Swedish treasury bills

According to the Swedish issuer *Riksgälden*, treasury bills can be defined in the following manner,

> "Interest securities with short maturity that we are using to manage the fluctuation in the need of borrowed capital. We usually issue treasury bills with a maturity up to six months [32]."

Furthermore, treasury bills are, when issued according to the long-term plan, of two different categories. Either it is a six-month maturity bill, in which case it should be issued in the IMM months[1]. The second alternative is a treasury bill that matures within three months, in which case it should be issued during one of the remaining months. Moreover, the financial markets fluctuate, which in turn can call upon actions to keep the financial situation at bay. Hence, it is common that more short-term treasury bills are issued to balance the outflow. These treasury bills are usually maturing with respect to an already given maturity time, although this implies that the time period for which the specific treasury bill is active is shorter than what previously have been stated as guidelines [30]. At the 21$^{\text{st}}$ of March 2018, the current treasury bills follow from table 1.

Table 1: Swedish treasury bills 2018-03-21 [33].

Name	Coupon (%)	Issue date	Expire date	Time to maturity (years)	Yield today (%/year)
STB	-	2017-09-01	2018-03-21	$= 0.00$	-0.659
STB	-	2018-01-12	2018-04-18	≈ 0.08	-0.664
STB	-	2018-02-02	2018-05-16	≈ 0.16	-0.784
STB	-	2017-12-08	2018-06-20	≈ 0.25	-0.657
STB	-	2018-03-02	2018-09-19	≈ 0.50	-0.698

Swedish government bonds

The most notable difference between a treasury bill and a government bond is that the time to maturity amongst the latter is significantly further away in time when issued. Furthermore, the government bond does usually also pay a yearly coupon on the notional. In general, this coupon is of a fixed character and payed annually to the bond holder. There also exist subcategories within the concept government bonds, namely nominal and real bonds. The difference between these two are that the real government bond compensates for the inflation. Hence, these bonds are not exposed to the risk of value erosion [28]. The time to maturity for the current Swedish government bonds is in a wide spectrum, spanning from two up to more than 20 years. Despite this being the case, the number of expiration dates are restricted to just a few. The number of bonds currently active are constantly changing. This depends on, for example, how the short-term plan changes due to the financial climate, both domestic and abroad. Although *Riksgälden* is continuously handing out bonds through auctions, the maturity date and coupon payments are predetermined to the currently existing bonds. This is especially relevant when considering government bonds because of the relatively low frequency in which new sorts of bonds are created. At the 21$^{\text{st}}$ of March 2018, the current government bonds follows from table 2.

Table 2: Swedish government bonds 2018-03-21 [33].

Name	Coupon (%)	Issue date	Expire date	Time to maturity (years)	Yield today (%/year)
SGB 1052	4.25	2007-11-26	2019-03-12	≈ 0.98	-0.739
SGB 1047	5.00	2004-02-02	2020-12-01	≈ 2.70	-0.451
SGB 1054	3.50	2011-02-14	2022-12-01	≈ 4.20	-0.112
SGB 1057	1.50	2012-10-22	2023-11-13	≈ 5.65	0.166
SGB 1058	2.50	2014-02-03	2025-05-12	≈ 7.15	0.387
SGB 1059	1.00	2015-05-22	2026-11-12	≈ 8.65	0.600
SGB 1060	0.75	2017-01-27	2028-05-12	≈ 10.15	0.771
SGB 1056	2.25	2012-03-20	2032-06-01	≈ 14.20	1.110
SGB 1053	3.50	2009-03-30	2039-03-30	≈ 21.03	1.409

[1] IMM stands for International Money Market and includes March, June, September and December.

Primary market

At the primary market, the government bonds and treasury bills are traded by auctions with minimum amount of one million Swedish kronor [35]. Hence, it is unusual for specific individuals to participate in these auctions, which are held every other Wednesday. As stated in table 1 and table 2, there are at the moment five treasury bills and nine government bonds active on the Swedish market. *Riksgälden* decides which bonds or bills that will be able to purchase and the conditions for the auctions are communicated one week before. The nature of the auction is that *Riksgälden* provides a limited number of bonds/bills available for trade. The tools for distinguishing and competing between the auctioneers are then the interest rate. The participants which are able to offer the lowest interest rate with respect to the specific bond will be victorious and are able to realize the purchase. This is often in economical literature referred to as a Vickrey auction [14]. Although *Riksgälden* acts as the issuer of the bond, they themselves do not act as the reseller. This task is divided upon different capital markets, including Barclays, Danske Markets, Handelsbanken Markets, Nordea Markets, NatWest Markets, SEB and Swedbank. These actors are commonly referred to as market makers [29].

Secondary market

When entering a contract consisting of government bonds and treasury bills, the buyer ties money to a specific investment and for a specific time of maturity. Although the buyer is well aware of the circumstances surrounding the contract, in order to make the bond market reach its full potential, a secondary market needs to exist. Countless events may occur from now up until maturity which force the buyer to sell the bond. An effective secondary market provides this possibility and also enables individual investors to trade within government bonds [6]. Moreover, if private investors wish to participate in the bond market, this is usually done by buying bond funds. As mentioned, due to the negative yield connected to the short-term securities, investors buy them with the incentive to sell them on the secondary market. Without a secondary market, one can argue that there would be close to zero investors willing to buy these short-term securities today.

2.2 Fair bond pricing

The preliminary thought behind fair bond pricing is that the price of the bond, $B(t,T)$, at time t should be equal to the discounted notional value L that the holder of the bond gets paid at maturity time $T \geq t$ [19]. If the bond then pays a coupon at time $t+dt$, $T \geq t+dt \geq t$, the value of the coupon today should be discounted from the interest rate at time $t+dt$, and not at the time of maturity of the bond. Example 2.1 illustrates how to price a bond, when the interest rate is assumed to be deterministic and known.

Example 2.1. *A bond with notional value $L = \$100$ pays an annual coupon of 8%, with a maturity of $T = 5$ years. The interest rate of each year is known and presented in table 3.*

Table 3: Deterministic interest rate.

Year	Interest rate per year (%) (continuously compounded)
1	4.2
2	5.2
3	6.0
4	6.4
5	6.8

The fair price at time $t = 0$ is then the discounted value of each payoff, i.e., the present value

$$price = 8e^{-0.042\cdot 1} + 8e^{-0.052\cdot 2} + 8e^{-0.060\cdot 3} + 8e^{-0.064\cdot 4} + 108e^{-0.068\cdot 5} = \$104.63.$$

Note that in example 2.1, we assumed the interest rate to be continuously compounded, if it instead would have been annually compounded, the price would be

$$8(1-0.042)^1 + 8(1-0.052)^2 + 8(1-0.060)^3 + 8(1-0.064)^4 + 108(1-0.068)^5 = \$103.58.$$

Further on in this thesis, we will assume the interest rate to be continuously compounded. However, in example 2.1, the interest rate is assumed to be deterministic, which is not in accordance with reality. Instead, the scope of this thesis is to model the future interest rate as a stochastic process.

2.2.1 Pricing Zero-coupon bonds

A zero-coupon bond (ZCB) is a financial contract that at maturity time, T, guarantees the owner of the contract the predetermined amount L, i.e., the notional value, without any intermediate dividends [4]. It still holds that the fair price of a ZCB is the discounted notional value. However, since the spot rate that determines the discount rate is stochastic, the fair price is the expected value of the discounted notional value, as described in definition 2.1 [16].

Definition 2.1. *The fair price $B(t,T)$ of a ZCB at time t that matures at time $T \geq t$, with notional value L is*

$$B(t,T) = \mathbb{E}[L\frac{D(T)}{D(t)}|R(t)], \tag{1}$$

where

$$D(t) = e^{-\int_0^t R(\tau)d\tau}$$

is the discount process based on the stochastic spot rate $R(t)$ [4].

One easy way of modelling $R(t)$ is to assume that it follows a geometric Brownian motion, i.e.,

$$R(t) = R(0)e^{\alpha t + \sigma W(t)}.$$

However, this does not allow for a negative interest rate, which is a desired feature in this thesis and hence, this model is rejected. It is also clear that due to the linearity of the expected value, the notional amount L can be moved outside of the parenthesis in equation (1) and assumed to be equal to 1, without loss off generality. Hence, further on in this thesis, we assume that $L = 1$ in the case of every bill and bond. Furthermore, since most known models for $R(t)$ are Markov processes [5], we can state a general pricing function

$$B(t,T) = v\big(t,T,R(t)\big), \tag{2}$$

where $v\big(t,T,R(t)\big)$ is a function that depends on the time left until maturity, $T-t$, and the current spot rate $R(t)$.

2.2.2 Pricing coupon bonds

If the bond pays a coupon c_i at time $t_i \geq t$, this coupon could be modelled as a ZCB with notional value c_i and hence a price $c_iB(t,t_i)$. The price of the coupon bond $B_c(t,T)$ is then the sum of all discounted coupons in $(c_1, c_2, .., c_{n-1}, c_n)$ together with the discounted final notional value 1 at maturity (note that $t_n = T$), i.e.,

$$B_c(t,T) = B(t,T) + \sum_{i=j+1}^{n} c_iB(t,t_i), \text{ for } t \in [t_j, t_{j+1}), j = 0, 1, ..., n-1. \tag{3}$$

If we further assume (as in the case with Swedish government bonds) that $c_i = c \; \forall \; i = 1, ..., n$, equations (2) and (3) then combine into a complete model for pricing coupon bonds, namely

$$B_c(t,T) = v\big(t,T,R(t)\big) + c\sum_{i=j+1}^{n} v\big(t,t_i,R(t)\big), \text{ for } t \in [t_j, t_{j+1}), j = 0, 1, ..., n-1. \tag{4}$$

2.3 Yield to maturity

The yield is a common expression in the world of finance and can vary in its interpretation. In this thesis, the yield considered will be yield to maturity (YTM), which is expressed annually. The yield of a bond is reliant on the interest rate and the cash flow connected to the bond. One can describe the YTM of the bond as the interest rate that calculates the net present value of the coupons and the notional amount paid at maturity [22], i.e., YTM is the same as the annual internal rate of return on an investment, with the investment representing any given bond.

Depending on whether the payoff of an investment is known or not, the uncertainty in the yield differs. When considering bonds, the payoff is known from the beginning but due to the credit rating of the issuer, the yield of these bonds tend to differ. This is because the buyer undertakes a higher risk when entering into an agreement with an issuer with low credit rating and therefore expects higher payoff. Furthermore, to incentivize risk aversion, high-risk bonds are commonly cheaper. Investors who buy bonds with the incentive to own them up until time of maturity, know in advance the amount of money they will earn and can therefore easily discount the forthcoming cash-flow to obtain the present value. However, if the investors buy the bonds with incentive to sell them before the time of maturity, the price can fluctuate due to the surrounding environment within the world economy. When a bond is sold at the secondary market, the market price $B(t, T)$ decides the yield. As stated before, the yield of a ZCB is the annual interest rate that the holder of the bond obtains. Hence, definition 2.1 can be used conversely to calculate the yield $Y(t, T)$ of any given ZCB at time t that matures at time T as

$$B(t,T) = e^{-Y(t,T)(T-t)} \Rightarrow Y(t,T) = -\frac{\log B(t,T)}{T-t}, \tag{5}$$

when the time to maturity T, the price of the bond $B(t, T)$ and the payoff, i.e., the notional amount $L = 1$, are all known. Then, by using the properties from equation (3) and (5), the yield of coupon bonds can also be calculated, using the formula

$$B_{\mathrm{c}}(t,T) = e^{-Y_{\mathrm{c}}(t,T)(T-t)} + c \sum_{i=j+1}^{n} e^{-Y_{\mathrm{c}}(t,T)(t_i-t)}, \text{ for } t \in [t_j, t_{j+1}), j = 0, 1, ..., n-1. \tag{6}$$

When we know this, we can calculate the yield $Y_{\mathrm{c}}(0, 5)$ at time $t = 0$ of the coupon bond in example 2.1 as

$$8e^{-Y_{\mathrm{c}}(0,5)\cdot 1} + 8e^{-Y_{\mathrm{c}}(0,5)\cdot 2} + 8e^{-Y_{\mathrm{c}}(0,5)\cdot 3} + 8e^{-Y_{\mathrm{c}}(0,5)\cdot 4} + 108e^{-Y_{\mathrm{c}}(0,5)\cdot 5} = \$104.63 \Rightarrow Y_{\mathrm{c}}(0,5) = 6.65\%.$$

In conclusion, the yield depends on the relation between coupons and the notional value, as well as on the interest rate between every year from time t until maturity at time $T \geq t$.

Yield curve

The yield curve can be used to represent the connection between the yield of several bonds with different maturities. This graphical representation has the yield at the y-axis and time to maturity at the x-axis. A yield curve can only include bonds with same quality. It is therefore not appropriate to use government bonds and corporate bonds within the same yield curve. To create a yield curve that illustrates the Swedish government bonds and treasury bills, one has to interpret the existing instruments available on the market. As stated above, there are a total of 14 currently active instruments on the Swedish market that together creates this yield curve. Figure 1 illustrates the yield curve, based on these 14 government bonds and treasury bills available on the market[2] on the 21$^{\text{st}}$ of March 2018. For further explanation regarding different shapes and implications of the yield curve, see Appendix A.

Let us observe government bond SGB 1053 in table 2 that currently has 21.03 years left to maturity. This bond is today used as a benchmark for the yield curve at 21.03 years to maturity. As years go by, this specific type of bond will be used as benchmark for different maturities. After 20 years, this same bond will have about 1.03 years left to maturity. This means that in 20.03 years the last coupon payment from SGB 1053 will emerge. Hence, the SGB 1053 will drift closer to origin when illustrating the yield curve over the years for which the bond is active. After that, this bond can be modelled as a treasury bill up until maturity. Therefore, there is a possibility

[2] Which essentially is 13, since the first treasury bill in table 1 expired at the day of our calibration

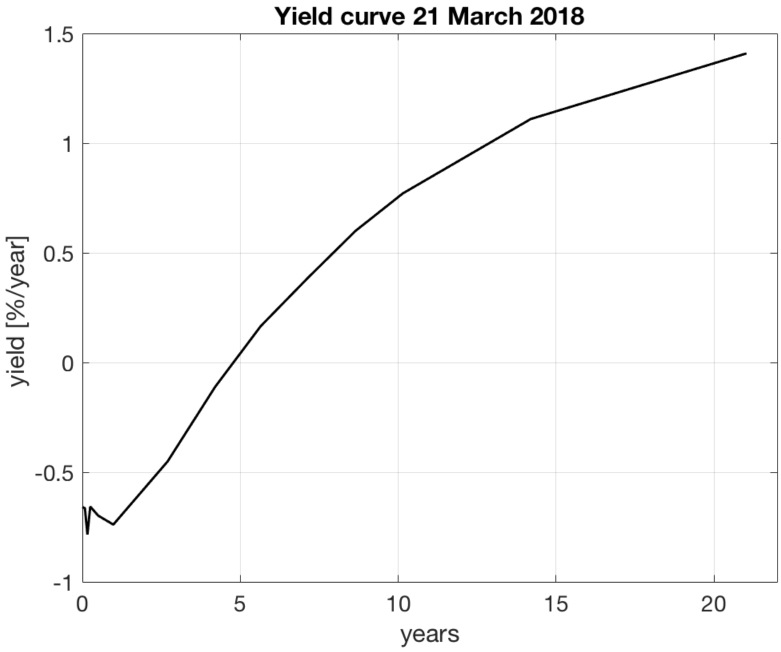

Figure 1: Yield of Swedish government bonds and treasury bills, 21 of March 2018.

that for some intervals there is an absence of instruments acting as benchmarks. In other words, the data that *Riksbanken* presents that correspond to, say five years to maturity, is sometimes a calculated average value from years nearby.

2.4 Forward rate

The forward rate is the projection of the future interest rate, based on either the spot rate of today or the yield curve. At a given point in time t, two parties entering an interest rate agreement know what the interest rate will be for time $t+\tau$, $\tau \geq 0$ as well as for time $T \geq t+\tau$ [25]. Hence, the general idea behind the forward rate is that, when buying a bond at time t that matures at time $t+\tau$ and furthermore reinvest the payoff in a new bond that matures at time T, then this should be equivalent to buying a bond at time t that matures at time T, i.e., the two investment alternatives should produce the same yield. Given that this holds, the following must be true

$$e^{Y(t,t+\tau)\tau}e^{Y(t+\tau,T)(T-(t+\tau))} = e^{Y(t,T)(T-t)}, \tag{7}$$

where, e.g., $Y(t,t+\tau)$ is the yield for the first time period $[t,t+\tau]$. Furthermore, the relations between the yield and the forward rate is that the sum of all forward rates in any given time interval must equal the yield expressed in terms of return on investment [16], i.e.,

$$Y(t,T)(T-t) = \int_t^T F(t,v)dv.$$

Hence, equation (7) extends to

$$e^{\int_t^T F(t,v)dv} = e^{\int_t^{t+\tau} F(t,v)dv}e^{\int_{t+\tau}^T F(t+\tau,v)dv},$$

which has a solution

$$F(t,T) = \lim_{\tau\to 0} F(t+\tau,T) = -\frac{\partial \log B(t,T)}{\partial T}. \tag{8}$$

The derivation of equation (8) is found in Appendix A. $F(t,T)$ is referred to as the instantaneous forward rate at time t, which is a continuous process. If we want to express the forward rate in a discrete manner, the same logic still holds, which gives us the forward rate $F_{\mathrm{D}}(t, t+\tau, T)$ in discrete terms of time as a function of the yield, i.e.,

$$F_{\mathrm{D}}(t, t+\tau, T) = \frac{e^{-Y(t,t+\tau)\tau} - e^{-Y(t,T)(T-t)}}{e^{-Y(t,T)(T-t)}(T-(t+\tau))} \;\forall\, T \geq t+\tau \geq t. \tag{9}$$

In terms of bond pricing, equation (9) can further be simplified, using equation (5), as follows

$$F_{\mathrm{D}}(t, t+\tau, T) = \frac{B(t, t+\tau) - B(t, T)}{B(t, T)\big(T-(t+\tau)\big)}. \tag{10}$$

2.5 Interest rate caps and floors

A borrower faces two options when entering a contract with a lender, either to let the interest rate of the contract loan float freely or to lock the interest rate to a specific level. Although the former allows for changes within the interest rate, these changes are evaluated at the start of a time interval called tenor, denoted as $\rho = t_i - t_{i-1}$, i.e., interest rate periods. The specific interest rate of the first day in a period statues the level for the remaining part of the tenor. There are some obvious pros and cons with both alternatives, for instance if the borrower thinks that the interest rate will be higher in the future it is reasonable to lock the interest to some predetermined level. This usually results in a scenario in which the rate is locked at a higher value than the current level. If the borrower decides to let the interest rate float freely, the risk connected to the loan increases due to greater uncertainties. The borrower can hedge against these uncertainties by buying interest rate caps, whereas the lender wishes to hedge from another perspective, namely a decrease in the interest rates. This is then done through buying interest rate floors. These interest rate derivatives are commonly traded over the counter (OTC) and consist of sequences of n caplets and floorlets, spread across the time interval $[t_1, t_2, .., t_n]$, with the tenor ρ being the equidistant time partition [37]. That is to say, one can summarize the caplets and floorlets by

$$Cap = \sum_{i=1}^{n} Caplet_i$$

and

$$Floor = \sum_{i=1}^{n} Floorlet_i$$

when calculating the price. The contexture of caps and floors makes the derivatives fairly similar to European call and put options [20]. The caplet is an agreement between two parties in which the seller promises the buyer insurance against higher interest rates. That is to say that, if the interest rate $R(t)$ on the exercise date is higher than an agreed-upon limit K, referred to as the cap rate, the seller pays the part of the interest rate that exceeds the cap rate [16].

Given that a tenor ρ represents a time period of length $t_i - t_{i-1} \;\forall\; i = 1, ..., n$, the interest rate at the last day of the tenor is used as comparison value towards the cap/floor rate K. If $R(t)$ exceeds K, the caplet is in the money and if it falls below K, the floorlet is in the money. If these instruments are agreed upon at time t, with a tenor ρ, the first caplet/floorlet is evaluated at time $t+\rho$, and further on payed at time $t+2\rho$, depending on if the derivative is in the money or not. The payoff for a caplet at time $t+\rho$, with tenor ρ and cap rate K, with notional amount L (which, without loss of generality, is set to $L = 1$ for simplicity), is then

$$P^{\text{Caplet}}(K, t+\rho, R(t+\rho), \rho) = \rho \max(R(t+\rho) - K, 0),$$

whereas the payoff for a floorlet is

$$P^{\text{Floorlet}}(K, t+\rho, R(t+\rho), \rho) = \rho \max(K - R(t+\rho), 0).$$

The fair price of these caplets and floorlets can be calculated by a modified version of the Black-Scholes formula, namely the Black model or the Black-76 model. The foundation of this formula is that the price at time t equals the risk neutral expected payoff at time $t+\rho$ and it is calculated by [39] as

$$\Pi^{\text{Caplet}}(K,t,R(t),\rho) = \rho B(t,t+2\rho)\Big(F_D(t,t+\rho,t+2\rho)\Phi(d_1) - K\Phi(d_2)\Big) \tag{11}$$

and

$$\Pi^{\text{Floorlet}}(K,t,R(t),\rho) = \rho B(t,t+2\rho)\Big(K\Phi(-d_2) - F_{\text{D}}(t,t+\rho,t+2\rho)\Phi(-d_1)\Big), \tag{12}$$

where

$$d_1 = \frac{\log(\frac{F_{\text{D}}(t,t+\rho,t+2\rho)}{K}) + \sigma^2_{R(t+\rho)}\frac{\rho}{2}}{\sigma_{R(t+\rho)}\sqrt{\rho}}$$

and

$$d_2 = d_1 - \sigma_{R(t+\rho)}\sqrt{\rho},$$

where $\sigma_{R(t+\rho)}$ is the variance of the underlying interest rate at time $t+\rho$, $F_{\text{D}}(t,t+\rho,t+2\rho)$ is the discrete forward rate defined in equation (9), $\Phi(x)$ the cumulative normal distribution of x and $B(t,t+\rho)$ the bond price defined in equation (1). Furthermore, the observant reader may notice that neither d_1 nor d_2 contains the interest rate $R(t)$ at time t. This in fact is aligned with the logic of how the forward price in a risk-neutral market already accounts for the risk-free rate for future values [40]. Furthermore, the price of a cap at time t, with tenor ρ, cap rate K, that matures at time T is then the sum of all caplets defined in equation (7), i.e.,

$$\Pi^{\text{Cap}}(K,t,T,R(t),\rho) = \sum_{i=1}^{\frac{T-t}{\rho}-1} \rho B(t,t+(i+1)\rho)\Big(F_D(t,t+i\rho,t+(i+1)\rho)\Phi(d_1) - K\Phi(d_2)\Big), \tag{13}$$

where $t_n = T$. Moreover, the price of a floor follows the same structure,

$$\Pi^{\text{Floor}}(K,t,T,R(t),\rho) = \sum_{i=1}^{\frac{T-t}{\rho}-1} \rho B(t,t+(i+1)\rho)\Big(K\Phi(-d_2) - F_D(t,t+i\rho,t+(i+1)\rho)\Phi(-d_1)\Big). \tag{14}$$

It is then notable that depending on what model is used to estimate the interest rate, d_1 and d_2 will depend on the time, which is taken into account in section 2.6.2.

2.6 The Vasicek model

A general model for the spot rate $R(t)$ is given by the stochastic differential equation

$$dR(t) = \alpha(r - R(t))dt + \sigma R(t)^\gamma dW(t). \tag{15}$$

Furthermore, in the case when $\gamma = 0$, this is a Gaussian process known as the Ornstein-Uhlenbeck process and also called the Vasicek model (for $\gamma = \frac{1}{2}$ we have the CIR model, which is along with other models described in Appendix D). Moreover, α implies the speed of which the short rate $R(t)$ is reverting towards the average short rate r, and σ describes the instantaneous volatility of $R(t)$. $dW(t)$ is the only stochastic term that governs the model, where $W(t)$ is a standard Wiener process [8]. For further explanation regarding the parameters in the Vasicek model, see Appendix A. Vasicek has gained acknowledgement with his model as the first one to capture mean reversion, a concept that statutes a theory regarding how the interest rate eventually will move towards an average value, i.e., when the short rate $R(t) < r$, it will have an upwards drift towards r (at a rate of α), and downwards when the short rate $R(t) > r$ [13]. Equation (15) for $\gamma = 0$ then has the continuous solution

$$R(t) = R(0)e^{-\alpha t} + r(1 - e^{-\alpha t}) + \sigma e^{-\alpha t}\int_0^t e^{\alpha s} dW(s). \tag{16}$$

Also note that $\mathbb{E}[R(t)] \to r$ as $t \to \infty$, which is a feature of the mean reversion. From this, theorem 2.1 follows.

Theorem 2.1 (Mean and variance for the Vasicek model)**.** *The interest rate $R(t)$, based on the current spot rate $R(0)$, modelled by the Vasicek model is normally distributed with mean $m(t)$ and variance $Var(t)$, [16] i.e.,*

$$R(t) \sim N\big(m(t),\, Var(t)\big),$$

where

$$m(t) = R(0)e^{-\alpha t} + r(1 - e^{-\alpha t}),$$

and

$$Var(t) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}).$$

For proof, see Appendix A. However, we only prove the mean and the variance, not that it is normally distributed since this follows from [16].

2.6.1 Pricing ZCBs with the Vasicek model

Following the same pricing formula defined in equation (2), we can use the parameters α, r and σ in order to price any ZCB with the Vasicek model. The price $B_{\text{Vasicek}}(t, T)$ at time t for a ZCB with face value 1, that expires as time T can be calculated using theorem 2.2.

Theorem 2.2 (ZCB-pricing with the Vasicek model)**.** *The Vasicek price of a ZCB at time t that matures at time T is calculated as*

$$B_{Vasicek}(t,T) = \mathbb{E}[e^{-\int_t^T R(s)ds}|R(t)] = e^{-R(t)C(T-t)-A(T-t)}.$$

Where

$$C(\tau) = \frac{1}{\alpha}(1 - e^{-\alpha\tau}),$$

and

$$A(\tau) = r\tau - \frac{r(1-e^{-\alpha\tau})}{\alpha} - \frac{\sigma^2}{\alpha^2}\Big(\frac{(1-e^{-2\alpha\tau})}{4\alpha} - \frac{(1-e^{-\alpha\tau})}{\alpha} + \frac{\tau}{2}\Big).$$

For proof, see Appendix A. Moreover, if we denote $T - t$ as τ, it is notable that

$$\lim_{\tau\to 0} B_{\text{Vasicek}}(t, t+\tau) = 1,$$

thus visualizing that the logic that today's price of a ZCB that matures today is just the face value of the ZCB [8]. Furthermore, theorem 2.2 will be significant in using the Vasicek model for estimating yields further on in this project.

2.6.2 Pricing interest rate caps/floors with the Vasicek model

Using the Vasicek pricing function from theorem 2.2, alongside with equation (10), (13) and (14), we can calculate the price of caps and floors as defined in section 2.5. The price $\Pi^{\text{Cap}}_{\text{Vasicek}}(K, t, T, R_t, \rho)$ of a cap at time t, that matures at time T, with tenor ρ and rate K is then given by

$$\Pi^{\text{Cap}}_{\text{Vasicek}}(K,t,T,R_t,\rho) = \rho \sum_{i=1}^{\frac{T-t}{\rho}-1} e^{-R_tC((i+1)\rho)-A((i+1)\rho)}\Big(F_{\text{Vasicek}}(i\rho,(i+1)\rho)\Phi(d_{1_i}) - K\Phi(d_{2_i})\Big), \tag{17}$$

where $t_n = T$, $C(i\rho)$ and $A(i\rho)$ are defined as in theorem 2.2. Furthermore,

$$F_{\text{Vasicek}}(i\rho,(i+1)\rho) = (\frac{e^{-R_tC(i\rho)-A(i\rho))}}{e^{-R_tC((i+1)\rho-A((i+1)\rho)}} - 1)\rho^{-1},$$

is the Vasicek time discrete forward rate, with

$$d_{1_i} = \frac{\log(\frac{F_{\text{Vasicek}}(i\rho,(i+1)\rho)}{K}) + \text{Var}(i\rho)\frac{i\rho}{2})}{\sqrt{\text{Var}(i\rho)i\rho}} \text{ and } d_{2_i} = d_{1_i} - \sqrt{\text{Var}(i\rho)i\rho},$$

where $\mathrm{Var}(i\rho)$ is the variance of a future time $i\rho$, defined in theorem 2.1. Moreover, the price of a floor with the same parameters is

$$\Pi^{\text{Floor}}_{\text{Vasicek}}(K,t,T,R_t,\rho)=\rho\sum_{i=1}^{\frac{T-t}{\rho}-1}e^{-R_tC((i+1)\rho)-A((i+1)\rho)}\Big(K\Phi(-d_{2_i})-F_{\text{Vasicek}}(i\rho,(i+1)\rho)\Phi(-d_{1_i})\Big). \tag{18}$$

3 Method

In order to predict the future interest rate using the Vasicek model, this thesis uses the maximum likelihood estimation (MLE) to estimate values for the parameters α, r and σ. First, we present the general foundations of this method, and then we establish how to use this method on a time series where the interest rate is assumed to follow the Ornstein-Uhlenbeck process defined by the Vasicek model.

3.1 Maximum likelihood estimation

The maximum likelihood method (MLM) is a general method for estimating the unknown parameters in any probability distribution. Assume we have a sample x with n independent and identically distributed (i.i.d.) random variables, then the MLM maximizes the probability of obtaining the same sample once again. By $x=\{x_1,x_2,...,x_n\}$ we denote our sample of size n, and by $\theta=\{\theta_1,\theta_2,...,\theta_m\}$ we denote the m different parameters in the probability density function (pdf) $f(x;\theta)$. Under the assumption that the random variables are i.i.d., the joint probability distribution (jpd) $f_{jpd}(x_1,x_2,...,x_n;\theta)$ can be expressed as[3]

$$\begin{aligned}f_{jpd}(x_1,x_2,...,x_n;\theta)&=\mathbb{P}(X_1=x_1\cap X_2=x_2\cap...\cap X_n=x_n)\\&=f(x_1;\theta)\cdot f(x_2;\theta)\cdot...\cdot f(x_n;\theta)\\&=\prod_{i=1}^{n}f(x_i;\theta).\end{aligned} \tag{19}$$

The objective is then to find the parameters θ that maximizes the likelihood function defined in equation (19). This can either be done numerically or analytically, here we show how to do it analytically. The first step is to simplify the algorithm by taking the logarithm of the likelihood function and thus obtain the log-likelihood function which we denote by $L(\theta)$. Since the logarithm is a monotonic function, the values that maximizes $L(\theta)$ also maximizes the likelihood function (i.e., $f_{jpd}(x_1,...,.x_n;\theta)$ for a given sample x) [9]. Thus we define $L(\theta)$ as

$$L(\theta)=\log\prod_{i=1}^{n}f(x_i;\theta)=\sum_{i=1}^{n}\log f(x_i;\theta). \tag{20}$$

Since it is easier to differentiate a sum than a product, this facilitates the next step of the algorithm, which is to differentiate and set to zero, i.e.,

$$\frac{\partial L(\theta)}{\partial\theta}=0. \tag{21}$$

Thus, each θ_i in the gradient vector $\frac{\partial L(\theta)}{\partial\theta}$ that renders this condition, represents the mean of each parameter estimate. Theorem 3.1 tells us more about the distribution of the parameter estimates.

Theorem 3.1. *Let $x=\{x_1,...,x_n\}$ denote our sample of size n, $\theta=\{\theta_1,...,\theta_m\}$ our m different parameters and $L(\theta)$ define the log-likelihood function for some given probability distribution. For n large it then holds that each ML-estimator $\hat{\theta}_i$ $(i=1,..,m)$ is multivariate normally distributed, i.e.,*

$$\hat{\theta}_i\sim N\big(\theta_i,\big(-I_{\mathbb{E}}(\theta_i)\big)^{-1}\big),$$

[3] If, for some reason, the random variables can not be assumed to be i.i.d., this of course changes the jpd, but that is beyond the scope of this project. See Appendix B for further explanation.

where

$$\theta_i = \arg\max_{\theta_i \in \theta} L(\theta),$$

and

$$I_{\mathbb{E}}(\theta) = \begin{bmatrix} \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_1^2}] & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_2}] & \dots & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_1 \partial \theta_m}] \\ \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_2 \partial \theta_1}] & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_2^2}] & \dots & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_2 \partial \theta_m}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_m \partial \theta_1}] & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_m \partial \theta_2}] & \dots & \mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \theta_m^2}] \end{bmatrix}.$$

$\big(- I_{\mathbb{E}}(\theta_i)\big)^{-1}$ *then represents the i:th diagonal element of the inverse of the negative expected information matrix* $I_{\mathbb{E}}(\theta)$ *[9].*

3.2 MLE on the Normal distribution

Using the properties from both equation (20) and theorem 3.1, one can provide a parameter estimation of the standard normal distribution. Given the general definition of the normal distribution

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

the log-likelihood function for a sample $x = \{x_1, x_2, ..., x_n\}$ follows as

$$L(\theta) = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{22}$$

To illustrate the principle behind MLE, we differentiate equation (22) for μ and σ^2 in the following manner

$$\frac{\partial L(\theta)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu), \qquad \frac{\partial L(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{23}$$

Furthermore, deriving the $I_{\mathbb{E}}(\mu,\sigma^2)$ by first calculating

$$\frac{\partial^2 L(\theta)}{\partial \mu^2} = -\frac{n}{\sigma^2}, \tag{24}$$

and

$$\frac{\partial^2 L(\theta)}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{n}{2(\sigma^2)^2} - \frac{n}{(\sigma^2)^2} = -\frac{n}{2(\sigma^2)^2}. \tag{25}$$

Since no term is stochastic in either equation (24) or (25), the expected value is not to be considered significant. Although when investigating

$$\frac{\partial^2 L(\theta)}{\partial \mu \partial \sigma^2} = -\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu),$$

one clearly notice that the stochastic term x_i still remains. Considering that $\mathbb{E}[x_i] = \mu$, its easy to see that the $\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \mu \partial \sigma^2}] = 0$. This, together with setting the equations in (23) to 0, provides the following expression for parameter estimation and the negative expected information matrix

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \;\; \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2, \;\; -I_{\mathbb{E}}(\mu;\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}. \tag{26}$$

This same calculation logic will be used in Appendix B when deriving the parameters for Vasiceks model.

3.3 MLE on the Vasicek model

Using theorem 2.1 alongside with the pdf for the normal distribution, we can derive an expression for $L(\theta)$. Assume we have a time series $R = \{R_{t_0}, R_{t_1}, ..., R_{t_n}\}$ consisting of $n+1$ different points with equidistant time partition $dt = t_i - t_{i-1}$ in a given time interval $t_0, t_1, ..., t_n$. Theorem 2.1 can than be deduced in discrete terms where the expected value of each R_{t_i} is then given by

$$\mathbb{E}[R_{t_i}] = R_{t_{i-1}} e^{-\alpha dt} + r(1 - e^{-\alpha dt}),$$

whereas the variance is given by

$$\mathrm{Var}[R_{t_i}] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha dt}).$$

The log-likelihood function $L(\theta)$, which is an extension of the log-likelihood function on the normal distribution in equation (22), following the same in theorem 2.1, is then given by

$$\begin{aligned} L(\theta) = L(\alpha, r, \sigma^2) = &-\frac{n}{2} \log\Big(\frac{\sigma^2}{2\alpha}\big(1 - e^{-2\alpha dt}\big)\Big) - \frac{n}{2}\log 2\pi \\ &- \frac{\alpha}{\sigma^2(1 - e^{-2\alpha dt})} \sum_{i=1}^{n} \Big(R_{t_i} - R_{t_{i-1}} e^{-\alpha dt} - r\big(1 - e^{-\alpha dt}\big)\Big)^2. \end{aligned} \tag{27}$$

Using the log-likelihood function in (27), we can then derive each estimator for the parameters α, r and σ^2 as presented in theorem 3.2.

Theorem 3.2. *If R_{t_0} is the first given interest rate in the time series and R_{t_n} is the current interest rate, the parameters α, r and σ^2 in the Vasicek models are given by the ML-estimators*

$$\hat{\alpha} = -\frac{1}{dt} \log\Big(\frac{n\sum_{i=1}^{n} R_{t_i} R_{t_{i-1}} - \sum_{i=1}^{n} R_{t_i} \sum_{i=1}^{n} R_{t_{i-1}}}{n\sum_{i=1}^{n} R_{t_{i-1}}^2 - (\sum_{i=1}^{n} R_{t_{i-1}})^2}\Big),$$

$$\hat{r} = \frac{1}{n(1 - e^{-\hat{\alpha} dt})} \Big(\sum_{i=1}^{n} R_{t_i} - e^{\hat{\alpha} dt} \sum_{i=1}^{n} R_{t_{i-1}}\Big)$$

and

$$\hat{\sigma}^2 = \frac{2\hat{\alpha}}{n(1 - e^{-2\hat{\alpha} dt})} \sum_{i=1}^{n} \Big(R_{t_i} - R_{t_{i-1}} e^{-\hat{\alpha} dt} - \hat{r}(1 - e^{-\hat{\alpha} dt})\Big)^2.$$

Where $dt = t_i - t_{i-1} \ \forall \ i = 1, ..., n$. Furthermore, the presented conclusion is supported by [41], [1] and [13]. However, see Appendix B for our proof. The expected information matrix $I_{\mathbb{E}}(\theta)$ is presented in theorem 3.3, which almost coincides with the matrix presented by [1] and [2], but is in disagreement with the incomplete matrix presented by [13].

Theorem 3.3. *In the Vasicek model, the negative expected information matrix $-I_{\mathbb{E}}(\theta) = -I_{\mathbb{E}}(\alpha, r, \sigma)$ has the following expression: $-I_{\mathbb{E}}(\alpha, r, \sigma) =$*

$$\begin{bmatrix} \frac{2\alpha(dt)^2 e^{-2\alpha dt} \sum_{i=1}^{n}(R_{t_{i-1}} - r)^2}{\sigma^2(1 - e^{-2\alpha dt})} + \frac{n e^{-4\alpha dt}(e^{2\alpha dt} - 2\alpha dt - 1)^2}{2\alpha(1 - e^{-2\alpha dt})^2} & -\frac{2\alpha dt \sum_{i=1}^{n}(R_{t_{i-1}} - r)}{\sigma^2(1 + e^{\alpha dt})} & -\frac{n(1 - e^{-2\alpha dt}(2\alpha dt + 1))}{\sigma\alpha(1 - e^{-2\alpha dt})} \\ -\frac{2\alpha dt \sum_{i=1}^{n}(R_{t_{i-1}} - r)}{\sigma^2(1 + e^{\alpha dt})} & \frac{2n\alpha(1 - e^{-\alpha dt})}{\sigma^2(1 + e^{-\alpha dt})} & 0 \\ -\frac{n(1 - e^{-2\alpha dt}(2\alpha dt + 1))}{\sigma\alpha(1 - e^{-2\alpha dt})} & 0 & \frac{2n}{\sigma^2} \end{bmatrix}.$$

In comparison with [1] and [2], elements (2,1) and (1,2), the expression differs in the denominator, where they claim the denominator of these elements is $\sigma^2(1 + e^{-\alpha dt})$, whereas we claim that it is $\sigma^2(1 + e^{\alpha dt})$. However, see Appendix B for our derivation of this matrix.

3.3.1 The biased parameter α

Similar studies to ours have shown that the maximum likelihood parameter for α is biased for values close to zero. For instance, [42] and [41] have both shown that when the speed of reversion, i.e., α is small, this does not just affect the estimation $\hat{\alpha}$, but also the estimations for $\hat{r}$ and $\hat{\sigma}^2$. The impact on $\hat{\sigma}^2$ however, is so small that it is disregarded.

To provide sufficient information on the bias of $\hat{\alpha}$, we simulated an Ornstein-Uhlenbeck process as a discrete Vasicek model that follows an autoregressive process of the order 1 as

$$R_t = R_{t-dt} + \alpha(r - R_{t-dt})dt + \sigma\sqrt{dt}\epsilon_t,$$

where dt is the equidistant time partition between every point in time, chosen as $dt = \frac{1}{12}$ to be in accordance with reality, and $\epsilon_t \sim$ i.i.d.N$(0,1)$ $\forall\ t$. Figure 2 shows the result from two of these simulations, whereas table 4 gives sufficient information about the difference between the two simulations. Furthermore, the simulations ran for 200 different values of $\alpha \in (-1, 1)$ with 500 different random number generators for each $\hat{\alpha}$ and a total sample size $n = 240$. The y-axes of figure 2 thus represents the mean of these 500 estimated values of $\hat{\alpha}$ for each α.

Table 4: Principles for figure 2

Graph	Arbitrary chosen parameters (I)	Parameters chosen from data in section 4.1 (II)
R_0	0	4.51% (1998 March spot rate)
r	0.05	-0.0218
σ	0.01	0.0059

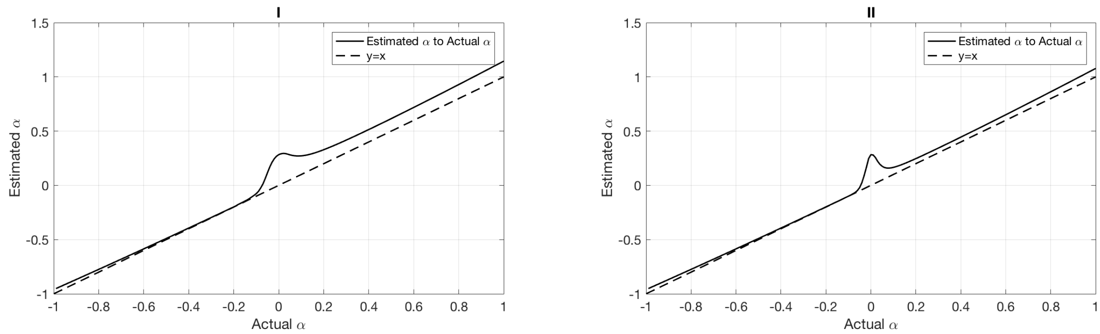

Figure 2: Parameter estimation based on simulation.

As figure 2 shows, regardless of the parameters r and σ, the estimator $\hat{\alpha}$ for α seems to be biased around values close to zero. In fact, this phenomenon is ubiquitous for every combination of the parameters r and σ. [41] mentions this as a problem with the MLE of the α-parameters, and states theorem 3.4.

Theorem 3.4. *For a stationary Vasicek process, with $dt = t_i - t_{i-1}$ fixed and $n \to \infty$, the estimator $\hat{\alpha}$ from theorem 3.2 is biased. In particular,*

$$\mathbb{E}[\hat{\alpha}] = \alpha + \frac{5 + 2e^{\alpha dt} + e^{2\alpha dt}}{2ndt} + O(n^{-2}),$$

where $O(n^{-2})$ tends to zero as n becomes sufficiently large.

Hence, the "true" value for α can be found by approximating $e^{\alpha dt}$ as $1 + \alpha\ dt$ for small values on $dt = \frac{T}{n}$, where T is defined as the "length of time expressed in years" of the data set. This then leads to the conclusion that

$$\frac{n(\mathbb{E}[\hat{\alpha}] - \frac{4}{T})}{n+2} \approx \alpha. \tag{28}$$

However, when dt is not sufficiently small, as in the case in section 4.1, where $dt = \frac{20}{240} = \frac{1}{12}$, the equation in theorem 3.4 needs to be solved numerically to obtain a fair value of α. It is then notable that some references, [42] for instance, claims that the speed reversion parameter α needs to be positive in order for the Vasicek model to make sense qualitatively. In this thesis however, this remark is overlooked, due to the fact that α and r both need to be either positive or negative to create a comprehensible yield curve, and MLE on all available Swedish bond data always returns a negative estimation of r.

3.4 Monte Carlo simulation for caps and floors

Given the complexity of the present interest rate environment, the risk that current models, together with our estimates, will not be able to properly provide a suitable price calculation for caps and floors, also leads this thesis into the subject of Monte Carlo (MC) simulations. This section of the thesis is done in a numerical manner using MATLAB, first simulating the interest rate according to equation (15). The parameters to be used in this MC simulation will be based on the outcome of MLE for the Vasiceks model, which will further be presented in the result section. However, explaining the principles behind MC simulation is out of the scope for this bachelor thesis. Hence, we assume that the reader is familiar with the concept of MC simulation. The MATLAB code used for numerical calculations is to be found in Appendix E.

Although the basics behind MC is assumed familiar, one still needs to have a discussion regarding the size, m, of the simulation. Using the properties from the central limit theorem together with several numerical test runs, its reasonable to draw a conclusion regarding what size of n that is sufficiently large. The interest rate simulation will be stored in a matrix as follows

$$A = \begin{bmatrix} R_{1,1} & R_{1,2} & \dots & R_{1,m} \\ R_{2,1} & R_{2,2} & \dots & R_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n,1} & R_{n,2} & \dots & R_{n,m} \end{bmatrix},$$

where every column represents an independent simulation of the interest rate according to equation (15). Note that every value in the top row is the same, since the starting point for the simulation is the current spot rate and hence known for all $R_{i,1}$. Furthermore, the simulation algorithm used in order to provide a reasonable estimation for the cap is

$$\begin{aligned} \Pi^{\text{Cap}}(K,t,T,R_t,\rho) &= \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{\frac{ndt}{\rho}-1} P^{\text{Caplet}}(K,t+i\rho,A(\frac{i\rho}{dt},j),\rho) \\ &= \frac{1}{m}\sum_{j=1}^{m}\rho\sum_{i=1}^{\frac{ndt}{\rho}-1} \max\Big(A(\frac{i\rho}{dt},j)-K,0\Big)e^{\sum_{k=1}^{i+\frac{\rho}{dt}} A(k,j)dt}. \end{aligned} \tag{29}$$

The parameters $dt = \frac{T}{n}$ and ρ from equation (29) represents the time between interest rate changes and the tenor size, as stated in section 2.5. The exact same logic holds during simulation of the floor price, with the exception of a payoff according to P^{Floorlet}.

4 Results

The following section aims to present our findings within the scope of this thesis. This mainly includes parameter estimation, yield curve modelling and pricing of caps and floors.

4.1 Data selection

Since the Vasicek model is a short rate model, both [3] and [24] use bonds with short maturity to model their yield curves, whereas [2] also uses the short rate to model the value of investment projects with the Vasicek model. Hence, we have limited the data sets to treasury bills with one month until maturity. The objective is then to choose the data sample that gives the best fit to the 13 different bonds presented in tables 1 and 2. After arbitrarily choosing several different sets of data, we found that the monthly averages of the one month maturity interest rate from the last 20 years gave the best fit to the real yield curve. In Appendix C, we present the results for different data sets. This data is also chosen as a foundation for graph II in figure 2. The data we choose is presented in figure 3. Also note that the vast deterioration in 2008 portraits the aftermath of the financial crisis, which could be seen as a source of error.

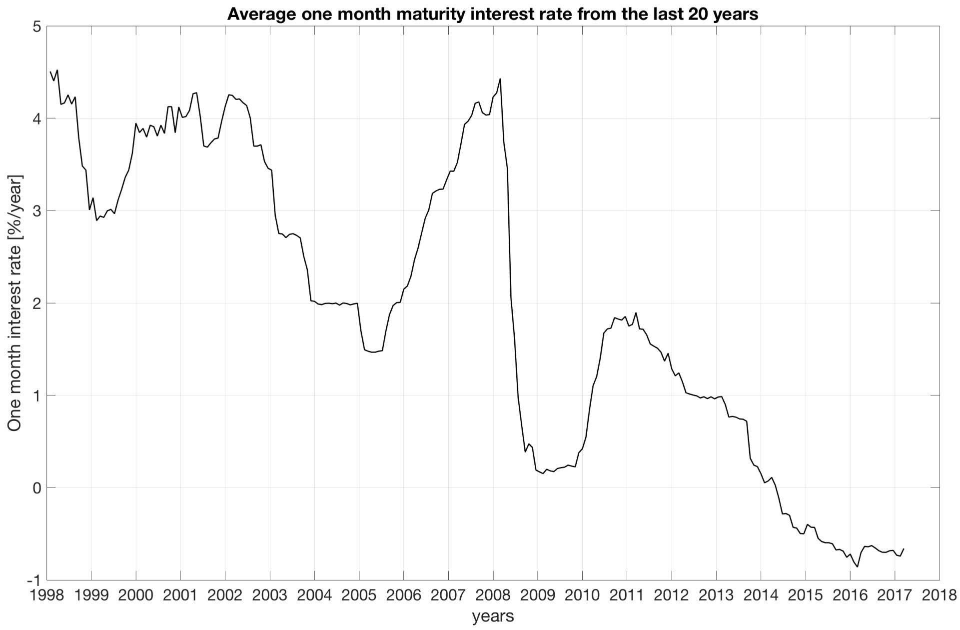

Figure 3: 20 years of one month maturity treasury bills [27].

4.2 Parameter estimation

On the foundations of theorem 3.2 and the data presented in figure 3, we used MATLAB to estimate the parameters α, r and σ. The results from the MLE are presented in table 5.

Table 5: MLE on the data from figure 3.

Parameter	Estimation	Standard error	95% Confidence interval
α	0.0630	0.08070	(-0.0951, 0.2212)
r	-0.0218	0.05670	(-0.1329, 0.0893)
σ	0.0059	0.00027	(0.0054, 0.0064)

First it is worth noticing the large standard errors of the parameter α and r. This of course creates ambiguity in interpreting the validity of this method. However, on the basis of section 4.3, we accept the parameter estimation at face value. Furthermore theorem 3.4 is essential in analyzing this outcome, since the α-value presented in table 5 is biased. Numerically solving the equation in theorem 3.4 for $\mathbb{E}[\hat{\alpha}] = 0.0630$ then gives the almost unbiased parameter $\alpha = -0.1358$. To further investigate this matter, we simulate two time series with the two different values for α presented whilst r and σ are the ones presented in table 5. Moreover, the simulations ran for 10000 different random number generators, start value $R_{t_0} = 4.51\%$ (March 1998 spot rate), $dt = \frac{1}{12}$ and $n = 240$. The results are presented in table 6.

Table 6: Parameter estimation based on different values for α.

Input parameter:	$\alpha = 0.0630$	$\alpha = -0.1358$
Estimated parameters:	$\hat{\alpha} = 0.1560$	$\hat{\alpha} = -0.1353$
	$\hat{r} = -0.0094$	$\hat{r} = -0.0231$
	$\hat{\sigma} = 0.0059$	$\hat{\sigma} = 0.0058$

As evident from figure graph II in figure 2, the estimation $\hat{\alpha}$ is clearly more biased for $\alpha \in [-0.05, 0.1]$. However it is interesting to see that the input parameter α also affects the estimation $\hat{r}$. Hence, the conclusion from this is to use the values $\alpha = -0.1358$, $r = -0.0218$ and $\sigma = 0.0059$ to model the yield of Swedish government bonds and interest rate caps/floors. However, in qualitative terms it is hard to interpret the fact that α is negative. For instance, [42] and [16] and several other references states that α has to be larger than zero, [16] even states that r has to be larger than zero, which of course is not in cohesion with the reality we are trying to model. In fact, some of these references base this claim on the assumption that the interest can only be positive, which is not the case. [38] claims that it is completely feasible to have a negative α as long as the current interest rate is larger than the mean reversion r, which is the case here. Furthermore, it is worth mentioning that in order to obtain a concave curve as in the of reality, both α and r need to be either positive or negative. In qualitative terms it would of course make more sense to have both parameters positive, but here we accept the results from theorems 3.2 and 3.4 at face value.

4.3 Yield curve modelling

On the basis of definition 2.1 and equation (6), the estimated yield can be calculated at every point in time that is equal to the maturities of the bonds in tables 1 and 2. The result is presented in figure 4, which gives a comparison between the actual yield curve presented in figure 1 from the 21^{st} of March 2018, and the yield curve generated with the Vasicek model. As figure 4 portraits,

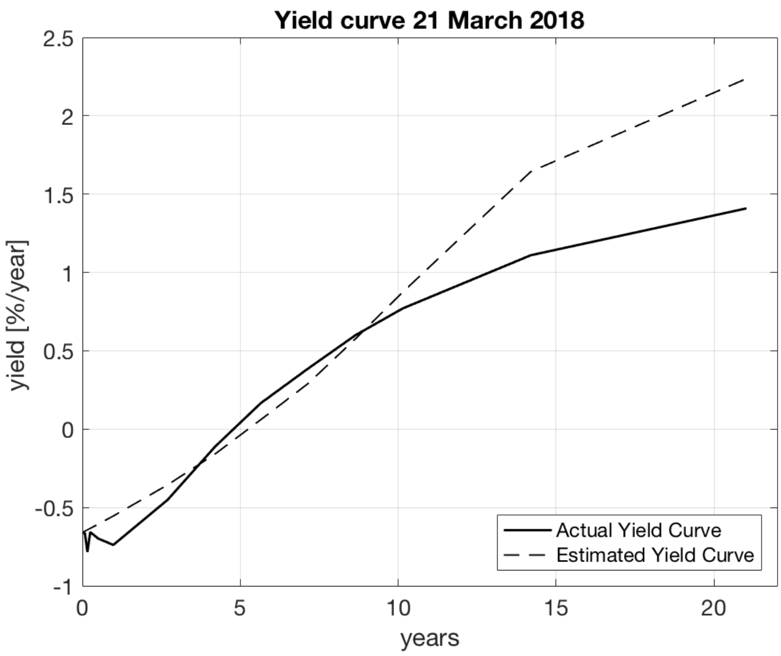

Figure 4: Yield curve comparison [27] [18].

the model works very well for bonds with maturities of ten years or less, in fact, it completely manages to capture the property of having negative interest rate for bonds with maturities of five years or less. However, for bonds with longer maturities our model seems to overestimate the yield. To adjust the estimation, several different sets of data were used, but none gave a better fit than the one presented in figure 4. Estimation of several other time series and a concluding comparison between all possible time series with time spans between 19 and 21 years can be found in Appendix C. However, the data from the last 20 years provided the best fit and hence it is the one we present here.

4.4 Pricing caps and floors

The results we obtained from the pricing functions presented in equation (17) and (14) were so unrealistic that we have chosen not the present them (some prices were actually negative, which should be virtually impossible). In fact, it turns out that the Black model presented in section 2.5 does not work for negative interest rates since negative interest rates implies negative forward rates. The negative forward rate leads to a risk neutral scenario where, depending on the rate K, one has to take the logarithm of a negative term, which is undefined. In fact, [11] mentions this as a serious drawback of the model and propose other models such as the Hull-White model and the Bachelier model to price these caps and floors. To elaborate upon this matter however, we chose to use Monte Carlo simulation to price these derivatives. The foundation for these simulations is described in section 3.4. However, before presenting our results, we restate that the price of a caplet at time t, that matures at time $t+\rho$, with rate K and tenor ρ is the expected payoff at time $t+\rho$, namely

$$\Pi^{\text{Caplet}}(K, t, R(t), \rho) = \rho\,\mathbb{E}[\max(R(t+\rho) - k, 0)].$$

The future interest rate $R(t+\rho)$ is then modeled with the Vasicek model on the same basis as the simulation presented in section 3.3.1.

Pricing caps and floors according to Monte Carlo simulation

To determine a reasonable sample size n for the MC simulation, we used the estimated Vasicek parameters from table 5 for the numerical price calculation with respect to both the cap and floor price. These together with the rest of the parameters for simulation, are to be found in table 7,

Table 7: Vasicek variables for cap and floor simulation.

Parameter	Cap	Floor
α	-0.1358	-0.1358
r	-0.0218	-0.0218
σ	0.0059	0.0059
dt	1/240	1/240
ρ	3/12	3/12
T	5	5
R_{t_0}	-0.0066	-0.0066
K	-0.01	0.01

where it is noticeable that the only parameter that differs between caps and floors is K. This is due to the context of the payoff function. The price of the cap/floor as a function of the amount of simulations n is shown in figure 5, which clearly states that a trial size of $n = 5000$ is sufficient for the MC simulation. As stated in section 3.4, this follows by the CLT.

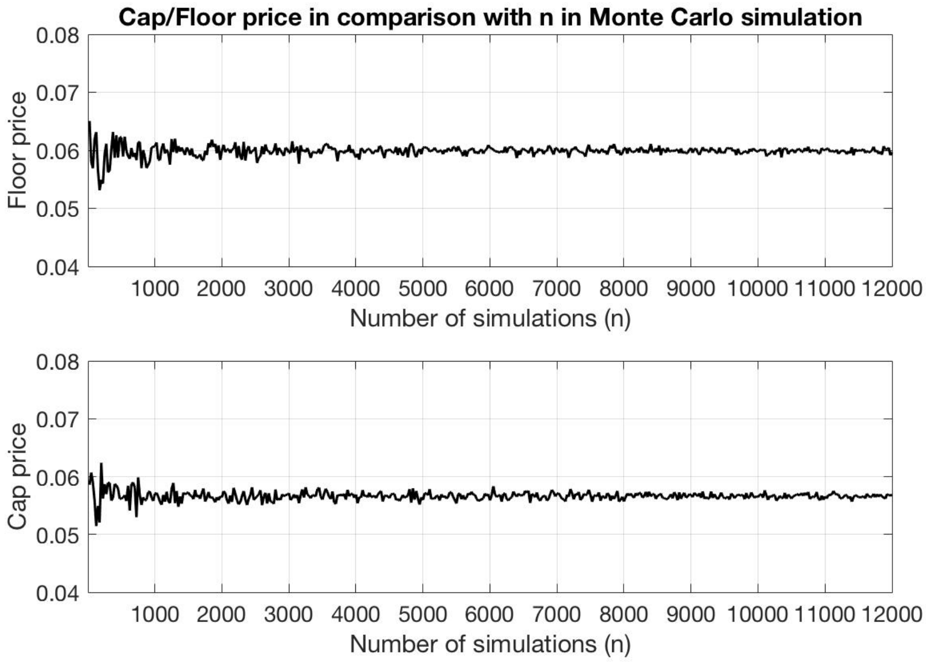

Figure 5: Cap/floor price with respect to sample size n.

Hence, the discussion shifts focus in able to actual estimate a price for both caps and floors. Since this clearly depends on the value of K, we found it reasonable to plot both the cap and floor price together with the values from table 7 but with $K \in [-0.05, 0.05]$. The result is presented below.

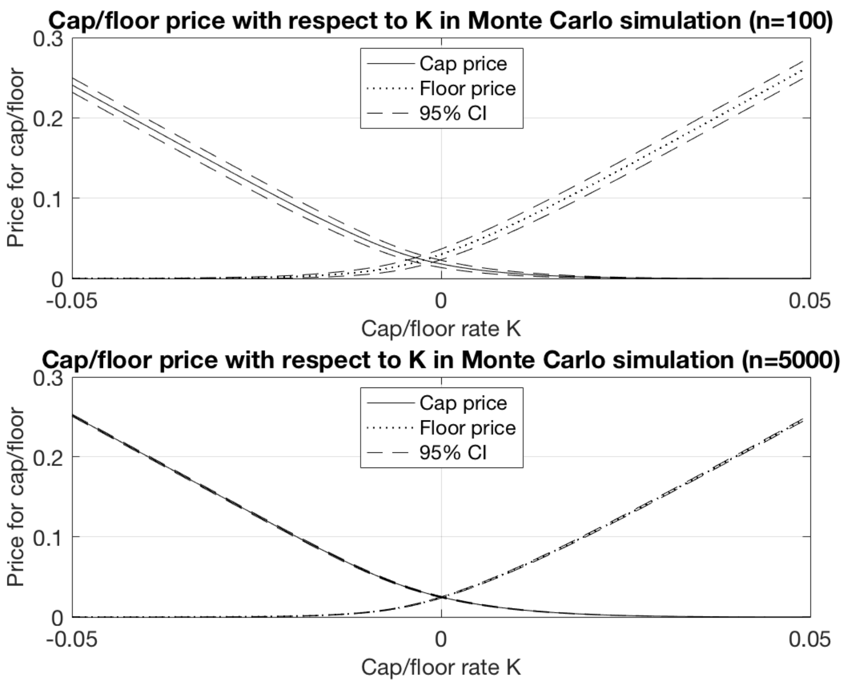

Figure 6: Cap/floor price in comparison with strike rate K for different n.

The correlation between the level of K and the final price for both the cap and the floor is expected to behave in the manner of Figure 6. Furthermore, the MC simulation proves in practice that our estimated parameters from the Vasicek model actually can be used to model the cap and floor price. The confidence interval has been created according to the normal distribution, which again follows by the CLT. The fact that a larger sample size, n, is resulting in a narrower CI, is also this aligned with the CLT. For a more detailed proof of this claim, see Appendix C. The narrow confidence intervals further strengthens the fact that $n = 5000$ is sufficient for a MC simulation.

5 Conclusion

The purpose of this report was to analyze how the Vasicek model could reproduce the yield curve of Swedish government bonds and treasury bills and thus evaluate whether or not the Vasicek model would be a suitable model for pricing Swedish bonds and interest rate derivatives. To investigate this purpose, all programming language was written in MATLAB.

When the Vasicek model was first introduced, the novelty of the model was its ability to incorporate mean reversion, i.e.

$$\lim_{t\to\infty} \mathbb{E}[R(t)] = \lim_{t\to\infty} \Big(R(0)e^{-\alpha t} + r(1 - e^{-\alpha t}) \Big) = r.$$

This however, is only true for $\alpha > 0$, which is not the case in our model. Instead, the maximum likelihood estimation resulted in a negative α-value after the bias of the estimation had been accounted for. Essentially this means that there is only a convergence towards r if $r = R(0)$, which in qualitative terms would mean that the expected value of the future interest rate is the same as the current spot rate. Due to the negative interest rates of today in Sweden, this is clearly not a feasible scenario. The more reasonable assumption is that the future spot rate will tend to positive values, as has been the case in history[4]. This is actually a feature that our model seems to encompass, e.g. the expected value of the March 2023 spot rate is (where $t = 0$ is defined as March 2018)[5]

$$\mathbb{E}[R(5)] = 0.008173.$$

Hence, the conclusion regarding the parameter estimation in the Vasicek is that, in quantitative terms, our parameters manage to capture the macroeconomic essence of today, at least in the short sighted future. However due to their definitions it would, in qualitative term, make more sense to obtain positive values for α and r. In fact, using the model with negative values for α and r redefines the meaning of the these parameters, i.e. due to the fact there is no mean reversion, r could obviously not be considered as the long-term interest rate and α, since it is negative, does not imply a realistic speed of reversion.

Regarding longer time spans however, the lack of a mean reversion leads to an every increasing future expected spot rate, e.g.,

$$\mathbb{E}[R(20)] = 0.208,$$

which clearly does not make sense. In fact, this is evident in figure 4, in which our model overestimates the yield of all bonds with maturities of 10 years or more. A possible explanation for this is the fact that the Vasicek model is a one-factor model. Since all parameters are assumed to be independent of time and since σ is the only term that determines the impact of the Wiener process, the model fails to account for all underlying factors in the long run. This is most likely the reason to why similar studies such as e.g. [24] and [2] do not use longer maturities than ten years for their estimations. In fact, even though other models are beyond the scope of this thesis, we consider it feasible to mention the existence of substitutes, as can be found in Appendix D.

Regarding the pricing of interest rate caps/floors, the lack of data on Swedish interest rate derivatives makes it impossible to collate the prices presented in figure 6 to actual market prices

[4]In fact, this is actually the case. At the current date, 14$^{\text{th}}$ of May 2018, the intersection between the x-axis and the yield curve from figure 1 has shifted from five to four years, thus visualizing that the negative interest rates seems to be diminishing.

[5]Furthermore, the model expects the spot rate to reach zero in 2.66 years.

of these derivatives. However, arguing from a qualitative perspective, the structure of the graph makes sense, as it is equable to the payoff structure of standard call and put options. Furthermore, on the basis of the evaluation from the Vasicek model, it is plausible to assume that our pricing method is not suitable for very long maturities.

Finally, in comparison with papers such as [13], [2] and [1], none of these mentions the bias of estimating α as a problem, instead they accept the estimations from the MLE at face value, with exception for [2], who evaluate their estimation with t-testing. Accounting for the bias in our estimation of α proved to be monumental in our calibration of the parameters. In comparison with the similar project MVEX01-18-82, who instead calibrated their model with the generalized method of moments, our results proved to be preferable. This is most likely due to the wider existence of papers such as [41] who investigate the bias of the MLE for any time series that is assumed to follow the Vasicek model. In fact, [41] also state that there is a bias in the estimation of σ. This was overlooked in our report however, since the bias did not account for any significant difference in the parameter estimation. We found that simulating time series with known parameters was a very useful method for analyzing the bias of estimators, which is a strategy we would recommend to further studies within the field of applied mathematics in finance.

Appendix A - Theory

Yield curve

Shape of the Yield Curve

By basic knowledge in economics one may think that the shape of the yield curve always should increase. This theory is further clarified by an example in [8]. Lets say that two persons with the same credit rating wants to borrow money, the first one for a year and the second one for ten years, the second person would most likely be charged at a higher rate. One reason to this is that both the alternative cost and the risk connected to the second loan is higher. Another reason, which in the past has been considered significant, is the inflation. By analyzing historical yield curves it is easy to see that this theory is not true for all times. Equation (5(and (6) increases the understanding in the set off between the yield and the price. The following subsections aims to explains in more detail the differences between various shapes and which financial expectations that are connected to them. If nothing else is specified, the source to the following subsections is [8]. Furthermore, the shape of every yield is presented in figure 7.

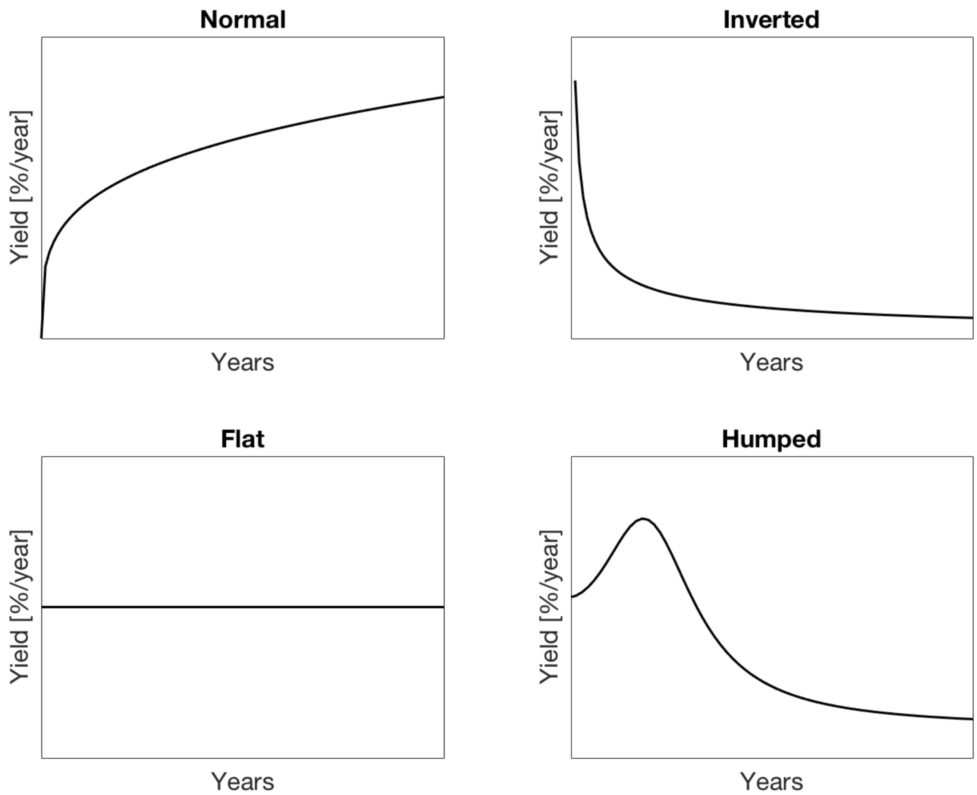

Figure 7: Shape of yield curves.

Normal

When long-term bonds yield are higher than the short-term bonds, the corresponding yield curve is referred to as normal. As the name implies, this is the most common shape which is seen as the average yield curve and it is a sign that there is a belief that the inflation and interest rates will rise [21]. Therefore, investors are less likely to buy long-term bonds with the current interest rate and subsequently, miss out on the higher future interest rate. The demand for short-term bonds rises, followed by an increase within the price, which creates a lower yield. The opposite happens to the long-term bonds. The normal yield curve indicates that the economy has been going strong for a while and that there are expectations from the investors that, in the Swedish case, *Riksbanken* will increase the interest rate in a near future [22].

Inverted

The inverted yield curve occurs in the market when the short-term yield is higher than the long-term yield. This is due to a belief that the rate will soon be lowered by *Riksbanken* which creates a rise in demand for long-term bonds. The investors sees an opportunity to lock-in the current rate, which by belief is supposed to be lower in a near future. The increase in demand for long-term bonds correspondingly rises the price and the yield goes down.

Flat

By the name it is easy to visualize the flat yield curve but the occurrence of them are rare. In fact, it may be considered as a phase that must be passed between every recession and boom in the economy. A flat yield curve means that long-term and short-term investors gets equally paid for owing bonds on a yearly basis. Hence, there are few incentives for the long-term bond holder that compensates them for the higher risk and the alternative cost.

Humped

When short-term and long-term bonds both yield lower than the medium-term bonds there is an increase in the middle of the yield curve. This scenario is uncommon and does not often occur. It can arise due to beliefs that the interest rate is increasing at first and then going to be lowered in the near future.

Application of the Yield Curve

Yield curves can be utilized in many different ways which are described in detail in [4] and this section contains a short summary of these discussions.

As a benchmark for other bonds

They yield curve of government bonds are used as a guideline when pricing other securities in the debt market. That is, the yield curve of bonds from various issuer but with the same maturity can be compared to the yield curve of government bonds with same maturity. The yield of these other bonds should be higher than the yield of the government bonds. To clarify this, if a government bond with maturity of ten years has a yield of eight percent, then other bonds with maturity of 10 years should yield at least eight percent or higher. The difference in the yield is commonly referred to as the spread.

As a predictor of future Yields

All participants in the debt market, even Central banks and government treasury departments, uses yield curves to predict future interest rates and inflation levels. As described above, the expectations of the future are different based on the current shape of the yield curve. So dependent on whether the present yield curve is normal, inverted, flat or humped, different decisions are made by the financial stakeholders.

Description of the Vasicek model parameters

The parameters in question for Vasiceks model are α, r and σ. The explanation follows below.

Description of the α parameter

As stated in section 2.6, the parameter α implies the speed of which the short rate $R(t)$ is reverting towards the average short rate r. Putting this in a context of more general term, one can see α as the factor which over time pulls the interest rate towards its long-term value r. Hence, it follows from this logic that a greater value of α makes the interest rate revert faster [15]. Furthermore, α excepts both positive and negative values, although one condition for $\alpha < 0$ is that the initial interest rate is larger than the avarage short rate [38].

Description of the r parameter

The average short rate r implies the value for which the interest rate converges over time. One simple way of illustrating this is to use mathematical software and perform Monte Carlo simulation of the interest rate. The more time you simulate the interest rate and calculates the average from all this simulation, the closer to the average short rate r should the result become.

Description of the σ parameter

σ describes the instantaneous volatility of the short term rate $R(t)$, which is assumed to be constant [8]. This assumption can be questioned due to the rapid turns in the world economy that in turn affects the short rate $R(t)$.

Proof of equation (8)

Proof. The instantaneous forward rate at time t and maturity T follows from

$$e^{\int_t^T F(t,v)dv} = e^{\int_t^{t+\tau} F(t,v)dv} e^{\int_{t+\tau}^T F(t+\tau,v)dv},$$

which is equivalent to

$$e^{\int_{t+\tau}^T F(t+\tau,v)dv} = e^{\int_t^T F(t,v)dv - \int_t^{t+\tau} F(t,v)dv} = e^{\int_t^T F(t,v)dv} e^{-\int_t^{t+\tau} F(t,v)dv}$$

where

$$e^{\int_t^T F(t,v)dv} = B(t,T)^{-1}$$

and

$$e^{-\int_t^{t+\tau} F(t+\tau,v)dv} = B(t,t+\tau)$$

which leads to

$$e^{\int_{t+\tau}^T F(t+\tau,v)dv} = \frac{B(t,t+\tau)}{B(t,T)}. \tag{A.1}$$

Solving equation (A.1) with respect to $F(t+\tau,T)$ gives

$$\int_{t+\tau}^T F(t+\tau,v)dv = log(B(t,t+\tau)) - log(B(t,T)) \Leftrightarrow$$

$$F(t+\tau,T) = -\frac{\partial}{\partial T}\Big(log(B(t,T)) - log(B(t,t+\tau))\Big) = -\frac{\partial log(B(t,T)}{\partial T},$$

which concludes the proof. □

Proof of theorem 2.1

Proof. Given equation (16) and denoting $m(t) = R(0)e^{-\alpha t} + r(1 - e^{-\alpha t})$, one can start calculation $\mathbb{E}[R(t)]$ accordingly,

$$\mathbb{E}[R(t)] = \mathbb{E}[m(t)] + \sigma e^{\alpha t}\,\mathbb{E}[\int_0^t e^{-\alpha s}\partial W(s)]. \tag{A.2}$$

First of all, one can notice that it does not exist any stochastic term in $m(t)$, hence $\mathbb{E}[m(t)] = m(t)$. Moreover, using the definition of Itô integrals, the second term of equation (A.2) becomes

$$\mathbb{E}[\sigma e^{\alpha t}(e^{-\alpha t}W(t) - \int_0^t e^{\alpha s}W(s)ds)].$$

Furthermore, using the property that the expected value of a Brownian motion is zero, clearly states that

$$\mathbb{E}[R(t)] = m(t) = R(0)e^{-\alpha t} + r(1 - e^{-\alpha t})$$

and hence concludes the first part of the proof. Continuing, the variance can be described as

$$\text{Var}[R(t)] = \mathbb{E}[R(t)^2] - \mathbb{E}[R(t)]^2 = \mathbb{E}[(m(t) + \sigma e^{\alpha t} \int_0^t e^{-\alpha s} \partial W(s))^2] - \mathbb{E}[R(t)]^2. \tag{A.3}$$

Using the same logic as in the first part of the proof, together with developing equation (A.3), we are left with

$$\begin{aligned} &\mathbb{E}[(\sigma e^{\alpha t}(e^{-\alpha t} W(t) - \int_0^t e^{\alpha s} W(s) ds))^2] \\ = (\sigma e^{\alpha t})^2 (\mathbb{E}[(e^{-\alpha t} W(t))^2] - 2\,\mathbb{E}[e^{-\alpha t} W(t) \int_0^t e^{\alpha s} W(s) ds] + \mathbb{E}[(\int_0^t e^{\alpha s} W(s) ds))^2]), \end{aligned} \tag{A.4}$$

which can be elaborated using basic stochastic principles for integration to

$$\begin{aligned} &\sigma^2 e^{-2\alpha t}(e^{-2\alpha t} t - 2\alpha e^{-\alpha t} \int_0^t e^{\alpha s} s ds + \alpha^2 \int_0^t \int_0^t e^{\alpha(s+\tau)} \min(s,\tau) ds d\tau) \\ = \sigma^2 e^{-2\alpha t}(e^{-2\alpha t} t - 2\alpha e^{-\alpha t} \int_0^t e^{\alpha s} s ds + \alpha^2 \int_0^t e^{\alpha \tau} (\int_0^\tau e^{\alpha s} s ds + \int_\tau^t e^{\alpha s} \tau ds) d\tau). \end{aligned}$$

Furthermore, using integration by parts, one is finally provided with the expression for the variance, namely

$$\text{Var}[R(t)] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$$

and hence the proof is completed. □

Proof of theorem 2.2

Proof. The price of a ZCB with maturity T at time t, according to the risk-neutral valuation framework, is derived as

$$B(t,T) = \mathbb{E}\left[e^{-\int_t^T R(s) ds} |\ R(t)\right],$$

which according to [23] is equivalent to

$$\begin{aligned} B(t,T) =& \exp\Big(\mathbb{E}[-\int_t^T R_s(R(t)) ds] - \frac{1}{2}\text{Var}[-\int_t^T R_s(R(t)) ds]\Big) \\ =& \exp\Big(\frac{R(t) - r}{\alpha}(1 - e^{-\alpha(T-t)}) - r(T-t) \\ &+ \frac{\sigma^2}{4\alpha^3}(2\alpha(T-t) - 3 + 4e^{-\alpha(T-t)} - e^{-2\alpha(T-t)}\Big) \\ =& \exp\Big(-\frac{1 - e^{\alpha(T-t)}}{\alpha} R(t) + r(\frac{1 - e^{-\alpha(T-t)}}{)}\alpha \\ &- (T-t)) - \frac{\sigma^2}{2\alpha^2}(\frac{1 - e^{-\alpha(T-t)}}{\alpha}) + \frac{\sigma^2}{2\alpha^2}(T-t) \\ &- \frac{\sigma^2}{4\alpha}(\frac{1 - 2e^{-\alpha(T-t)} + e^{-2\alpha(T-t)}}{\alpha^2})\Big) \\ =& \exp\Big(-C(\tau)R(t) + rC(\tau)^2 - r\tau - \frac{\sigma^2}{2\alpha^2}C(\tau) \\ &+ \frac{\sigma^2}{2\alpha^2}(\tau)\frac{\sigma^2}{4\alpha}C(\tau)^2\Big) \\ =& e^{-C(\tau)R(t) - A(\tau)} \end{aligned} \tag{A.5}$$

where

$$C(\tau) = \frac{1 - e^{-a(\tau)}}{a}$$

and

$$A(\tau) = r\tau - \frac{r(1 - e^{-\alpha\tau})}{\alpha} - \frac{\sigma^2}{\alpha^2}\Big(\frac{(1 - e^{-2\alpha dt})}{4\alpha} - \frac{(1 - e^{-\alpha\tau})}{\alpha} + \frac{\tau}{2}\Big).$$

□

Appendix B - Method

Log-likelihood function for dependent random variables

If it is not feasible to assume that the random variables are i.i.d., there are other ways to present the likelihood function. One example is to assume that the random variables are dependent of each other, but hold Markov properties. The jpd $f_{jpd}(x_1, x_2, ..., x_n; \theta)$ is then described as

$$\begin{aligned} f_{jpd}(x_1, x_2, ..., x_n; \theta) &= \mathbb{P}(X_1 = x_1 \cap X_2 = x_2 \cap ... \cap X_n = x_n) \\ &= f(x_1; \theta) \cdot f(x_2|x_1; \theta) \cdot ... \cdot f(x_n|x_{n-1}, x_{n-2}, .., x_1; \theta) \\ &= f(x_1; \theta) \prod_{i=2}^{n} f(x_i|x_{i-1}; \theta). \end{aligned} \tag{B.1}$$

Then $L(\theta)$ can be expressed as

$$\begin{aligned} L(\theta) &= \log f(x_1 \theta) \prod_{i=2}^{n} f(x_i|x_{i-1}; \theta) \\ &= \log f(x_1; \theta) + \sum_{i=2}^{n} \log f(x_i|x_{i-1}; \theta). \end{aligned} \tag{B.2}$$

Proof of theorem 3.2

Firstly, we state that the derivation for this proofs was done more than two months before the deadline of the project. During this time period, the notations in the main rapport have changed somehow. Hence, what is referred to as R_{t_0} in theorem 3.2 is denoted as R_1 in this proof, likewise what is referred to as R_{n+1} in theorem 3.2 is referred to as R_{t_n} here. Hence, the sum $\sum_{i=1}^{n} R_i$ in this proof is the same as the sum $\sum_{i=1}^{n} R_{t_{i-1}}$ in theorem 3.2. Furthermore, when the Vasicek model is described as a continuous process as in e.g. theorem 2.1, the spot rate today, which is denoted as $R(0)$ in that case, is the same as R_{t_n}, since it is custom to address the time now t as $t = 0$.

Proof. To derive the maximum likelihood estimators for the Vasicek model, we start by simplifying the model and expressing the parameters as follows

$$\begin{aligned} A &= e^{-\alpha t} \\ B &= r(1 - e^{-\alpha t}) \\ C &= \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}). \end{aligned}$$

The log-likelihood function $L(\theta)$ in equation (27) can then be described as

$$L(\theta) = L(A, B, C) = -\frac{n}{2} \log C - \frac{n}{2} \log 2\pi - \frac{1}{2C} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)^2 \tag{B.3}$$

in the terms of A, B and C. Furthermore, the MLE-algorithm can be elaborated using simple differentiation rules for the Vasicek parameters α,r and σ^2,

$$\begin{aligned} \frac{\partial L(\theta)}{\partial \alpha} &= \frac{\partial L(\theta)}{\partial A} \frac{\partial A}{\partial \alpha} + \frac{\partial L(\theta)}{\partial B} \frac{\partial B}{\partial \alpha} + \frac{\partial L(\theta)}{\partial C} \frac{\partial C}{\partial \alpha} \\ \frac{\partial L(\theta)}{\partial r} &= \frac{\partial L(\theta)}{\partial B} \frac{\partial B}{\partial r} \\ \frac{\partial L(\theta)}{\partial \sigma^2} &= \frac{\partial L(\theta)}{\partial C} \frac{\partial C}{\partial \sigma^2}. \end{aligned}$$

Proof for σ^2

Since A nor B depend on σ^2, C is the only relevant parameter for differentiate the estimator for σ^2. This is done similar to the logic in section 3.2 regarding the normal distribution, specifically

$$\frac{\partial L(\theta)}{\partial \sigma^2} = \Big(- \frac{n}{2C} + \frac{1}{2C^2} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)^2 \Big) \frac{1}{2\alpha}(1 - e^{-2\alpha t}) = 0. \tag{B.4}$$

Since $\alpha \neq 0$ by the definition of the Vasicek model, this provides the first linear combination (which essentially is the proof for the third term in theorem 3.2)

$$C = \frac{1}{n} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)^2 \Rightarrow \hat{\sigma}^2 = \frac{2\alpha}{n(1 - e^{-2\alpha t})} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)^2. \tag{B.5}$$

Proof for α and r

Furthermore, the parameter r does only exist within B, hence the same logic as for σ^2 can be applied. Namely

$$\frac{\partial L(\theta)}{\partial r} = \frac{(1 - e^{-\alpha t})}{C} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big) = 0.$$

Since $(1 - e^{-\alpha t}) \neq 0$ and $C \neq 0$, this provides the second linear combination

$$\sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big) = 0. \tag{B.6}$$

In order of expressing B as a function of the other variables, it follows from elaborating equation (B.6),

$$B = \frac{1}{n} \sum_{i=1}^{n} \big(R_{i+1} - AR_i\big). \tag{B.7}$$

Finally for α,

$$\begin{aligned}
\frac{\partial L(\theta)}{\partial \alpha} &= (-\frac{te^{-\alpha t}}{C}) \sum_{i=1}^{n} \Big(R_i(R_{i+1} - B) - AR_i^2\Big) \\
&+ \Big(\frac{1}{C} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)\Big) \frac{\partial B}{\partial \alpha} \\
&+ \Big(- \frac{n}{2C} + \frac{1}{2C^2} \sum_{i=1}^{n} \big(R_{i+1} - AR_i - B\big)^2 \Big) \frac{\partial C}{\partial \alpha}.
\end{aligned}$$

Although the expression for $\frac{\partial L(\theta)}{\partial \alpha}$ at first glance not provides a concrete solution, using the properties from the first two linear combinations, equation (B.4) and (B.6), and that they equal zero, in turn removes the second and third term from $\frac{\partial L(\theta)}{\partial \alpha}$ expression (also note that the term $-\frac{te^{-\alpha t}}{C} \neq 0$ and therefore does not affect the linear combination), namely

$$\sum_{i=1}^{n} \Big(R_i(R_{i+1} - B) - AR_i^2\Big) = 0. \tag{B.8}$$

Elaborating the above expression, one is provided with

$$\sum_{i=1}^{n} R_i R_{i+1} = B \sum_{i=1}^{n} R_i + A \sum_{i=1}^{n} R_i^2. \tag{B.9}$$

The linear combinations in equation (B.6) and (B.8) can then be solved in order to obtain an estimator for A and B. Start with substituting B in equation (B.9) with B from equation (B.7),

$$\sum_{i=1}^{n} R_i R_{i+1} = \Big(\frac{1}{n}\sum_{i=1}^{n}\big(R_{i+1} - AR_i\big)\Big)\sum_{i=1}^{n} R_i + A\sum_{i=1}^{n} R_i^2 \Rightarrow$$

$$\sum_{i=1}^{n} R_i R_{i+1} - \frac{1}{n}\sum_{i=1}^{n} R_{i+1}\sum_{i=1}^{n} R_i = A\Big(\sum_{i=1}^{n} R_i^2 - \frac{1}{n}\big(\sum_{i=1}^{n} R_i\big)^2\Big) \Rightarrow$$

$$A = \frac{n\sum_{i=1}^{n} R_i R_{i+1} - \sum_{i=1}^{n} R_i \cdot \sum_{i=1}^{n} R_{i+1}}{n\sum_{i=1}^{n} R_i^2 - (\sum_{i=1}^{n} R_i)^2}. \tag{B.10}$$

As stated earlier in this proof, $A = e^{-\alpha t}$. Thus, equation (B.10) can be used to express the estimator for α, namely

$$\hat{\alpha} = -\frac{1}{t}\log A = -\frac{1}{t}\log\Big(\frac{n\sum_{i=1}^{n} R_i R_{i+1} - \sum_{i=1}^{n} R_i \cdot \sum_{i=1}^{n} R_{i+1}}{n\sum_{i=1}^{n} R_i^2 - (\sum_{i=1}^{n} R_i)^2}\Big). \tag{B.11}$$

Furthermore, $B = r(1 - e^{-\alpha t})$ and then equation (B.7) can be used to express r as a function of α, namely

$$\hat{r} = \frac{B}{(1 - e^{-\alpha t})} = \frac{1}{n(1 - e^{-\alpha t})}\Big(\sum_{i=1}^{n} R_{i+1} - e^{-\alpha t}\sum_{i=1}^{n} R_i\Big). \tag{B.12}$$

Thus, equations (B.5), (B.11) and (B.12) express the full conclusion of this proof. □

Proof of theorem 3.3

Proof. The proof for theorem 3.3 follows the same initial simplification as for theorem 3.2. Furthermore by the definition of the variance for the maximum likelihood estimators, it follows that the variance of each estimator is the negative expected value of the inverse of the second derivative, as explained in theorem 3.1. Note that we derive the variance of $\hat{\sigma}$ and not of $\hat{\sigma}^2$, although we in the previous proof derived the estimator of $\hat{\sigma}^2$. However, in that case it does not matter since the maximized probability with respect to $\hat{\sigma}$ is the same as the one with respect to $\hat{\sigma}^2$. The same logic holds for the covariance calculation.

Proof for $\frac{\partial^2 L(\theta)}{\partial \sigma^2}$

The second derivative of the log-likelihood function in equation (B.3) with respect to σ^2 can then be expressed as

$$\begin{aligned}\frac{\partial^2 L(\theta)}{\partial \sigma^2} &= \frac{\partial^2 L(\theta)}{\partial C^2}\Big(\frac{\partial C}{\partial \sigma}\Big)^2 \\ &= \frac{\partial}{\partial C}\Big[-\frac{n}{2C} + \frac{1}{2C^2}\sum_{i=1}^{n}(R_{i+1} - AR_i - B)^2\Big]\Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2 \\ &= \Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2\Big(\frac{n}{2C^2} - \frac{1}{C^3}\sum_{i=1}^{n}(R_{i+1} - AR_i - B)^2\Big).\end{aligned} \tag{B.13}$$

Hence, the expected value of equation (B.13) then needs to be calculated. Since the only stochastic term in the sum is R_{i+1}, this provides the following

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial \sigma^2}\Big] = \Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2\Big(\frac{n}{2C^2} - \frac{1}{C^3}\sum_{i=1}^{n}\mathbb{E}\big[(R_{i+1} - AR_i - B)^2|R_i\big]\Big). \tag{B.14}$$

Moreover, at any time t_i the interest rate R_i is given and the expected interest rate at time t_{i+1} depends on the known interest rate R_i. Hence

$$\begin{aligned}\mathbb{E}\left[(R_{i+1} - AR_i - B)^2|R_i\right] &= \mathbb{E}\left[(R_{i+1} - (AR_i + B))^2|R_i\right] \\ &= \mathbb{E}\left[R_{i+1}^2|R_i\right] - 2(AR_i + B)\,\mathbb{E}\left[R_{i+1}|R_i\right] + (AR_i + B)^2.\end{aligned} \tag{B.15}$$

Note that $\mathbb{E}[R_{i+1}^2|R_i]$ is the same as the squared mean plus the variance of R_{i+1}. Whilst $\mathbb{E}[R_{i+1}|R_i]$ is the mean of R_{i+1}. Hence

$$\mathbb{E}\left[R_{i+1}^2|R_i\right] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) + \Big(R_i e^{-\alpha t} + r(1 - e^{-\alpha t})\Big)^2 = C + (AR_i + B)^2$$

and

$$\mathbb{E}\left[R_{i+1}|R_i\right] = R_i e^{-\alpha t} + r(1 - e^{-\alpha t}) = AR_i + B.$$

Hence, equation (B.15) gives us that $\mathbb{E}\left[(R_{i+1} - AR_i - B)^2|R_i\right] = C$. Thus, this can be replaced in equation (B.14), which provides us with

$$\begin{aligned}\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial \sigma^2}\Big] &= \Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2\Big(\frac{n}{2C^2} - \frac{1}{C^3}\sum_{i=1}^{n} C\Big) = -\frac{n}{2C^2}\Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2 \\ &= -\frac{n}{2\big(\frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})\big)^2}\Big(\frac{\sigma}{\alpha}(1 - e^{-2\alpha t})\Big)^2 = -\frac{2n}{\sigma^2}.\end{aligned} \tag{B.16}$$

Since the variance is the inverse of the negative expected value of the second derivative, equation (B.16) then gives that

$$\frac{\partial^2 L(\theta)}{\partial \sigma^2} = -\Big(-\frac{1}{\frac{2n}{\sigma^2}}\Big) = \frac{\sigma^2}{2n},$$

which concludes the first part of the proof.

Proof for $\frac{\partial^2 L(\theta)}{\partial r^2}$

Furthermore, the variance of $\hat{r}$ is calculated in the same manner, namely

$$\frac{\partial^2 L(\theta)}{\partial r^2} = \frac{\partial^2 L(\theta)}{\partial B^2}\Big(\frac{\partial B}{\partial r}\Big)^2$$

$$\frac{\partial^2 L(\theta)}{\partial r^2} = \frac{\partial}{\partial B}\Big[\frac{1}{C}\sum_{i=1}^{n}(R_{i+1} - AR_i - B)\Big](1 - e^{-\alpha t})^2 = -\frac{n}{C}(1 - e^{-\alpha t})^2. \tag{B.17}$$

Substituting C in equation (B.17) then gives us that

$$\frac{\partial^2 L(\theta)}{\partial r^2} = -\frac{2n\alpha(1 - e^{-\alpha t})^2}{\sigma^2(1 - e^{-2\alpha t})}. \tag{B.18}$$

As stated before, the variance is the inverse of the negative expected value of the second derivative and therefore equation (B.18) can be used to express the variance as

$$\frac{\partial^2 L(\theta)}{\partial r^2} = -\left(-\frac{1}{\frac{2n\alpha(1 - e^{-\alpha t})^2}{\sigma^2(1 - e^{-2\alpha t})}}\right) = \frac{\sigma^2(1 - e^{-2\alpha t})}{2n\alpha(1 - e^{-\alpha t})^2},$$

which concludes the second part of the proof.

Proof for $\frac{\partial^2 L(\theta)}{\partial \alpha^2}$

The differentiation for $\frac{\partial^2 L(\theta)}{\partial \alpha^2}$ is quite obscure and will therefore be done in sex separate steps. First of all we conclude that

$$\frac{\partial^2 L(\theta)}{\partial \alpha^2} = \frac{\partial^2 L(\theta)}{\partial A^2}\Big(\frac{\partial A}{\partial \alpha}\Big)^2 + \frac{\partial^2 L(\theta)}{\partial B^2}\Big(\frac{\partial B}{\partial \alpha}\Big)^2 + \frac{\partial^2 L(\theta)}{\partial C^2}\Big(\frac{\partial C}{\partial \alpha}\Big)^2 +$$
$$2\frac{\partial^2 L(\theta)}{\partial A \partial B}\Big(\frac{\partial A}{\partial \alpha}\Big)\Big(\frac{\partial B}{\partial \alpha}\Big) + 2\frac{\partial^2 L(\theta)}{\partial A \partial C}\Big(\frac{\partial A}{\partial \alpha}\Big)\Big(\frac{\partial C}{\partial \alpha}\Big) + 2\frac{\partial^2 L(\theta)}{\partial B \partial C}\Big(\frac{\partial B}{\partial \alpha}\Big)\Big(\frac{\partial C}{\partial \alpha}\Big). \tag{B.19}$$

Subsequently, the expected value of each of these derivatives can be calculated as follows

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial A^2}\Big(\frac{\partial A}{\partial \alpha}\Big)^2\Big] = -\frac{1}{C}\sum_{i=1}^{n}(R_i^2)(-tA)^2 \tag{B.20}$$

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial B^2}\Big(\frac{\partial B}{\partial \alpha}\Big)^2\Big] = -\frac{(rtA)^2 n}{C} \tag{B.21}$$

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial C^2}\Big(\frac{\partial C}{\partial \alpha}\Big)^2\Big] = -\frac{n}{2C^2}\Psi^2 \tag{B.22}$$

where $\Psi = \frac{\partial C}{\partial \alpha} = -\frac{\sigma^2 A^2(-2\alpha t + A^{-2} - 1)}{2\alpha^2}$

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial A \partial B}\Big(\frac{\partial A}{\partial \alpha}\Big)\Big(\frac{\partial B}{\partial \alpha}\Big)\Big] = \frac{r}{C}(tA)^2\sum_{i=1}^{n} R_i \tag{B.23}$$

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial A \partial C}\Big(\frac{\partial A}{\partial \alpha}\Big)\Big(\frac{\partial C}{\partial \alpha}\Big)\Big] = \mathbb{E}\Big[2(tA)\Psi\sum_{i=1}^{n}(AR_i^2 - R_i(R_{i+1} - B))\Big] = 0. \tag{B.24}$$

As previously shown, $\mathbb{E}[R_{i+1}|R_i] = AR_i + B$. Thus, the expected value of equation (B.24) will just be zero since $\mathbb{E}\Big[\sum_{i=1}^{n}(AR_i^2 - R_i(R_{i+1} - B))\Big] = \sum_{i=1}^{n}(AR_i^2 - R_i(AR_i + B - B) = 0$. Moreover, the next line will follow in i similar manner as we will see that

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial B \partial C}\Big(\frac{\partial B}{\partial \alpha}\Big)\Big(\frac{\partial C}{\partial \alpha}\Big)\Big] = \mathbb{E}\Big[(rtA)\Psi(-\frac{1}{C^2})\sum_{i=1}^{n}(R_{i+1} - AR_i - B)\Big] = 0. \tag{B.25}$$

Again using the properties of the expected value of R_{i+1}, the summation turns out as follows

$$\mathbb{E}\Big[\sum_{i=1}^{n}(R_{i+1} - AR_i - B)\Big] = \sum_{i=1}^{n}(AR_i + B - AR_i - B) = 0.$$

Hence,

$$\begin{aligned}\frac{\partial^2 L(\theta)}{\partial \alpha^2} &= -\Big((B.20) + (B.21) + (B.22) + 2\cdot(B.23)\Big)^{-1} \\ &= \Big(\frac{(tA)^2}{C}\sum_{i=1}^{n} R_i^2 + \frac{n(rtA)^2}{C} + \frac{n\Psi^2}{2C^2} - 2\frac{r(tA)^2}{C}\sum_{i=1}^{n} R_i\Big)^{-1} \\ &= \Big(\frac{(tA)^2}{C}(\sum_{i=1}^{n}(R_i - r)^2 + \frac{n\Psi^2}{2C(tA)^2})\Big)^{-1} \\ &= \Big(\frac{2\alpha t^2 e^{-2\alpha t}}{\sigma^2(1 - e^{-2\alpha t})}\sum_{i=1}^{n}(R_i - r)^2 + \frac{ne^{-4\alpha t}(e^{2\alpha t} - 2\alpha t - 1)^2}{2\alpha(1 - e^{-2\alpha t})^2}\Big)^{-1},\end{aligned} \tag{B.26}$$

which concludes this part of the proof.

Proof for $\frac{\partial^2 L(\theta)}{\partial r \partial \sigma}$

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial r \partial \sigma}\Big] = \mathbb{E}\Big[\frac{\partial}{\partial r}\big[\frac{\partial L}{\partial C}\frac{\partial C}{\partial \sigma}\big]\Big] = \mathbb{E}\Big[\Big(\frac{\partial^2 L}{\partial A \partial L}\frac{\partial A}{\partial r} + \frac{\partial^2 L}{\partial B \partial C}\frac{\partial B}{\partial r} + \frac{\partial^2 L}{\partial C^2}\frac{\partial C}{\partial r}\Big)\frac{\partial C}{\partial \sigma}\Big].$$

Since neither A nor C contains r, both $\frac{\partial A}{\partial r}$ and $\frac{\partial C}{\partial r}$ becomes zero. Furthermore, $\mathbb{E}\big[\frac{\partial^2 L}{\partial B \partial C}\big]$ also equals zero due to equation (B.25), hence

$$\mathbb{E}\Big[\frac{\partial^2 L(\theta)}{\partial r \partial \sigma}\Big] = 0.$$

Proof for $\frac{\partial^2 L(\theta)}{\partial r \partial \alpha}$

First, note that $\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial B \partial C}] = 0$ from equation (B.25), which will be used later in this proof. Moreover,

$$\begin{aligned}
\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial r \partial \alpha}] &= \mathbb{E}\Big[\frac{\partial}{\partial \alpha}\big[\frac{\partial L(\theta)}{\partial B}\frac{\partial B}{\partial r}\big]\Big] \\
&= \mathbb{E}\Big[\Big(\frac{\partial^2 L(\theta)}{\partial A \partial B}\frac{\partial A}{\partial \alpha} + \frac{\partial^2 L(\theta)}{\partial B^2}\frac{\partial B}{\partial \alpha} + \frac{\partial^2 L(\theta)}{\partial B \partial C}\frac{\partial C}{\partial \alpha}\Big)\frac{\partial B}{\partial r}\Big] \\
&= \Big(\frac{tA}{C}\sum_{i=1}^{n} R_i - \frac{rntA}{C}\Big)(1 - e^{-\alpha t}) = \frac{tA(1 - e^{-\alpha t})}{C}\sum_{i=1}^{n}(R_i - r) \\
&= \frac{2\alpha t e^{-\alpha t}(1 - e^{-\alpha t})}{\sigma^2(1 - e^{-2\alpha t})}\sum_{i=1}^{n}(R_i - r) = \frac{2\alpha t}{\sigma^2(1 + e^{\alpha t})}\sum_{i=1}^{n}(R_i - r).
\end{aligned} \tag{B.27}$$

Since $I(\hat{\alpha}, \hat{r}, \hat{\sigma})$ contains of the negative seconds derivatives, we only need to take the negative value of equation (B.27), which concludes this part of the proof.

Proof for $\frac{\partial^2 L(\theta)}{\partial \alpha \partial \sigma}$

First note that $\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial B \partial C}] = 0$ and that $\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial A \partial C}] = 0$ from equation (B.24) and (B.25), which will be used later. Also note that this is the derivative with respect to σ and not σ^2. Hence,

$$\begin{aligned}
\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial \alpha \partial \sigma}] &= \mathbb{E}\Big[\frac{\partial}{\partial \alpha}\big[\frac{\partial L(\theta)}{\partial C}\frac{\partial C}{\partial \sigma}\big]\Big] \\
&= \mathbb{E}\Big[\Big(\frac{\partial^2 L(\theta)}{\partial A \partial C}\frac{\partial A}{\partial \alpha} + \frac{\partial^2 L(\theta)}{\partial B \partial C}\frac{\partial B}{\partial \alpha} + \frac{\partial^2 L(\theta)}{\partial C^2}\frac{\partial C}{\partial \alpha}\Big)\frac{\partial C}{\partial \sigma}\Big].
\end{aligned} \tag{B.28}$$

Now, also note from equations (B.13), (B.15) and (B.16) that $\mathbb{E}[\frac{\partial^2 L(\theta)}{\partial C^2}] = -\frac{n}{2C^2}$. Hence (B.28) equals (where Ψ is defined as in equation (B.22))

$$\begin{aligned}
-\frac{n}{2C^2}\Psi\frac{\sigma(1 - e^{-2\alpha t})}{\alpha} &= -\frac{n}{2\big(\frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})\big)^2}\Big(-\frac{\sigma^2 e^{-2\alpha t}(e^{2\alpha t} - 2\alpha t - 1)}{2\alpha^2}\Big)\frac{\sigma(1 - e^{-2\alpha t})}{\alpha} \\
&= \frac{n}{\sigma\alpha}\frac{1 - e^{-2\alpha t}(2\alpha t + 1)}{1 - e^{-2\alpha t}}.
\end{aligned} \tag{B.29}$$

As stated in the previous section, the matrix contains the negative second derivative, thus this completes the final part of the proof. ☐

Appendix C - Further results

Yield curve estimating

Here we provide estimation data for eight arbitrarily chosen time series with time spans between five and 25 years. Note that in the cases when the unbiased estimation of α and the estimation of r are not both positive or negative, we also provide an estimated yield curve using the biased estimator of α. Each yield estimation will follow without further explanation about its time series. Figure 16 also shows a final comparison of nine different time series (note that we only present estimation data for the time series of length 19, 20 and 21 years, not the once in between).

Table 8: Monthly 25 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.1593	0.0589	(0.0438, 0.2747)
α (unbiased)	-0.00074		
r	0.0022	0.0136	(-0.0245, 0.0288)
σ	0.0074	0.00030	(0.0068, 0.0080)

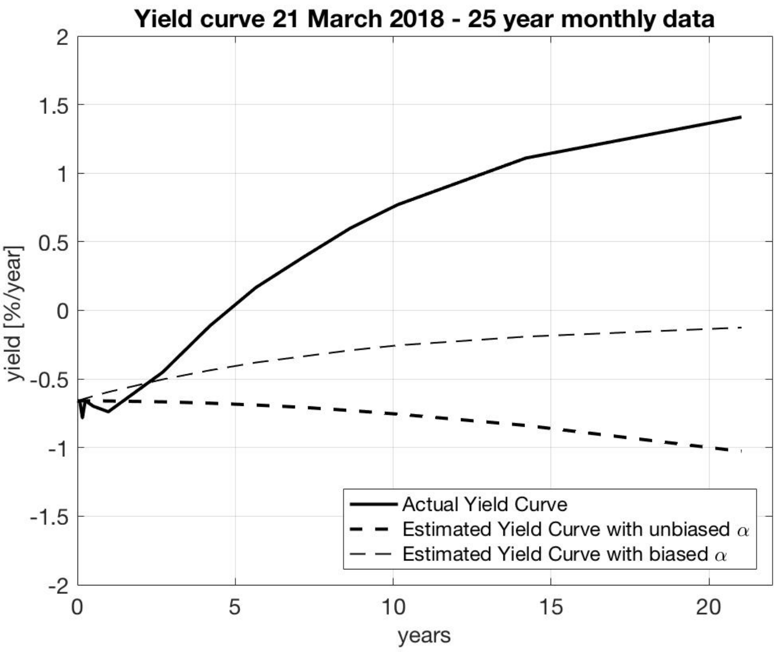

Figure 8: Estimated yield curve based on 25 years of monthly data.

Table 9: Monthly 22 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.1988	0.0761	(0.0496, 0.3481)
α (unbiased)	0.0169		
r	0.0035	0.0097	(-0.0156, 0.0226)
σ	0.0063	0.00027	(0.0057, 0.0068)

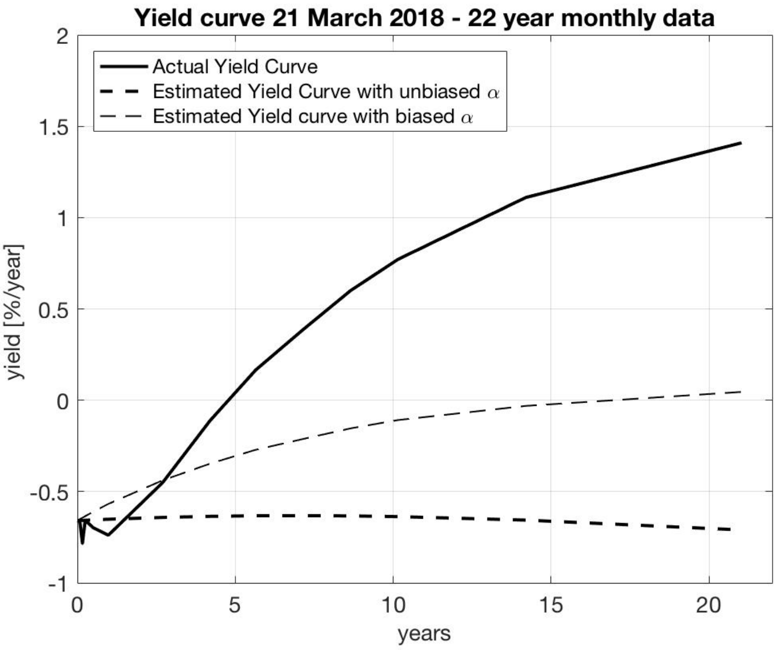

Figure 9: Estimated yield curve based on 22 years of monthly data.

Table 10: Monthly 21 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.0385	0.0777	(-0.1137, 0.1907)
α (unbiased)	-0.1508		
r	-0.0394	0.1253	(-0.2850, 0.2062)
σ	0.0059	0.00026	(0.0054, 0.0064)

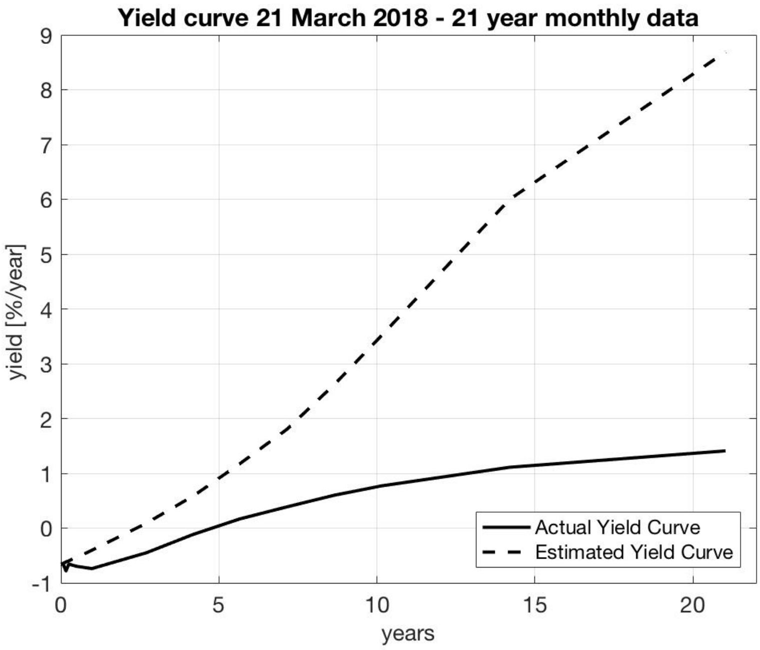

Figure 10: Estimated yield curve based on 21 years of monthly data.

Table 11: Monthly 19 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.0180	0.0823	(-0.1434, 0.1794)
α (unbiased)	-0.1908		
r	-0.0931	0.5140	(-1.1004, 0.9143)
σ	0.0058	0.00027	(0.0052, 0.0063)

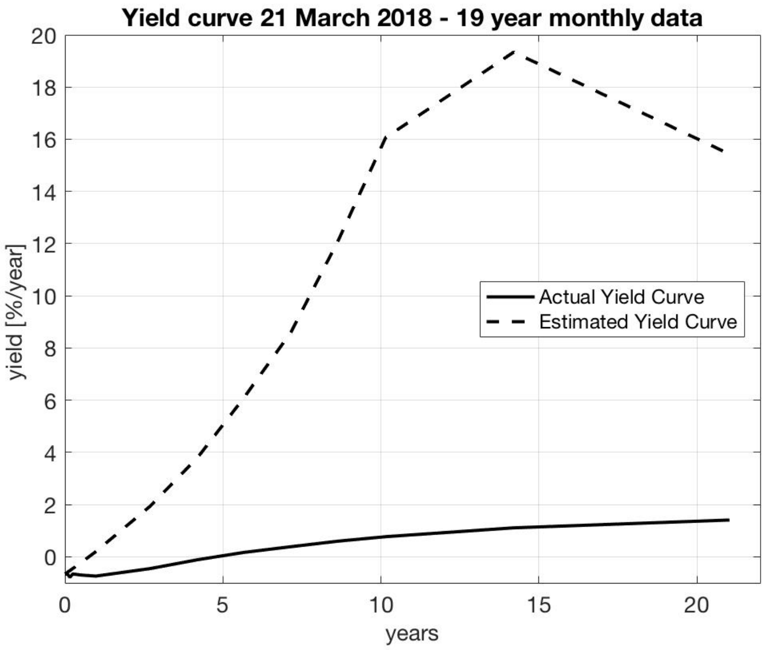

Figure 11: Estimated yield curve based on 19 years of monthly data.

Table 12: Weekly 20 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.0539	0.0718	(-0.0869, 0.1946)
α (unbiased)	-0.1459		
r	-0.0292	0.0681	(-0.1627, 0.1044)
σ	0.0053	0.00012	(0.0053, 0.0055)

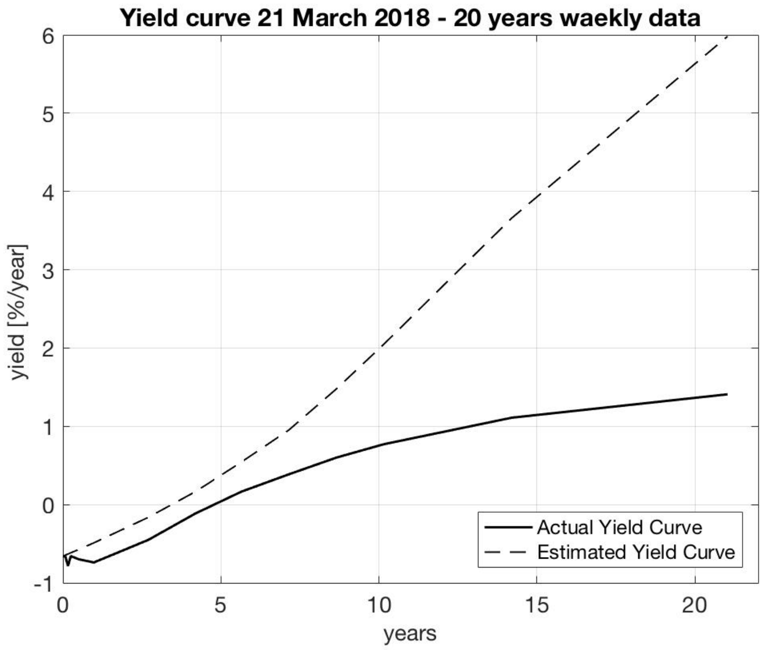

Figure 12: Estimated yield curve based on 20 years of weakly data.

Table 13: Weekly 15 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.0713	0.0957	(-0.1163, 0.2590)
α (unbiased)	-0.1950		
r	-0.0295	0.0690	(-0.1647, 0.1058)
σ	0.0061	0.00013	(0.0058, 0.0064)

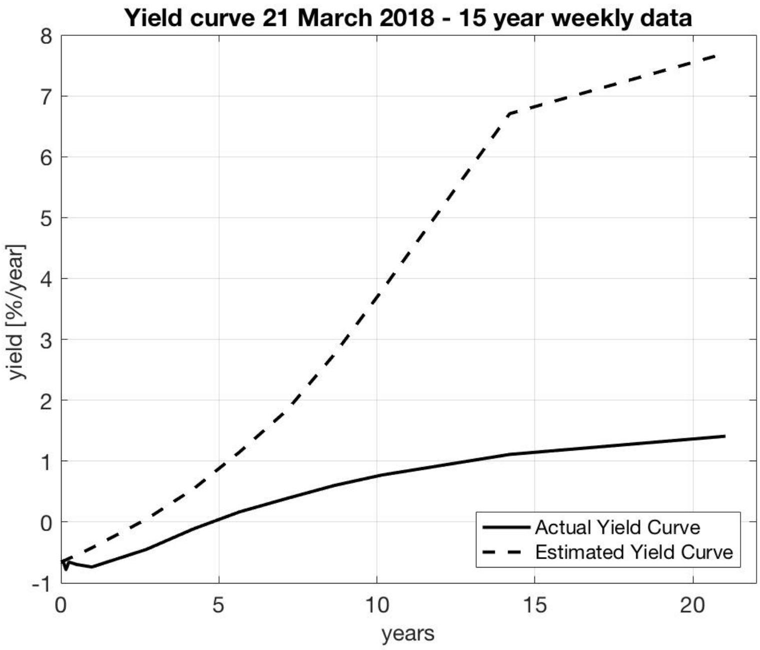

Figure 13: Estimated yield curve based on 15 years of weekly data.

Table 14: Weekly 10 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.4488	0.1555	(0.1439, 0.7536)
α (unbiased)	0.0486		
r	0.0039	0.0057	(-0.0150, 0.0072)
σ	0.0061	0.00019	(0.0057, 0.0065)

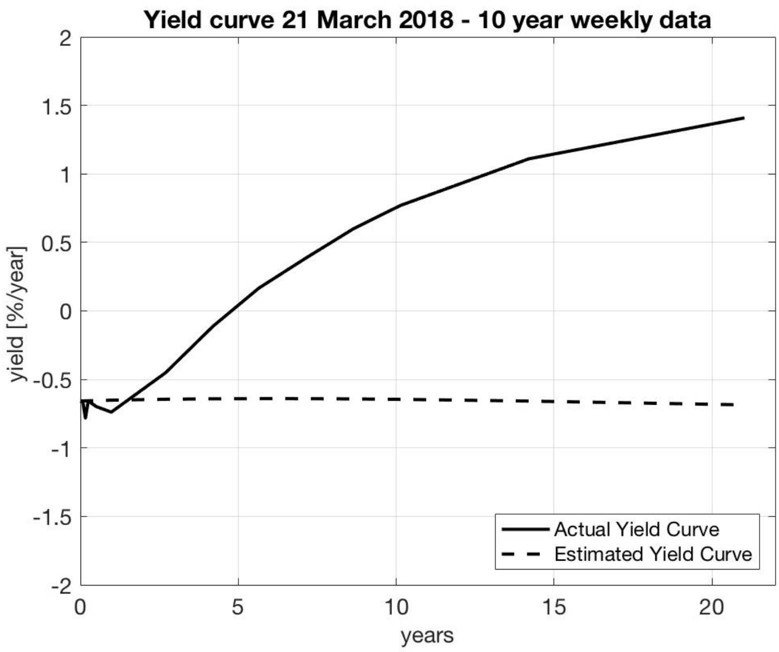

Figure 14: Estimated yield curve based on 10 years of weekly data.

Table 15: Weekly 5 years

Parameter	Estimation	Standard error	95% Confidence interval
α (biased)	0.4100	0.2876	(-0.1537, 0.9737)
α (unbiased)	-0.3870		
r	-0.0091	0.0072	(-0.0232, 0.0051)
σ	0.0042	0.00018	(0.0038, 0.0045)

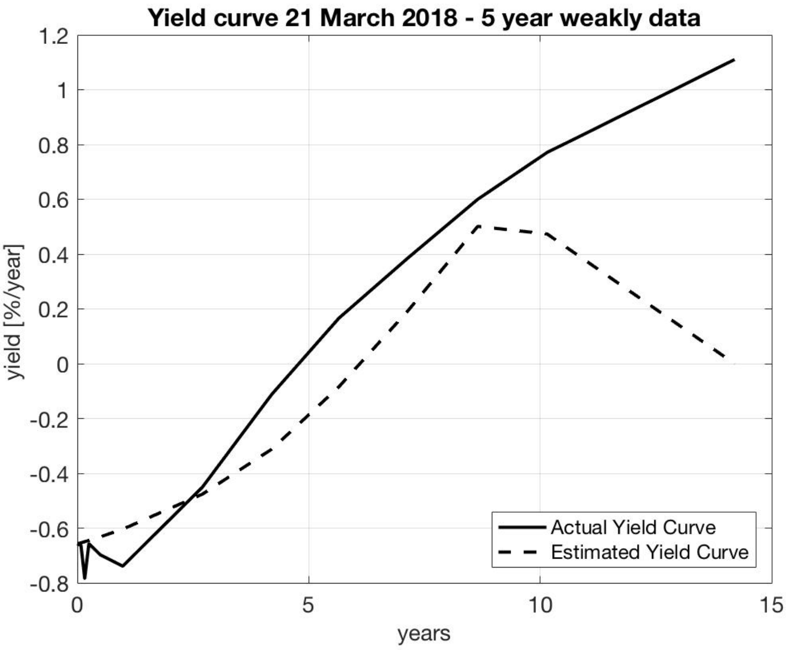

Figure 15: Estimated yield curve based on 5 years of weakly data.

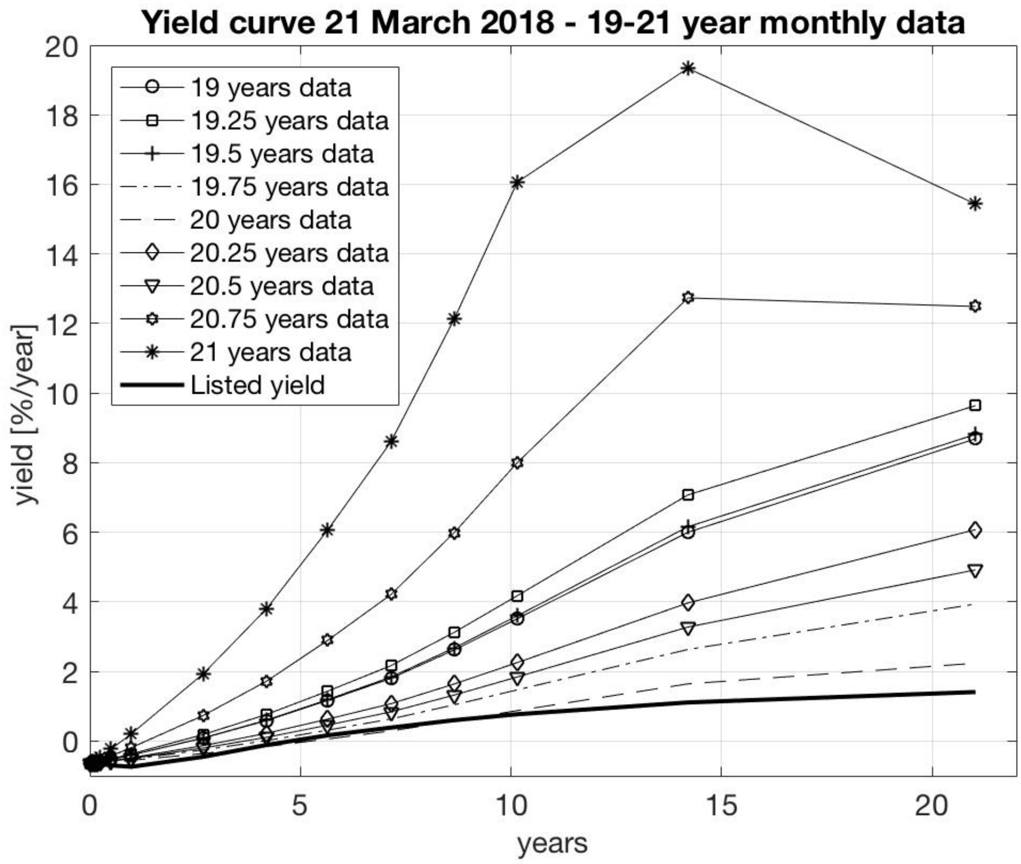

Figure 16: Comparison of estimated yield curves to actual yield curve

Monte Carlo simulation

The confidence interval that was calculated in figure 6 was created according to

$$\Pi_{\text{C/F for 95\%}} = \bar{\Pi} \pm \frac{\sigma}{\sqrt{n}}. \tag{C.1}$$

This follows from the CLT assumption of the MC simulations sample being normally distributed. To illustrate this fact, we performed a QQ-plot with respect to the normal distribution. This could for obvious reasons only be performed upon one derivative type at the time with a fixed strike value K. Figure 17 illustrates how a QQ-plot for a cap could look like (the exact same logic holds for floors).

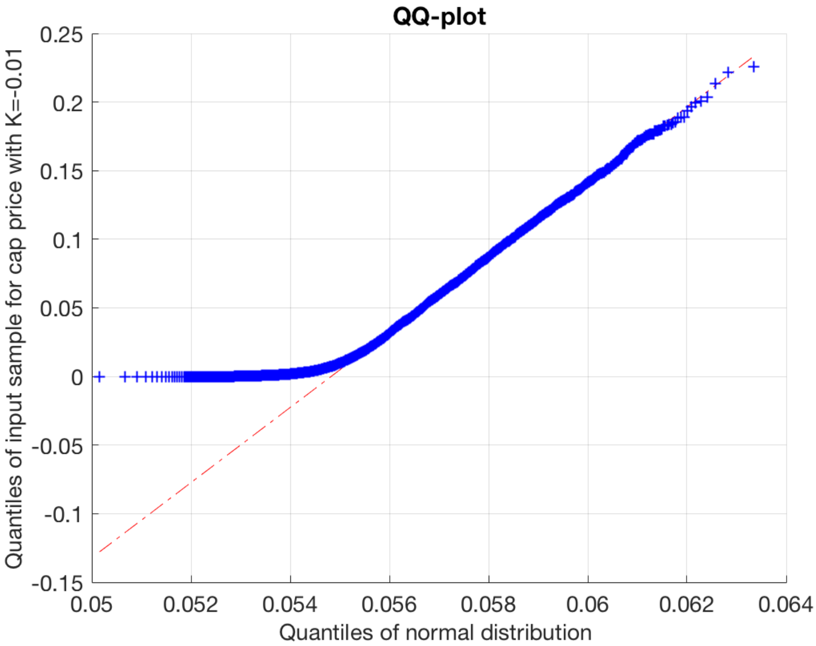

Figure 17: QQ-plot for MC simulation sample.

At first, the skewness from Figure 17 together with the commonly known theory regarding QQ-plot indicates for the educated reader that this needs some modification before considered normally distributed. However, considering the formulas used in equation (29), one clearly notice that the skewness is derived from the fact that negative values are considered as zero. Hence, it follows from this that the skewness is heavely correlated to the value of K. For Figure 17, the values from table 7 was used with the exception of K being fixed to -0.01. Following the logic from section 4.4, the price of a cap increases with a diminishing value of K, and with similar logic the price of a floors increases with a increasing value of K. Hence, it also follows that the values of K providing the highest price for either caps or floors, will provide less skewness in the QQ-plot. It is also these values that are most relevant for discussion and hence a confidence interval accordingly to equation (C.1) is reasonable in approximation.

Appendix D - Conclusion

One-factor models

In the following models, the short rate is the stochastic factor which will determine all the future interest rates. Since it is the short rate that implies the uncertainty, the following expression is used to describe it generally [8].

$$dr(t) = \mu r(t)dt + \sigma r(t)dW.$$

Where σ is the standard deviation and μ the instantaneous drift rate of r, the short rate. This functions are both functions of r but they are assumed to be independent of time. In a one factor model it is assumed that the interest rates all have the same direction [8]. These models use a Gaussian interpretation, since it gives solutions that are numerical and is straightforward when used to price instruments. But since Gaussian models allows for negative interest rates in specific cases there are opinions that these models does not represent the reality and should therefore not be used. But the models remain popular just for this specific reason, that they are able to handle negative interest rates [8].

The Dothan model

Published in 1978 [7], Dothan presented his model with the background in driftless geometric Brownian motion, implying that the average does not change as the starting point shifts [34]. Under the objective probability measure Q_0, the interest rate can be described as

$$dR(t) = \sigma R(t)dW^0(t), \quad R(0) = R_0.$$

R_0 and σ are both positive constants. Furthermore, Dothan presented a constant risk price. This in turns is equivalent to assuming risk-neutral dynamics according to

$$dR(t) = aR(t)dt + \sigma R(t)dW(t), \quad a = \text{constant}.$$

Without delving deeper into the characteristics of the model, Dothan possess the ability to only simulate positive R_t. This is due to the fact that R is lognormal distributed [4].

The Cox, Ingersoll and Ross (CIR) model

Published in 1985, the CIR model, under the condition of the risk-neutral measure Q, can be described as

$$dR(t) = k(\theta - R(t))dt + \sigma\sqrt{R(t)}dW(t), \quad R(0) = R_0,$$

with R_0, k, θ, σ all being positive constants. Subsequently, putting the CIR model in context towards the Vasicek model, one will notice that CIR allows for analytical tractability and similar to the Dothan model, also an instantaneous short rate with exclusively positive values [4]. A similarity with Vasicek, CIR and Hull-White model is that they are all models with mean reversion. The risk-free price for a ZCB in the CIR model is given by

$$B(t,T) = C(t,T)e^{-A(t,T)R(t)}$$

where

$$C(t,T) = \frac{2(e^{z(T-t)} - 1)}{(z+a)(e^{z(T-t)} - 1) + 2z}$$

$$A(t,T) = \Big[\frac{2ze^{(z+a)(T-t)/2}}{(z+a)(e^{\lambda(T-t)} - 1) + 2z}\Big]^{2ab/\sigma^2}$$

and

$$z = \sqrt{a^2 + 2\sigma^2} \quad [8].$$

The Hull-White / Extended Vasicek Model

Noticing the flaws and improvement areas, Hull and White published an extended version of the Vasicek model in 1990 with the main attribute being a time-varying parameter as follows

$$dR(t) = (\theta(t) - \alpha R(t))dt + \sigma dW(t).$$

With α and σ as positive constant and the mean reversion level $\theta(t)/\alpha$ [8]. The time dependent parameter $\theta(t)$ is chosen to fit the term structure of interest rates in the real market [4]. Under certain conditions, the Hull-White model no longer remains Hull-White

$$\text{Hull-White} = \begin{cases} \text{Simplified Hull-White model if } \alpha \neq 0 \\ \text{Vacisek model if } \alpha \neq 0 \text{ and } \theta(t) = r \cdot \alpha \text{ [8].} \end{cases}$$

A disadvantage with the Hull-White model is that it is possible for extreme negative values, but the probability for this is very small [11].

Multi-factor models

Since the one-factor models has some drawbacks it is sometimes more beneficial to use multi-factor models and more specified, two-factor models since they allow for a more realistic correlation pattern. These models are in general more useful when the correlation is very important or when one is seeking a higher precision [4]. If we were to use the G2 model instead of Vasicek, we obtain the following equation for the bond price $B(\tau)$,

$$B(\tau) = A(\tau)\text{exp}(-C^x(\tau)x_t - B^y(\tau)y_t).$$

Where $C(\tau)$ and $A(\tau)$ is the same as in theorem 2.2. The two-factor model can easily be extended to contain three or more factors but an arising problem is here how to decide the number of factors so the implementation will remain practical. The trade-off what has to be made is whether it is more desired to represent realistic correlation patterns or maintain a more numerically-efficient implementation [4].

HJM Framework

The Heath, Jarrow and Morton framework was developed from the single factor short rate models drawbacks. It is similar to the Hoo-Lee model, which models the yield in a binomial tree, but instead of being discrete, Heath, Jarrow and Morton translated the Ho-Lee model into continuous time. One significant difference between the HJM model and one-factor short rate models is that the stochastic evolution of the yield curve is obtained by deriving an arbitrage-free framework with continuously compounded interest rates at time T , compared to simply choose the forward rate $f(t, T)$ [8].The bond price $B(t, T)$ is provided by following equation

$$f(t,T) = -\frac{\partial ln B(t,T)}{\partial T}$$

$$B(t,T) = e^{-\int_t^T f(t,s)ds}.$$

The conclusions from the HJM framework highligts that the short rate is not Markov. The short rate process therefore depends on the path between t and today for the short rate including the value at time t. To represent the structure movements, Monte Carlo simulation have to be used which gives a binomial nonrecombining tree which quickly gets extremely large (with 2^n nodes results in roughly 1 billion when $n = 30$). Depending on how the instantaneous standard deviation, $s(t, T, \Omega)$ is chosen the HJM framework becomes other interest rate models [4].

$$s(t,T,\Omega) = \begin{cases} \text{constant}, \sigma & \text{leads to} & \text{Ho-Lee Model} \\ \sigma e^{-a(T-t)} & \text{leads to} & \text{Hull-White Model} \end{cases}$$

Both Ho-Lee Model and Hull-White Model are particular Markov examples of HJM [16].

Appendix E - MATLAB code

MLE for Vasicek

```
function [parameters,expectedinfomatrix,lvalue] = Mle_Vasicek(data,years)
  disp('alpha,r,sigma') %shows the order of the parameters
  n = length(data)-1;
  %'n' corresponds to the amount of changes, since this is discrete.
  %Note that we always have n+1 points in our sample.
  dt = years/(n+1); % size of 'each step'.
  parameters = zeros(1,3);
  expectedinfomatrix = zeros(3,3);
  S0 = 0; % inital for sum of all R(t(i-1))
  S1 = 0; % inital for sum of all R(t(i))
  S00 = 0; % inital for sum of all R(t(i-1))^
  S01 = 0; %inital for sum of all R(t(i-1))*R(t(i))
  for i = 1:n %calculates every sum
    S0 = S0+data(i);
    S1 = S1+data(i+1);
    S00 = S00+data(i)*data(i);
    S01 = S01+data(i)*data(i+1);
  end
  ahat = -(1/dt)*log((n*S01-S0*S1)/(n*S00-S0^2)); %adresses a value to alpha-hat
  rhat = 1/(n*(1-exp(-ahat*dt)))*(S1-exp(-ahat*dt)*S0);
  %adresses a value to r-hat
  sigmapart = 0; %inital for the sum in our estimator of sigma
  for j = 1:n %calculates the sum in our estimator of sigma
     sigmapart = sigmapart+(data(j+1)-data(j) *exp(-ahat*dt)-rhat *(1-exp(-ahat*dt)))
  end
  sigma2 = 2*ahat*sigmapart/(n*(1-exp(-2 *ahat*dt)));
  %adresses a value to sigma2-hat
  parameters(1) = ahat;
  parameters(2) = rhat;
  parameters(3) = sqrt(sigma2);
  A = exp(-ahat*dt); %to simplify (see proof for thm 3.2 and 3.3)
  C = sigma2/(2*ahat)*(1-exp(-2*ahat*dt)); %to simplify (see proof)
  dCdalpha = -(sigma2*A^2*(-2 *ahat *dt+exp(2*ahat*dt)-1))/(2 *ahat^);
  %derivative to simplify (see proof)
  d2La2 = -1/C*S00 *dt^2*A^; %derivative to simplify see proof)
  d2Lb2 = -n/C*(rhat*dt*A)^; %derivative to simplify (see proof)
  2Lc2 = -dCdalpha^{*n}/(2*C^2); %derivative to simplify (see proof)
  d2Lab = rhat*(dt*A)^2*S0/C; %derivative to simplify (see proof)
  2Lalpha2 = -(d2La2+d2Lb2+d2Lc2+2*d2Lab);
  %complete second order derivative (see proof)
  d2Lr2 = (2*n*ahat*(1-exp(-ahat*dt))^2)/(sigma2 *(1-exp(-2 *ahat*dt)));
  %complete secondorder derivative (see proof)
  d2Lsigma22 = (2*n)/sigma2; %complete second order derivative (see proof)
  %the following seven lines completes the negative expected informationmatrix
  expectedinfomatrix(1,1) = d2Lalpha2;
  expectedinfomatrix(2,2) = d2Lr2;
  expectedinfomatrix(3,3) = d2Lsigma22;
  expectedinfomatrix(2,1) = -2*ahat*dt*(S0-n*rhat)/(sigma2*(1+exp(ahat*dt)));
  expectedinfomatrix(1,2) = expectedinfomatrix(2,1);
  expectedinfomatrix(3,1) = -n/(sqrt(sigma2)*ahat)*(1-exp(-2*ahat*dt)*(2*ahat*dt+1))/
  (1-exp(-2*ahat*dt));
  expectedinfomatrix(1,3) = expectedinfomatrix(3,1);
```

```
  lvalue = -n/2*log(sigma2/(2*ahat)*(1-exp(-2 *ahat*dt)))-n/2*log(2*pi)-
  (sigma2/ahat*(1-exp(-2*ahat*dt)))^(-1)*sigmapart;
  %adresses a value to the log-likelihood based on the estimators
end
```

Price with Vacisek

```
function price = vasicekPrice(alpha,r,sigma,R0,maturity,decimal,timeNow,coupon)
  %Note that
  C = @(T,t) (1-exp(-alpha*(T-t)))/alpha; %part of the pricing function
  A = @(T,t) r*(T-t)-r*(1-exp(-alpha*(T-t)))/alpha-sigma^2*(1-exp(-2*alpha*(T-t)))/
  (4*alpha^3)+sigma^2*(1-exp(-alpha*(T-t)))alpha^3-sigma^2*(T-t)/(2*alpha^2);
  %part of the pricing function
  if maturity > 1
    %if the maturity > 1, the payoff will be 1+coupon, else 1.
    q1=coupon;
  else
    q1=0; end
    t=timeNow; %adresses timeNow-value, this is always set to zero
  if decimal > 0
    price=q1*exp(-(R0*C(decimal,t)+A(decimal,t)));
     %price for the first payoff.
  else
    price=0;
  end
  for i=t+1:maturity-1
    price=price+q1 *exp(-(R0*C(i+decimal,t)+A(i+decimal,t)));
  end
  price=price+(1+q1)*exp(-(R0*C(maturity+decimal,t)+A(maturity+decimal,t)));
  end
```

Yield with Vacisek

```
function yield = yieldII(price,integer,decimal,timeNow,coupon)
  %Note that this code only works for two decimals, as this was the case
  %in our project, i.e.  it can calculate the yield of a bond that matures
  %in 21.54 years but not one that matures in 21.541 years (however, the
  %difference is almost none).
  %Also note that integer + decimal = maturity

  t = timeNow; %usually set to zero
  if decimal > 0
  %if there is a decimal, the forthcomming matrix will be long, i.e.  5
  %elements turns to 50 if the decimal
  %is in the order of ten and 500 if the decimal is in the order of a hundred
    if floor(10*(integer+decimal)) == 10*(integer+decimal)
    %checks decimal order
      matrix = zeros(1,10*(integer-t+decimal)+1);
      %if true, creates a matrix of the order of 10*(integer+decimal)+1
      for i = 1:integer %adresses each value in the discounted payoff-matrix
        matrix(length(matrix)-(10*decimal+1*(i-1))) = coupon;
      end
    else
      matrix = zeros(1,100*(integer-t+decimal)+1);
      %if the last if-statement is not true,creates a matrix of the order
```

```
        %of 100*(integer+decimal)+1
        for i = 1:integer
          matrix(length(matrix)-(100*decimal+100*(i-1))) = coupon;
          %adresses each value
        end
      end
    else
      matrix = zeros(1,integer-t+1);
      %if the first if-statement is not true, this creates a simple
      %matrix of the order integer+1 (this is the most simple case)
      for i = 2:length(matrix)-1
        matrix(i) = coupon; %adresses each value
      end
    end
    matrix(length(matrix)) = -price;
    %sets price at t=0 (last element is the first element in
    %MATLABs roots-function
    if integer > 1 %if integer > 1 the payoff at maturity will be 1+coupon,
    %else it will just be 1 (this is regarded in the price aswell)
      matrix(1) = 1+coupon;
    else
      matrix(1)=1;
    end
    if decimal > 0
    %solves discounted payoff-matrix based on the
    order of decimals
      if floor(10*decimal) == 10*decimal
      %checks order of decimal (ten or hundred)
        exp10yield = roots(matrix);
        %solves the matrix (i.e.  polynomial) set to zero
        for j = 1:length(exp10yield)
          if isreal(exp10yield(j)) == 1  exp10yield(j) > 0
          %since roots returns a full matrix containing both negative
          %roots and
          yield = -10*log(exp10yield(j));
          %ten since we extended each element by the order of ten,
          %for example x^0.5 -> x^5
          end
        end
      else
        exp100yield = roots(matrix);
        for j = 1:length(exp100yield)
          if isreal(exp100yield(j)) == 1  exp100yield(j) > 0
            yield = -100*log(exp100yield(j));
          end
        end
      end
    else
      expyield = roots(matrix);
      for j = 1:length(expyield)
        if isreal(expyield(j)) == 1  expyield(j) > 0
          yield = -log(expyield(j));
        end
      end
    end
  end
```

capfloorPrice

```
function [priceVectorCap,priceVectorFloor] = capfloorPrice(Matrix,p,K,dt)

dim=size(Matrix);
nrTrials=dim(2);
%Obtaining the number of columns in the matrix
lengther=dim(1);
%Obtaining the number of rows in the matrix
priceVectorCap=zeros(nrTrials,1); %vector for the price of caps
priceVectorFloor=zeros(nrTrials,1); %vector for the price of floors
discounter=0;
tenorGap=p/dt;%Declaring the size of each tenor

for i=1:nrTrials
   for j=tenorGap:lengther
      for k=1:tenorGap
      %Creating three foor-loops that will be used to calculate multiple caps/floors.
      % Each row in the column will act as a cap/floor
         discounter=discounter+Matrix(j-k+1,1)*dt;
         %Discounter that will change with respect to the number of steps
      end
      checkValue=Matrix(j,i);
      priceVectorCap(i)=priceVectorCap(i)+p*max(checkValue-K,0)*exp(-discounter);
      priceVectorFloor(i)=priceVectorFloor(i)+p*max(K-checkValue,0)*exp(-discounter);
      %Calculates the prices of the caplets and the floorlets as in equation
      (11) and (12)

   end
   discounter=0;
end

end
```

MCsimulateVasicek

```
function [interestRate]=MCsimulateVasicek(a,r,sigma,dt,T,Rt0)

nrTrials=T/dt;
interestRate=zeros(T,1);
interestRate(1)=Rt0;
%Placing the initial spot rate into the vector interetRate
for i=2:nrTrials %Simulating the spot rate up to nrTrials number of times.
   interestRate(i)=interestRate(i-1)+a*(r-interestRate(i-1))*dt+sigma*sqrt(dt)*randn;
   %fills the vector interestRates with the spot rate
end
end
```

strikeShifter

```
function [finalCapVector,finalFloorVector,xvalues,finalCapPriceUpperCI,
finalCapPriceLowerCI,finalFloorPriceUpperCI,finalFloorPriceLowerCI]=
strikeShifter(a,r,sigma,dt,p,T,Rt0,nrTrials,Ktrials)
%Simulates different cap and floors prices depending upon strike level

finalCapVector=zeros(Ktrials,1);
finalFloorVector=zeros(Ktrials,1);

Z=1.96; %confidence interval of 95%

finalCapPriceUpperCI=zeros(Ktrials,1);
finalCapPriceLowerCI=zeros(Ktrials,1);

finalFloorPriceUpperCI=zeros(Ktrials,1);
finalFloorPriceLowerCI=zeros(Ktrials,1);

for i=1:Ktrials
  K=-0.05+(i-1)/(1000);
  %Makes shifts in striek value   xvalues(i)=K;
  matrix=simulateInterestRate(a,r,sigma,dt,T,Rt0,nrTrials);
  %Provides simulations of the interest rate
  [capPriceVector,floorPriceVector]=capfloorPrice(matrix,p,K,dt);
  %Calculates the price according to simulation

  finalCapPrice=mean(capPriceVector);
  finalFloorPrice=mean(floorPriceVector);

  finalCapVector(i)=finalCapPrice;
  finalFloorVector(i)=finalFloorPrice;

  sqrtN=sqrt(length(capPriceVector));
  stdevCPV=std(capPriceVector);
  stdevFPV=std(floorPriceVector);

  %Calculates CI finalCapPriceUpperCI(i)=finalCapPrice+Z*stdevCPV/sqrtN;
  finalCapPriceLowerCI(i)=finalCapPrice-Z*stdevCPV/sqrtN;

  finalFloorPriceUpperCI(i)=finalFloorPrice+Z*stdevFPV/sqrtN;
  finalFloorPriceLowerCI(i)=finalFloorPrice-Z*stdevFPV/sqrtN;
end
```

simulateInterestRate

```
function simulateMatrix=simulateInterestRate(a,r,sigma,dt,T,Rt0,nrTrials)

for i=1:nrTrials
  rng(i);
  simulateMatrix(:,i)=MCsimulateVasicek(a,r,sigma,dt,T,Rt0);
  %simulates the interest rate with Vasicek model end

end
```

Test for size of n

```
a=-0.1358;
r=-0.0218;
sigma=0.0059;
dt=1/240;
p=3/12;
T=5;
Rt0=-0.0066;
K=-0.01;
ticker=0;

for i=25:25:12000
   %Simulates for different sample size, to determine a suitable level
   nrTrials=i;
   ticker=ticker+1;
   matrix=simulateInterestRate(a,r,sigma,dt,T,Rt0,nrTrials);
   [capPriceVector,floorPriceVector]=capfloorPrice(matrix,p,K,dt);
   finalCapPrice(ticker)=mean(capPriceVector);
   finalFloorPrice(ticker)=mean(floorPriceVector);
   xvector(ticker)=i;
   %Saves the simulated values
end
```

Run program - MC simulation for different K

```
a=-0.1358;
r=-0.0218;
sigma=0.0059;
dt=1/240;
p=3/12;
T=5;
Rt0=-0.0066;
nrTrials=5000;
Ktrials=100;
%The following section plots different strike values

[finalCapVector,finalFloorVector,xvalues,finalCapPriceUpperCI,finalCapPriceLowerCI,
finalFloorPriceUpperCI,finalFloorPriceLowerCI]
=strikeShifter(a,r,sigma,dt,p,T,Rt0,nrTrials,Ktrials);
plot(xvalues,finalCapVector,'black','LineWidth',0.5)
hold on
plot(xvalues,finalFloorVector,':','LineWidth',1,'color','black')
hold on
plot(xvalues,finalCapPriceUpperCI,'-','LineWidth',0.25,'color','black')
hold on
plot(xvalues,finalFloorPriceUpperCI,'-','LineWidth',0.25,'color','black')
hold on
plot(xvalues,finalCapPriceLowerCI,'-','LineWidth',0.25,'color','black')
hold on
plot(xvalues,finalFloorPriceLowerCI,'-','LineWidth',0.25,'color','black')
hold on

title('Cap/floor price with respect to K in Monte Carlo simulation');
legend('Cap price','Floor price','95% CI','location','north')
set(gcf,'color','w');
set(gca,'FontSize',14);
```

```
ylabel('Price for cap/floor','FontSize',14);
xlabel('Cap/floor rate K','FontSize',14);
grid on
```

Appendix F - Sectioning

Section	Writers
Populärvetenskaplig presentation	All
Abstract	All
Sammanfattning	All
1 Introduction	All
Backround	Oscar
Purpose	All
Outline	Axel, Fredrik
2 Theory	All
The bond market	Axel, Fredrik
Fair bond pricing	Justin, Oscar
Yield to maturity	Axel, Oscar
Forward rate	Fredrik, Justin
Interest rate caps and floors	Axel, Justin
The Vasicek model	Justin
3 Method	All
Maximum likelihood estimation	Axel, Oscar, Justin
MLE on the Normal distribution	Axel, Oscar
MLE on the Vasicek model	Fredrik, Justin
Monte Carlo simulation for caps and floors	Fredrik
4 Results	All
Data selection	Axel, Oscar
Parameter estimation	Axel, Fredrik, Justin
Yield curve modeling	Axel, Justin, Oscar
Pricing caps and floors	Fredrik, Justin
5 Conclusion	All
Appendix A-Theory	All
Appendix B-Method	Fredrik, Justin
Appendix C-Result	Fredrik, Justin
Appendix D-Conclusion	Oscar
Appendix E-MATLAB-code	All
Appendix F-Sectioning	All
Appendix G-Glossary	Oscar

Appendix G - Glossary

Arbitrage - To buy and sell an asset and make a profit from a price difference.

Bond - A type of fixed income investment where the investor lends money to, for example the government, for a fixed interest rate over a defined period of time.

Cap rate - The measure level used to determine if a caplet/floorlet is in the money.

Coupon - The interest expressed as percentage of the principal value of a bond.

Discounted notional value - The total value of an asset discounted to the value of today.

European call options - If the owner of the call option with strike price \$30 has the opportunity to buy the underlying asset for \$30 at maturity. If the value of the underlying asset is \$50 the owner can buy the asset and exercise the option and therefore make a profit of \$20 per share.

European options - A financial contract that gives the owner the possibility at maturity to exercise the option if the price of the underlying asset is in the money.

European put options - If the owner of the put option with strike price \$30 has the opportunity to sell the underlying asset for \$30 at maturity. If the value of the underlying asset is \$20 the owner can sell the asset and exercise the option and therefore make a profit of \$10 per share.

Fair price - The price which both parties, both buyer and seller agrees up on. The market value often represents the fair price.

Fixed income security - Investment that pays fixed and predetermined periodic returns and the principal at maturity.

Forward rate - The interest rate connected to a financial transaction in the future.

Government bond - A debt security issued by a government.

Interest rate caplets - Is a European style call option that enables hedging towards higher interest rates. If an investor buys a caplet he gets paid if the interest increases more than a determined strike price.

Interest rate floorlets - Is a European style put option that enables hedging towards lower interest rates. If an investor buys a floorlet he gets paid if the interest decreases more than a determined strike price.

Internal rate of return - The discount rate that equals the net present value of all future cash flows to zero.

Maximum likelihood method (MLM) - Is a method to estimate the parameters in a statistical model when the observations is known. The estimation seeks the values for the maximum of the likelihood function.

Mean reversion - Is a theory stating that returns and prices will move towards the mean or average eventually.

Net present value - The difference between the present cash inflows and outflows over a time period.

Nominal bond - The payments of this bond is of fixed amount instead of a fixed real value.

Notional Value - The total value of an asset.

Over the counter - When a security or asset gets traded in less formal exchanges as New York Stock Exchange.

Real bond - A bond which has inscurance against the inflation.

Risk aversion - When an investor chooses to invest in a more predictable payoff with lower risk rather than a possibly higher payoff with a higher grade of uncertainty.

Securities - A tradeable financial asset of various kind. The most common types are bonds, stocks or options.

Skewness - A measurement of how assymetrical an probability distribution is of an random variable around its mean.

Speed of reversion - How fast the prices move towards mean returns and prices, see Mean reversion.

Spot rate - The price of a settlement on a security.

Spread - Difference between the asking price and the bidding price of an asset.

Vickrey auction - Is a auction where the bidders submit sealed bids and therefor doesn't know the others bids in the auction.

Yield - The income return from an investment. Can be both dividends and interest from an asset.

Zero-coupon bond (ZCB) - An asset which does not pay coupon and pays the nominal value at maturity.

References

[1] L. Ballestra, G. Pacelli and D. Radi. *A note on Fergusson and Platen: "Application of maximum likelihood estimation stochastic short-rate models"*. Ann. Finan. Econ., 2016.

[2] L. Ballestra, G. Pacelli and D. Radi. *Valuing investment projects under interest rate risk: empirical evidence from European firms*. Applied Economics, 2017.

[3] D. Bayazit. *Yield curve estimation and prediction with Vasicek model*. Department of Financial Mathematics, The Middle Easy Technical University, 2004.

[4] D. Brigo and F. Mercurio. *Theory and Practice*. Springer Finance, 2007.

[5] Christina Sklibosios Nikitopoulos Carl Chiarella Xue-Zhong He. *Derivative Security Pricing*. 2015.

[6] R. Chami, S. Sharma and C. Fullenkamp. *Framework for Financial Market Development*. International Monetary Fund, 2009.

[7] K. CHAN et al. *An Empirical Comparison of Alternative Models of the Short-Term Interest Rate*. URL: `https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1992.tb04011.x`.

[8] M. Choudhry. *Analysing and Interpreting the Yield Curve*. John Wiley & Sons, Incorporated, 2011.

[9] S. Coles. *An introduction to statistical modelling of extreme values*. Springer series in statistics, 2001.

[10] S. Compagnone. *NEGATIVE INTEREST RATES: how the existent financial models can fit with this new scenario? A focus on Vasicek and CIR*. URL: `https://run.unl.pt/bitstream/10362/27857/1/Compagnone_2017.pdf`.

[11] Deloitte.com. *Interest rate derivatives in the negative-rate environment - Pricing with a shift*. URL: `https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-be-aers-fsi-pricing-with-negative-rates.pdf`.

[12] ECB.com. *Life below zero: Learning about negative interest rates*. URL: `https://www.ecb.europa.eu/press/key/date/2014/html/sp140909.en.html`.

[13] K. Fergusson and E. Platen. *Application of maximum likelihood esitmation stochastic short-rate models*. Ann. Finan. Econ., 2015.

[14] O. Granstand. *INDUSTRIAL INNOVATION ECONOMIC*. Svenska Kulturkompaniet, 2007.

[15] J. Hull and A. White. *Pricing Interest-Rate-Derivative Securities*. URL: `https://academic.oup.com/rfs/article/3/4/573/1585938#113912441`.

[16] John C. Hull. *Options, Futures, and Other Derivatives*. Pearson, 2015.

[17] ig.com. *Vad är statsobligationer?* URL: `https://www.ig.com/se/what-are-government-bonds-`.

[18] Investing.com. *Sweden - Government Bonds*. URL: `https://www.investing.com/rates-bonds/sweden-government-bonds`.

[19] Investopedia.com. *Bond*. URL: `https://www.investopedia.com/terms/b/bond.asp`.

[20] Investopedia.com. *Interest Rate Caps and Floors*. URL: `https://www.investopedia.com/exam-guide/cfa-level-1/derivatives/interest-rate-caps-floors.asp`.

[21] Investopedia.com. *Normal Yield Curve*. URL: `https://www.investopedia.com/terms/n/normalyieldcurve.asp`.

[22] David G. Luenberger. *Investment Science*. Oxford University Press, 2013.

[23] R. Mamon. *Three Ways to Solve for Bond Prices in the Vasicek Model*. URL: `http://www.kurims.kyoto-u.ac.jp/EMIS/journals/HOA/JAMDS/8/11.pdf`.

[24] B. Márton Süli. *Yield curve modelling*. Faculty of Science, Eötvös Loránd University, 2014.

[25] nasdaq.com. *Forward Rate*. URL: `https://www.nasdaq.com/investing/glossary/f/forward-ratef`.

[26] PWC.com. *Negativa räntor i Sverige – innebörd och effekter på redovisningen.* URL: https://www.pwc.se/sv/redovisning/negativa-rantor-i-sverige.html.

[27] Riksbanken. *Referensränta.* URL: https://www.riksbank.se/sv/statistik/sok-rantor--valutakurser/referensranta/.

[28] Riksgalden. *Handbok om statspapper.* Riksgalden, 2007.

[29] Riksgalden.se. *Återförsäljare.* URL: https://www.riksgalden.se/sv/For-investerare/Statspapper/Aterforsaljare/.

[30] Riksgalden.se. *Emissionsbeslut.* URL: https://www.riksgalden.se/sv/For-investerare/policy-regelverk/policy-forsaljning/Emissionsbeslut/.

[31] Riksgalden.se. *Premieobligation.* URL: https://www.riksgalden.se/sv/global/ordlista/P/Premieobligation/.

[32] Riksgalden.se. *Statsskuldväxel.* URL: https://www.riksgalden.se/sv/global/ordlista/S/Statsskuldvaxel//.

[33] Riksgalden.se. *Utestående statspapper.* URL: https://www.riksgalden.se/sv/For-investerare/Statspapper/Utestaende-statspapper/.

[34] K. C. CHAN G. ANDREW KAROLYI FRANCIS A. LONGSTAFF ANTHONY B. SANDERS. *An Empirical Comparison of Alternative Models of the Short-Term Interest Rate.* The journal of finance, 1992.

[35] Swedbank.se. *Statsobligationer.* URL: https://www.swedbank.se/foretag/spara-och-placera/aktier-rantor-och-andra-placeringar/ranteplaceringar/statsobligationer/.

[36] Y. Ueno. *Term Structure Models with Negative Interest Rates.* URL: https://www.imes.boj.or.jp/research/papers/english/17-E-01.pdf.

[37] R. Venkata Subramani. *Accounting for Investments: Fixed Income Securities and Interest Rate Derivatives: A Practitioner's Guide, Volume 2.* John Wiley & Sons, 2011.

[38] B. Wang and Y. Zhao. *Short-Term Interest Rate Models - An Application of Different Models in Multiple Countries.* Lund University School of Economics and Management, 2017.

[39] Paul Wilmott. *Paul Wilmott Introduces Quantitative Finance.* 2001.

[40] YieldCurve.com. *Interest Rate Derivatives: An Introduction to the Pricing of Caps and Floors.* URL: http://www.yieldcurve.com/mktresearch/files/alidualeh_irpricingmar03final.pdfp.

[41] C. Yong Tang and S. Xi Chen. *Parameter estimation and bias correction for diffusion processes.* Journal of Econometrics, 2009.

[42] J. Yu. *Bias in the estimation of the mean reversion parameter in continuous time models.* Journal of Econometrics, 2012.