Stockholms universitet

Forecasting US Interest Rates Using Vasicek Model

Li Wu

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2024:10
Matematisk statistik
Maj 2024

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:10**
http://www.math.su.se

Forecasting US Interest Rates Using Vasicek Model

Li Wu*

May 2024

Abstract

This thesis focuses on enhancing our comprehension of interest rates in the United States while developing a forecasting model based on the Vasicek approach. Simultaneously, it aims to delve into the mathematical and statistical aspects of this model. Leveraging historical data, we strive to estimate key parameters and evaluate the model's predictive accuracy.

The methodology begins by establishing a comprehensive understanding of interest rate theory and related mathematical and statistical concepts. Subsequently, we introduce the Vasicek model, a widely recognized framework for interest rate forecasting. Utilizing Maximum Likelihood Calibration, we estimate model parameters, harnessing insights from historical data to ensure precise estimations.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: catrina.wuli@gmail.com. Supervisor: Taras Bodnar, Jan Olov Persson.

Sammanfattning

I detta examensarbete är vårt primära mål att fördjupa vår förståelse för räntor i USA och att utveckla en prognosmodell med hjälp av det klassiska modell Vasicek-modellen. Samtidigt syftar den till att utforska de matematiska och statistiska aspekterna av denna modell. Genom att utnyttja historiska data strävar vi efter att uppskatta nyckelparametrar och utvärdera modellens prediktiva noggrannhet.

Metodiken inleds med att etablera en grundläggande förståelse för ränteteorin och relaterade matematiska och statistiska begrepp. Därefter introducerar vi Vasicek-modellen, en välkänd ram för prognostisering av räntor. Genom att använda Maximum Likelihood Calibration uppskattar vi modellparametrar och utnyttjar insikter från historiska data för att säkerställa precisa uppskattningar.

Acknowledgements

I am deeply grateful to several individuals whose support and guidance were instrumental in the completion of this thesis. Firstly, I extend my sincere appreciation to my supervisors at Stockholm University, Taras Bodnar, and Jan Olov Persson, for their invaluable support, guidance, and feedback throughout the process. Their expertise and encouragement significantly contributed to the development of both the subject matter and methodology of this thesis.

I would also like to express my heartfelt gratitude to my family. To my husband and daughter, I am indebted for their support, constant encouragement, and motivation. Their belief in me and their understanding of the challenges I faced were indispensable throughout this journey.

Additionally, I want to express my gratitude to ChatGPT for its invaluable assistance in grammar checking, providing code suggestions, and offering Latex support.

Contents

1 Introduction **5**
1.1 Background . . . 5
1.2 Objective . . . 5
1.3 Disposition . . . 5

2 Mathematical Background of Interest Rate Theory **6**
2.1 Discount Factors . . . 6
2.1.1 Constant Interest Rate . . . 6
2.1.2 Continuous Interest Rate . . . 7
2.2 Spot Rate . . . 8

3 Mathematical and Statistical Background **10**
3.1 Diffusion Process . . . 10
3.1.1 Mathematical and Statistical Preliminaries of Diffusion Processes . . . 10
3.2 Examples of Diffusion Processes . . . 11
3.2.1 Brownian Motion . . . 11
3.2.2 Ornstein-Uhlenbeck Process . . . 11
3.3 Stochastic Calculus . . . 12

4 One-factor Spot Rate Model: Vasicek Model **13**
4.1 Assumptions . . . 13
4.2 Standard Vasicek Model . . . 13
4.3 Maximum Likelihood Calibration . . . 15
4.4 Forecasting and Accuracy . . . 18

5 Dataset **19**

6 The Model **19**
6.1 Data Selection . . . 19
6.1.1 Training Data . . . 20
6.1.2 Testing Data . . . 21
6.2 Parameters Estimations in Vasicek Model . . . 21
6.3 Forecasting US Interest Rates . . . 22
6.3.1 1-Step-Ahead Forecast . . . 22
6.3.2 Bootstrap Confidence Interval . . . 22

7 Discussion **24**

1 Introduction

1.1 Background

The pricing of fixed-income securities, such as zero-coupon bonds, is a fundamental aspect of financial markets and is heavily influenced by prevailing interest rates. Consequently, accurate modeling of interest rates has long been a priority in financial mathematics. One model that has been extensively utilized for this purpose is named after its creator, Vasicek, who introduced it in 1977 [13]. The Vasicek model depicts interest rate evolution as a stochastic process characterized by a single source of randomness, hence its designation as a one-factor model. While this model offers simplicity and a closed-form solution, its applicability to real-world financial markets is limited, as these markets are influenced by multiple sources of randomness.

Despite its limitations in capturing the complexity of real financial markets, the Vasicek model provides a valuable framework for understanding the dynamics of interest rates and serves as an illustration of the application of mathematical and statistical theory in modeling real-world phenomena. Moreover, advancements in financial mathematics have led to the proposition of alternative models, such as the Cox-Ingersoll-Ross (CIR) model[11], which is an extension of the Vasicek model and is specifically designed to address some of the limitations of the Vasicek model, particularly in capturing the volatility of interest rates.

1.2 Objective

This thesis aims to provide a comprehensive exploration of interest rates, focusing on their mathematical and statistical descriptions. We will delve into the core concepts of interest rates and investigate how mathematical and statistical tools can be used to characterize them. A key aspect of our study will be the examination of the Vasicek model, a foundational framework introduced in 1977. We will analyze why this model remains widely used today and examine the underlying mathematical and statistical principles. Additionally, we will conduct a forecasting analysis to evaluate the model's predictive accuracy against real-world data. Through these efforts, this thesis seeks to deepen our understanding of interest rate dynamics and contribute to the advancement of financial modeling techniques.

1.3 Disposition

This thesis is structured as follows: Sections 2 and 3 delve into the theory of interest rates, along with mathematical and statistical concepts pertinent to the subject matter. Section 4 provides a detailed analysis of the standard Vasicek model, focusing on its parameter estimation using Maximum Likelihood Estimation (MLE). Finally, sections 5 and 6 encompass the practical application of the Vasicek model to historical interest rate data, including model fitting and interpretation of forecasting results.

2 Mathematical Background of Interest Rate Theory

In the field of finance, various types of interest rates coexist. One commonly discussed rate is the federal funds rate, also known as the interbank rate in USA, controlled by the central bank. However, this thesis will specifically delve into another category of interest rates—zero-coupon interest rates, also known as yields. These rates essentially represent the return or yield obtained from investments in government bonds. This choice stems from the recognition that most financial studies involve the returns, rather than the prices, of assets, because return series are easier to handle than price series, and the former possess more favorable statistical properties. The interest rate we are discussing here is actually the return of a bond investment, which is why it is also referred to as yields.

This thesis centers on the examination of zero coupon interest rates (or yields). Several compelling reasons support this specific focus, each contributing to a thorough understanding of financial dynamics:

-**Risk-Free Rate:** Zero coupon interest rates provide invaluable insights into the risk-free rate, a foundational element for various financial analyses.

-**Usage in Discounted Cash Flow Analysis:** A thorough comprehension of zero coupon interest rates is indispensable for the application of discounted cash flow analysis—a fundamental financial technique employed for evaluating the present value of future cash flows.

-**Pricing Fixed Income Securities:** Zero coupon interest rates play a pivotal role in the pricing of fixed income securities, exerting a substantial influence on their valuation within the market.

In this section, we introduce the main definitions and concepts related to interest rate that will be utilized throughout the thesis.

2.1 Discount Factors

Discount factors play a pivotal role in financial analysis, serving as fundamental components utilized across various domains such as pricing and risk assessment. Regardless of whether interest rates are assumed to be constant or modeled as a continuous process, the underlying concept of discount factors remains consistent. These factors are instrumental in determining the present value of assets at different time points. In order to gain familiarity with this concept, it is necessary to introduce some relevant definitions and notations.[2]

Definition: ***Bank account (Money-market account)*** *We define $B(t)$ to be the value of a bank account at time $t \geq 0$. We assume $B(0) = 1$ and that the bank account evolves according to the following differential equation:*

$$dB(t) = r_t B(t)dt,$$

where r_t is a positive function of time. As a consequence,

$$B(t) = \exp\left(\int_0^t r_s ds\right).$$

2.1.1 Constant Interest Rate

Constant interest rate refers to a fixed rate that remains unchanged over a specified period of time. It is commonly employed in financial calculations and models where the interest rate is assumed to remain constant throughout the duration of the investment or analysis. We will derive the discount factors under the assumption of a constant interest rate.

The interest rate, denoted as i, is a constant. For example, if the interest rate remains i yearly all the time, then $B(0) = 1$ unit of money invested at your *bank account* at time 0, which represents today. After one year, the value of $B(1)$ will be

$$B(1) = B(0)(1+i)$$

and after t years

$$B(t) = B(0)(1+i)^t.$$

Here, we introduce a notation r and generalize the example above:

$$r = \ln(1+i),$$

$$\exp(r) = (1+i),$$

$$\exp(rt) = (1+i)^t,$$

$$B(t) = B(0)\exp(rt).$$

Here, $B(t)$ is the future bank account value at time t. If we want to derive its present value, namely $t = 0$, we have

$$B(0) = \exp(-rt)B(t),$$

where r is a constant.

Subsequently, we determine the discount factor in the case of a constant interest rate, denoted by $\exp(-rt)$.

2.1.2 Continuous Interest Rate

Continuous interest rate refers to a theoretical concept where interest is compounded continuously. We will explain more about later. The derivation of the discount factor in this case is analogous to the constant interest case; the only difference is that, the interest rate i is a function of time s in continuous time. This implies that

$$B(t) = B(0)\prod_{s=0}^{t}(1+i(s))$$

where

$$(1+i(s)) = \exp(r_s).$$

Because s represents continuous time here, we can rewrite the equation as follows:

$$B(t) = B(0)\exp(\int_0^t r_s ds),$$

and

$$B(0) = B(t)\exp(-\int_0^t r_s ds).$$

Subsequently, we determine the discount factor in the case of a continuous interest rate, denoted by $\exp(\int_0^t -r_s ds)$. This r_s is identical to the r_s in the definition of *Bank account*, and is often referred to as the force of interest.

2.2 Spot Rate

In this section, we delve into the core concept of interest rate theory: Spot rate. We will attempt to model it, examine its mathematical properties, and analyze its trends to gain insight into future interest rates in later sections. To develop a basic understanding of spot rate, it is necessary to introduce several definitions and notations[8].

Definition: ***Zero coupon bond*** *A zero coupon bond with maturity date T (also called a T-bond) is a contract which pays one unit at time T. The price at time $t \in [0,T]$ is denoted by $P(t,T)$. Throughout, we take $P(t,t) = 1$.*
Assuming that the prices of zero-coupon bonds are deterministic at any time t, where $0 \leq t \leq T' \leq T$. This implies that

$$P(t,T) = \exp(-\int_t^T r_s ds)P(T,T),$$

$$P(t,T) = \exp(-\int_t^{T'} r_s ds) \cdot \exp(-\int_{T'}^T r_s ds),$$

$$P(t,T) = P(t,T')P(T',T).$$

Often, the contracts defined above are referred to as *default-free zero coupon bonds*. This emphasizes the low likelihood of these bonds defaulting, meaning that the issuers are not likely to go bankrupt. However, it is important to note that analogous bonds exist in the markets, and when we mention zero coupon bonds in the thesis, we specifically refer to bonds issued by the U.S. government.

Definition: ***Continuously compounded spot rate*** *The continuously compounded zero coupon yield (or the continuously compounded spot rate) $R(t,T)$ is defined by*

$$R(t,T) = -\frac{\log P(t,T)}{T-t}.$$

It follows directly from this definition that the price of the zero coupon bond can be expressed in terms of the continuously compounded spot rate as follows:

$$P(t,T) = \exp(-R(t,T)(T-t)),$$

so the equation above can be interpreted as a discount factor obtained by using the (constant) interest rate $R(t,T)$ during the interval $[t,T]$.

Now consider $0 \leq t \leq T' \leq T$. We are interested in finding a deterministic rate at time t (today) for a future investment made at time T' and terminating at time T. This quantity is precisely the forward rate. We cannot use the yield $R(T',T)$ since it is defined in terms of $P(T',T)$, which, in general, is not known at time t. Instead, one introduces the concept of forward rates.

Definition: ***Continuously compounded forward rate*** *The continuously compounded forward rate at time t for the period $[T',T]$ is defined by*

$$f(t,T',T) = -\frac{\log P(t,T) - \log P(t,T')}{T-T'}.$$

Mathematically, it is the same as $R(T',T)$,

$$R(T',T) = -\frac{\log P(T',T)}{T-T'}.$$

$$R(T',T) = -\frac{\log P(t,T)/P(t,T')}{T-T'} = -\frac{\log P(t,T) - \log P(t,T')}{T-T'}.$$

In finance, they represent different concepts. The spot rate refers to the interest rate applicable today or at a specific historical time, while the forward rate represents the interest rate applicable to a future period of time, agreed upon today.

3 Mathematical and Statistical Background

In this section, we outline the mathematical and statistical foundations crucial for understanding the content presented in this thesis. The first part explores diffusion processes, among which the Vasicek model assumes that the spot rate follows the Ornstein-Uhlenbeck process, which is an example of a diffusion process. The second part delves into stochastic calculus, pivotal for solving stochastic differential equations within the Vasicek model framework. We will examine various definitions and theorems to deepen our comprehension of the Vasicek model.

3.1 Diffusion Process

"Diffusion processes are continuous-time, continuous-state processes whose sample paths are everywhere continuous but nowhere differentiable. In the fields where diffusion has been applied, it has been used to model phenomena evolving randomly and continuously in time under certain conditions, for example, bond price fluctuations in a perfect market, variations of population growth in ideal conditions, and communication systems with noise"(Ibe, 2013, p. 295).

3.1.1 Mathematical and Statistical Preliminaries of Diffusion Processes

Consider a continuous-time, continuous-state Markov process X_t, $t \geq 0$, $t > s$ whose transition probability distribution is given by

$$F(X_t \mid X_s) = P(X_t \leq y \mid X_s = x) \tag{1}$$

If the derivative

$$f(X_t \mid X_s) = \frac{\partial}{\partial y} F(X_t \mid X_s) \tag{2}$$

exists, then it is called the transition density function of the diffusion process. Since X_t is a Markov process, $f(X_t \mid X_s)$ satisfies the Chapman-Kolmogorov equation:

$$f(X_t, y \mid X_s, x) = \int_{-\infty}^{\infty} f(X_t, y \mid X_u, z) f(X_u, z \mid X_s, x) dz \tag{3}$$

We assume that the process $X(t), t \geq 0$ satisfies the following conditions:

1. $P(X_{t+\Delta t} - X_t \mid X_t = x) \geq \varepsilon = o(\Delta t)$, for $\varepsilon > 0$, which states that the sample path is continuous; alternatively, we say that the process is continuous in probability.
2. $E[X_{t+\Delta t} - X_t \mid X_t = x] = \mu(t,x)\Delta t + o(\Delta t)$ so that
$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E[X_{t+\Delta t} - X_t \mid X_t = x] = \mu(t,x)$$
3. $E[(X_{t+\Delta t} - X_t \mid X_t = x)^2 | X_t = x] = \sigma^2(t,x)\Delta t + o(\Delta t)$ is finite so that
$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E[(X_{t+\Delta t} - X_t \mid X_t = x)^2 | X_t = x] = \sigma^2(t,x)$$

A Markov process that satisfies these three conditions is called a diffusion process. The function $\mu(t,x)$ is called the instantaneous (or infinitesimal) mean (or drift) of X_t, and the function $\sigma^2(t,x)$ is called the instantaneous (or infinitesimal) variance of X_t. Let the small increment in X_t over any small interval dt be denoted by dX_t. Then it can be shown that if W_t is a standard Brownian motion we can incorporate the above properties into the following stochastic differential equation:

$$dX_t = \mu(t,x)dt + \sigma(t,x)dW_t \tag{4}$$

where dW_t is the increment of W_t over the small interval $(t, t+\Delta t)$(Ibe, 2013, p. 296-297).[6]

Definition: ***A stochastic differential equation*** *A stochastic differential equation (SDE) is a differential equation in which one or more of the terms is a stochastic process, resulting in a solution which is also a stochastic process.* [9]

3.2 Examples of Diffusion Processes

The following are two diffusion processes closely related to the Vasicek model. These processes differ only in their values of the infinitesimal mean and infinitesimal variance.

3.2.1 Brownian Motion

A well-known example of a diffusion process is Brownian motion. In the Brownian view, diffusion is a process that causes particle mixing because of random collisions among themselves or with other particles. In Brownian motion, we model the movement of a single particle, whereas with diffusion, we are dealing with a larger number of particles.

Definition: ***Standard Brownian motion*** *Standard Brownian motion starting at level zero is a process* $W(t), 0 \leq t$ *satisfying the conditions*(Brockwell,1991, p. 37-38)[3]
(a) $W(0) = 0$,
(b) $W(t_2) - W(t_1), W(t_3) - W(t_2), ..., W(t_n) - W(t_{n-1})$, are independent for every $n \in 3, 4, ...$ and every $t = (t_1, ..., t_n)'$ such that $0 \leq t_1 < t_2 < ... < t_n$,
(c) $W(t) - W(s) \sim N(0, t-s)$ for $s \leq t$.

Theorem: *The Brownian motion exists on a probability space.*
We check some properties of the Brownian motion. In general, a stochastic process $X_t, t \geq 0$ is said to be *adapted* if X_t is a function of the past history $\mathscr{F}_s, s \leq t$. An adapted process X_t is called a Markov process if the conditional distribution of X_t given $\mathscr{F}_s$ depends only on X_s for $s < t$. An adapted process X_t is called a *martingale* if $E[X_t|\mathscr{F}_s] = X_s$ for $s < t$.[10]

Theorem: *The Brownian motion itself is a Markov process and martingale.*

3.2.2 Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck process is a diffusion process that was first introduced to help model the velocity of a particle undergoing Brownian motion.
It is a stochastic process that satisfies the following stochastic differential equation:[7]

$$dX_t = a(b - X_t)dt + \sigma dW_t, \quad t \in [0, \infty). \tag{5}$$

where W_t is a standard Brownian motion on $t \in [0, \infty)$.
The constant parameters are:

- $a > 0$ is the rate of mean reversion;
- b is the long-term mean of the process;
- $\sigma > 0$ is the volatility or average magnitude, per square-root time, of the random fluctuations that are modelled as Brownian motions.

Mean-reverting property: If we ignore the random fluctuations in the process due to dW_t, then we see that X_t has an overall drift towards a mean value b. The process X_t reverts to this mean at rate a, with a magnitude in direct proportion to the distance between the current value of X_t and b.

This can be seen by looking at the solution to the ordinary differential equation $dx_t = a(b - x_t)dt$ which is:

$$\frac{1}{ab - ax_t} dx_t = 1dt, \tag{6}$$

and

$$x_t = c_1 e^{-at} + b. \tag{7}$$

For this reason, the Ornstein-Uhlenbeck process is also referred to as a mean-reverting process when $a > 0$. However, if a is negative, the process does not revert to the long-term mean; instead, it diverges to infinity as t approaches infinity, following the properties of the exponential function. In such cases, the mean-reverting property is not preserved. Therefore, in the Vasicek model, we exclude estimates with negative values of a, which we will discuss further.

3.3 Stochastic Calculus

We now discuss the concept of a stochastic integral, ignoring the various technical conditions that are required to make our definitions rigorous. In this section, we write X_t to emphasize that the quantities in question are stochastic. Itô's formulas, named after the Japanese mathematician Kiyoshi Itô, are fundamental results in stochastic calculus. They provide a method for computing the differentials of functions of stochastic processes, such as Brownian motion. These formulas are essential for solving stochastic differential equations (SDEs). In this thesis, we will utilize Itô's Isometry theorem.

Definition: ***Elementary Function*** *We say a process, $h_t(\omega)$, is elementary if it is piece-wise constant so that there exists a sequence of stopping times $0 = t_0 < t_1 < \cdots < t_n = T$ and a set of flirtations $\mathscr{F}_{t_i}$-measurable functions, $e_i(\omega)$, such that*

$$h_t(\omega) = \sum_i e_i(\omega) \mathbf{I}_{[t_i, t_{i+1})}(t)$$

where $\mathbf{I}_{[t_i, t_{i+1})}(t) = 1$ if $t \in [t_i, t_{i+1})$ and 0 otherwise.

Definition: ***The Itô Integral*** *The stochastic integral of an elementary function, $h_t(\omega)$, with respect to a Brownian motion, W_t, is defined as*

$$\int_0^T h_t(\omega)\, dW_t(\omega) := \sum_{i=0}^{n-1} e_i(\omega) \left(W_{t_{i+1}}(\omega) - W_{t_i}(\omega) \right)$$

Definition: ***The Itô Isometry*** *Let $W : [0, T] \times \Omega \to \mathbb{R}$ denote the canonical real-valued Brownian motion defined up to time $T > 0$, and let $X : [0, T] \times \Omega \to \mathbb{R}$ be a stochastic process that is adapted to the natural filtration $\mathscr{F}_{t_i}^W$ of the Brownian motion. Then*

$$\mathrm{E}\left[\left(\int_0^T X_t\, dW_t \right)^2 \right] = \mathrm{E}\left[\int_0^T X_t^2\, dt \right],$$

where E *denotes expectation with respect to Brownian motion.* [10]

4 One-factor Spot Rate Model: Vasicek Model

We are interested in forecasting the yield (or the continuously compounded spot rate) for 1-year maturity Treasury Securities. With data available in daily frequency, the Vasicek model, commonly utilized for short-rate modeling, can be adapted to capture the daily changes in yields at this specific maturity.

4.1 Assumptions

First, the spot rate $r(t)$, namely the instantaneously continuously compounded spot rate, is presumed to conform to a continuous-time stochastic process and is regarded as possessing the Markov property. Second, the market price of risk is considered constant. Third, the price $P(t,s)$ of a discount bond is determined by assessing the segment $\{r(\tau), t \geq \tau \geq s\}$ of the spot rate process over the term of the bond. Additionally, the market is assumed to be efficient, implying the absence of transaction costs, universal availability of information to all investors simultaneously, and rational behavior by all investors, ensuring that no riskless arbitrage opportunities exist.

In the Vasicek model, interest rates are directly modeled rather than being derived from observed bond prices. In contrast, in the real world, market participants often observe bond prices and use them to infer implied interest rates.

Moreover, the rate $r(t)$ can have a negative value with positive probability, which constitutes a major drawback of the Vasicek model. However, its analytical tractability is a notable feature. The Vasicek model, as proposed in financial literature, is time-homogeneous, implying that the assumed short-rate dynamic depends solely on constant coefficients.[13]

4.2 Standard Vasicek Model

Vasicek(1977) model specifies that the instantaneous spot rate under the real-world measure evolves as an Ornstein-Uhlenbeck process with constant coefficients. The stochastic differential equation for the process is given by

$$dr(t) = a(b - r(t))dt + \sigma dW_t,$$

where W_t is a standard Brownian motion under the risk-neutral framework, which represents the continuous inflow of randomness into the system[2]. The typical parameters b, a, and σ, together with the initial condition r_0, where $r(0) = r_0$, completely characterize the dynamics, and can be quickly characterized as follows, assuming a to be non-negative:

- b: "long term mean level". All future trajectories of r will evolve around a mean level b in the long run.
- a: "speed of reversion". a characterizes the velocity at which such trajectories will regroup around b in time.
- σ: "instantaneous volatility", measures instant by instant the amplitude of randomness entering the system. Higher σ implies more randomness.

By integration, we obtain, between two any instants s and t where t is greater than s,

$$r(t) = r(s)e^{-a(t-s)} + b(1 - e^{-a(t-s)}) + \sigma \int_s^t e^{-a(t-u)} dW_u.$$

Given the available information $\mathscr{F}$ at time s, $r(t) \mid \mathscr{F}_s$ is normally distributed. This result stems from the property of a standard Brownian motion, where the increment in a standard Brownian motion follows a normal distribution with an expected value of zero and a variance equal to the time difference between increments, so $r(t) \mid \mathscr{F}_s$ is normally distributed with mean and variance given, respectively, by

$$E[r(t) \mid \mathscr{F}_s] = r(s)e^{-a(t-s)} + b(1-e^{-a(t-s)})$$

$$Var[r(t) \mid \mathscr{F}_s] = \frac{\sigma^2}{2a}\left(1-e^{-2a(t-s)}\right).$$

To validate these two equalities, we compute

$$\begin{aligned} E[r(t) \mid \mathscr{F}_s] &= E[r(s)e^{-a(t-s)}] + E[b(1-e^{-a(t-s)})] + E[\sigma \int_s^t e^{-a(t-u)} dW_u] \\ &= E[r(s)e^{-a(t-s)}] + E[b(1-e^{-a(t-s)})] + 0 \text{ (by the Itô integral)} \\ &= r(s)e^{-a(t-s)} + b(1-e^{-a(t-s)}) \end{aligned}$$

and

$$\begin{aligned} \text{Var}(r(t) \mid \mathscr{F}_s) &= \text{Var}\left(\sigma \int_s^t e^{-a(t-u)} dW_u\right) = \sigma^2 e^{-2at} \mathbb{E}\left[\left(\int_s^t e^{au} dW_u\right)^2\right] \\ &= \sigma^2 e^{-2at} \int_s^t e^{2au} du \text{ (by the Itô Isometry)} \\ &= \frac{\sigma^2}{2a}\left(1-e^{-2a(t-s)}\right). \end{aligned}$$

For illustrative purposes, we conduct two simulations in the Vasicek model. The only difference between the simulations lies in the value of the parameter a: one simulation has $a = 0.3$, while the other has $a = 0.01$. This parameter a represents the velocity at which trajectories will regroup around the long-term mean b over time. With a higher value of a, trajectories regroup much faster around the long-term mean, which is 3 in our case. Conversely, with a lower value of a, the paths will drift more before regrouping. Figure 1 illustrates simulations depicting the effect of mean reversion.

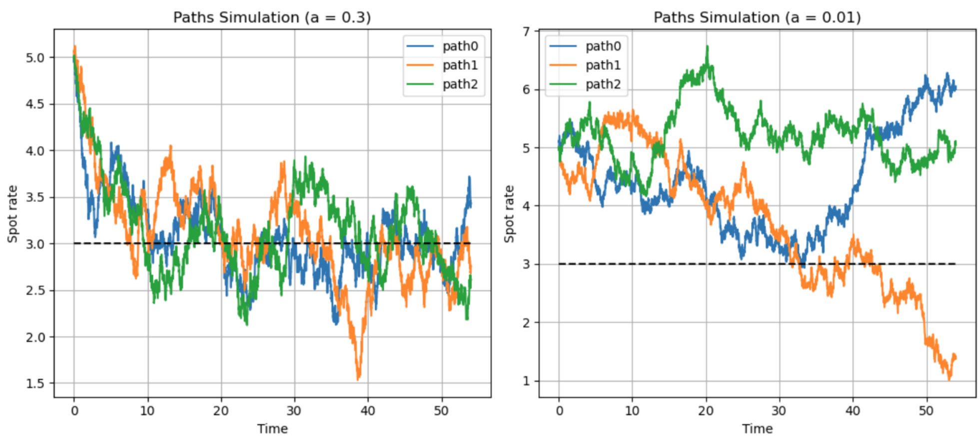

Figure 1: Three paths of the Vasicek model together with its long-term mean. The parameters chosen for simulations are, $a = 0.3/0.01$, $b = 3$, $r_0 = 5$, $\sigma = 0.3$.

4.3 Maximum Likelihood Calibration

In order to forecast the spot rate by using the Vasicek model, this thesis employs Maximum Likelihood Estimation to estimate the parameters a, b and σ which are positive constants. Let R_{t_i} represent spot rates during the time period between i and $i-1$, replacing the notation $r(t)$ and also providing additional information regarding the position i, where i takes non-negative integer values. Meanwhile, R_{t_i} is the increment in a standard Brownian motion, in other words, it is normally distributed.[4]
A random process $R_{0:n} = \{R_{t_0}, R_{t_1}, \ldots, R_{t_n}\}$, which consists of $n+1$ different points with uniform time intervals $dt = t_i - t_{i-1}$ at given time points $t_0, t_1, \ldots, t_n$, follows a continuous Markov process by the assumption in Vasicek model. As result we have the joint density function as below:

$$f(R; a, b, \sigma) = f(R_{t_n}, R_{t_{n-1}}, \ldots, R_{t_0})$$
$$= f(R_{t_n} \mid R_{t_{n-1}}, R_{t_{n-2}}, \ldots, R_{t_0}) \cdot f(R_{t_{n-1}}, R_{t_{n-2}}, \ldots, R_{t_0})$$
$$= f(R_{t_n} \mid R_{t_{n-1}}) \cdot f(R_{t_{n-1}} \mid R_{t_{n-2}}), \ldots, \cdot f(R_{t_0})).$$

The probability density function of $R_{t_i} \mid R_{t_{i-1}}$ is normally distributed and takes the form of

$$f(R_{t_i} \mid R_{t_{i-1}}) = \frac{1}{\sqrt{\pi \frac{\sigma^2}{a}(1 - e^{-2adt})}} e^{-\frac{(R_{t_i} - R_{t_{i-1}} e^{-adt} - b + b e^{-adt})^2}{\frac{\sigma^2}{a}(1 - e^{-2adt})}}.$$

with the expected value and variance given by:

$$E[R_{t_i} \mid R_{t_{i-1}}] = R_{t_{i-1}} e^{-adt} + b(1 - e^{-adt})$$

and

$$Var(R_{t_i} \mid R_{t_{i-1}}) = \frac{\sigma^2}{2a}(1 - e^{-2adt}).$$

So, the log-likelihood function for the process is:

$$\mathscr{L} = \ln f(R_{t_0}) + \sum_{t=1}^{n} \ln f(R_{t_i} \mid R_{t_{i-1}})$$

or

$$\mathscr{L} = -n \cdot \ln\left(\frac{\sigma^2}{a}(1 - e^{-2adt})\right) - \frac{\sum_{t=1}^{n} (R_{t_i} - R_{t_{i-1}} e^{-adt} - b(1 - e^{-adt}))^2}{\frac{\sigma^2}{a}(1 - e^{-2adt})}.$$

Maximum likelihood estimators are

$$\hat{a} = -\frac{1}{dt} \ln\left(\frac{n \sum_{t=1}^{n} R_{t_i} R_{t_{i-1}} - \sum_{t=1}^{n} R_{t_i} \sum_{t=1}^{n} R_{t_{i-1}}}{n \sum_{t=1}^{n} R_{t_{i-1}}^2 - (\sum_{t=1}^{n} R_{t_{i-1}})^2}\right),$$

$$\hat{b} = \frac{1}{n(1 - e^{-\hat{a}dt})} \left(\sum_{t=1}^{n} R_{t_i} - e^{-\hat{a}dt} \sum_{t=1}^{n} R_{t_{i-1}}\right),$$

and

$$\hat{\sigma^2} = \frac{2\hat{a}}{n(1 - e^{-2\hat{a}dt})} \sum_{t=1}^{n} (R_{t_i} - R_{t_{i-1}} e^{-\hat{a}dt} - \hat{b}(1 - e^{-\hat{a}dt}))^2.$$

To derive the maximum likelihood estimators for a, b, σ, we can apply an important property of the maximum likelihood estimate:

Invariance of the MLE [5]
Let$\hat{\theta}_{\text{ML}}$ be the maximum likelihood estimator (MLE) of θ, and let $\phi = h(\theta)$ be a one-to-one transformation of θ. The MLE of ϕ can be obtained by inserting $\hat{\theta}_{\text{ML}}$ into $h(\theta)$: $\hat{\phi}_{\text{ML}} = h(\hat{\theta}_{\text{ML}})$.
Utilizing this property within likelihood theory enables us to simplify the calculations significantly when compared to the derivations presented in Malmström et al. [4].

We express the log-likelihood as

$$\mathscr{L} = -n \cdot \ln C - \frac{\sum_{t=1}^{n}(R_{t_i} - R_{t_{i-1}}A - B)^2}{C},$$

Where

$$A = e^{-adt},$$

$$B = b(1 - e^{-adt}),$$

and

$$C = \frac{\sigma^2}{a}(1 - e^{-2adt}).$$

The maximum likelihood estimators of A, B, and C are one-to-one transformations of (a, b, σ^2). Then we have

$$\frac{\partial \mathscr{L}}{\partial A} = 0 - \frac{1}{C}\sum_{t=1}^{n} -R_{t_{i-1}}(R_{t_i} - R_{t_{i-1}}A - B) = 0, \tag{8}$$

$$\frac{\partial \mathscr{L}}{\partial B} = 0 - \frac{1}{C}\sum_{t=1}^{n} -1(R_{t_i} - R_{t_{i-1}}A - B) = 0, \tag{9}$$

and

$$\frac{\partial \mathscr{L}}{\partial C} = -\frac{n}{C} + \frac{1}{C^2}\sum_{t=1}^{n}(R_{t_i} - R_{t_{i-1}}A - B)^2 = 0. \tag{10}$$

By calculating equations (8)-(10), we obtain:

$$A = \frac{\sum_{t=1}^{n} R_{t_i}R_{t_{i-1}} - B \cdot \sum_{t=1}^{n} R_{t_{i-1}}}{\sum_{t=1}^{n} R_{t_{i-1}}}, \tag{11}$$

$$B = \frac{1}{n}(\sum_{t=1}^{n} R_{t_i} - A \cdot \sum_{t=1}^{n} R_{t_{i-1}}), \tag{12}$$

and

$$C = \frac{\sum_{t=1}^{n}(R_{t_i} - R_{t_{i-1}}A - B)^2}{n}. \tag{13}$$

Then we solve the equation system (11)-(13) and apply Invariance of the MLE, we have

$$A = \frac{n\sum_{t=1}^{n} R_{t_i}R_{t_{i-1}} - \sum_{t=1}^{n} R_{t_i} \cdot \sum_{t=1}^{n} R_{t_{i-1}}}{n\sum_{t=1}^{n} R_{t_{i-1}}^2 - (\sum_{t=1}^{n} R_{t_{i-1}})^2} = e^{-\hat{a}dt}$$

and, consequently

$$\hat{a} = -\frac{1}{dt}\ln(\frac{n\sum_{t=1}^{n} R_{t_i}R_{t_{i-1}} - \sum_{t=1}^{n} R_{t_i}\sum_{t=1}^{n} R_{t_{i-1}}}{n\sum_{t=1}^{n} R_{t_{i-1}}^2 - (\sum_{t=1}^{n} R_{t_{i-1}})^2}).$$

As $B = \frac{1}{n}\sum_{t=1}^{n}(R_{t_i} - R_{t_{i-1}}A) = b(1 - e^{-adt})$ and $A = e^{-adt}$, we obtain

$$\hat{b} = \frac{1}{n(1 - e^{-\hat{a}dt})}(\sum_{t=1}^{n} R_{t_i} - e^{-\hat{a}dt}\sum_{t=1}^{n} R_{t_{i-1}}).$$

As $C = \frac{\sigma^2}{a}(1-e^{-2adt}) = \frac{\sum_{t=1}^{n}(R_{t_i}-R_{t_{i-1}}A-B)^2}{n}$, we obtain

$$\hat{\sigma^2} = \frac{2\hat{a}}{n(1-e^{-2\hat{a}dt})}\sum_{t=1}^{n}(R_{t_i}-R_{t_{i-1}}e^{-\hat{a}dt}-\hat{b}(1-e^{-\hat{a}dt}))^2.$$

Here We define a function to estimate parameters from historical data in python, drawing inspiration from the MIT open-source framework. [1]:

```
"""
Arguments of the function
data1: array-like sequence of R_t_(i-1)
data2: array-like sequence of R_t_i
n: SampleSize-1
dt: increments of time
"""
def para_estimator(data1,data2,n,dt):
    Sx = sum(data1) #sum of sequence of R_t_(i-1)
    Sy = sum(data2) #sum of sequence of R_t_i
    Sxx = sum(data1 * data1) #sum of sequence of R_t_(i-1)^2
    Sxy = sum(data1 * data2) #sum of sequence of R_t_(i-1)*R_t_i
    Syy = sum(data2 * data2) #sum of sequence of R_t_i^2

    #calculate a hat
    a_1 = (n * Sxy - Sx * Sy)/(n * Sxx - Sx**2)
    a = -np.log(a_1)/dt

    #calculate b hat
    b_1 = np.exp(-a*dt)
    b = (Sy - b_1*Sx)/(n*(1- b_1)) #

    #calculate sigma hat
    sigma_square_1 = (Syy - 2*b_1* Sxy + b_1*b_1*Sxx - 2*b*(1-b_1)*(Sy-b_1*Sx)+n*b*
    sigma_square = (2*a*sigma_square_1)/(n-n*(b_1*b_1))
    sigma = np.sqrt(sigma_square)

    #return the estimates of a,b,sigma
    return a,b,sigma
```

4.4 Forecasting and Accuracy

To forecast interest rates using the Vasicek model, the involved parameters a, b and σ need to be calibrated. This calibration process is carried out using maximum likelihood estimation in last section. Once calibrated, the parameter estimates will be utilized to forecast future expected interest rates through the following conditional expectation formula [12]

$$E[R_{t_i} \mid R_{t_{i-1}}] = R_{t_{i-1}} e^{-adt} + b(1 - e^{-adt}).$$

In order to measure the accuracy of our approach, we compute the square root of the mean square error (RMSE), denoted as ε, defined as

$$\varepsilon = \sqrt{\frac{1}{n} \sum_{t=1}^{n} {e_h}^2},$$

where $e_h = R_h - \hat{R}_h$ represents the residual between the actual interest rate R_h and the corresponding fitted values $\hat{R}_h$.

Here is the Python code for predicting interest rates and calculating their RMSE.

```
#Function to calculate R_i+1 considering a one-step forecast horizon
def r_Vasicek(n,a,b):
    if n == 0 : # base case
        return 4.99 #interest rate 2024-03-15
    else:
        return r_Vasicek(n-1,a,b)*np.exp(-a) + b*(1-np.exp(-a))

#Function to calculate RMSE
def calculate_rmse(actual_values, predicted_values):
    # Arguments actual_values and predicted_values is a array-like sequence.

    # Ensure both arrays have the same length
    if len(actual_values) != len(predicted_values):
        raise ValueError("Actual and predicted arrays must have the same length.")

    # Calculate the mean square error (MSE)
    mse = np.mean((actual_values - predicted_values)**2)

    # Calculate the square root of the mean square error (RMSE)
    rmse = np.sqrt(mse)

    return rmse
```

5 Dataset

This section provides details of the dataset used for forecasting interest rates. Our dataset records daily interest rates, known as Market Yield, with 1-year maturities, sourced from the Federal Reserve Economic Data for the USA.

Figure presents a time plot of U.S. daily interest rates, specifically highlighting the 1-year Treasury constant maturity rates, with shaded areas representing U.S. recessions.

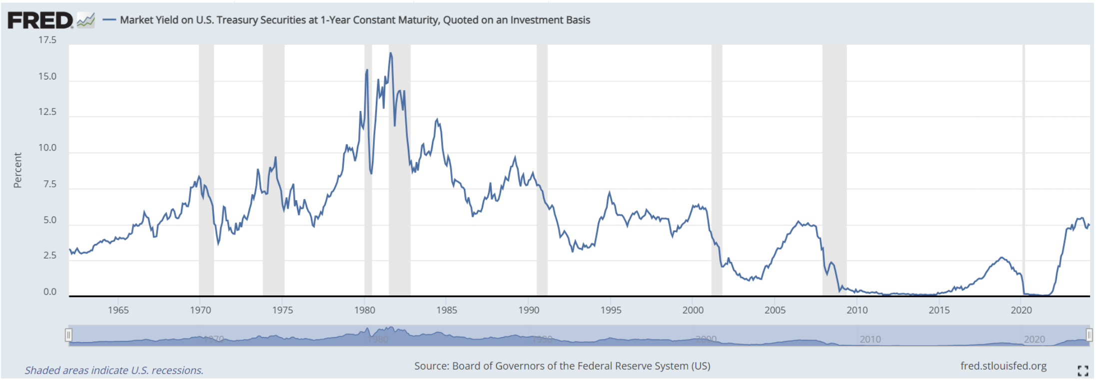

Figure 2: Time plots of daily U.S. interest rate from February 1962 to February 2024 : 1-year maturity rate.

6 The Model

6.1 Data Selection

In this section, we delve into various data selection methods for estimating parameters using Maximum Likelihood Estimators. Given the uncertainty surrounding the optimal amount of data required for accurate parameter estimation, our aim is to identify the best datasets that yield the most reliable forecast results. We will employ the square root of the mean square error (RMSE) to quantify this accuracy in the next session. Since interest rates may be sensitive to recent trends, two out of three methods include data from the most recent year. Furthermore, to maintain the mean-reverting property in the Vasicek model, parameter a is should be a positive constant. As a result, estimates with negative values of a will be excluded.

We examine three distinct data selection methods: Conventional Few-Year Data, Partitioned Dataset, and Rolling Window Technique.

6.1.1 Training Data

All data points from 1962 to 2024-02-13 will be used for training, resulting in approximately 100 subsets.

Method 1: Conventional Few-Year Data This method involves utilizing data from the past few years, typically less than five years, to estimate parameters. After applying this method, we obtained the following parameter estimates:

Partitions	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
2021 to 2024	0.0066	11.62	0.184
2022 to 2024	0.060	5.34	0.220
2023 to 2024	0.102	5.070	0.232

Table 1: Parameter estimates for Conventional Few-Year Data Method

Method 2: Partitioned Dataset Here, we utilize the entire dataset and partition it into multiple segments, progressively excluding data from the oldest year in each subsequent partition. This approach allows us to assess the impact of dataset size on parameter estimation accuracy and explore whether specific partitions yield superior results. After applying this method, we obtained the following parameter estimates. Here, we display only a subset of the results, but all the valid parameter estimates will be used in forecasting.

Partitions	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
1968 to 2024	0.0026	4.78	0.26
1969 to 2024	0.0027	4.58	0.26
1970 to 2024	0.0029	4.11	0.26
1971 to 2024	0.0026	4.89	0.26
1972 to 2024	0.0026	5.02	0.26
1973 to 2024	0.0026	4.66	0.26
1974 to 2024	0.0027	4.15	0.27
1975 to 2024	0.0027	4.08	0.26

Table 2: Parameter estimates for Partitioned Dataset Method

Method 3: Rolling Window Technique In this method, we implement rolling window techniques with a fixed window size of 10. By continuously updating the dataset window, we aim to capture changes in parameter estimates over time and identify potential improvements in estimation accuracy. After applying this method, we obtained the following parameter estimates. Here, we display only a subset of the results, but all the valid parameter estimates will be used in forecasting.

Through these comprehensive investigations, our goal is to determine whether specific subsets of the data provide advantages in achieving accurate parameter estimation or if they all result in similar estimates.

Partitions	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$
1962 to 1972	0.007	6.07	0.13
1963 to 1973	0.007	7.37	0.15
1964 to 1974	0.009	7.15	0.19
1965 to 1975	0.015	6.52	0.21
1966 to 1976	0.014	6.14	0.22
1967 to 1977	0.016	6.67	0.22
1968 to 1978	-0.0009	-10.52	0.22
1969 to 1979	-0.002	-1.29	0.26
1970 to 1980	-0.001	-10.43	0.36
1971 to 1981	0.006	13.30	0.43

Table 3: Parameter estimates for Rolling Window Techniques

6.1.2 Testing Data

The test data comprises actual daily interest rates spanning from February 13, 2024, to April 13, 2024, constituting a single subset.

6.2 Parameters Estimations in Vasicek Model

All valid estimates from previous section are utilized to compute the corresponding RMSE with the test data. We select the four estimates with the lowest RMSE values and plot their fitted values alongside the real data in the figure below.

Partitions	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	RMSE
2009 to 2019	0.0009	5.28	0.057	0.0395
1985 to 1995	0.01	5.06	0.208	0.04
1972 to 2024	0.0026	5.02	0.26	0.0397
2023 to 2024	0.102	5.070	0.232	0.0679

Table 4: Parameter estimates and the corresponding RMSE for Vasicek models

In the plot below, each of the four curves represents the fitted values derived from different sets of estimated parameters, showcasing the variance between these fitted values and the actual data. It is evident that the fitted values fail to fully capture the volatility observed in the actual data, as the random term is omitted in our fitted models. This omission stems from the utilization of the conditional expectation formula detailed in Section 4.

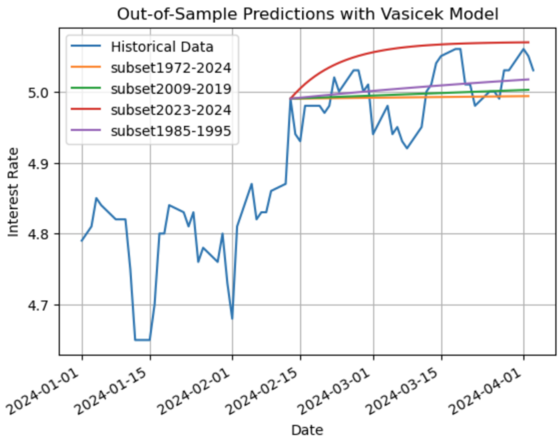

Figure 3: Plotting the curves with the lowest RMSE against the actual data

6.3 Forecasting US Interest Rates

6.3.1 1-Step-Ahead Forecast

Considering a one-step forecast horizon and using the estimated parameters from the subset 2023-2024, where i corresponds to *2024-04-30* and $i-1$ to *2024-04-29*, the expected interest rate value on *2024-04-30* is

$$E[R_{t_i} \mid R_{t_{i-1}}] = R_{t_{i-1}} e^{-adt} + b(1 - e^{-adt}) = 5.20 \cdot e^{-0.102} + 5.07 \cdot (1 - e^{-0.102}) = 5.187.$$

Here, we have the forecasted interest rate of 5.187 for *2024-04-30*, which closely aligns with the actual value of 5.25. It's worth noting that at the time of writing this thesis, *2024-04-30* was still in the future.

6.3.2 Bootstrap Confidence Interval

In this section, we conduct an analysis through simulating 1000 paths of the Vasicek model

$$R_{t_i} \mid R_{t_{i-1}} = R_{t_{i-1}} e^{-a} + b(1 - e^{-a}) + \sigma \int_{t_{i-1}}^{t_i} e^{-a(t-u)} dW_u,$$

using the bootstrap technique. The parameters chosen for the simulation are the same as 1-Step-Ahead Forecast, namely $a = 0.102$, $b = 5.07$ and $\sigma = 0.232$.
The bootstrap simulation technique draws bootstrap samples by sampling from the estimated function instead of the real function, which is not known in our case. We perform parameters estimation for every simulated path and use the results to compute our 1-Step-Ahead Forecast. After repeating the process for 1000 paths, we will obtain 1000 forecasts for the US interest rate for *2024-04-30*.

We compute statistics of interest, specifically the 95 percent confidence interval of the forecasts and the distributions of our estimates of the three parameters. The 95 percent confidence interval lies in $(5.165, 5.208)$. Our findings indicate that the result 5.187 from the last section falls within the interval, implying a 95 percent confidence level in the accuracy of our estimate.

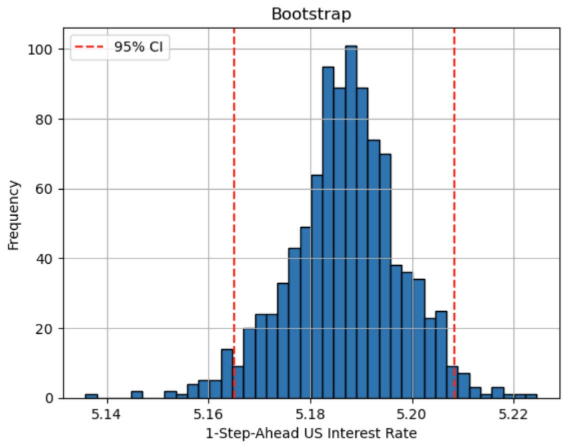

Figure 4: The plot showing 95 percent confidence intervals for the forecast

The distributions of the three estimated parameters are depicted in the figure below. Upon examination of the plots, it is apparent that the parameter estimates used in our model also lie within the 95 percent confidence interval, indicating a 95 percent confidence level in the accuracy of parameter estimation.

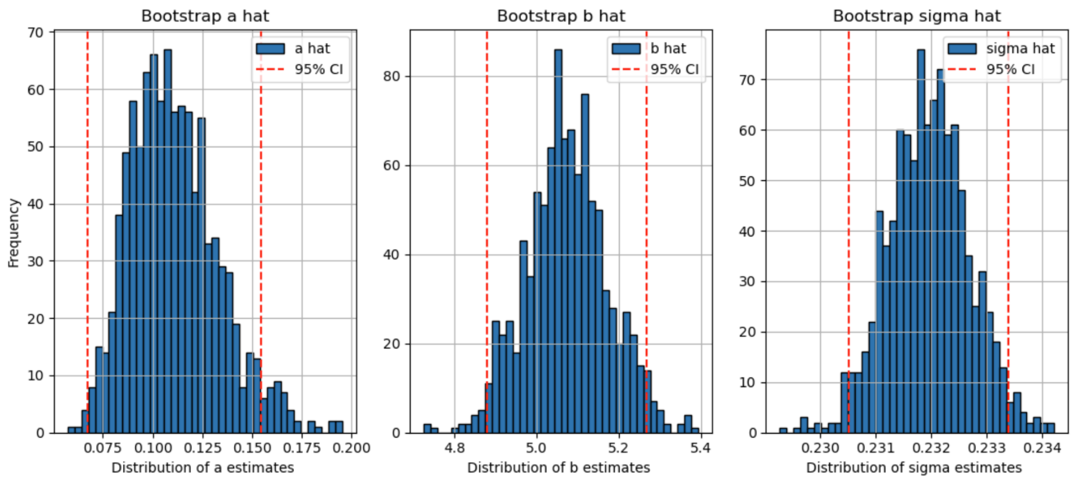

Figure 5: The plot showing the distributions and 95 percent confidence intervals for the parameter estimation

7 Discussion

Based on the Vasicek model, we have inferred that the one-step horizon interest rate in the USA is approximately 5 percent. While our thesis primarily focuses on the mathematical and statistical aspects of the Vasicek model, it is important to acknowledge that we have not constructed a yield curve or derived bond prices, which are essential for comprehensive interest rate forecasting. However, these elements could be incorporated by applying a similar methodology to various treasuries and bonds. Despite its potential, this specific aspect falls outside the scope of our thesis, as our primary emphasis lies in the mathematical and statistical application of the Vasicek model. This limitation highlights the inherent uncertainty in our predictions, although our methodology provides a robust framework.

Upon reflection, it becomes evident that there may be alternative statistical methods capable of providing more optimal parameter estimates. Our assumption of constant parameters has constrained our ability to capture historical data from a dynamic perspective. Exploring statistical techniques that address this limitation could potentially lead to more accurate predictions.

This process has provided valuable insights into the role of mathematics and statistics in applied science. Despite encountering challenges arising from the absence of advanced mathematical concepts such as stochastic calculus, this experience has offered an opportunity to translate theoretical knowledge into practical applications.

References

[1] Vasicek one factor. open source modelling MIT, 2022.

[2] Damiano Brigo and Fabio Mercurio. Other interest-rate models. *Interest Rate Models Theory and Practice*, pages 369–374, 2001.

[3] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 1991.

[4] Axel Gerebrink, Justin Lundgren, Fredrik Malmström, and Oscar Thorén. Maximum likelihood calibration of the vasicek model to the swedish interest rate market. 2019.

[5] Leonhard Held and Daniel Sabanés Bové. Likelihood and bayesian inference. *Statistics for Biology and Health. Springer, Berlin, Heidelberg,*, 2020.

[6] Oliver Ibe. *Markov processes for stochastic modeling*. Newnes, 2013.

[7] Martin Jacobsen. Laplace and the origin of the ornstein-uhlenbeck process. *Bernoulli*, 2(3):271–286, 1996.

[8] Thomas Møller and Mogens Steffensen. *Market-valuation methods in life and pension insurance*. Cambridge University Press, 2007.

[9] Benjamin Neander and Victor Mattson. An attempt at pricing zero-coupon bonds under the vasicek model with a mean reverting stochastic volatility factor, 2023.

[10] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[11] Giuseppe Orlando, Rosa Maria Mininni, and Michele Bufalo. Interest rates calibration with a cir model. *The Journal of Risk Finance*, 20(4):370–387, 2019.

[12] Giuseppe Orlando, Rosa Maria Mininni, and Michele Bufalo. Forecasting interest rates through vasicek and cir models: A partitioning approach. *Journal of Forecasting*, 39(4):569–579, 2020.

[13] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.