NTNU

Norwegian University of
Science and Technology

Evaluating the predictive properties of the Markov-switching jump diffusion LIBOR Market Model

Truls Brubak

Eivind Berg Fosse

Master of Science in Physics and Mathematics
Submission date: June 2017
Supervisor: Jacob Laading, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

We would like to thank our supervisor Jacob Laading for productive discussions and constructive feedback. A special thanks goes to Lea Steinrücke for kindly responding to our email and explaining and providing details on her work on the LIBOR market model.

Abstract

This thesis investigates the predictive properties of the Markov-switching jump diffusion LIBOR market model. A numerical scheme obtaining forward LIBOR rate forecasts is specified by letting the parameters in a jump diffusion LMM, specified under a marked Poisson point process, be dependent on an underlying continuous time Markov chain and then asserting an Euler scheme. The final objective is to see whether adding a Markov-switch component will significantly affect the predictions regarding counterparty risk and kurtosis. Further, an ordinary log-normal LIBOR market model is implemented in addition to a jump diffusion LMM such that the results can be properly compared. Procedures for calibrating the MSJD to market data is described in detail. Finally, crude forms for backtesting is done for the risk measure potential future exposure, in addition to considerations on kurtosis.

Sammendrag

I denne masteroppgaven undersøker vi egenskapene til en Markov-switching jump diffusion LIBOR market model. Vi spesifiserer en numerisk metode for å simulere fremtidige distribusjoner av forward-renter. Dette gjøres ved å la parameterne i en jump diffusion LMM, speisifsert ved en marked Poisson point prosess, være avhengig av en underliggende Markov-kjede. Det endelige målet med oppgaven er å se hvordan denne Markov-switching kompontenten med tilhørende spesifisert kalibreringsrutine vil påvirke prediksjonene i form av kurtose og beregning av motpartsrisiko. For å undersøke dette, implementerer vi også en log-normal LMM, i tillegg til en jump diffusion LMM for å sammenligne modelloppførselen. Prosedyrer for kalibrering for å kunne simulere med modellen er beskrevet i detalj. Til slutt utfører vi en grov backtesting.

Contents

1 Introduction

Late fall 2016 the outstanding notional amount in the over-the-counter (OTC) derivatives market totaled just short of 500 trillion USD [1]. In comparison, the total market capitalization of the global stock market was 65 trillion USD [2]. Of the derivatives in the OTC market, 79 % of them was on interest rates. With these numbers, the need for models that manage to capture some of the dynamics of these interest rates is clear [3]. However, derivatives on interest rates are especially difficult to model, compared to other financial instruments, since a whole family of underlyings has to be modeled. Thus interest rate modelling presents some unique challenges for researchers in mathematical finance [3].

Over the years, several different mathematical models for describing interest rate dynamics have been proposed. The first widely used methods were so-called one-factor models, where only the rate of shortest maturity determine the evolution of the whole yield curve. Interest rates of different maturities can be calculated from market prices on instruments such as bonds. Thus, the theoretical yield curve from a one-factor model will most likely contradict the current market prices [4]. In the late 1980's, there was a break-through as Heath et al. introduced a multi-factor model based on instantaneous forward rates (HJM) that was able to reproduce the yield curves quoted in markets [5]. Later came the LIBOR market model (LMM) based on the same principle as the HJM model, the absence of arbitrage, but using simple rates [5]. Here, the diffusion is often assumed constant, which results in the model producing log-normal forward rates. This assumption means that calibration from Black's formula [6] to market prices can be done with ease. As stated in the book of Glasserman [7], the LMM is gaining rapid acceptance in the financial industry.

However, the assumption of constant diffusion has been widely problematized in recent years. As pointed out numerous times, see fore example Rebanato [5], Johannes [8] and Das [9], there is empirical evidence for leptokurtosis in interest rates. That is, the historical distributions have fat tails, something the log-normal LMM has been failing to properly reproduce. Further, it is also unable to reproduce volatility smiles[10]. The leptokurtic distributions observed in markets are often credited to two phenomena: Jumps in interest rates and shifts in economic periods. Empirical evidence of jumps is presented in several papers such as Babbs and Webber [11] and El-Jahel [12], whilst the existence of different economic regimes with different volatilites in interest rates is put forward in Rebanoato and Joshi [13]. Hence, a natural improvement to the log-normal LIBOR Market Model is to incorporate forms of diffusion that allows for these phenomena. Glasserman and Kou (2003) [14] describes a market model with jumps based on marked point process. Further, Jamishidan

(1999) [15] based on the working paper of Glasserman and Kou, defines a more general jump process using semi-martingales [15]. Finally, to include different economic regimes, Steinrücke et al. (2014) [16] let the parameters in the jump diffusion model be dependent on an underlying Markov chain.

Whereas Steinrücke et al. [16] focus on calibration and analytic pricing formulae for liquid instruments, in this thesis we will rather investigate the forward LIBOR forecast distributions of a Markov-switching jump diffusion LMM. The goal is then to see whether adding a Markov-switch will make the LMM reproduce the heavier tails observed in markets more accurately. To this end we will obtain distributions of all three variations of the LMM to evaluate the benefit of adding more complexity. Glasserman and Merner (2004) [10] developed numerical schemes for the jump diffusion LMM using MPPs. Thus we will augment these schemes with the Markov-switching component described in Steinrücke et al. [16] to obtain forecast distributions from a MSJD LMM. The calibration routines of Steinrücke is not directly applicable since we will not calibrate the model to caps. Thus we propose an alternative, though highly similar routine to obtain our necessary parameters.

Interest rates models are often used to evaluate the risk exposure associated with a portfolio of interest rate derivatives, where the distribution tails are especially important [17]. A model is hence highly interesting in a risk context if it is able to reproduce the tails seen in the markets more accurately [18]. A crude backtest for the risk measure PFE will be implemented to evaluate model performance were we compare the MSJD to the log-normal LMM and jump-diffusion LMM. Our final goal will be to forecast said potential future exposure, with all the above variations of LMM and observe potential differences.

This thesis is divided into 8 chapters. First we state the necessary mathematical preliminaries and briefly describe essential concepts in financial markets. Then we describe the models, their schemes and calibration procedures in chapter 4. In chapter 5 we analyze our data and show estimated jumps and switching routines. In chapter 6 we present the results from the backtesting and compare the forecasted potential future exposures for a given portfolio of derivatives. Finally we present the discussion and conclusion in chapter 7 and 8.

2 Mathematical Preliminaries

To derive the models, calibration routines and numerical schemes in this thesis we will make use of several mathematical concepts. For the sake of rigour, this section will briefly state the key results and definitions in stochastic processes in finance, probability measures and Bayesian hierarchical models that will be used in this thesis. We will avoid overly theoretical and general definitions.

2.1 Stochastic processes in finance

The diffusion in the different specifications of the LIBOR market model make use of different stochastic processes. However, they all contain the Wiener process.

Definition 2.1 (**Wiener Process**)**.** A Wiener process W_t is a continuous-time stochastic process with the following properties:

- $W_{t+u} - W_t, u \geq 0$ is independent of $W_s, s \leq t$.
- $(W_{t+u} - W_t) \sim \mathcal{N}(0, u)$.
- W is continuous with probability 1.

Further, The wiener process can be viewed as a limit of the well-known random walk by letting the number of steps tend to infinity and is sometimes called standard Brownian motion.

The jump-diffusion extension will in addition make use of marked point processes. Simply put, a marked point process(MPP) associates marks to a point process. The formal definition of a general point process is rather involved, thus we will demonstrate the concept through the Poisson point process on the real line as this will be the most relevant for this thesis.

Definition 2.2 (**Poisson Point Process**)**.** We consider the homogenous Poisson point process with intensity parameter where $\lambda > 0$. Let $N(a, b)$ be a Poisson random variable denoting the number of points occurring in the interval $(a, b]$ and n be some counting number. Then we have

$$P[N(a,b) = n] = \frac{(\lambda(b-a))^n}{n!} e^{-\lambda(b-a)}.$$

In other words $N(a, b) \sim Poisson(\lambda(b - a))$.

Note that $N(\cdot)$ is integer valued and that it represents a random measure[19]. Further, the Poisson point process has some nice properties; The number of points in each finite interval has a Poisson distribution and further letting C and D be disjoint intervals, then $N(C)$ and $N(D)$ are independent random variables. Following Johannes et al. [8] these properties yield that for a small enough interval $(a, b]$, $P(N(a, b) = 1) \approx (b - a)\lambda$ and $P(N(a, b) = 0) \approx 1 - (b - a)\lambda$.

We now define the marked Poisson point process used in this thesis.

Definition 2.3 (**Marked Poisson Point Process**)**.** We define the marked Poisson point process as a sequence $\{(\tau_j, X_j)\}_{j=1,2...}$, where τ_j follows a Poisson point process, and $X_j \in \mathbb{R}$ is called a mark.

The Markov-switching extension will additionally require a Markov process.

Definition 2.4 (**Markov processes**)**.** A Markov process is a stochastic process that satisfy the Markov property. The Markov property is defined as follows

$$P(X_{t+1} \mid X_1, \ldots, X_t) = P(X_{t+1} \mid X_t).$$

This imply that the history of observations, except the last, does not affect the probability of the next observation. The Markov chain is a special case of a Markov processes. The discrete-time Markov chain can be described by a set of random variables $X_1, X_2, X_3 \ldots X_n$, that satisfy the Markov property, where the values of X_i form a countable set which represents the state space of the chain. To specify a discrete-time Markov chain we will use a $n \times n$ matrix stating the probability of moving from any pair of states and a an initial state. The continuous time Markov chain is similarly defined where the Markov property for a homogeneous process, meaning the intensity parameter is constant, now becomes $P(X(t + s) = j \mid X(s) = i) = P(X(t) = j \mid X(0) = 1)$. Instead of a transition matrix, a continuous time Markov chain has a transition rate matrix A, also called infinitesimal generator, where the elements q_{ij} of row i and column j satisfy the following properties

- $0 \leq -q_{ii} < \infty$
- $0 \leq q_{ij} : i \neq j \quad \forall\, i$
- $\sum_j q_{ij} = 0 \quad \forall\, i$

Unlike ordinary Markov models, where the states are observable, the states in a hidden Markov Model (HMM) are unobserved, and are thereby described as hidden. However, the hidden state is observed through

another stochastic process, and inference about the hidden state is made through the observable states [20].

Definition 2.5 (**Hidden Markov Processes**)**.** Figure 1 shows an illustration of how a first-order HMM is constructed.

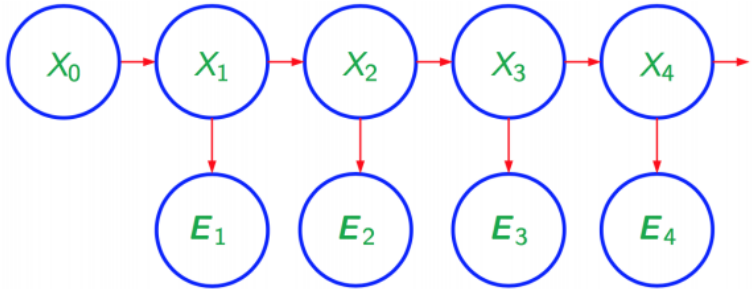


Figure 1: Illustration of a first-order hidden Markov model. X-nodes are unobservable states and E-nodes are observable states.

There is a set of properties embedded in a HMM. Let $X_1, \ldots, X_n$ be the discrete stochastic variables that are unobserved, while $E_1, \ldots, E_n$ are the observable variables, called the evidence. These variables equal one of the states $s_1, \ldots, s_k$. Moreover, a first-order HMM satisfies

- **The first-order Markov property**

$$P(X_{t+1} \mid X_1, \ldots, X_t) = P(X_{t+1} \mid X_t).$$

  As described above, the history of observations, except the last, does not affect the probability of the next observation.

- **Stationary process** The conditional probability $P(X_{t+1} \mid X_t)$ remains the same for all t. In other words, the probability of an observation, given the previous observation, is independent of time.

- **The sensor Markov assumption**

$$P(E_{t+1} \mid X_1, \ldots, X_t, E_1, \ldots, E_t) = P(E_{t+1} \mid X_t).$$

  As long as the previous state of X is known, the previous evidence will not provide any addition information to the probability of a future observation.

Let A be a transition matrix describing the probability of moving between the possible states. The probability of moving from state i to state j is then denoted a_{ij}. That is,

$$a_{ij} = P(X_{t+1} = s_j \mid X_t = s_i), \quad \text{where} \quad a_{ij} \geq 0 \quad \text{and} \quad \sum_j a_{ij} = 1 \ \ \forall\ i, j.$$

Moreover, the initial state distribution of a HMM is defined as

$$\pi_i = P(s_1 = i), \quad \forall\ i \in S,$$

where S is the set of possible states.

There are four inference tasks of interest in HMM's

- **Filtering**

  The aim is to determine the probability of X_t based on the evidence $e_1, \ldots, e_t$: $P(X_t \mid e_1, \ldots, e_t)$.

- **Prediction**

  The probability of a future X based on the evidence $e_1, \ldots, e_t$: $P(X_{t+l} \mid e_1, \ldots, e_t)$, where l is a positive integer.

- **Smoothing**

  Updating the probability of previous values of X based on $e_1, \ldots, e_t$ to improve the estimates: $P(X_l \mid e_1, \ldots, e_t)$, where $0 \leq k < t$.

- **Most likely path**

  Finding the most probable sequence $x_1, \ldots x_t$ based on the evidence $e_1, \ldots, e_t$:

  $\underset{X_1, \ldots, X_t}{\arg\max}\ P(x_1, \ldots, x_t \mid e_1, \ldots, e_t)$.

Central to the interest rate models described in this thesis is the concept of Martingales.

Definition 2.6 (**Martingale**)**.** A discrete-time martingale is a stochastic process described by stochastic variables $X_1, X_2, ..., X_n$ where the following should hold

$$E(X_{n+1} \mid X_1, X_2, ..., X_n) = X_n$$

$$E(|X_n|) < \infty$$

This simply means that given a realized sequence of a martingale, the expectation of the next value is equal

to the last value in the sequence. Martingales are hence closely related to the Markov property. A process is semi-martingale if it can be written as the sum of local martingales.

Similarly we define the continuous time martingale

Definition 2.7 (**Continous-time Martingale**)**.** A continuous time martingale with respect to the stochastic process X_t satisfy:

$$E(Y_t \mid \{X_\tau, \tau \leq s\}) = Y_s \quad \forall s \leq t$$
$$E(|Y_t|) < \infty$$

If $P(Y_t > 0) = 1 \ \forall t$, we say the martingale is strictly positive

2.2 Bayesian Hierarchical Models

Estimating parameters when there are several levels in model can be done with Bayesian Hierarchical models.

Definition 2.8 (**Bayesian Hierarchical Models**)**.** Let $\boldsymbol{x}$ be the data, $f(\cdot)$ denote a probability density function. Further, let $\boldsymbol{x} \mid \theta \sim f(\boldsymbol{x} \mid \theta)$, called the likelihood function, and $\theta \sim f(\theta)$, called the prior distribution. Then, by Bayes' rule

$$f(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x}\theta) f(\theta),$$

which is the posterior distribution. This can be exploited to find joint posteriors, such as

$$f(\theta, \gamma \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \theta, \gamma)\, f(\theta, \gamma)\, f(\theta).$$

One can select the prior distribution such that is in the same family as the posterior

Definition 2.9 (**Conjugate Priors**)**.** If the posterior $f(\theta \mid \boldsymbol{x})$ is in the same family as the prior $f(\theta)$, then the prior is conjugate.

If the posterior distribution is not parametric, one can use Markov Chain Monte Carlo methods for sampling. A special case is called Gibbs sampling

Definition 2.10 (**Gibbs Sampling**)**.** Given a joint distribution $p(x_1, \ldots, x_n)$ we aim to obtain samples $\boldsymbol{X^{(i)}} = (x_1^{(i)}, \ldots, x_n^{(i)})$. This can be done with the following the procedure:

- Set initial values $X^{(i)}$
- Now sample $x_j^{(i+1)} \forall j$ by conditioning the joint distribution on the rest of the parameters, i.e. $p(x_j^{(i+1)} \mid \bar{x}^{(i)})$.

Here $\bar{x}^{(i)}$ represents all observations except $x_j^{(i)}$. Repeating this procedure k times will produce k samples. If the conditional joint distribution $p(x_j^{(i+1)} \mid \bar{x}^{(i)})$ is parametric, the procedure is straightforward.

The samples can be used for parameter estimation through Monte Carlo estimation. First we define:

Definition 2.11 (**Central Limit Theorem**)**.** Let X be a random variable with mean $\mu < \infty$ and variance $0 < \sigma^2 < \infty$. Consider the independent, identically distributed random sample $\boldsymbol{X}$ of size n. Then, the central limit theorem provides that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

That is, the sample mean of $\boldsymbol{X}$ approaches the normal distribution with mean μ and variance σ^2/n.

Definition 2.12 (**The Weak Law of Large Numbers**)**.** Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Furthermore, let $\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$. The weak law of large numbers then states that

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1, \tag{2.2.1}$$

for every $\epsilon > 0$.

We are now ready to define Monte Carlo simulation. The idea behind Monte Carlo methods is the relationship between probability and volume [7]. More precisely, an event is one of the possible outcomes in a sample space, and the probability of a specific event is its volume relative to the sample space. Monte Carlo simulations are thereby performed by drawing random samples from a given sample space and then find the fraction of outcomes that fall in a given set as an estimate of the volume of the set. By the weak law of large numbers, equation (2.2.1), this estimate will converge to the true value as the number of draws increases. Moreover, the central limit theorem from equation (2.11) indicates the error made in from this estimation when a finite number of samples are drawn.

Definition 2.13 (**Monte Carlo**). To illustrate Monte Carlo simulations, consider the integral

$$\alpha = \int_0^1 f(x)\,dx, \tag{2.2.2}$$

for an arbitrary function $f(x)$. This coincide with the expected value of $f(U)$, with U being uniformly distributed from 0 to 1. Let $U_1, \ldots, U_n$ be independent, uniformly distributed samples from $[0, 1]$. The estimated mean of $f(U)$ is

$$\hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i). \tag{2.2.3}$$

Provided that f is integrable in the interval $[0, 1]$, the weak law of large numbers from equation (2.2.1) ensures that

$$\lim_{n\to\infty} \hat{\alpha}_n \to \alpha, \tag{2.2.4}$$

with probability 1. Moreover, assume f is twice integrable and define

$$\sigma_f^2 = \int_0^1 (f(x) - \alpha)^2\,dx, \tag{2.2.5}$$

and the quantity $\hat{\alpha}_n - \alpha$ to be estimation error. The error will approach a zero-mean normal distribution with standard deviation $\sigma_f/\sqrt{n}$ as n grows. Even though σ_f can be difficult to access, it might be estimated by the unbiased sample standard deviation

$$S_f = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(f(U_i) - \hat{\alpha}_n)^2}. \tag{2.2.6}$$

Hence, the function f evaluated at $U_1, \ldots U_n$ estimates α and additionally provides the estimation error.

The LMM specifications contain the derivative of continuous-time stochastic variables. Itô's lemma relates small changes in a function of a random variable to the small change in the random variable itself and is typically used to find the necessary derivatives in interest rate models.

Lemma 2.1 (**Itô's Lemma**). *Consider a smooth function f of a random variable G, where G is described by a stochastic differential equation*

$$dG = a(X, t)dX + b(G, t)dt. \tag{2.2.7}$$

By applying Itô's lemma, $f(G)$ can be expressed as

$$df = a\frac{df}{dG}dX + \left(b\frac{df}{dG} + \frac{1}{2}a^2\frac{d^2f}{dG^2}\right)dt. \tag{2.2.8}$$

Finally we formalize kurtosis.

Definition 2.14 (**Kurtosis**)**.** Kurtosis can loosely be desribed as a measure of how heavy the tails of the respective distribution is. More formaly it is defined as the fourth standardized moment of a random variable.

$$Kurt[\boldsymbol{X}] = \frac{E[(\boldsymbol{X} - \mu)^4]}{(E[(\boldsymbol{X} - \mu)^2])^2}$$

A distribution is leptokurtic if it has greater kurtosis than the normal distribution, which always has a kurtosis of 3.

2.3 Change of measure

Central in the derivation of interest rate models is the concept of changing random measures to obtain different dynamics. For a formal definition of random measures we refer to the text book by Albert Shiryaev[19].

Definition 2.15 (Probability measure)**.** Let μ denote a function, if μ satisfy the following, it can be considered a probability measure

- μ must return results in the unit interval [0,1] where 1 is returned for the whole interval and the empty set for 0.
- μ must satisfy the property of countable additivity where all countable collections of $\{E_i\}$ of pairwise disjoint sets we have that $\mu\left(\bigcup_{i\in I}\right) = \sum_{i\in I}\mu(E_i)$

Two measures might be considered equivalent.

Definition 2.16 (Equivalent measures)**.** Let μ, θ be two signed measures where $\mu, \theta \to \mathbb{R}$. The measure μ is considered equivalent to θ if they are absolutely continuous with respect to each other: $\mu \ll \theta \ll \mu$

Grisanov's theorem describes how the dynamics of a stochastic process is changed when the measure is changed to another equivalent measure. We will state the theorem in relation to the Wiener process due to the relevance in finance.

Theorem 2.2 (Grisanov's theorem)**.** *Let W_t be a Wiener process and let X_t be a measurable process where a natural filtration $\mathcal{F}_t$ is adapted to W_t with $X_0 = 0$. Now define $\mathcal{E}([X]_t) = exp\,(X_t - 0.5[X]_t)$. Further if $\mathcal{E}(X)_t$ is a strictly positive martingale, then we can define a probability measure Q such that the Radon-Nikodyn derivative is*

$$\frac{dQ}{dP} \mid_{\mathcal{F}_t} = \mathcal{E}(X)_t$$

The measure Q for each t restricted to the fields $_t$ is now an equivalent measure to P restricted on $\mathcal{F}_t$.

Finally we define the Radon-Nikodym derivative in the

Definition 2.17 (Radon-Nikodym Derivative)**.** Given a measurable space (X, Σ) and a σ-finite measure ν that is absolutely continuous with respect to a similar measure μ on the same space, then there exists a measurable function $f : X[0, \infty)$ such that

$$\nu(A) = \int_A f d\mu \tag{2.3.1}$$

where $A \subset X$ is some measurable space.

3 Introducing Interest Rate Modelling and Risk Management

In this section we will give a brief introduction to interest rate modelling and risk management. Further we will state basic concepts from finance and the necessary preliminaries related to interest rate derivatives and relate the theory of probability measures to its applications in mathematical finance.

3.1 Interest Rates and Financial Instruments

The most elementary notion in this thesis is interest rates.

Definition 3.1 (**Interest Rate**)**.** Interest, sometimes referred to as yield, is the cost of borrowing capital or the compensation for setting money at the disposal of others. The simple interest rate refers to the cost as a fraction of the amount of capital borrowed that has to be paid at the when the loan ends while the compounded interest rate can be defined more mathematically: Given a deposit of 1, n interest payments and an annual rate of r/n. At time t, the deposit's value after T years equals

$$C(t) = \left(1 + \frac{r}{n}\right)^{n(T-t)}.$$

Many models make use of the continuously compounded rate where the value is obtained by letting $n \to \infty$ giving

$$C(t) = e^{r(T-t)}, \tag{3.1.1}$$

As the interest rate may vary during the time period, we let

$$C(t) = e^{\int_t^T r(\tau)\, d\tau}. \tag{3.1.2}$$

An example of a widely used interest rate is The London Interbank Offered Rate (LIBOR).

Definition 3.2 (**LIBOR**)**.** The London Interbank Offered Rate represents the average of interest rates estimated by each of the leading banks in London that it would be charged were it to borrow from other banks. LIBOR is quoted daily in several currencies for several maturities.

The LIBOR rate is a reference rate, meaning that is often used as an benchmark to determine the prices of certain financial instruments.

Definition 3.3 (**Financial Instrument**)**.** Financial instruments are loosely defines as monetary contracts between a certain numeber of parties. They can often be traded in a market.

We will mainly focus on one such type of instrument called derivatives.

Definition 3.4 (**Derivatives**)**.** A derivative is a contract where the value is derived from an underlying. The underlying may refer to several entities such as stocks or interest rates. An interest rate derivative simply refers to a derivative with a reference rate as underlying. The most basic and widely traded derivatives are often called "vanilla"- derivatives. The term "exotic" would refer to the opposite.

One of the most basic and widely traded interest rate derivatives is the interest rate cap. In general, caps are traded to reduce risk exposure regarding upward movements in interest rates.

Definition 3.5 (**Cap**)**.** An interest rate cap is a contract that generates yield when the interest is above an agreed strike price. Cap contracts agrees upon a certain interest rate reference, such as the 6-month LIBOR. Additionally, the contract also contains a notional value and payment times. A Cap is a portfolio of caplets. That is, the total lifespan is divided into subsequent time periods, where each period has a related caplet. The payment of caplet number i, C_i, at its maturity T_i is given by

$$C_i(r, T_i) = N\alpha_i \max(r_R(T_i) - r_S, 0), \tag{3.1.3}$$

where N is the notional value, α_i is the day count fraction corresponding to period i, r_R is the reference rate, like LIBOR, and r_S is the strike rate. The holder often receives the payment at the end of the period, i.e. at time T_{i+1}.

Since a cap consists of caplets, its payout is dependent on the level of the reference rate at several points in time in the future. If the strike rate is set equal to today's forward rate, the caplet is said to be "at the money" (ATM). An interest rate *floor* works very similar to a cap where the payout is given when the LIBOR rate goes *below* the strike rate. Hence the price becomes

$$F_i(r, T_i) = N\alpha_i \max(r_S - r_R(T_i), 0).$$

Another important financial instrument in this thesis is bonds.

Definition 3.6 (**Bonds**)**.** A bond is a fixed-income contract paid in advance. The issuer of the bond receives a payment, called notional, from the now holder of the bond. The contact then yields a predetermined amount

to the holder at a given date called the maturity date. Further, a bond may be coupon-bearing, meaning that it also pays cash dividends at fixed times. Without these intermediate coupon payments, the bond is called a zero-coupon bond. Major governments usually issue bonds with a fixed set of maturities, e.g. 5-year, 10-year and 30-year bonds.

From a more pragmatic perspective, the issuer of the bond received a loan from the holder, equalling the notional, that is paid back with interest at the maturity date. The majority of bonds are possible to trade in a secondary market and the prices of bonds issued buy large governments are quoted daily. Often the annual interest rate on the bonds is quoted rather than the prices. If you were to plot the different yields of the bonds of different maturities quoted on the same day, you would obtain a yield curve, also called term structure. The prices of the bonds can be obtained from the yields.

Definition 3.7 (**Zero Coupon bond**)**.** Let $Z(t;T)$ denote the value of a zero-coupon bond at time t and F be its face value. $Z(t;T)$ must satisfy

$$Z(t;T) = Fe^{-\int_t^T r(\tau)d\tau}, \tag{3.1.4}$$

where the interest rate r is supposed to be known at all times $t < T$.

These bond prices are often seen in the context of forward rates.

Definition 3.8 (**Forward rates**)**.** A forward rate is the interest rate for borrowing or lending at a future time[7]. More specifically, Let $f_r(t, T_1, T_2)$ denote a forward rate based on simple interest, which is set at time t and applies in the time interval $[T_1, T_2]$. The forward rates can be calculated using the prices of zero coupon bonds

$$f_r(t, T_1, T_2) = \frac{1}{T_2 - T_2} \frac{Z(t, T_1) - Z(t, T_2)}{Z(t, T_2)}, \tag{3.1.5}$$

Proof: If a loan of 1 is taken out at T_1 and repaid at T_2, the amount repaid at T_2 is $1 + f_r(t, T_1, T_2) \cdot (T_2 - T_1)$. This loan could be replaced by buying and issuing zero-coupon bonds. Let $Z(t, T_1)$ and $Z(t, T_2)$ be the price of zero-coupon bonds bought at time t and maturing at time T_1 and T_2, respectively. The lender needed 1 at T_1, which can be attained by buying the zero-coupon bond $Z(t, T_1)$. However, to make this payment, an amount k of the zero-coupon bond $Z(t, T_2)$ are issued. The following must hold

$$kZ(t, T_2) - Z(t, T_1) = 0 \iff k = \frac{Z(t, T_1)}{Z(t, T_2)}$$

As this construction was made to reproduce the loan described above, the payment at T_1 is 1 and $1 +$

$f_r(t, T_1, T_2) \cdot (T_2 - T_1)$ at T_2. Thereby,

$$k = 1 + f_r(t, T_1, T_2)(T_2 - T_1).$$

Inserting the ladder expression for k yields equation 3.1.5

The above definition of forward rates introduces an important concept in interest modeling: By quoting prices for bonds with maturity T_1 and T_2, the market indicate an interest rate applying between the dates T_1, T_2. Suppose you obtain the forward rate applying in the window from 3 to 4 years in the future using the bond prices quoted in the market to day. This forward rate should represent the markets prediction for the 1-year interest rate occurring in 3 years. This interest rate is particularly interesting if you have a portfolio of interest derivatives which payout are directly dependent on the rate set in this exact window. However, interest rate derivatives are dependent on reference rates and not directly on bond prices. An idea is thus to use the forward rates calculated from bond prices as a proxy for future LIBOR rates[5]. These proxies are called the forward LIBOR rates.

Definition 3.9 (**Forward LIBOR rates**)**.** Let δ denote the accrual period of a LIBOR rate, which is often fixed. Equation (3.1.5) for forward LIBOR rates $L(t, T)$ becomes

$$L(t, T) = \frac{1}{\delta} \frac{Z(t, T) - Z(t, T + \delta)}{Z(t, T + \delta)}. \tag{3.1.6}$$

From the above discussion the distinction between the actual LIBOR rates set by a group of banks and the forward LIBOR rates calculated from bond prices should be clear. In this thesis we will almost exclusively refer to the forward LIBOR rates and the context should make it clear if the actual LIBOR rate is referred.

Given a set of forward rates, the future payouts of caps and floors are discounted by the estimated set of forward rates $\hat{\mathbf{L}}$ in the following way under the spot measure

Definition 3.10 (**Deflating Caps**)**.** The price of a cap is found by summing over all its caplets where each caplet should be discounted

$$C_i(\hat{\mathbf{L}}, T_i) = N\alpha_i \max(L_i(T_i) - r_S, 0) \prod_{j=0}^{i-1} \frac{1}{1 + \delta_j \hat{L}_j(T_J)} \tag{3.1.7}$$

similar for floors we have

Definition 3.11 (**Deflating Floors**)**.** The price of a floor is found by summing over all its caplets where each caplet should be discounted

$$F_i(\hat{\mathbf{L}}, T_i) = N\alpha_i \max(r_s - L_i(T_i), 0) \prod_{j=0}^{i-1} \frac{1}{1 + \delta_j \hat{L}_j(T_J)} \tag{3.1.8}$$

3.2 Stochastic interest rate models and change of Numeraires

Interest rates are often assumed to be stochastic [5]. Arguably the simplest and most general way of modelling the future course of the interest rate $r(t)$ is by treating it as a random variable following a random walk described by

$$dr = w(r,t)\, dX + u(r,t)\, dt, \tag{3.2.1}$$

where $w(r,t)$ and $u(r,t)$ are functions explaining the behaviour of the short rate $r(t)$ [21]. Models on this form are hence one-factor short rate models since they use a univariate stochastic variable. The short rate is the interest rate received by the shortest possible deposit. One of the most common one-factor models for short rates is the Vasicek model.

Definition 3.12 (**Vasicek**)**.** The Vasicek model describes the evolution of the interest rate by the following equation

$$dr_t = \alpha(b - r_t)\, dt + \sigma dW_t, \tag{3.2.2}$$

where r_t is the interest rate at time t, σ is the volatility, dW_t is a wiener processs, b is the long-run mean level, α is the speed of reversion.

Further assuming that α and b are positive constants, this model is clearly mean reverting. The Vasicek model can thus be shown to satisfy two nice properties for short rate models: Always positive, and mean reverting[21]. The definition of the one-factor short rate model does not include forward rates calculated from bond prices. Hence as discussed in Wilmot [21] the theoretical yield curve from a one-factor model will most likely contradict the current market prices. The parameters may, though, be fitted to an observed yield curve [21].

In contrast to short rate models, one of the main principals in modern interest rate models such as the LMM is the absence of arbitrage [3].

Definition 3.13 (**Arbitrage**)**.** Arbitrage is the possibility of an instantaneous risk-free profit and often

arise due to price differences in markets.

Since efficient markets are supposed to be free of arbitrage, it is natural to impose the same property on the model[3]. In the context of interest rate modeling, absence of arbitrage implies that the model should be martingale under the chosen numeraire[7].

Definition 3.14 (**Numeraire**)**.** An asset with strictly positive prices that denominate other assets.

Here asset can refer to any financial instrument. This means that the when choosing a numeraire, we also choose the discount factor to be dependant on the asset chosen as numeraire. The theory of measure changes now demonstrates its utility. Choosing an asset S_d as numeraire involves defining a new probability measure P_{S_d} using the original measure P_β together with the Radon-Nikodyn theorem.

$$\frac{dP_{S_d}}{dP_\beta} = \frac{S_d(t)}{\beta(t)} \Big/ \frac{S_d(0)}{\beta(0)} \tag{3.2.3}$$

Here $S_d(t)/\beta(t)$ is assumed to be positive martingale and thus dividing by the initial values defines a change of measure [7].

When estimating parameters straight from empirical data the term "real-world"measure is often used [7]. The "physical" measure refers to the same concept. If the bond of shortest maturity is used as numeraire, the measure is called the "spot" measure. If any of the other bonds maturing at a later time T_f is used as numeraire, the measure is called the "forward" measure. A LIBOR market model is usually specified to be martingale, and thus free from arbitrage, under one of these measures [5].

3.3 Volatility and Risk Measures

In finance, the term volatility with respect to an instrument usually refers to the variance of the price of the respective instrument.

Definition 3.15 (**Volatility**)**.** Loosely defined volatility is the annualized variance of the price of a financial instrument.

The concept of implied volatility is central in financial mathematics and arises when the price of a derivative is a function of the volatility of the underlying. A classic example of this is Black's formula [6] for pricing an interest rate cap. Given all other parameters is known, we can simply solve the formula with respect to the volatility.

Definition 3.16 (**Implied volatility**)**.** Implied volatility is the volatility found by using a pricing formula that contains volatility as one of the variables.

Thus implied volatility is a function of current market prices on the respective derivative. These prices often reflects investors predictions of future volatility [5]. There will often be events, political etc., that will change the volatility of a market. By simply looking at the historic volatility for the last 2 years, say, you will marginalize these recent events. However, if the derivative is highly liquid, recent market prices will reflect these events, thus the implied volatility will also reflect these events. This is why implied volatility is often preferred to historical when forecasting future derivative prices [5].

Closely related to volatility is the notion of risk. In finance, managing risk is central. The Basel Committee on Banking Supervision provide regulations and for Banks and trading desks regarding risk exposure[22].

Definition 3.17 (Basel Committee)**.** The main role of The Basel Committee on Banking Supervision is setting capital adequacy requirements where the required capital is dependent on the level of risk exposure. Procedures for calculating risk exposure is outlined in many of their documents.

Two types of risk defined in the Basel documents that is relevant for this thesis is market risk and counterparty risk.

Definition 3.18 (**Market Risk**)**.** Market risk is the risk associated with future losses arising from movements in market prices [22].

Thus there will be market risk associated with a portfolio of assets and and derivatives on assets. Further counterparty risk is defined as

Definition 3.19 (**Counterparty Risk**)**.** The risk associated with a counterparty defaulting and not fulfilling their contract.

When evaluating levels of risk exposure, measures of risk is needed. There are certain properties that are desirable in a risk measure.

Definition 3.20 (**Coherent Risk Measure**)**.** Let $Z_1, Z_2 \in S$ where S is the set of all stochastic processes. Further let the function $\phi : S \to \mathbb{R}$ define a coherent risk measure. Then ϕ satisfy the following properties

- Monotonicity: $\forall\, Z_1, Z_2 \in S$ where $Z_2 \leq Z_1$, then $\phi(Z_1) \leq \phi(Z_2)$. This is best explained with a portfolio of in the money call options. "in the money" call options mean that the strike price is lower than the market price of the underlying. If the options in Z_1 has a lower strike price than the ones in Z_2, meaning they are expected to have a higher yield, then the risk associated with Z_1 should be

lower than the one associated with Z_2

- Translation Invariant: Let λ be a risk-free yield, $\forall\, Z_1 \in S$ and $\forall\, \lambda \in \mathbb{R}$ then $\phi(Z_1 + \lambda) = \phi(Z_1) - \lambda$
- Subadditive: $\forall Z_1, Z_2 \in S$ we have that $\phi(Z_1 + Z_2) \leq \phi(Z_1) + \phi(Z_2)$. This means that adding portfolios will never generate higher risk.
- Positive Homogeneous: $\forall\, Z_1 \in S$ we have that $\phi(cZ_1) = c\phi(Z_1)$ for a positive constant c.

A common risk measure defined in the Basel documents is "Value at Risk"(VaR).

Definition 3.21 (**Value at Risk**)**.** If $\eta(t)$ represents the portfolio value at time t, then

$$VaR_\alpha(\eta) = inf\{x \in \mathbb{R} : P(\eta < -x) \leq 1 - \alpha\}$$

Put simply; VaR is the smallest number x such that the probability that the portfolio value η exceeds -x is at least α. VaR does, however, not have the sub-additive property, meaning that it does not give the proper incentives to diversify a portfolio.

Another risk measure defined in the Basel documents is "Excpected Shortfall"(ES).

Definition 3.22 (**Expected Shortfall**)**.**

$$ES_\alpha = E[-\eta : \eta > VaR_\alpha] = \frac{1}{\alpha}\int_0^\alpha VaR_\alpha(\eta)$$

The expected shortfall is simply the expectancy of the values at risk. In contrast to VaR, ES is a coherent risk measure. The final risk measures from the Basel documents we define is the potential future exposure(PFE) and expected exposure(EE)

Definition 3.23 (**Potential Future Exposure**)**.** The potential future exposure is the exposure of a portfolio over its entire lifetime calculated at some level of confidence. PFE(t) is thus a function of time. The upper quantile of the portfolio value is used.

The risk represented by PFE(t) is due to a counter party might defaulting its debt, thus it is considered a part of counterparty risk. Do note that the PFE in this thesis does not consider the probability of the counterparty defaulting, it considers the risk **given** that the counter party defaults some time in the future. The upper quantile is used since it represents scenarios where others are indebted to you. This measure can be explained more intuitively by an example. Consider an asset, say a stock, that follows a stochastic

process similar to the Wiener process. Then the estimated quantile, say 95%, will naturally rise the further into the future you forecast, since the variance of the Wiener process increases. Calculating potential future exposure associated with a portfolio consisting of such a stock at a specific time t would be very similar to the calculating the Value at Risk where the opposite tail is used. PFE(t) should for this portfolio be a strictly increasing function of time. If however, a portfolio consisting of an instrument such as an interest rate cap is considered, then the exposure related to individual caplets will gradually disappear at certain future time points. An ATM cap will typically have a rising PFE(t) in the beginning as uncertainty rise, and a declining PFE(t) at the end as most of the caplets have rolled off and no longer represent exposure. Expected exposure (EE) is similar to PFE(t) but measures the *expected* counterparty exposure.

Stochastic interest models are often backtested where one consider how well the model determine exposure [23].

Definition 3.24 (Backtesting Risk Measures)**.** Let the respective dataset contain observations from the time interval $[A, F]$ and let the Portfolio P consist of interest rate derivatives. Backtesting risk measures on a stochastic interest model involves calibrating the model using historic data from a period $[B, C]$ where $A \leq B < C$ then using the model to forecast the value of P into the interval $[D, E]$ where $C < D < E \leq F$. The quantiles predicted for the value of P by the model for the interval $[D, E]$ is then compared to the historic prices of the derivatives. Statistical tests then may be used to determine if the quantiles given by the model seems to fit the actual observations[18].

One of the simpler methodologies for backtesting VaR is based on the binomial distribution and uses the test statistic

$$Z = \frac{x - Np}{\sqrt{Np(1-p)}}$$

Here x is the number of failures N the number of observations and $p = 1-$VaR level, where VaR level is given by the model. Further x is a failure when it is below the VaR level. There are several more sophisticated way of testing VaR like Christoffersen's test where the dependency between consecutive days is taken into consideration.

We now outline some challenges linked to backtesting risk measures on interest rate models. First, consider a portfolio of stocks and a stochastic model for stock prices. The portfolio would then contain objects that was directly modeled by the model and the VaR level predicted by the model is directly comparable to the observed stock prices. Now consider a portfolio of interest rate derivatives. An interest rate model

do not model the prices of derivatives directly, they are after all derived prices. The type of instrument contained in the portfolio would most likely change the estimated VaR-level, thus obscuring the back-testing procedure; you are in fact backtesting both the portfolio and the model. Further, consider backtesting the measure PFE(t) stated at different points t_i where $i = 1 \dots M$. As stated above, PFE(t) equals VaR when conditioned on t, hence one could imagine a procedure involving the test statistic for VaR above at all t_i [18]. However, since the estimated VaR levels of the different time points would be correlated, finding the proper statistic, and thus determining if x failed or not, is not straight forward. The exact testing procedures of risk measures linked to portfolios of derivatives used by banks is rather involved [23].

In this thesis backtesting is done using PFE as risk measure. We consider a crude version based on the book of Wedong [23] and the paper by Ng [17]. A portfolio of caps and floors is considered such that both downward and upward interest movements are considered under different stressed periods where the significance of the results will then be discussed.

4 LIBOR Market Models

In this chapter we derive the LMM variations, their respective numerical schemes and describe calibration procedures. Due to the close relation between the LIBOR market models, we will start with the log-normal LMM, add jump diffusion, and finally consider the Markov-switching jump diffusion. This way we gradually increase the model complexity and facilitate thorough understanding. First, we motivate the models by briefly outlining the history of LIBOR market models.

While there exist several interest rate models, variations of the LIBOR Market Model (LMM) are some of the most commonly used [7]. In the literature, as discussed by Rebonato [5], the name LIBOR market model is actually slightly ambiguous. For the sake of rigour, we will use the same definition as given in Glasserman's book [7]: "LIBOR market models are based on simple rates, and describe the arbitrage-free dynamics of the term structure of interest rates through the evolution of forward rates."

4.1 Log-normal LMM

Before introducing the elemental dynamics in the LIBOR market models, we recall equation (3.1.6) where the forward LIBOR rates are defined through zero-coupon bonds

$$L(t,T) = \frac{Z(t,T) - Z(t,T+\delta)}{\delta Z(t,T+\delta)},$$

where $L(t,T)$ is the forward LIBOR rate at time t with a maturity of $T+\delta$. In the remainder of this thesis, the set of maturities are fixed to

$$0 = T_0 < T_1 < \cdots < T_M < T_{M+1},$$

where the distance between all maturities are equal. That is, $\delta = T_{i+1} - T_i$, for $i = 0, \ldots, M$. The most common notation for forward LIBOR rates is letting $L_n(t)$ be the forward LIBOR rate at time t with maturity at T_n. Equation (3.1.6) is rewritten by applying the same notation for zero-coupon bonds

$$L_n(t) = \frac{Z_n(t) - Z_{n+1}(t)}{\delta Z_{n+1}(t)}, \qquad 0 \le t \le T_n, \quad n = 0, \ldots, M. \tag{4.1.1}$$

The constraints on t and n implies that forward LIBOR rate $L_n(t)$ is not valid for $t > T_n$. More explicitly, after its maturity, $L_n(t)$ expires. Hence, as time increases, the number of valid forward LIBOR rates will decrease. In the paper by Brace, Gatarek and Musiela [24], these forward rates are assumed to follow the simple dynamics

$$\frac{dL_n(t)}{L_n(t)} = \mu_n(t)\,dt + \sigma_n(t)^\intercal d\mathbf{W}(t), \qquad 0 \le t \le T_n, \quad n = 1, \ldots, M, \tag{4.1.2}$$

where $\mathbf{W}(t)$ is a d dimensional standard Brownian motion and $\sigma_n(t)^\intercal, \mu(t) \in \mathbb{R}^d$. Hence, equation 4.1.2 represents coupled stochastic differential equations , with drift and diffusion. This specification is also called the log-normal LMM since it results in the forward rates being log-normally distributed.

We will now briefly outline how to derive the arbitrage-free drift term of the log-normal LMM, based on Glasserman's book [7]. We start by finding a suitable numeraire asset for the forward LIBOR rates. Consider the following investment strategy: spend 1 on buying $1/Z_1(0)$ zero-coupon bonds today which matures at T_1 (at time $t = 0$), at T_1 spend the resulting cash on zero-coupon bonds $Z_2(T_1)$ maturing at T_2, and proceed in the same manner. The strategy generates a simple interest of $L_i(T_i)$ for every time interval $[T_i, T_{i+1}]$. Hence, when only considering times coinciding of maturities, the cash flow by the investment strategy $C(T_n)$, at time T_n, is given by

$$C(T_i) = \prod_{j=0}^{n-1} (1 + \delta_i L_j(T_i)), \tag{4.1.3}$$

where $\delta_i = T_{i+1} - T_i$ and $n = i + 1, \ldots, M + 1$. This relationship defines the numeraire asset is based on the bond with the shortest maturity, hence it induces the the spot measure in the LMM.

In order to make the LMM dynamics defined in equation (4.1.2) free from arbitrage under the spot measure, we invert equation (4.1.1) with respect to the zero-coupon bond prices at maturity dates

$$Z_n(T_i) = \prod_{j=i}^{n-1} \frac{1}{1 + \delta L_j(T_i)}, \qquad n = 1, \ldots, M. \tag{4.1.4}$$

The zero-coupon bond prices need to be martingales when normalized by the numeraire asset [7]. Hence, we

divide equation (4.1.4) by equation (4.1.3)

$$D_n(T_i) = \frac{\prod_{j=i}^{n-1}(1+\delta L_j(T_i))^{-1}}{\prod_{j=0}^{i-1}(1+\delta L_j(T_j))}, \qquad n = 1, \dots, M.$$

The term $D_n(T_i)$ is called the deflated zero-coupon bond price. As argued by Jamshidian [25], the deflated zero-coupon bond prices must be positive, which is shown by Glasserman [7] to impose the following expression for the drift

$$\mu_n(L_n(T_i), T_i) = \sum_{j=\eta(t_i)}^{n} \frac{\delta L_j(T_i)\sigma_n(T_i)^\intercal \sigma_j(T_i)}{1+\delta L_j(T_i)}. \tag{4.1.5}$$

Consequently, the LMM dynamics under the spot measure reads

$$\frac{dL_n(T_i)}{L_n(T_i)} = \sum_{j=\eta(t_i)}^{n} \frac{\delta L_j(T_i)\sigma_n(T_i)^\intercal \sigma_j(T_i)}{1+\delta L_j(T_i)}\, dt + \sigma_n(T_i)^\intercal\, dW(T_i), \qquad n = 1, \dots, M. \tag{4.1.6}$$

The function $\eta(t)$ represents the index of the rate maturing next. It can be chosen to be left- or right continuous, where Glasserman and Zhao [26] recommends using right-continuous for smallest discretisation error.

4.1.1 Simulation

To simulate from the dynamics of the log-normal LMM, we discretize equation (4.1.6). As shown by Glasserman and Zhao [26], there are various ways to discretize the dynamics, but in this thesis we choose an ordinary Euler scheme. Moreover, we choose to apply the scheme on $\log(L_n(t))$ rather than $L_n(t)$. A more detailed discussion on this topic can be found in Glasserman's text book [7], we note however that using the logarithm of the forward rates ensure that they stay positive. We obtain

$$\hat{L}_n(T_{i+1}) = \hat{L}_n(T_i)\, \exp\left(\left[\mu_n(\hat{L}_n(T_i), T_i) - \frac{1}{2}||\sigma_n(T_i)||^2\right](T_{i+1}-T_i) + \sqrt{T_{i+1}-T_i}\,\sigma_n(T_i)^\intercal\, Z_{i+1}\right), \tag{4.1.7}$$

for $n = 1, \dots, M$. Z_{i+1} is a standard Brownian motion, $\sigma_n(T_i)$ is the volatility factor of maturity n at maturity date T_i, and $\mu_n(T_i)$ is defined by equation (4.1.5). The simulations are initialized using equation

(4.1.1) for $t = 0$, that is

$$\hat{L}_n(0) = \frac{Z_n(0) - Z_{n+1}(0)}{\delta Z_{n+1}(0)},$$

for $n = 1, \ldots, M$.

Note that due to the properties of the Wiener process, using the scheme 4.1.7, one can scale the time grid without increasing the discritization error [7]. For example, if you aim to price a cap, you only need each forward rate *at* its maturity, i.e. $L_i(t_i)$. Then your time grid could only consist of points at $T_1 \ldots T_M$ since all the time points in between the maturities are skipped by simply scaling the variance. The time grid used in this thesis however will be daily since we will not only price the caps.

4.1.2 Parameter estimation

Common for all our models is that only the parameters related to diffusion is estimated from data. The drift is completely determined from the no-arbitrage condition as seen in the previous section. We now state the volatility structure that will be used throughout the thesis.

The $\sigma_n(t)$ in equation (4.1.2) is called the volatility factors. Following [7], we assume $\sigma_n(t)$ to be constant between tenor dates, i.e. $\sigma_n(t)$ remains constant for $t \in [T_i, T_{i+1}]$. In a one-factor case, $\sigma_n(t)^\intercal \in \mathbb{R}$, the volatility structure looks like [7]

$$\begin{pmatrix} \sigma_1(T_0) & & & \\ \sigma_2(T_0) & \sigma_2(T_1) & & \\ \vdots & \vdots & \ddots & \\ \sigma_M(T_0) & \sigma_M(T_1) & \ldots & \sigma_M(T_{M-1}) \end{pmatrix}$$

This structure can be interpreted in the following manner. When $t \in [0, T_0)$, $L_1(t)$ has volatility $\sigma_1(T_0)$, whilst for $t \geq T_0$, $L_1(t)$ has expired. Further, $L_2(t)$ has a volatility of $\sigma_2(T_0)$ for $t \in [0, T_0)$ and $\sigma_2(T_1)$ for $t \in [T_0, T_1)$. The same structure applies for the rest of the remaining maturities and for all relevant times. This is the rationale behind the lower triangular shape.

Moreover, we let the volatility factors be stationary, causing the $\sigma_n(t)$ to depend merely on the time to maturity. That is, the difference between n and t meaning $\sigma_n(t) = \sigma(n - t)$ [7]. For instance, let $t = T_3$ and

$n = 5$. Then, $\sigma_n(t) = \sigma_5(3) = \sigma(5-3) = \sigma(2)$. The volatility structure then reads

$$\begin{pmatrix} \sigma(1) & & & \\ \sigma(2) & \sigma(1) & & \\ \vdots & \vdots & \ddots & \\ \sigma(M) & \sigma(M-1) & \dots & \sigma(1) \end{pmatrix}.$$

Hence, when $t \in [T_0, T_1)$, $L_1(t)$ behaves as the LIBOR rate of the lowest maturity, while $L_2(t)$ has the properties of a LIBOR rate of the second lowest maturity, and so forth. Once the time reaches T_1, $L_2(t)$ inherits the volatility $\sigma(1)$. In the same manner, $L_3(t)$ gets $\sigma(2)$ as volatility, and so forth.

In this thesis, the correlation between all forward rates are taken into consideration, causing $\sigma_n(t)^\intercal \in \mathbb{R}^M$. We now state the relation between the covariance of the forward rates, and the volatility factors

$$\text{Cov}(L_i, L_j) = \sigma_i(t)^\intercal \sigma_j(t). \tag{4.1.8}$$

The correlation then becomes

$$\text{Cor}(L_i, L_j) = \frac{\sigma_n(t)^\intercal \sigma_j(t)}{||\sigma_n(t)||\,||\sigma_j(t)||} \tag{4.1.9}$$

For the log-normal and jump-diffusion LMM we will weight the covariance- and correlation matrix such that more recent observations becomes more prominent. The notion of focusing on more recent data when forecasting with LMM is described in Rebonato(2002)[5]. To obtain the weighted matrices, a technique called exponentially weighted moving averages will be used. Let $l_{i,k}$ be the log-returns of forward rate i at time k, that is $l_{i,k} = log(L_i(t_{k+1})/L_i(t_k))$. As discussed in [27], the weighted matrix is given by

$$\Sigma_{i,j} = \frac{1-\lambda}{1-\lambda^N} \sum_{k=0}^{N-1} \lambda^k (l_{i,N-k} - \bar{l_i})(l_{j,N-k} - \bar{l_j})$$

where $\bar{l_i} = \frac{1}{N} = \sum_{k=1}^{N} l_{i,k}$.

The smoothing parameter λ determines the weighting and can be chosen by the practitioner. It is usually between 0.75 and 0.97 and in this thesis we will use $\lambda = 0.94$ [27].

Since the drift in our model is completely determined by arbitrage arguments, we only need to estimate the volatility factorsσ_n, for $n = 1, \dots, M$. The volatility factors describe the variance of the log-returns. Hence,

we calculate the historical correlations and covariance between the log-returns, and then estimate σ_n from equation (4.1.8) and (4.1.9). The system of equations is solved by a built-in numerical optimization routine in MATLAB.

4.2 Jump diffusion LMM

We now follow Glasserman and Kou [7] and Glasserman and Merner [10], and introduce the jump diffusion LMM. Intuitively, the interest rate dynamic posed in this section is based on the same principles as that of the log-normal LMM where additionally random jumps is allowed to occur of a random magnitude. We start out with a general description and gradually state the model assumptions.

The jumps can be represented through a marked point process as described in definition 2.3. We thus have a sequence of tuples $\{(\tau_j, X_j)\}_{j=1,2\ldots}$, where τ_j is a time point, and X_j denotes a jump mark. Note that τ_j is not yet assumed Poisson as in definition 2.3. Further, we stress that the magnitude of the jump is not the same as the mark X_j, however they are related through the map $h(\cdot)$. Similar to [10] we define the jump magnitude $J(t)$ occurring at τ_i to be given by a map $h(\cdot)$, which translates the mark to a jump magnitude in $\mathbb{R}$

$$J(t) = \sum_{j=1}^{N_t} h(X_j). \tag{4.2.1}$$

Here, N_t, similarly to 2.2, is the number of marks in the time interval $(0, t]$. We let N_t have deterministic arrival rate $\hat{\lambda}(t)$. Note that N_t is not necessarily a Poisson process under the desired measure and is considered a general counting process for the time being. Further, we let the marks X_j be independent, with each having the probability density function $g(x)$. The intensity of the ith MPP is defined as $\nu^i(x,t) = g_i(x)\hat{\lambda}_i(t)$. For the next sections it is important to distinguish between the intensity of a MPP and the corresponding intensity of the actual jump. We will need M marked point processes to model rates of M different maturities. We thereby define, as in [10], the jump term for the nth forward rate to be

$$J_n(t) = \sum_{i=1}^{r} \sum_{j=1}^{N_t^{(i)}} H_{ni}(X_j^i).$$

We are now ready to augment the dynamics of equation (4.1.2) as in [14]

$$\frac{dL_n(t)}{L_n(t-)} = \alpha_n(t)\,dt + \gamma_n(t)\,dW(t) + dJ_n(t), \qquad 0 \le t \le T_n \quad n = 1, \ldots, M, \tag{4.2.2}$$

where $t-$ denotes the left limit in time. Intuitively, $L_n(t-)$ represents the level of the rate just before it may jump. The arbitrage-free drift term will now have to incorporate the jump diffusion as well. Under the spot measure, Glasserman and Kou [14] shows the general drift to be given by

$$\begin{aligned}\alpha_n = &\sum_{j=\eta(t)}^{n} \frac{\delta\gamma_j(t)^\intercal\gamma_n(t)L_j(t-)}{1+\delta L_j(t-)} \\ &-\int_0^\infty \sum_{i=1}^{r} H_{ni}(L(t-),x,t) \prod_{k=\eta(t)}^{n} \frac{1+\delta L_k(t-)}{1+\delta L_k(t-)(1+H_{ki}(L(t-),x,t))}\nu^i(x,t)\,dx,\end{aligned} \tag{4.2.3}$$

where $\eta(t)$ is defined as the index of the next maturity when you are at time t. More formally,

$$\eta(t) = \inf\{j \geq |\ T_j \geq t\}$$

The purpose of $\eta(t)$ becomes clear when t is not equal one of the maturing dates. When t is equal one of the maturing dates this parametrization is equivalent to the stationary parametrization used for the log-normal LMM.

So far, the jump diffusion model is quite abstract and difficult to simulate due to the presence of the marked point processes that is not necessarily Poisson. We therefore introduce some assumptions that enables us to formulate the dynamics without MPPs. First we define $I_n(t)$ to be the set of MPPs that can affect the rate $L_n(t)$ at time t. For instance, for $t = 4$, the LIBOR rate $L_3(4)$ has expired and is not affected by any of the MPPs any more; hence $I_3(4) = \{\}$. Now, let the translation from marks to jump magnitudes be

$$H_{ni}(x) = \begin{cases} x-1, & i \in I_n(t) \\ 0, & \text{otherwise.} \end{cases} \tag{4.2.4}$$

This simply imply that the mark x, at a jump time τ, is translated into an actual jump size of $x-1$. This applies for all the rates affected by that MPP. We additionally impose that the ith MPP is Poisson with rate $\hat{\lambda}_i(t)$, and the mark distributions $g_i(x)$ are log-normal. Following Glasserman and Kou [7], it can be shown from here that the arbitrage-free dynamics of equation (4.2.2) and (4.2.3), under *its own* forward measure, may be written as

$$\frac{dL_n(t)}{L_n(t-)} = -\lambda_n m_n dt + \gamma_n(t)\,dW(t) + d\left(\sum_{j=1}^{N_t} Y_j - 1\right) \qquad 0 \leq t \leq T_n \quad n = 1,...,M. \tag{4.2.5}$$

Here, Y_j has probability density distribution f_n, and N_t is Poisson distributed with rate λ_n. We follow [10] and let the jumps be log-normally distributed, imposing

$$f_n(a_n, s_n) \sim \exp\left(\mathcal{N}(a_n, s_n^2)\right). \tag{4.2.6}$$

This notation highlights the relation between the log-normal and the normal distribution, where a_n is the mean and s_n^2 is the variance of the corresponding normal distribution. To lighten the notation we will denote the first moment of the log-normal distribution as $m_n + 1 = \int_0^\infty x f_n(a_n, s_n)\, dx$. From equation (4.2.5), we see drift has become $-\lambda_n m_n$. This is the same as $E[\sum_{j=1}^{N_t}(Y_j - 1)]$ and subtracting this term clearly keeps the forward LIBOR rates $L_n(t)$ martingale under *its own* forward measure.

Considering a single forward rate $L_n(t)$ the dynamics of equation (4.2.5) bears a lot of intuition. First of all, there is no abstract marked point processes involved, and the jumps are sampled directly through Y_j. The jump term $\sum_{j=1}^{N_t}(Y_j - 1)$ is simply found by summing over all the jump magnitudes for the jumps that occur. Finally, part of the assumption is that the arrival of jumps are Poisson. We thereby have a model which seems quite tractable in terms of simulation. However, now considering the multifactor case, as shown in Glasserman [7], when the jumps are Poisson for all the rates under their own forward measures, as in equation (4.2.5), they cannot all be Poisson under the *same* forward measure. This is problematic when pricing derivatives, since they may require each forward rate to be comparable [16]. This is why we made the effort of formulating the MPP in the first place: Glasserman and Kou [7] showed that by constructing a thinning process from a random Poisson measure, we are able to sample from our desired MPPs. In other words, we aim to reintroduce the MPPs in our dynamics and then specify how to make these tractable by stating a Poisson thinning process.

We start by assuming the following relation between the intensity of the MPPs and the intensities in the actual jumps

$$\sum_{i \in I_n(t)} \nu_i(y, t) = \prod_{k=\eta(t)}^{n} \frac{1 + \delta y L_k(t-)}{1 + \delta L_k(t-)} \lambda_n f_n \tag{4.2.7}$$

In other words, we are relating the abstract MPP intensities ν_i as described above to the physical intensity in the nth forward rate $\lambda_i f_i(y)$.

To make our jump diffusion comparable to the log-normal LMM, we aim to obtain the equivalent of a Musiela parameterization. To this end we state that $\lambda_n(t) = \lambda_{n+1-\eta(t)}$, $f_n(t) = f_{n+1-\eta(t)}$ and $\gamma_n(t) = \gamma_{n+1-\eta(t)}$.

Under this specification, it is seen from equation (4.2.8) that $I_n(t)$, the set of which MPPs L_n is sensitive to, becomes

$$I_n(t) = \{n + 1 - \eta(t), n + 2 - \eta(t), \ldots, M\}$$

.

Together with the stationary specification given by $I(t)$ Glasserman and Merener[10] obtain

$$\nu_i(y,t) = \prod_{j=\eta(t)}^{i+\eta(t)-1} \frac{1+\delta y L_j(t-)}{1+\delta L_j(t-)} \left[\lambda_i f_i(y) - \lambda_{i+1} f_{i+1}(y) \frac{1+\delta y L_{j+\eta(t)}(t-)}{1+\delta L_{j+\eta(t)}(t-)}\right] \tag{4.2.8}$$

Applying Theorem 3.1 in Kou [7] now enables us to state the dynamics of the jump diffusion LMM

$$\begin{aligned}\frac{dL_n(t)}{L_n(t-)} =& -\lambda_{n+1-\eta(t)} m_{n+1-\eta(t)}\, dt + \sum_{j=\eta(t)}^{n} \frac{\delta \gamma_n(t)^\intercal \gamma_n(t) L_j(t-)}{1+\delta L_j(t-)}\, dt \\ &+ \gamma_{n+1-\eta(t)}\, dW(t) + d\left(\sum_{i=n+1-\eta(t)}^{M} \sum_{j=1}^{N_t^{(i)}} (Y_j^{(i)} - 1)\right), \qquad 0 \le t \le T_n, \quad n = 1, \ldots, M.\end{aligned} \tag{4.2.9}$$

Here, $N_t^{(i)}$ is the intensity of the ith MPP and $Y_j^{(i)} - 1$ is the translated jump magnitude. Observe that the last term in equation (4.2.9) implies that if the rate L_k jumps, then all rates maturing earlier also jump. Since $N_t^{(i)}$ is not Poisson, we are still not able to simulate the dynamics. This motivates the introduction of Proposition 4.1 from [7], which enables the use a Poisson random measure to simulate equation (4.2.9). The dynamics are rewritten

$$\begin{aligned}\frac{dL_n(t)}{L_n(t-)} =& -\lambda_{n+1-\eta(t)} m_{n+1-\eta(t)}\, dt + \sum_{j=\eta(t)}^{n} \frac{\gamma_n(t) \delta \gamma_j(t)^\intercal L_j(t-)}{1+\delta L_j(t-)}\, dt \\ &+ \gamma_{n+1-\eta(t)}(t)\, dW(t) + \int_0^\infty \int_0^1 (y-1) \sum_{i=k+1-\eta(t)}^{M} \theta_i(y, u, L(t-), t)\, p(dy \times dx, dt)\end{aligned} \tag{4.2.10}$$

where $p(dy \times dx, dt)$ is a Poisson random measure with intensity $\hat{\lambda}_0 f(y)$. $\hat{\lambda}_0$ is related to the jump rate via $\hat{\lambda}_0 = \lambda_1(2 + m_1)$. Furthermore, $f(y) = \frac{f_1(y) + y f_1(y)}{2+m_1}$ where $f_1(y)$ has mean $m_1 + 1$ and variance s_1.

Finally,

$$\theta_i(y,u,L(t-),t) = \begin{cases} 1, & \sum_{j=1}^{i-1} \nu_j(y,L_1(t-),...,L_M(t-),t) \leq uf(y)\lambda_0 < \sum_{j=1}^{i} \nu_j(y,L_1(t-),...,L_M(t-),t) \\ 0, & \text{otherwise,} \end{cases} \tag{4.2.11}$$

where u is uniformly distributed between 0 and 1. Requiring all $\nu_j \geq 0$, only one of the MPPs can be non-zero at the same point in time. As discussed in Glasserman [7], this is needed for a meaningful process. Keeping the intensities non-zero yields restrictions on the parameter set. In Glasserman and Kou[14] they sate a simple way to check these restrictions by assuming that $s_{n+1} < s_n$, i.e. that the jump variance decrease as maturity increase.

$$4b_n^2 \left(d_n - \log\left(\frac{\lambda_{n+1}}{\lambda_n} \right) \right) \geq \max(c_n^2, (c_n - 1)^2), \tag{4.2.12}$$

where

$$b_n = \frac{1}{2}\left(\frac{1}{s_{n+1}^2} - \frac{1}{s_n^2} \right), \quad c_n = \frac{a_n}{s_n} - \frac{a_n + 1}{s_{n+1}}, \quad d_n = -\frac{1}{2}\left(\frac{a_n^2}{s_n^2} - \frac{a_{n+1}^2}{s_{n+1}^2} \right) + \log\left(\frac{s_{n+1}}{s_n} \right).$$

To conclude, this section demonstrates that the jumps can be represented by M MPPs with Poisson arrival rate. Further, the the intensity of the MPPs are related to the intensity of the jump. This means we can estimate the jump arrival rate and jump mean from historical data, then estimate the MPPs from these parameters, and finally use the MPPs to forecast the jump term again through a thinning measure.

4.2.1 Simulation

As stated by Glasserman and Merener [10], there is no way of generating exact sample paths for the model in equation (4.2.10). Hence, we briefly follow [10] to develop a numerical scheme that are can handle the combination of both standard diffusion and jumps. The following dynamics are considered

$$dX(t) = a(X(t)) + b(x(t))\, dW(t) + \int_E c(X(t), z)\, p(dz, dt) \tag{4.2.13}$$

where $W(t)$ is scalar Brownian motion, and $p(dz, dt)$ represents a Poisson random measure on $E \times [0, T]$. Its intensity is $\lambda_0 h(z)$, and the jump marks z are distributed with $h(z)$ as its probability density function. The idea is now to construct solutions to equation (4.2.13) at a set of discrete points in time $\{\tau_i\}$. Derivation

of the first order Euler scheme can be found in [10]. We merely state the result here, where the scheme reads

$$\hat{L}_n(t_{i+1}^-) = \hat{L}_n(t_i)\exp[\hat{a}_n(t_i)(t_{i+1}-t_i) + \gamma_k(t_i)(W_{t_{i+1}} - W_{t_i})], \tag{4.2.14}$$

where $\hat{a}_k(t_i)$ is defined as

$$\hat{a}_n(t_i) = -\lambda_{n+1-\eta(t_i)} m_{n+1-\eta(t_i)} - \frac{1}{2}\gamma_n(t_i)^2 - \sum_{j=\eta(t_i)}^{n} \frac{\delta_j \gamma_k(t)^\intercal \gamma_j \hat{L}_j(t_i^-)}{1+\delta_j \hat{L}_j(t_i^-)}. \tag{4.2.15}$$

As can be seen from the left limit t_{i+1}^- above, this scheme is without the jumps. Actually, if $\lambda_0 = 0$, then this scheme is identical to the log-normal scheme from equation (4.1.7). After calculating the rates without regarding jumps, we are ready to incorporate the jumps

$$\hat{L}_n(t_{i+1}) = \hat{L}_n(t_{i+1}^-)\left[1 + \int_0^\infty \int_0^1 (y-1) \sum_{j=n+1-\eta(t_i)}^{M} \theta_j(y, u, \hat{L}(t_{i+1}^-), t_{i+1}^-) p(dy \times du, t_{i+1})\right], \tag{4.2.16}$$

where θ_j denotes a marked point process (MPP). The double integral is found by the thinning procedure outlined in in equation (4.2.11). However, as pointed out in Glasserman and Merener [7], the notation for the thinning process is obscuring a quite simple simulation routine.

We now quickly describe how we use equation (4.2.14) and (4.2.16) in practice. At each time step we firstly apply equation (4.2.14) to find t_{i+1} $\hat{L}_n(t_{i+1}^-)$, for $n = 1, \ldots, M$. That is, the rates without jumps. Now we sample a jump path J, i.e. we sample a vector $\mathbf{J}$ of 1's and 0's, indicating jump or no jump, respectively. By the MPP construction, this path is Poisson with intensity λ_0. When $\Delta t = \tau_{i+1} - \tau_i$ is sufficiently small, the probability of more than 1 jump becomes negligible, and we have $P(\mathbf{J}_i = 1) \approx \lambda_0 \Delta t$ [8].

We briefly pause to point out the implications of this detail. In a log-normal LMM, Δt is usually set equal to the accrual period, due to the properties of the Wiener Process, see equation (2.1). There is simply no gain in simulating the process between the maturity dates in equation (4.1.6). This logic is, however, not applicable for the jump diffusion LMM, since the probability for more than one jump between maturity dates is most likely significant. That is, we focus on daily time increments for the jump diffusion LMM when utilizing the scheme from equation (4.2.16).

Continuing on the thinning process. After the jump path $\mathbf{J}$ is acquired, we sample a mark y from the distribution

$$f(y) = \frac{f_1(y) + y f_1(y)}{2 + m_1},$$

whenever $\mathbf{J_i} = 1$. Firstly, we recognize that $f(y)$ is a mixture of two distributions. We therefore sample from $\mathbf{Z_1} = f_1(y)$ with probability $(2 + m_1)^{-1}$ and from $\mathbf{Z_2} = yf_1(y)$ with probability $(1 + m_1)(2 + m_1)^{-1}$. Remembering from equation (4.2.6) that f_1 is log-normally distributed, noted in Galsserman and Merener [10] that when $\mathbf{Z_2} = yf_1(y)$ we have $\mathbf{Z_2} \sim \exp(\mathcal{N})$ where the first moment is $E[\mathbf{Z_2}] = E[\mathbf{Z_1}] + \text{Var}[\mathbf{Z_1}]$ and $\text{Var}[\mathbf{Z_2}] = \text{Var}[\mathbf{Z_1}]$.

After we have sampled a jump y from $f(y)$ for each jump time, the actual thinning starts. We begin with computing all the partial sums needed in equation (4.2.11), i.e. $S_i = \sum_{j=1}^{i} \nu_j$ for $i = 1 \dots M$ which can be found by using 4.2.8. Now, if $S_{i-1} < uf(y)\lambda_0 < S_i$, where $u \sim \mathcal{U}(0, 1)$, the jump is accepted and all rates maturing at T_i and earlier will accept the jump y. If $uf(y)\lambda_0 > s_M$, the jump is rejected. Using this thinning process, the integral in equation (4.2.16) is replaced with y. That is $\hat{L}_k(t_{i+1}) = \hat{L}_k(t_{i+1}^-)\, y$. This concludes the simulation procedure, where we have followed Glasserman and Merener [7] to demonstrate how to simulate the LMM with jump diffusion under the spot measure using the forward specification of jump arrivals.

4.2.2 Parameter estimation

The procedure for estimating the parameters in equation (4.2.9) is involved. We cannot estimate the diffusion with the same routine as in the previous section, since it does not distinguish between what is standard diffusion and what is jump diffusion. Further, we also need to estimate the jump mean and variance, in addition to the arrival rate of jumps. Similarly to Steinrücke et al. [16], we start by following the dynamics given in Johannes et al. [8]

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t + d\left(\sum_{j=1}^{N_t} S_{\tau_j -}(\exp(Z_j) - 1)\right),$$

where $S_t \in \mathbb{R}^k$, W_t is a vector standard Brownian motion and $\sigma\sigma^\intercal = \Sigma \in \mathbb{R}^k \times \mathbb{R}^k$. Here, Σ is the covariance matrix for the diffusion. N_t is a Poisson process with constant rate λ and the jump sizes $Z_j \in \mathbb{R}^k$ are multivariate normal, with mean μ_J and covariance matrix Σ_J. Estimating the parameters in equation (4.2.2) captures the movements of correlated asset prices with jumps. This means that the arbitrage-free requirement of the log-normal LMM given in equation (4.2.3) is not imposed. However, Steinrücke et al. [16] proposes to apply the dynamics in order to estimate the jump diffusion only and discard the estimated drift as it does not meet the arbitrage-free requirements.

Moreover, the forward specification of the jump diffusion LMM given in equation (4.2.9) does not explicitly require the correlation between jumps, i.e. there is no correlation matrix for jumps. Additionally, we have imposed restrictions for the jump parameters given in equation 4.2.12. Hence our procedure will be the following: Use the one-dimensional case of equation (4.2.2) to estimate the jump parameters of each historical forward rate *separately*. Now if the assumption in 4.2.12 is not kept, a reasonable adjustment of the parameter must be made. The assumption in (4.2.12) given by [10] that the variance and intensity is higher for lower maturities is not arbitrary. As we will see in the data analysis section, the log-returns of shorter maturities is significantly more fluctuating. Thus, even though we have no guarantee that our parameter estimation will yield variances and intensities that are strictly decreasing for higher maturities, we may assume that they are downward trending. Thus we argue that in the context of comparing the forecast distributions of the Jump diffusion LMM with other models, a potential slight adjustment of the jump parameters will not invalidate our results.

Even though the jumps in the rates are assumed to be uncorrelated in equation (4.2.9), the rates themselves are assumed to be correlated similarly to the log-normal case. Consequently, using the one-dimensional version of equation (4.2.2) for parameter estimation leaves us with no proper estimate of the volatility factors $\gamma_n(t)$. In their calibration routine, Steinrücke et al. [16] corrects their estimate of the implied volatility by removing the diffusion that is attributed to jump diffusion. Similarly, if we remove the points from the historical data that is estimated to represent jumps, the resulting log-returns of the series would represent the ordinary diffusion term. Now, we could simply estimate the covariance matrix from the new data set and calculate $\gamma_n(t)$ in equation (4.2.9) for $n = 1, \ldots, M$ the same way we did for the log-normal model. From this discussion it is clear that we only need to estimate the following parameters from equation (4.2.2): The jump path $\mathbf{J}$, the jump arrival rate λ, the jump mean m_J and the jump variance v_J for each maturity separately.

The jump path $\mathbf{J}$ is simply a vector of 0's or 1's for each time point in our data, indicating jump and no jump, respectively. The main difference between the estimation procedure above and the procedure presented in Steinrücke et al. [16] regards the diffusion. In this thesis we are not calibrating to a specific instrument, and will therefore not concern ourselves with implied volatilities.

We are now ready to state the Bayesian hierarchical model for the estimation in the dynamics of equation (4.2.2). Again, we closely follow Johannes et al. [8]. Solving the differential equation in equation (4.2.2), we

can write

$$\log\left(\frac{S_t}{S_{t-1}}\right) = \mu + \sigma(W_{t+1} + W_t)\, d\left(\sum_{j=1}^{N_t} Z_j\right),$$

where the term μ represents the drift after integration. As Johannes et al. [8], we consider this model over a small time frame, and as pointed out in the preliminaries, when the time step is sufficiently small, the probability of more than 1 jump becomes negligible. We thereby obtain $P(\mathbf{J}_t = 1) \approx \lambda_0 \Delta t$. Solving this differential equation one can write

$$Y_t \equiv \log\left(\frac{S_t}{S_{t-1}}\right) = \mu + \sigma\sqrt{\Delta t}\,\epsilon_t + J_t Z_t, \tag{4.2.17}$$

where ϵ_t is standard normal. The changes in diffusion is given by the properties of the Wiener process, where the variance is proportional to the time step.

Our dynamics are now appropriately simple and the rest of the section will specify how to estimate the parameters in equation (4.2.17), in addition to the jump path $\mathbf{J}$ and the jump sizes $\mathbf{Z}$. As proposed in [8], the procedure is carried out in the following Gibbs sampling steps, where k is the iteration index

1. Sample $\mathbf{\Theta}^{(k)}$ from $p(\mathbf{\Theta} \mid \mathbf{Y}, \mathbf{Z}^{(k-1)}, \mathbf{J}^{(k-1)})$
2. Sample $\mathbf{Z}^{(k)}$ from $p(\mathbf{Z} \mid \mathbf{Y}, \mathbf{\Theta}^{(k)}, \mathbf{J}^{(k-1)})$
3. Sample $\mathbf{J}^{(k)}$ from $p(\mathbf{J} \mid \mathbf{Y}, \mathbf{\Theta}^{(k)}, \mathbf{Z}^{(k)})$

In order to perform step 1, we firstly apply equation (4.3.7) and obtain

$$p(\mathbf{\Theta} \mid \mathbf{Y}, \mathbf{Z}, \mathbf{J}) \propto p(\mathbf{Y}, \mathbf{J}, \mathbf{Z} \mid \mathbf{\Theta})\, p(\mathbf{\Theta}), \tag{4.2.18}$$

Furthermore, we assume the parameters of $\mathbf{\Theta}$ are a priori independent which yields

$$p(\hat{\alpha}, \delta^2, m_J, v_J^2, \lambda) = p(\hat{\alpha})\, p(\delta^2)\, p(m_J)\, p(v_J^2)\, p(\lambda)$$

Following Johannes et al. [8], we have that $p(\mu, \sigma|J, Z, Y)$, $p(\mu_j, \Sigma_j|J, Z)$ and $p(\lambda|J)$ characterizes $p(\Theta|X, Y)$, while $p(Z_t|\Theta, J_t, Y_t)$ and $p(J_t|\Theta, Z_t, Y_t)$ characterize $p(J, Z|\Theta, Y)$.

The input data of the estimation procedure will be a vector of observed historical rates, transformed according to equation (4.2.17). From equation (4.2.17), we know that the observations in $\mathbf{Y}$ are independent conditional

on $\mathbf{J}$, $\mathbf{Z}$. Also, they are normally distributed, and we attain the likelihood

$$p(\mathbf{Y} \mid \mathbf{\Theta}, \mathbf{J}, \mathbf{Z}) = \prod_{t=1}^{T} \phi(y_t, \mu + J_t Z_t, \sigma\sqrt{\Delta t}),$$

where $\phi(a, b, c)$ is defined as the normal density function evaluated at a, with mean b and variance c.

The prior distributions of the parameters of $\mathbf{\Theta}$ are chosen to be conjugate [8]. The details of the priors and their hyperpriors are found in [16], and we will only show the main results here. For $\hat{\alpha}$, we choose a normal prior distribution. By assuming $\hat{\alpha}$ is conditionally independent of the jump parameters, we get

$$p(\hat{\alpha} \mid \mathbf{Y}, \delta^2, \mathbf{J}, \mathbf{Z}) \propto p(\mathbf{Y} \mid \delta^2, \mathbf{J}, \mathbf{Z})\, p(\hat{\alpha}), \tag{4.2.19}$$

producing a normal distribution. Next, the prior distribution of δ^2 is inverse-gamma [16]. As δ^2 is also assumed to be conditionally independent of jump parameters, the posterior distribution becomes

$$p(\delta^2 \mid \hat{\alpha}, \mathbf{Y}, \mathbf{Z}, \mathbf{J}) \propto p(\mathbf{Y} \mid \delta^2, \mathbf{J}, \mathbf{Z})\, p(\delta^2), \tag{4.2.20}$$

which is inverse-gamma distributed.

The jump size mean m_J variance v_J^2 are assumed to be dependent only on each other the jump path and the jump sizes. More precisely, only the jump sizes at times which jump occurs. We attain

$$p(m_J \mid v_J^2, \mathbf{Z}, \mathbf{J}) \propto p(\mathbf{Z} \mid m_J, v_J^2)\, p(m_J), \tag{4.2.21}$$

and

$$p(v_J^2 \mid m_J, \mathbf{Z}, \mathbf{J}) \propto p(\mathbf{Z} \mid m_J, v_J^2)\, p(v_J), \tag{4.2.22}$$

where the likelihood has independent observations that are assumed normal [16]. Thus,

$$p(\mathbf{Z} \mid m_J, v_J^2) = \prod_{t=1}^{T} \phi(Z_t, m_J, v_J^2), \tag{4.2.23}$$

where ϕ is still the normal density function. The jump mean is assumed to have a normal distribution as its prior, while the variance prior distribution is inverse-gamma [16]. The posteriors are thereby normal and inverse-gamma, respectively.

The last parameter is the jump arrival rate λ. The probability of a jump in a time interval is λ, and the jumps are assumed to arrive independently [16]. Let $N_{\mathbf{J}}$ denote the number of jumps in the jump path $\mathbf{J}$. The probability of having $N_{\mathbf{J}}$ jumps over T time steps is therefore

$$p(\mathbf{J} \mid \lambda) = \lambda^{N_{\mathbf{J}}}(1-\lambda)^{T-N_{\mathbf{J}}}$$

Using equation (4.3.7), we obtain

$$p(\lambda \mid \mathbf{J}) \propto p(\mathbf{J} \mid \lambda)\, p(\lambda), \tag{4.2.24}$$

where the prior distribution $p(\lambda)$ is chosen to be beta, yielding a beta posterior [8].

Moving on to step 2, we start by using equation (4.3.7) as in [16] yet another time

$$p(Z_t \mid \mathbf{Y}, \mathbf{\Theta}, \mathbf{J}) \propto p(Y_t \mid Z_t, \mathbf{\Theta}, J_j)\, p(Z_t \mid \mathbf{\Theta}). \tag{4.2.25}$$

The likelihood is still normally distributed, and together with a normally distributed prior for Z_t, the posterior is normal.

Finally, in step 3, we exploit the fact that J_t can either be 0 or 1. Hence, as presented in [16], the posteriors are

$$\begin{aligned} p(J_t = 1 \mid \mathbf{\Theta}, Z_t, Y_t) &\propto p(Y_t \mid J_t = 1, \mathbf{\Theta}, Z_t)\, p(J_t = 1 \mid \mathbf{\Theta}) \\ p(J_t = 0 \mid \mathbf{\Theta}, Z_t, Y_t) &\propto p(Y_t \mid J_t = 0, \mathbf{\Theta}, Z_t)\, P(J_t = 0 \mid \mathbf{\Theta}), \end{aligned} \tag{4.2.26}$$

where $p(J_t = 1 \mid \mathbf{\Theta}) = \lambda$ and $p(J_t = 0 \mid \mathbf{\Theta}) = 1 - \lambda$.

The parameter estimation procedure can now be summed up in the following steps

1. Sample the hyperpriors a_0, b_0^2, f_0, g_0, $a_{J,0}$, $b_{J,0}$, $f_{J,0}$, $g_{J,0}$, u_0, w_0.
2. Sample priors
   - $\hat{\alpha}^{(0)} \sim \mathcal{N}(a_0, b_0^2)$.
   - $\delta^{2(0)} \sim IG(f_0, g_0)$.
   - $m_J^{(0)} \sim \mathcal{N}(a_{J,0}, b_{J,0})$.
   - $v_J^{2(0)} \sim IG(f_{J,0}, g_{J,0})$.
   - $\lambda^{(0)} \sim \mathcal{B}(u_0, w_0)$.

The remaining steps are repeated for $k = 1, \ldots, R$

3. Sample the parameters of $\boldsymbol{\Theta}^{(k)}$

   - Sample $\hat{\alpha}^{(k)} \sim N(a^{(k)}, b^{2(k)})$.
   - Sample $\delta^{2(k)} \sim IG(f^{(k)}, g^{(k)})$.
   - Sample $m_J^{(k)} \sim N(a_J^{(k)}, b_J^{(k)})$.
   - Sample $v_J^{2(k)} \sim IG(f_J^{(k)}, g_J^{(k)})$.
   - Sample $\lambda^{(k)} \sim \mathcal{B}(u^{(k)}, w^{(k)})$.

   For details on the hyperpriors, see Steinücke et al. [16].

4. Sample $Z_t^{(k)} \sim N(R_j^{(k)}, S_j^{2(k)})$ for $t = 1, \ldots, T$.

5. Sample J_t for $t = 1, \ldots, T$ using equation (4.2.26).

The first R_B iterations are discarded. The estimated values of $\hat{\alpha}$, δ^2, m_J, v_J^2, λ, $\mathbf{Z}$ and $\mathbf{J}$ are set to the respective average values over the remaining iterations. For $\mathbf{J}$, the average value is rounded to the closest integer.

For a verification of the implementation of this procedure, see Appendix B.

4.3 Markov-Switching Jump Diffusion LMM

Finally we consider the Markov-switching jump diffusion LIBOR market model derived by Steinrücke et al. [16] where all the parameters of a general jump-diffusion LMM switches between states according to a Markov chain. We start by briefly summarizing their result.

Jamishidan (1999) [15] considers a generalized version of the jump specification in Glasserman and Kou [14] using semi-martingales. From Jamishidan they obtain the existence of the the following dynamics under the

terminal measure Q^N

$$\begin{aligned}\frac{dL_n(t)}{L_n(t-)} &= \sum_{j=i+1}^{n} \frac{\delta\gamma_j(t)^\intercal\gamma_n(t)L_j(t-)}{1+\delta L_j(t-)} + \sigma_n(t)^\intercal\, d\mathbf{W}^N(t) \\ &\quad - \int_E \psi_n(t,z)\left(\prod_{j=i+1}^{N-1}\left(1+\frac{\delta L_n(t-)\psi_n((t,z)}{1+\delta L_n(t-)}\right)-1\right)\nu^N\,(dt,dz) \\ &\quad + \int_E \psi_n(,t,z)(\mu-\nu^N)\,(dt,dz),\end{aligned} \tag{4.3.1}$$

where they have assumed the following

- W_t standard Brownian motion
- μ is an integer-valued random measure on $\mathbb{R}^+ \times \mathbb{R}^k$

The terminal measure Q^N is simply the measure induced by using the bond of highest maturity as numeraire. Further E is a euclidean space, ν^N is the compensator of the random measure μ under the terminal measure and $\psi_i(t,z)$ and $\gamma_n(t)$ some deterministic functions as in Belomestny and Schoenmakers (2011) [28]. It is noted that the random measure associated with a MPP with time-homogeneous arrival rates is integer-valued(see [19]). Thus the semi-martingale allows for the MPP specification of a jump process that follow a standard Brownian motion, demonstrating the that semi-martingales are the more general process as stated in Jamishidan [15]. Steinrücke et al. [16] states that the framework in Jamishidan [25] now allows for the parameters $\psi_i(t,z)$ and $\gamma_n(t)$ to depend on the continuous-time homogeneous finite Markov chain with associated infinitesimal generator $\mathcal{A}$ taking values in $E = \{e_1,\ldots,e_M\}$ and obtain

$$\begin{aligned}\frac{dL_n^X(t)}{L_n^X(t-)} &= \sum_{j=i+1}^{n} \frac{\delta\gamma_j(t,X_{t-})^\intercal\gamma_n(t,X_{t-})L_j(t-)}{1+\delta L_j(t-)} + \sigma_n(t)^\intercal\, d\mathbf{W}^N(t) \\ &\quad - \int_E \psi_n(X_{t-},t,z)\left(\prod_{j=i+1}^{N-1}\left(1+\frac{\delta L_n(t-)\psi_n(X_{t-}(t,z)}{1+\delta L_n(t-)}\right)-1\right)\nu^N_{X_{t-}}\,(dt,dz) \\ &\quad + \int_E \psi_n(X_{t-},t,z)(\mu-\nu^N_{X_{t-}})\,(dt,dz),\end{aligned} \tag{4.3.2}$$

Note that the only difference between equations (4.3.2) and (4.3.1) is that the parameters in (4.3.2) depend on a markov chain. To make the dynamic above sufficiently tractable and ease developing Fourier pricing

formulae, they further state that

- The spot W^Q Wiener process and the compensated jump measure $\mu - \nu^Q_{X_{t-}}$ are independent given X
- The only source of randomness volatilities dependent on the underlying Markov chain, meaning that the volatilities are state dependent.
- The state dependent jump compensator $\nu^Q_{X_{t-}}(dt, dz)$ of μ is predictable by a jump intensity $\lambda(X_{t-})^Q$ and a jump marker distribution $k^Q(X_{t-}, z)$ through the relation

$$\nu^Q_{X_{t-}} = k^Q(X_{t-}, z)\, \lambda^Q(X_{t-})\, dt$$

- The function $\psi_n(t, X_{t-}, z)$ is predictable and state dependent. For $t < T_n$, we set $\psi_n(t, X_{t-}, z) = 0$.

Now they propose the model (4.3.2) exists under the measure change from Q^N to Q^{i+1}. That is, similar to the specification in (4.2.5), they consider the rates free of arbitrage under their own forward measure. The new semi-martingale specification of the LIBOR rates become

$$\frac{dL^X_n(t)}{L^X_n(t-)} = d\mathbf{W}^{i+1}(t) + \int_E \psi_n(X_{t-}, t, z)(\mu - \nu^N_{X_{t-}})\,(dt, dz) \tag{4.3.3}$$

Remember that in the equation (4.2.5), W_t is defined as standard Brownian and N_t is Poisson with constant rate λ_n. Hence the random variable N_t has an associated integer-valued measure, see [19]. Thus (4.2.5) fulfills the assumptions as given by Steinrücke [16] and the parameters can be state dependent on an underlying continuous-time homogeneous Markov chain. Further we note as in Steinrücke [16] that this is the only source of randomness in the parameters.

We rewrite 4.2.5 for clarity

$$\frac{dL^X_n(t)}{L^X_n(t-)} = -\lambda_n^{X_{t-}} m_n^{X_{t-}} dt + \gamma_n(t)^{X_{t-}}\, dW(t)^{n+1} + d\left(\sum_{j=1}^{N^{X_{t-}}} Y_j^{X_{t-}} - 1\right) \qquad 0 \le t \le T_n \quad n = 1, ..., M. \tag{4.3.4}$$

Further Steinrücke et al. [16] shows that a measure change from Q^{i+1} to Q^i does not affect the underlying Markov chain. Thus as in section 4.2 the measure of equation (4.3.4) can be changed to the spot martingale

$$\begin{aligned}\frac{dL_n(t)}{L_n(t-)} =& -\lambda_{n+1-\eta(t)}^{X_{t-}}(t)m_{n+1-\eta(t)}^{X_{t-}}(t)\,dt + \sum_{j=\eta(t)}^{n}\frac{\gamma_n(t,X_{t-})\delta\gamma_j(t,X_{t-})^\intercal L_j(t-)}{1+\delta L_j(t-)}\,dt \\ &+ \gamma_{n+1-\eta(t)}(t,X_{t-})\,dW(t) + \int_0^\infty\int_0^1 (y-1)\sum_{i=k+1-\eta(t)}^{M}\theta_i^{X_{t-}}(y,u,L(t-),t).\end{aligned} \tag{4.3.5}$$

The definitions of all the terms are identical to what was presented in 4.2.1. However, we elaborate the state dependency for the thinning process more explicitly to avoid ambiguity. The state dependent θ_i is defined as

$$\theta_i^{X_{t-}}(y,u,L(t-),t) = \begin{cases} 1, & \sum_{j=1}^{i-1}\nu_j^{X_{t-}}(y,L_1(t-),...,L_M(t-),t) \le uf(y,X_{t-})\lambda_0^{X_{t-}} < \sum_{j=1}^{i}\nu_j^{X_{t-}}(y,L_1(t-),...,L_M(t-),t) \\ 0, & \text{otherwise.}\end{cases}$$

Here,

$$\lambda_0^{X_{t-}} = \lambda_1^{X_{t-}}(2+m_1^{X_{t-}}) \qquad \text{and} \qquad f(y,X_{t-}) = \frac{f_1(y,X_{t-}) + yf_1(y,X_{t-})}{2+m_1^{X_{t-}}}.$$

Moreover, the state dependent compensators follow

$$\sum_{i\in I_n(t)}\nu_i(y,t) = \prod_{k=\eta(t)}^{n}\frac{1+\delta yL_k(t-)}{1+\delta L_k(t-)}\lambda_n^{X_{t-}}f_n(X_{t-}).$$

The function $f_n(X_{t-})$ is as described in subsection 4.2.1, but is now state dependent through

$$f_n(X_{t-}) \sim \exp\Big(\mathcal{N}\big(a_n^{X_{t-}}, s_n^{X_{t-}}\big)\Big).$$

For the mark y, the state dependency is not made explicit. Nevertheless, it is sampled using the same approach as in subsection 4.2.1, given the state dependent jump parameters.

4.3.1 Simulation

The simulation procedure for 4.3.5 is rather similar to the procedure used for simulating jumps. Given a state X_t, in the Markov chain at time t. Then we have a set of parameters associated to that state $\mathbf{\Theta}_{X_t}$ = [Volatility factors(γ_i), jump intensities(λ_i), the jump means(m_i) and the jump variances(s_i)] for all maturities, similar to Steinrücke[16]. To find the LIBOR rates at time $t+1$ the exact same procedure as in (4.2.16) can be used where the parameters are given by $\mathbf{\Theta}_{X_t}$. Hence the we assume that the parameter

set $\mathbf{\Theta}_{X_t}$ applies between t and $t+1$. Now the state of the Markov chain might change between the times $t+1$ to $t+2$. Thus to obtain X_{t+1}, similar to Steinrücke [16], given the infinitesimal generator $\mathcal{A}$ of the continuous time Markov chain, we sample this Markov chain where the transition matrix $C_{\Delta t}$ is approximated by $C_{\Delta t} \approx \exp(\mathcal{A}\Delta t)$. Here Δt is as usual our time step. For this discrete time Markov chain to to be a reasonable approximation of the continuous time Markov chain, the time step has to be small enough such that the probability for changing state more than once during Δt is negligible. Note that all the maturities are under the same regime at all times.

To summarize the steps:

- Set the initial state X_0 equal to the state of the last known historical date
- For each future time steps T
  - Sample next state X_t conditioned on X_{t-1} using transition matrix $C_{\Delta t}$.
  - Use parameter set $\mathbf{\Theta}_{X_t}$ together with equation (4.2.16) to simulate the LIBOR forward rates Δt time ahead. Repeat for all maturities that have not expired.

The steps are repeated using Monte Carlo iterations, and the rates are saved for each iteration. Further more we choose to model the economy by two regimes, thus using two states in the Markov chain.

4.3.2 Parameter estimation

The strategy for estimating the parameters in this frame work follows that of Steinrücke [16]. We split our data set according to which regime they belong to, state 1 or state 2, and then we simply run the same parameter estimation as described in the jump diffusion section on each of these data sets. Thereby the parameter sets $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ is obtained. Hence we start by classifying the historical data in to the two regimes. To this end, following Steinrücke et al. [16] a Markov-switch algorithm is constructed using a hidden Markov model together with a one-factor Vasicek model where additionally the transition matrix C is estimated. Steinrücke et al. [16] argues that historical caplet volatilities is a good proxy for the state of the economy. They associate each historical caplet volatility to either a regime 1 or 2 and then they take the average over all these states to obtain a single time series containing 1's and 2's indicating which state the economy is in. This series is termed the over all most likely path. Then they use the over all most likely path to split the historical forward LIBOR rates. Then finally the jump parameters are estimated using the log-returns of the forward rates as in section 18 at the split sets separately. As discussed earlier, Steinrücke et al [16] do

not find the volatility factors using historic LIBOR rates, rather they use the caplet prices to find implied volatility and thereby calibrate the model to interest rate caps.

In this thesis we will use a similar procedure as Steinrücke et al. [16]. However, we will not calibrate the model any specific instrument as different instruments will be priced with the same algorithm. Thus, to split the data set we would rather use the Markov switch algorithm on the historic log-returns as these represents the volatility in our specification. Similar to Steinrücke [16] we find the overall most likely path by averaging the estimated paths of all maturities and then round to either 1 or 2 where the only difference is that the data is split on log-returns and not cap volatilites. We now do the same procedure as with jump diffusion where we remove the identified jumps from all the time series before the covariance and correlation matrix for each state is found. Note that EWMA has not been used this time. The volatility factors for all maturities for each state is calculated by the numerical procedure in MATLAB. Finally, we note that the initial state of our Markov chain is already given by the last state of our over all most likely path. Thus, the algorithm works somehow similar to the EWMA where the last historic volatility is weighted.

The Markov-Switch algorithm

The derivation of the Markov-switch algorithm is inspired by the work of Frühwirth-Schnatter from 2006 [29] and Steinrücke et al. from 2013 [16]. We construct a first-order hidden Markov model as described in section 2.5, with two possible states. State one represent low volatility, whereas state two represent high volatility. The historical rate data represent the evidence, while the corresponding volatility level is the hidden state.

The main idea is to consider a Vasicek model as defined in equation (3.2.2). However, as there are two possible states, the parameters involved in the model need to be calibrated for both states. Consequently, two parameter sets are to be determined. In order to perform the calibration, equation (3.2.2) is approximated at the discrete, equidistant time points $t_0 < t_1 < \cdots < t_N$, where the distance is denoted as Δt. The process is furthermore discretized using an Euler scheme

$$r_j = r_{j-1} + \alpha_{X_{t-1}}(b_{X_{j-1}} - r_{j-1})\Delta t + \sigma_{X_{j-1}}\mathcal{N}(0, \Delta t),$$

where the subscript X_{j-1} indicates the state of X at time point $j-1$. By introducing the variables $\gamma_{X_{j-1}} = \alpha_{X_{j-1}} b_{X_{j-1}} \Delta t$ and $\delta_{X_{j-1}} = 1 - \Delta t\, \alpha_{X_{j-1}}$, the discrete state-dependent version of the Vasicek

model reads

$$r_j = \gamma_{X_{j-1}} + \delta_{X_{j-1}} r_{j-1} + \mathcal{N}(0, \Delta t\, \sigma^2_{X_{j-1}}). \tag{4.3.6}$$

As Frühwirth-Schnatter points out in her book on finite mixture and Markov-switching models [29], there are three main problems to asses in order to make Markov-switching models useful. Firstly, the number of possible states in the HMM needs to be determined. Next, the parameter sets for the possible states are to be estimated. The transition matrix indicating the probability of moving between the states must be estimated in this step as well. The last problem is to estimate the sequence of the hidden states of the HMM needs to be determined. Since the number of states is already chosen, the last two steps makes up the Markov-switching algorithm that is applied throughout the thesis.

We chose a Bayesian approach to estimate the parameters, in line with the theory of Bayesian hierarchical models. Let $\mathbf{\Theta} = (\theta_1, \theta_2)$ denote the parameter sets for state 1 and state 2, respectively. Furthermore, $\mathbf{r} = r_1, \ldots, r_N$ be the N historical interest rates, and $\mathbf{X} = X_1, \ldots, X_N$ be the sequence of the hidden states of the HMM. That is, $\mathbf{X}$ is a sequence consisting of 1's and 2's, indicating the state at the time. Lastly, let C be the transition matrix containing the probabilities for moving between the states.

Since the objective is to use the data $\mathbf{r}$ to obtain estimates for the parameters $\mathbf{\Theta}$, the transition matrix C and the sequence of hidden states $\mathbf{X}$, we will focus on finding the posterior distribution $p(\mathbf{\Theta}, C, \mathbf{X} \mid \mathbf{r})$. As Steinücke et al. [16], we use a Bayesian MCMC method. Let $\mathrm{p}(\mathbf{\Theta}, C)$ be the prior of the joint distribution of $\mathbf{\Theta}$ and C. Applying equation (4.3.7), with $p(\mathbf{r} \mid \mathbf{\Theta}, C, \mathbf{X})$ being the likelihood, we obtain the posterior distribution

$$p(\mathbf{\Theta}, C, \mathbf{X} \mid \mathbf{r}) \propto p(\mathbf{r} \mid \mathbf{\Theta}, C, \mathbf{X})\, p(\mathbf{X} \mid \mathbf{\Theta}, C)\, p(\mathbf{\Theta}, C). \tag{4.3.7}$$

We continue as demonstrated in [16] and [29] by augmenting the data. That is, treating the hidden sequence $\mathbf{X}$ as unknown, which will allow for Gibbs from the posterior distribution. We simplify the posterior by assuming the parameter sets $\mathbf{\Theta}$ and the transition matrix C to be independent [16]. Then the prior from equation (4.3.7) becomes

$$p(\mathbf{\Theta}, C) = p(\mathbf{\Theta})\, p(C).$$

Moreover, the two parameter sets are assumed to be independent conditional on the hyperparameters. Let ω denote the set of hyperparameters, where ω is the same for both states [29]. Thereby, the joint probability

distribution of the parameter sets and the hyperparameters become

$$p(\theta_1, \theta_2, \delta) \propto p(\delta)\, p(\theta_1 \mid \delta)\, p(\theta_2 \mid \delta). \tag{4.3.8}$$

From the Vasciek model in equation (4.3.6), the collection of the parameter sets becomes

$$\boldsymbol{\Theta} = (\theta_1, \theta_2) = (\{\gamma_1, \delta_1, \sigma_1^2\}, \{\gamma_2, \delta_2, \sigma_2^2\}).$$

The most common way of estimating these parameters is to first condition on knowing the sequence $\mathbf{X}$ [29], which is also the case for the transition matrix C. Applying equation (4.3.7) for the distribution of $(\boldsymbol{\Theta}, C)$ then yields the posterior distribution

$$p(\boldsymbol{\Theta}, C \mid \mathbf{X}, \mathbf{r}) \propto p(\mathbf{X}, \mathbf{r} \mid \boldsymbol{\Theta}, C)\, p(\boldsymbol{\Theta}, C).$$

Repeating this for the likelihood distribution $p(\mathbf{X}, \mathbf{r} \mid \boldsymbol{\Theta}, C)$, the posterior takes the form

$$p(\boldsymbol{\Theta}, C \mid \mathbf{X}, \mathbf{r}) \propto p(\mathbf{r} \mid \mathbf{X}, \boldsymbol{\Theta}, C)\, p(\mathbf{X} \mid \boldsymbol{\Theta}, C)\, p(\boldsymbol{\Theta}, C). \tag{4.3.9}$$

As Steinrücke et al. [16] argues, a next state $j+1$ will only depend on the transition matrix C and the state j. Thus, $p(\mathbf{X} \mid \mathbf{r})$

$$p(\mathbf{X} \mid \boldsymbol{\Theta}, C) = p(X_0 \mid C) \prod_{s=1}^{2} C_{s,1}^{N_{s,1}(\mathbf{X})}\, C_{s,2}^{N_{s,2}(\mathbf{X})}, \tag{4.3.10}$$

where $N_{s,j}(\mathbf{X})$ is the number of transitions from state s to state j in the sequence $\mathbf{X}$, and $p(X_0 \mid C)$ indicates the steady state probability based on C. For the likelihood distribution $p(\mathbf{r} \mid \mathbf{X}, \boldsymbol{\Theta}, C)$, we obtain

$$p(\mathbf{r} \mid \mathbf{X}, \boldsymbol{\Theta}, C) = \prod_{j=1}^{N} p(r_j \mid \theta_{X_{j-1}}), \tag{4.3.11}$$

where the subscript of θ indicates the state at time $j-1$. The dependency on C and the other states than X_{j-1} is removed as the probability of an observation r_j will follow the Vasicek model. The same argument determines the form of equation (4.3.11). As observation r_j is assumed to be normally distributed with mean

$\gamma_m + \delta_m v_{j-1}$ and variance $\sigma_m^2 \Delta t$, if $X_{j-1} = s$, the likelihood reads

$$p(\mathbf{r} \mid \mathbf{X}, \mathbf{\Theta}, C) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_{X_{j-1}}\Delta t}} \exp\left(-\frac{(r_j - \gamma_{X_{j-1}} - \delta_{X_{j-1}} r_{j-1})^2}{2\,\sigma^2_{X_{j-1}}\Delta t} \right). \tag{4.3.12}$$

Lastly, most of the parameters considered independent [16]. Also, the initial state X_0 is assumed to be independent of the parameters, making

$$p(\gamma, \delta, \boldsymbol{\sigma^2}, C, X_0) \propto p(\gamma, \delta)\, p(\delta)\, p(C)\, p(X_0). \tag{4.3.13}$$

Now, equation (4.3.9) only needs priors for $\mathbf{\Theta}$ and C in order to be tractable. The prior for the transition matrix is chosen to be Dirichlet [29], [30], with number of transitions between the different states as argument. Row s of C reads

$$(C_{s,1}, C_{s,2}) \sim \mathcal{D}(e_{s,1}, e_{s,2}),$$

where $e_{s,j}$ is the prior number of transitions from state s to state j. Following [29], the rows of C are assumed independent a priori, making them the a posteriori independent by letting

$$C_{s,1}, C_{s,2} \mid \mathbf{X} \sim \mathcal{D}(e_{s,1} + N_{s,1}(\mathbf{X}), e_{s,2} + N_{s,2}(\mathbf{X})), \tag{4.3.14}$$

where $N_{s,j}(\mathbf{X})$ is the number of transitions from s to j in the sequence $\mathbf{X}$.

For a detailed derivation of the priors distributions of the parameters $\mathbf{\Theta}$, see Steinrücke et al. [16]. The idea is to choose priors that are conjugate priors for the likelihood distribution $p(\mathbf{r} \mid \mathbf{\Theta}, C, \mathbf{X})$ [29]. The pair (γ_s, δ_s) is chosen to be have a bivariate normal distribution, while σ_s^2 has an inverse gamma prior distribution. That is,

$$\begin{aligned} p(\gamma_s, \delta_s) &\sim \mathcal{N}(\mu_0, \Sigma_0) \\ p(\sigma_s^2) &\sim IG(f_0, g_0), \end{aligned}$$

where μ_0, Σ_0, f_0 and g_0 are also defined in [16]. As all priors are conjugate to the likelihood, we obtain the following posteriors

$$\begin{aligned} \gamma_s, \delta_s \mid \mathbf{X}, \sigma_s^2 &\sim \mathcal{N}(\mu_s, \Sigma) \\ \sigma_s^2 &\sim IG(f_s, g_s) \end{aligned} \tag{4.3.15}$$

where $s = 1, 2$. Again, μ_s, Σ, f_s and g_s are defined in [16]. Hence, we now have a way of estimating the

parameters $\mathbf{\Theta}$ and C.

The final step is to estimate the unknown sequence $\mathbf{X}$. As explained in subsection 10.5.6 of [29], the estimation is done through a single-move Gibbs sampler. The samples are drawn from $p(\mathbf{X} \mid \mathbf{\Theta}, C, \mathbf{r})$, and state number j makes use of the other states. By applying equation (4.3.7), we attain the following relationship

$$p(X_j \mid \mathbf{X}_{-j}, \mathbf{\Theta}, C, \mathbf{r}) \propto p(\mathbf{r} \mid \mathbf{X}, \mathbf{\Theta}, C)\, p(\mathbf{X} \mid C), \tag{4.3.16}$$

where $\mathbf{X}_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N)$. The likelihood $p(\mathbf{r} \mid \mathbf{X}, \mathbf{\Theta}, C)$ is calculated using equation (4.3.12), and for $p(\mathbf{X} \mid C)$ we utilize equation (4.3.10). By inserting this into equation (4.3.16) and disregarding all proportionality constants, the situation is simplified

$$p(X_j = s \mid \mathbf{X}_{-j}, \theta_s, C, \mathbf{r}) \propto p(r_j \mid r_{j-1}, \theta_s, C)\, C_{X_{j-1},s} C_{s,X_{j+1}}. \tag{4.3.17}$$

The states are sampled sequentially from $j = 0$ to $j = N$, implying that X_{j+1} is not known when estimating $p(X_j \mid \mathbf{X}_{j-1}, \mathbf{\Theta}, C, \mathbf{r})$. Since we are using a MCMC approach, we find X_{j+1} from the previous Monte Carlo iteration. Hence, equation (4.3.17) takes the following form for iteration $k > 0$,

$$p(X_j^{(k)} = s \mid \mathbf{X}_{-j}^{(k-1)}, \theta_s^{(k)}, C^{(k)}, \mathbf{r}) \propto p(r_j, \mid r_{j-1}, \theta_s^{(k)})\, C^{(k)}_{X_{j-1}^{(k)},s}\, C^{(k)}_{s,X_{j+1}^{(k)}} \tag{4.3.18}$$

There are two edge cases for which this expression does not hold. At $j = 0$, X_{j-1} does not exist, and at $j = N$, X_{j+1} is not available. Thus,

$$\begin{aligned} p(X_0^{(k)} = s \mid \mathbf{X}_{-0}^{(k-1)}, \theta_s^{(k)}, C^{(k)}, \mathbf{r}) &\propto p(S_0 = s \mid C)\, C^{(k)}_{s,X_{j+1}^{(k-1)}} \\ p(X_N^{(k)} = s \mid \mathbf{X}_{-N}^{(k-1)}, \theta_s^{(k)}, C^{(k)}, \mathbf{r}) &\propto p(r_N \mid r_{N-1}, \theta_s^{(k)})\, C^{(k)}_{X_{N-1}^{(k)},s}. \end{aligned} \tag{4.3.19}$$

Now we have all the ingredients needed for the Markov-switch algorithm. The Markov-switch algorithm is summarized

1. Set hyperparameters μ_0, Σ, f_0, g_0, $e_{1,1}$, $e_{1,2}$, $e_{2,1}$, $e_{2,2}$.
2. Sample priors for $s = 1$ and $s = 2$
   - $(\gamma_s^{(0)}, \delta_s^{(0)}) \sim \mathcal{N}(\mu_0, \Sigma_0)$

- $\sigma_s^{2(0)} \sim IG(f_0, g_0)$
- $(C_{s,1}^{(0)}, C_{s,2}^{(0)}) \sim \mathcal{D}(e_{s,1}, e_{s,2})$

3. Set an initial sequence $\mathbf{X}^{(0)}$ by sampling from C

   The following steps are repeated for $k = 1, \ldots, R$, and for $s = 1, 2$

4. Sample row s of the transition matrix $C^{(k)}$ using equation (4.3.14) with $\mathbf{X}^{(k-1)}$
5. Sample $(\gamma_s^{(k)}, \delta_s^{(k)})$ using equation (4.3.15) with $\mathbf{X}^{(k-1)}$ and $\sigma_s^{2(k-1)}$
6. Sample $\sigma_m^{2\ (k)}$ using equation (4.3.15) with $\mathbf{r}$, $\mathbf{X}^{(k-1)}$, $\gamma_s^{(k)}$, $\delta_s^{(k)}$
7. Sample $\mathbf{X}^{(k)}$ using equation (4.3.18) and (4.3.19) with the updated parameters $\mathbf{\Theta}^{(k)}$ and $C^{(k)}$.

The first R_B are considered as burn-in iterations and are disregarded. The parameter values are set according to

$$\bar{\theta}_j = \frac{1}{R - R_\mathrm{B}} \sum_{k=R_\mathrm{B}+1}^{R+R_\mathrm{B}} \theta_j^{(k)}, \tag{4.3.20}$$

where θ_j is one of the parameters in $\mathbf{\Theta}$ and $\theta_j^{(k)}$ is the sampled value for the parameter for Monte Carlo iteration k. The same procedure is applied to estimate the most likely sequence $\mathbf{X}$, where the result is rounded to the closest integer. The transition matrix C is estimated based on $\mathbf{X}$ in the following manner

$$C_{i,j} = \frac{\sum_{n=1}^{N-1} I(X_n = i)\, I(X_{n+1} = j)}{\sum_{n=1}^{N} I(X_n = i)}, \qquad i, j \in \{1, 2\}, \tag{4.3.21}$$

which concludes the Markov-switching algorithm.

A verification of the robustness of the algorithm and a verification of the implementation is found in Appendix B.

5 Data Analysis

The data set used in this thesis contains daily US treasury bond rates with different maturities. This data is provided by the US treasury and is publicly available. The first rates are given at 02.12.1990 and the last are given at 15.06.2017. The maturities quoted in the market are one, two, three, five, seven and ten years. In order to obtain forward rates at an annual basis, we apply linear interpolation, i.e we get the rates having maturity of four, and six years from interpolating. This section will present and discuss behaviour of the forward LIBOR rates in the context of jumps and shifting periods.

5.1 Historical LIBOR rates

We begin by focusing on the full data set, and give a complete data analysis. Firstly, we use the provided rates to price zero-coupon bonds, using equation (3.1.4), given a face value of 1 and assuming constant interest rate between the tenor dates. Next, we apply the zero-coupon bond prices to find historical forward LIBOR rates through equation (3.1.6), where the accrual period is set to 1 year. This is repeated for all maturities and for all historical rates. The daily historical forward LIBOR rates are shown in Figure 2, and descriptive statistics of the rates are presented in Table 1. Further, the volatility observed during the last 500 trading days are found using the EWMA technique described in section 4.1.2. The corresponding covariances and correlations between the log-returns of the rates are presented in Table 5 and 3, respectively.

Table 1: Descriptive statistics of the historical forward LIBOR rates and their log-returns.

Maturity	Mean (%)	Variance (%)	Log-returns mean (%)	Log-return variance (%)
1Y	3.8913	0.0639	-0.0248	0.1905
2Y	4.2796	0.0541	-0.0223	0.1220
3Y	4.6739	0.0462	-0.0201	0.0548
4Y	5.1186	0.0435	-0.0182	0.0553
5Y	5.2469	0.0378	-0.0182	0.0344

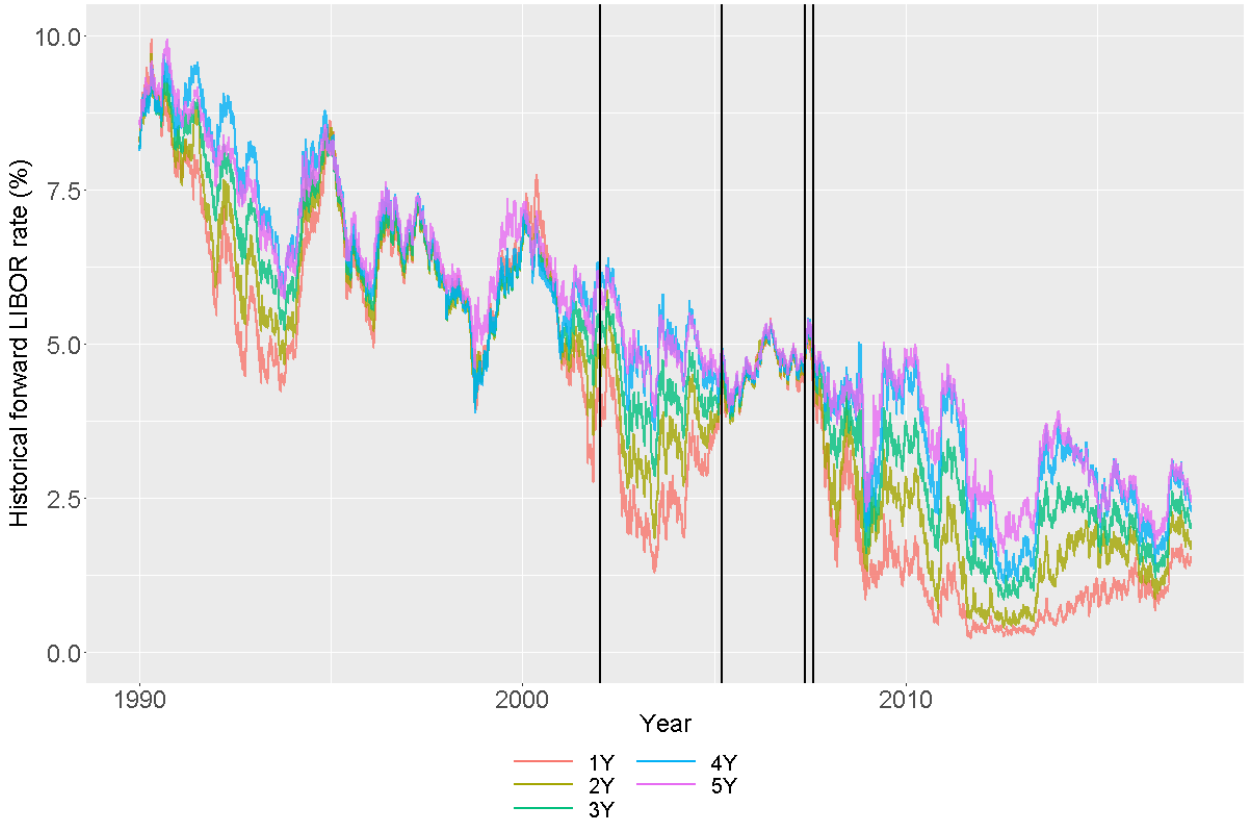


Figure 2: Historical forward LIBOR rates in the US market. The vertical lines indicate the end of period I, II, III and IV, respectively.

We first observe from Figure 2 that the time series are far from stationary, where the volatility seem to vary and the rates from time to time have sudden large hikes and drops. Furthermore, the rates seem to follow each other rather closely, which is also seen from the correlation matrix in Table 3. Here, the rates of higher maturity have greater correlations. Table 1 shows that the variance of the rates tends to decrease with increasing maturity, however the mean is higher for the higher maturities. The same applies for the variance of the log-returns, while the means of log-returns are close to zero.

Considering the vertical distance of the rates at a fixed time point gives a yield curve. The shape of these yield curves is assumed to have significant macro economic implications [4]. During the period 2002 to 2005 the forward rates operate at rather different levels giving an upward sloping yield curve implying that investors demand lower yields on short term investments. In macro economics this is considered a sign that the current economic regime involves stability and growth [21]. From 2005 to 2008 the rates are almost the same giving a flat or even inverted yield curve, implying investors consider the short term investment risky

[21]. This period coincides with the financial crisis that greatly affected the American economy[4]. There has also been a slight spike in recent times, specifically the autumn of 2016. The spike happened the same day as election day in the US, where some [31] [32] attribute the change of presidents for the rise in interest.

Table 2: Covariance matrix of the log-returns of the historical forward LIBOR rates.

	1Y	2Y	3Y	4Y	5Y
1Y	0.00190	0.00079	0.00072	0.00064	0.00045
2Y	0.00079	0.00122	0.00059	0.00052	0.00042
3Y	0.00072	0.00059	0.00055	0.00054	0.00035
4Y	0.00064	0.00052	0.00054	0.00055	0.00034
5Y	0.00045	0.00042	0.00035	0.00034	0.00034

Table 3: Correlation matrix of the log-returns of the historical forward LIBOR rates.

	1Y	2Y	3Y	4Y	5Y
1Y	1.00000	0.52001	0.70078	0.62624	0.55414
2Y	0.52001	1.00000	0.72189	0.63696	0.64818
3Y	0.70078	0.72189	1.00000	0.98680	0.80373
4Y	0.62624	0.63696	0.98680	1.00000	0.78002
5Y	0.55414	0.64818	0.80373	0.78002	1.00000

Table 4: Covariance matrix estimated via EWMA for the log-returns of the most recent 500 historical forward LIBOR rates.

	1Y	2Y	3Y	4Y	5Y
1Y	0.00061	0.00029	0.00031	0.00030	0.00022
2Y	0.00029	0.00068	0.00040	0.00039	0.00034
3Y	0.00031	0.00040	0.00038	0.00041	0.00026
4Y	0.00030	0.00039	0.00041	0.00044	0.00027
5Y	0.00022	0.00034	0.00026	0.00027	0.00028

5.2 Economic states

We now represent the different economic regimes. The Markov-switching algorithm described in section 4.3.2 is applied at the log returns of the historical LIBOR rates of all maturities. Thus we obtain five paths of states describing the historical economic states. As described in the last section an overall most likely path across all maturities is found by taking the average and then rounding. The resulting overall most likely path for the whole data set is presented in Figure 3 together with the historical log-returns of the forward LIBOR rates for all maturities.

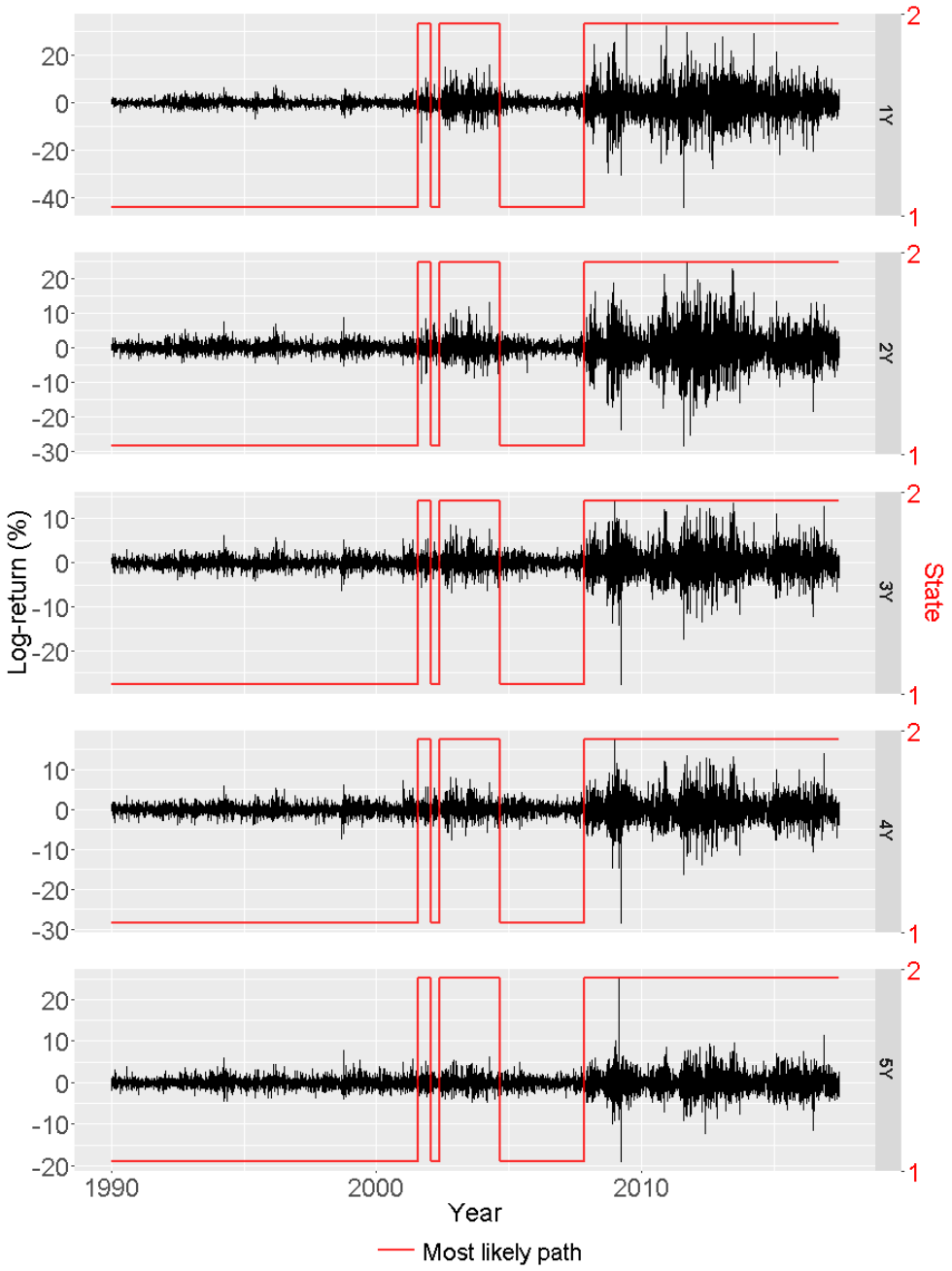


Figure 3: Log-returns with the overall most likely path.

All the maturities seem to have comparable patterns in terms of periods of low and high volatility. Further the overall most likely path seems to identify these periods in a reasonable manner. The transition matrix

according to the overall most likely state path in Figure 3 was estimated to

$$\hat{C} = \begin{bmatrix} 0.994 & 0.006 \\ 0.009 & 0.991 \end{bmatrix}, \tag{5.2.1}$$

which yields a stationary distribution of $[0.608, 0.392]^\intercal$. Consequently, the probability of staying in the same state is high. However, the probability is higher for staying in state one than state two, which gives the difference in the steady state distribution. The data in state one is characterized with low volatility, and the data of state two has high volatility as seen in Figure 5. Note that EWMA is not used with the split dataset. The steady state vector implies that the regime of a stability occurs the most, while there are slightly shorter periods of volatile regimes. This is similar to the results seen in [16]. The two volatilities are clearly different, which testaments the different periods documented in [12]. Note however that we are considering the log returns of the rates; the exact macro economic interpretation of the states is not immediate. However, in terms of model behavior it is.

Table 5: Variance of the data in state one and state two.

Maturity	Variance state one	Variance state two
1Y	0.00027	0.00422
2Y	0.00020	0.00266
3Y	0.00013	0.00113
4Y	0.00015	0.00112
5Y	0.00012	0.00065

5.3 Existence of Jumps

The jump parameter calibration routine described in section 4.2.2 is applied to the log-returns of the historical forward rates depicted in Figure 2. The algorithm estimates the jump times, the intensity of jumps, in addition to the mean and the variance of the jump sizes. Figure 4 shows the data at which the algorithm is applied, along with the estimated jumps; indicated with red circles. Moreover, the estimated parameters from the algorithm is listed in Table 6.

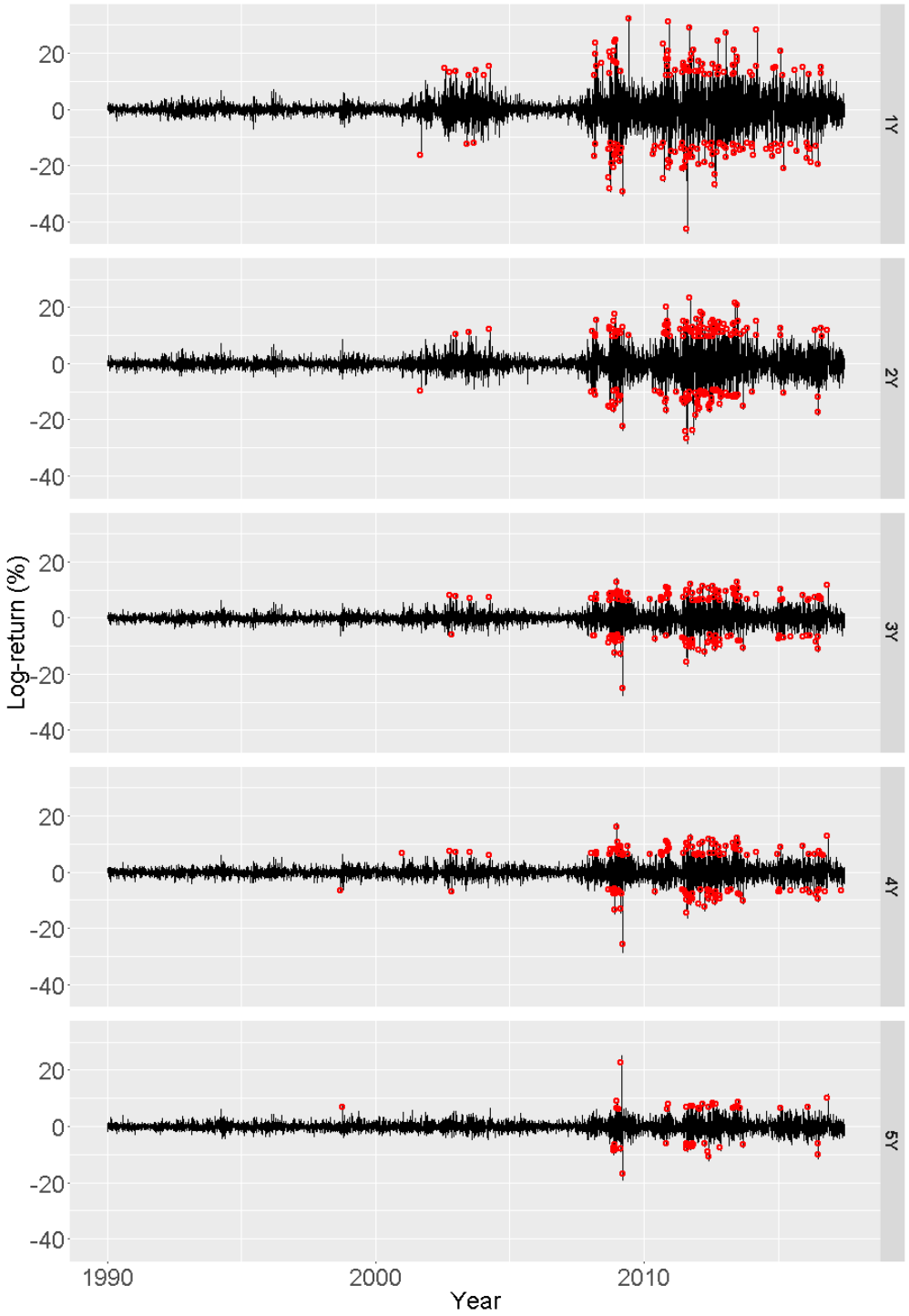


Figure 4: Log-returns of the historical forward LIBOR rates, including estimated jumps from the full data set. The red circles indicate jumps.

Figure 4 shows that the algorithm succeeds at identifying the jumps. The frequency of jump occurrences

seems to decrease with increasing maturity. This is also seen from the decreasing values of λ in the left of Table 6. The same tendency is found in the jump size variance. Lastly, the mean of the jump size is small for all maturities. All together, this supports the assumptions on the jump parameters made in 4.2.2. Moreover, the jump variance is clearly larger than the historic volatility found in Table 5 consequently when a jump occurs it should affect the interest rate in relatively large degree.

Table 6: Estimated jump mean, variance and intensity for the different data sets.

	Full data set			**Data set state 1**			**Data set state 2**		
Maturity	m_J	v_J	λ	m_J	v_J	λ	m_J	v_J	λ
1Y	0.0003	0.0292	0.0252	0.0016	0.0024	0.0295	-0.0029	0.0218	0.0465
2Y	0.0015	0.0171	0.0249	0.0045	0.0031	0.0032	0.0078	0.0147	0.0342
3Y	0.0014	0.0073	0.0220	0.0016	0.0025	0.0022	0.0074	0.0132	0.0094
4Y	0.0017	0.0073	0.0220	0.0041	0.0022	0.0015	0.0063	0.0132	0.0105
5Y	0.0032	0.0070	0.0070	0.0082	0.0002	0.0007	0.0004	0.0260	0.0021

Next, we calculate the volatilities of the log-returns after removing the identified jumps showed in Figure 4. In order to perform the calculation, we find the jump time of all jumps, for all maturities, and remove the data points found to these times. Then the EWMA estimation is applied to find the covariance matrix, which is shown in Table 7.

Table 7: Covariance matirx of the log-returns after removing identified jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00062	0.00027	0.00030	0.00029	0.00022
2Y	0.00027	0.00062	0.00033	0.00031	0.00030
3Y	0.00030	0.00033	0.00032	0.00033	0.00022
4Y	0.00029	0.00031	0.00033	0.00036	0.00022
5Y	0.00022	0.00030	0.00022	0.00022	0.00027

The covariance matrix in Table 5 and Table 7 are very similar. The first one is for the full data set, while the second is for the same data set, but without jumps. The fact that they are so similar is explained by studying the jump plots in Figure 4. Since the covariance is found by using EWMA from the last 500 days,

and these days contains very few jumps, the data set are approximately the same. Thus, we observe the similar covariance matrices.

When simulating under the MSJD LMM model, we need two parameter sets - one for each economic state. Therefore, we split our data set according to the overall most likely path showed in Figure 3. Next, we apply the jump calibration routine for each of the data sets, which results in the jump parameters presented in the center and right column of Table 6. The routine also estimates jumps as shown in Figure 5 and 6 for state one and two, respectively. The routine clearly manages to identify the jumps for both of these data sets too.

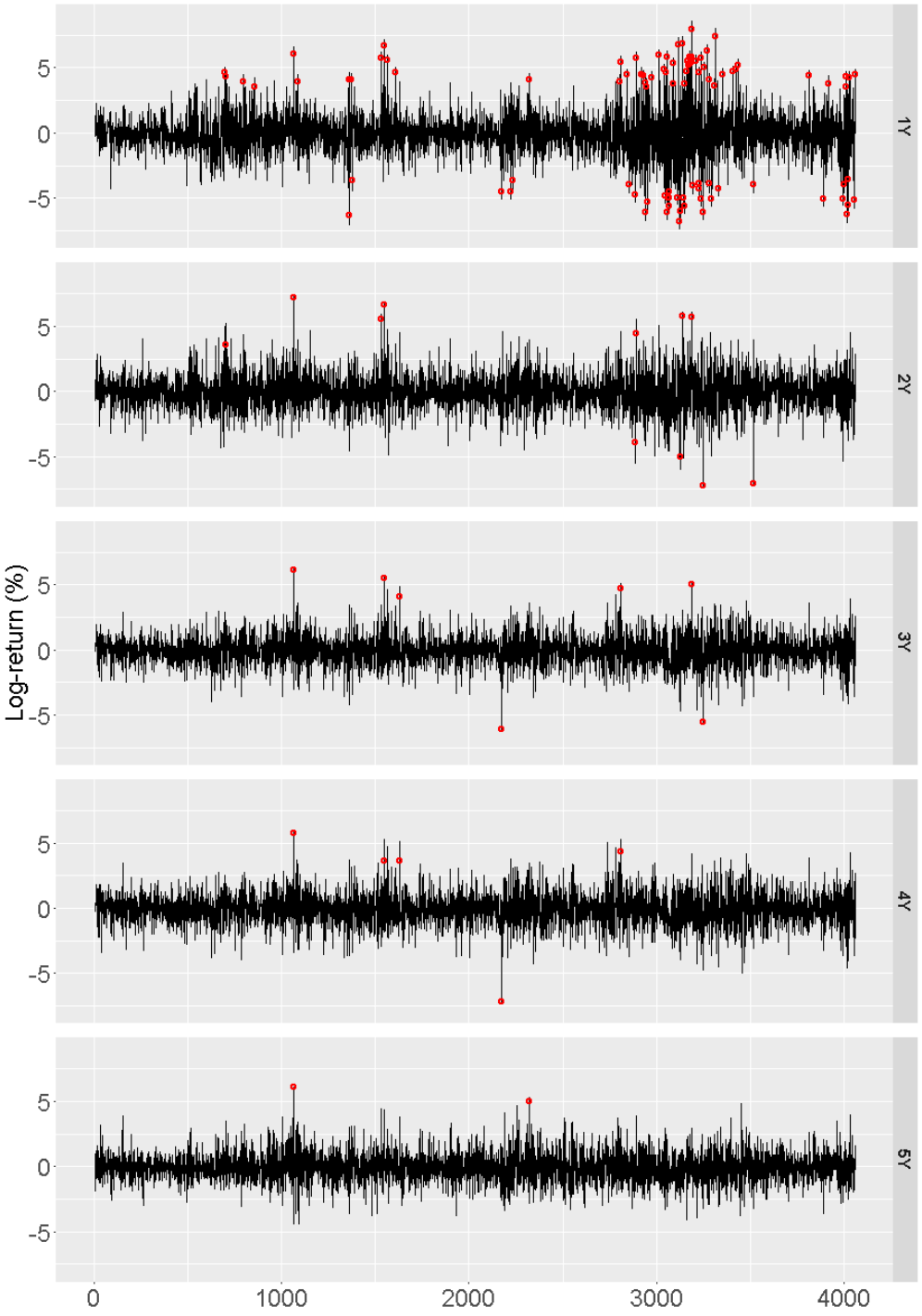


Figure 5: Estimated jumps from the data in state one. The red circles indicate jumps.

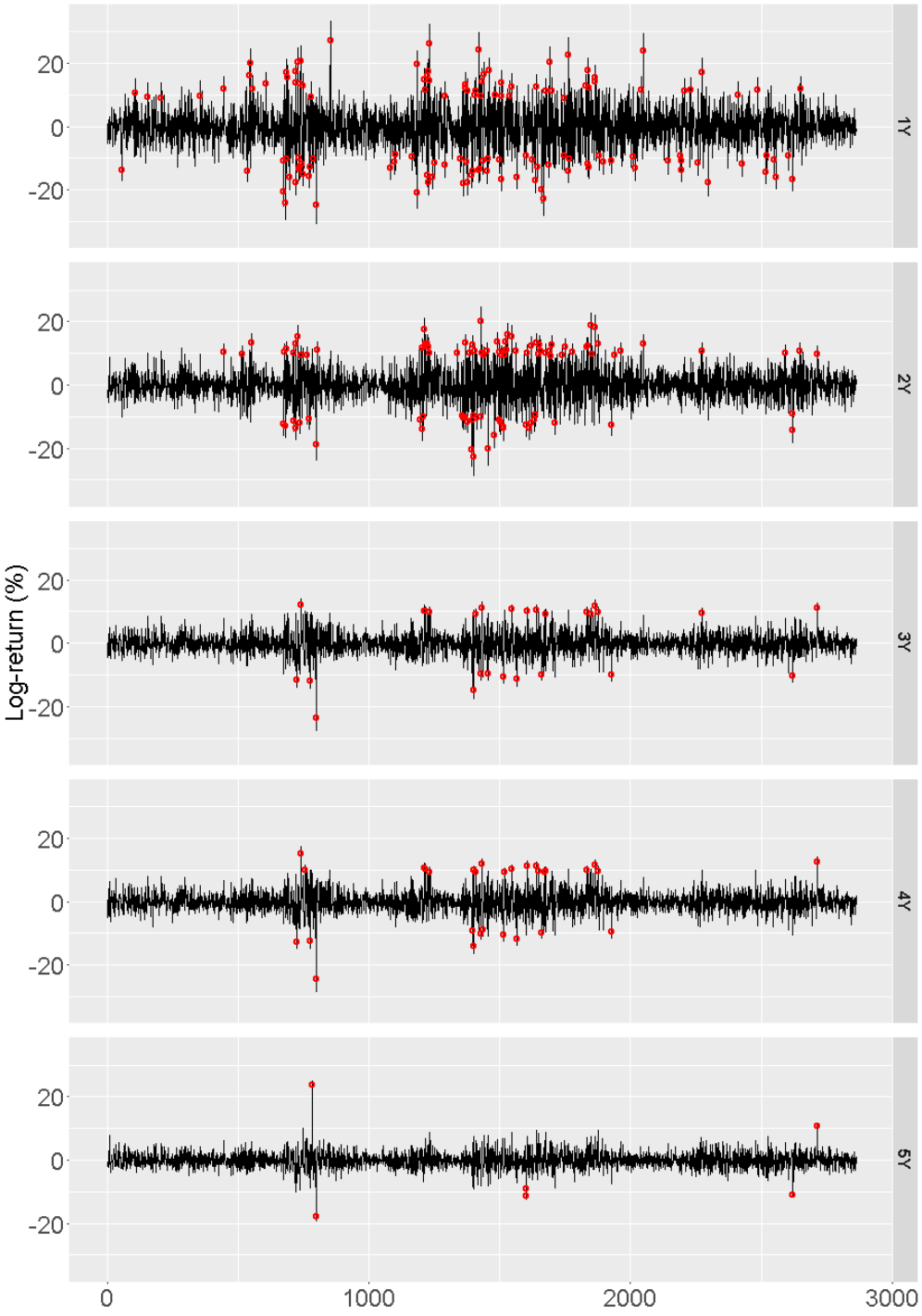


Figure 6: Estimated jumps from the data in state two. The red circles indicate jumps.

Firstly, when examining the estimated jump parameters in Table 6, we observe that the intensities of jumps,

and the jump size variances, are greater in state two than in state one. Typically, the jump intensity of the full data set is between the intensity of state one and two. The variance of the jump sizes, however, is greatest for the full data set. These considerations coincides with what was expected from the construction of the Markov-switching algorithm. Namely, the data of state one and two should have low and high volatility, respectively. As jumps cause the volatility to rise, these will inherently, often end up in state two. Since the full data set contains jumps from both states, it is sensible that the jump variances lie between the ones from the split data sets.

As for the full data set, the jump means are small for both the split data sets. Also, we generally find the pattern of decreasing variance and intensity for increasing maturity.

When finding the covariance of the split data sets, we focus on the full period in state one and two. That is, not only the last 500 trading days of each set. However, we are also removing the jumps to find the covariance in the ordinary diffusion of the MSJD LMM. The resulting covariance matrix for state one and two is presented in Table 8 and 9, respectively.

Table 8: Covariance matrix of the log-returns of the data in state one, after removing jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00015	0.00009	0.00009	0.00009	0.00007
2Y	0.00009	0.00015	0.00008	0.00008	0.00008
3Y	0.00009	0.00008	0.00010	0.00011	0.00007
4Y	0.00009	0.00008	0.00011	0.00013	0.00007
5Y	0.00007	0.00008	0.00007	0.00007	0.00011

Table 9: Covariance matrix of the log-returns of the data in state two, after removing jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00190	0.00073	0.00069	0.00061	0.00044
2Y	0.00073	0.00119	0.00056	0.00049	0.00043
3Y	0.00069	0.00056	0.00060	0.00060	0.00038
4Y	0.00061	0.00049	0.00060	0.00062	0.00037
5Y	0.00044	0.00043	0.00038	0.00037	0.00042

First of all, the variances found in both tables are significantly smaller than their respective jump size variances. We also notice that the covariances of state one are smaller than those of state two, while the level of the covariance matrix of the full data set without jumps is between these values. Therefore, we can conclude that state one and two clearly have different characteristics, in terms of both ordinary diffusion and jumps. Moreover, the full data represents a composition of the split data sets. Consequently, the Markov-switch algorithm together with the jump calibration routine has yielded explainable results.

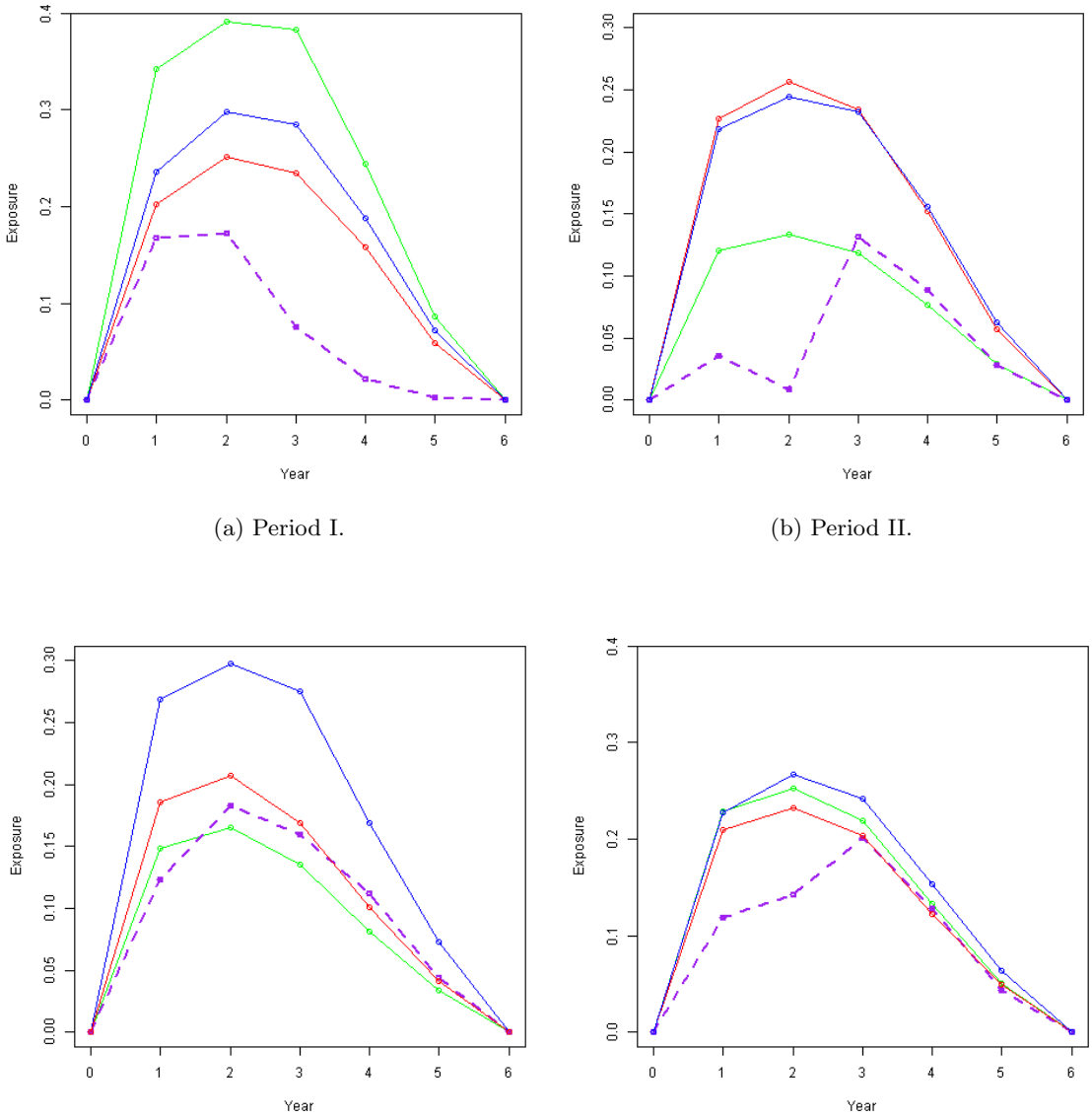
6 Results

We begin this section with displaying the results from the backtesting of potential future exposure described in section 3, as a crude measure of model performance. Secondly, we simulate forward LIBOR rates using the three models presented in section 3. Lastly, we compare the potential future exposure associated with a portfolio of caps and floors, based on the simulations of the forward LIBOR rates. Discussion will follow when all results are presented.

The portfolio used throughout consists of three caps and three floors, all having a strike rate equal to the mean of the last known historical rates. The maturities of the caps and floors are three, four and five years. That is, the portfolio has pairs of caps and floors at equal strike rates, but with different maturities. This construction follows Ng [23] for the purpose of reasonable backtesting. The price formulas given in 3.1.7 is used.

6.1 Backtesting

We have backtested the three models calibrating the model to historical data from the periods starting at 1990 and ending at the different vertical lines seen in Figure 2. For each period, we forecast forward LIBOR rates for each day 5 years ahead, that is where the all derivatives have matured, and then estimate the potential future exposure with yearly increments. The forecasts are then compared to the calculated actual PFE for the portfolio during that period.



(a) Period I.

(b) Period II.

(c) Period III.

(d) Period IV.

Figure 7: The estimated 95 % potential future exposure under all three models, in four different periods. The green line indicates the log-normal LMM, the red line is the jump diffusion LMM and the blue line is the MSJD LMM. Further, the purple line is the observed exposure.

6.2 Forward LIBOR rates

The three models described in section 3 are applied to forecast LIBOR rates from the end of the dataset; june 2017. We choose to make predictions five years into the future, using annually predictions. The results are

found using the simulation strategies described in section 4.1.1, 4.2.1 and 4.3.1. All models were simulated by 30000 Monte Carlo iterations, with the initial rates equal to the last known historical rates. The distribution of the forward LIBOR rates for the three models are shown for the rates $L_1(1)$, $L_3(3)$ and $L_5(5)$ in Figure 8. The corresponding descriptive statistics are listed in Table 10.

When the log-normal LMM was simulated, we found the volatilities from the covariances presented in Table 5. For the jump diffusion LMM, the jump parameters from the left of Table 6 were applied for the jump part, while the covariances from Table 7 governed the ordinary diffusion part of the process. Finally, for the MSJD LMM, we used the estimated transition matrix of (5.2.1), the jump parameters in the middle and to the right of Table 6, and lastly the covariances of Table 8 and 9.

Table 10: Descriptive statistics of the simulated forward rates.

Model	**Maturity**	**Mean** (%)	**Variance** (%)	**95 % VaR**
Log-normal LMM	1Y	1.5070	0.0038	2.6660
	2Y	1.7994	0.0124	3.8876
	3Y	2.0735	0.0224	4.8892
	4Y	2.3813	0.0413	6.0949
	5Y	2.4413	0.0492	6.4337
Jump diffusion LMM	1Y	1.5196	0.0096	3.3522
	2Y	1.8696	0.0346	5.1740
	3Y	2.3051	0.0812	7.0648
	4Y	2.7892	0.1958	9.0492
	5Y	2.8942	0.2051	9.7948
MSJD LMM	1Y	1.5431	0.0085	3.1876
	2Y	1.9133	0.0251	4.6322
	3Y	2.2515	0.0421	5.7746
	4Y	2.6338	0.0720	7.2406
	5Y	2.7636	0.0885	7.6585

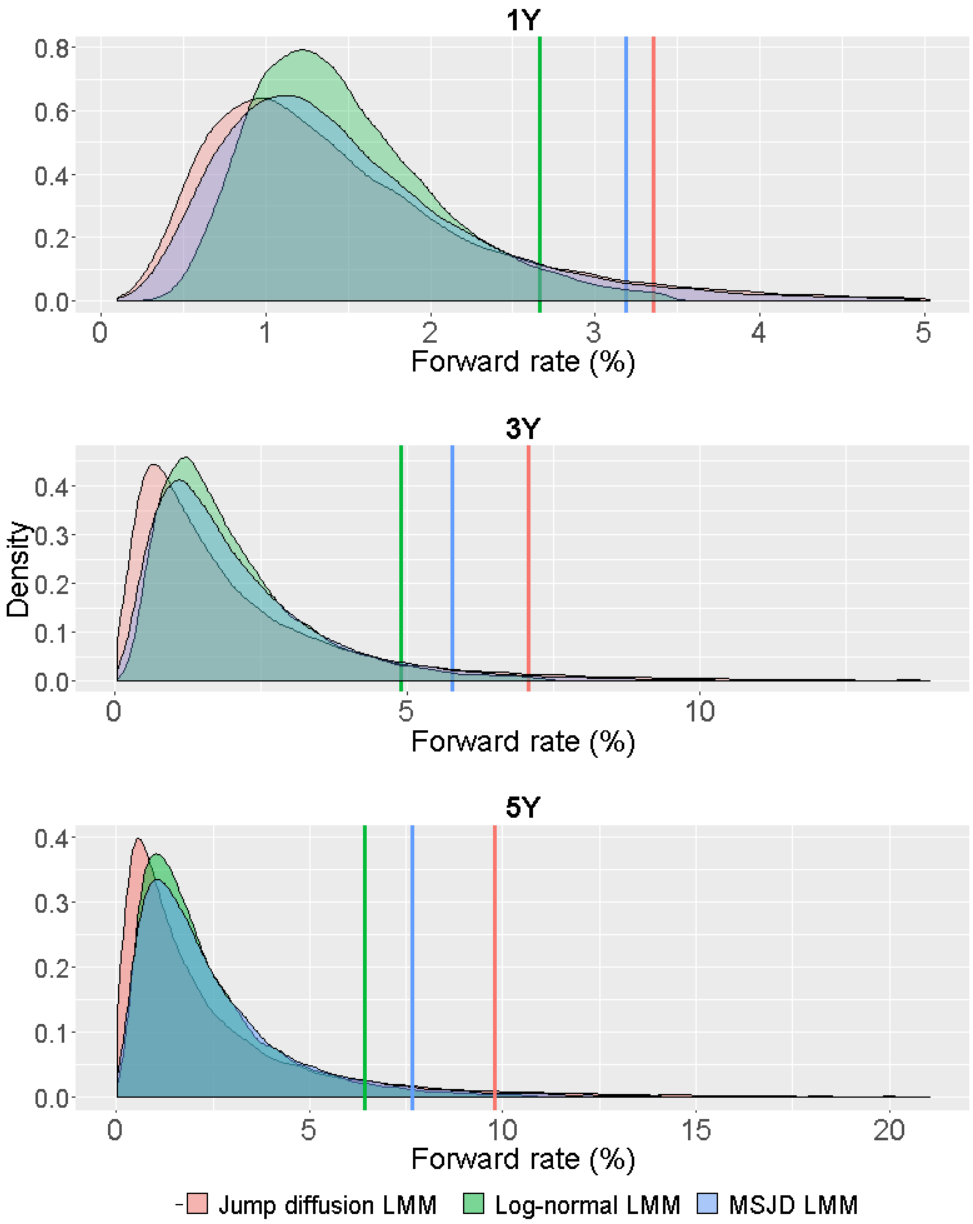


Figure 8: Simulated forward LIBOR rates under the three different models. The vertical lines indicate 95 % VaR, where the colour of the lines and the distributions corresponds.

In table 11 we see the historical kurtosis caluculated from the first-differenced actual 1-year US dollar LIBOR rate compared to the kurtosis of the simulated forward LIBOR rate L_1 first-differenced for the different periods.

Table 11: Historical kurtosis for the four periods, and the estimated kurtosis for the three models.

Period	Historical	Log-normal LMM	Jump diffusion LMM	MSJD LMM
I	4.3878	3.4700	3.1292	4.5378
II	7.7175	3.0253	21.4632	8.8913
III	7.5488	3.0460	18.0561	23.6177
IV	6.6881	3.1383	16.8138	14.8126

6.2.1 Forecasted Potential Future Exposure

The simulated of forward rates are applied to find the 95 % PFE associated with a portfolio. The results are shown in Figure 9.

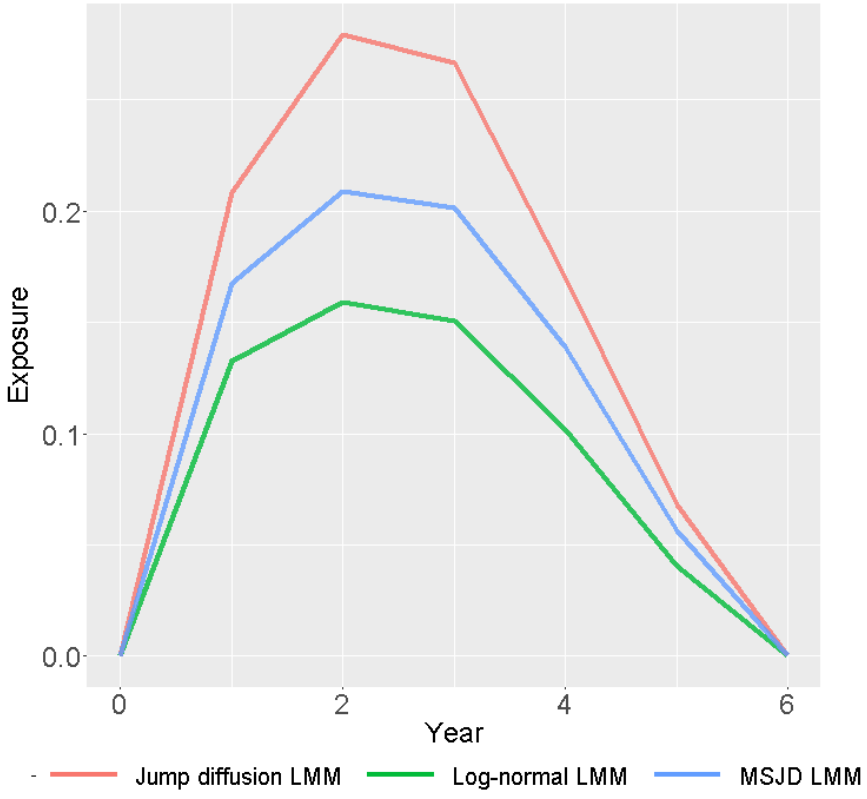


Figure 9: 95 % potential future exposure from the three models.

7 Discussion

We start by evaluating Figure 8 and Table 10, where forecast distributions have been obtained by calibrating to the entire data set. It is evident that the mean of the forward rates are relatively close to the rate level for the initial rates. This is sensible considering the arbitrage free drift specification of the LIBOR market model where the drift related to the chosen measure is relatively slight. The variance of the forward rates increases with higher maturity. This is expected, since the further we simulate into the future, the greater the uncertainty. Further, as seen in the Table 10, the forecast distributions under the jump diffusion LMM yields highest variance, whilst the MSJD LMM simulations have higher variance than the log-normal LMM simulations. The vertical lines representing 95% VaR follows the same pattern.

In Figure 9 one can observe the PFE related to the simulations shown in Figure 8. The PFE is calculated annually over then entire life time of the portfolio. It is evident that the exposure profile for all models peaks towards the center of the considered forecasting time. Also, the shapes of the profiles from the models are rather similar. The PFE of a portfolio of caps and floors with caplets and floorlets that roll off as time goes by should have have a shape similar to what we observe, where the risk becomes zero when all derivatives have matured. The start exposure is relatively close to zero due to the strike rates being close to the initial rates.

We furthermore see that the jump diffusion LMM produces highest PFE values for all future times, and the log-normal LMM yields the lowest PFE. Based on the high relative magnitudes of the variance and the 95 % quantile in the forward rates in Figure 8, this result is as expected. Further, the PFE estimates in Figure 9 are higher for the jump diffusion LMM than for the MSJD LMM. By recalling the data analysis made in section 5, we are able to indicate the rationale behind this relative behaviour. First of all, a higher PFE estimate might be caused by larger elements in the covariance matrix. When studying the covariance matrices from Table 7, 8 and 9, we see that the covariance matrix of the full data set, without jumps, is closer to the covariances of state two than state one. Consequently, the ordinary diffusion in the jump diffusion LMM is closer to the ordinary diffusion of state two of the MSJD LMM. Additionally, the jump parameters presented in Table 6 are in most cases higher for the full data than for data set one. Therefore, the jump diffusion LMM will have more variance than state one of the MSJD LMM. For state two of the MSJD LMM, however, we see that the covariance matrix are larger than the EMWA estimate used for jump diffusion. Further many of the jump parameters in state two are higher than jump calculated on the

whole data set. Still, since these parameters are sufficiently close, and that the stationary distribution of the estimated transition matrix in (5.2.1) is towards state one, the variance of the jump diffusion LMM is naturally predicted to be higher than for the MSJD LMM.

The results of a crude backtest can be seen in Figure 7 where 95% PFE is calculated by calibrating the data on four different periods. The dotted line then represent the actual PFE calculated using historic data in the time interval we forecast into to. We see breaks in the estimated PFE. That is, some models underestimate the exposure that was observed during the simulation period. This is the case for the log-normal LMM in period II, III and IV. For the jump diffusion LMM, it underestimates the PFE in period III and IV. The MSJD LMM, on the other hand, estimates sufficiently large counterparty risk at all these periods. This is not in any way a rigours back test, but it demonstrates some promise for the MSJD LMM.

Further, the 95 % PFE profiles seen in Figure 7 are quite different. For period I, the PFE estimate under the log-normal LMM is higher than for the jump diffusion LMM and the MSJD LMM. This stems from the occurrence of many jumps during the end of this period. Remember that EWMA will be very sensitive to the most recent fluctuations. In the jump diffusion LMM and the MSJD LMM, we calibrate the ordinary diffusion by removing the jumps from the data set, and then find the remaining covariance. Thus, when the jumps are removed from the most recent data, this shows to have a major effect on the jump diffusion model. The MSJD is not as sensitive to the procedure as it does not use EWMA, but rather the last state of the over all most likely path. For the log-normal LMM, however, these jumps are included in the estimation of covariance and hevily weighted by EMWA. Thus, the covariance applied in the log-normal LMM is high compared to the covariances applied in the other models. Even though the other models have several types of diffusion, they do not exceed the variance of the log-normal LMM in this scenario. This remarkably high 95 % PFE from the log-normal LMM gives an ex post motivation for our model extensions. The covariance estimated for the log-normal LMM yields simulations which overestimates the PFE by quite a lot. Thus we obtain a certain perspective: The jump diffusion LMM with EWMA is not as sensitive to sudden fluctuations in interest rates as the LMM, rather it tries to balance having updated volatility factors and at the same time include the notion of occasional jumps. The MSJD might be even less prone to recent fluctuations, because it would rather see it as a change of economic regime in a broader historic perspective.

The opposite is seen in Period III. Here, the MSJD LMM estimates higher exposure than the jump diffusion LMM. As discussed for the full data set, this stems from the relative parameters driving the variance of the models. In this case, we see from Table 15 that the covariance of the ordinary diffusion, i.e. without

jumps, is close to the covariance matrix of state one, see Table 16. Besides, the jump parameters seen in Table 14 are usually higher for state two. So, although the transition matrix of (A.0.1) implies a stationary distribution that is clearly in favor of state one, the variance from state two of the MSJD LMM makes the variance significantly higher than for the jump diffusion LMM.

Based on these cases, we can propose that the relative model behaviour between the jump diffusion LMM and the MSJD LMM is dependent on the parameters governing the variance in the jump diffusion LMM. If the parameters are close to the parameter values of state two in the MSJD LMM, we will expect the exposure estimation of the jump diffusion LMM to be higher than for the MSJD LMM. Given the opposite, where the jump parameters are close to those of state one, the MSJD LMM is expected to yield a higher estimated exposure.

The examples considered in this thesis are by no means exhaustive, but they illustrate clearly that the estimated PFE from the models can vary considerably. The crude backtesting of the models seen in Figure 7 furthermore shows that the traditional log-normal LMM, and even the jump diffusion extension, underestimates the PFE at 95 % level. As the MSJD LMM did not violate this percentile, it would be interesting to investigate this model further, in a more rigorous manner. However, as discussed in [23], there are yet no wide-spread adopted methodology for backtesting of PFE in terms of pure statistical tests. Testing the exposure estimation for more periods and a variety of portfolios would still be interesting in terms of exploring the model flexibility and behaviour.

In Table 11 we see the historic kurtosis against the predicted kurtosis of the models. The leptokurtic properties of the MSJD LMM and the jump diffusion LMM is clearly displayed, however with the given calibration they seem to be over predicting the kurtosis for some periods. The Log-normal clearly under predicts the kurtosis as expected. Similar to Yang [33] the first- differenced rates are used. The reason for this is that the 1-year LIBOR rate for the American market is actually not particularly leptokurtotic. From Yang [33] we see that significant leptokurtosis is find in the SHIBOR rate not differenced as well. Thus one might conclude that it would be particularly interesting to test several different markets showing leptokurtosis. Further, it should be noted that the shorter maturity LIBOR rates are more leptokurtotic thus that would be an interesting study as well [33].

8 Conclusion

In this thesis we have closely followed the papers from Steinrucke et al. [16], Glasserman and Merener [10] and Glasserman and Kou [14] to develop and implement a scheme that yields forecast distributions from the Markov-Switching Jump Diffusion LIBOR market model. A log-normal LMM and a jump diffusion LMM was also implemented for comparison. We have outlined in detail how one can calibrate the MSJD to historical forward LIBOR rates and why these parameters are reasonable in a comparison setting. The proposed calibration routine satisfactory identified regimes of different volatilities and spikes that could be considered jumps. When calculating potential future exposure with the different models we obtain clearly different results with the different models during different periods. Even though the backtesting for PFE is rather crude and it is not necessarily possible to conclude whether the MSJD captures tail risk more accurately, this testing demonstrates a particular behaviour. From this we extract a perspective on the MSJD as a model that focuses on including historical events rather than being highly sensitive to recent market events. The conclusion is thus that the Markov-switching jump diffusion extension is an interesting addition to the LIBOR market model family. It is of especially interest to practitioners who wants to increase the kurtoisis in their estimate in an empirically sound way.

Further research on the model can clearly be to search for a market and periods that are especially leptokurtotic. Together with a rigorous backtesting using developed statistical tests, this could better determine the usefulness of MSJD. Steinrücke also finds evidence in her paper that more than two regimes appear in markets, hence an infinite hidden Markov model could be used to find the optimal number of economic states. Further, different data sets could be used to identify the state of the economy such as the caplet volatilities used by Steinrucke et al[16].

References

[1] B. of International Settlements, "Http://www.bis.org/statistics/derstats.htm", 2017.

[2] T. W. Bank, "Http://data.worldbank.org/indicator/cm.mkt.lcap.cd", 2017.

[3] A. Papapantoleon, "Old and new approaches to LIBOR modeling", *Statistica Neerlandica*, vol. 64, pp. 257–275, 2010.

[4] P. Wilmott, S. Howison, and J. Dewynne, *The mathematics of financial derivatives*. Cambridge University Press, 1995.

[5] R. Rebanato, *Modern pricing of interest-rate derivatives: The libor market model and beyond*. Princeton University Press, 2002.

[6] F. Black, "The pricing of commodity contracts", *Journal of Financial Economics*, vol. 3, no. 1-2, pp. 167–179, 1976.

[7] P. Glasserman, *Monte carlo simulations in financial engineering*. Springer, 2004.

[8] Johannes, "The statistical and economic role of jumps in continuous-time interest rate models", 2004.

[9] S. R. Das, "The surprise element: Jumps in interest rates", 1999.

[10] Glasserman and Merener, "Numerical solution of jump-diffusion libor market models", 2002.

[11] Babbs and Webber, "Term structure modelling under alternative official regimes; in mathematics of derivative securities", 1997.

[12] L. El-Jahel and W. Perraudin, "Interest rate distributions, yield curve modelling and monetary policy", 1997.

[13] Rebonato and Joshi, "A joint empirical and theoretical investigation of the modes of deformation of swaption matrices: Implications for model choice", 2002.

[14] S.G.KOU and P. Glasserman, "THE TERM STRUCTURE OF SIMPLE FORWARD RATES WITH JUMP RISK", *Mathematical Finance*, vol. 13, no. 1, pp. 383–410, 2003.

[15] F. Jamshidian, "Libor market model with semimartingales", *Manuscript NetAnalytic Limited*, 1999.

[16] L. Steinrücke, R. Zagst, and A. Swishchuk, "The LIBOR Market Model: A Markov-Switching Jump Diffusion Extension", in *Hidden Markov Models in Finance: Further Developments and Applications, Volume II*, R. S. Mamon and R. J. Elliott, Eds. Boston, MA: Springer US, 2014, pp. 85–116, ISBN: 978-1-4899-7442-6. DOI: 10.1007/978-1-4899-7442-6_4.

[17] K. H. NG and Y. SUN, "Commercial banking risk management", ?

[18] Wernz, "Interest rates and how they work", 2007.

[19] J. Jacod and A. N. Shiryaev, "Limit theorems for stochastic processes", 2003.

[20] O. Cappé, E. Moulines, and T. Rydén, "Inference in hidden markov models", 2009.

[21] P. Wilmott, *Paul wilmott introduces quantitative finance*, 2nd ed. John Wiley Sons, 2007.

[22] BIS, "Basel iii: A global regulatory framework for more resilient banks and banking systems", 2011.

[23] W. Tian, "Commercial banking risk management", 2017.

[24] A. Brace, D. Gatarek, and M. Musiela, "The market model of interest rate dynamics", *Mathematical Finance*, vol. 7, no. 2, pp. 127–147, 1997.

[25] F. Jamshidian, "Libor and swap market models and measures", *Finance and Stochastics*, vol. 1, no. 4, pp. 293–330, 1997.

[26] P. Glasserman and X. Zhao, "Arbitrage-free discretization of lognormal forward Libor and swap rate models", *Finance and Stochastics*, vol. 4, no. 1, pp. 35–68, 2000.

[27] A. Carol, "Moving average models for volatility and correlation, and covariance matrices", 2007.

[28] D. Belomestny and J. Schoenmakers, "A jump-diffusion Libor model and its robust calibration", *Quantitative Finance*, vol. 11, no. 4, pp. 529–546, 2011.

[29] S. Frühwirh-Schnatter, *Finite mixture and markov switching models*. Springer, 2006.

[30] S. Chib, "Calculating posterior distributions and modal estimates in markov mixture models", 1996.

[31] CNBC, "Http://www.cnbc.com/2017/01/05/bond-prices-rise-as-fed-gears-up-to-raise-rates-to-counter-trump-inflation.
2016.

[32] Bloomberg, "Http://www.economist.com/news/finance-and-economics/21711036-recent-falls-bond-prices-and-rises-yiel
2016.

[33] B. Yang, "Regime switching in dynamics of risk premium: Evidence from shibor", 2011.

A Data analysis of Period III

This section contains a data analysis performed on a subset of the data in Period III. The estimations and calibrations presented here are performed in the same manner as for the full data set, presented in the Data Analysis section.

Period III also starts at the beginning of the data set, 02.01.1990, but ends at 21.05.2007. The historical forward LIBOR rates for this period can be seen from Figure 2, to the left of the vertical line indicating the end of this period. Descriptive statistics for this period's historical forward LIBOR rates and their log-returns is listed in Table 12.

Table 12: Descriptive statistics of the historical forward LIBOR rates and their log-returns, for Period III.

Maturity	Mean (%)	Variance (%)	Log-returns mean (%)	Log-return variance (%)
1Y	5.4268	0.0314	-0.0126	0.0456
2Y	5.6984	0.0251	-0.0130	0.0288
3Y	5.9709	0.0219	-0.0122	0.0177
4Y	6.3306	0.0224	-0.0123	0.0193
5Y	6.3833	0.0196	-0.0129	0.0152

Bt studying Figure 2, we see that the rate level usually increases for higher maturities. This is also seen From Table 12, where the mean of the rates is found to increase with maturity. The variance, on the other hand, will in most cases decreases with maturity. The same holds for the variance of the log-returns, whilst the mean of the log-returns does not follow the same clear trend, but are all close to zero. This is as expected from the transformation to log-returns.

The covariance of the last 500 trading days is found using EWMA with a smoothing parameter of 0.94, and can be seen in Table 5. We see that the rate of lowest maturity has the highest variance. This coincides with what was observed for the log-returns in Table 12. The variance level, however, is lower in this case. This relative relationship between the variance of the log-returns holds for all maturities.

Table 13: Covariance matrix of the log-returns of the historical forward LIBOR rates from Period III.

	1Y	2Y	3Y	4Y	5Y
1Y	0.00017	0.00009	0.00010	0.00009	0.00007
2Y	0.00009	0.00010	0.00005	0.00005	0.00006
3Y	0.00010	0.00005	0.00009	0.00010	0.00005
4Y	0.00009	0.00005	0.00010	0.00012	0.00005
5Y	0.00007	0.00006	0.00005	0.00005	0.00005

Next, the most likely path is found using the Markov-switch algorithm of section 4.3.2. The transition matrix from the overall most likely state path was estimated to

$$\hat{C} = \begin{bmatrix} 0.989 & 0.011 \\ 0.028 & 0.972 \end{bmatrix}, \tag{A.0.1}$$

implying a stationary distribution of $[0.713, 0.287]^\intercal$. The probability of staying in the same state is high for both states, albeit highest for state one. This causes the stationary distribution to be in favour of being in state one. As this is the more stable state, this is what we expected from the construction of the Markov-switching algorithm.

Jumps are identified, and jump parameters are estimated as explained in section 4.2.2. This is performed both on the full data set in this period, and also the data sets that were split according to the overall most likely path. The estimated parameters from the algorithm is shown in Table 14.

Table 14: Estimated jump mean, variance and intensity for the full data set of Period III, and in the split data sets.

	Full data set			**Data set state 1**			**Data set state 2**		
Maturity	m_J	v_J	λ	m_J	v_J	λ	m_J	v_J	λ
1Y	0.0022	0.0041	0.0568	0.0067	0.0224	0.0023	0.0143	0.0212	0.0728
2Y	0.0050	0.0033	0.0126	0.0001	0.0437	0.0005	0.0204	0.0126	0.0335
3Y	0.0045	0.0024	0.0136	0.0031	0.0260	0.0002	0.0102	0.0105	0.0096
4Y	0.0065	0.0023	0.0089	0.0022	0.0296	0.0001	0.0097	0.0091	0.0092
5Y	0.0168	0.0007	0.0053	0.0035	0.0264	0.0002	0.0058	0.0074	0.0040

From Table 6, we observe that the jump size variance v_J and jump intensities λ for the full data set are decreasing with maturity. This is for the data of state two, and in most cases for the data of state one. More precisely, the jump size variance in the data of state one is higher for the second lowest maturity than for the lowest. However, since the jump intensities of the second lowest maturity, and the remaining maturities, are small compared to first maturity, we model these rates without jumps. That is, we interpret the jump intensity to be zero for all maturities, except for the lowest. Hence, the assumptions on v_J and λ from section 4.2.2 are not violated.

The jump intensities from state one are, for all maturities, lower than for the full data set and state two. Moreover, the jump intensities of the full data set are typically smaller than the for state two.

After removing the jumps, the covariance of the remaining data is found using EWMA. The covariance of the full data set is seen in Table 15. For the split data, however, we do not apply EWMA, but rather find the covariance from all of the data in the respective data set. The resulting covariance matrix for state one and two is presented in Table 16 and 17, respectively.

Table 15: Covariance matrix of the log-returns after removing identified jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00015	0.00008	0.00009	0.00009	0.00006
2Y	0.00008	0.00010	0.00005	0.00004	0.00005
3Y	0.00009	0.00005	0.00008	0.00010	0.00004
4Y	0.00009	0.00004	0.00010	0.00012	0.00004
5Y	0.00006	0.00005	0.00004	0.00004	0.00005

Table 16: Covariance matrix of the log-returns of the data in state one, after removing jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00016	0.00010	0.00010	0.00010	0.00008
2Y	0.00010	0.00015	0.00009	0.00008	0.00008
3Y	0.00010	0.00009	0.00011	0.00012	0.00008
4Y	0.00010	0.00008	0.00012	0.00014	0.00008
5Y	0.00008	0.00008	0.00008	0.00008	0.00011

Table 17: Covariance matrix of the log-returns of the data in state two, after removing jumps

	1Y	2Y	3Y	4Y	5Y
1Y	0.00113	0.00064	0.00047	0.00038	0.00034
2Y	0.00064	0.00059	0.00036	0.00030	0.00026
3Y	0.00047	0.00036	0.00032	0.00031	0.00020
4Y	0.00038	0.00030	0.00031	0.00031	0.00018
5Y	0.00034	0.00026	0.00020	0.00018	0.00023

We notice that the variances found in Table 15, 16 and 17 are considerably lower than the jump size variances for the corresponding data sets. Hence, the identified jumps clearly represent data points where the observed value is not due to ordinary diffusion. Furthermore, the variances of the data in state two are significantly higher than in state one, as expected.

The covariances of the data with and without jumps in the full data set are very similar. To interpret this, we need to keep in mind that they are found using EWMA for the last 500 trading days. This stems from the fact that there are very few jumps occurring during the last 500 days. This covariances would still be very much the same if the window of 500 days was increased, given a constant smoothing parameter of 0.94 in the EWMA. Even if the more distant data contains a lot of jumps, the weight would be so small that the effect almost disappears.

B Verification

B.1 Verification of the jump diffusion implementation

A one-factor case of equation (4.2.5), with Poisson intensity, is used to simulate log-returns of rates, given the following parameter set which was chosen at random.

Table 18: Actual and estimated parameters for the jump calibration routine.

	$\hat{\alpha}$	δ^2	mJ	v_J	λ
Actual	0.000700	0.000004	0.000500	0.000200	0.050000
Estimated	0.000752	0.000004	0.004540	0.000222	0.048395

The number of realizations were 3000 and Δt was set to $1/252$.

Running the jump diffusion algorithm described in section 4.2 for the simulated data will estimate the jump sequence, the corresponding jump sizes, and the parameter values. Figure 10 shows the result, and we clearly see that the algorithm identifies the jumps and estimate the jump sizes correctly.

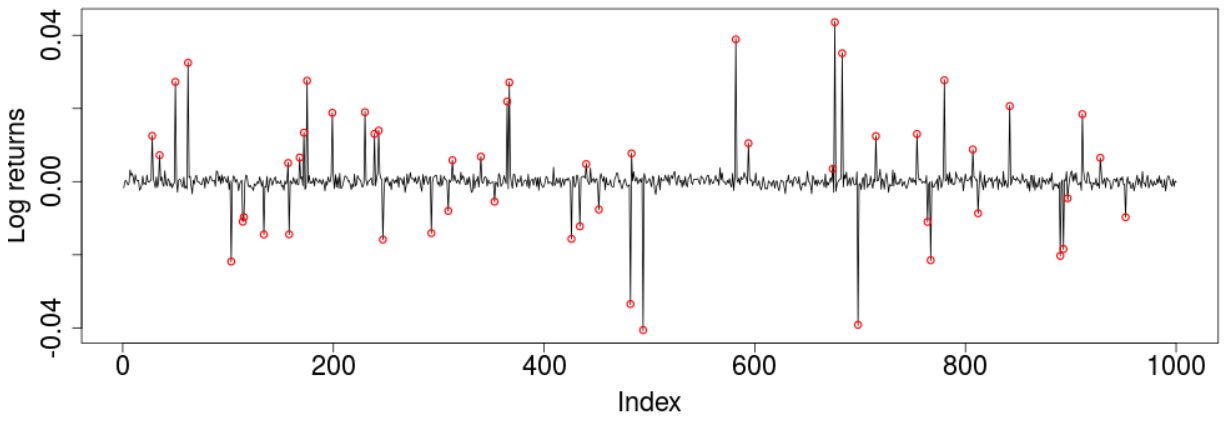


Figure 10: Log-returns for the simulated rates. The red circles indicate jumps.

Comparing the estimated values to the true values, we can conclude that the implementation is correct and the algorithm succeeded.

B.2 Verification of the Markov-switch implementation

This section shows the robustness of the Markov-switching algorithm, as well as it includes a verification of the implementation. The main idea is to select two parameter sets, choose a state path and insert this into the two-state Vasicek model defined in equation (3.2.2). Then the algorithm is applied at the data from the Vasicek model, and tries to reproduce the path chosen initially. If the estimated path and the actual path coincide, the implementation is successful.

A transition matrix and an initial condition is chosen

$$C = \begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix}, \qquad v_0 = 0.2\,\%.$$

Moreover, $\Delta t = 1/252$. The chosen parameter sets for the Vasicek process are listed in Table 19.

Table 19: MS Parameters.

	True parameters			**Simulated parameters**		
State	γ	δ	σ^2	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\sigma}^2$
1	0.0075	0.85	0.0015	0.00762	0.84872	0.00149
2	0.0080	0.91	0.0020	0.00633	0.93433	0.00178

Firstly, a path of state is sampled. We set the initial state to 1, and then sample from C to generate a path of states with length $N = 1260$. Since there are two possible states, the path will consist of 1's and 2's, indicating the state at the time. The path is shown in the middle of Figure 11. Now, this path, along with the initial condition v_0 and the parameter sets are inserted into the two-state discrete Vasicek process given in Equation 4.3.6. The result is shown at the top of Figure 11.

Only the Vasicek data $\boldsymbol{v}$ is inserted into the algorithm in order to estimate the parameters and find the path from which $\boldsymbol{v}$ is produced. The algorithm is run using the following priors

$$\boldsymbol{b}_0 = (0,0), \quad \boldsymbol{B}_0 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad f_0 = 0.5, \quad g_0 = \mathrm{Var}(\boldsymbol{v}), \quad \sigma_1^2 = 0.003, \quad \sigma_2^2 = 0.001, \quad C = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$$

The estimated path from the algorithm is shown in the lowermost plot of Figure 11. The algorithm clearly reproduces the actual path, and with 10000 Monte Carlo iterations, where 1000 are burn-in iterations, the paths differ at only 34 out of 1260. The algorithm also estimates the Vasicek parameters, which are listed in Table 19. By comparing the true and simulated parameters, it is clear that the Markov-switch algorithm is successful. Most importantly, the transition matrix was estimated to $\hat{C} = \begin{pmatrix} 0.9790 & 0.0210 \\ 0.1048 & 0.8952 \end{pmatrix}$, which is close to the transition matrix that was used to produce the data.

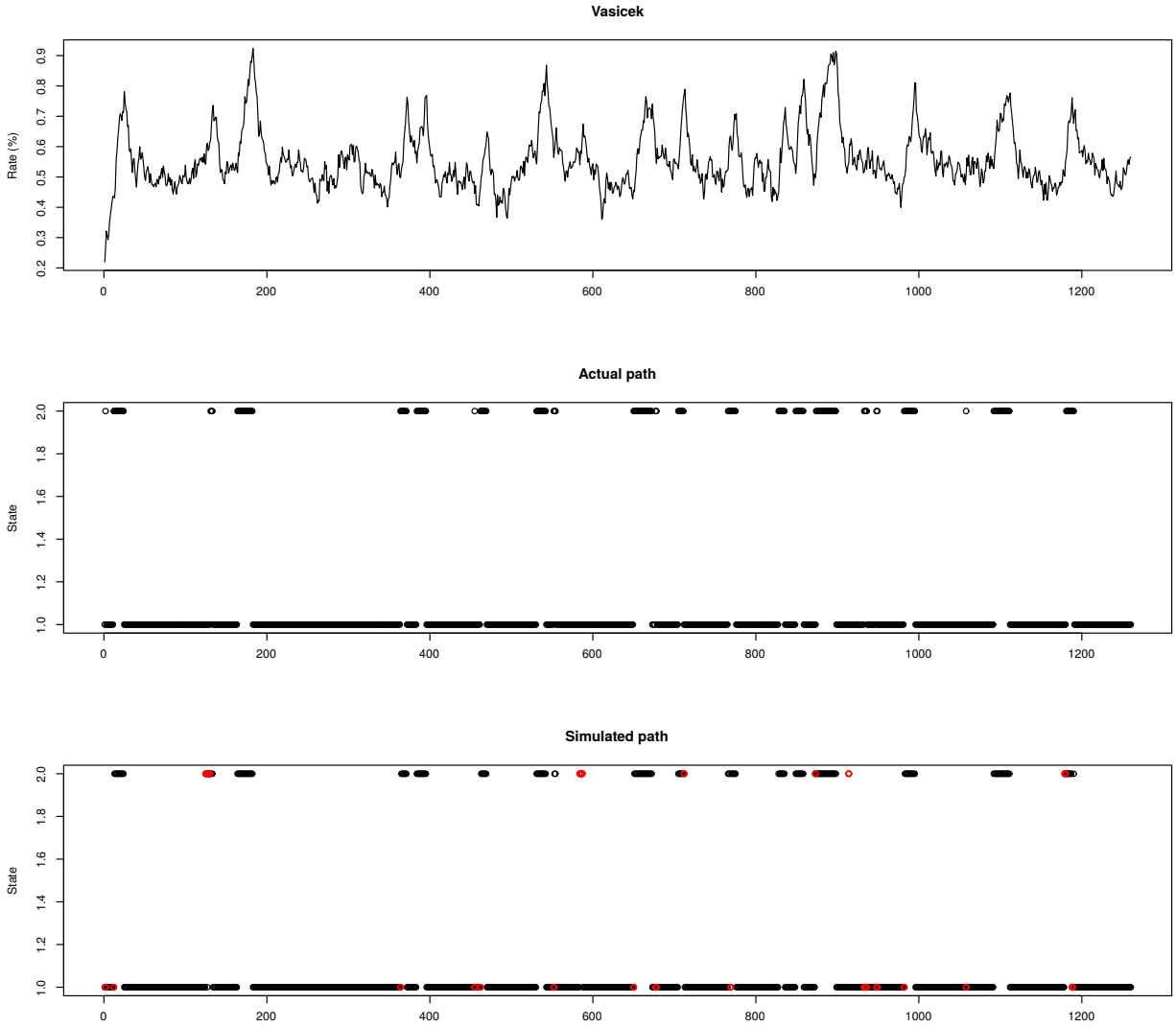


Figure 11: Verification of the implementation of the Markov-switch algorithm. The red circles in the lowermost plot indicates the wrong states.