Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Calibration of Different Interest Rate Models for a Good Fit of Yield Curves

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

H.H.N. AMIN

Delft, the Netherlands
September 2012

MSc THESIS APPLIED MATHEMATICS

"Calibration of Different Interest Rate Models for a Good Fit of Yield Curves"

H.H.N. AMIN

Delft University of Technology

Daily supervisor

Dr. J.A.M. van der Weide

Responsible professor

Prof. dr. F.H.J. Redig

Other thesis committee members

Prof. dr. ir. A.W. Heemink

Dr. R.J. Fokkink

Drs. J. Hommels

September 11, 2012

Delft, the Netherlands

Preface

One of the first mathematical models to describe the interest rate over time was the Vasicek model (1978). Soon after, the Cox Ingersoll Ross (CIR) model (1985) was introduced. The Vasicek model and the CIR model belong to the family of short interest rate models. Through transformation these models can be applied to compute the interest rate values. Classical techniques such as the Maximum Likelihood Estimate (MLE) and Least Squares Method (LSM) are used to estimate the parameters in the short rate model from the historical data.

One must understand that the short rate values cannot be observed from the financial market. However, we can observe the bond prices and from these we can compute the interest rates values. It turns out that when the short interest rate values are not known, MLE and LSM cannot be used to estimate the parameters in the CIR model. Rabobank encountered this problem.

My assignment was to estimate the parameters in the CIR model using the historical data collected by Rabobank. When doing calibration using MLE or LSM for the Vasicek model, it turns out that the drift parameters are estimated with very high bias. Rabobank uses the Long Term Quantile (LTQ) method, which is expected to have no bias. My assignment was to test this claim and see whether there is a bias and if so I had to eliminate this bias.

Acknowledgment

First of all I would like to thank Rabobank and in particular Jasper Hommels for giving me the opportunity to work on this challenging project, for his feedback and the excellent cooperation. Furthermore, I would like to thank my daily supervisor dr. J.A.M. van der Weide for his guidance and advice. Next, I would like to thank prof.dr.ir. A.W. Heemink for his feedback on the Kalman Filter method and dr.ir. R.J. Fokkink for taking over the tasks of my daily supervisor in the final weeks. A special thanks goes to my family and dearest friends who supported me through good and bad times and made it possible for me to be the person I am today.

Hassan Amin

Delft, September 2012

Contents

Introduction

The big financial institutes such as banks and pension funds hold a zero-coupon bond (also called pure discount bond) contract, which guarantees the payment of one unit of currency at maturity[1] of the contract, with no intermediate payments. This means that one party, in case of a government bond this is the government, agrees to pay a fixed interest rate at maturity time to the holder of the contract. In mathematical terms, we define the contract value at time t as $P(t,T)$ where T is maturity, hence the contract payment is $P(T,T) = 1$. The length of the contract can vary from one month up to 30 years, the constant interest rate increases as the length of the contract increases.

Thus, for a contract of 1 year the interest rate may be 1% whereas for the same contract over a period of 10 years one might get a constant interest rate of 3%. One of the reasons behind this is that often it is assumed that the interest rate will rise. Secondly, if the interest rate return were equal for a contract of 1 month and a contract of 30 years then there would be no advantage in entering a contract of a duration of 30 years, for the simple reason that contract over a period of 30 years must have a higher interest rate. When entering a contract there is a fixed interest rate agreed on by the two parties. It is very important to compute the "fairest" interest rate for each contract. If one were able to predict the future the fairest interest rate would simply be the average of the daily measured interest rate up to the maturity. Since this is not the case, mathematical models come into play. Therefore, it is extremely important to find a good model that models the changes in the interest rate over time closely. First we define the (continuously compounded spot) interest rate by $R(t,T)$. The bond price is defined as,

$$P(t,T) = e^{-\tau R(t,T)}, \tag{1}$$

with $\tau = T - t$, see [4]. The compound interest rate $R(t,T)$ is from a family of affine term structure models. Affine term structure models are interest rate models where the compound interest rate $R(t,T)$ is an affine function in the short rate $r(t)$[2]. Thus, we have

$$R(t,T) = \alpha(t,T) + \beta(t,T)r(t), \tag{2}$$

where $\alpha(t,T)$ and $\beta(t,T)$ are deterministic functions of time [4]. Therefore, the zero-coupon bond price can be rewritten as,

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}, \tag{3}$$

[1] Maturity here is defined as the expiry date of the contract

[2] Unless stated otherwise, $\tau = T - t$, T = maturity, t = time, $R(t,T)$ = compound interest rate and $r(t) = r_t$ = short interest rate.

where

$$\alpha(t,T) = \frac{-\ln(A(t,T))}{\tau},$$

$$\beta(t,T) = \frac{B(t,T)}{\tau}.$$

It is impossible to model $R(t,T)$ exactly. Each model, no matter how good, comes with assumptions, no model can exist without assumptions. For this reason a model can never be perfect. Therefore, modeling $R(t,T)$ is like forecasting the weather. The best way to predict the weather for tomorrow is by measuring the weather today and by using mathematical models to predict the movement of the clouds. The assumption made here is that the weather tomorrow depends on the weather today, which is true for most cases but not always. The more assumptions you make the easier the model, hence results into less accurate predictions. Models are the backbone of our society, because of models we know that there are four different seasons and if it is 30 degrees today it is almost 100% certain that it will not snow tomorrow, but it might rain depending on the movement of the clouds. These predictions are not 100% accurate; unfortunately, however, a farmer has to rely on these predictions, since he has no better alternatives. Keep in mind that models improve day by day and who knows what will happen in the future, the prediction will never be 100% accurate but may be very close.

In the world of finance we have the same problem. Mathematical models are used to describe the future by using historical information. Similar to the mathematical models used in the weather forecasting, in the mathematical models used in the financial world certain scenarios can be ruled out with a high certainty. Poor predictions may result in big losses. This is why, we observe an increase in the number of mathematicians working for a financial institute, to improve the models used today to tackle the financial crisis.

In this thesis we will compare three different models; the Vasicek model [21] (1978), the Cox Ingersoll Ross (CIR) model [5] (1985) and the Nelson-Siegel model (2006) [6] (2006). The first two methods are well-known and used for decades in the financial world. However, the Nelson-Siegel model, which originates from [16] (1987) and was reintroduced by Diebold and Li (in 2006), is used by the majority of Central Banks [3]. Each model of these three models comes with the assumption that the historical movements of the interest rate is the best prediction for the future interest rates. The models are described by a fixed number of parameters that depend on the collected historical data. It is extremely important to estimate these parameters without any bias, because a bias in the estimation may result in an huge loss over the short term and the long term, as we will show in Chapter 2. Estimation of these parameters is called *calibration*. The classic methods such as the Maximum Likelihood Estimator (MLE) and the Least Square Method (LSM) are commonly used for the calibration of short rate models. Recently, M. de Ruijter [17] and E. van Elen [9] both applied these methods for calibration of the short rate models in their final theses for Rabobank and Netspar respectively. However, doing estimations using MLE and LSM result in a large bias of up to 400% in one of the estimated parameters. This phenomenal was first addressed by Merton [14] (1980) and later confirmed by many others like Ball and Torous [2] (1996) and Yu and Phillips [23] (2001). Despite this bias these methods are still used today. Yu and Phillips [24] (2006) proposed a jackknife method to reduce the bias by a factor of 5 up to 10, however this

comes with an increase in the variance. Aït-Sahalia [1] (2008) and Tang and Chen [20] (2009) used bootstrap method combined with MLE to reduce the bias by the same factor, which does not increase the variance. E. van Elen did not use any bias reduction techniques. Therefore, the bias remains in the estimated parameters and the consequences of this in computing the bond price is shown in Chapter 2. Rabobank uses the Long Term Quantile (LTQ) method as a better alternative for MLE and LSM to reduce the bias in the estimated parameters. This is why M. de Ruijter applied the LTQ method to reduce the bias, however no tests were performed to measure the bias reduction with respect to LSM and MLE.

In our work we reviewed LTQ and found that we are still have a bias of around 60%, which can have a significant impact on the bond price as we shall see in Chapter 2. Furthermore, we propose the Kalman filter, which is usually applied to estimate parameters for the CIR model [15] in order to reduce the bias in the estimated parameters. The Kalman filter is often used as a last resort when methods such as LSM and MLE fail, such as in the case when estimating parameters in the CIR model [15]. We shall show in this thesis why the Kalman filter should be applied in all cases as this method outperforms LSM, MLE and even LTQ. In the final thesis of E. van Elen the CIR model is estimated using MLE. This can only be done if there is no distinction is made between the $R(t,T)$ and $r(t)$, which is a common pitfall. However, one must understand that the historical data collected from the market is $R(t,T)$ and that $r(t)$ cannot be collected from the market. Therefore, a certain transformation is always needed when dealing with short rate models as we will show in Chapter 1. For this reason LSM and MLE fail in case of calibration of the CIR model.

After the calibration, we compare our results with the initial historical yield curve. The yield curve for day x is defined as the compound interest rate for 9 different contracts, with 9 different maturities (1 month, 3 months, 6 months, 1 year, 2 years, 5 years, 10 years, 20 years, 30 years), at day x. The compound interest rate is computed from the bond prices for each contract, which are observed by Rabobank.

1

Interest Rate Models

In this chapter we will describe the models that we will use. In this thesis we will work with the Vasicek model, the CIR model and the Nelson-Siegel model. In Section 1.1 we derive the partial differential equation for the zero-coupon bond price for short rate models. This is done using the same approach as when deriving the Black-Scholes partial differential equation. The partial differential equation for the zero-coupon bond price allows us to derive the zero-coupon bond price for the Vasicek model and the CIR model, in Section 1.2 and Section 1.3 respectively.

Most of the banks today use the Vasicek model or the extended version of the model called the Hull White model, which was introduced by Hull and White [11] (1987), to model the evolution of short interest rates. There are two common reasons for this. The first and most obvious reason is that these were the first models introduced describing r_t but that alone is not enough. The important reason is because of its simplicity as we will see in Section 1.2. The latter is probably the reason why the model is still used after decades. The major upset when using this model is that one encounters negative interest rate values. This is due the fact that the Vasicek model is normally distributed as we will show in Section 1.2. It is obvious that no one will put money in his/her saving account knowing that it will be worth less next year than it is today, so working with negative interest rate is not realistic.

Therefore, a better alternative is the Cox Ingersoll Ross (CIR) model as this model does not generate negative interest rate values and if the Feller condition holds, see Section 1.3, the interest rate values will be strictly positive. The CIR model is a more realistic model compared to the Vasicek model. It turns out that the CIR model is more complicated. Deriving the bond price for the CIR model is much harder as we will see in Section 1.3. Estimating the parameters for the CIR model cannot be done using classical methods such as MLE and LSM and the reasons for this will be given in Section 1.3.

Even though the Vasicek model, or an extended version of this model, is widely used in the financial world, nine of the thirteen central banks that inform the Bank for International Settlements of their estimations use the model developed by Nelson and Siegel [6] or its extended version, proposed by Svensson to model the yield curve [3]. In Section 1.4 the Nelson-Siegel model is explained, the idea behind this model and why this model is very popular by the Central Banks.

1.1 Deriving Partial Differential Equation for Bond Price

The Vasicek model and the CIR model assume that the interest rate follows a Markov process. We define $W(t)$ as a Wiener process on the risk-neutral probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the short interest rate follows the following stochastic differential equation [18],

$$\mathrm{d}r(t) = \mu(r,t)\mathrm{d}t + \sigma(r,t)\mathrm{d}W(t). \tag{1.1}$$

In this section we relax the notation $r_t = r(t)$ and we will work with $r = r(t)$, because later on we will use partial derivatives and we do not want to cause confusion.

We assume that we have a market which is perfectly liquid. This means that it is possible to purchase or sell a bond, or their fractions in any amount, at any given time. We define the bond price, $P(t,T)$, as a function of t and r, hence $P(t,T) = f(t,r)$. Note that $f(t,r)$ is at least twice continuously differentiable. Therefore, we can apply Itô's lemma to get,

$$\begin{aligned}
\mathrm{d}f(t,r) &= f_t\mathrm{d}t + f_r\mathrm{d}r + \tfrac{1}{2}f_{rr}(\mathrm{d}r)^2 \\
&= f_t\mathrm{d}t + f_r\big(\mu(r,t)\mathrm{d}t + \sigma(r,t)\mathrm{d}W(t)\big) + \frac{1}{2}f_{rr}\big(\mu(r,t)\mathrm{d}t + \sigma(r,t)\mathrm{d}W(t)\big)^2 \\
&= \big(f_t + f_r\mu(r,t) + \tfrac{1}{2}f_{rr}\sigma^2(r,t)\big)\mathrm{d}t + f_r\sigma(r,t)\mathrm{d}W(t),
\end{aligned} \tag{1.2}$$

where f_t is the derivative of $f = f(t,r)$, etcetera.

We want to create a risk-free *self-financing* investment, $\prod(t)$, at time t [19][1]. We define b_1 and b_2 as two arbitrary zero-coupon bonds with different maturities with bond price $f_1(t,r)$ and $f_2(t,r)$ respectively. Furthermore, we define $\triangle_1$ and $\triangle_2$ as the (fraction) amount of bonds for b_1 and b_2 that we have to buy in our risk-free portfolio. The return on our risk-free portfolio is the sum of proportional returns in each of the two underlying bonds. In other words,

$$\frac{\mathrm{d}\prod}{\prod} = \triangle_1\frac{\mathrm{d}f_1(t,r)}{f_1(t,r)} + \triangle_2\frac{\mathrm{d}f_2(t,r)}{f_2(t,r)}. \tag{1.3}$$

We can substitute Equation (1.2) into Equation (1.3) to get,

$$\begin{aligned}
\frac{\mathrm{d}\prod}{\prod} &= \triangle_1\frac{\big(f_{1_t} + f_{1_r}\mu(r,t) + \frac{1}{2}f_{1_{rr}}\sigma^2(r,t)\big)\mathrm{d}t + f_{1_r}\sigma(r_t,t)\mathrm{d}W(t)}{f_1(t,T)} \\
&\quad + \triangle_2\frac{\big(f_{2_t} + f_{2_r}\mu(r,t) + \frac{1}{2}f_{2_{rr}}\sigma^2(r,t)\big)\mathrm{d}t + f_{2_r}\sigma(r,t)\mathrm{d}W(t)}{f_2(t,T)} \\
&= \triangle_1\hat{\mu_1}\mathrm{d}t + \triangle_1\hat{\sigma_1}\mathrm{d}W(t) + \triangle_2\hat{\mu_2}\mathrm{d}t + \triangle_2\hat{\sigma_2}\mathrm{d}W(t) \\
&= \big(\triangle_1\hat{\mu_1} + \triangle_2\hat{\mu_2}\big)\mathrm{d}t + \big(\triangle_1\hat{\sigma_1} + \triangle_2\hat{\sigma_2}\big)\mathrm{d}W(t),
\end{aligned} \tag{1.4}$$

[1] In the book of Seydel the self-pricing portfolio is used to derive the Black-Scholes equation, the steps are almost identical only in our work we are using it to derive the differential equation for the bond price.

with

$$\begin{aligned}
\hat{\mu_1} &= \frac{\big(f_{1_t} + f_{1_r}\mu(r_t,t) + \frac{1}{2}f_{1_{rr}}\sigma^2(r,t)\big)\mathrm{d}t}{f_1(t,T)},\\
\hat{\sigma_1} &= \frac{f_{1_r}\sigma(r,t)\mathrm{d}W(t)}{f_1(t,T)},\\
\hat{\mu_2} &= \frac{\big(f_{2_t} + f_{2_{r_t}}\mu(r,t) + \frac{1}{2}f_{2_{rr}}\sigma^2(r,t)\big)\mathrm{d}t}{f_2(t,T)},\\
\hat{\sigma_2} &= \frac{f_{2_r}\sigma(r,t)\mathrm{d}W(t)}{f_2(t,T)}.
\end{aligned}$$

Since we want a risk-free portfolio we need to eliminate the risk factor, which is the Wiener process. As $\triangle_1$ and $\triangle_2$ are the only two variables which can be changed, they must be chosen in such a way that we can eliminate the Wiener process. In other words,

$$\begin{aligned}
\triangle_1 &+ \triangle_2 &&= 1,\\
\triangle_1\hat{\sigma_1} &+ \triangle_2\hat{\sigma_2} &&= 0.
\end{aligned}$$

Solving this results in

$$\begin{aligned}
\triangle_1 &= \frac{-\hat{\sigma_2}}{\hat{\sigma_1} - \hat{\sigma_2}},\\
\triangle_2 &= \frac{\hat{\sigma_1}}{\hat{\sigma_1} - \hat{\sigma_2}}.
\end{aligned}$$

If we then substitute this into our Equation (1.3) we get

$$\frac{\mathrm{d}\prod}{\prod} = \left(\hat{\mu_1}\frac{-\hat{\sigma_2}}{\hat{\sigma_1} - \hat{\sigma_2}} + \hat{\mu_2}\frac{\hat{\sigma_1}}{\hat{\sigma_1} - \hat{\sigma_2}}\right)\mathrm{d}t. \tag{1.5}$$

Our portfolio, $\prod$, is risk-free, and hence must offer the same return as any other risk-free investment. A safe investment in this case might be a saving account on a bank. Therefore, if we put our money in a savings account the expected return is

$$\frac{\mathrm{d}\prod}{\prod} = r\mathrm{d}t. \tag{1.6}$$

Substituting Equation (1.5) into Equation (1.6) and then solving this will give

$$r(t) = \frac{\hat{\mu_2}\hat{\sigma_1} - \hat{\mu_1}\hat{\sigma_2}}{\hat{\sigma_1} - \hat{\sigma_2}}. \tag{1.7}$$

We can rewrite this as follows,

$$\frac{\hat{\mu_1} - r}{\hat{\sigma_1}} = \frac{\hat{\mu_2} - r}{\hat{\sigma_2}}. \tag{1.8}$$

Let us now define

$$\lambda(t) = \frac{\hat{\mu_t} - r(t)}{\hat{\sigma_t}}. \tag{1.9}$$

This parameter, $\lambda(t)$, is called the *market risk* parameter [5][2]. If we now substitute the values $\hat{\sigma_t}$ and $\hat{\mu_t}$ back into Equation (1.9) we get

$$\lambda(t) = \frac{f_t + f_r\mu(r,t) + \frac{1}{2}f_{rr}\sigma^2(r,t)}{f_r\sigma(r,t)} - \frac{rf(t,r)}{f_r\sigma(r,t)}.$$

This can be rewritten as,

$$f_t(t,r) + (\mu(r,t) - \lambda(t)\sigma(r,t))f_r(t,r) + \tfrac{1}{2}\sigma^2(r,t)f_{rr}(t,r) - rf(t,r) = 0. \tag{1.10}$$

Let us put back $f(t,r) = P(t,T)$, then we get

$$P_t(t,T) + (\mu(r,t) - \lambda(t)\sigma(r,t))P_r(t,r) + \tfrac{1}{2}\sigma^2(r,t)P_{rr}(t,r) - rP(t,T) = 0. \tag{1.11}$$

We want to compute the bond price in the following form,

$$P(t,T) = e^{C(t,T) - B(t,T)r(t)}. \tag{1.12}$$

Note that this is simply Equation (3), see introduction, only now $A(t,T) = \exp(C(t,T))$. Note that the functions $B(t,T)$ and $C(t,T)$ depend on the variable t only (because T is fixed). Therefore, we can define $B(t,T)$ and $C(t,T)$ as functions with one variable τ, hence $B(t,T) = B(\tau)$ and $C(t,T) = C(\tau)$. For the derivatives of $B(t,T)$ and $C(t,T)$ with respect to τ we find

$$B_\tau(\tau) = -B_t(\tau),$$

$$C_\tau(\tau) = -C_t(\tau).$$

Next, we introduce

$$\tilde{\mu} = \mu(r,t) - \lambda_t\sigma(r,t),$$

$$\tilde{\sigma} = \sigma(r,t).$$

[2] A negative value for λ means that the risk premium for holding longer term bonds is positive [15].

Rewrite Equation (1.11) using $\tilde{\mu}$ and $\tilde{\sigma}$ yields

$$P_t + \tilde{\mu}P_r + \tfrac{1}{2}\tilde{\sigma}^2 P_{rr} - rP = 0. \tag{1.13}$$

Then we compute P_t, P_r and P_{rr}. This gives us

$$\begin{aligned} P_\tau(\tau) &= \big(-C_\tau(\tau) + B_\tau(\tau)r\big)P(\tau), \\ P_r(\tau) &= -B(\tau)P(\tau), \\ P_{rr}(\tau) &= B^2(\tau)P(\tau). \end{aligned}$$

Substituting this in Equation (1.13) we get

$$-C_\tau(\tau) - \tilde{\mu}B(\tau) + \tfrac{1}{2}\tilde{\sigma}^2 B^2(\tau) - \big(1 - B_\tau(\tau)\big)r = 0. \tag{1.14}$$

Note that if there exists a solution then this solution is unique [10]. In Section 1.2 and Section 1.3 we will solve Equation (1.14) for the Vasicek model and the CIR model.

1.2 The Vasicek model

The stochastic differential equation of the Vasicek model is,

$$\mathrm{d}r_t = \kappa(\mu - r_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{1.15}$$

where κ, μ and σ are constants and W_t is a Wiener process[21][3].

It can easily be shown using Itô's lemma [18], with $Y_t = e^{\kappa t} r_t$, that the exact discrete model corresponding to (1.15) is,

$$r_{t+\triangle t} = e^{-\kappa \triangle t} r_t + \mu(1 - e^{-\kappa \triangle t}) + \sigma e^{-\kappa(t+\triangle t)} \int_t^{t+\triangle t} e^{-\kappa t} \mathrm{d}W_t. \tag{1.16}$$

This means that r_t is normally distributed with Equation (1.17) as its mean and Equation (1.18) as its variance,

$$\mathbb{E}\big[r_{t+\triangle t} | r_t\big] = e^{-\kappa \triangle t} r_t + \mu(1 - e^{-\kappa \triangle t}), \tag{1.17}$$

$$\mathrm{Var}\big[r_{t+\triangle t} | r_t\big] = \left(\frac{\sigma^2}{2\kappa}\right)(1 - e^{-2\kappa \triangle t}). \tag{1.18}$$

If κ goes to zero then the expectation goes to r_t and the variance to zero. However, if κ goes to infinity we see that the expectation goes to μ and the variance to $\frac{\sigma^2}{2\kappa}$.

We want to compute the zero-coupon bond price for the Vasicek model. This means we have to solve Equation (1.14) with

[3] We assume that the risk parameter $\lambda = 0$.

$$\begin{aligned}\tilde{\mu} &= \kappa(\mu - r),\\ \tilde{\sigma} &= \sigma.\end{aligned}$$

We get

$$-\frac{\partial C(\tau)}{\partial \tau} - \kappa(\mu - r)B(\tau) + \tfrac{1}{2}\sigma^2 B^2(\tau) - \left(1 - \frac{\partial B(\tau)}{\partial \tau}\right)r = 0. \tag{1.19}$$

We can rewrite this as

$$-\frac{\partial C(\tau)}{\partial \tau} - \kappa\mu B(\tau) + \tfrac{1}{2}\sigma^2 B^2(\tau) - \left(1 - \kappa r B(\tau) - \frac{\partial B(\tau)}{\partial \tau}\right)r = 0. \tag{1.20}$$

Because Equation (1.20) holds for every r and every τ we must have,

$$-\frac{\partial C(\tau)}{\partial \tau} - \kappa\mu B(\tau) + \tfrac{1}{2}\sigma^2 B^2(\tau) = 0, \tag{1.21}$$

and

$$\kappa B(\tau) + \frac{\partial B(\tau)}{\partial \tau} = 1. \tag{1.22}$$

Furthermore, we have that

$$\begin{aligned}P(T,T) &= P(\tau = 0)\\ &= e^{C(0) - B(0)r(T)},\end{aligned}$$

which implies that $C(0) = B(0) = 0$. We first solve the following Equation (1.22) with boundary condition $B(0) = 0$. It is easy to see that $\frac{\partial B(\tau)}{\partial \tau} = e^{-\kappa\tau}$, solving this with $B(0) = 0$ we get

$$B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}. \tag{1.23}$$

Now substitute Equation (1.23) into Equation (1.21) and then integrate with respect to t, over an interval $[t, T]$. This gives

$$
\begin{aligned}
C(\tau) &= \int_t^T \kappa\mu B(T-s) + \tfrac{1}{2}\sigma^2 B^2(T-s)\mathrm{d}s \\
&= \int_t^T \kappa\mu\frac{1-e^{-\kappa(T-s)}}{\kappa}ds + \int_t^T \tfrac{1}{2}\sigma^2\left(\frac{1-e^{-\kappa(T-s)}}{\kappa}\right)^2\mathrm{d}s \\
&= \int_t^T \kappa\mu\frac{1-e^{-\kappa(T-s)}}{\kappa}ds + \int_t^T \frac{\sigma^2}{2\kappa}\big(1-2e^{-\kappa(T-s)}+e^{-2\kappa(T-s)}\big)\mathrm{d}s \\
&= -\mu(T-t) + \mu\frac{1-e^{-\kappa(T-t)}}{\kappa} + \frac{\sigma^2}{2\kappa}(T-t) - \frac{\sigma^2}{2\kappa}\frac{1-e^{-\kappa(T-t)}}{\kappa} + \frac{\sigma^2}{2\kappa}\frac{1-e^{-2\kappa(T-t)}}{2\kappa} \\
&= -\mu(\tau) + \mu B(\tau) + \frac{\sigma^2}{2\kappa}(\tau) - \frac{\sigma^2}{2\kappa}B(\tau) + \frac{\sigma^2}{2\kappa}\left(\frac{1-e^{-2\kappa(T-s)}}{2\kappa}\right) \\
&= \frac{\big(B(\tau)-\tau\big)(\kappa^2\mu-\frac{\sigma}{2})}{\kappa^2} - \frac{\sigma^2}{4\kappa}B^2(\tau).
\end{aligned}
$$

Therefore, we get for the bond price

$$
\begin{aligned}
P(t,\tau) &= A(\tau)e^{-B(\tau)r(t)} \\
&= e^{-\tau R(t,\tau)},
\end{aligned}
\tag{1.24}
$$

with

$$
\begin{aligned}
A(\tau) &= \exp\left\{\left(\mu-\tfrac{\sigma^2}{2\kappa^2}\right)(B(\tau)-\tau) - \tfrac{\sigma^2}{4\kappa}B^2(\tau)\right\}, \\
B(\tau) &= \frac{1-e^{-\kappa\tau}}{\kappa}.
\end{aligned}
$$

Recall, the values were determined under the risk neutral measure. We get the following stochastic differential equation for $R(t,\tau)$,

$$
\mathrm{d}R_{(t,\tau)} = \kappa\left(\frac{B_\tau\mu}{\tau} - \frac{\ln(A_\tau)}{\tau} - R_{(t,\tau)}\right)\mathrm{d}t + \frac{B_\tau}{\tau}\sigma\mathrm{d}W_t, \tag{1.25}
$$

which can be rewritten as,

$$
\mathrm{d}R_{(t,\tau)} = \widehat{\kappa}(\widehat{\mu} - R_{(t,\tau)})\mathrm{d}t + \widehat{\sigma}\mathrm{d}W_t, \tag{1.26}
$$

with

$$\begin{aligned}\widehat{\kappa} &= \kappa,\\ \widehat{\mu} &= \frac{B_\tau \mu}{\tau} - \frac{\ln(A_\tau)}{\tau},\\ \widehat{\sigma} &= \frac{B_\tau}{\tau}\sigma.\end{aligned}$$

Note that both r_t and $R_{(t,\tau)}$ are normally distributed in the Vasicek model. We can calibrate both stochastic differential equations using the same distribution function but with different parameters. We collected $R_{(t,\tau)}$ from the historical data. After estimating $\widehat{\kappa}$, $\widehat{\mu}$ and $\widehat{\sigma}$ using the historical data, we can find the parameters κ, μ and σ for the short rate model. Suppose through calibration we estimated $\widehat{\kappa}$, $\widehat{\mu}$ and $\widehat{\sigma}$, then we get

$$\begin{aligned}\kappa &= \widehat{\kappa},\\ B(\tau) &= \frac{1-e^{-\kappa\tau}}{\kappa},\\ \sigma &= \frac{\tau}{B(\tau)}\widehat{\sigma},\\ A(\tau) &= \exp\left\{\tau\big(\tau - B_\tau\big)\left(\frac{B_\tau}{\tau}\left[\frac{\sigma^2 B_\tau^2}{4k\big(B_\tau-\tau\big)} + \frac{\sigma^2}{2k^2}\right] - \hat{\mu}\right)\right\},\\ \mu &= \big(1-B_\tau\big)\left(\frac{\sigma^2 B_\tau^2}{4k\big(B_\tau-\tau\big)} + \frac{\sigma^2}{2k^2}\right) - \hat{\mu}.\end{aligned}$$

1.3 The Cox Ingersoll Ross model

Now we consider the CIR model. The stochastic differential equation of the CIR model is

$$\mathrm{d}r_t = \kappa(\mu - r_t)\mathrm{d}t + \sigma\sqrt{r_t}\mathrm{d}W_t, \tag{1.27}$$

where κ, μ and σ are constants and W_t is a Wiener process [5][4].

Using Itô's lemma again, as for the Vasicek model, we can show the exact discrete model corresponding to (1.27) is,

$$r_{t+\triangle t} = e^{-\kappa\triangle t} r_t + \mu(1-e^{-\kappa\triangle t}) + \sigma e^{-\kappa(t+\triangle t)} \int_t^{t+\triangle t} e^{-\kappa t}\sqrt{r_t}\mathrm{d}W_t. \tag{1.28}$$

Now we can easily compute the expectation and the variance of the process. This gives us,

[4]We assume that the risk parameter $\lambda = 0$.

$$\mathbb{E}\big[r_{t+\triangle t}|r_t\big] = e^{-\kappa\triangle t}r_t + \mu(1-e^{-\kappa\triangle t}), \tag{1.29}$$

$$\mathrm{Var}\big[r_{t+\triangle t}|r_t\big] = e^{-\kappa\triangle t}r_t\left(\frac{\sigma^2}{\kappa}\right)(1-e^{-\kappa\triangle t}) + \mu\left(\frac{\sigma^2}{2\kappa}\right)(1-e^{-\kappa\triangle t})^2. \tag{1.30}$$

If κ goes to zero then the expectation goes to r_t and the variance to zero. However, if κ goes to infinity we see that the expectation goes to μ and the variance to $\mu\frac{\sigma^2}{2\kappa}$. It can be shown that the process r_t features a noncentral chi-squared distribution function [5].

We want to compute what the zero-coupon bond price for the CIR model is. This means we have to solve Equation (1.14) with

$$\begin{aligned}\tilde{\mu} &= \kappa(\mu - r),\\ \tilde{\sigma} &= \sigma\sqrt{(}r).\end{aligned}$$

Therefore, we get

$$-\frac{\partial C(\tau)}{\partial\tau} - \kappa(\mu-r)B(\tau) + \tfrac{1}{2}r\sigma^2B^2(\tau) - \left(1-\frac{\partial B(\tau)}{\partial\tau}\right)r = 0. \tag{1.31}$$

We can rewrite this as,

$$-\frac{\partial C(\tau)}{\partial\tau} - \kappa\mu B(\tau) - \left(1-\tfrac{1}{2}\sigma^2B^2(\tau) - \kappa rB(\tau) - \frac{\partial B(\tau)}{\partial\tau}\right)r = 0. \tag{1.32}$$

Using the same argument as for the Vasicek model we get

$$-\frac{\partial C(\tau)}{\partial\tau} - \kappa\mu B(\tau) = 0, \tag{1.33}$$

and

$$\kappa B(\tau) + \frac{\partial B(\tau)}{\partial\tau} + \tfrac{1}{2}\sigma^2B^2(\tau) = 1, \tag{1.34}$$

with $B(0) = 0$ and $C(0) = 0$. We easily solved the partial differential equation for $B(\tau)$ in the Vasicek model, see Section 1.2. For the CIR model this is slightly harder. A point of notice, we decided to derive the solution for $C(\tau)$, but not for the $B(\tau)$. The reasons for this are: deriving the solution for $C(\tau)$ is harder and a very similar derivation for $B(\tau)$ is provided in [10]. We know that there exists a solution, see [5]. The solution is

$$B(\tau) = \frac{2(e^{\lambda\tau}-1)}{2\lambda + (\lambda+\kappa)(e^{\tau\lambda}-1)}, \tag{1.35}$$

with[5]

$$\lambda = \sqrt{\kappa^2+2\sigma^2}.$$

This gives us,

[5]Note in this case λ is not defined as the risk parameter.

$$
\begin{aligned}
B(\tau) &= \frac{2(e^{\lambda\tau}-1)}{2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)},\\
B^2(\tau) &= \frac{4(e^{\lambda\tau}-1)^2}{\big(2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)\big)^2},\\
B_\tau(\tau) &= \frac{4\lambda^2 e^{\lambda\tau}}{\big(2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)\big)^2}.
\end{aligned}
$$

As expected this solution indeed solves Equation (1.34), and since there exists only one [10], the proposed solution is unique. Now substitute Equation (1.35) into Equation (1.33) and then integrate with respect to t, over an interval $[t,T]$. Therefore, we get

$$
\begin{aligned}
C(\tau) &= \int_t^T -\kappa\mu B(T-s)\mathrm{d}s\\
&= \int_t^T -\kappa\mu\bigg(\frac{2(e^{\lambda(T-s)}-1)}{2\lambda+(\lambda+\kappa)(e^{(T-s)\lambda}-1)}\bigg)\mathrm{d}s.
\end{aligned}
$$

To solve this integral, we start with first using the following transformation

$$x(s) = e^{\lambda(T-s)},$$

with

$$\mathrm{d}s = -\frac{1}{\lambda x(s)}\mathrm{d}x.$$

This gives us,

$$
\begin{aligned}
C(\tau) &= \int_t^T -2\kappa\mu\bigg(\frac{(e^{\lambda(T-s)}-1)}{2\lambda+(\lambda+\kappa)(e^{(T-s)\lambda}-1)}\bigg)\mathrm{d}s\\
&= \int_{e^{\lambda(T-t)}}^1 -2\kappa\mu\bigg(\frac{-(x-1)}{\lambda x\big(2\lambda+(\lambda+\kappa)(x-1)\big)}\bigg)\mathrm{d}x\\
&= \int_{e^{\lambda(T-t)}}^1 \frac{2\kappa\mu}{\lambda}\bigg(\frac{1}{2\lambda+(\lambda+\kappa)(x-1)}-\frac{1}{x\big(2\lambda+(\lambda+\kappa)(x-1)\big)}\bigg)\mathrm{d}x\\
&= \int_{e^{\lambda(T-t)}}^1 \frac{2\kappa\mu}{\lambda}\bigg(\frac{1}{2\lambda+(\lambda+\kappa)(x-1)}-\frac{(\lambda+\kappa)}{(\kappa-\lambda)\big(2\lambda+(\lambda+\kappa)(x-1)\big)}-\frac{1}{x(\lambda-\kappa)}\bigg)\mathrm{d}x.
\end{aligned}
$$

Next, we apply the transformation

$$y(x) = 2\lambda + (\lambda + \kappa)(x - 1),$$

with

$$\mathrm{d}x = \frac{1}{(\lambda + \kappa)}\mathrm{d}y.$$

This leads to,

$$\begin{aligned}
C(\tau) &= \int_{2\lambda+(\lambda+\kappa)(e^{\lambda(T-t)}-1)}^{2\lambda-(\lambda+\kappa)} \frac{2\kappa\mu}{\lambda}\left(\frac{1}{(\lambda+\kappa)y} - \frac{1}{(\kappa-\lambda)y}\right)\mathrm{d}y - \frac{2\kappa\mu}{\lambda}\int_{e^{\lambda(T-t)}}^{1} \frac{1}{(\lambda-\kappa)x}\mathrm{d}x \\
&= \frac{2\kappa\mu}{\lambda}\left(\left[\tfrac{1}{(\lambda+\kappa)}\ln(y) - \tfrac{1}{(\kappa-\lambda)}\ln(y)\right]_{2\lambda+(\lambda+\kappa)(e^{\lambda(T-t)}-1)}^{2\lambda-(\lambda+\kappa)}\right) - \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda-\kappa)}\ln(x)\right]_{e^{\lambda(T-t)}}^{1} \\
&= \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda+\kappa)}\ln(2\lambda+(\lambda+\kappa)(x-1)) - \tfrac{1}{(\kappa-\lambda)}\ln(2\lambda+(\lambda+\kappa)(x-1))\right]_{e^{\lambda(T-t)}}^{1} \\
&\quad - \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda-\kappa)}\ln(x)\right]_{e^{\lambda(T-t)}}^{1} \\
&= \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda+\kappa)}\ln(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1)) - \tfrac{1}{(\kappa-\lambda)}\ln(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1))\right]_{t}^{T} \\
&\quad - \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda-\kappa)}\ln(e^{\lambda(T-s)})\right]_{t}^{T}.
\end{aligned}$$

Instead of just substituting the boundaries into the integral and solving it, we will first simplify the equation so that eventually we can have a nice simple formula that we can work with. Simplifying the above equation gives

$$
\begin{aligned}
C(\tau) &= \frac{2\kappa\mu}{\lambda}\left[\ln\left(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1)\right)^{-\frac{2\lambda}{(\kappa^2-\lambda^2)}}\right]_t^T - \frac{2\kappa\mu}{\lambda}\left[\tfrac{1}{(\lambda-\kappa)}\ln\left(e^{\lambda(T-s)}\right)\right]_t^T \\
&= \frac{2\kappa\mu}{\lambda}\left[\ln\left(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1)\right)^{-\frac{2\lambda}{(\kappa^2-\lambda^2)}}\right]_t^T - \frac{2\kappa\mu}{\lambda}\left[\ln\left(e^{\frac{(T-s)(\kappa+\lambda)}{2}}\right)^{-\frac{2\lambda}{(\kappa^2-\lambda^2)}}\right]_t^T \\
&= \frac{-4\kappa\mu}{\kappa^2-\lambda^2}\left[\ln\left(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1)\right)-\ln\left(e^{\frac{(T-s)(\kappa+\lambda)}{2}}\right)\right]_t^T \\
&= \frac{2\kappa\mu}{\sigma^2}\left[\ln\left(2\lambda+(\lambda+\kappa)(e^{\lambda(T-s)}-1)\right)-\ln\left(e^{\frac{(T-s)(\kappa+\lambda)}{2}}\right)\right]_t^T \\
&= \frac{2\kappa\mu}{\sigma^2}\left[\ln(2\lambda)-\ln\left(2\lambda+(\lambda+\kappa)(e^{\lambda\tau}-1)\right)+\ln\left(e^{\frac{\tau(\kappa+\lambda)}{2}}\right)\right] \\
&= \frac{2\kappa\mu}{\sigma^2}\ln\left(\frac{2\lambda e^{(\lambda+\kappa)\frac{\tau}{2}}}{2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)}\right).
\end{aligned}
$$

Now we have a formula for the zero-coupon bond price for the CIR model which is given as

$$
\begin{aligned}
P(t,\tau) &= A(\tau)e^{-B(\tau)r(t)} \\
&= e^{-\tau R(t,\tau)}.
\end{aligned} \tag{1.36}
$$

with

$$
\begin{aligned}
A(\tau) &= \left(\frac{2\lambda e^{(\lambda+\kappa)\frac{\tau}{2}}}{2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)}\right)^{\frac{2\kappa\mu}{\sigma^2}}, \\
B(\tau) &= \frac{2(e^{\lambda\tau}-1)}{2\lambda+(\lambda+\kappa)(e^{\tau\lambda}-1)}, \\
\lambda &= \sqrt{\kappa^2+2\sigma^2}.
\end{aligned}
$$

Now we shall repeat the same steps as for the Vasicek model. Under the risk neutral measure we get the following stochastic differential equation for $R_{(t,\tau)}$,

$$
\mathrm{d}R_{(t,\tau)} = \frac{B_\tau}{\tau}\left[\kappa\left(\mu-\frac{\tau R_{(t,\tau)}+\ln(A_\tau)}{B_\tau}\right)\mathrm{d}t+\sigma\sqrt{\frac{\tau R_{(t,\tau)}+\ln(A_\tau)}{B_\tau}}\mathrm{d}W_t\right]. \tag{1.37}
$$

We cannot rewrite $R_{(t,\tau)}$ in the following form;

$$\mathrm{d}R_{(t,\tau)} = \widehat{\kappa}(\widehat{\mu} - R_{(t,\tau)})\mathrm{d}t + \widehat{\sigma}\sqrt{R_{(t,\tau)}}\mathrm{d}W_t, \tag{1.38}$$

and from this we can conclude that r_t and $R_{(t,\tau)}$ do not follow the same distribution, as it was the case in the Vasicek model, this means classical methods such as MLE and LSM fail. Therefore, we must apply the Kalman filter algorithm which has been used in the paper of Duan and Simonata [7] (1995) dealing with the estimation of affine term structure models.

1.4 The Nelson-Siegel model

Diebold and Li (2006) reintroduced the Nelson and Siegel model, which originates from (1987). They provided evidence, in their paper [6], that the model can also be a valuable tool for forecasting the yield curve. Even though the yield curve is almost perfectly fitted (unlike the case with the Vasicek model and the CIR model, as we will see in Chapter 3) the Nelson-Siegel model does allow arbitrage. Despite this huge disadvantage, the Bank of International Settlements (BIS, 2005) reports that currently nine out of thirteen Central Banks which report their curve estimation methods to the BIS use the Nelson-Siegel to construct zero-coupon yield curves. For this reason, we describe this model briefly. Let us define $r(t,\tau) = r_t(\tau)$ forward rate curve at time t over a time interval τ. Then in the Nelson-Siegel model $r_t(\tau)$ is given by the following equation,

$$r_t(\tau) = \beta_{1,t} + \beta_{2,t}e^{-\lambda_t\tau} + \beta_{3,t}\lambda_t\tau e^{-\lambda_t\tau}. \tag{1.39}$$

The compound interest rate, $R(t,\tau) = R_t(\tau)$, is the average of the forward rate curve,

$$R_t(\tau) = \frac{1}{\tau}\int_0^\tau r_t(s)\mathrm{d}s. \tag{1.40}$$

Substituting Equation (1.39) into (1.40) we get

$$R_t(\tau) = \beta_{1,t}\kappa_1 + \beta_{2,t}\kappa_2 + \beta_{3,t}\kappa_3, \tag{1.41}$$

with

$$\begin{aligned}
\kappa_1 &= 1, \\
\kappa_2 &= \frac{1 - e^{-\lambda_t\tau}}{\lambda_t\tau}, \\
\kappa_3 &= \frac{1 - e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}.
\end{aligned}$$

There are several reasons why this model is widely used. The most important reason is that it provides a very good approximation of the yield curve by using a small number of parameters. Together, the three parameters $\kappa_1, \kappa_2, \kappa_3$, give the model enough flexibility to fit all different types shapes. We want to explain the parameters $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$. Diebold and

Li (2006) define the parameters $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ as long-term, short-term and medium-term components respectively. The following limits are well-defined,

$$\lim_{\tau \to 0} R_t(\tau) = \beta_{1,t} + \beta_{2,t}, \tag{1.42}$$

and

$$\lim_{\tau \to \infty} R_t(\tau) = \beta_{1,t}. \tag{1.43}$$

The first compound, $\beta_{1,t}$, is referred to as the long-term compound, because the loading on $\beta_{1,t}$ is κ_1 which is equal to 1 and hence does not decay to zero in the limit as $t \to \infty$. The long-term compound at time t is chosen in such a way that it is equal to the interest rate of the contract with the longest maturity [6] at time t. For our specific case we have,

$$R_t(\tau = 30 \text{ years}) \approx \beta_{1,t}. \tag{1.44}$$

The second compound, $\beta_{2,t}$, is referred to as the short-term compound, because the loading on $\beta_{2,t}$ is κ_2 which starts at 1 but decays monotonically and quickly to zero as $t \to \infty$. Diebold and Li define the yield curve slope as Equation (1.42) minus Equation (1.43), which gives $-\beta_{2,t}$. For our specific case we have,

$$R_t(\tau = 30 \text{ years}) - R_t(\tau = 1 \text{ month}) \approx -\beta_{2,t}. \tag{1.45}$$

The final compound, $\beta_{3,t}$, is referred to as the medium-term compound, because the loading on $\beta_{3,t}$ is κ_3 which starts at 0, increases to its maximum, which is determined by the decay parameter λ_t, before returning zero as $t \to \infty$. The medium-term compound is closely related to the yield curve curvature, which is defined as twice the two-year yield minus the sum of the limits in Equation (1.42) and Equation (1.43) [6]. For our specific case we have,

$$2R_t(\tau = 30 \text{ years}) - \big(R_t(\tau = 30 \text{ years}) - R_t(\tau = 1 \text{ month})\big) \approx \beta_{2,t}. \tag{1.46}$$

Thus, if we define l_t, s_t and c_t as the yield curve level, yield curve slope and yield curve curvature respectively, we must have,

$$\operatorname{corr}\big[\beta_{1,t}, l_t\big] \approx 1, \tag{1.47}$$

$$\operatorname{corr}\big[\beta_{2,t}, s_t\big] \approx -1, \tag{1.48}$$

$$\operatorname{corr}\big[\beta_{3,t}, c_t\big] \approx 1. \tag{1.49}$$

We may choose to estimate the parameters $\theta = \{\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \lambda_t\}$ by nonlinear least squares for each time step. However Diebold and Li, instead fix λ_t at a prespecified value, which allows them to compute the values of $\beta_{1,t}, \beta_{2,t}, \beta_{3,t}$ by using LSM for each step t. Therefore, we will also fix λ_t before using LSM to estimate the parameters $\beta_{1,t}, \beta_{2,t}$, and $\beta_{3,t}$. We must choose λ_t in such a way that the error between the actual yield curve and the estimated yield curve, using the Nelson-Siegel model, is minimized, for details see Chapter 3.

2

Calibration of Interest Rate Models

2.1 Calibration with MLE and LSM

We are given the historical data and we want to find the parameters κ, μ and σ for the Vasicek model. To do so, we will use two methods LSM and MLE. We start with LSM. The Euler discretization of the Vasicek model is given by,

$$R_{(t+\triangle t,\tau)} = R_{(t,\tau)} + \kappa(\mu - R_{(t,\tau)})\triangle t + \sigma\big(W_{t+\triangle t} - W_t\big). \tag{2.1}$$

By rewriting Equation (2.1) we can use LSM to estimate the drift parameters by solving the following equation,

$$\big(\hat{\kappa}, \hat{\mu}\big) = \arg\min_{\kappa,\mu} \sum_{i=1}^{N-1} \left(R_{(i+1,\tau)} - R_{(i,\tau)} - \kappa\mu\triangle t + \kappa R_{(i,\tau)}\triangle t \right)^2. \tag{2.2}$$

Once we have solved (2.2) (in our case with MATLAB), we then compute the standard deviation of residuals and use that as estimator for the parameter σ which we define as $(\hat{\sigma})$. Thus, we have an estimator for (κ, μ, σ) which is, $(\hat{\kappa}, \hat{\mu}, \hat{\sigma})$. We can do the transformations in Section 1.2 to get the parameters for r_t. Note that for LSM we do not use the conditional density function or the density function. Any discretization suffices, in our case we used the Euler discretization to estimate the parameters.

Next, we will use MLE to estimate the parameters (κ, μ, σ). Since R_t follows a Markov process it follows that $(R_{(t+\triangle t,\tau)}|R_{(t,\tau)})$ is independent and identically distributed with conditional density function and density function [18]. Therefore, MLE is given by

$$\begin{aligned} \mathcal{L}(\theta; R_{(2,\tau)}|R_{(1,\tau)}, \ldots, R_{(N,\tau)}|R_{(N-1,\tau)}) &= f(R_{(2,\tau)}|R_{(1,\tau)}, \ldots, R_{(N,\tau)}|R_{(N-1,\tau)}; \theta) \\ &= \textstyle\prod_{i=1}^{N-1} f(R_{(i+1,\tau)}|R_{(i,\tau)}; \theta), \end{aligned} \tag{2.3}$$

with $\theta = (\kappa, \mu, \sigma)$. In our case it is more convenient to work with the log-likelihood, which is the logarithm of the likelihood function $\mathcal{L}$. Thus, we get

$$\begin{aligned}\ell(\theta; R_{(2,\tau)}|R_{(1,\tau)}, \ldots, R_{(N,\tau)}|R_{(N-1,\tau)}) &= \ln\left(\mathcal{L}(\theta; R_{(2,\tau)}|R_{(1,\tau)}, \ldots, R_{(N,\tau)}|R_{(N-1,\tau)})\right) \\ &= \textstyle\sum_{i=1}^{N-1} \ln\left(f(\theta; R_{(2,\tau)}|R_{(1,\tau)}, \ldots, R_{(N,\tau)}|R_{(N-1,\tau)})\right).\end{aligned} \tag{2.4}$$

Many articles, such as Durham and Gallant (2002), have approximated the likelihood function of a continuous system analytically and computed it by numerical methods. Despite its good asymptotic properties, MLE can produce a huge bias in the estimated parameters in the short rate models. It is well known that the bias is mainly found in the drift parameters see for instance Merton [14], Ball and Torous [2] and Yu and Phillips [23]. Merton first discovered the bias when estimating parameters with the Black-Scholes model. The difficulty in the drift parameter comes from the estimation approaches, including MLE. The problem increases when the process has a lack of dynamics, which happens when κ is small. As mentioned by Yu and Phillips [24] and in our test results, see Section 2.4, MLE and LSM for κ can produce a bias of more than 400%. The prices of bond options and other derivative securities depend mainly on the value of unknown parameters. Yu and Phillips [24] introduced the jackknife method to reduce the bias in parameter estimation, however this comes with an increase in the variance. Aït-Sahalia [1] and Tang and Chen [20] used bootstrapping combined with MLE to reduce the bias in the parameters, this reduces the bias without increasing the variance.

2.2 Calibration with Kalman Filter

In 1960, Rudolph E. Kalman published his famous paper [12] introducing a powerful linear filtering technique named after him. The Kalman filter is a method that provides an efficient recursive algorithm to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful because it supports estimations of past, present and future states, this can be done even when exact parameters of the model are not known.

A good introduction of the Kalman filter is given in [22]. We will try to follow this paper briefly but give a more detailed introduction of the Kalman filter. The Kalman filter addresses the general problem of trying to estimate the state transition space $x \in \mathbb{R}^n$ of a discrete time controlled process that is governed by the linear stochastic differential equation,

$$x_t = Ax_{t-1} + Bu_{t-1} + w_t, \tag{2.5}$$

with a measurement space $z \in \mathbb{R}^m$ that is observed. Thus,

$$z_t = Hx_t + v_t. \tag{2.6}$$

The variable $u_{t-1} \in \mathbb{R}^n$ is the optional control input, we assume that u_{t-1} is equal to the identity matrix. The random variables w_t and v_t represent the process and measurement noise respectively. They are assumed to be independent and following a normal distribution. Thus, $w_t \sim \mathcal{N}(0, Q_t)$ and $v_t \sim \mathcal{N}(0, R_t)$, where

$$Q_t = \begin{pmatrix} q_{t_1}^2 & 0 & \cdots & 0 \\ 0 & q_{t_1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{t_1}^2 \end{pmatrix},$$

and

$$R_t = \begin{pmatrix} \rho_{t_1}^2 & 0 & \cdots & 0 \\ 0 & \rho_{t_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_{t_3}^2 \end{pmatrix}.$$

The covariance matrices Q_t and R_t are time dependent, therefore, they may change with each time step or measurement. However, we will take the covariance matrix R_t constant. Minimizing this will result in a more reliable estimated parameters as we will show later in this section. In Section 2.4 we will also demonstrate this when testing the Kalman filter. We know that since $w_t \sim \mathcal{N}(0, Q_t)$,

$$p(x_t|x_{t-1}) \sim \mathcal{N}(Ax_{t-1} + B, Q_t), \tag{2.7}$$

and since $v_t \sim \mathcal{N}(0, R_t)$ we have

$$p(z_t|x_t) \sim \mathcal{N}(Hx_t, R_t). \tag{2.8}$$

It follows that,

$$p(x_t|z_t) \sim \mathcal{N}(\mu_t, \sigma_t^2), \tag{2.9}$$

for details see [22]. We want to find μ_t and σ_t^2. We define $\hat{x}_{t|t-1}$ to be the estimated value of x_t at time t, using information up to $t-1$ and $\mathbb{E}[\hat{x}_{t|t-1}]$ is the expected value of this estimate. The estimated error at time t is given by,

$$\begin{aligned} \triangle x_{t|t-1} &= x_t - \hat{x}_{t|t-1} \\ &= x_t - \mathbb{E}[\hat{x}_{t|t-1}] \\ &= Ax_{t-1} + B + w_t - (A\hat{x}_{t-1|t-1} + B) \\ &= A(x_{t-1} - \hat{x}_{t-1|t-1}) + w_t \\ &= A\triangle\hat{x}_{t-1|t-1} + w_t. \end{aligned}$$

We define the covariance of estimated error at time t given information up to $t-1$ as $P_{\{t|t-1\}}$. If we want to compute the covariance of estimated error $P_{\{t|t-1\}}$, then it suffices to compute $\mathbb{E}[\triangle x_{t|t-1} \triangle x_{t|t-1}^T]$[1], because $\mathbb{E}[\triangle x_{t|t-1}] = 0$. Thus, we get

[1]In this section T = transpose and not the maturity time, hence A^T = the transpose of a matrix A.

$$
\begin{aligned}
P_{\{t|t-1\}} &= \mathbb{E}\Big[\triangle x_{t|t-1}\triangle x_{t|t-1}^T\Big] \\
&= \mathbb{E}\Big[\Big(A\triangle x_{t-1|t-1}+w_t\Big)\Big(A\triangle x_{t-1|t-1}+w_t\Big)^T\Big] \\
&= \mathbb{E}\Big[\Big(A\triangle x_{t-1|t-1}+w_t\Big)\Big(\triangle x_{t-1|t-1}^T A^T+w_t^T\Big)\Big] \\
&= \mathbb{E}\Big[A\triangle x_{t-1|t-1}\triangle x_{t-1|t-1}^T A^T + A\triangle x_{t-1|t-1}w_t^T + w_t\triangle x_{t-1|t-1}^T A^T + w_t w_t^T\Big] \\
&= A\Big(\mathbb{E}\Big[\triangle x_{t-1|t-1}\triangle x_{t-1|t-1}^T\Big]\Big)A^T + \mathbb{E}\Big[w_t w_t^T\Big] \\
&= AP_{\{t-1|t-1\}}A^T + Q_t.
\end{aligned}
$$

Next, we define $\epsilon_{t|t-1}$, also called *innovation* or *measurement residual*, as the difference between our observed data at time t and the data we expected to observe at time $t-1$, that is,

$$\epsilon_{t|t-1} = z_t - H\hat{x}_{t|t-1}. \tag{2.10}$$

Define K as the Kalman gain matrix. The Kalman gain matrix is a correction added to the estimate, $\hat{x}_{t|t-1}$, that is proportional to the measurement residual. Therefore, we have

$$
\begin{aligned}
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + K\epsilon_{t|t-1} \\
&= \hat{x}_{t|t-1} + K(z_t - H\hat{x}_{t|t-1}).
\end{aligned} \tag{2.11}
$$

The matrix K is chosen in such a way that it minimizes, $P_{\{t|t\}}$, the covariance of estimated error at time t. So let us first find an expression for $P_{\{t|t\}}$, before we give an expression for K

$$
\begin{aligned}
P_{\{t|t\}} &= \mathbb{E}\Big[\triangle x_{t|t}\triangle x_{t|t}^T\Big] \\
&= \mathbb{E}\Big[\Big(x_t-\hat{x}_{t|t}\Big)\Big(x_t-\hat{x}_{t|t}\Big)^T\Big] \\
&= \mathbb{E}\Big[\Big(x_t-\hat{x}_{t|t-1}-K\epsilon_{t|t-1}\Big)\Big(x_t-\hat{x}_{t|t-1}-K\epsilon_{t|t-1}\Big)^T\Big] \\
&= \mathbb{E}\Big[\Big(x_t-\hat{x}_{t|t-1}-K(z_t-H\hat{x}_{t|t-1})\Big)\Big(x_t-\hat{x}_{t|t-1}-K(z_t-H\hat{x}_{t|t-1})\Big)^T\Big] \\
&= \mathbb{E}\Big[\Big(x_t-\hat{x}_{t|t-1}-K(Hx_t+v_t-H\hat{x}_{t|t-1})\Big)\Big(x_t-\hat{x}_{t|t-1}-K(Hx_t+v_t-H\hat{x}_{t|t-1})\Big)^T\Big] \\
&= \mathbb{E}\Big[\Big((I-KH)(x_t-\hat{x}_{t|t-1})-Kv_t\Big)\Big((I-KH)(x_t-\hat{x}_{t|t-1})-Kv_t\Big)^T\Big] \\
&= (I-KH)\mathbb{E}[(x_t-\hat{x}_{t|t-1})(x_t-\hat{x}_{t|t-1})^T](I-KH)^T+K_t\mathbb{E}[v_tv_t^T]K^T \\
&= (I-KH)P_{\{t|t-1\}}(I-KH)^T+KR_tK^T \\
&= P_{\{t|t-1\}}-P_{\{t|t-1\}}H^TK^T-KHP_{\{t|t-1\}}+KHP_{\{t|t-1\}}H^TK^T+KR_tK^T,
\end{aligned}
$$

where I = identity matrix. So we have

$$P_{\{t|t\}} = P_{\{t|t-1\}}-P_{\{t|t-1\}}H^TK^T-KHP_{\{t|t-1\}}+KHP_{\{t|t-1\}}H^TK^T+KR_tK^T. \quad (2.12)$$

If we want to minimize the covariance of $P_{\{t|t\}}$, we can use the mean square error measure

$$
\begin{aligned}
\mathbb{E}\Big[\ \Big|x_t-\hat{x}_{t|t}\Big|^2\ \Big] &= \mathrm{tr}\Big(\mathbb{E}[\triangle x_{t|t}\triangle x_{t|t}^T]\Big) \\
&= \mathrm{tr}(P_{\{t|t\}}).
\end{aligned}
$$

If we then take the trace of both sides and find the gradient with respect to K we get

$$\nabla_K \mathrm{tr}(P_{\{t|t\}}) = -P_{\{t|t-1\}}H^T-P_{\{t|t-1\}}H^T+2KHP_{\{t|t-1\}}H^T+2KR_t.$$

By setting the left-hand side equal to zero and making K the subject, we find the value K that minimizes $P_{\{t|t\}}$. Therefore, the Kalman gain matrix equals

$$K = P_{\{t|t-1\}}H^T(HP_{\{t|t-1\}}H^T+R_t)^{-1}. \quad (2.13)$$

Note that when the covariance matrix R_t, approaches zero the Kalman gain matrix weights the residual more heavily. In other words, the measurement space z_t is assumed to be more and more reliable. Our measurement in the short interest rate models is the compound interest rate, the historical data. The historical data must be trusted, therefore, we ought to minimize R_t, the variance of error term v_t in Equation (2.6), to get a reliable estimated parameters. However, if the estimated covariance matrix $P_{\{t|t-1\}}$ approaches zero we see that,

$$\lim_{P_{\{t|t-1\}}\to 0} K = 0, \tag{2.14}$$

which means that the Kalman gain matrix weights the residuals less heavily. In other words, the measurement space z_t is assumed to be less and less reliable. What is left to update is the covariance of estimated error at time t, $P_{\{t|t\}}$. We can do this by putting Equation (2.13) into Equation (2.12). This gives us

$$P_{\{t|t\}} = (I - KH)P_{\{t|t-1\}}. \tag{2.15}$$

So now we know the distribution of Equation (2.9), namely

$$p(x_t|z_t) \sim \mathcal{N}(\hat{x}_{t|t}, P_{\{t|t\}}). \tag{2.16}$$

The log-likelihood function is given by

$$\ell(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\ln\left[\det\left(P_{\{t|t\}}\right)\right] - \frac{1}{2}\epsilon_t\left(P_{\{t|t\}}\right)^{-1}\epsilon_{t|t-1}^{T}, \tag{2.17}$$

with n as the dimension space of x_t as defined at the start and N = the number of the collected data. The Equation (2.17) is also called *Quasi Maximum logLikelihood estimator* which best explains the observed values of x_t.

Putting all the steps together we have the Kalman filter algorithm:

1. Choose initial values for $\hat{x}_{t-1|t-1}$ and $P_{\{t-1|t-1\}}$,
2. $\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1} + B$,
3. $P_{\{t|t-1\}} = AP_{\{t-1|t-1\}}A^T + Q_t$,
4. $\epsilon_{t|t-1} = z_t - H\hat{x}_{t|t-1}$,
5. $K = P_{\{t|t-1\}}H^T(HP_{\{t|t-1\}}H^T + R_t)^{-1}$,
6. $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K\epsilon_{t|t-1}$,
7. $P_{\{t|t\}} = (I - KH)P_{\{t|t-1\}}$.

In this thesis we will only show how the Kalman filter is applied to the CIR model, because the steps for the Vasicek model are almost identical. The only differences are the choice of the matrices in Equation (2.5) and the moments.

2.3 Calibration with the Long Term Quantile method

In this section we will explain the Long Term Quantile (LTQ) method, which is the method used by Rabobank for calibration purposes. The major assumption in this model is that the quantiles from the historical data are representative for quantiles in the future. Therefore, a 95% confidence interval is taken from the historical data and the parameters in the short interest rate model are chosen such that in 95% of the cases the generated interest rates will fall within the confidence interval taken from the historical data. Sometimes the best way to explain a method is by given an example. We will show how this method works for the Vasicek model. We know that for the Vasicek model the compound interest model is normally distributed. Recall,

$$R(t,\tau) = \alpha_\tau + \beta_\tau r(t), \tag{2.18}$$

and

$$\mathrm{d}R_{(t,\tau)} = \kappa\bigg(\beta_\tau\mu - \alpha_\tau - R_{(t,\tau)}\bigg)\mathrm{d}t + \beta_\tau\sigma\mathrm{d}W_t. \tag{2.19}$$

Therefore, we get

$$\begin{aligned}
\mathbb{E}\big[R_{(t+\triangle t,\tau)}|R_{(t,\tau)}\big] &= \mathbb{E}\big[\alpha_\tau + \beta_\tau r_{t+\triangle t}|R_{(t,\tau)}\big] \\
&= \alpha_\tau + \beta_\tau\mathbb{E}\big[r_{t+\triangle t}|R_{(t,\tau)}\big] \\
&= \alpha_\tau + \beta_\tau e^{-\kappa\triangle t}r_t + \mu(1-e^{-\kappa\triangle t}),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\big[R_{(t+\triangle t,\tau)}|R_{(t,\tau)}\big] &= \mathrm{Var}\big[\alpha_\tau + \beta_\tau r_{t+\triangle t}|R_{(t,\tau)}\big] \\
&= \mathrm{Var}\big[\beta_\tau r_{t+\triangle t}|R_{(t,\tau)}\big] \\
&= \beta_\tau^2\mathrm{Var}\big[r_{t+\triangle t}|R_{(t,\tau)}\big] \\
&= \beta_\tau^2\big(\tfrac{\sigma^2}{2\kappa}\big)(1-e^{-2\kappa\triangle t}).
\end{aligned}$$

Therefore, if $t\to\infty$ we get

$$\lim_{t\to\infty} R_{(t,\tau)} \sim \mathcal{N}\bigg(\beta_\tau - \alpha_\tau, \beta_\tau^2\frac{\sigma^2}{2\kappa}\bigg). \tag{2.20}$$

Another key assumption that is made in the LTQ method is that the drift term is ignored. There is a logical reason behind this, Rabobank collects the data on a daily bases which means that $\mathrm{d}t = \triangle t$ is very small and hence can be ignored. This means,

$$\mathrm{d}R_{(t,\tau)} = \beta_\tau\sigma\mathrm{d}W_t. \tag{2.21}$$

Using Equation (2.21) with the collected historical data we can solve $\beta_\tau\sigma$. Suppose we collected N different data this gives us,

$$\beta_\tau\sigma = \sqrt{\frac{\sum_{t=1}^{N}\big(R_{(t,\tau)} - R_{(t-1,\tau)}\big)}{\big(N-1\big)\triangle t}}. \tag{2.22}$$

Thus, for the 95% confidence interval we have

$$\mathrm{P}\bigg(\beta_\tau - \alpha_\tau - q\beta_\tau^2\frac{\sigma^2}{2\kappa} < \lim_{t\to\infty} R_{(t,\tau)} < \beta_\tau - \alpha_\tau + q\beta_\tau^2\frac{\sigma^2}{2\kappa}\bigg) = q, \tag{2.23}$$

with $q = 0.95$. We define (x, y, z) as follows,

$$\begin{aligned} x &= \beta_\tau\sigma, \\ y &= \beta_\tau - \alpha_\tau - q\beta_\tau^2\frac{\sigma^2}{2\kappa}, \\ z &= \beta_\tau - \alpha_\tau + q\beta_\tau^2\frac{\sigma^2}{2\kappa}. \end{aligned}$$

We can solve x using Equation (2.21) and (y, z) can be collected from the historical data. Now we can estimate the parameters κ, μ and σ. This gives us

$$\begin{aligned} \kappa &= 2\bigg(\frac{xq}{z-y}\bigg)^2, \\ \sigma &= \frac{x}{\beta_\tau}, \\ \mu &= \frac{y+z}{2} + \frac{\sigma^2 - x\sigma}{2\kappa^2} - \frac{x^2\tau}{4\kappa}. \end{aligned}$$

2.4 Testing and Comparing the Methods

In the previous sections we described the calibration of 4 different methods. Before using any of the methods it is wise to test the accuracy of the methods. Some may suggest using Monte Carlo, but we do not want to use Monte Carlo simulation for the simple reason that we only collect the data for 9 different contracts and not thousands of different contracts. The collected data from Rabobank is from 1 January 2001 up to 1 September 2011. That is approximately 10 years of data. Therefore, we simulate 10 years of data using Euler's forward scheme. We want to eliminate (or minimize to almost zero) any bias caused by the discretization. Before we explain how we will do that, we first describe the model that we will use for this test. In Chapter 1 we discussed three different models.

For the test we will use the Vasicek model. There are two reasons for this. Firstly, the Vasicek model is the easiest model to work with. Secondly, the parameters in this model can be estimated using any method, unlike the CIR model which parameters can only be fitted with the Kalman filter. Therefore, the Vasicek model is the only "fair" model to do this test. We want to eliminate the bias in the discretization. This can be done as follows. First, we choose the parameters $\kappa = 0.1$, $\mu = 0.05$ and $\sigma = 0.02$, with initial value $r_0 = 0.06$. Then, we simulate the short rates using Euler's forward scheme with $\triangle t = \frac{1}{12}$ (assuming that we are measuring the data once every month) $\frac{\triangle t}{10}$ for 9 different contracts defined as $\tau = \{1 \text{ month}, 3 \text{ months}, 6 \text{ months}, 1 \text{ year}, 2 \text{ years}, 5 \text{ years}, 10 \text{ years}, 20 \text{ years}, 30 \text{ years}\}$. Thus,

$$r_{t+\frac{\triangle t}{10}} = r_t + \kappa(\mu - r_t)\frac{\triangle t}{10} + \sigma\sqrt{r_t}\big(W_{t+\frac{\triangle t}{10}} - W_t\big). \tag{2.24}$$

Then, we take the compound interest as a function of the 10th short interest rates,

$$R(t + \triangle t, \tau) = \alpha_\tau + \beta_\tau r(t + \triangle t). \tag{2.25}$$

We want to show the strength of the Kalman filter method. Therefore, we first estimate the parameters for one contract $\tau = 1$ month. We define $\hat{\kappa}$, $\hat{\mu}$ and $\hat{\sigma}$ as the estimated values for κ, μ and σ respectively. If we look at Table 2.1, Table 2.2 and Table 2.3, we see that the diffusion parameter σ comes with the smallest bias. This is because the Euler discretization of the Vasicek model is only confined to the diffusion part. In general, discretization schemes are likely to encounter this kind of systematic bias [13]. However, if $\triangle t \to 0$ the bias in the diffusion parameter goes to zero.

Parameter	σ	$\hat{\sigma}$	Bias	Bias (%)
LSM	0.02	0.0176	0.0024	12
MLE	0.02	0.0186	0.0014	7
LTQ	0.02	0.0180	0.0020	10
Kalman Filter	0.02	0.0130	0.0070	34

Table 2.1: Estimated σ values for each method using only $\tau = 1$ month.

If we look at Table 2.2 and 2.3 we see that there is a large bias in the estimated parameter when using LSM and MLE, especially when estimating the κ parameter. As mentioned before the bias may be very large. The LTQ method is the best calibration method when using just one contract (e.g., $\tau = 1$). Note that the Kalman filter in this case estimates the parameters with a large bias just like LSM and MLE.

Parameter	κ	$\hat{\kappa}$	Bias	Bias (%)
LSM	0.1	0.6506	0.5506	551
MLE	0.1	0.6689	0.5689	569
LTQ	0.1	0.1582	0.0582	58
Kalman Filter	0.1	0.5162	0.4162	416

Table 2.2: Estimated κ values for each method using only $\tau = 1$ month.

If we now estimate the parameters using all the contracts, $\tau = \{1 \text{ month}, 3 \text{ months}, 6 \text{ months}, 1 \text{ year}, 2 \text{ years}, 5 \text{ years}, 10 \text{ years}, 20 \text{ years}, 30 \text{ years}\}$. Then we see that the bias in the

Parameter	μ	$\hat{\mu}$	Bias	Bias (%)
LSM	0.05	0.0843	0.0343	69
MLE	0.05	0.0843	0.0343	69
LTQ	0.05	0.0775	0.0275	55
Kalman Filter	0.05	0.0747	0.0247	49

Table 2.3: Estimated μ values for each method using only $\tau = 1$ month.

Kalman filter disappears. In other words, the independent error term v_t from Equation (2.6) has variance zero. Note this is only the case in a test scenario, because when dealing with the real data we would never encounter the independent error term v_t with variance zero, simply because no theoretical model can fit the historical data exactly. Therefore, when estimated the real data in Chapter 3 the reliability of our estimated parameters depends heavily on the variance of the independent error term.

For MLE the bias in the parameter σ increases rapidly, this is because as the duration of the contract increases, $\tau = \{10 \text{ years}, 20 \text{ years}, 30 \text{ years}\}$, the variance of $R(t, \tau)$ increases and hence the bias in the σ sigma parameter increases. This is simultaneously the reason behind the bias increase for LSM even though the increase is much smaller there. We can see that the bias in the drift parameter μ decreases by a small margin. This is simply because for each contract the long term mean must converge to μ. Therefore, the more data we collect the easier it is to estimate the long term mean. If we had used the Monte Carlo method to do this test, then the bias in the parameter μ would go to zero.

The parameter κ is by far the most difficult one to estimate when using classic techniques like MLE and LSM, the bias reduction techniques are, therefore, mainly designed to reduce the bias in this parameter. Like in most cases the more data collected does not necessarily mean a better estimate. When dealing with only one contract we have seen that LTQ outperforms the other methods including the Kalman filter. However, when we put in all the collected data we see that the Kalman filter is by far the best method, because any sort of biasness is eliminated. Therefore, the Kalman filter is by far the best calibration method. We strongly advise this method above the alternatives (MLE, LSM and LTQ). The LTQ method is still a better alternative than MLE and LSM as the bias for the κ estimator is reduced by a factor of 10 and when dealing with more contracts there is no increase in the bias for the parameter σ.

Parameter	σ	$\hat{\sigma}$	Bias	Bias (%)
LSM	0.02	0.0138	0.0062	31
MLE	0.02	0.0510	0.0310	155
LTQ	0.02	0.0209	0.0009	5
Kalman Filter	0.02	0.02	0	0

Table 2.4: Estimated σ values using all the simulated data for τ.

Why does the Kalman filter method outperform the other methods? That is because the Kalman filter exploits all the collected data! The methods LSM, MLE and LTQ estimate the parameters of $R(t+\triangle t, \tau)$ separately for each τ and average them to get the results. Therefore, it does not matter if one has the collected data for just one contract or for 9 contracts because the methods do not use the collected data for different contracts efficiently. The Kalman filter

Parameter	κ	$\hat{\kappa}$	Bias	Bias (%)
LSM	0.1	0.6506	0.5506	551
MLE	0.1	0.6690	0.5690	569
LTQ	0.1	0.1582	0.0582	58
Kalman Filter	0.1	0.1	0	0

Table 2.5: Estimated κ values using all the simulated data for τ.

Parameter	μ	$\hat{\mu}$	Bias	Bias (%)
LSM	0.05	0.0741	0.0241	48
MLE	0.05	0.0723	0.0223	45
LTQ	0.05	0.0713	0.0213	43
Kalman Filter	0.05	0.05	0	0

Table 2.6: Estimated μ values using all the simulated data for τ.

on the other hand distinguishes itself from the other methods mainly because the Kalman filter uses each collect contract efficiently. As we have seen, when only using one contract for the calibration (instead of 9 contracts) the Kalman filter comes with a large bias.

Next, we plot the yield curve, see Figure 2.1, and see what the error in the approximation does to our yield curve. Therefore, we use the estimated parameters for each method and plot them together with the actual yield curve, which has parameters $\kappa = 0.1$, $\mu = 0.05$ and $\sigma = 0.02$. We can see that even though LTQ parameters have a relatively small bias compared to LSM and MLE the yield curve of LTQ is closer to the estimated yield curve of LSM and MLE, which is way off, rather than the actual yield curve. Therefore, a relatively small biases in the parameters can still have a huge impact on the actual yield curve. This problem can be avoided by using the Kalman filter.

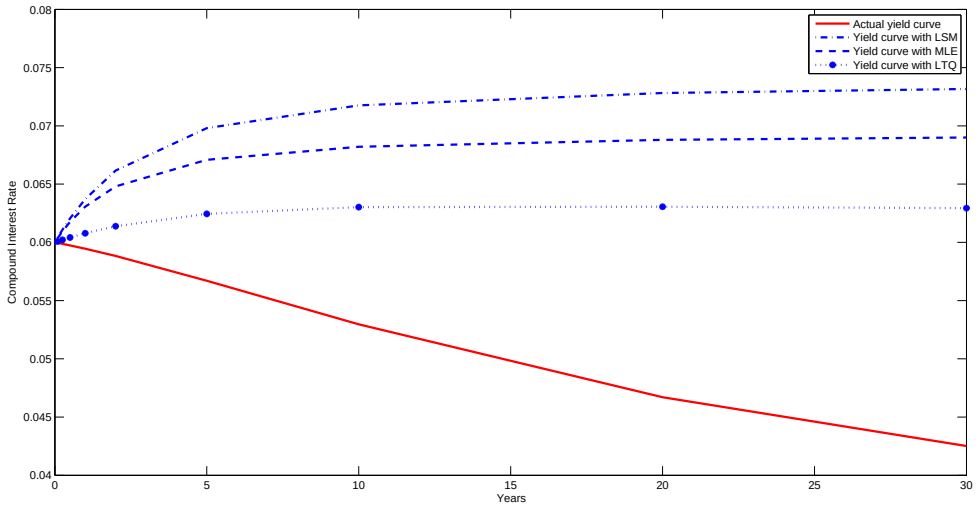

Figure 2.1: A plot of the actual initial test yield curve against estimated initial yield curves using MLE, LSM and LTQ.

The reason for calibration is to eventually try to predict the future and compute the fairest bond price. We do not want to overestimate the bond price as this means that we are selling the bond price higher than it actually is. If another company can lower their prices with the same risk, then we lose out. However if we underestimate the bond price, this means that we sell a bond price at a lower cost than it actually is, losing money and risking bankruptcy. We will compute the bond price using Monte Carlo. Our initial value is $r_0 = 0.06$ and we simulate 1000 paths using Eueler discretization,

$$r_{t+\frac{\triangle t}{10}} = r_t + \kappa(\mu - r_t)\frac{\triangle t}{10} + \sigma\sqrt{r_t}\big(W_{t+\frac{\triangle t}{10}} - W_t\big).$$

We want to see the impact of the bias on the estimated parameters on the short and long term. Therefore, we plot in Figures 2.2, 2.3 and 2.4 the bond price for a duration of 1 month, 1 year and 10 years respectively. We can see for a contract with a duration of 1 month the bias in the estimated parameters using LSM and LTQ results in an underestimate of the bond price, whereas estimation using MLE overestimates the bond price. For a contract with a duration of 1 year or 10 years, the bias in the estimated parameters using LSM, MLE and LTQ all result in an underestimate of the bond price. The gab in the "real" bond price and the estimated bond prices, using LSM, MLE and LTQ, increases as the years increase. This results in a loss which would have been prevented using Kalman filter. Therefore, we can conclude that the Kalman filter is the best method for calibration. In Section 3.2 we will only work with the Kalman filter to estimate the parameters in the CIR model. This is not only due to the fact that the Kalman filter is a better alternative but also because LSM, MLE and LTQ cannot be applied to the CIR model as we have shown in Chapter 1.

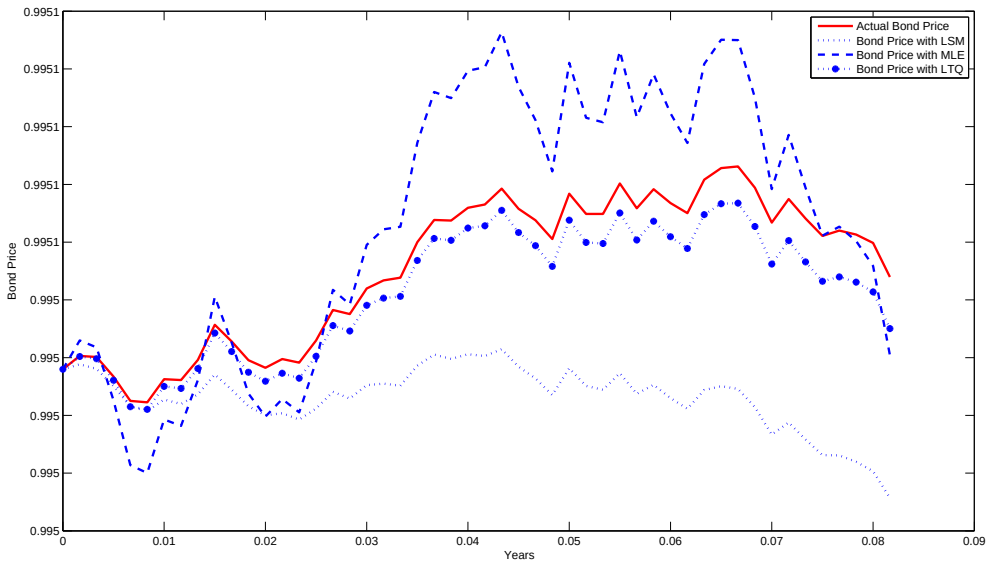

Figure 2.2: A plot of the actual bond price of $\tau = 1$ month over a period of 1 month compared to a plot using MLE, LSM and LTQ.

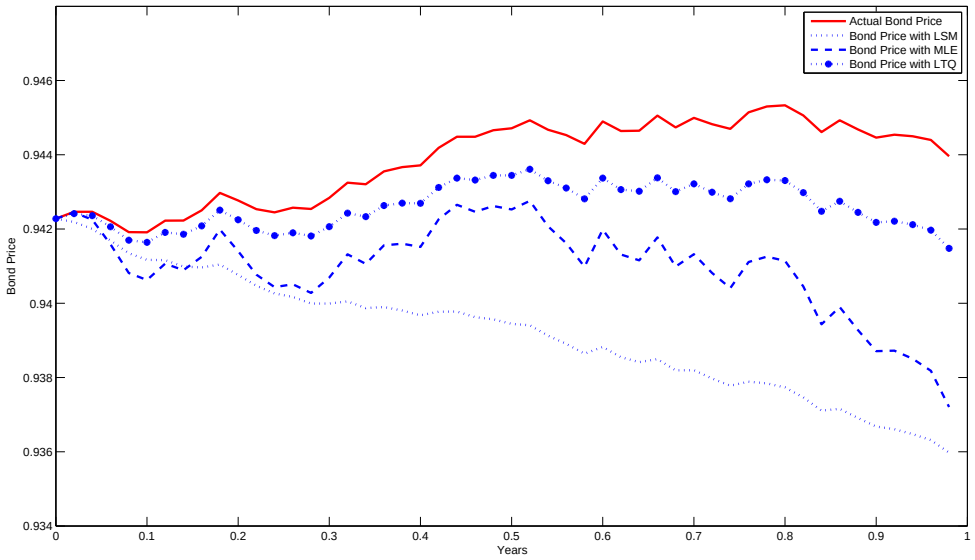

Figure 2.3: A plot of the actual bond price of $\tau = 1$ month over a period of 1 year compared to a plot using MLE, LSM and LTQ.

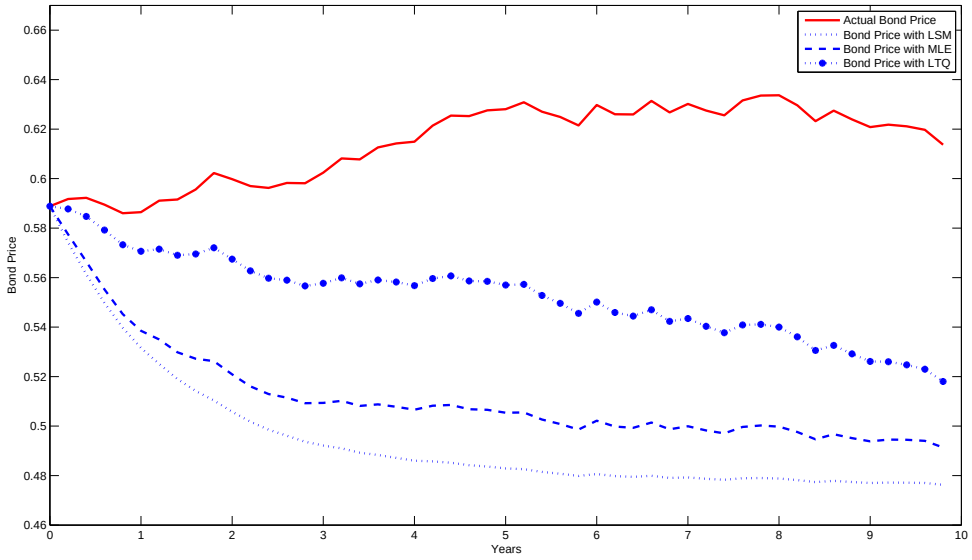

Figure 2.4: A plot of the actual bond price of $\tau = 1$ month over a period of 10 years compared to a plot using MLE, LSM and LTQ.

3

Forecasting

We have seen in Chapter 2 that the Kalman filter estimates the parameters in the Vasicek model without any bias. The same argument holds for the CIR model. In this chapter we will model the interest rate using the Vasicek model, the CIR model and the Nelson-Siegel model.

3.1 Modeling data with the Vasicek model

To keep the order as in Chapter 1 we start again with the Vasicek model. As descriped before, the major handicap in this model is that it generates negative values for r_t. However, we still want to use the Vasicek model as comparison for the CIR model later on in this chapter. Even though we strongly advise working with Kalman filter for calibration, we want to see what we get if we did the calibration using, LSM, MLE or LTQ. Table 3.1 shows the results.

Parameter	κ	μ	σ
LSM	0.3856	0.0264	0.0069
MLE	0.3906	0.0238	0.0325
LTQ	0.0904	0.0368	0.0113
Kalman Filter	0.2039	0.0492	0.0045

Table 3.1: Estimated parameters, for the Vasicek model, of the historical data (from 1 Januray 2001 - 1 September 2011) for EUR.

The dimension space of τ is 9, because we have the historical data for 9 different contracts. This means that the dimension space of the variance of the independent error term v_t of Equation (2.6) has dimension space of 9×9, with non zero values only at the diagonal. Recall, the values at the diagonal were defined as $\rho_1, \cdots, \rho_n$. The corresponding $\rho_1, \cdots, \rho_n$, with the estimated parameters of the Vasicek model are given in Table 3.2.

Next, we plot the yield curves. Unlike in Section 2.4 we are now dealing with 9 different values for $R(0, \tau)$ (we take $t = 0$ to be equal to 1 January 2001). This is because, in Section 2.4 we choose one initial value for r_t, hence we had one initial value for $R(0, \tau)$ that corresponded to all maturities and plotted our yield curves for this starting value. This is not the case for

Variance	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9
	0.003	0.001	0	0.001	0.003	0.004	0.004	0.005	0.006

Table 3.2: The variance of the independent error term v_t of Equation (2.6) for the Vasicek model.

the historical data and this is also why short rate models give a poor fit of the historical yield curves as we will see later on in this section. Our 9 initial values are

$$\begin{aligned} R(0,\tau_1) &= 0.0478, \\ R(0,\tau_2) &= 0.0473, \\ R(0,\tau_3) &= 0.0460, \\ R(0,\tau_4) &= 0.0444, \\ R(0,\tau_5) &= 0.0455, \\ R(0,\tau_6) &= 0.0484, \\ R(0,\tau_7) &= 0.0527, \\ R(0,\tau_8) &= 0.0568, \\ R(0,\tau_9) &= 0.0572, \end{aligned}$$

with τ_1 = contract for 1 month, τ_2 = contract for 3 months, ..., τ_9 = contract for 30 years. In other words, instead of one initial value we have 9 different initial values and, hence we get 9 different yield curves. Therefore, we use the estimated parameters for each method and plot them together with the actual initial yield curve. See Figures 3.1, 3.2, up to, 3.11 for a plot of the yield curves using the Kalman filter, LSM, MLE and LTQ. We can see that the Kalman filter estimation is the closest to the actual initial yield curve for each initial value of $R(0,\tau_i)$, with $i = 1,\ldots,9$ followed by LTQ, as was the case in Section 2.4.

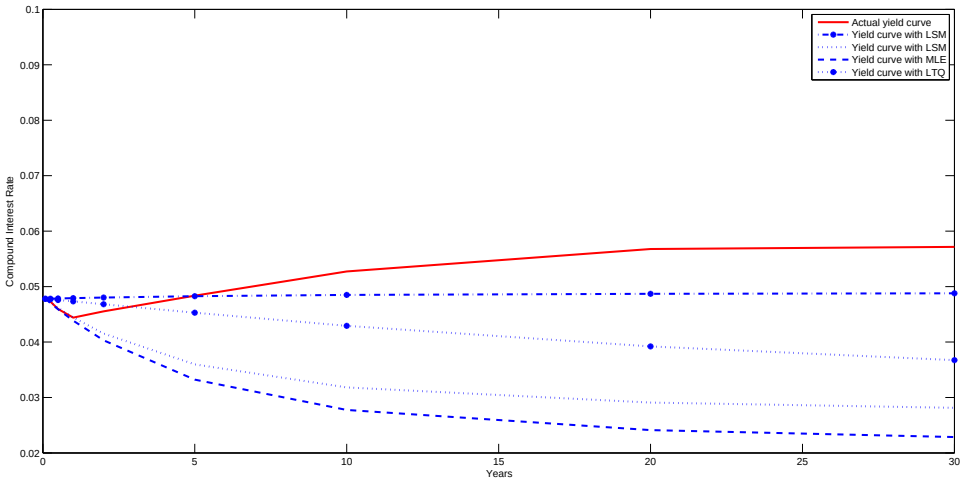

Figure 3.1: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_1)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

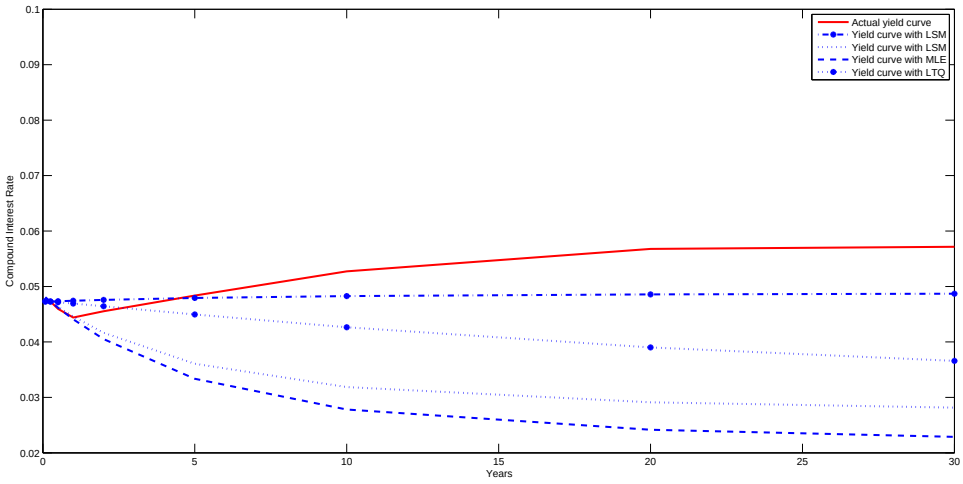

Figure 3.2: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_2)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

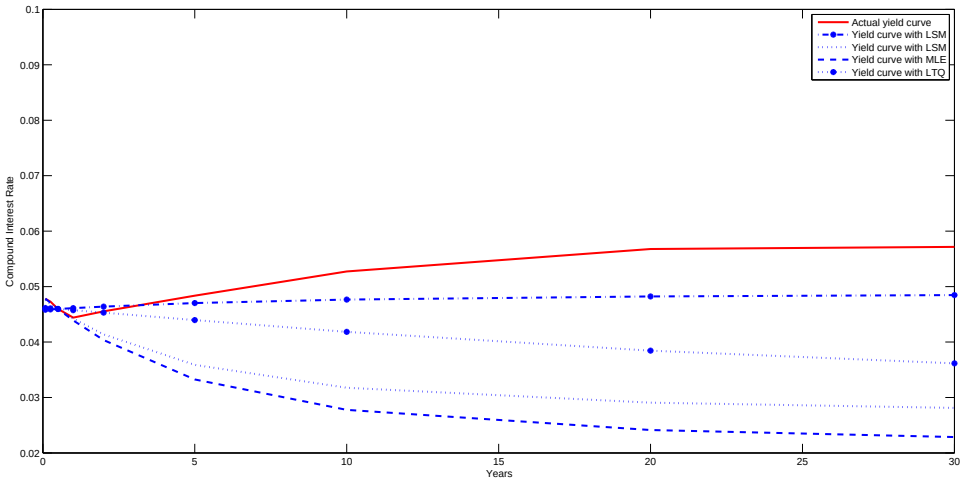

Figure 3.3: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_3)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

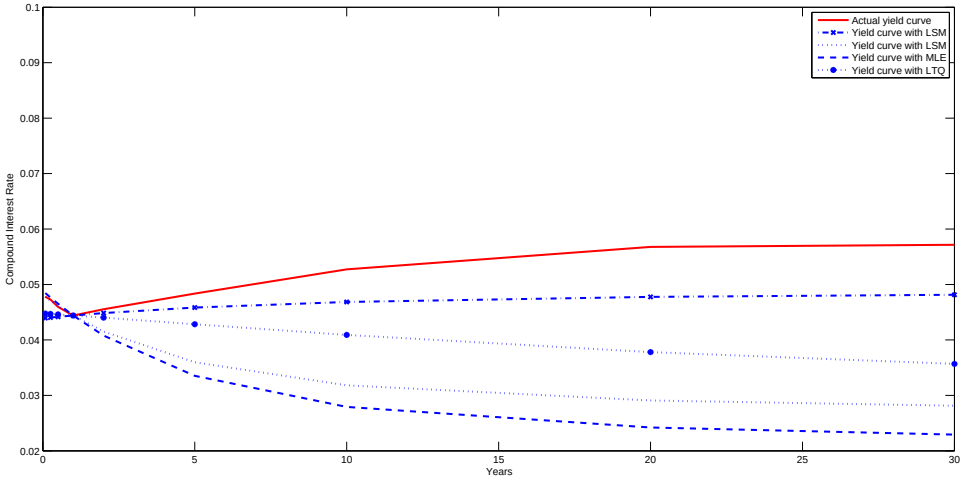

Figure 3.4: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_4)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

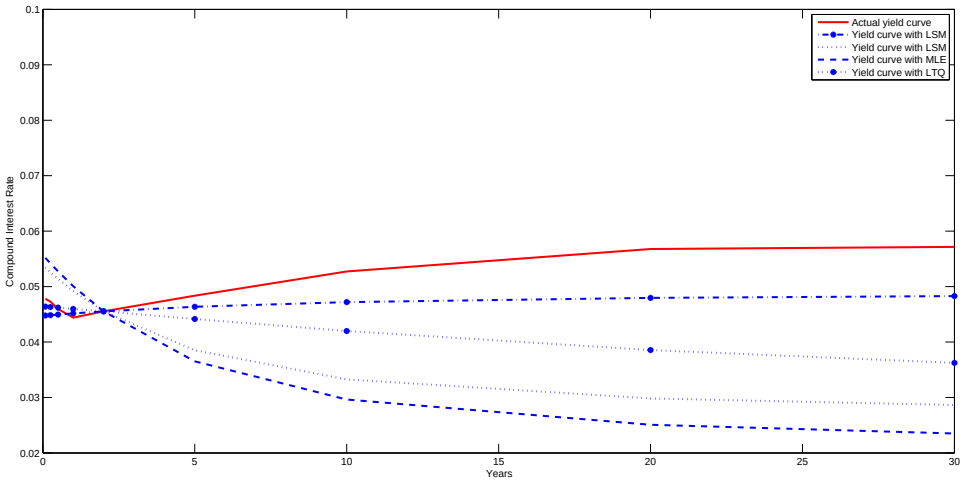

Figure 3.5: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_5)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

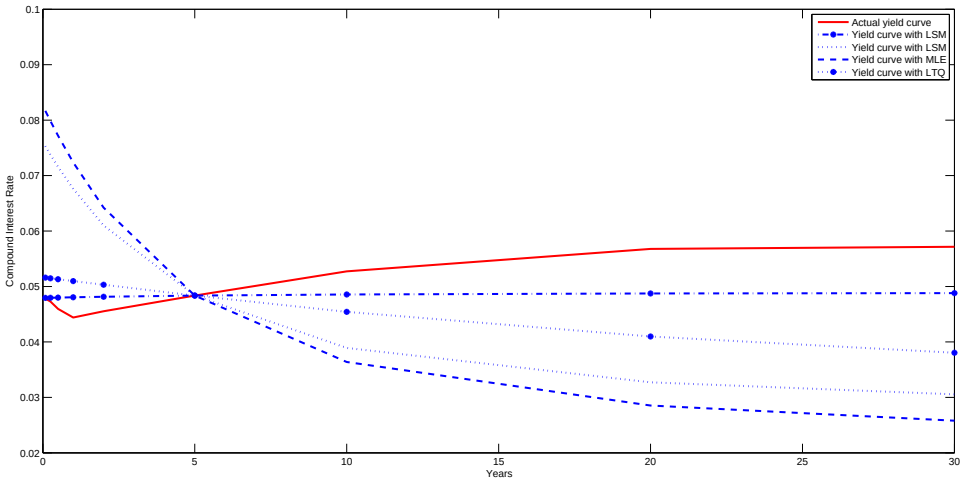

Figure 3.6: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0,\tau_6)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

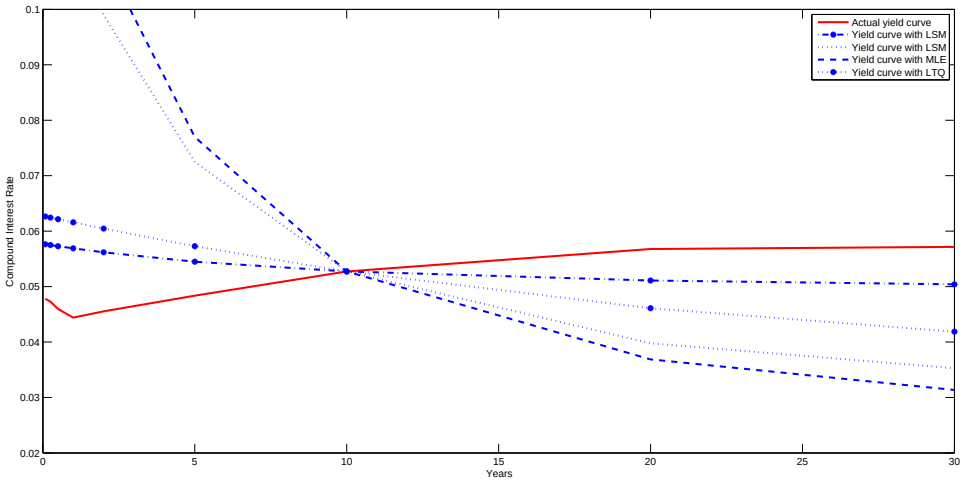

Figure 3.7: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0, \tau_7)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

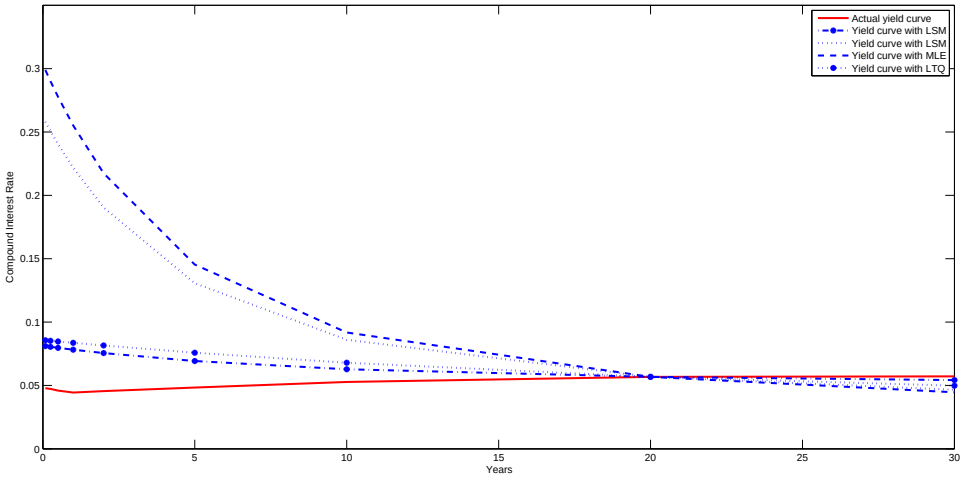

Figure 3.8: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0, \tau_8)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

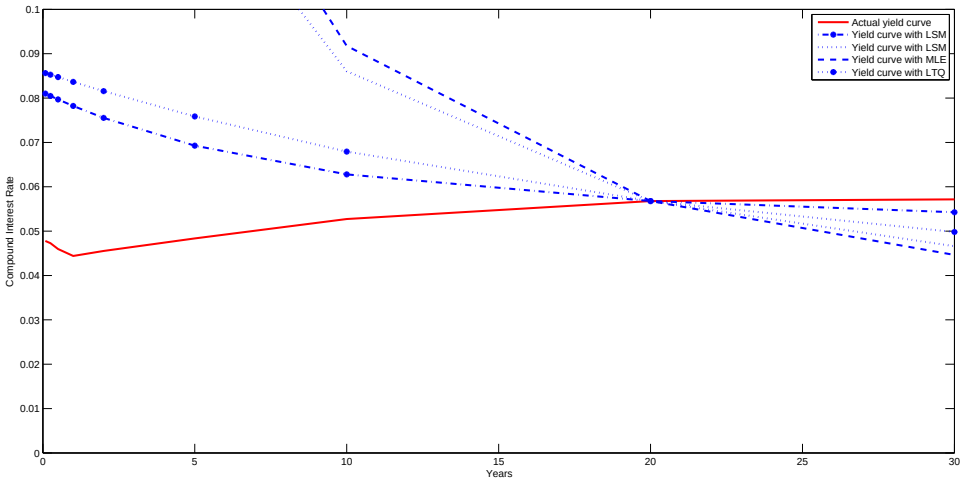

Figure 3.9: A closed up plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0, \tau_9)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

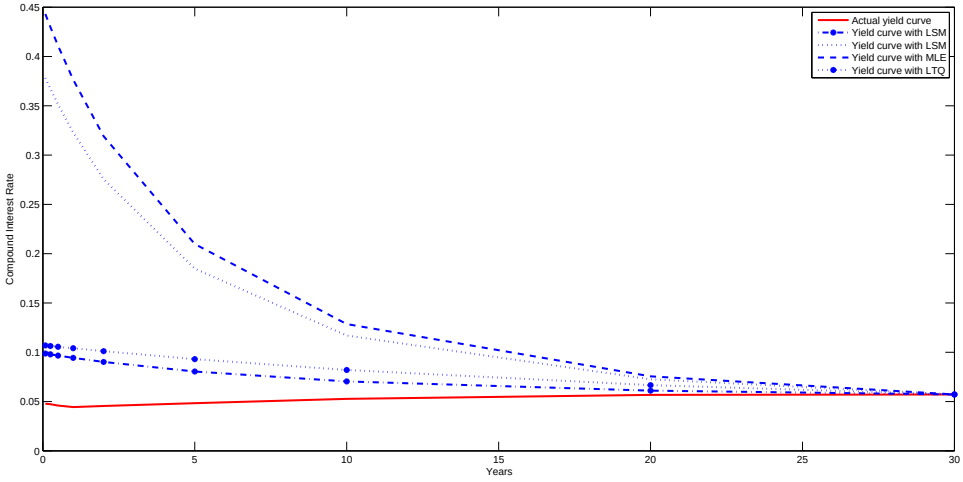

Figure 3.10: A plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0, \tau_8)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

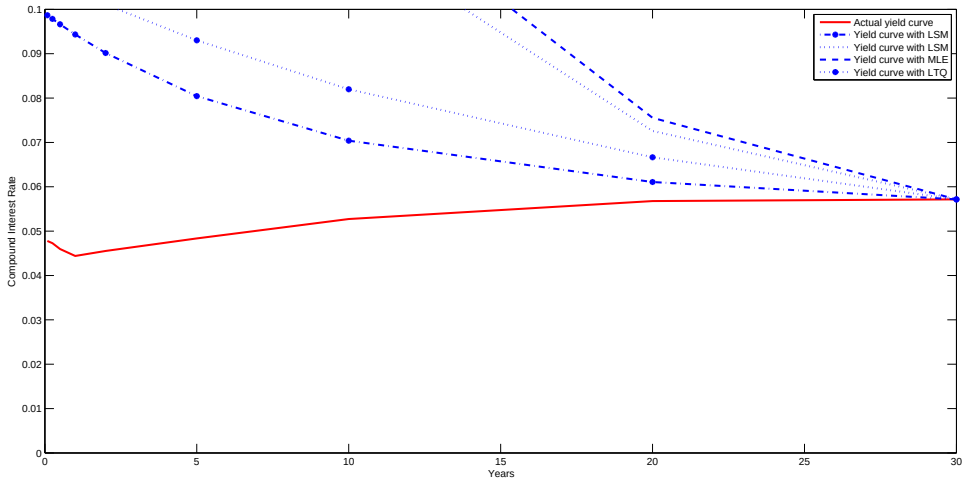

Figure 3.11: A closed up plot of the actual initial yield curve, for the Vasicek model, against estimated initial yield curves, with $R(0, \tau_9)$, using the Kalman Filter (KF), MLE, LSM and LTQ.

Now we give a plot of the KF estimated parameters for the for several different time steps, see Figures 3.12, up to, 3.16.

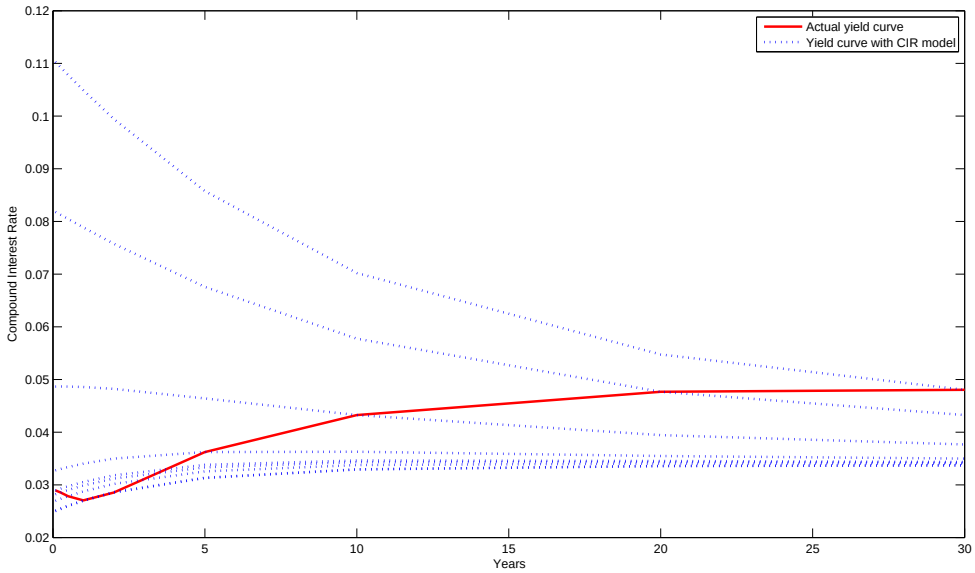

Figure 3.12: A plot of the actual yield curve on 01/01/2003, for the Vasicek model, against estimated yield curves, using the Kalman Filter (KF).

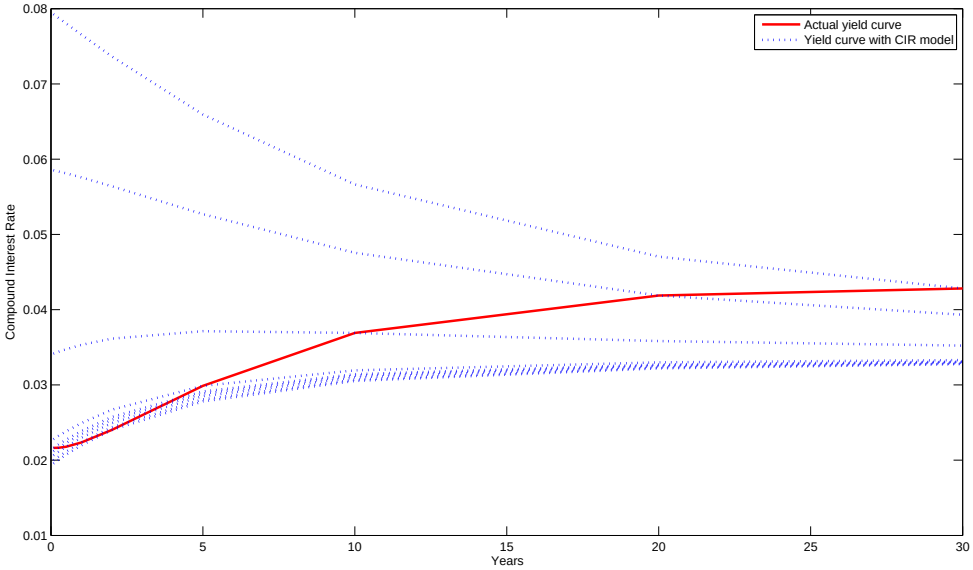

Figure 3.13: A plot of the actual yield curve 03/01/2005, for the Vasicek model, against estimated yield curves, using the Kalman Filter (KF).

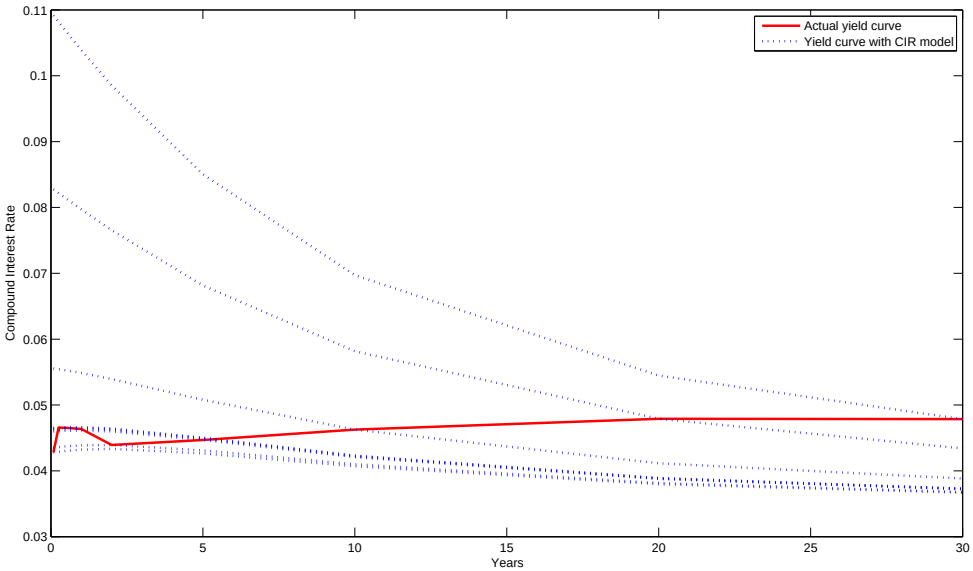

Figure 3.14: A plot of the actual yield curve 01/01/2008, for the Vasicek model, against estimated yield curves, using the Kalman Filter (KF).

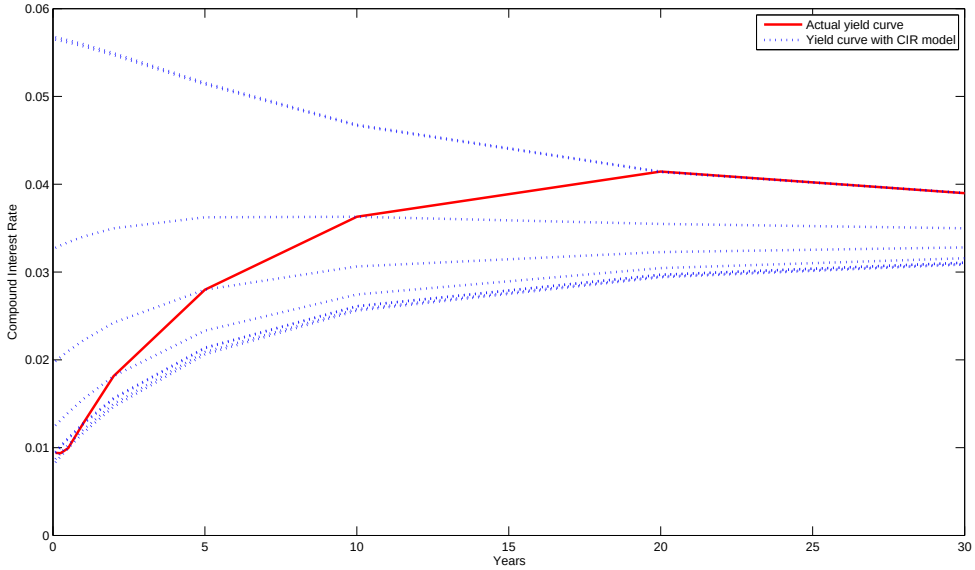

Figure 3.15: A plot of the actual yield curve 01/01/2010, for the Vasicek model, against estimated yield curves, using the Kalman Filter (KF).

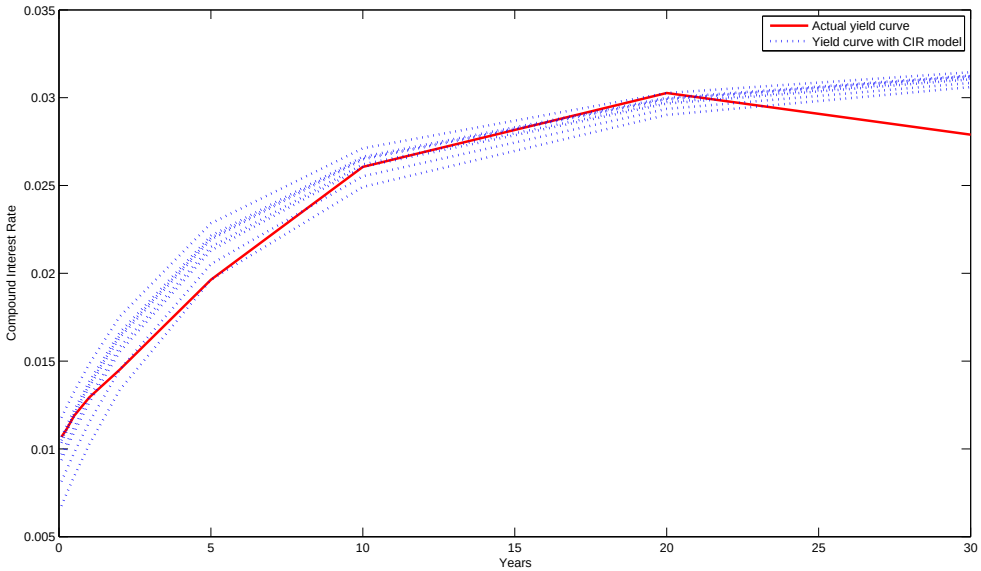

Figure 3.16: A plot of the actual yield curve 03/01/2011, for the Vasicek model, against estimated yield curves, using the Kalman Filter (KF).

Next, we plot the estimated bond price for a contract with maturity of 1 month, 1 year and 10 years over a period of 1 month, 1 year and 10 years respectively for each method. We compute our estimated bound price in the same way as in Chapter 2 using Monte Carlo simulation, in Figures 3.17, 3.18 and 3.19 we can see the results. Note that such as in Section 2.4, the LTQ method is closed to the Kalman filter estimation.

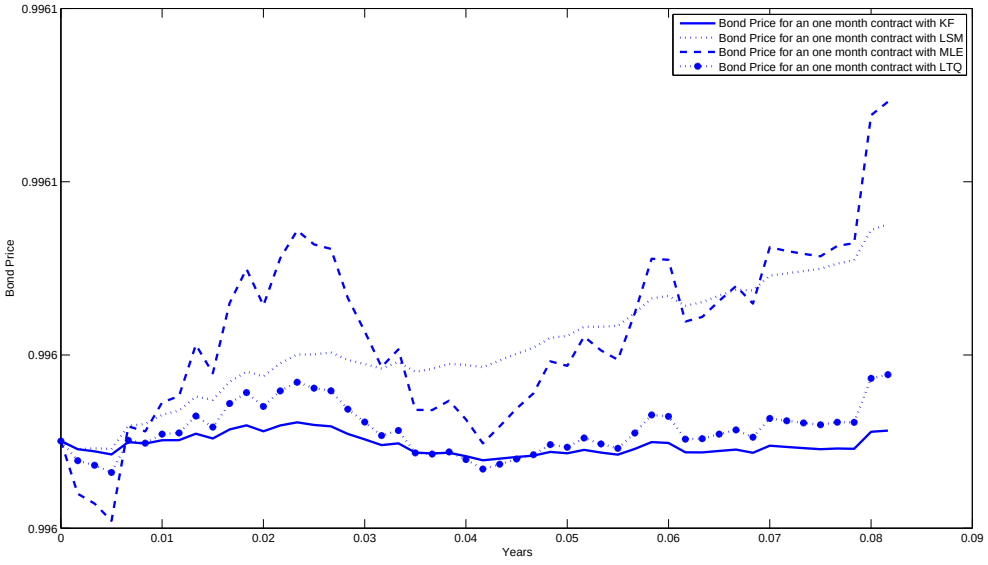

Figure 3.17: A plot of the simulated bond price for $\tau = 1$ month with initial value $R(0, \tau_1)$, for the Vasicek model, for 1 month compared to a plot using the Kalman Filter (KF) MLE, LSM and LTQ.

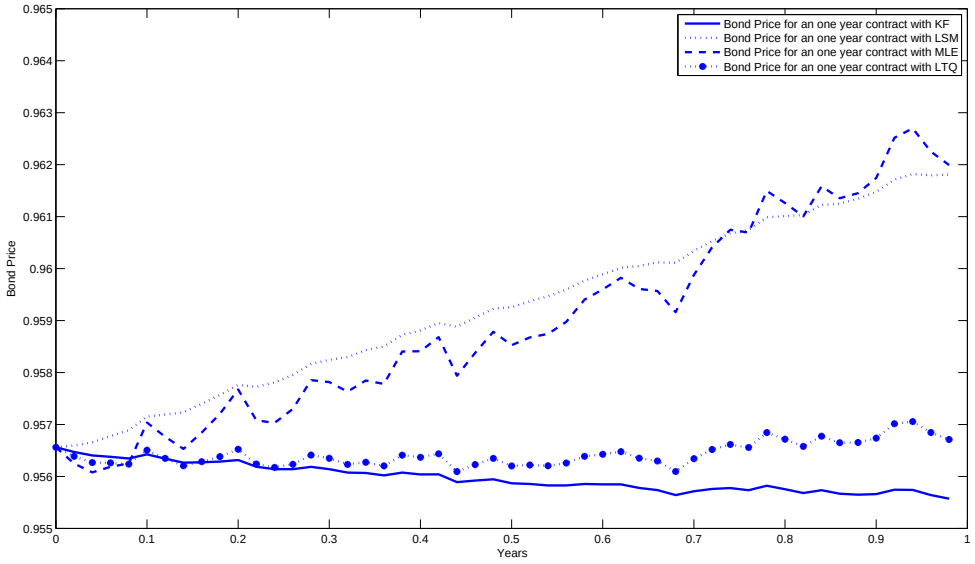

Figure 3.18: A plot of the simulated bond price for $\tau = 1$ year with initial value $R(0, \tau_4)$, for the Vasicek model, for 1 year compared to a plot using the Kalman Filter (KF) MLE, LSM and LTQ.

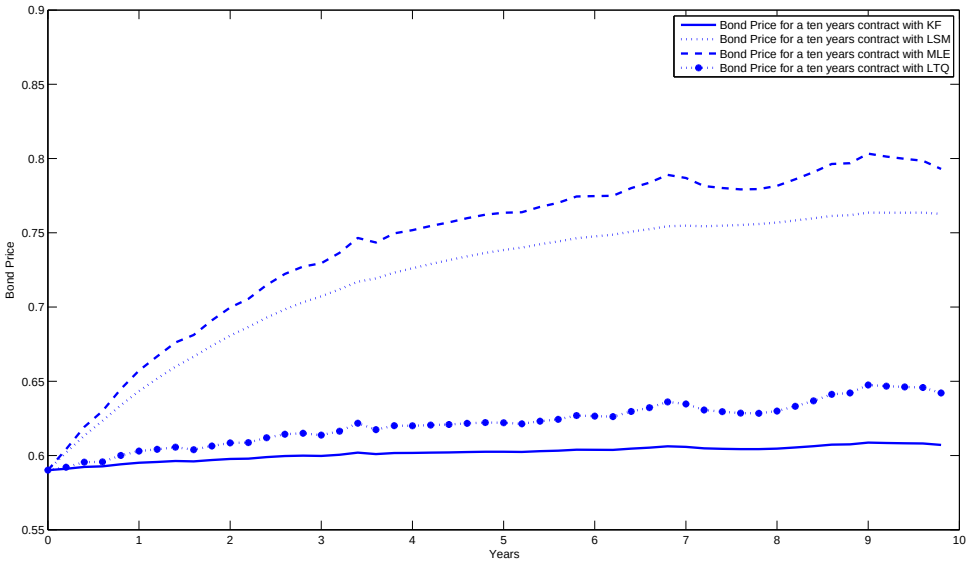

Figure 3.19: A plot of the simulated bond price for $\tau = 10$ years with initial value $R(0, \tau_7)$, for the Vasicek model, for 10 years compared to a plot using the Kalman Filter (KF) MLE, LSM and LTQ.

3.2 Modeling data with the Cox Ingersoll Ross model

In this section we will model the historical data using the CIR model. The parameters in this model can only be estimated using the Kalman filter. We will give a brief explanation of the Kalman filter applied to the CIR model. There is a relationship between compound interest rate $R_{(t,\tau)} = R_t(\tau)$ and the short interest rate r_t, as shown in the previous chapters. We define the short interest rate, r_t, as transition space, with w_t defined as in Equation (2.5). Therefore, we have

$$r_t = G + Hr_{t-1} + w_t, \tag{3.1}$$

with

$$\begin{aligned} G &= \mu(1 - e^{-\kappa\triangle t}), \\ H &= e^{-\kappa\triangle t}, \\ w_t &= r_t e^{-\kappa\triangle t}\left(\frac{\sigma^2}{\kappa}\right)(1 - e^{-\kappa\triangle t}) + \mu\left(\frac{\sigma^2}{2\kappa}\right)(1 - e^{-\kappa\triangle t})^2. \end{aligned}$$

The measure space is defined as

$$R_t(\theta) = C(\theta) + D(\theta)r_t + v_t, \tag{3.2}$$

where $\theta = (\kappa, \mu, \sigma)$, $D(\theta)$ and $C(\theta)$ are vectors with $C(\theta) = c_i = \frac{-\ln(A(\tau_i))}{\tau_i}$ and $D(\theta) = d_i = \frac{B(\tau_i)}{\tau_i}$ and v_t defined as in Equation (2.6)[1].

Now that we have described the space representation of the CIR model. We will apply the Kalman filter. The first step in the algorithm is to choose the initial mean as $\mathbb{E}[r_t|r_0\big] = \mu$ and initial variance as $\mathrm{Var}\big[r_t|r_0\big] = \frac{\mu\sigma^2}{2\kappa}$. See below, were we outlined the updating equations used in the Kalman filter.

$$R_t = C(\theta) + D(\theta)\mathbb{E}[r_t|r_{t-1}\big], \tag{3.3}$$

$$\epsilon_t = R_t - R_{t-1} \tag{3.4}$$

$$\mathrm{Var}\big[R_t|R_{t-1}] = D(\theta)\mathrm{Var}\big[r_t|r_{t-1}\big]D(\theta)^T + v_t, \tag{3.5}$$

$$K_t = \mathrm{Var}\big[r_t|r_{t-1}\big]D(\theta)^T(\mathrm{Var}\big[R_t|R_{t-1}]\big)^{-1}, \tag{3.6}$$

$$\mathbb{E}[r_{t+1}|r_t\big] = \mathbb{E}[r_{t+1}|r_t\big] + K_t\epsilon_t, \tag{3.7}$$

$$\mathrm{Var}\big[r_{t+1}|r_t\big] = \mathrm{Var}\big[r_t|r_{t-1}\big] - K_tD(\theta)\mathrm{Var}\big[r_t|r_{t-1}\big], \tag{3.8}$$

[1]It is important to point out that if the parameters κ, μ and σ are not known then we cannot compute $A(\tau)$ or $B(\tau)$, because $A(\tau) = A(\tau, \kappa, \mu, \sigma)$ and $B(\tau) = B(\tau, \kappa, \mu, \sigma)$.

with K_t = the Kalman gain matrix at time t. Under the assumption that the errors, ϵ_t, are normally distributed, we can construct the Quasi Maximum logLikelihood estimator,

$$\ell(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln\left[\det\left(\mathrm{Var}\left[R_{t_i}|R_{t_i-1}\right]\right)\right] - \frac{1}{2}\epsilon_{t_i}\left(\mathrm{Var}\left[R_{t_i}|R_{t_i-1}\right]\right)^{-1}\epsilon_{t_i}^{T}, \quad (3.9)$$

with n = number of contracts and N = number of collected data for each contract. Therefore, the best estimator for θ is found by maximizing (3.9). We will do so using MATLAB optimization function *fmincon*. This functions finds a constrained minimum of a function of several variables specified by,

$$\min_x \ f(x) \ \text{ such that } = \begin{cases} c(x) \leq 0, & \text{(nonlinear constraint)} \\ ceq(x) = 0, & \text{(nonlinear constraint)} \\ A \cdot x \leq b, & \text{(linear constraint)} \\ Aeq \cdot x = beq, & \text{(linear constraint)} \\ lb \leq x \leq ub, & \text{(bounds)} \end{cases}$$

x, b, beq, lb, and ub are vectors, A and Aeq are matrices, $c(x)$ and $ceq(x)$ are functions that return vectors, and $f(x)$ is a function that returns a scalar. Furthermore, $f(x), c(x)$, and $ceq(x)$ can be nonlinear functions. We want to maximize the function in Equation (3.9), this is the same as minimizing $-\ell(\theta)$ in Equation (3.9).

Table 3.3 shows the estimated parameters for the CIR model and Table 3.4 the ρ_n values that correspond to them.

Parameter	κ	μ	σ
Kalman Filter	0.1990	0.0497	0.0354

Table 3.3: Estimated parameters, for the CIR model, of the historical data (from 1 Januray 2001 - 1 September 2011) for EUR.

Variance	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9
	0.003	0.001	0	0.001	0.003	0.004	0.004	0.005	0.006

Table 3.4: The variance of the independent error term v_t of Equation (2.6) for the CIR model.

Now we plot the yield curves. This is done by the same way as the Vasicek model, we first compute $R(0, \tau_i)$ for $i = 1, \ldots, 9$. This gives us,

$$
\begin{aligned}
R(0,\tau_1) &= 0.0478,\\
R(0,\tau_2) &= 0.0473,\\
R(0,\tau_3) &= 0.0460,\\
R(0,\tau_4) &= 0.0444,\\
R(0,\tau_5) &= 0.0455,\\
R(0,\tau_6) &= 0.0484,\\
R(0,\tau_7) &= 0.0527,\\
R(0,\tau_8) &= 0.0568,\\
R(0,\tau_9) &= 0.0572.
\end{aligned}
$$

We plotted all these yield curves in Figures 3.20, up to, 3.25. In Section 3.4 we will give a plot of the bond price comparing this with the Vasicek model.

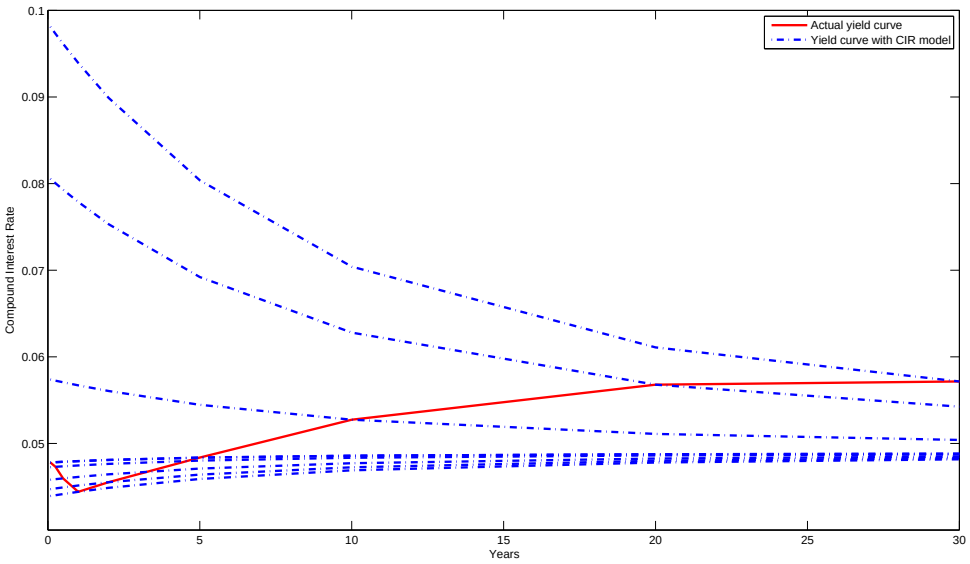

Figure 3.20: A plot of the actual initial yield curve, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

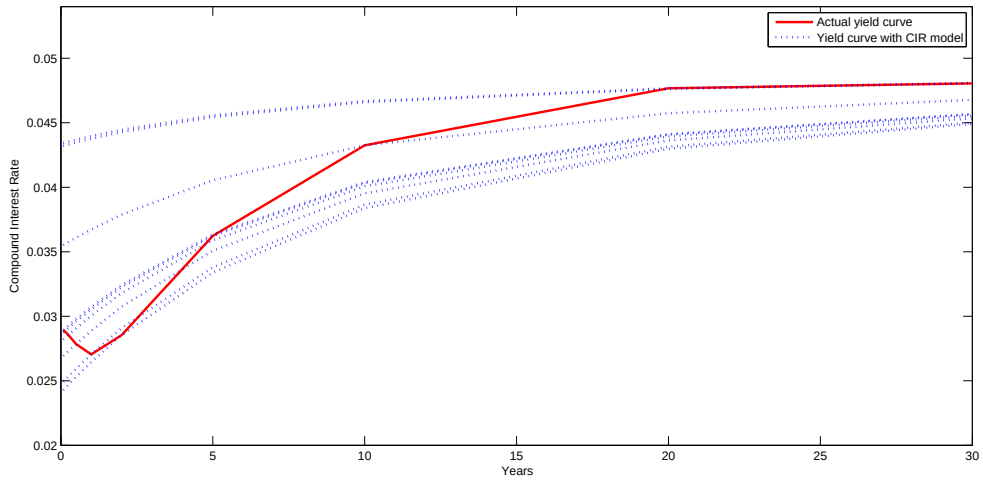

Figure 3.21: A plot of the actual yield curve on 01/01/2003, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

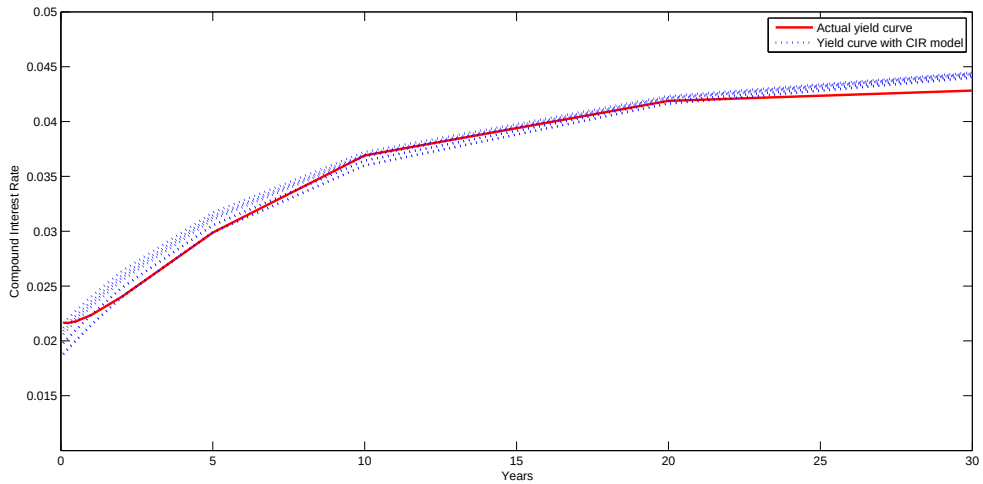

Figure 3.22: A plot of the actual yield curve on 03/01/2005, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

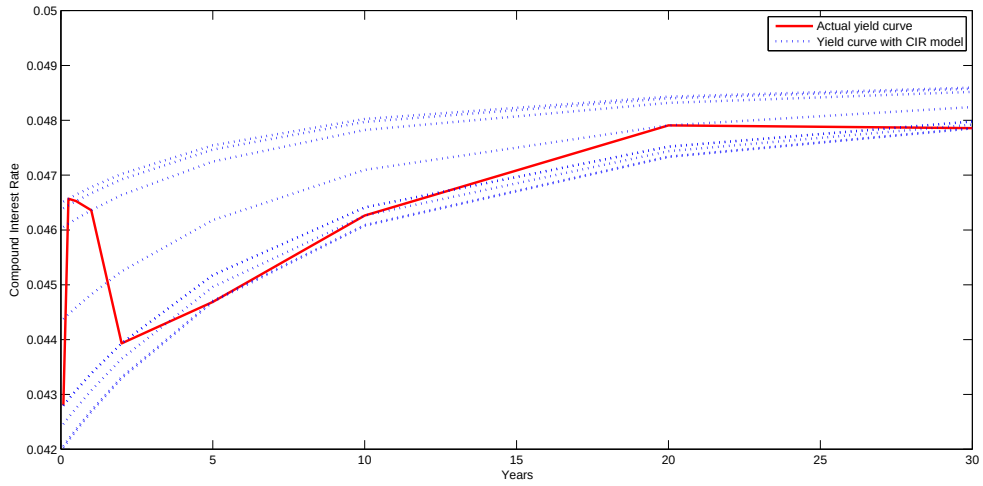

Figure 3.23: A plot of the actual yield curve 01/01/2008, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

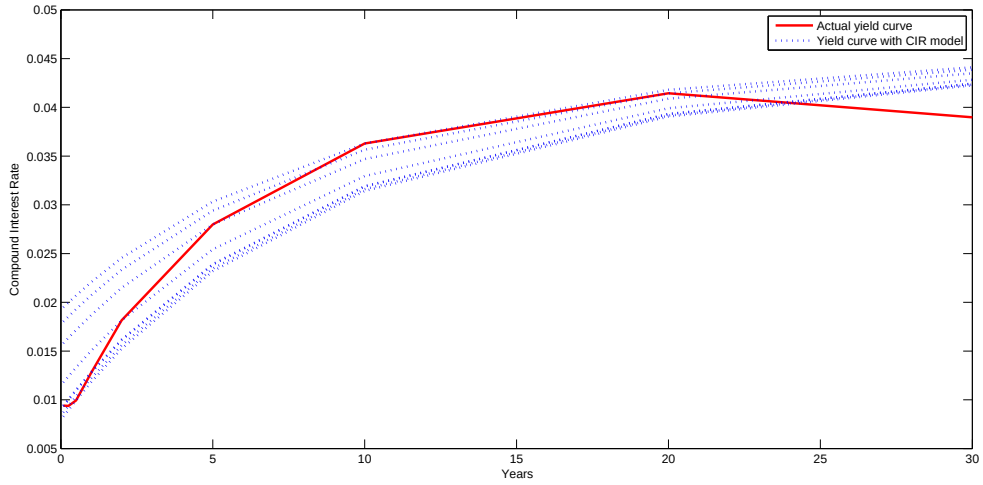

Figure 3.24: A plot of the actual yield curve 01/01/2010, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

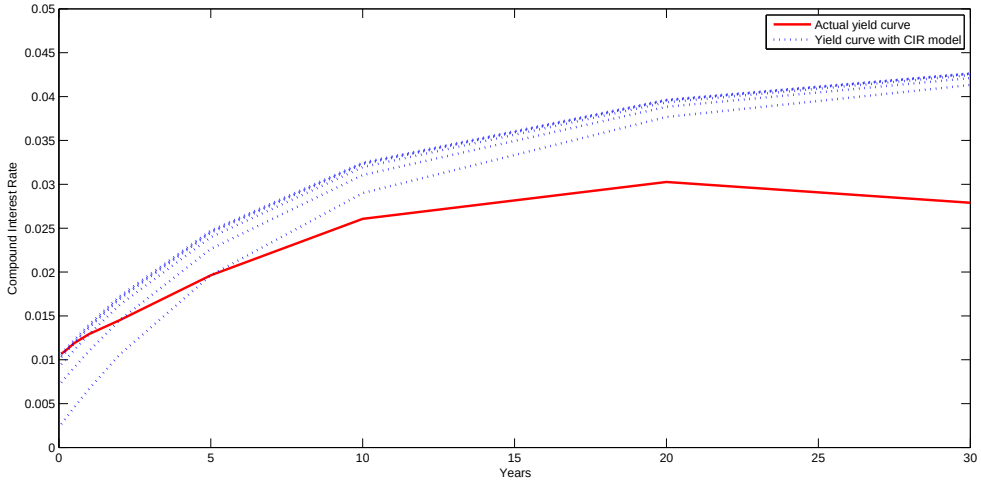

Figure 3.25: A plot of the actual yield curve 03/01/2011, for the CIR model, against estimated yield curves, using the Kalman Filter (KF).

3.3 Modeling data with the Nelson-Siegel model

The Nelson-Siegel model is the easiest to calibrate, as described in Chapter 1. We have

$$R_t(\tau) = \beta_{1,t}\kappa_1 + \beta_{2,t}\kappa_2 + \beta_{3,t}\kappa_3, \tag{3.10}$$

with

$$\begin{aligned} \kappa_1 &= 1, \\ \kappa_2 &= \frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}, \\ \kappa_3 &= \frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}. \end{aligned}$$

Recall, λ_t is fixed. Therefore, we first compute λ_t using MATLAB (optimization function *fmincon*) and get $\lambda_t = 0.6807$. This now allows us to estimate $\beta_{1,t}, \beta_{2,t}, \beta_{3,t}$, for each time step, see Figures 3.26, 3.27, 3.28. We define $\hat{\beta}_{1,t}$, $\hat{\beta}_{2,t}$ and $\hat{\beta}_{3,t}$ as the average value of $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$ respectively, see Table 3.5.

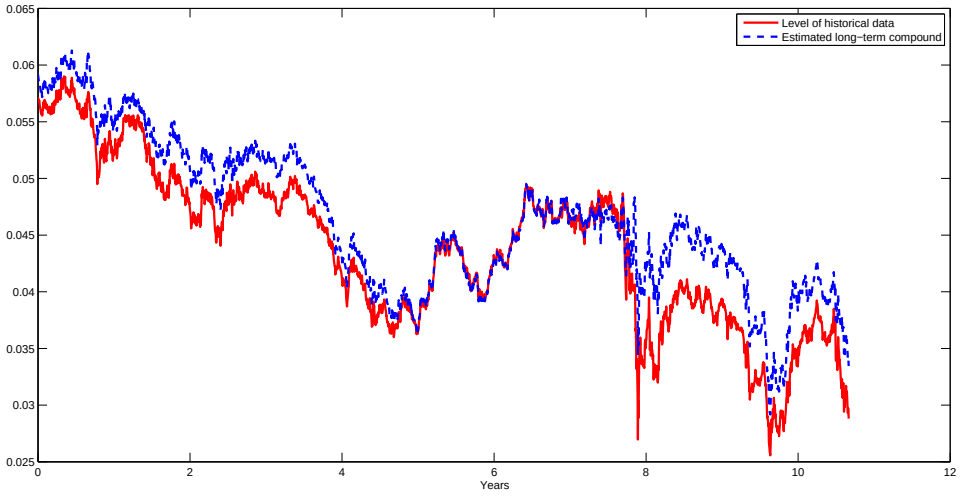

Figure 3.26: A plot of the yield curve level and the parameters $\beta_{1,t}$ of the Nelson-Siegel model.

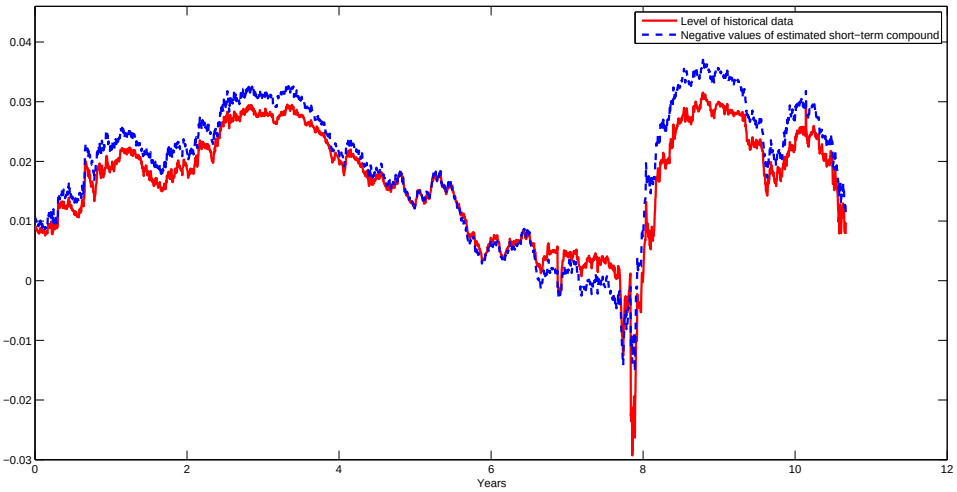

Figure 3.27: A plot of the yield curve slope and the parameters $-\beta_{2,t}$ of the Nelson-Siegel model.

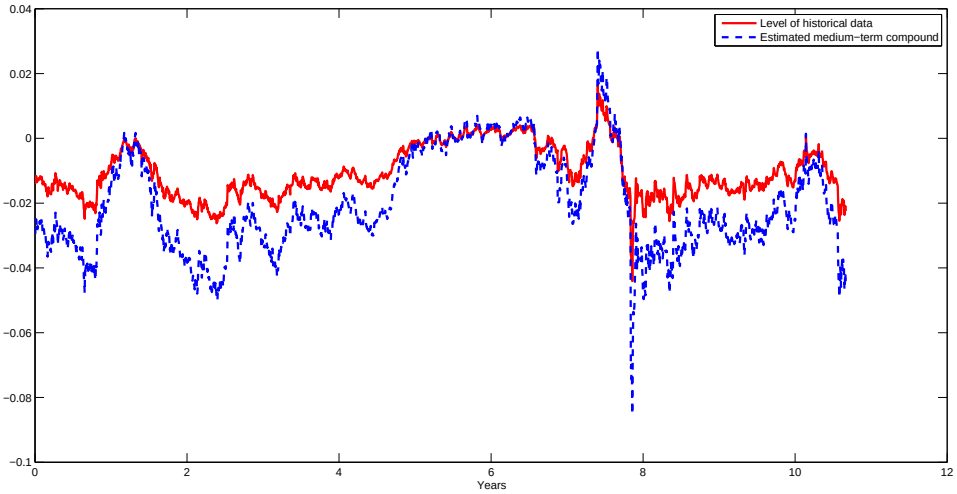

Figure 3.28: A plot of the yield curve curvature and the parameters $\beta_{3,t}$ of the Nelson-Siegel model.

Parameter	$\hat{\beta}_{1,t}$	$\hat{\beta}_{2,t}$	$\hat{\beta}_{3,t}$
	0.0463	-0.0187	-0.0203

Table 3.5: Estimated parameters of the Nelson-Siegel model of the historical data (from 1 Januray 2001 - 1 September 2011) for EUR.

Next, we want to plot the initial yield curve. We do so using the initial values of $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$, see Table 3.6. See Figure 3.29 for a plot of the yield curve. We can see that the actual initial yield curve is estimated perfectly with the Nelson-Siegel model. This is clearly the reason why this model is so popular by the majority of the Central Banks.

Parameter	corr$[\beta_{1,t}, l_t]$	corr$[\beta_{2,t}, s_t]$	corr$[\beta_{1,t}, l_t]$
	0.9598	-0.9755	0.9946

Table 3.6: Initial estimated parameters of the Nelson-Siegel model of historical data (from 1 Januray 2001 - 1 September 2011) for EUR.

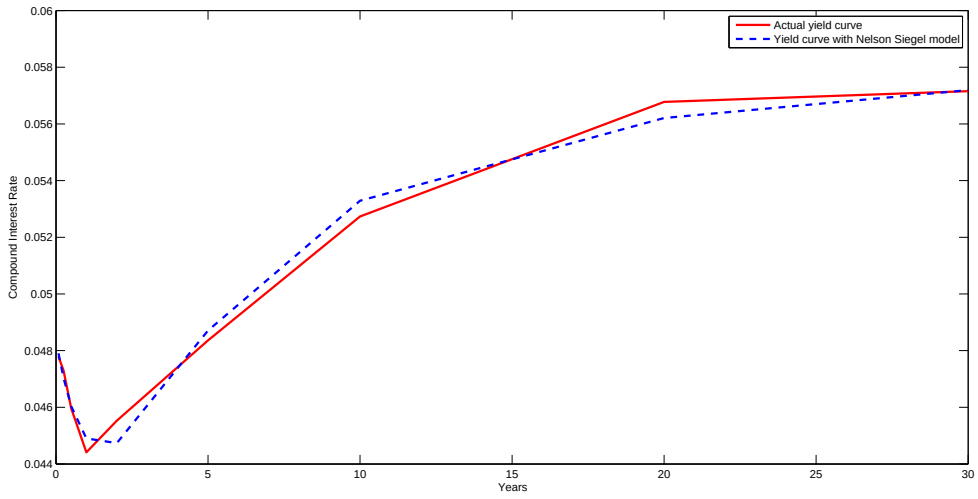

Figure 3.29: A plot of the actual initial yield curve level and the estimated yield curve of the Nelson-Siegel model.

3.4 Comparing the Results

Now that we have estimated the parameters for the Vasicek model, the CIR model and the Nelson-Siegel model. We will give a plot of the yield curves to compare these models. For the Vasicek model we will use the estimated values computed with the Kalman filter. See Figure 3.30 for the result. Note, that the Vasicek model and the CIR model both give the same initial yield curve. However, the CIR model estimates the yield curve better when different time steps are taken into account, as we have seen in Section 3.2. Plus the fact that in the CIR model the short rate values cannot turn negative, it is better to use this model.

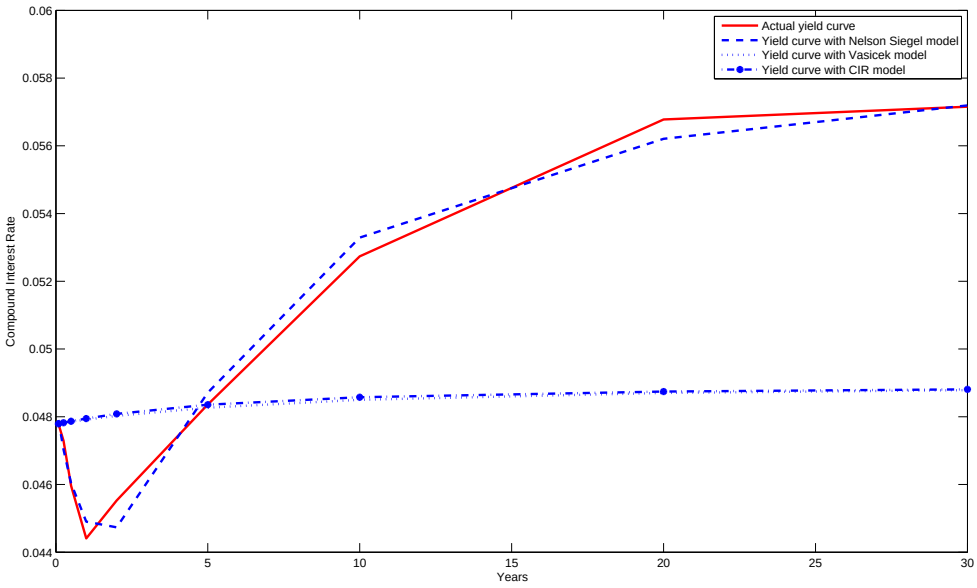

Figure 3.30: A plot of the actual initial yield curve and the estimated initial yield curves with the Vasicek model, the CIR model and the Nelson-Siegel model.

Next, we compare the estimated bond price for the Vasicek model and CIR model for a duration of one month, one year and ten years. We do this using 1000 Monte Carlo simulations. See Figure 3.31, 3.32 and 3.33 for the results. We can see that the bond price of the Vasicek model is more like a straight line, because the interest rates are normally distributed and the estimated σ is very small, looking at the historical data this is highly unlikely, therefore, the values remain around the initial value. Whereas, the bond price for the CIR model decreases over time this is due the fact that the estimated σ here is relatively much larger, more realistic looking at the historical data. Therefore, we can conclude that the CIR model is as expected a better model to describe the interest rate.

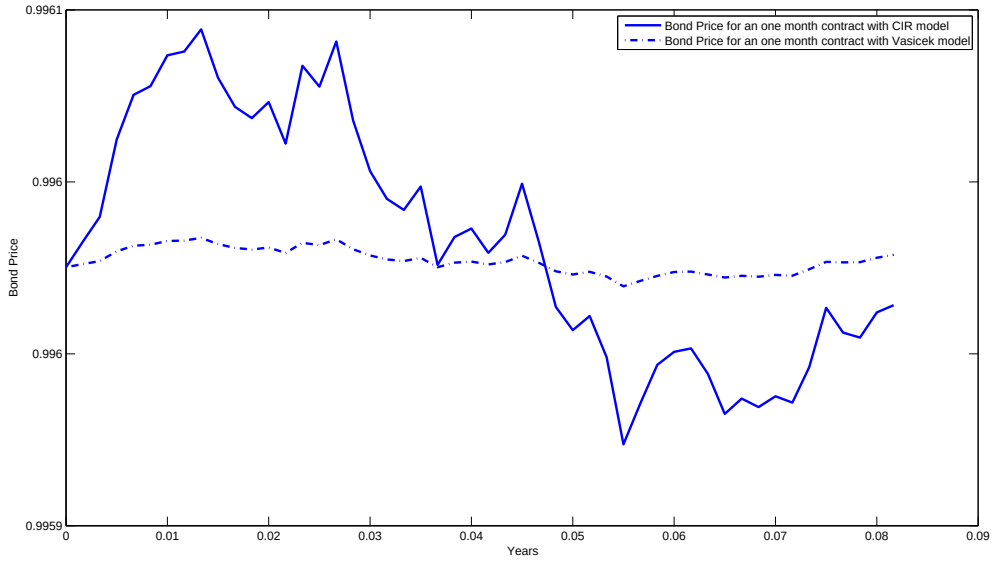

Figure 3.31: A plot of the bond price with the Vasicek model and the CIR model for a duration of 1 month.

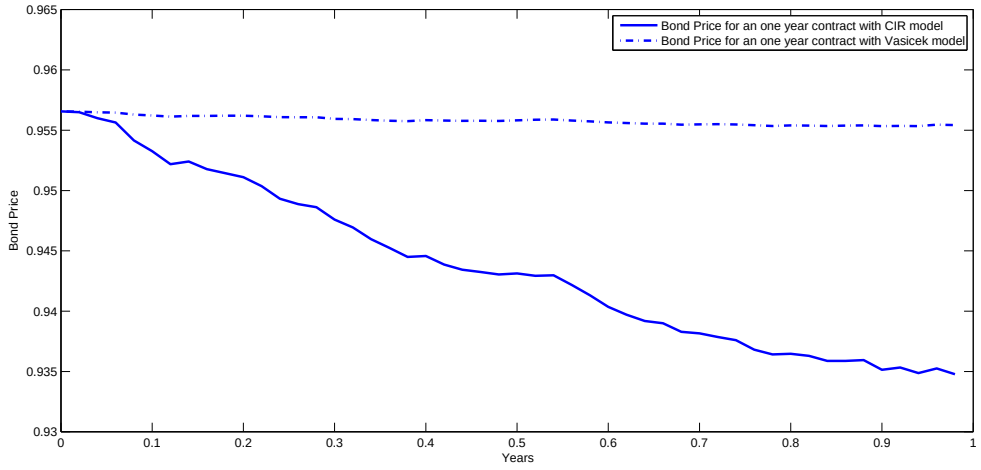

Figure 3.32: A plot of the bond price with the Vasicek model and the CIR model for a duration of 1 year.

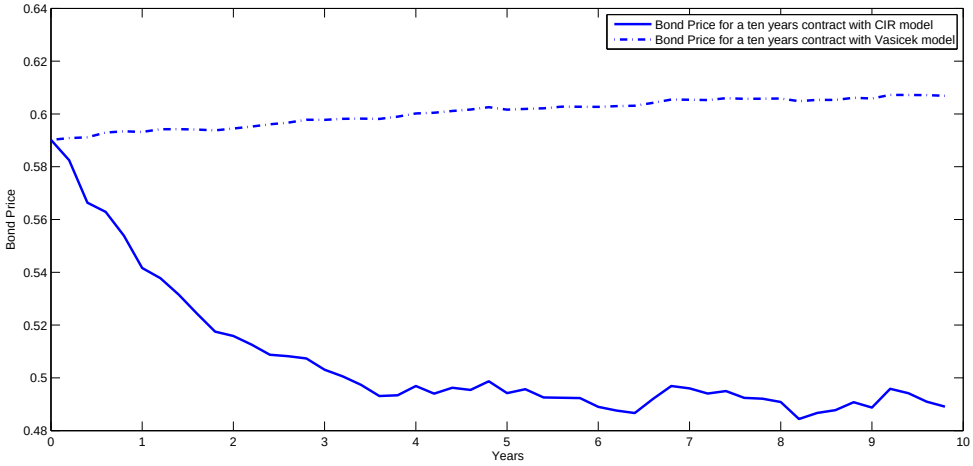

Figure 3.33: A plot of the bond price with the Vasicek model and the CIR model for a duration of 10 years.

Conclusions and Further Research

Conclusions

We have shown that calibration with MLE and LSM results in a huge bias. In the past years there have been methods developed to reduce the bias. Methods such as jackknife, bootstrap combined with MLE and LTQ, all are good alternatives as they reduce the bias with a factor between 5 and 10. However, the bias on one of the drift parameters still remains high, around 60%. We have used the Kalman filter to eliminate the bias and from this we can conclude that the Kalman filter outperforms LSM, MLE and LTQ. The strength in the Kalman filter lays in the way the method handles the collected historical data for different contracts as we have shown in Chapter 2. Therefore, we recommand Rabobank to apply the Kalman filter for calibration when dealing with short rate models, such as the Vasicek model or the CIR model (or extended versions of these models).

We have shown that short rate models, such as the Vasicek model and the CIR model, fit the initial yield curve poorly. This is due to the fact that we have 9 different initial values from the historical data and we can only apply one initial value at time zero. This results in 9 different yield curves, each fitting the yield curve poorly. However, unlike the Vasicek model, the CIR model estimates the yield curve better at different time steps as shown in Section 3.2. The Nelson-Siegel model on the other hand, fits all the yield curves perfectly. However, this raises another concern as the arbitrage rule is neglected.

The CIR model estimates the yield curves relatively much better compared to the Vasicek model. Plus the fact that in the CIR model the short rate values cannot turn negative, it is better to use this model. Even though the Nelson-Siegel model fits the yield curves perfectly, it does allow arbitrage and, therefore, must be ignored. Thus, we conclude that the CIR model is more preferable when dealing with interest rate.

Further Research

For Further research we would like to recommand Rabobank to look at the two factor CIR model as defined in [18]. We expect that this results in a more accurate estimation of the yield curves. The problem that will arise is that the partial differential equation of the bond price does not have an exact solution. However, the Kalman filter can still be applied for calibration.

In this thesis we have given a brief description of the Nelson-Siegel model. This model gives a good fit of the yield curves, however, the arbitrage rule is not obeyed. Diebold and Li have recently expanded the Nelson-Siegel model to obey the no arbitrage rule. This resulted

in a poor fit of the yield curves. The choice is really clear, one can have a perfect fit of the yield curves with the arbitrage rule obeyed or a poor fit of the yield curve with the arbitrage rule upheld. Therefore, we strongly recommend Rababank to do some research on this model and see how the arbitrage rule can be obeyed while maintaining a perfect fit of the yield curves.

Bibliography

[1] Aït-Sahalia Y., 2008. *Closed-form likelihood expansions for multivariate diffusions.* Annals of Statistics 36, 906-937.

[2] Ball C. and Torous W., 1996. *Unit roots and the estimation of interest rate dynamics.* Journal of Empirical Finance 3, 215-238.

[3] Bank for International Settlements (BIS), 2005. *Zero-coupon yield curves: technical documentation.* Papers, No. 25, Basel.

[4] Brigo D. and Mercurio F., 2006. *Interest Rate Models: Theory and Practice with Smile, Inflation and Credit*, Heidelberg, Springer Verlag, 2nd Edition.

[5] Cox, J., Ingersoll J. and S. Ross, 1985. *A theory of the term structure of interest rates.* Econometrica, Vol. 53, 385-407.

[6] Diebold F.X. and Li C., 2006. *Forecasting the Term Structure of Government Bond Yields.* Journal of Econometrics, 130, 337-364.

[7] Duan J.C. and Simonato J.G., 1995. *Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter," Discussion paper*, Centre universitaire de recherche et analyze des organizations (CIRANO).

[8] Durham G. and A. R. Gallant, 2002. *Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes.* Journal of Business and Economic Statistics, 20, 297-316.

[9] Elen E.A.L.J. van, 2010. *Term structure forecasting.* Rabobank.

[10] Haberman R., 2003. *Applied Partial Differential Equations.* Vol. 4.

[11] Hull J. and White A., 1987. *The Pricing of Options with Stochastic Volatilities.* Journal of Finance, Vol.42, 281-300.

[12] Kalman R.E., 1960. *A new approach to linear filtering and prediction theory.* Journal of Basic Engineering, Transactions ASME, Series D.

[13] Lo A., 1988. *Maximum likelihood estimation of generalized Ito processes with discretely sampled data.* Econometric Theory 4, 231-247.

[14] Merton R.C., 1980. *On estimating the expected return on the market: An exploratory investigation.* Journal of Financial Economics 8, 323-361.

[15] Miller J., Edelman D. and Appleby J., 2007. *Numerical Methods for Finance.* Chapman & Hall/CRC Financial Mathematics Series.

[16] Nelson C.R. and A.F. Siegel, 1987. *Parsimonious Modeling Of Yield Curves.* Journal of Business, 60, 473-489.

[17] Ruijter M.C.A. de, 2010. *An interest rate model for counterparty credit risk.* Rabobank.

[18] Shreve S., 2004. *Stochastic Calculus for Finance II*, Vol. 2.

[19] Seydel R., 2003. *Tools for Computational Finance*, 3th edition, Springer Verlag, Berlin.

[20] Tang C.Y and Chenb S.X, 2009. *Parameter estimation and bias correction for diffusion processes.*, Journal of Econometrics 149, (2009) 65-81.

[21] Vasicek O., 1977. *An equilibrium characterization of the term structure.* Journal of Financial Economics, Vol. 5, 177-186.

[22] Welcha G. and Bishop G., 2001. *An Introduction to the Kalman Filter.*, `http://www.cs.unc.edu/~tracker/media/pdf/SIGGRAPH2001_CoursePack_08.pdf`.

[23] Yu J. and Phillips P., 2001. *A Gaussian approach for estimating continuous time models of short term interest rates.* The Econometrics Journal 4, 211-225.

[24] Yu J. and Phillips P., 2006. *Jackknifing bond option prices. Review of Financial Studies*, Vol. 18, 707-742.