



On Calibrating an Extension of the Chen Model

MSc Degree Project in Mathematical Finance

Martin Möllberg

SF299X Degree Project in Mathematical Statistics, Second
Cycle

THE DEPARTMENT OF MATHEMATICS AT KTH

March 4, 2015

Abstract

There are many ways of modeling stochastic processes of short-term interest rates. One way is to use one-factor models which may be easy to use and easy to calibrate. Another way is to use a three-factor model in the strive for a higher degree of congruency with real world market data. Calibrating such models may however take much more effort. One of the main questions here is which models will be better fit to the data in question. Another question is if the use of a three-factor model can result in better fitting compared to one-factor models.

This is investigated by using the Efficient Method of Moments to calibrate a three-factor model with a Lévy process. This model is an extension of the Chen Model. The calibration is done with Euribor 6-month interest rates and these rates are also used with the Vasicek and Cox-Ingersoll-Ross (CIR) models. These two models are calibrated by using Maximum Likelihood Estimation and they are one-factor models. Chi-square goodness-of-fit tests are also performed for all models.

The findings indicate that the Vasicek and CIR models fail to describe the stochastic process of the Euribor 6-month rate. However, the result from the goodness-of-fit test of the three-factor model gives support for that model.

Acknowledgements

First of all, the author of this report wants to thank Boualem Djehiche and Henrik Shahgholian who supervised the project. I want to thank them for their guidance and for taking their time. I want to thank them for their support, their advice and fruitful discussions. They kindly shared their knowledge and for that I am grateful. Boualem Djehiche shared important insights about Mathematical Statistics and Estimation Theory (and its applications) while supervising. Henrik Shahgholian initially supervised this project and his ideas were very important for the initiation of it.

There are more staff members at the department of Mathematics who gave advice and support during this project. I also want to thank them for this and for taking their time, they are Alexander Aurell, Bengt Lindberg and Johan Karlsson.

I also want to thank Krister Alvelius and Henrik Segerberg who work at Nordea, they also gave me some advice for this Master Thesis Project. Earlier I worked together with them on another project.

Lastly, I want to thank Henrik Eriksson who gave lectures in Calculus to us students at KTH. He is a dedicated teacher and lecturer and he encouraged me to continue studying Mathematics while I was an undergraduate student.

With Gratitude,

Martin Möllberg

Contents

1	Introduction	5
2	Theory & Background	7
2.1	The Vasicek Model	7
2.2	The CIR model	8
2.3	A Goodness-of-fit test for SDE models	10
2.4	Lévy processes	13
2.5	The Efficient Method of Moments	13
2.6	The extended Short Rate Model by Andersen et al.	16
2.7	Estimating the covariance matrix	19
2.8	Conversion of rates	22
3	Implementation	23
3.1	Processing the historical data	23
3.2	Calibration of the Vasicek model and chi-square testing	24
3.3	Calibration of the CIR model and chi-square testing	24
3.4	Modeling the Poisson process	25
3.5	Smoothing of the rates	27
3.6	The first step of the EMM procedure	27
3.7	The second step of the EMM procedure	30
4	Results	31
4.1	Calibration of the Vasicek model and chi-square testing	31
4.2	Calibration of the CIR model and chi-square testing	32
4.3	The first step of the EMM procedure	32
4.4	The estimated covariance matrix	35
4.5	The second step of the EMM procedure	35
5	Discussion	37
5.1	The estimated covariance matrix	37
5.2	The implementation	37
5.3	Comparing the models	39
6	Conclusions	42
	Appendices	43

1 Introduction

During the last few decades multi-factor Short-Term Interest Rate Models have emerged. The purpose of their development has been to provide us with 'rich' models to capture certain characteristics of the interest rate that we can observe. Such characteristics include mean reversion, time-dependent volatility and discontinuities and models that account for mean reversion have been around for many years. The Vasicek model was the first one to capture this characteristic and it is a one-factor model. However, this model has its limitations as it is theoretically possible to obtain negative values which would make no sense. Also the volatility is a constant in this model and some may wish to use a model with time-dependent volatility instead.

Another model that deals with this characteristic is the Cox-Ingersoll-Ross (CIR) model (one-factor) which is an extension of the Vasicek model. The level dependent volatility in this model means that it is less likely that negative values of the rate will be obtained. However, some would find more complexity well justified in the strive for a higher degree of congruency with real world market data. Therefore a model with stochastic volatility may be used instead.

A model that may be viewed as a generalization of the CIR model is the Chen model, the latter is a three-factor model with stochastic mean and stochastic volatility. Indeed it is a model that captures much of the complexity of the real world but there is still one characteristic that it does not account for – discontinuities.

Therefore it would seem quite natural to introduce a Lévy process in a Short Rate Model as a next step. And this is what Andersen, Benzoni and Lund [2] in fact did as they extended the Chen model. In their paper they describe how they calibrate the model by using the Efficient Method of Moments (EMM) and weekly US 3-month Treasury-Bill rates. They also describe how the use of this model means an improvement compared to one-factor models for instance.

In this report a project is described where the model presented by Andersen et al. is calibrated by using Euribor 6-month rates. This model is in Section 2.6 referred to simply as the 'extended Short Rate Model' for the sake of brevity. The Vasicek and CIR models are also calibrated and then some chi-square tests are performed for goodness-of-fit testing. The intention is to try to shed more light on any possible improvements made when using a three-factor model instead of a one-factor model.

Against the background of this description of the project the problem formulation is as follows:

1. Calibrate the Vasicek model by using Euribor 6-month rates and Maximum Likelihood Estimation.
2. Calibrate the CIR model by using the same rates and Maximum Likelihood Estimation.
3. Calibrate the extended Short Rate Model presented by Andersen et al. [2] by using the same rates and EMM.
4. Perform goodness-of-fit (chi-square) tests with the models mentioned above.

The investigations described in this report are made to see if it would be possible to find support for the findings of Andersen et al. [2]. In their report the p-values indicate that the extended Short Rate Model is a better fit to the observations than the CIR model (for instance).

Next follows a section where the theory to be applied is described, i.e. theory for solving the problems stated above. The implementation of this project is also described in this report and the results are presented. Lastly the results are discussed and a presentation of the conclusions drawn is given.

2 Theory & Background

2.1 The Vasicek Model

The Vasicek model is explained in a book by Hull [7]. Under this model the risk-neutral process for r is given by the following Stochastic Differential Equation (SDE):

$$dr_t = a(b - r_t)dt + \sigma dW_t.$$

This model has been named after the Czech mathematician Oldřich Vašíček (born in 1942) who founded it. This is also the SDE of an Ornstein-Uhlenbeck process. The mean reversion of this model is incorporated by letting the drift term be dependent on r in the above equation. Here a , b and σ are constants and W_t is a Wiener process (σ is the volatility). These constants may be estimated by using Maximum Likelihood Estimation (MLE), this procedure is often referred to as *calibration*.

By applying the theory of Stochastic Calculus the SDE above may be explicitly solved. This leads to

$$S_{i+1} = S_i e^{-\lambda\delta} + \mu(1 - e^{-\lambda\delta}) + \sigma \sqrt{\frac{1 - e^{-2\lambda\delta}}{2\lambda}} N(0, 1).$$

The density function of a normally distributed variable (with mean μ and the standard deviation equal to σ) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right].$$

Therefore the conditional probability density function of an observation S_{i+1} – given a previous observation S_i – is given by

$$f(S_{i+1}|S_i; \mu, \lambda, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \exp\left[-\frac{(S_i - S_{i-1}e^{-\lambda\delta} - \mu(1 - e^{-\lambda\delta}))^2}{2\hat{\sigma}^2}\right],$$
$$\hat{\sigma}^2 = \sigma^2 \frac{1 - e^{-2\lambda\delta}}{2\lambda}.$$

Now the log-likelihood function reads:

$$\begin{aligned} \mathcal{L}(\mu, \lambda, \hat{\sigma}) &= \sum_{i=1}^n \ln f(S_i S_{i-1}; \mu, \lambda, \sigma) = \\ &= -\frac{n}{2} \ln(2\pi) - n \ln(\hat{\sigma}) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n [S_i - S_{i-1} e^{-\lambda\delta} - \mu(1 - e^{-\lambda\delta})]^2. \end{aligned}$$

This function shall be maximized during the calibration. Given the estimated parameters and a discretization of the SDE the short rates may be simulated. By using an Euler scheme the equation becomes:

$$r_{i+1} = r_i + a(b - r_i)\Delta t_{i+1} + \sigma x_{i+1} \sqrt{\Delta t_{i+1}}.$$

Here the variable x_i is a random variable with standard normal distribution. In Section 4.1 the result from the calibration is given and in the same section the result of a chi-square test may be found.

2.2 The CIR model

The model to be described in this section is the Cox-Ingersoll-Ross model and this model was named after the economists who founded it. The model is given by the following SDE and this may also be found in the book by Hull [7]:

$$dr_t = a(b - r_t)dt + \sigma\sqrt{r_t}dW_t.$$

This looks pretty much like the SDE of the Ornstein-Uhlenbeck process but the difference is that $\sqrt{r_t}$ appears as a factor in the martingale part of the equation. This implies that the volatility is level dependent, just as stated in the introduction of this report. The stochastic process in this case is referred to as a 'CIR process' and it also incorporates mean reversion, since the drift term is of the same kind as in the Ornstein-Uhlenbeck SDE.

Modeling with level dependent volatility may be advantageous compared to the case with the Vasicek model. The reason for this is that if the value of the rate would decrease to a number close to zero then there would not be much

white noise. This is so since the factor $\sqrt{r_t}$ appears in the martingale part in the SDE, therefore the variance of the Itô integral becomes small for small rates. Consequently the evolution of the rate becomes dominated by the drift factor when the value of the rate is small enough. Typically the values of the parameters imply that the drift factor is positive for small rates, hence the value of the rate will increase instead of decrease to a negative number.

It is important that the rates never become negative during numerical computations using an Euler scheme. The reason is that if the rate would become negative during an iteration then the value in the next iteration will become a complex number – the reason for this is the factor $\sqrt{r_t}$ in the martingale part – which would not make any sense. There is a certain condition that should be imposed on the parameters during calibration and this condition is known as the Feller condition:

$$2ab > \sigma^2. \quad (1)$$

This condition is also imposed during the calibrations in this project. The SDE of the CIR process may be discretized and this leads to:

$$r_{i+1} = r_i + a(b - r_i)\Delta t_{i+1} + \sigma\sqrt{r_i}x_{i+1}\sqrt{\Delta t_{i+1}}.$$

This is obtained by using an Euler scheme and this result is also used for simulations. Again, x_i is standard normally distributed.

Just as MLE may be used for calibrating the Vasicek model this method may also be used for calibrating the CIR model. In order to do so the log-likelihood function must be derived first, in an article by Cox, Ingersoll and Ross [3] the following probability density function is given:

$$f(r_s; r_t) = c \cdot e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}),$$

$$c = \frac{2a}{\sigma^2(1 - \exp[-a(s-t)])},$$

$$u = c \cdot r_t \exp[-a(s-t)],$$

$$v = c \cdot r_s,$$

$$q = \frac{2ab}{\sigma^2} - 1.$$

This is the density function of the interest rate at time s conditional on its value at the current time t . Here I_q is the modified Bessel function of the first kind of order q . Now the log-likelihood may be derived by the following steps, first of all the likelihood function for a series of interest rates with N observations is

$$L(\theta) = \prod_{i=1}^{N-1} f(r_{i+1}|r_i).$$

Then, the log-likelihood function is given by

$$\ln L(\theta) = \sum_{i=1}^{N-1} \ln f(r_{i+1}|r_i).$$

The following analytical expression for the log-likelihood function of the CIR process may now be derived:

$$\begin{aligned} \ln L(\theta) = & (N - 1)\ln(c) + \\ & + \sum_{i=1}^{N-1} \left\{ -u_i - v_{i+1} + 0.5q \cdot \ln \left(\frac{v_{i+1}}{u_i} \right) + \ln(I_q(2\sqrt{u_i v_{i+1}})) \right\}. \end{aligned}$$

This expression is needed for the MLE.

CIR processes are not only used for the CIR model, one such example is the Heston volatility model which is used for modeling stochastic volatility. It will appear later in Section 2.6 when the extended Short Rate Model will be studied.

2.3 A Goodness-of-fit test for SDE models

In this section a chi-square (goodness-of-fit) test for one-factor SDE models is presented, based on Allen [4]. We assume that a collection of data (observations) is available and this goodness-of-fit test is used to test if there is a lack-of-fit between the SDE model and this data.

Suppose that the observations are given at times t_0, t_1, \dots, t_{N-1} where $t_i = i\Delta t$ for a constant $\Delta t > 0$. The N observations of the process are also given: x_0, x_1, \dots, x_{N-1} . Suppose that the SDE is

$$dX(t) = f(t, X)dt + g(t, X)dW(t).$$

In the goodness-of-fit procedure M simulations are performed under the SDE model. These are calculated from time t_{i-1} until time t_i starting at x_{i-1} . Using an Euler scheme with K steps leads to

$$\begin{aligned} X_{j+1,i}^{(m)} &= X_{j,i}^{(m)} + f\left(t_{i-1} + \frac{j\Delta t}{K}, X_{j,i}^{(m)}\right) \frac{\Delta t}{K} \\ &\quad + g\left(t_{i-1} + \frac{j\Delta t}{K}, X_{j,i}^{(m)}\right) \sqrt{\frac{\Delta t}{K}} \eta_{j,i}^{(m)}, \end{aligned}$$

for $j = 0, 1, \dots, K - 1$ and $m = 1, 2, \dots, M$ with $X_{0,i}^{(m)} = x_{i-1}$ and where $\eta_{j,i}^{(m)}$ is standard normally distributed for each i, j and m .

Let $X_i^{(m)} = X_{K,i}^{(m)}$ be the m th simulated value at t_i for $m = 1, 2, \dots, M$ and for $i = 1, 2, \dots, N - 1$. Now define

$$s_i^{(m)} = \begin{cases} 1, & \text{if } x_i \geq X_i^{(m)}, \\ 0, & \text{if } x_i < X_i^{(m)}. \end{cases}$$

Further, let

$$z_i = 1 + \sum_{m=1}^M s_i^{(m)} \quad \text{for } i = 1, 2, \dots, N - 1.$$

Then, z_i is the rank of value x_i as compared with the M simulated values $X_i^{(m)}$, $1 \leq m \leq M$ for $i = 1, 2, \dots, N - 1$. It should be noted that $1 \leq z_i \leq M + 1$ for $i = 1, 2, \dots, N - 1$.

The null hypothesis is that the SDE model describes the stochastic process. Under this null hypothesis the ranks z_i have equally likely values between 1

and $M + 1$. A chi-square goodness-of-fit test is used to test this hypothesis. To perform the test the observed and expected frequencies are needed. Let

$$I_{i,q} = \begin{cases} 1, & \text{if } z_i = q, \\ 0, & \text{if } z_i \neq q, \end{cases}$$

for $i = 1, 2, \dots, N - 1$ and

$$\Omega(q) = \sum_{i=1}^{N-1} I_{i,q} \quad \text{for } q = 1, 2, \dots, M + 1.$$

Here $\Omega(q)$ is the observed frequency that the rank equals the value q , therefore

$$\sum_{q=1}^{M+1} \Omega(q) = N - 1.$$

The expected frequency under the null hypothesis is $(N - 1)/(M + 1)$. The test statistic is

$$Q_M = \sum_{q=1}^{M+1} \frac{\left(\Omega(q) - \frac{N-1}{M+1}\right)^2}{\left(\frac{N-1}{M+1}\right)}.$$

Under the null hypothesis this statistic is approximately distributed as a chi-square random variable with M degrees of freedom. A large value of this test statistic indicates a lack-of-fit between the stochastic model and the data. Specifically, if $P(\chi^2(M) \geq Q_M)$ is smaller than a preset level of significance then the null hypothesis is rejected, indicating a lack-of-fit of the SDE model with the observed data. However, the chi-square approximation fails if the expected frequencies under the null hypothesis are too small. A rule-of-thumb often applied is that the expected frequencies should be no less than 5. Applying this rule gives

$$\frac{N - 1}{M + 1} \geq 5 \quad \text{i.e.} \quad M \leq \frac{N - 6}{5}.$$

In Section 4 the results from performing this chi-square test are presented for the Vasicek and CIR models.

2.4 Lévy processes

In this section some of the essentials of Lévy processes will be described. This is fundamental since it is a constituent of the extended Short Rate Model which is one of the subjects of study. Lévy processes were named in honor of the French mathematician Paul Lévy (1886-1971) and Tankov [5] describes him as one of the founding fathers of the modern theory of stochastic processes.

Lévy processes in full may be described as something more general than the Brownian motion with drift, since Lévy processes have discontinuous paths. Therefore, simply thinking of a random walk with the discontinuities as an additional feature may help in order to get a better grasp of this concept. Tankov describes them as simple enough to study and rich enough for applications at the same time, or at least to be used as building blocks for more realistic models. This is in fact what Andersen et al. [2] used them for in their extended Short Rate Model.

The mathematical details are deferred to the appendix.

2.5 The Efficient Method of Moments

The Efficient Method of Moments (EMM) is a very important method, it will be used to calibrate the extended Short Rate Model. As stated in Andersen, Chung and Sørensen [8] this method was introduced by Bansal et al. and Gallant and Tauchen. In that paper EMM is described as a Method of Moments procedure that often provides a viable approach to estimation when Maximum Likelihood Estimation is computationally intensive or infeasible. Some may resort to simulation-based approximations to the likelihood or avoid direct dependence on the likelihood altogether. The latter applies to EMM but it still seeks to mimic the efficiency of likelihood-based inference. Chung et al. state that the key insight is that a careful selection of moment conditions – guided by the characteristics of the observed data – will allow for efficient estimation via a standard Generalized Method of Moments (GMM) procedure. EMM is a variant of simulation-based GMM.

Further, Chung et al. state that Maximum Likelihood Estimation itself may

be interpreted as a Method of Moments procedure with the derivative of the log-likelihood function, *the score vector*, providing the (exactly identifying) moment conditions. EMM involves using an auxiliary model – or *score generator* – that allows for a closed-form expression for the associated (quasi-) score vector. This auxiliary model is given by a conditional density parameterized by the auxiliary parameter vector ξ . The conditional density may be expressed as $f(y_t|Y_{t-1}, \xi)$ where $Y_{t-1} = \{y_{t-1}, \dots, y_1\}$ are referred to as *lagged endogenous variables*.

The first step in the EMM procedure is to estimate ξ by Quasi-Maximum Likelihood Estimation (QMLE). And this ensures that the QML estimator $\hat{\xi}_T$ satisfies the associated first-order conditions:

$$\frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \xi} \ln f(y_t|Y_{t-1}, \hat{\xi}_T) = \frac{1}{T} \sum_{t=1}^T s_f(Y_t, \hat{\xi}_T) = 0. \quad (2)$$

Here $s_f(Y_t, \hat{\xi}_T) = (\partial/\partial \xi) \ln f(y_t|Y_{t-1}, \hat{\xi}_T)$ denotes the quasi-score function. Even if the auxiliary model would be misspecified standard QML theory implies that (under suitable regularity) $\hat{\xi}_T \rightarrow \xi_0$ where the limiting value ξ_0 is denoted the quasi-true value of ξ . The QML function used is $g_T = (1/T) \sum_{t=1}^T \ln f(y_t|Y_{t-1}, \xi)$, therefore the moment condition in equation (2) is a condition imposed on the gradient of the QML function.

The next step is to invert the auxiliary parameter estimate – or rather the associated score function in equation (2). This is done in order to obtain a consistent estimate of the structural parameter ψ in a second GMM-based step. The left-hand side of equation (2) is simply the sample average of the quasi-score function evaluated at $\hat{\xi}_T$. Therefore it provides an estimate of the expected value of the auxiliary score. EMM means using the corresponding (population) expectation under the structural model of the score from the auxiliary model as moment conditions. The moments depend directly on the structural parameters, since the expected quasi-score is defined under the probability measure $P(Y_t|\psi)$ induced by the structural model. Identification requires that the dimension of the quasi-score – i.e. the number n_ξ of parameters in ξ – exceeds that of the structural parameter vector n_ψ . But otherwise the auxiliary model needs not have anything to do with the structural model. However, the choice of moments is critical for efficiency (as with any GMM-based procedure).

The population moments that identify the structural parameters are given as

$$m(\psi, \xi_0) = E_\psi[s_f(Y_t, \xi_0)] = \int s_f(Y_t, \xi_0) dP(Y_t, \psi).$$

Here the presence of a latent variable renders an analytical expression for this moment condition infeasible. Therefore, the sample moments are computed by using Monte Carlo integration. Hence, the second EMM step is effectively an application of the Simulated Method of Moments. A simulated series $\hat{y}_n(\psi)$ ($n = 1, \dots, N$) is generated from the structural model for a given ψ and used in evaluating the sample moments at the fixed QML estimate $\hat{\xi}_T$. This leads to

$$m_N(\psi, \hat{\xi}_T) = \frac{1}{N} \sum_{n=1}^N \frac{\partial}{\partial \xi} \ln f(\hat{y}_n(\psi) | \hat{Y}_{n-1}(\psi), \hat{\xi}_T).$$

It may be shown that $m_N(\psi, \hat{\xi}_T) = m(\psi, \hat{\xi}_T)$ almost surely as $N \rightarrow \infty$. Thus, for a large enough simulated sample the Monte Carlo error becomes negligible. This error is ignored in the following derivation. Since $\hat{\xi}_T$ is available from the QML step and the quasi-score is given in analytic form the evaluation of the expression above is straightforward. The dimension of the score vector typically exceeds that of the structural parameter vector and in that case the score vector cannot be forced to zero. Instead the GMM criterion in the moment vector is minimized to obtain the EMM estimator of ψ :

$$\hat{\psi}_T = \underset{\psi}{\operatorname{argmin}} [m_N(\psi, \hat{\xi}_T)' \hat{I}_T^{-1} m_N(\psi, \hat{\xi}_T)],$$

where \hat{I}_T denotes a consistent estimator of the asymptotic covariance matrix I of the sample quasi-score vector. The quasi-information matrix that is. If the auxiliary model is expanded to the point where it accommodates all main systematic features of the data, likelihood theory implies that the quasi-scores constitute a (near) martingale difference sequence. And a convenient estimator of the quasi-information matrix is obtained from the outer product of the scores:

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \xi} \ln f(y_t | Y_{t-1}, \hat{\xi}_T) \frac{\partial}{\partial \xi} \ln f(y_t | Y_{t-1}, \hat{\xi}_T)'$$

This matrix may be obtained directly from the first QML step, avoiding the need for computation of the weighting matrix during the second GMM-based estimation step.

Further, a test of the over-identifying restrictions may be obtained directly from the criterion function. The chi-square test is performed by first forming the product of T and the minimized value of the EMM objective function. Under the null hypothesis of correct model specification this product has a chi-square distribution with $n_\eta - n_\rho$ degrees of freedom. Therefore the test statistic may be used to calculate a p-value and a large value of the statistic indicates a lack of fit.

2.6 The extended Short Rate Model by Andersen et al.

The model that Andersen et al. [2] present is the following:

$$\begin{aligned} dr_t &= \kappa_1 \left(\mu_t - \mu_J \frac{\lambda}{\kappa_1} - r_t \right) dt + \sqrt{V_t} dW_{1,t} + Z_t dq_t, \\ dV_t &= \kappa_2 (\alpha - V_t) dt + \eta_1 \sqrt{V_t} dW_{2,t}, \\ d\mu_t &= \kappa_3 (\theta - \mu_t) dt + \eta_2 \sqrt{\mu_t} dW_{3,t}, \end{aligned}$$

$$Z_t \sim N(\mu_J, \sigma_J), \quad \text{Prob}(dq_t = 1) = \lambda dt.$$

This means that Z_t is normally distributed with zero mean and that the probability that $dq_t = 1$ (during the time period dt) is equal to λdt . Here q is a Poisson process uncorrelated with W_i and W_i are in turn uncorrelated Brownian motions. Therefore there is a Lévy process in the first equation and dq_t is often referred to as the 'jump part'. Here follows all the parameters that may be estimated in a calibration of the model:

$$\psi = (\kappa_1, \kappa_2, \kappa_3, \alpha, \theta, \eta_1, \eta_2, \sigma_J, \lambda) \in \mathbb{R}^9 \quad (3)$$

For brevity the following presentation of the model may be chosen, since the mean μ_J of Z_t may be set to zero:

$$\begin{aligned} dr_t &= \kappa_1 (\mu_t - r_t) dt + \sqrt{V_t} dW_{1,t} + Z_t dq_t, \\ dV_t &= \kappa_2 (\alpha - V_t) dt + \eta_1 \sqrt{V_t} dW_{2,t}, \\ d\mu_t &= \kappa_3 (\theta - \mu_t) dt + \eta_2 \sqrt{\mu_t} dW_{3,t}, \end{aligned}$$

$$Z_t \sim N(0, \sigma_J), \quad \text{Prob}(dq_t = 1) = \lambda dt.$$

As described in the previous section an auxiliary model is used for estimation by EMM. In this case the model is an ARMA-EGARCH model including a Hermite expansion, the Semi-Non-Parametric (SNP) density function is as follows:

$$f_K(r_t | x_t; \xi) = \left(\nu + (1 - \nu) \times \frac{[P_K(z_t, x_t)]^2}{\int_{\mathbb{R}} [P_K(z_t, x_t)]^2 \phi(u) du} \right) \frac{\phi(z_t)}{r_{t-1}^\delta \sqrt{h_t}}, \quad \nu = 0.01.$$

Here ϕ denotes the standard normal density function, $x_t = \{r_1, r_2, \dots, r_{t-1}\}$ the information set and ξ the SNP density parameter vector:

$$\begin{aligned} z_t &= \frac{r_t - \mu_t}{r_{t-1}^\delta \sqrt{h_t}}, \\ \mu_t &= \phi_0 + \sum_{i=1}^s \phi_i r_{t-i} + \sum_{i=1}^u \zeta_i (r_{t-i} - \mu_{t-i}), \\ \ln h_t &= \omega \left(1 - \sum_{i=1}^p \beta_i \right) + \sum_{i=1}^p \beta_i \ln h_{t-i} + \\ &\quad (1 + \alpha_1 L + \dots + \alpha_q L^q) [\theta_1 z_{t-1} + \theta_2 (b(z_{t-1}) - \sqrt{2/\pi})], \\ b(z) &= |z| \quad \text{if } |z| \geq \frac{\pi}{2K}, \quad b(z) = \frac{1}{K} \left(\frac{\pi}{2} - \cos(Kz) \right) \quad \text{for } |z| < \frac{\pi}{2K}, \\ K &= 100, \quad P_K(z, x) = \sum_{i=0}^{K_z} a_i(x) z^i = \sum_{i=0}^{K_z} \left(\sum_{|j|=0}^{K_x} a_{ij} x^j \right) z^i, \quad a_{00} = 1, \end{aligned}$$

where j is a multi-index vector, $x^j \equiv (x_1^{j_1}, \dots, x_M^{j_M})$ and $|j| \equiv \sum_{m=1}^M j_m$.

The function $b(z)$ is twice-differentiable. It closely approximates the absolute value function in the EGARCH variance equation. The reason for Andersen et al. to use $b(z)$ instead of simply using the absolute value of z (which is used for the 'standard' EGARCH model) is quite obvious, since it is necessary to calculate partial derivatives of the natural logarithm of the SNP density function for EMM. Those partial derivatives are necessary for calculating the function values of the score function. And the absolute value function is not differentiable at the origin. Therefore $b(z)$ may be described as a 'smoothing' of the absolute value function.

In Andersen et al. [2] a model selection procedure is described and their results indicate that an ARMA(4,1)-Level-EGARCH(2,1)- $K_z(6)$ - $K_x(0)$ model would be the model of choice. This is also the model that is used for the project described in this report.

Some restrictions on the parameters in this ARMA-EGARCH model are used to make sure that the time series is stationary. If none of the roots of the following polynomial satisfies the equality $|z| = 1$ then the ARMA process for the conditional mean is stationary:

$$1 - \sum_{k=1}^s \phi_k z^k.$$

Also the ARMA process is casual if the inequality $|z| > 1$ is satisfied. Invertibility is assured by estimating the parameters ζ_k such that all the roots of the following polynomial lie outside the unit circle:

$$1 + \sum_{k=1}^u \zeta_k z^k.$$

Similarly stationarity of the EGARCH process is enforced by ensuring that the roots of the following polynomial lie outside the unit circle:

$$1 - \sum_{k=1}^p \beta_k z^k.$$

These are restrictions to enforce in the first step of the EMM procedure, i.e. during the QMLE. In the second step some restrictions on the parameters in the extended Short Rate Model are important to enforce. In this model the

stochastic mean and the stochastic volatility are described by CIR processes, therefore the model may be described as having the Heston volatility model as an integrated part of it. The restrictions used to ensure positive CIR processes are:

$$\begin{aligned} 2\kappa_2\alpha &> \eta_1^2, \\ 2\kappa_3\theta &> \eta_2^2. \end{aligned}$$

All parameters in the extended Short Rate Model are also assumed to be positive.

For the EMM procedure it is necessary to discretize the Stochastic Differential Equations so that the numerical computations can be done. Andersen et al. [2] state that an Euler scheme may be used. The result of the discretization is as follows:

$$\begin{aligned} r_{i+1} &= r_i + \kappa_1(\mu_i - r_i)\Delta t_{i+1} + \sqrt{V_i}x_{1,i+1}\sqrt{\Delta t_{i+1}} + Z_i dq_i, \\ V_{i+1} &= V_i + \kappa_2(\alpha - V_i)\Delta t_{i+1} + \eta_1\sqrt{V_i}x_{2,i+1}\sqrt{\Delta t_{i+1}}, \\ \mu_{i+1} &= \mu_i + \kappa_3(\theta - \mu_i)\Delta t_{i+1} + \eta_2\sqrt{\mu_i}x_{3,i+1}\sqrt{\Delta t_{i+1}}. \end{aligned}$$

Here x_k are random variables drawn from the standard normal distribution.

2.7 Estimating the covariance matrix

As described in Section 2.5 an estimate of the covariance matrix is required for EMM. Andersen et al [2] describe how they use the outer product of the gradient as an estimate of this matrix. However, in many applications of Estimation Theory it turns out that this estimate of the covariance matrix is numerically ill-conditioned. This is also the case here, here it turns out to be very ill-conditioned.

To address the problem of the ill-conditioned matrix an alternative estimator may be used. The type of estimator used for this project is described by Ledoit and Wolf [1]. In their paper they describe this estimator as both well-conditioned and asymptotically more accurate than what they call the 'sample covariance matrix'. In this context the scores will constitute the

'sample' and they state that one way to get a well-conditioned structured estimator is to impose the condition that all variances are the same. And that all the covariances are zero but the estimator they recommend is a weighted average of this structured estimator and the sample covariance matrix. They describe this estimator as inheriting the good conditioning properties of the structured estimator and that it is more accurate than both the sample covariance matrix and the structured estimator. This is achieved by choosing the weight optimally according to a quadratic loss function.

A difficulty is that the true optimal weight depends on the true covariance matrix, which clearly is unobservable. This is solved by finding a consistent estimator of the optimal weight and the authors also show in their paper that replacing the true optimal weight with a consistent estimator makes no difference asymptotically.

Let X denote an $n \times p$ matrix of independent and identically distributed observations on a system of p random variables with mean zero and covariance matrix Σ . That the observations will have zero mean is assured when the moment condition in equation (2) in Section 2.5 is satisfied, this was also explained in that section.

Further, a norm is used that is similar to the Frobenius norm but with the difference that there is a division by p in this expression:

$$\|A\| = \sqrt{\text{tr}(A^t A)/p}. \quad (4)$$

The goal here is to find the linear combination $\Sigma^* = \rho_1 I + \rho_2 S$ of the identity matrix I and the sample covariance matrix $S = X^t X/n$ whose expected quadratic loss $E[\|\Sigma^* - \Sigma\|^2]$ is minimum. Ledoit and Wolf further states that the optimality result that will be obtained will come at a price, Σ^* will not be a *bona fide* estimator. This is so because it will require hindsight knowledge of four scalar functions of the true (and unobservable) covariance matrix Σ . However, what is presented is a *bona fide* estimator S^* with the same properties as Σ^* asymptotically as the number of observations and the number of variables go to infinity together.

Associated with the norm given in equation (4) is the inner product $\langle A_1, A_2 \rangle = \text{tr}(A_1^t A_2)/p$. And four scalars have a central role here: $\mu = \langle \Sigma, I \rangle$, $\alpha^2 = \|\Sigma - \mu I\|^2$, $\beta^2 = E[\|S - \Sigma\|^2]$ and $\delta^2 = E[\|S - \mu I\|^2]$. It is necessary to assume that the random variables have finite fourth moments so that β^2 and δ^2 are finite.

We have:

Theorem 2.1 Consider the optimization problem:

$$\begin{aligned} \min_{\rho_1, \rho_2} \quad & E[|\Sigma^* - \Sigma|^2], \\ \text{s.t.} \quad & \Sigma^* = \rho_1 I + \rho_2 S, \end{aligned}$$

where the coefficients ρ_1 and ρ_2 are nonrandom. Its solution verifies:

$$\begin{aligned} \Sigma^* &= \frac{\beta^2}{\delta^2} \mu I + \frac{\alpha^2}{\delta^2} S, \\ E[|\Sigma^* - \Sigma|^2] &= \frac{\alpha^2 \beta^2}{\delta^2}. \quad \square \end{aligned}$$

Now the subscript n will be used because all results hold asymptotically and this is introduced in the four scalars: $\mu_n = \langle \Sigma_n, I_n \rangle_n$, $\alpha_n^2 = \|\Sigma_n - \mu_n I_n\|_n^2$, $\beta_n^2 = E[\|S_n - \Sigma_n\|_n^2]$ and $\delta_n^2 = E[\|S_n - \mu_n I_n\|_n^2]$. Several lemmas and theorems are presented in the paper by Ledoit and Wolf but many of them will be left out in this report, instead some more scalars will be defined. And then the sought estimator will be given. Define:

$$\begin{aligned} m_n &= \langle S_n, I_n \rangle_n, \quad d_n^2 = \|S_n - m_n I_n\|_n^2, \quad \bar{b}_n^2 = \frac{1}{n^2} \sum_{k=1}^n \|x_k^n (x_k^n)^t - S_n\|_n^2, \\ b_n^2 &= \min(\bar{b}_n^2, d_n^2), \quad a_n^2 = d_n^2 - b_n^2. \end{aligned}$$

Now when those scalars are given the unobservable scalars in the formula defining Σ_n^* are replaced with the given consistent estimators. The result is the sought *bona fide* estimator of the covariance matrix:

$$S_n^* = \frac{b_n^2}{d_n^2} m_n I_n + \frac{a_n^2}{d_n^2} S_n.$$

In Section 4.4 the results are given after computing the estimator to be used for the EMM procedure.

2.8 Conversion of rates

The Short Rate Models are calibrated using historical data of the 6-month Euribor rate, these rates are quoted on an actual/360 day count convention. However, for the calibrations using MLE and the EMM procedure it is necessary to use rates given on continuous compounding. When a spot rate r_d is given on an actual/360 day count basis the discount factor may be calculated as follows:

$$D_f = \frac{1}{1 + r_d T}.$$

If the time to maturity is six months and the actual number of days during this interest period is 181 for instance, then $T = 181/360$. And when a rate is given on continuous compounding then the discount factor is given by the following formula instead:

$$D_f = e^{-r_c T}.$$

Therefore the sought rate may be calculated as follows:

$$r_c = -\frac{1}{T} \ln \left(\frac{1}{1 + r_d T} \right).$$

In Section 3.1 a program developed with .Net is mentioned and the conversion of the rates is one of the things that this program is used for.

3 Implementation

In this section the implementation of the project is presented. What follows is a description of how the estimations of the parameters in the Vasicek and CIR models are performed. Also the implementation for the estimation of the parameters in the SNP density function and the second step of EMM (Efficient Method of Moments) is presented. This is for the calibration of the extended Short Rate Model. But first the data processing of the interest rates is described.

3.1 Processing the historical data

Before carrying out MLE and EMM it is necessary to process the historical data of the rates and convert them to continuous compounding. Here something similar to the procedure described in Andersen et al. [2] is done, first the day of the week with the least missing observations is selected. A table is created with the number of occurrences, i.e. a table that may be used to tell how the rates are distributed over the days of the week. As shown in table 1 below wednesday is the day with the most observations, just as in the case with the work by Andersen et al. [2]. The reason for missing observations is obvious, since there are public holidays which may occur during one of the first five days of the week. Such days are not trading days and therefore there is no data for those days.

Weekday	Occurrence
Monday	812
Tuesday	825
Wednesday	827
Thursday	825
Friday	813

Table 1: Occurrences

For those weeks for which the rates of wednesday are missing the rates of tuesday may be selected in most of the cases. This is described in Andersen et al. [2]. But for some of the weeks it is not even possible to select the rates of tuesday, nor the rates of wednesday. The reason for this becomes clear after examining those weeks and having a look at the calendar. It turns out that for some years Christmas day occurs on a tuesday and both Christmas day and the following day are public holidays. Those weeks are listed in table 2 and for those weeks the rates of thursday are selected instead.

Date	Week-No
2001-12-26	52
2007-12-26	52
2012-12-26	52

Table 2: Dates without observations

The observed Euribor rates are given on an actual/360 day count convention and they are converted to continuous compounding, in congruence with Section 2.8. This conversion and the data processing described above is made possible by first developing a Windows application with .Net and C#. And then by using this application, the data series processed is the set of Euribor rates from 1998-12-30 to 2015-01-06. This processing leads to a series of 836 weekly rates which is used for the calibrations.

3.2 Calibration of the Vasicek model and chi-square testing

The calibration of the Vasicek model is done by using Maximum Likelihood Estimation, as explained in Section 2.1. This estimation is performed by using Matlab and a function that may be found – at this writing – at the official web site of Sitmo (www.sitmo.com). This Matlab function is a function for performing MLE for the calibration of the Vasicek model. The input consists of the rates and the time step and the output values are the estimated parameters. The time step is set to $dt = 1/52$ in congruence with Section 3.1.

After this calibration the chi-square test explained in Section 2.3 follows next. For this a Matlab function is written, the results of the calibration and the chi-square test are given in Section 4.1. Simulations are needed for the chi-square test and for them the time step $dt = (1/52) \cdot (1/25)$ is used.

3.3 Calibration of the CIR model and chi-square testing

The calibration of the CIR model is also performed by using MLE and Matlab. Some Matlab functions written by Kamil Kladvko are used for this calibration and those functions are available – at this writing – at the official web site of MathWorks. In Section 2.2 the log-likelihood function was derived but for the implementation another analytical expression is used. But

this expression is equivalent:

$$\begin{aligned} \ln L(\theta) &= (N-1)\ln(c) + \\ &+ \sum_{i=1}^{N-1} \left\{ -u_i - v_{i+1} + 0.5q \cdot \ln\left(\frac{v_{i+1}}{u_i}\right) + \ln(I_q(2\sqrt{u_i v_{i+1}})) \right\} = \\ &= (N-1)\ln(c) + \\ &\sum_{i=1}^{N-1} \left\{ -u_i - v_{i+1} + 0.5q \cdot \ln\left(\frac{v_{i+1}}{u_i}\right) + \ln(I_q[2\sqrt{u_i v_{i+1}}] \exp[-2\sqrt{uv}] + 2\sqrt{uv}) \right\}. \end{aligned}$$

This expression is used because it is problematic to perform MLE with the Bessel function in the expression. The reason is that this function may diverge to plus infinity and to avoid this trouble a scaled Bessel function is used instead – according to the expression above. The time step $dt = 1/52$ is also used for this calibration. During this MLE the Feller condition (1) given in Section 2.2 is imposed.

After this MLE is performed the chi-square test follows next, here the same Matlab function is used as in the case with the Vasicek model. In Section 4.2 the results from the calibration and the chi-square test are presented. For the simulations the time step $dt = (1/52) \cdot (1/25)$ is used.

3.4 Modeling the Poisson process

Andersen et al. [2] describe in their paper how they model the Poisson process for their numerical implementation. It is necessary to do so and to 'smooth' the discontinuities of this process. They describe how they first approximate the differential of the Poisson process by replacing dq_t with a random variable Y that has a Binomial distribution. For this variable $\text{Prob}\{Y = 1\} = \lambda(t)dt$ and $\text{Prob}\{Y = 0\} = (1 - \lambda(t)dt)$. They generate a random variable U with the distribution of Uniform(0,1) and 'smooth' the discontinuity of Y over an interval centered around $1 - \lambda(t)dt$:

$$Y = \begin{cases} 0 & \text{if } 0 \leq U < 1 - \lambda(t)dt - h/2, \\ g(X) & \text{if } 1 - \lambda(t)dt - h/2 \leq U < 1 - \lambda(t)dt + h/2, \\ 1 & \text{if } 1 - \lambda(t)dt + h/2 \leq U \leq 1, \end{cases}$$

$$X = U - (1 - \lambda(t)dt - h/2),$$

$$g(X) = -\frac{2}{h^3 X^3} + \frac{3}{h^2 X^2} \quad \text{for } 0 \leq X \leq h.$$

As the interpolation interval length h goes to zero the function g becomes steeper and it is a C^∞ function. They describe how they fine-tune h by choosing the smallest possible interval size that eliminates the numerical problems in the EMM criterion function.

However, there are some issues with this fine-tuning. For some values it may happen that the function above will take on negative values and this may happen for small values of h . It is observed that in those cases the function above does not lead to an approximation of the Poisson process at all.

To avoid those problems with the fine-tuning a completely different approximation is used for this project. This makes things less complicated. It may be noted that the function being approximated here is essentially a step function and to the scholars of distribution theory the Heaviside step function is well-known. And the following function is a smooth approximation of the Heaviside step function:

$$f(x) = \frac{1}{1 + e^{-2kx}}.$$

The larger the value of k chosen here the sharper will the transition be at $x = 0$. But in this case the transition shall not be at $x = 0$. Since the smoothing of Y explained above is over an interval centered around $1 - \lambda(t)dt$ the transition shall also be at $1 - \lambda(t)dt$. Therefore the following function is used:

$$f(x) = \frac{1}{1 + \exp(-2k[x - (1 - \lambda(t)dt)])}.$$

Here some tuning is also necessary, k must be chosen large enough so that this function will lead to an accurate approximation of the Poisson process. This parameter should be chosen so that the curve at the transition is very steep, it should almost coincide with a straight vertical line in the graph. It is also important to keep in mind that for the values of x that satisfy the inequality $x < 1 - \lambda(t)dt$ the function value may be positive but very tiny. This value could be of the order 10^{-110} for instance, or even less.

To avoid problems with cancellations – that may occur while computing – the approximated numerical values of $Z_i dq_i$ are rounded to numbers that are not smaller than 10^{-15} . Numbers smaller than that are so small that they may be neglected. One more advantage of using this function is that this means that the Matlab code may be vectorized. It is known that Matlab executes slowly if many iterations are used and that vector calculations are quicker. In Section 4.5 the result of the fine-tuning of this function is presented.

3.5 Smoothing of the rates

When using the computer program to simulate short rates under the extended Short Rate Model it may occasionally happen that the values take on negative or large values. In order to handle this Andersen et al. [2] describe how they use a certain function, given a simulated rate r_t they compute the smoothed rate \hat{r}_t according to

$$\hat{r}_t = \begin{cases} C_1 \exp(C_2 r_t) + C_3 & \text{if } r_t \leq r_{min}, \\ r_t & \text{if } r_{min} < r_t < r_{max}, \\ 0.5(r_t + r_{max} + \ln(1 + r_t - r_{max})) & \text{if } r_{max} \leq r_t, \end{cases}$$

$$r_{min} = 0.02\%, \quad r_{max} = 75\%, \quad C_1 = 0.5r_{min}\exp(-2), \quad C_2 = 2/r_{min}, \\ C_3 = 0.5r_{min}.$$

These smoothed rates are passed in an argument to the objective function in the second EMM step.

3.6 The first step of the EMM procedure

The first step of the EMM procedure is to estimate the parameters in the auxiliary model, this is done by using Quasi-Maximum Likelihood Estimation (QMLE). The problem of this QMLE is formulated as an optimization problem and the estimation is done by using Matlab. The function *fmincon* may be used for this and the problems are specified by

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

Here b and beq are vectors, A and Aeq are matrices and $c(x)$ and $ceq(x)$ are functions that return vectors. The function $f(x)$ returns a scalar and this function can be nonlinear, the latter also applies to $c(x)$ and $ceq(x)$. The bounds lb and ub may be passed as vectors and this is also done during this implementation. The constraints are set according to the restrictions on the parameters described in Section 2.6, this is done by defining the function $c(x)$ and passing this function as an argument to $fmincon$.

However, it turns out that there are some issues when using $fmincon$ for this QMLE. The solver is converging to a value but then it stops, according to the message displayed the current step size is less than the default step size tolerance. In order to try to resolve this issue some attempts are made by adjusting this step size tolerance and also the error tolerance of the objective function. These attempts do not lead to the desired results.

It turns out that it is much easier to use another solver and a completely different algorithm. The function used instead is $fminsearchcon$ which is based on the Nelder-Mead algorithm, this method is gradient-free and by using this the issues with the step size tolerance are avoided. This function uses $fminsearch$ and $fminsearchcon$ was developed by John D'Errico, the Matlab files are available at the official home page of MathWorks (at this writing). This function is used to solve problems specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0, \\ ceq(x) = 0, \\ A \cdot x \leq b, \\ lb \leq x \leq ub. \end{cases}$$

The observed rates are passed as elements of a vector that is used as an argument. It is also necessary to use the first four rates as a presample, the reason for this is that the AR model contained in the ARMA-EGARCH model is of the order four – just as described in Section 2.6.

It turns out that it is too restrictive to impose all the constraints described in Section 2.6, therefore the constraint for causality is relaxed. The reason is that when constraints are imposed during optimization the gradient of the objective function at the optimum might not be equal to the zero vector. There are special cases when it is equal to it but in the general case it is not equal to the zero vector. One of the moment conditions is that the gradient shall be equal to the zero vector and therefore this is checked. The optimization is performed and then the roots of the polynomial $1 - \sum_{k=1}^s \phi_k z^k$ are examined afterwards to check that none of these lie on the unit circle. It is still important that the ARMA process is both stationary and invertible and the constraint on β_i is also imposed.

A starting point has to be used and the values of β_1 and β_2 from the paper by Andersen et al [2] are used as elements in this vector. It is reasonable to use them as those parameters satisfies the inequalities described above. It is also necessary to initiate the EGARCH model which means that a presample conditional variance is needed. The natural logarithm of this presample variance is included as one of the parameters estimated, the estimated unconditional variance of the historical data is used for the initial guess of this conditional variance. The formula for the natural logarithm of the estimated unconditional variance is:

$$h = \ln \left(\frac{1}{n} \sum_{i=1}^n y_i^2 \right).$$

Here all the observed rates are included in the sum, also the presample. In the presample conditional mean each mean is set equal to the corresponding rate. And therefore all the presample innovations (error terms) are set equal to zero, in congruence with those means.

The lower bounds for all the parameters are set to -4 and the upper bounds are set to 4. The only exception is the parameter ϕ_0 for which the lower bound is set so that this parameter shall be greater than zero. The upper bound is also set to 4 for this parameter.

It is also necessary to use values for the initial guess of the parameters in the ARMA model contained in the ARMA-EGARCH model. These parameters are assumed to be close to the parameters obtained after fitting the ARMA model to the observed rates, by minimizing the sum of the squared residuals. Therefore the Ordinary Least Squares (OLS) estimator is calculated here. All other coordinates of the starting point are set to 0.5, this results in a vector which satisfy all the constraints.

3.7 The second step of the EMM procedure

Given the estimated parameters obtained by QMLE – described in Section 3.6 – the second step of the Efficient Method of Moments procedure follows next. In this step the parameters given in (3) in Section 2.6 are estimated. This problem is also formulated as an optimization problem and the Matlab function *fmincon* is also used for this step (initially). However, the issue with the step size tolerance – as explained in the previous section – appears again. Therefore the function *fminsearchcon* is used again.

For this step an initial guess is also needed and the elements in this vector are set so that the conditions described in Section 2.6 are satisfied. Some experimentation with the values leads to the vector presented in Section 4.5.

The number of simulated rates in each iteration is equal to 10000 but the first 5000 rates are discarded. The reason for this is that the values used to initiate the numerical calculations should not affect the result. Therefore the last 5000 simulated rates are the only ones used for calculating the value of the objective function.

4 Results

In this section the results are presented, what follows are the results from calibration and chi-square testing for the Vasicek, CIR and the extended Short Rate Model.

4.1 Calibration of the Vasicek model and chi-square testing

The calibration of the Vasicek model and chi-square testing is performed in congruence with Section 3.2, the estimated (and rounded) parameters are given in table 3.

Parameter	Estimate
a	-0.037374
b	0.071924
σ	0.003871

Table 3: The estimates

The optimal value is approximately equal to 3862.38 here. With these parameters given the chi-square test follows next and it is performed in congruence with Section 2.3, the result is given in table 4.

Quantity	Value
Statistic	509.12
M	164
p-value	0

Table 4: χ^2 test result

With the value of the test statistic this large the p-value is tiny, it is practically equal to zero. Therefore the null hypothesis is rejected, it does not matter if the significance level would be set as low as equal to 1% since the p-value is smaller than that. In other words, the hypothesis that the Vasicek model describes the stochastic process of the Euribor 6-month rate is rejected.

If this chi-square test would be repeated then another value of the test statistic would be obtained. But none of the trials led to a value that would imply a p-value even close to 1%, none of those values of the test-statistic were small enough.

4.2 Calibration of the CIR model and chi-square testing

The calibration of the CIR model and chi-square testing is performed in congruence with Section 3.3, the estimated (and rounded) parameters are given in table 5.

Parameter	Estimate
a	0.086978
b	0.002910
σ	0.022499

Table 5: The estimates

The value of the log-likelihood function is approximately 6.335416 at this optimum. And the result from the chi-square test is given in table 6.

Quantity	Value
Statistic	315.77
M	164
p-value	1.0988e-11

Table 6: χ^2 test result

Here the test statistic is much smaller compared to the case with the Vasicek model, but it is still not small enough for implying nothing but a tiny p-value. Therefore the hypothesis that the CIR model describes the stochastic process of the Euribor 6-month rate is also rejected.

Just as in the case with the chi-square test of the Vasicek model different test statistics will be obtained in different trials. But again, no observation was ever made of a test statistic that would imply a p-value large enough.

4.3 The first step of the EMM procedure

As described in Section 3.6 an initial guess is needed for the QMLE, therefore the natural logarithm of the estimated unconditional variance is calculated. Also the OLS estimator of the parameters in the ARMA model is computed, as this estimator is used for the initial guess. The elements of the vector containing the initial guess are given in table 7 below, rounded to four decimals. The last parameter in this table is the natural logarithm of the estimated presample conditional variance.

Parameter	Initial guess
ϕ_0	$9.370 \cdot 10^{-4}$
ϕ_1	1.9770
ϕ_2	-1.0687
ϕ_3	0.1086
ϕ_4	-0.0173
ζ_1	-0.6507
ω	0.5000
α_1	0.5000
β_1	1.6288
β_2	-0.6315
θ_1	0.5000
θ_2	0.5000
δ	0.5000
a_{01}	0.5000
a_{02}	0.5000
a_{03}	0.5000
a_{04}	0.5000
a_{05}	0.5000
a_{06}	0.5000
η	2.1037

Table 7: The initial guess

With this vector given the QMLE procedure is carried out, the result is given in table 8 with rounded numbers.

It takes slightly less than 2.5 hours to run this QMLE script. It is observed that the value of the objective function – i.e. the negative of the Quasi-Maximum Likelihood function value – is approximately equal to $g = -1.9681$ (rounded). It is also important to calculate the gradient of the QML function at this optimum to verify that the elements are close to zero. This is important since this is once of the moment conditions during the second step, as explained in Section 2.5. Even if the elements are not equal to zero it turns out that this is good enough.

The roots of the polynomial $1 - \sum_{k=1}^s \phi_k z^k$ are given in table 10, it is important to examine them since none of them should lie on the unit circle. As it turns out, none of them lie there.

Parameter	Estimate
ϕ_0	0.0039269
ϕ_1	1.9632151
ϕ_2	-1.0797751
ϕ_3	0.1554109
ϕ_4	-0.0255158
ζ_1	-0.7808014
ω	-0.4682029
α_1	2.2085778
β_1	1.2986088
β_2	-0.2995694
θ_1	0.2771091
θ_2	0.3603787
δ	0.8590185
a_{01}	2.3562138
a_{02}	3.9864244
a_{03}	-1.1912557
a_{04}	1.9779675
a_{05}	-3.2476398
a_{06}	0.767797
η	1.581841

Table 8: The estimates

Parameter	Element
ϕ_0	-0.0305902
ϕ_1	-0.0320809
ϕ_2	-0.0557897
ϕ_3	-0.0698803
ϕ_4	-0.0057946
ζ_1	0.0031403
ω	0.0000376
α_1	-0.0042114
β_1	0.1612341
β_2	0.131776
θ_1	-0.0412521
θ_2	-0.0057746
δ	-0.001761
a_{01}	-0.0000244
a_{02}	-0.0002344
a_{03}	-0.0008068
a_{04}	-0.0013416
a_{05}	-0.0010697
a_{06}	-0.0008772

Table 9: The gradient

Root
$1.933896 + 5.364194i$
$1.933896 - 5.364194i$
1.284862
0.938117
0.938117

Table 10: Roots

4.4 The estimated covariance matrix

Given the estimated parameters from the QMLE the asymptotic covariance matrix is estimated, this is done in congruence with Section 2.7. First the sample covariance matrix S is calculated and then the estimator proposed by Ledoit and Wolf [1] is calculated. The latter is denoted by W and the results are given in table 11.

Matrix	Condition number
S	$4.4990 \cdot 10^{15}$
W	307.74

Table 11: Condition numbers

The matrix W is inverted and then this inverse is used to calculate the objective function values in the second EMM step.

4.5 The second step of the EMM procedure

Once the QMLE is done and the covariance matrix is estimated the second step of the EMM procedure follows. In this step a value for k is needed for the modeling of the Poisson process. As explained in Section 3.5 this parameter is used in the approximation of a step function and some experimentation leads to $k = 5 \cdot 10^6$. With this large value the approximation of the Poisson process is considered to be good.

For the second step of the EMM procedure a starting vector is also needed and the elements are given in table 12.

Parameter	Initial guess
κ_1	0.810282
α	0.00005
κ_2	1
η_1	0.002600
θ	0.050516
κ_3	1.039055
η_2	0.285115
σ_J	$6.16 \cdot 10^{-13}$
λ	1.633223

Table 12: The initial guess

This is a result of experimenting with the values trying to find a point close to the one of the optimum. And then an attempt is made to estimate the parameters by using *fmincon*, but after running for approximately 59 hours the solver stops. The message displayed is that the size of the current step is less than the default value of the step size tolerance. Then *fminsearchcon* and the Nelder-Mead algorithm are used instead. Then it takes an additional 15 hours (approximately) to find the optimum and the estimates are given in table 13.

Parameter	Estimate
κ_1	0.632735
α	$2.530487 \cdot 10^{-5}$
κ_2	0.822380
η_1	$4.363919 \cdot 10^{-3}$
θ	$5.817222 \cdot 10^{-2}$
κ_3	1.086908
η_2	0.275111
σ_J	$1.947530 \cdot 10^{-7}$
λ	1.615313

Table 13: The estimates

The optimal value is 0.002566 (rounded) and with this value the test statistic for the chi-square test may be calculated.

Quantity	Value
T	832
Statistic	2.14
$n_\eta - n_\rho$	10
p-value	0.9952

Table 14: χ^2 test results

This is a very high p-value and the significance level for a chi-square test is usually set from 1% to 5%. With such a significance level the null hypothesis will not be rejected. In other words, the hypothesis that the extended Short Rate Model describes the stochastic process in question will not be rejected.

5 Discussion

5.1 The estimated covariance matrix

With the results given in 4.4 it may be concluded that the estimator proposed by Ledoit and Wolf [1] means a vast improvement of the condition number of the matrix. Here this condition number is compared to the one of the sample covariance matrix and the latter has a condition number of the order of 10^{15} .

Maybe such a matrix can be described as 'useless' in the sense that it cannot be used for numerical computations. With a condition number this high a small numerical error will lead to a catastrophic error amplification and therefore the results of the computations will not be correct.

The well-conditioned estimator will also lead to error amplification since the condition number is not equal to one. But since the condition number is of the order of 10^2 this is clearly much better and this matrix may be used.

5.2 The implementation

As stated earlier in this report the implementation was done by first developing computer programs with .Net and Matlab. The former technology was used for developing the program used for the processing of the historical data. According to the author of this report C# is an excellent programming language for developing software used for data management. Having access to all the libraries in the .Net-framework with all the methods and all the classes makes problem solving easy.

Similarly there is an incentive for choosing Matlab for numerical computations, there are so many libraries in Matlab that may be used for solving problems in Numerical Analysis and Mathematics. This can be fantastic for solving many of the problems in those fields.

However, there are some drawbacks with using Matlab and this was experienced during this implementation. Maybe Matlab is not a good choice for problem solving with some of the simulation-based methods and maybe it is best to think twice before choosing it. This is a lesson learned the hard way during this project and it may be noted that Andersen et al. [2] used a completely different program. The program that they used was written in C++

instead and the Metropolis-Hastings algorithm was used. This algorithm is based on a Markov Chain Monte Carlo method and this is a completely different approach to solving the problem of estimation by EMM.

Writing a program in Matlab is often easy and often it is not necessary to write so many lines of code. But maybe this advantage and the advantage of having access to all the libraries does not outweigh the drawbacks with Matlab. Matlab has long been known for executing slowly when many iterations are used, but proponents of Matlab may claim that this is not the case anymore as the Just-In-Time accelerator has been implemented. But these claims are not consistent with the results and the observations made during the implementation of this project.

Those issues with slow iterations have been encountered during this implementation and sometimes it is very hard to avoid iterations. Sometimes iterations may be replaced by writing lines of code with vectorization and this can lead to faster execution. But vectorization is not always possible and this can be exemplified by the computations with the ARMA-EGARCH model presented earlier in Section 2.6. The reason is that the values of the conditional mean and variance are calculated by using recursive formulas.

It was observed that it took many seconds to calculate the values of the objective function in the second step of EMM. And this is far more problematic compared to the calculations in the first step. The reason is that the Jacobian must be calculated for the calculation of the scores in the second step. At the same time it was observed that only about 12% to 16% of the capacity of the CPU in the computer is used during this second step. And this is also the total usage including the usage of other programs running on the computer at the same time. The author of this report must say that this is really bad.

The computer used for this implementation has an Intel Core i7 CPU and 16.0 GB RAM which means that there are seven cores in the CPU. I (the author) think that this is really good and it should be possible to use the capacity of the computer better. Therefore some attempts were made to make use of all the cores of the CPU by using the Parallel Computing Toolbox in Matlab. The keyword *spmd* was used in some tests where values of the objective function was calculated and it was observed that more than 90% of the capacity of the CPU was used here. This may look much better but it turned out that this did not lead to faster calculations. Instead it took much longer time to calculate those values compared with the case when the Parallel Computing Toolbox was not used.

As there were no indications that usage of the Parallel Computing Toolbox would lead to faster calculations the second step was performed without this toolbox. It may be noted that in some cases parallelization cannot easily be applied to an algorithm for faster execution. This fact is known by the scholars of Parallel Computing and apparently this program for EMM is such an example. Clearly it is important to find a way to successfully apply parallelization if faster executions are desired, otherwise it will not even be possible to use a supercomputer for 'speeding up' the computations.

As the results presented in Section 4 show the second step took several days, but the QMLE was much faster in comparison. Before this some attempts were made to see if it would be possible to use the Matlab Coder to generate code in C++ from the Matlab code. However, some of the Matlab functions used in the program are not possible to translate using the Matlab Coder. Therefore it would have been necessary to rewrite the program first in that case, replacing all those function calls with something else. Therefore it was considered to be too complicated to try to find a better solution by first translating.

If C++ would have been chosen from the beginning it would certainly have meant that many more lines of code had to be written. This is so since a Matlab program can be written with few lines – in comparison – but maybe it would still have been worth it. This is a reflection after considering the drawbacks with Matlab, maybe it would have been easier to perform EMM with more simulated rates if C++ would have been the language of choice. But it is clearly non-trivial to write a very good program for estimation by EMM.

Another reflection is that maybe the second step of EMM would have been much quicker if the function *fminsearchcon* would have been used from the very beginning. After the author examined some other examples of GMM and EMM the conclusion is that in all cases the Nelder-Mead algorithm was used for the optimization. In none of those cases gradient-based methods were used and maybe such methods are not suitable for EMM.

5.3 Comparing the models

Summarizing the results of the chi-square tests – given in Section 4 – the conclusion is that there is only one null hypothesis that was not rejected. And that null hypothesis is that the extended Short Rate Model describes the stochastic process of the Euribor 6-month rate. In this case the p-value was very high but for the Vasicek and CIR models the p-values were very

low instead. Therefore these findings give no support to the Vasicek or CIR models for accurately describing the stochastic process in question. But there is support for the extended Short Rate Model, this indicates that it may be advantageous to use the latter.

These findings give support to the findings of Andersen et al. [2], in their report the p-value was quite high when the chi-square test was performed for the extended Short Rate Model. The models were calibrated using US 3-month T-Bill rates and the p-value was very small for the CIR model. Therefore the result that the null hypotheses for both the Vasicek and CIR models were rejected is maybe what one may expect, after reading [2].

That the value of the test statistic in the case of the Vasicek model was even larger than the one of the CIR model might come as no surprise. A large value of the test statistic indicates a lack-of-fit and it seems like the fit was even worse in the case of the Vasicek model. It seems like the cause is the constant volatility of the Vasicek model instead of letting the volatility be level-dependent – the only difference between those two models is the factor $\sqrt{r_t}$ in the martingale part in the SDE of the CIR model.

It makes sense that it is better to model with level-dependent volatility rather than with constant volatility. If the rate decreases to a number that is close to zero then the random shocks should be small, it is easy to realize this since negative rates rarely occurs. If the rate is assumed to be positive then it cannot decrease much more if it already is quite small. And if the rate is small and if the volatility is high at the same time then the risk is high that a negative rate will be obtained in a numerical computation. This characteristic with smaller random shocks when the rate is small is not captured in a model with constant volatility.

On the other hand, this characteristic is neither captured in the Chen model nor in the extended Short Rate Model. With these stochastic volatilities – which are not level dependent – negative rates may occur and this may be viewed as a major drawback with such models. However, these three-factor models may on overall lead to better fitting than the one-factor models and this is what the results in Section 4 indicate – at least in the case of comparing the extended Short Rate Model with these two one-factor models. Real world market data indicates that interest rates tend to revert to a mean that is not constant but varies by time. This characteristic can be captured with a stochastic mean such as the case with the Chen model and its extension.

It may be pointed out that there may be better ways of calibrating the Vasicek and CIR models compared to what was described here. Here one

rate per week was selected and the other observations were discarded, but there are Kalman filters that may be used with incomplete panel-data. With these filters it is possible to handle the missing observations on the weekends and with them it is not necessary to discard data. It is possible to cope with missing data and the fact that all days are not trading days, that is. Maybe the use of a Kalman filter would have resulted in a better fit. There was however a reason for not doing this, it is that the implementation explained in this report may be viewed as 'treating the models on equal footing'. Data was discarded when the extended Short Rate Model was about to be calibrated and then the same series should be used when the one-factor models are calibrated. This should make it easier to compare the models and this procedure of discarding some of the data and selecting the rates to keep is known as statistical *imputation*. There are other variants of imputation where it is not necessary to discard data.

One thing that would be very interesting to investigate is pricing of interest rate derivatives under the extended Short Rate Model. And if such pricing would lead to small or crude discrepancies with real world market data. If the discrepancies would be considerably smaller than the discrepancies with market data and pricing with one-factor models – for instance – then this would also give support to the extended Short Rate Model. In that case the findings would give support to the thesis that the latter gives a better description of the stochastic process in question, that is. Such pricing could be tested with Monte Carlo simulations.

However, there is something that can be very problematic with such an investigation and that is the difficulties with getting financial data. It is difficult to get this because the financial data feed services are very expensive and therefore one might find it hard to finance it. Financial institutions all over the world spend tens of thousands of US dollars per license and per year but such organizations have the resources they need. Therefore they can afford it but it is too expensive for a thesis project and this is very unfortunate. The Euribor 6-month rates used for the calibrations during this project are available to the public so anyone can get this data.

An impression the author got after talking with a Financial Analyst is that many of them may choose much simpler models than the extended Short Rate Model. The reason is that one-factor models are much easier to use and the Analysts can avoid the hurdles of calibrating three-factor models. When considering the experiences from this project this is not really surprising since it took so much more effort to calibrate the extended Short Rate Model than the one-factor models. Examples of how to calibrate the Vasicek and CIR models may readily be found on the Internet and such calibration is much

easier.

6 Conclusions

After considering the results and the discussion in previous sections one conclusion is that the Vasicek and CIR models do not describe the stochastic process accurately. The stochastic process of the Euribor 6-month rate that is and this is concluded since the null hypotheses were rejected when performing the chi-square tests for those models. The findings do however give support to the extended Short Rate Model and the null hypothesis that this model describes the stochastic process was not rejected. This indicates a better fit.

Much more effort is needed for calibrating a three-factor model compared to a one-factor model. In comparison it is very easy to write a program for calibrating the Vasicek and CIR models. And writing a program for calibration by using EMM is indeed non-trivial. It is important to consider a fast programming language like C++ for writing a program for performing EMM before choosing something else. It is important to take the drawbacks of Matlab into consideration before possibly choosing this instead. The Nelder-Mead algorithm is more suitable for EMM and makes it easier to find the optimum. This concludes this report.

Appendix: Lévy processes

This Appendix covers some of the essentials of the mathematical theory of Lévy processes. The sources of this theory are Tankov [5] and Billingsley [6], in the lecture notes by Tankov a presentation of the French mathematician Paul Lévy may also be found. He gave name to the Lévy process.

There is a type of function that needs to be defined now since this type is important in this context. This is known as *càdlàg* (an acronym for "continu à droite, limites à gauche" which is in French) and the following definition is found in Billingsley [6] but with a slightly different notation:

Definition 1. *Let $D = D[0, 1]$ be the space of real functions x on $[0, 1]$ that are right-continuous and have left-hand limits:*

- (i) For $0 \leq t < 1$, $x(t+) = \lim_{s \rightarrow t+} x(s)$ exists and $x(t+) = x(t)$.
- (ii) For $0 < t \leq 1$, $x(t-) = \lim_{s \rightarrow t-} x(s)$ exists. \square

Stating a definition of the Lévy process as a next step in this presentation is natural. This definition will also make more sense after the term 'càdlàg' has been defined:

Definition 2. *A stochastic process X is a Lévy process if it is càdlàg, satisfies $X_0 = 0$ and possesses the following properties:*

- Independent increments.
- Stationary increments.

From these properties it follows that

- X is continuous in probability: $\forall \epsilon > 0, \lim_{s \rightarrow 0} P[|X_{s+t} - X_t| > \epsilon] = 0$.
- At any fixed time, the probability of having a jump is zero: $\forall t, P[X_{t-} = X_t] = 1$. \square

Further, Tankov states that Lévy processes are essentially processes with 'jumps', because it can be shown that any Lévy process which has almost surely continuous trajectories is a Brownian motion with drift. This follows from the proposition below:

Proposition 1. *Let X be a continuous Lévy process. Then there exist $\gamma \in \mathbb{R}^d$ and a symmetric positive definite matrix A such that*

$$X_t = \gamma t + W_t$$

where W is a Brownian motion with covariance matrix A . \square

Those readers who want to immerse in their studies in mathematical theory can find a proof in the lecture notes by Tankov [5]. This proof will however not be presented in this report and this also applies to all other propositions in this text. Instead an example of a Poisson process follows next:

Definition 3. *Let $(\tau_i)_{i \geq 1}$ be a sequence of exponential random variables with parameter λ and let $T_n = \sum_{i=1}^n \tau_i$. Then the process*

$$N_t = \sum_{n \geq 1} 1_{t \geq T_n} \tag{5}$$

is called the Poisson process with parameter (or intensity) λ . \square

The following proposition is important because several characteristics of the Poisson process can be identified by it:

Proposition 2 (Properties of the Poisson process).

1. For all $t \geq 0$, the sum in (5) is finite a.s. (almost surely).
2. The trajectories of N are piecewise constant with 'jumps' of size 1 only.
3. The trajectories are càdlàg.
4. $\forall t > 0$, $N_{t-} = N_t$ with probability 1.
5. $\forall t > 0$, N_t follows the Poisson law with parameter λt :

$$P[N_t = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

6. The characteristic function of the Poisson process is

$$E[e^{iuN_t}] = \exp(\lambda t(e^{iu} - 1)).$$

7. The Poisson process is a Lévy process. \square

The Poisson process counts the events with exponential interarrival times.

The Poisson process itself is not adequate to model such things as the stochastic processes of asset prices or interest rates. The reason for this is the restriction that the jump size is always equal to 1 and clearly other values must be allowed in such models. However, after studying the following definition it can be concluded that the Poisson process can be used as a building block to construct richer models:

Definition 4 (*Compound Poisson process*). *The compound Poisson process with jump intensity λ and jump size distribution μ is a stochastic process $(X_t)_{t \geq 0}$ defined by*

$$X_t = \sum_{i=1}^{N_t} Y_i$$

where $\{Y_i\}_{i \geq 1}$ is a sequence of independent random variables with law μ and N is a Poisson process with intensity λ independent from $\{Y_i\}_{i \geq 1}$. \square

As Tankov concludes, a compound Poisson process is a piecewise constant process which jumps at jump times of a standard Poisson process and whose jump sizes are i.i.d. random variables with a given law.

References

- [1] Olivier Ledoit, Michael Wolf, *A well-conditioned estimator for large-dimensional covariance matrices*. Journal of Multivariate Analysis 88, 2004.
- [2] Torben G. Andersen, Luca Benzoni and Jesper Lund, *Stochastic Volatility, Mean Drift, and Jumps in the Short-Term Interest Rate*. April 2004.
- [3] John C. Cox, Jonathan E. Ingersoll Jr. and Stephen A. Ross, *A Theory of the Term Structure of Interest Rates*. Econometrica 53, 1985.
- [4] Edward Allen, *Modeling with Itô Stochastic Differential Equations*. Springer, Dordrecht in the Netherlands, 2007.
- [5] Peter Tankov, *Financial Modeling with Jump Processes* (Lecture notes).
- [6] Patrick Billingsley, *Convergence of Probability Measures*. John Wiley & Sons Inc., New York 1999, Second Edition.
- [7] John Hull, *Options, Futures and Other Derivatives*. Pearson Education Inc., New Jersey 2009, Seventh Edition.
- [8] Torben G. Andersen, Hyung-Jin Chung and Bent E. Sørensen, *Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study*. Journal of Econometrics 91, 1999.