CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

Modeling Swedish Interest Rates by Simulated Maximum Likelihood

ANDRÉ AYDIN

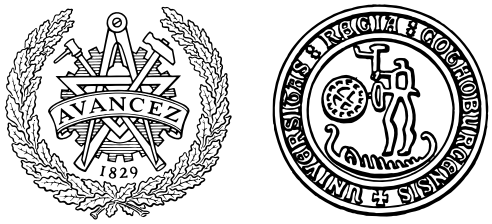
Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2007

Thesis for the Degree of Master of Science
(20 credits)

Modeling Swedish Interest Rates by Simulated Maximum Likelihood

André Aydin

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden
Göteborg, May 2007

Modeling Swedish Interest Rates by Simulated Maximum Likelihood

André Aydin

18th September 2007

Abstract

We[1] investigate a collection of continuous time stochastic differential equation (SDE) models for short term interests with respect to their match to Swedish 3-month Treasure bill rates.

It is well-known that estimation of parameters of SDE models are difficult in the absence of closed form expresion for the transition density. In this thesis we use simulated maximum likelihood (SML) to estimate parameters.

Our conclusion is that SDE models with the most general volatility specification gives the best fit to the data.

[1] Throughout this thesis we use the conventions We = I and we = I.

Acknowledgements

I would like to thank my supervisor Patrik Albin for his support during my work. He his an ideal supervisor for a student. I would also like to thank Jan Lennartsson, Jimmy Myhrman, Nikola Vorkapic, Ottmar Cronie, Anders Runeson and Thomas Ericsson for their help. On the personal level I would like to thank my family and friends who always are there for me.

Contents

1 Introduction

The dynamic behavior of interest rates is an important aspect of the financial environment. This is so because financial institutions and intermediaries such as banks and insurance companies has liabilities and assets that depend on interest rates. If these actors have poor information about interest rates, it can have very dramatic effects on their activities.

For example, in the savings and loan market in the 1980s in USA, interest rates increased rapidly and billions of dollars in profits turned over to huge losses and numerous insolvencies. The problem was that assets were primarily long term and fixed-rate mortgage while most liabilities were short term demand deposits. So, when interest rates increased rapidly the difference between the long term earnings used to pay the short term liabilities became very small or even negative. Effects like this make it very important to develop models such that risk management tools that can be used to decrease risks, see Ahlgrim et al. [1].

Nowadays SDE are commonly used to model financial and economic data. Estimation of parameters of SDE models has proved to be difficult. This is so because closed form expressions of transition densities are not known in general, so that the log likelihood function isn't either. However, recently new methods have been developed to approximate transition densities, such as a generalized method of moment (GMM) by Chan et al. [9], simulated method of moments by Duffie and Singleton, efficient method of moments (EMM) by Gallant and Tauchen, and approximation of transition densities using Hermite function by Ait-Shalia [3]. Another important method is Markov Chain Monte Carol (MCMC), a method which in turn is closely related to the simulated maximum likelihood (SML) method by Pedersen [16] that we will use in this thesis.

The idea behind SML is to approximate the unknown transition density by simulating $M-1$ unobserved points between two observations x_t and x_s say, where $s < t$. Since the simulated points are unobserved they have to be integrated out by means of th Chapman-Kolmogorov theorem. Choosing a big M gives better approximations, but at a cost of a higher dimensional integrals, since $M-1$ unobserved points correspond to a $M-1$ dimensional integral. In general, the only feasable way to calculate this integral is by Monte Carlo integration. There we face a trade off between the accuracy of the approximation and the simulation time required.

There are two approximation errors for the SML method, namely a bias error due to approximation of the transition density by for example an Euler scheme, and a variance error due to the Monte Carlo approximation of integrals. To make the method faster one employs different acceleration

methods, such as importance sampling for the Monte Carlo method and a finer approximation of the transition density.

The purpose of this thesis is to learn how to estimate parameters for SDE models when the transition density is not known in closed form. Then we apply our found method to fit weekly Swedish 3-month Treasure bill rates to the following nested SDE models proposed by Ait-Shalia [2]:

$$dX = \left(\alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 / X\right)\, dt + \sqrt{\beta_0 + \beta_1 X + \beta_2 X^{\beta_3}}\, dW.$$

The thesis is structured as follows: In Section 2 we introduce the SML method and some of its asymptotic properties.

In Section 3 we present the Monte Carlo integration method and methods for variance reduction.

In Section 4 we describe how to combine the simulation approach SML with acceleration methods.

In section 5 we evaluate the performance of our accelerated SML method by applying it to the CIR model. As this model has analytic solutions we can compare the results of our method with exact results.

In Section 6 we give a review of different interest rate models.

In Section 7 we present our data set and the test statistics we use to evaluate our fitted interest rate models.

In Section 8 we give the results of our fit of interest rate models to the Treasure bill rates data set.

Finally, in Section 9 we make conclusions from our work.

2 Simulated maximum likelihood (SML)

2.1 The SML method

The idea of the SML method goes as follows: Consider a time homogeneous SDE model

$$dX = \mu(X, \theta)\, dt + \sigma(X, \theta)\, dW, \quad X(t_0) = X_0. \tag{1}$$

Here θ is a parameter vector and W a standard Brownian motion.

Suppose that we have observed the values $\{X_i = X(t_i), i = 0, \ldots, n\}$ of the solution to (1). Let the transition density for $(X(t, \theta)|X(s, \theta))$ be given by $p(x_t, t; x_s, s, \theta)$ for $s < t$, so that (as we are working with a Markov process) the log likelihood is given by

$$l_n(\theta) = \sum_{i=1}^{n} \log\left(p(X_i, t_i; X_{i-1}, t_{i-1}, \theta)\right). \tag{2}$$

In many cases of practical interest a closed form expression for the transition density is not available, and thus this is also so for the log likelihood. Maximum likelihood (ML) estimators and with their statistical properties then cannot be calculated in a straightforward naive manner.

The first order Euler approximation of the solution to (1) is given by

$$dX_{i+1} = \mu(X_i, \theta)\Delta_i + \sigma(X_i, \theta)\Delta_i^{1/2}\epsilon_i, \quad \Delta_i = t_{i+1} - t_i, \quad \epsilon_i \sim N(0, 1). \tag{3}$$

The Euler approximation (3) corresponds to the first order approximated transition density

$$p^{(1)}(x_t, t; x_s, s, \theta) = \phi\left(x_t, x_s + \mu(x_s)(t - s), \sigma^2(x_s)(t - s)\right), \tag{4}$$

where $\phi(x; \mu, \sigma^2)$ is a Gaussian density probability density function with expected value μ and variance σ^2. From now on we call approximative transition densities *sub densities*.

The approximations (3) and (4) are good only if the length $h = t - s$ of the interval $[s, t]$ is small enough, and in the limit as $h \to 0$ the approximation becomes exact. To get a better approximation one may partition the interval in subintervals $s = \tau_0 < \tau_1 < \ldots < \tau_M = t$, all with equal length $\delta = \Delta/M$. When the process values $X(\tau_1), \ldots X(\tau_{M-1})$ has not been observed they must be integrated out using the Chapman-Kolmogorov theorem to get an approximated likelihood.

Theorem 1 (Chapman-Kolmogorov). *For the transition density function $p(y,t;x,s)$ of a diffusion process we have*

$$p(y,t;x,s) = \int_{-\infty}^{\infty} p(y,t;z,u)\, p(dz,u;x,s) \quad \textit{for } s<u<t. \tag{5}$$

Using (5) we can calculate the approximated transition density $p^{(M)}(x_t,t;x_s,s,\theta)$ recursively: The probability that $X(\tau_M)=x_t$ conditional on that $X(\tau_0)=x_s$ passing through $X(\tau_1)=u_1,\ldots,X(\tau_{M-1})=u_{M-1}$ can by (5) be written as

$$p^{(M)}(x_t,t;x_s,s,\theta) = \int_{\mathbb{R}} p(x_t,\tau_M;u_{M-1},\tau_{M-1},\theta)\, \underbrace{p(u_{M-1},\tau_{M-1};x,s,\theta)}_{=I}\, du_{M-1}. \tag{6}$$

As the one step ahead transition density is given by (4) we can replace $p(x_t,\tau_M;u_{M-1},\tau_{M-1},\theta)$ in (6) by $p^{(1)}(x_t,\tau_M;u_{M-1},\tau_{M-1},\theta)$. Treating I in (6) in the same way we get

$$\begin{aligned} &p(u_{M-1},\tau_{M-1};x,s,\theta) \\ &\qquad = \int_{\mathbb{R}} p(u_{M-1},\tau_{M-1};u_{M-2},\tau_{M-2},\theta)\, p(u_{M-2},\tau_{M-2};x,s,\theta)\, du_{M-2}. \end{aligned}$$

Here we approximate $p(u_{M-1},\tau_{M-1};u_{M-2},\tau_{M-2},\theta)$ by $p^{(1)}(u_{M-1},\tau_{M-1};u_{M-2},\tau_{M-2},\theta)$ ... . Repeating this procedure $M-1$ times we end up with the approximate transition density

$$\begin{aligned} p(x_t,t,x_s,s,\theta) &\approx p^{(M)}(x_t,t,x_s,s,\theta) \\ &= \int_{\mathbb{R}^{M-1}} \prod_{m=0}^{M-1} p^{(1)}(u_{m+1},\tau_{m+1};u_m,\tau_m,\theta)\, d\lambda(u_1,\ldots,u_{M-1}), \end{aligned}$$

where λ denotes Lebesgue measure, $u_0=x_s$ and $u_M=x_t$. In general, the only feasible way to calculate this integral numerically is by means of Monte Carlo integration.

2.2 Asymptotic condition

The SML method has some very nice asymptotic properties, two of which will be briefly reviewed here. For more detail information the interested reader is refered to Brandt and Santa-Clara [6], Durham and Gallant [12] and Pedersen [15, 16].

Assume that the densities $p(\cdot, t; x_s, s, \theta)$ and $p^{(M)}(\cdot, t; x_s, s, \theta)$ exist for all $s < t$, $x_s \in \sup[X(s, \theta^\circ)]$, $\theta \in \Theta$ and $M \geq 1$, where θ° is the unknown parameter vector and $X_i = X[t_i; \theta^\circ]$ are observations generated from the Euler approximation (4). Denote by $P_{\theta^\circ, n}$ the probability measure induced by $\{X_0, \ldots, X_n\}$, by $l_n^{(M)}(\theta)$ the approximated log likelihood

$$l_n^{(M)}(\theta) = \sum_{i=1}^{n} \log \left(p^{(M)}(X_i, t_i; X_{i-1}, t_{i-1}, \theta) \right)$$

and by $l_n(\theta)$ the exact likelihood

$$l_n(\theta) = \sum_{i=1}^{n} \log \left(p(X_i, t_i; X_{i-1}, t_{i-1}, \theta). \right)$$

Pedersen [16] shows under some special conditions that

$$\lim_{M \to \infty} p^{(M)}(\cdot, t; x_s, s, \theta) = p(\cdot, t; x_s, s, \theta) \quad \text{in } \mathbb{L}^1,$$

and that

$$\lim_{M \to \infty} l_n^{(M)}(\theta) = l_n(\theta) \quad \text{in probability under } P_{\theta^\circ, n} \text{ for } \theta \in \Theta.$$

3 Monte Carlo integration

To calculate the expectation of a function in one and multidimensional space approximatively one can use the Monte Carlo method. Here we give a brief introduction to this method.

3.1 The Monte Carlo method

Consider a random variable X with cummulative probability distribution function F_X. The expected value of a function $h(X)$ of X is given by

$$\mathbf{E}\{h(X)\} = \int h(x)\, dF_X(x).$$

We get the Monte Carlo estimate of $\mathbf{E}\{h(X)\}$ by taking a sample of $\{X_1, \ldots, X_n\}$ of observations of X and compute the mean of h over the sample, that is,

$$\tilde{h}_n(X) = \frac{1}{n} \sum_{i=1}^{n} h(X_i).$$

By the strong law of large numbers we have $\tilde{h}_n(X) \to \mathbf{E}\{h(X)\}$ as $n \to \infty$ almost surely. Note that the estimator $\tilde{h}_n(X)$ is unbiased, by a trivial calculation.

As always with simulations, we want our method to be computationally efficient. In our case the standard error is given by $\omega/\sqrt{n}$ where ω is the standard deviation of $h(X)$. There exist many methods for accelarating the convergence, some of which are presented in Cairns [8]. In this thesis we will use a random number scheme and importance sampling.

3.2 Random number scheme

Suppose that we have to generate an i.i.d. sample $\{X_1, \ldots, X_{2n}\}$ from a standard normal distributed random variable X. Instead of a sample of length $2n$ we generate an initial i.i.d. sample $\{X_1, \ldots, X_n\}$ and get the final sample by setting $\tilde{X}_1 = -X_1, \ldots, \tilde{X}_n = -X_n$. Without a random number scheme, the variance is given by $\sigma^2/(2n)$ where σ is the standard deviation of $h(X)$. But if we use the n variates X together with their n antithetic variates $\tilde{X}$ and set

$$\overline{h}(X) = \frac{1}{2n} \sum_{i=1}^{n} \frac{h(X_i) + h(\tilde{X}_i)}{2}$$

we get

$$\begin{aligned}\mathbf{Var}\{\overline{h}(X)\} &= \frac{\mathbf{Var}\{h(X)+h(\tilde{X})\}}{4n} \\ &= \frac{\mathbf{Var}\{h(X)+\mathbf{Var}\{h(\tilde{X})\}+2\,\mathbf{Cov}\{h(\tilde{X}),h(X)\}}{4n}.\end{aligned}$$

Therefore the variance will be smaller if $\mathbf{Cov}\{h(\tilde{X}),h(X)\} < 0$. This will typically (but not strictly always) be the case with our choice $\tilde{X} = -X$.

In this thesis we will use normalized variates $\hat{W}$, that is,

$$(\hat{W}_1,\ldots,\hat{W_{M-1}}) = \left(\frac{1}{M-1}\sum_{m=1}^{M-1} W_i^2\right)^{-1/2}(W_1,\ldots,W_{M-1}).$$

By normalization we can control the jaggedness of the sample path by letting each vector of Gaussian increments have sample variance one.

Since we maximize our log likehood function numerically we must calculate it for different parameter values and for each calculation we use the same random numbers. This will give approximated densities that are smooth functions of the parameters and will thus be of help when optimizing. The asymptotics results are also based on using the same random numbers repetedly.

3.3 Importance sampling

The idea of importance sampling is to draw random numbners from a part of a distribution that contribute the most to the integration. This is based on a trick to get an integrand that varies less than the original one over the integration region, and goes like this: For a random variable X with probability density function f and a function r, the expectation $\mathbf{E}\{r(X)\}$ can we written as

$$\mathbf{E}\{r(X)\} = \int r(x)f(x)\,dx = \int h(x)\,\frac{f(x)r(x)}{h(x)}\,dx = \mathbf{E}_h\Big\{\frac{r(X)f(X)}{h(X)}\Big\}$$

for probability density functions h, where the notation $\mathbf{E}_h$ means expectation with respect to the density h. This can be seen as changing of measure. The new Monte Carlo estimation is the given by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)r(X_i)}{h(X_i)} \quad \text{where } X_i \text{ have density } h. \tag{7}$$

It can been shown that the variance of the estimator (7) is minimized when h is proportional to rf, see Andersen [5]. However, there are some other properties the importance sampler should have possess as well: It should be easy to simulate values from h and it should be easy to compute $h(x)$ for all x we realize. Also, as rf and h roughly have the same shape we get trouble if the tails of h get thinner faster than those of rf, that is, if we realize a value X_i from the far tails the ratio $r(X_i)f(X_i)/h(X_i)$ will be much larger than values of $r(X)f(X)/h(X)$ that we actually observe.

In our particular application the importance sampling will work as follows: Fix $s < t$, x_t, x_t, θ and M and let $q(u_1, \ldots, u_{M-1}, \theta)$ denote a probability density on $\mathbb{R}^{M-1}$. Further, let $\{(u_{k,1}, \ldots, u_{k,M-1}), k = 1, \ldots, K\}$ be independent observations from q. We approximate (6) by

$$\begin{aligned} & p^{(M,K)}(x_t, t; x_s, s, \theta) \\ &= \frac{1}{K} \sum_{k=1}^{K} \frac{1}{q(u_{k,1}, \ldots, u_{k,M-1}, \theta)} \prod_{m=1}^{M} p^{(1)}(u_{k,m}, \tau_m; u_{k,m-1}, \tau_{m-1}, \theta), \end{aligned} \tag{8}$$

where $u_{k,0} = x_s$ and $u_{k,M} = x_t$ for all k. The density q is given by

$$\prod_{m=1}^{M-1} p_{\text{imp}}(u_m, \tau_m; u_{m-1}, \tau_{m-1}, \theta)$$

where $p_{\text{imp}}(u_m, \tau_m; u_{m-1}, \tau_{m-1}, \theta)$ is the transition density for the importance sampler.

If it holds that

$$\mathbf{E}\left\{ \frac{1}{q(U_1, \ldots, U_{M-1}, \theta)} \prod_{m=1}^{M} p^{(1)}(U_m, \tau_m; U_{m-1}, \tau_{m-1}, \theta) \right\} < \infty,$$

then the strong law of large numbers gives

$$\lim_{K \to \infty} \left| p^{(M,K)}(x_t, t; x_s, s, \theta) - p^{(M)}(x_t, t; x_s, s, \theta) \right| = 0 \quad \text{almost surely.}$$

4 SML with acceleration methods

In this section we introduce different simulation approaches for estimating the parameters of our processes due to Pedersen and Santa Clara (independent of each other). Then we describe acceleration methods to make speed up the method. Durham and Gallant [12] investigated different bias and variance reduction techniques on a one factor CIR process, and we will use the acceleration techniques that gave the best results according to their investigation.

4.1 Lamperti transformation

Before we present the bias and variance reduction techniques we transform the SDE (1) to make the diffusion term constant, following Shoji and Ozaki [20] and Ait-Shalia [3]. This will improve both the approximation of (4) and the acceleration of convergence of the Monte Carlo estimation of (6), because the transformed SDE is closer to a Gaussian process. The transformation we will use is often called Lamperti transformation and is given by

$$Y = G(X) = \int^{X_t} \frac{du}{\sigma(u;\theta)}, \tag{9}$$

where $G(X)$ satisfies $G'(X) = 1/\sigma(X;\theta)$. By Ito's lemma we have

$$dY = G'(X)\,dX + \frac{1}{2}G''(X)\sigma^2(X)\,dt,$$

which after inserting dX_t can be rewritten as

$$dY = \left(\frac{\mu(G^{-1}(Y))}{\sigma(G^{-1}(Y))} - \frac{1}{2}\sigma'(G^{-1}(Y))\right)dt + dW.$$

In some cases the integral in (9) is defined as $-Y$ to get $Y > 0$. If, for example, $\sigma(X;\theta) = X^p$ and $p > 1$, then we get $Y = G(X) = -X^{1-p}/(1-p)$. The transformed SDE in this case is equal to

$$d(-Y) = -\left(\frac{\mu(G^{-1}(Y))}{\sigma(G^{-1}(Y))} - \frac{1}{2}\sigma'(G^{-1}(Y))\right)dt - dW.$$

The relation between the new transition density p_Y and p is given by

$$p(x_t, t; x_s, s, \theta) = p_Y(y_t, t; y_s, s, \theta)\left|\frac{dy}{dx}\right|,$$

where $|dy/dx|$ is the Jacobian.

It is not always the case that the Lamperti transform has a closed form expression, so that one has to calculate the transformation numerically. This can be done by defining an equidistant grid $x_1, \ldots, x_g$ and calculate the corresponding values $y_1, \ldots, y_g$ from

$$\int_c^{X_t} du/\sigma(u;\theta)$$

by, for example, the Runge-Kutta method. The constant c has to be chosen less than x_1. For an observation x_0 we find the interval where it belongs, say $[x_i, x_{i+1}]$ and then interpolate between $[y_i, y_{i+1}]$. The forward transformation $Y = G(X)$ is straightforward and fast compared to the backward transformation $X = G^{-1}(Y)$. First we have to locate the interval for an observation y_0 by, for example, using bisection search, and this has to be done for all unobserved data which requires extra simulation time.

4.2 Pedersen and Santa Clara approach

The importance sampler used by Clara and Pedersen is the Euler scheme: For given $x_s = X(s)$ and $x_t = X(t)$ this sampler is given by the recursion

$$u_{m+1} = u_m + \mu(u_m;\theta)\delta + \sigma(u_m;\theta)\delta^{1/2}W_{m+1}, m = 0, \ldots, M-2,$$

where $W_1, \ldots, W_{M-1}$ are i.i.d. standard normal, $u_0 = x_s$ and $\delta = (t-s)/M$.

With the above setup (8) simplifies a lot. The first $M-1$ factors in the numerator cancel out since

$$q(u_{k,1}, \ldots, u_{k,M-1}, \theta) = \prod_{m=0}^{M-2} \phi(u_{k,m+1}; u_{k,m} + \mu(u_{k,m};\theta)\delta, \sigma^2(u_{k,m};\theta)\delta),$$

so that we end up with

$$p^{(M,K)}(x_t, t; x_s, s, \theta) = \frac{1}{K}\sum_{k=1}^{K} p^{(1)}(x_t, t; u_{k,M-1}, \tau_{M-1}, \theta),$$

where we draw $\{u_{k,M-1}, k = 1, \ldots, K\}$ from the $(M-1)$:st component of q. Durham and Gallant [12] show that this approach does not work well becuase samples are drawn from regions where the integrand has little mass.

4.3 Bias reduction: Shoji and Ozaki approach

We have two sources of approximation errors, namely a bias error due to approximation of the subdensity and a variance error from the Monte-Carlo integration.

We will use a method to decrease the bias error proposed by Shoji and Ozaki [20] called *local linearization method*. Their method approximates a non-linear SDE with a linear SDE. Since the linear SDE has a Gaussian transition density the parameters can be estimated by the ML method. They show the method is more efficient and much more numerically stable than the Euler approximation. To use the method one need constant volatility.

Given the SDE

$$dX = \mu(X,t)\,dt + \sigma\,dW$$

and a fixed x_s, an application of Ito's lemma to $\mu(X,t)$ gives

$$d\mu(X,t) = \left(\frac{\sigma^2}{2}\frac{\partial^2\mu(X,t)}{\partial X^2} + \frac{\partial u(X,t)}{\partial t}\right)dt + \frac{\partial u(X,t)}{\partial X}\,dX. \tag{10}$$

Next we linearize u wrt. x_t and t by assuming that

$$\frac{\partial^2\mu(x,t)}{\partial x^2}, \quad \frac{\partial u(x,t)}{\partial x} \quad \text{and} \quad \frac{\partial u(x,t)}{\partial t}$$

are constant. If t belongs to a small interval $[s, s+\Delta)$ this assumption will be reasonable. In this manner (10) can be approximated as

$$u(x_t,t) - u(x_s,s) = \left(\frac{\sigma^2}{2}\frac{\partial^2\mu(x,t)}{\partial x^2} + \frac{\partial u(x,t)}{\partial t}\right)(t-s) + \frac{\partial u(x,t)}{\partial x}(x_t - x_s).$$

We can write this equation as

$$u(x_t,t) = (L_s x_t + M_s t + N_s)\,dt + \sigma\,dW, \tag{11}$$

where

$$\begin{aligned} L_s &= \frac{\partial u(x_s,s)}{\partial x}, \\ M_s &= \frac{\sigma^2}{2}\frac{\partial^2\mu(x_s,s)}{\partial x^2} + \frac{\partial u(x_s,s)}{\partial t}, \\ N_s &= u(x,s) - \frac{\partial u(x_s,s)}{\partial x}x_s - \left(\frac{\sigma^2}{2}\frac{\partial^2\mu(x_s,s)}{\partial x^2} + \frac{\partial u(x_s,s)}{\partial t}\right)s. \end{aligned} \tag{12}$$

Next we transform (11) using Girsanov,s theorem, which gives

$$dx_t = L_s x_t\,dt + \sigma\,d\tilde{W}, \tag{13}$$

where

$$\tilde{W} = W_t - \int_s^t \gamma(u)\, du \quad \text{and} \quad \gamma(u) = -\frac{1}{\sigma}\left(M_s u + N_s\right).$$

Using Ito's lemma on $y_t = f(x_t, t) = e^{-L_s t} x_t$ we can solve (13) and the solution is given by

$$\begin{aligned} & y_t y_s + \sigma \int_s^t e^{-L_s u}\, d\tilde{B}_u \\ & \quad = y_s + \int_s^t \left(M_s u + N_s\right) e^{-L_s u}\, du + \sigma \int_s^t e^{-L_s u}\, dB_u. \end{aligned}$$

Putting things together we end up with a discretized process x_t given by

$$x_t = x_s + \left(\frac{u(x_s, s)}{L_s} + \frac{M_s}{L_s}\right)\left(e^{L_s(t-s)} - 1\right) + M_s\left(t - s\right) + \sigma \int_s^t e^{L_s(t-u)}\, dB_u,$$

where L_s and M_s are given by (12). The solution to our new SDE is an Ohrstein-Uhlenbeck process with subdensity given by

$$p(x_t, t; x_s, s, \theta) = \phi(x_t; \hat{\mu}, \hat{\sigma}^2)$$

where

$$\begin{aligned} \hat{\mu} &= x_s + \frac{\mu(x_s)}{\mu'(x_s)} K + \frac{\sigma^2 \mu''(x_s)}{2[\mu'(x_s)]^2}\left[K - \mu'(x_s)\Delta\right], \\ K &= e^{\mu'(x_s)\Delta} - 1, \\ \hat{\sigma}^2 &= \frac{\sigma^2}{2\mu'(x_s)}\left(e^{2\mu'(x_s)\Delta} - 1\right). \end{aligned}$$

Example of other subdensities is Elerian's based on a Milstein scheme. Kessler uses higher order Ito Taylor approximations. See Durham and Gallant [12] for more information.

4.4 Variance reduction by importance sampling

The importance sampler in thesis samples u_{m+1} from a Gaussian density function

$$\phi\left(u_{m+1}; u_m + \tilde{\mu}\delta, \tilde{\sigma}^2\delta\right),$$

where

$$\delta = \frac{t-s}{M}, \quad \tilde{\sigma}^2 = \frac{M - m - 1}{M - m}\hat{\sigma}^2 \quad \text{and} \quad \tilde{\mu} = \frac{u_M - u_m}{t - \tau_m}.$$

The resulting process is called a modified Brownian bridge. This process actually is a Brownian bridge if σ is constant, as

$$u_{m+1} = u_m + \frac{u_M - u_m}{t - \tau_m}\delta + \sqrt{\frac{M-m-1}{M-m}}\sigma\delta^{1/2}W_m$$

where W_m are i.i.d. standard normal.

Examples of other importance samplers are given in Durham and Gallant [12] and Pastorello and Rossi [17].

4.5 Log likelihood bias reduction

It is well-known that the approximated log likelihood $\log(p^{(M,K)}(x_t, t; x_s, s, \theta))$ is a biased estimator of the real log likelihood $\log(p(x_t, t; x_s, s, \theta))$. We use a bias reduction from Bruche [7] given by

$$\log\left(p^{(M,K)}(x_t, t; x_s, s, \theta)\right) = \log\left(\frac{1}{K}\sum_{k=1}^{K}\omega_k\right) + \frac{s_\omega^2}{2K}\left(\frac{1}{K}\sum_{k=1}^{K}\omega_k\right)^{-2}$$

where

$$\omega_k = \frac{1}{q(u_{k,1}, \ldots, u_{k,M-1}, \theta)}\prod_{m=1}^{M} p^{(1)}(u_{k,m}, \tau_m; u_{k,m-1}, \tau_{m-1}, \theta).$$

and s_ω is the standard deviation of ω. This transformation reduces the bias to $O(K^{-3/2})$.

In practice, our results turned out to be nearly the same with or without the log likelihood bias reduction.

5 SML applied to the CIR process

We test the SML method on the famous CIR interest rate model of Cox, Ingersoll and Ross [10]. Since the CIR model has a closed form transition density and thus a closed form log likelihood a test can be done.

The CIR process is given by

$$dX = \alpha(\theta - X)\,dt + \sigma\sqrt{X}\,dW, \tag{14}$$

where $\alpha, \theta, \sigma > 0$ are parameters such that $2\alpha\theta/\sigma^2 \geq 1$. In Figure 1 below a trajectoria of the CIR process is depicted.

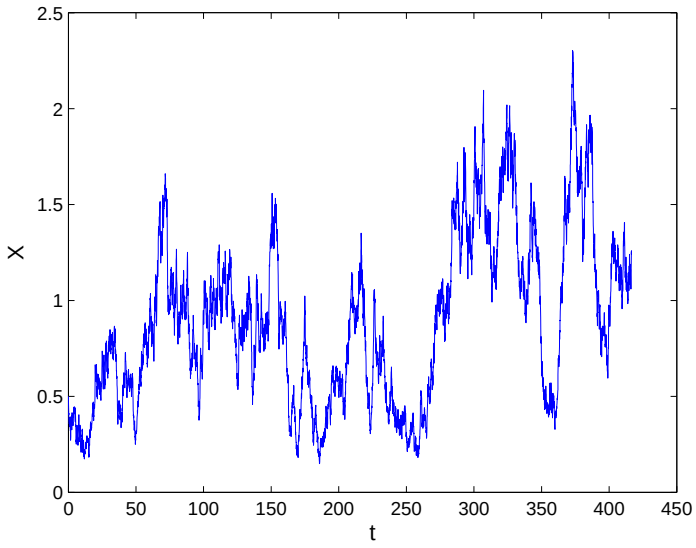


Figure 1: CIR trajectory.

If we let

$$\begin{aligned} c &= \frac{2\alpha}{\sigma^2\left(1 - e^{-\alpha(t-s)}\right)}, \\ q &= \frac{2\alpha\theta}{\sigma^2} - 1, \\ u &= cx(s)e^{-\alpha(t-s)}, \\ v &= cx(t), \\ Z &= 2cX, \end{aligned}$$

then $(Z_t|Z_s)$ has a non-central chi-squared distribution with $4\alpha\theta/\sigma^2$ degrees of freedom and non-centrality parameter $Z_s e^{\theta(t-s)}$. This is to say that

$$p(x_t, t; x_s, s, \theta) = ce^{-u-v}(v/u)^{q/2} I_q(2\sqrt{uv})$$

where $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

The Lamperti transform for the CIR model is given by $Y = 2\sqrt{X}/\sigma$. Applying Ito's lemma and inserting in (14) we get the transformed equation

$$dY = \left(-\frac{\alpha Y}{2} - \frac{1}{2Y}\left(1 - \frac{4\alpha\theta}{\sigma^2}\right)\right) dt + dW.$$

The relation between the transformed transition density $p_Y(y_t, t; y_s, s, \theta)$ and the original one is given by

$$p(x_t, t; x_s, s, \theta) = p_Y(y_t, t; y_s, s, \theta) \left|\frac{dy}{dx}\right| = \frac{p_Y(y_t, t; y_s, s, \theta)}{\sigma\sqrt{x}}.$$

Figures 2-5 below show the results when approximating transition densities for different M and K with parameters $dt = 1/12$, $(\alpha, \theta, \sigma) = (0.6, 0.05, 0.15)$, $X_s = 0.1$ and $X_t = [0.05, 0.15]$. The top panels display the approximated likelihood (circles) and exact log likelihood (line) for 124 repetitions. The bottom panels display the error between the approximated and exact log likelihood (line) together with the median and interquartiles (both dashed).

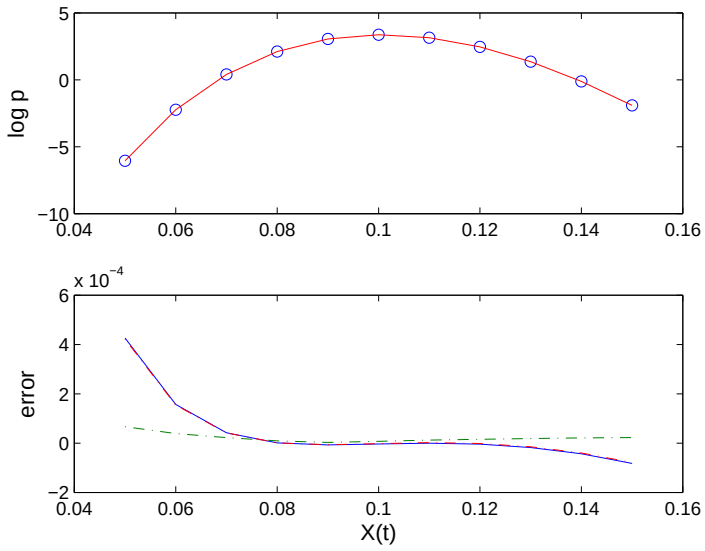


Figure 2: Log likelihoods and approximation errors for $M = 8$, $K = 32$.

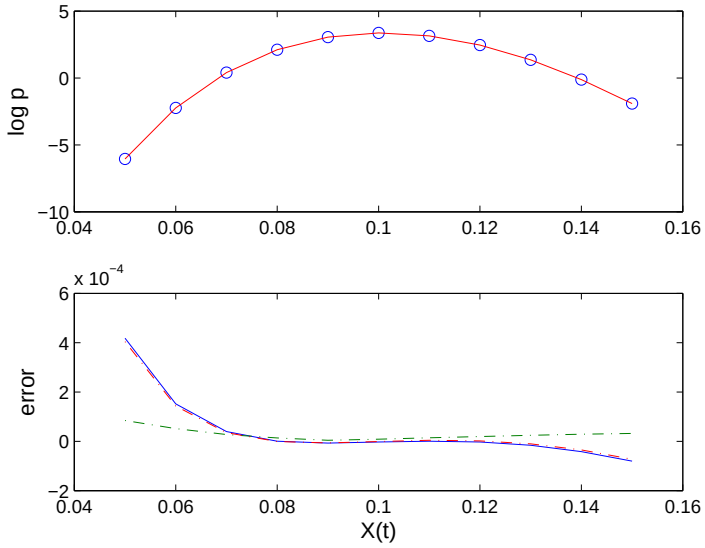


Figure 3: Log likelihoods and approximation errors for $M = 8$, $K = 16$.

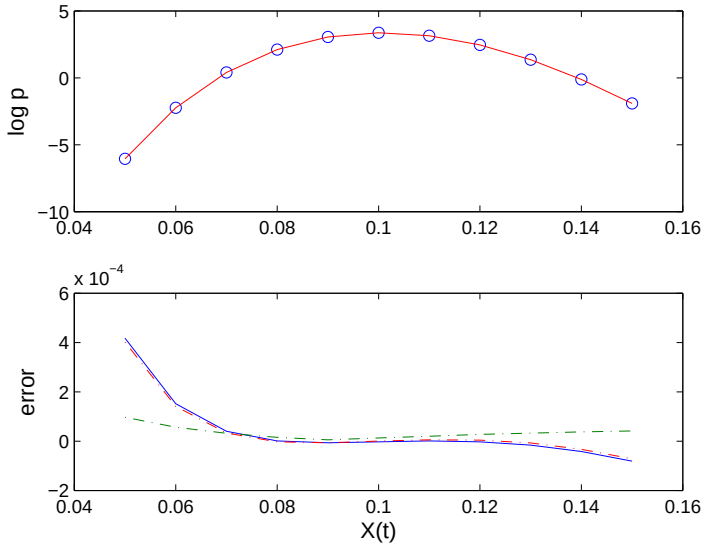


Figure 4: Log likelihoods and approximation errors for $M = 8$, $K = 8$.

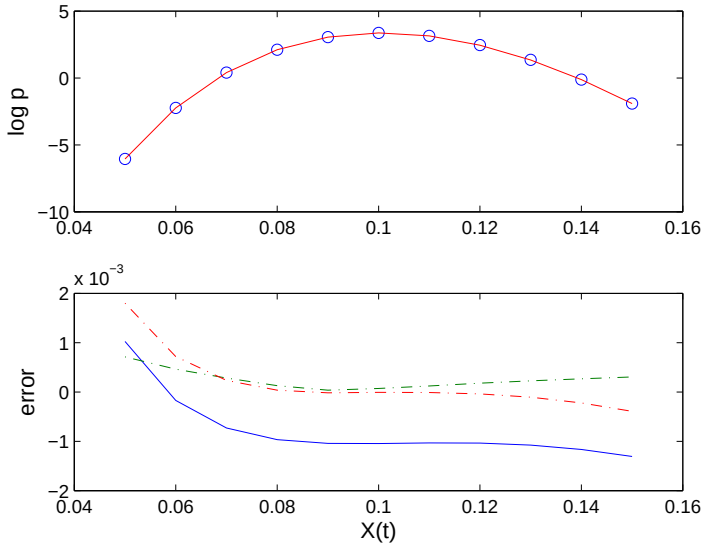


Figure 5: Log likelihoods and approximation errors for $M = 4$, $K = 8$.

We would like to see how well we can approximate the log likelihood for a larger sample. Therefore we use the same settings as above, except for the number of observations where we use $n = 5000$. The sample comes from a explicit discrete scheme proposed by Alfonsi [4], given by

$$x_{t_{i+1}} = \left(\frac{\sigma \left(W_{t_{i+1}} - W_{t_i} \right) + \sqrt{\Delta_{t_i}}}{2(1 + \alpha(t_{i+1} - t_i))} \right)^2,$$

$$\Delta_{t_i} = \sigma^2 (W_{t_{i+1}} - W_{t_i})^2 + 4(x_{t_i} + \left(\alpha\theta - \frac{\sigma^2}{2} \right)(t_{i+1} - t_i))(1 + \alpha(t_{i+1} - t_i)).$$

As starting point we choosed 0.5. The scheme will ensure positive values for the solution if, so that there are no problem with the square root. We used the RMSE measure proposed in Durham and Gallant [12] to calculate the error, which is given by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[\log \left(\tilde{p}(x_{i+1}|x_i) \right) - \log \left(p(x_{i+1}|x_i) \right) \right]^2}.$$

In Table 1 below the RMSE and CPU time[2] are presented.

[2]We used a 3GHz Windows machine together with Matlab software.

(M,K)	RMSE	CPU time (s)
$M = 8 \quad K = 32$	0.000183	29.92
$M = 8 \quad K = 16$	0.000175	15.95
$M = 8 \quad K = 8$	0.000248	8.47
$M = 4 \quad K = 8$	0.000472	8.11

Table 1: Test of approximation of log likelihood

6 Interest rate models

As mentioned in the introduction we model interest rates through the SDE

$$dX = \left(\alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3/X\right) dt + \sqrt{\beta_0 + \beta_1 X + \beta_2 X^{\beta_3}}\, dW. \qquad (15)$$

This SDE nests most models in the literature, as is shown in Table 2 below.

Model	$\mu(x;\theta)$	$\sigma(x;\theta)$
Merton	α	σ
Vasicek	$\alpha + \beta x$	σ
CIR	$\alpha + \beta x$	$\sigma\sqrt{x}$
Pearson-Sun	$\alpha + \beta x$	$\sigma\sqrt{x - \eta}$
Dothan	$-$	σx
GBM	βx	σx
Brennan-Schwartz	$\alpha + \beta x$	σx
CIR VR	$-$	$\sigma x^{3/2}$
CEV	βx	σx^{γ}
CKLS	$\alpha + \beta x$	σx^{γ}

Table 2: Examples of interest rate models nested in our model (15).

Our drift and volatility specifications are listed in Table 3 below (cf. Durham [11]). The Affine model gives affine zero-coupon bond prices and include CIR and Vasicek as special cases. The CEV1-CEV3 models have so called constant elasticity of their volatilities, while the GEN1-GEN3 models have more general volatility specifications.

Model	$\mu(x;\theta)$	$\sigma(x;\theta)$
Affine	$\alpha_0 + \alpha_1 x$	$\sqrt{\beta_0 + x\beta_1}$
CEV1	α_0	$\beta_0 x^{\beta_1}$
CEV2	$\alpha_0 + \alpha_1 x$	$\beta_0 x^{\beta_1}$
CEV3	$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$	$\beta_0 x^{\beta_1}$
GEN1	α_0	$\sqrt{\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}}$
GEN2	$\alpha_0 + \alpha_1 x$	$\sqrt{\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}}$
GEN3	$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x$	$\sqrt{\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}}$

Table 3: Interest rate models used in this thesis

7 Data analys and test statistics

Here we present our data set together with the test statistics we use.

7.1 Data

The data set we used is the weekly Swedish 3-month Treasure bill rates from January 1, 1983 to January 1, 2007, see Riksbanken [18]. As the five rates from week 18 1983, weeks 5-7 1990 and week 27 2004 were missing, we substituted these missing data with interpolated values. The total number of data is $n = 1247$. The data is plotted in Figure 6 below.

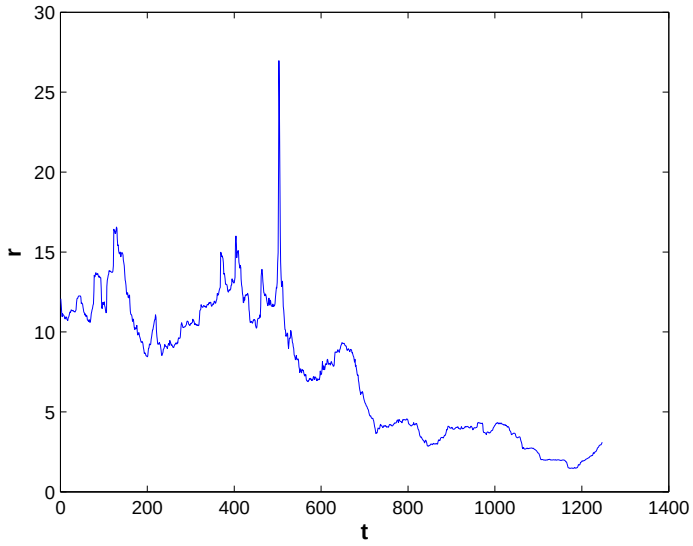


Figure 6: Weekly Treasure bill rates January 1, 1983 to January 1, 2007.

Table 4 below lists some stylized facts (descriptive statistics) for the data.

Mean	7.4912
Variance	17.1794
Standard Deviation	4.1448
Skewness	0.311
Kurtosis	2.3027

Table 4: Stylized facts of the data

Note that the skewness 0.3111 and kurtosis 2.3027 indicate some asymmetry and some rounded peaks within the data.

7.2 Test statistics

Since our models are nested in (15) and since likelihood values are available we can apply the well-known likelihood ratio (LR) test and together with the convenient asymptotic properties of this test. We also use Akaike information criterion (AIC) and Schwartz' Bayesian criterion (SC) to rank the models.

7.2.1 LR test

With the LR test we can compare if a reduced model gives a fit which is as good as a fuller model.

Consider a parameter vector θ and a null hypothesis H_0 that puts m restrictions on θ. For example, suppose that we are interested to test if a population has zero mean μ when $\theta = (\mu, \sigma)$, so that $m = 1$ since we have one restriction on θ. Let $\tilde{\theta}_{\mathrm{ML}}$ denote the ML estimator without restrictions and $\tilde{\theta}_{0,\mathrm{ML}}$ the restricted ML estimate. If the null hypothesis is true, then $\tilde{\theta}_{\mathrm{ML}}$ and $\tilde{\theta}_{0,\mathrm{ML}}$ both should be close to θ, which should imply that the corresponding likelihoods $L(\tilde{\theta}_{\mathrm{ML}})$ and $L(\tilde{\theta}_{0,\mathrm{ML}})$ are close.

The likelihood ratio test rejects H_0 if

$$2\left[\log(L(\tilde{\theta}_{ML})) - \log(L(\tilde{\theta}_{0,ML}))\right] \geq \chi^2_{\alpha,m}, \tag{16}$$

where $\chi^2_{\alpha,m}$ is the α upper probability value of the chi-square distribution with m degrees of freedom. The LR test (16) is only approximative, as it is obtained the limit test obtained when the sample size tends to infinity under some regularity conditions.

The asymptotic distribution of the ML estimator $\tilde{\theta}$ is well-known, see e.g., Pawitan [14]. It can be described in several ways, as, for example,

$$\tilde{\theta} \sim N(\theta, I(\tilde{\theta})^{-1})$$

where $I(\tilde{\theta})$ is the Fischer information, which is to say minus the Hessian, that is,

$$I_{ij}(\tilde{\theta}) = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log L(\theta) \mid_{\theta=\tilde{\theta}} .$$

The standard error for the parameter number i is given by $I_{ii}^{-1/2}(\tilde{\theta})$.

7.2.2 AIC and SC

AIC and SC are used for model selection and both depend on the likelihood. The AIC is defined as

$$\text{AIC} = \frac{-2\log(L(\tilde{\theta}_{ML}))}{n} + \frac{2K}{n},$$

while the SC is defined as

$$\text{SC} = -\frac{2\log(L(\tilde{\theta}_{ML}))}{n} + \frac{\log(n)K}{n}.$$

In both cases K is the number of free parameters and n the sample size. The model selection is made by minimizing each of AIC and SC. Both criteria depend in the same way on the likelihood, but differ in the second term which penalizes too many parameters. Therefore SC tend to favour models with fewer parameters than AIC. In practice the models selected by AIC and SC tend to be close or even identical, see Ruppert [19].

8 Parameter estimations

We $M = 8$ and $K = 16$ we estimated the parameters. This choice gave good approximation results for the CIR model at the same time as giving reasonable simulation times. Matlab was used as software, but I would like to recommend faster softwares such as C\C++ or Fortran.

The optimization of the log likelihood was not trivial since the Hessian was ill-conditioned, that is, close to a singular matrix.

In Tables 5 and 6 below we list the estimated drift and volatility parameters, respectively, together with their corresponding standard errors for the different nested models considered.

Model	α_0	α_1	α_2	α_3
Affine	0.22 ± 0.09	0.00 ± 0.05		
CEV1	0.00 ± 0.09			
CEV2	0.001 ± 0.000	-0.012 ± 0.000		
CEV3	-0.063 ± 0.000	-0.010 ± 0.000	-0.004 ± 0.000	0.187 ± 0.000
GEN1	0.000 ± 0.002			
GEN2	0.025 ± 0.000	-0.062 ± 0.000		
GEN3	-0.319 ± 0.000	0.000 ± 0.000	-0.001 ± 0.000	0.505 ± 0.000

Table 5: Estimated drift parameters with standard errors.

Model	β_0	β_1	β_2	β_3
Affine	-0.751 ± 0.008	0.515 ± 0.004		
CEV1	0.073 ± 0.002	1.428 ± 0.010		
CEV2	0.073 ± 0.000	1.427 ± 0.000		
CEV3	0.073 ± 0.000	1.426 ± 0.000		
GEN1	-0.064 ± 0.000	0.074 ± 0.000	0.000 ±0.000	7.520 ± 0.107
GEN2	-0.073 ± 0.001	0.075 ± 0.001	0.000 ± 0.000	7.488 ± 0.000
GEN3	-0.070 ± 0.000	0.073 ± 0.000	0.000 ± 0.000	7.518 ± 0.000

Table 6: Estimated volatility parameters with standard errors

Table 7 below presents the log likelihood together with the AIC, SC and LR tests. Note that the volatility is the sensitive part in model specification and thus the most general volatility specification is to be preferred for our data set.

Model	log likelihood	AIC	SC	LR test
Affine	254.	-0.40	-239.	rejected
CEV1	708.	-1.13	-698.	rejected
CEV2	709.	-1.13	-694.	rejected
CEV3	709.	-1.13	-688.	rejected
GEN1	860.	-1.37	-842.	not rejected
GEN2	863.	-1.37	-842.	not rejected
GEN3	863.	-1.37	-835.	-

Table 7: Estimations of log likelihood, AIC,SC and LR test

The Affine model (and thus Vasicek and CIR models) fit the data very poorly. This conclusions was also made by Chan et al. [9] and Durham [11] with different data sets. The reason that we have used this model is because of its analytical tractability. All interest rate models from Table 2 are rejected.

In Figure 7 below we have plotted the fitted drift for GEN1-GEN3 and the fitted volatility for GEN1-GEN3 and CEV3. Both the drift and volatility seem to have non-linearity properties as the non-linearity parameters are statistically significant.

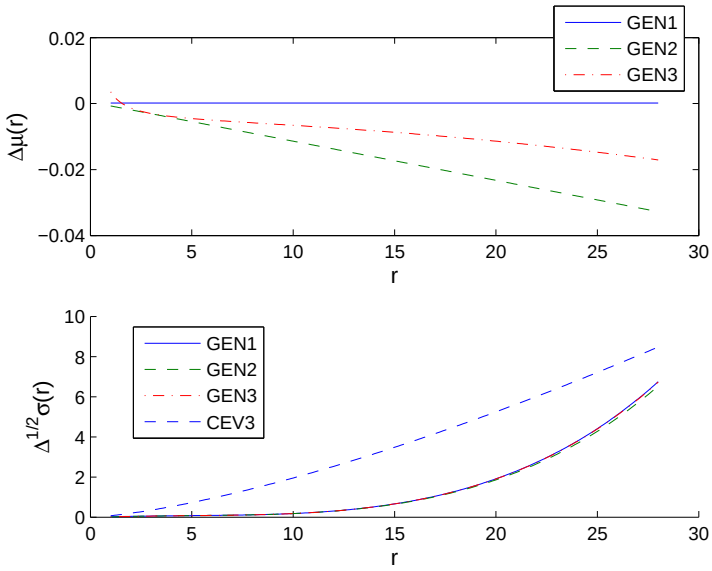


Figure 7: Fitted drift and volatility.

Our conclusion that a general volatility function is required is supported by Durham [11], who required general volatilities to model daily observations

of the U.S 3-month Treasury bill rate. Durham mentions that the CEV volatility function which for low interest rates is constrained to approach zero was unable to catch the relativetely high volatility found in the daily data at low interest rates. The result is that the model tries to catch the low values of the interest rates causing a volatility function with too little curvature.

In Figures 8-10 below we show simulations of our fitted interest rate models, to be compared with the plot of the real data set in Figure 6.

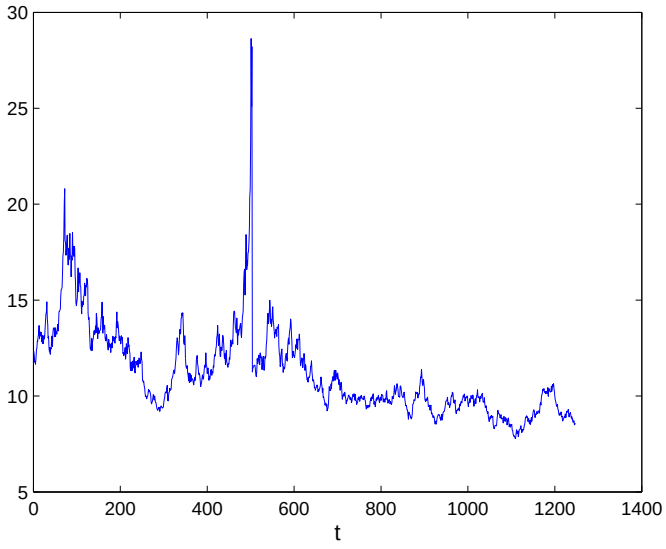


Figure 8: Trajectory of GEN1.

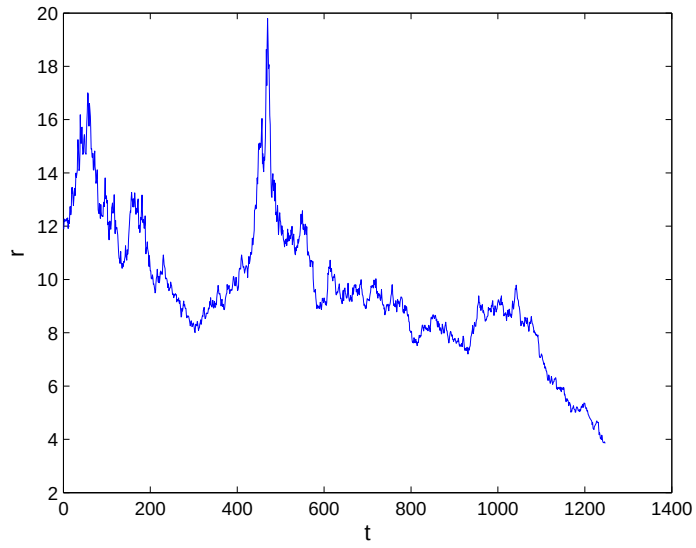


Figure 9: Trajectory of GEN2.

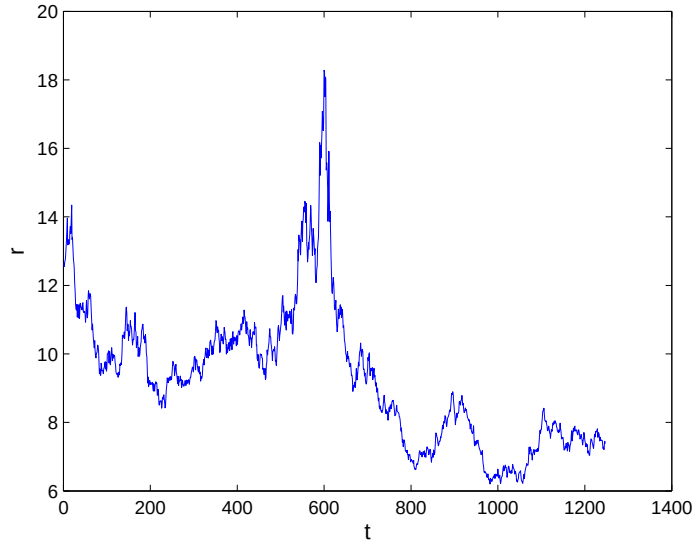


Figure 10: Trajectory of GEN3.

9 Conclusions

In this thesis we have shown that SDE interest rate models with a more flexible volatility function match the weekly Swedish 3-month Treasure bill rates from 1 January, 1983 to 1 January, 2007 much better than constant elasticity of volatility models and the Affine model, the latter of which includes, for example, the CIR and Vasicek models.

When using SML is seems favorable to use fast softwares such as C\C++ and Fortran. However, also the programmer highly affects the speed of the method. The SML method has the advantage to be robust for large time discretization steps. By increasing the parameters M and K of the method, the approximation error can be controlled and made arbitrarily small. This has to be balanced against available computing resources. In our case is was problematic to choose M and K sufficiently large to get small approximation errors as this resulted in simulation that took too long time.

We evaluated the performance of the SML method by calculating log likelihoods for the well-known CIR interest rate model. It turned out that the likelihood for a dataset with 5000 data could be approximatated very well with 8 seconds of computing time. Regarding the optimization of the log likelihood there is the disadvange with the use of SML that it is difficult to differentiate the log likelihood as the approximated density is not smooth enough. Further, when the Lamperti transformation does not exist in closed form, then this problem becomes even greater at the same time as significantly more simulation time is required. Although SML does not require Lamperti transformation, it has been shown by Durham and Gallant [12] and Stramer and Yan [21] that better results are obtained with transformation.

To estimate parameters the ML method was used. The well-known optimal properties of ML helped us analyze the fitted models. The LR test was used to test reduced models against non-reduced ones. We used AIC and SC to rank the different models. All these statistical procedures require likelihoods which is the reason we used the SML method with its good approximation properties for likelihoods.

References

[1] Ahlgrim, K.C., D'Arcy. S.P. and Gorvett, R.W. (1999). Parameterizing interest rate models. Unpublished manuscript.

`http://www.casact.org/pubs/forum/99sforum/99sf001.pdf`

[2] Ait-Shalia, Y. (1996). Testing continous-time models of the spot interest rate. *Review of Financial Studies* **9** 385-426.

[3] Ait-Shalia, Y. (2001). Maximum likelihood estimation of discretly sampled diffusions: a closed-form approximation approach. *Econometrica* **70** 223-262.

[4] Alfonsi, A. (2005). On the discretization schemes for the CIR (and Bessel squared) processes. Unpublished manuscript.

`http://cermics.enpc.fr/reports/CERMICS-2005/CERMICS-2005-279.pdf`

[5] Andersen, C.E. (1999). Monte Carlo methods and importance sampling. Unpublished manuscript.

`http://ib.berkeley.edu/labs/slatkin/eriq/classes/guest_lect/mc_lecture_notes.pdf`

[6] Brandt, M.W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics* **63** 161-210.

[7] Bruche, M. (2005). Estimating structural bond pricing models via simulated maximum likelihood. Unpublished manuscript.

`http://fmg.lse.ac.uk/~max/paper2.pdf`

[8] Cairns, A.J.G. (2004). *Interest rate models. An introduction.* Princeton University Press.

[9] Chan, K.C., Karolyi, G,A., Longstaff, F.A. and Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance* **47** 1209-1227.

[10] Cox, J.C., Ingersoll Jr., J.E. and Ross, S.A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385-407.

[11] Durham, G.B. (2002). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics* **70** 463-487.

[12] Durham, G.B. and Gallant, A.R. (2002). Numerical techniques for simulated maximum likelihood estimation of stochastic differential equations. *Journal of Buisness and Economic Statistics* **20** 297-338.

[13] Heath, M.T. (2002). *Scientific Computing: An Introductory Survey, 2^{nd} Ed.* McGraw-Hill, New York.

[14] Pawitan, Y. (2001). *In all Likelihood: Statistical Modelling and Inference using Likelihood.* Oxford Science Publications.

[15] Pedersen, A.R. (1995a). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1** 257-279.

[16] Pedersen, A.R. (1995b). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* **22** 55-71.

[17] Pastorello, S. and Rossi, E. (2004). Efficient importance sampling maximum likelihood estimation of stochastic differential equations. Unpublished manuscript.

`http://www.cide.info/conf_old/papers/1147.pdf`

[18] Riksbanken. The homepage of Sveriges Riksbank. [In Swedish]

`http://www.riksbank.se/templates/Page.aspx?id=15963`

[19] Ruppert, D. (2004). *Statistics and Finance. An Introduction.* Springer-Verlag, New York.

[20] Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications* **16** 733-752.

[21] Stramer, O. and Yan, J. (2006). On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation.

`http://www.stat.uiowa.edu/techrep/tr371.pdf`