CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

Evaluation of Some Methods for Parameter Estimation for Stochastic Differential Equations

OSCAR HAMMAR

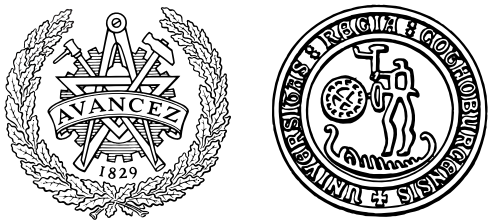
Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2005

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

Evaluation of Some Methods for Parameter Estimation for Stochastic Differential Equations

Oscar Hammar

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden
Göteborg, September 2005

Evaluation of Some Methods for Parameter Estimation for Stochastic Differential Equations

Oscar Hammar

5th September 2005

Contents

1 Introduction **2**

2 Volatility induced stationarity, vis **3**

3 Simulation **4**

4 Estimation methods **5**
4.1 Generalized method of moments 5
4.2 Martingale estimation function 6
4.3 Approximate likelihood estimation 7

5 Inconsistency of the Generalized method of moments and the Martingale method **9**

6 Practical considerations **10**
6.1 Approximate likelihood method 10
6.1.1 Choice of the grid 10
6.1.2 Experimental order determination 10
6.1.3 Optimization issues 12
6.2 Generalized method of moments and Martingale method 12
6.2.1 Optimization issues 12

7 Evaluation **14**

8 Conclusions **15**

Chapter 1

Introduction

A diffusion process is the solution to a stochastic differential equation (SDE). An SDE has the form

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t, \quad X_0 = x_0,$$

where W_t is Brownian motion, b and σ are the drift and σ the diffusion, respectively, while θ is a parameter. Written in integral form, this SDE takes the form

$$X_t = \int_0^t b(X_s;\theta)ds + \int_0^t \sigma(X_s;\theta)dW_s. \tag{1.1}$$

The particular diffusion process that will be studied in this paper is the CKLS model, introduced by Chan, Karolyi, Longstaff and Sanders [1], where $\theta = (\alpha, \beta, \sigma, \gamma), b(X_t, \theta) = \alpha + \beta X_t$ and $\sigma(X_t;\theta) = \sigma X_t^\gamma$, so that

$$dX_t = (\alpha + \beta X_t)dt + \sigma X_t^\gamma dW_t, X_0 = x_0 \tag{1.2}$$

In practice one can only observe the process at discrete time points. Ideally inference about θ from these discrete observations should be based on the likelihood function. However, the likelihood function for discrete observation is a product of transition densities, which are not known, except in special cases. Some approximation of the likelihood function must be made. Two approximations are presented in this paper. The first is to use that the unknown transition density solves the ordinary differential equation (ODE) called the Fokker-Plank equation. Then the problem is reduced to solve an ODE. The second is to construct a martingale by compensating for the error introduced by using the known likelihood function in the continuous case. In addition, the generalized method of moment approach used in [1] is evaluated. To evaluate the methods we simulate the SDE with two choices of θ and apply the methods. The first choice θ = [0.2 -2 0.2 0.8], is the same as in the evaluation of the methods in [2], for which we get similar results to those in [2]. In the second parameter vector, θ = [0.2 -2 0.2 1.1], γ is changed to 1.1. This small change makes inference much harder. One reason for this the introduction of vis, volatility induced stationarity, a phenomena explained in the next chapter.

Chapter 2

Volatility induced stationarity, vis

Usually when one describes the solution to an SDE, one splits the SDE into two parts. The first integral (the drift) in (1.1) is viewed as the solution to an ODE, and the second integral (the diffusion) in (1.1) is viewed as an white noise added to the solution to the ODE. For some parameter values of the CKLS model (1.2) this interpretation is not correct. The reason for that is vis, volatility induced stationarity. For SDE:s with vis not only the drift ensures that the process returns to a stationary level (has a solution), but also the diffusion. The diffusion influences the level of the process in the following way: At large values of the process the diffusion increases the volatility of the process. With a high volatility the process will not stay at the current level for long. It either moves back to the stationary level or to an even higher level, where the volatility gets even higher and thus increases the chance that the process returns to a lower level even more. Sooner or later the process returns. (In fact, because of vis, this will rather happen soner than later.)

The parameter values for which the SDE (1.2) has vis are

$$\{\tfrac{1}{2}<\gamma<1, \alpha>0, \beta=0\} \cup \{\gamma=1, \alpha>0, 0\leq\beta<\tfrac{1}{2}\sigma^2\} \cup \{\gamma>1, \alpha>0\}$$
$$\cup\{\gamma>1, \alpha=0, \beta>0\}.$$

In particular, for θ = [0.2 -2 0.2 1.1], the SDE has vis, and for θ = [0.2 -2 0.2 0.8] it does not.

Chapter 3

Simulation

Ideally one would use the transition probabilities to simulate weak solutions to the SDE. But since these probabilities are not known in explicit form one has to use numerical methods. In this paper a family of Euler schemes are used.

Let $[0,T]$ be a fixed time interval, let $0 = t_0 < t_1 < ... < t_N = T$ be equidistant time points with spacing Δ, and let $\Delta W_n = W_{n+1} - W_n$ be the increment of the Brownian motion over the interval $[t_n, t_{n+1}]$. For $0 \leq \Theta_\sigma \leq 1$ and $0 \leq \Theta_b \leq 1$ a family of Euler schemes is defined by $Y_0 = y_0$, while, at time t_{n+1},

$$\begin{aligned} Y_{n+1} &= Y_n + (\Theta_b \bar{b}(Y_{n+1}) + (1-\Theta_b)\bar{b}(Y_n))\Delta \\ &+ (\Theta_\sigma \sigma(Y_{n+1}) + (1-\Theta_\sigma)\sigma(Y_n))\Delta W_n \end{aligned}$$

for $n < N$. Here $\bar{b} = b - \Theta_\sigma \sigma \sigma'$ is a correction term to ensure convergence to the Itô solution to the SDE.

For $\Theta_b = \Theta_\sigma = 0$ one gets an explicit Euler scheme. For the CKLS model the explicit scheme converges to the true solution if $\gamma < 1$, for example according to [7]. To improve stability for more general processes one can use an implicit scheme ($\Theta_b > 0, \Theta_\sigma > 0$). In this paper the full implicit Euler method is used ($\Theta_b = 1, \Theta_\sigma = 1$). Given Y_n, Y_{n+1} is given by the solution of the non-linear equation $f(Y_{n+1}) = Y_n$, where

$$f(y) = y - \Theta_b(\alpha + \beta y - \Theta_\sigma \gamma \sigma^2 y^{2\gamma-1})\Delta - \Theta_\sigma \sigma y^\gamma \Delta W_n.$$

This requires that $f(Y_{n+1}) = Y_n$ has a unique solution for all possible values of Y_n. Unfortunately, the uniqueness of the solution depends on the size of the increment of the Brownian motion. For large positive values of ΔW_n, f is not monotone for all possible Y_n. In these cases, changing the values of Θ_b and Θ_σ gives a unique solution. In this paper we follow the suggestion in [7] and take an explicit step in those cases, which means to set $\Theta_b = \Theta_\sigma = 0$. It is argued in [7] that this gives a stable scheme also for larger values of γ.

Chapter 4

Estimation methods

4.1 Generalized method of moments

The change of the value of the process between times t_i and t_{i+1} is given by

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} \alpha + \beta X_t dt + \int_{t_i}^{t_{i+1}} \sigma X_t^\gamma dW_t. \tag{4.1}$$

Observing that $t_{i+1} - t_i = \Delta$, for small Δ, the drift in (4.1) is approximated well by

$$\int_{t_i}^{t_{i+1}} \alpha + \beta X_t dt = \Delta(\alpha + \beta X_{t_i}).$$

In a similar manner, when there is no vis, the diffusion is approximated well by

$$\int_{t_i}^{t_{i+1}} \sigma X_t^\gamma dW_t = \sigma X_{t_i}^\gamma \int_{t_i}^{t_{i+1}} dW_t.$$

Let $\epsilon_{t_{i+1}} = X_{t_{i+1}} - X_{t_i} - \Delta(\alpha - \beta X_{t_i})$. With the above approximations, we have

$$\epsilon_{t_{i+1}} = \sigma X_t^\gamma \int_{t_i}^{t_{i+1}} dW_t$$

and therefore:

$$\begin{aligned} \mathbf{E}(\epsilon_{t_{i+1}} | \mathcal{F}_{t_i}) &= 0, \\ \mathbf{E}(\epsilon_{t_{i+1}}^2 | \mathcal{F}_{t_i}) &= \Delta \sigma^2 X_t^{2\gamma}. \end{aligned} \tag{4.2}$$

We seek inference for θ: Let

$$f_{t_i}(\theta) = \begin{pmatrix} \epsilon_{t_{i+1}} \\ \epsilon_{t_{i+1}} X_{t_i} \\ \epsilon_{t_{i+1}}^2 - \Delta \sigma^2 X_{t_i}^{2\gamma} \\ (\epsilon_{t_{i+1}}^2 - \Delta \sigma^2 X_{t_i}^{2\gamma}) X_{t_i} \end{pmatrix}$$

For the true parameter θ_0, we have $\mathbf{E}(f_{t_i}(\theta_0)) = 0$. We estimate $\mathbf{E}(f_{t_i}(\theta_0))$ by

$$g_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} f_{t_i}(\theta) \tag{4.3}$$

The estimate of θ is then given by the θ which statisfies $g_N(\theta) = 0$. This is the method used in [1].

4.2 Martingale estimation function

We will consider the following kind of estimating function $G_n(\theta)$, where the dependence of the data is suppressed in the notation:

$$G_n(\theta) = \sum_{i=1}^{n} g(X_{t_{i-1}}, X_{t_i}; \theta) \tag{4.4}$$

It is particularly easy to work with an estimating function that is a martingale. Under the true value θ_0, the expectation of each term in the martingale function is zero, independent on the previous ones. The estimate is obtained by equating the estimating function to zero and solving the equations with respect to the parameters.

The ideal martingale estimating function is the score function, the derivative of the likelihood function. Usually the score function is unknown. The idea is to construct a martingale estimating function based on an approximation of the score function. This approximation to the unknown discrete score function is based on the known continuous likelihood function.

The following way to derive the martingale estimating function is used in [6]. If σ does not depend on θ and under some additional conditions, the *continuous-time* log-likelihood function is

$$l_t(\theta) = \int_0^t \frac{b(X_s, \theta)}{\sigma^2(X_s)} dX_s - \frac{1}{2}\int_0^t \frac{b^2(X_s; \theta)}{\sigma^2(X_s)} ds.$$

An *approximate* discrete score function is obtained by using Riemann and Itô sums and differentiating with respect to θ:

$$\begin{aligned}\partial_\theta \tilde{l}_n(\theta) = &\sum_1^n \frac{\partial_\theta b(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}(X_{i\Delta} - X_{(i-1)\Delta}) \\ &- \Delta \sum_1^n \frac{b(X_{(i-1)\Delta}; \theta)\partial_\theta b(X_{(i-1)\Delta}; \theta)}{\sigma^2(X_{(i-1)\Delta})}.\end{aligned}$$

More inaccuracy of the approximation is contributed by the fact that usually one is interested in processes where σ *does* depend on the parameter θ, so that

$$\begin{aligned}\partial_\theta \tilde{l}_n(\theta) = &\sum_1^n \frac{\partial_\theta b(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta}, \theta)}(X_{i\Delta} - X_{(i-1)\Delta}) \\ &- \Delta \sum_1^n \frac{b(X_{(i-1)\Delta}; \theta)\partial_\theta b(X_{(i-1)\Delta}; \theta)}{\sigma^2(X_{(i-1)\Delta}, \theta)}.\end{aligned} \tag{4.5}$$

The above approximations have introduced bias, so that $\mathbf{E}_\theta(\partial_\theta \tilde{l}_n)$ is not 0. The non-zero expectation has to be compensated for. To that end, let $F(x; \theta) = \mathbf{E}(X_\Delta | X_0 = x)$ and note that the conditional expectation of each term in the sum $\partial_\theta \tilde{l}_n$ is given by

$$\begin{aligned}\mathbf{E}(\partial_\theta \tilde{l}_i(\theta) - \partial_\theta \tilde{l}_{i-1}(\theta) | F_{i-1}) = &\frac{\partial_\theta b(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}(F(X_{(i-1)\Delta}; \theta) - X_{(i-1)\Delta}) \\ &- \Delta \frac{b(X_{(i-1)\Delta}; \theta)\partial_\theta b(X_{(i-1)\Delta}; \theta)}{\sigma^2(X_{(i-1)\Delta})}.\end{aligned}$$

Summing these expectations and removing them from $\partial_\theta \tilde{l}_n$ gives the following estimating function, which is a zero-mean martingale:

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n} \frac{\partial_\theta b(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta}, \theta)} \{X_{i\Delta} - F(X_{(i-1)\Delta}; \theta); \theta\}.$$

This estimating function is called a linear estimating function. It works well for SDE:s with a diffusion that does not dependent on θ. For more complicated diffusions, more advanced martingale estimating functions are needed, as the g-function has to reveal information about the volatility. Quadratic estimating functions can then be used. The g-function of a quadratic martingale estimating function takes the form

$$\begin{aligned} & g(X_{t_{i-1}}, X_{t_i}; \theta) \\ &= A[X_{t_i} - F(X_{(i-1)\Delta}; \theta)] + B[(X_{t_i} - F(X_{(i-1)\Delta}; \theta)^2 - \phi(X_{(i-1)\Delta}; \theta)], \end{aligned} \quad (4.6)$$

where

$$\phi(X_{t_{i-1}}, X_{t_i}; \theta) = \mathbf{Var}(X_{t_i} | X_{t_{i-1}}).$$

To get an estimator with minimal variance it is shown in [3] that the coefficients A and B in (4.6) shall be chosen in such a way that

$$\begin{aligned} G_n(\theta) = \sum_{i=1}^{n} \{ & \frac{\partial_\theta b(X_{t_{i-1}}; \theta)}{\sigma^2(X_{t_{i-1}}; \theta)} [X_{t_i} - F(\Delta, X_{t_{i-1};\theta})] \\ & + \frac{\partial_\theta v(X_{t_{i-1}}; \theta)}{2\sigma^4(X_{t_{i-1}}; \theta)\Delta} [(X_{t_i} - F(\Delta, X_{t_{i-1};\theta}))^2 - \phi(\Delta, X_{t_{i-1};\theta})]\}. \end{aligned} \quad (4.7)$$

4.3 Approximate likelihood estimation

Since the transition density, ϕ, is unknown we can not directly use maximum likelihood estimation. Though ϕ is unknown it is still known that, under some regularity conditions, ϕ solves the following equation:

$$\frac{\partial}{\partial t}\phi(t, x, y) = -\frac{\partial}{\partial y}(\mu(t, x, y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)\phi(t, x, y)): \quad (4.8)$$

This equation is known as the Fokker-Planck, or the Kolmogorov, or the forward equation. With $\phi(t, x, y)$ denoting the density function of X_t given $X_0 = x$, the initial condition for the PDE is $\phi(0, x, y) = \delta(y - x)$, where δ is the Dirac δ-function. This initial condition can be problematic since the δ-function is not a function in the usual sense, when trying to solve the PDE numerically. Practical considerations are discussed in 6

Following the approach in [4], we rewrite (4.8) as

$$\phi_t(t, y) = a(y)\phi + b(y)\phi_y + c(y)\phi_{yy}, \quad (4.9)$$

where

$$\begin{aligned} a(y) &= (\sigma)^2 + \sigma\sigma_{yy} - \mu_y, \\ b(y) &= 2\sigma\sigma_y - \mu, \\ c(y) &= \frac{1}{2}\sigma^2. \end{aligned}$$

To find an approximate solution, the Crank-Nicolson finite difference method is used. Consider a time/space grid. The diffusion is time homogeneous, so we let the grid values in the time direction go from 0 to t. Since the solution we seek is a transition density, the values in the space direction should cover all probable values of X in the interval $[0, t]$.

Let k be the length of the interval between two grid points in time direction, and h the length of the interval between two grid points in space direction. We approximate ϕ, at the grid point with number n in the time direction and number m in the space direction, by $\phi(nk, mh) = v_m^n$, where

$$\frac{v_m^{n+1} - v_m^n}{k} = a\delta_0(h) + b\delta_1(h) + c\delta_2(h).$$

Here the δ:s are difference operators given by

$$\begin{aligned}
\delta_0(h) &= \frac{1}{2}v_m^{n+1} + \frac{1}{2}v_m^n, \\
\delta_1(h) &= \frac{1}{2}\frac{v_{m+1}^{n+1} - v_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{v_{m+1}^{n} - v_{m-1}^{n}}{2h}, \\
\delta_2(h) &= \frac{1}{2}\frac{v_{m+1}^{n+1} - 2v_m^{n+1} + v_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{v_{m+1}^{n} - 2v_m^{n} + v_{m-1}^{n}}{2h}.
\end{aligned}$$

In order to find the density we need an initial condition and boundary conditions in addition to the approximation of the ODE. The problematic initial Dirac δ-function is handled by means of introducing a normal density at time k

$$v_m^1 = \phi^N(y_0 + mh; x + \mu(x)k, \sigma(x)k),$$

where ϕ^N denotes the normal density function. The boundary conditions are received by putting the boundary of the grid in space direction sufficiently far apart so that ϕ can be assumed to be zero on the boundaries.

Chapter 5

Inconsistency of the Generalized method of moments and the Martingale method

In [2] it is shown that the GMM method is not consistent. In fact the GMM-estimates α_{GMM} and β_{GMM} converge to $(\exp(\Delta\beta) - 1)/\Delta$ and $\alpha(\exp(\Delta\beta) - 1)/(\Delta\beta)$. The reason is that the conditional second moment $\mathbf{E}(X_{t_{i+1}}^2|\mathcal{F}_{t_i})$ is not known, and therefore approximated by (4.2). In [2] a new approximation of $\mathbf{E}(X_{t_{i+1}}^2|\mathcal{F}_{t_i})$ is presented. However, this approximation is too complicated for practical use, and the estimates based on it are not clearly improved. (In fact, only the estimates of γ are accurate, and variances of the estimates are poor.) Also the version of quadratic martingale estimating function used in this paper is based on the inaccurate approximation (4.2). A version of a quadratic martingale estimating function based on the new approximation is also presented in [2]. There is no clear improvement for the martingale method either.

Chapter 6

Practical considerations

6.1 Approximate likelihood method

6.1.1 Choice of the grid

The way to choose the position of the grid involves some challenges. Firstly one has to decide how far away from the current x-value one shall put the boundaries. The fluctuation of the process in a small time-interval depends mainly on the diffusion, σ. It is therefore natural to base the position of the boundaries on σ. The choice of the boundaries in [4] is the current x-value $\pm$ 6σ. The advantage of having the boundaries far apart is that the approximation of the zero density at the boundaries is more accurate. The disadvantage is that much of the computer effort is used to approximate ϕ for extreme values, rather than for values that are of any real interest. In this paper we follow the choice of [4] and put the boundaries on $x \pm 6\sigma$.

The next consideration is what to do when the current x-value $- 6\sigma$ is less than zero. This is a common situation for the parameters chosen in the evaluation of the methods in this paper It is not reasonable to have non-zero densities on negative values. Therefore we put the lower boundary to $\max(0, x - 6\sigma)$.

6.1.2 Experimental order determination

We need some experimental verification of the numerical method, since the true transition density is not know. For this we use a method developed by Østerby [8], that is also utilized in [4]. Let u be a function of *one variable*, and v a discrete approximation based on steps of length h. We say that v is globally first order accurate if

$$v(x) = u(x) - hc(x) - h^2 d(x) - h^3 f(x) - \ldots ,$$

where c, d and f are smooth functions. If also $c = 0$, then we say that the approximation is (at least) globally second order accurate. Now consider the numerical solutions with step sizes h, $2h$ and $4h$:

$$v_1(x) = u(x) - hc(x) - h^2 d(x) - h^3 f(x) - \ldots ,$$

$$v_2(x) = u(x) - 2hc(x) - 4h^2 d(x) - 8h^3 f(x) - \ldots ,$$
$$v_3(x) = u(x) - 4hc(x) - 16h^2 d(x) - 64h^3 f(x) - \ldots .$$

We combine these numerical solutions to get

$$\frac{v_3 - v_2}{v_2 - v_1} = 2\frac{c + 6hd + 28h^2 f + \ldots}{c + 3hd + 7h^2 f}. \tag{6.1}$$

Assume that h is small. If the ratio in (6.1) is close to 2.0, then the conclusion is that $c \neq 0$, and thus the approximation is first order accurate. If $c = 0$, then the terms involving hd will dominate the terms involving $h^2 f$, and the ratio will be approximately 4.0.

Now consider the situation in our case, where the function u depends on two variables h and k. We perform the determination of the order separately in each direction, keeping the other variable fixed. Let y_0 and y_M denote the lower and upper boundary of the grid in space direction. In the experimental order determination in [4] y_0 and y_M are set to 0.05 and 0.15. With this choice of y_0 and y_M, and the parameter values used in the evaluation of the method, a pattern appears, much like the one in [4]. The time between observations Δ is $\frac{1}{12}$, and the current value of the process x is 0.1, which is the stationary level of the process.

θ = [0.2 -2 0.2 0.8], $y_0 = 0.05, y_M = 0.15$

y	0.080	0.085	0.090	0.095	0.100	0.105	0.110	0.115	0.120
h-ratio	3.5	3.7	3.9	4.0	4.1	4.1	4.3	4.1	4.0

θ = [0.2 -2 0.2 0.8], $y_0 = 0.05, y_M = 0.15$

y	0.080	0.085	0.090	0.095	0.100	0.105	0.110	0.115	0.120
k-ratio	5.3	4.0	3.8	4.0	3.9	3.8	3.9	3.9	5.0

However, in practice it is not reasonable to let the grid points in space direction go from 0.05 to 0.15. The method is based on the fact that we know the density on the boundaries, namely the grid is set so wide that the density can be assumed to be zero on the boundaries. Therefore, in the evaluation of the method, we take $y_0 = \max(0, x - 6\sigma)$ and $y_M = x + 6\sigma$. Now the picture change:

θ = [0.2 -2 0.2 0.8], $y_0 = \max(0, 0.1 - 6\sigma) = 0, y_M = 0.1 + 6\sigma = 0.2902$

y	0.080	0.085	0.090	0.095	0.100	0.105	0.110	0.115	0.120
h-ratio	5.2	3.7	3.6	3.9	3.7	3.2	3.8	3.7	3.1

θ = [0.2 -2 0.2 0.8], $y_0 = \max(0, 0.1 - 6\sigma) = 0, y_M = 0.1 + 6\sigma = 0.2902$

y	0.081	0.084	0.090	0.096	0.099	0.105	0.110	0.116	0.119
k-ratio	0.9	1.3	2.4	4.0	4.9	7.4	-19.1	-2.7	-3.5

For the second parameter vector, $\theta = [0.2\ \text{-}2\ 0.2\ 1.1]$, the k-ratios look nice. But the h-ratios are really bad:

$\theta = [0.2\ \text{-}2\ 0.2\ 1.1]$, $y_0 = 0.05, y_M = 0.15$

y	0.080	0.085	0.090	0.095	0.100
h-ratio	$-2.2 * 10^8$	$3.3 * 10^4$	0.7	0.2	0.5

$\theta = [0.2\ \text{-}2\ 0.2\ 1.1]$, $y_0 = 0.05, y_M = 0.15$

y	0.105	0.110	0.115	0.120
h-ratio	1.6	3.9	4.7	4.8

6.1.3 Optimization issues

The Matlab optimization routine for constrained nonlinear multivariable functions, *fmincon* was utilized for the approximate maximum likelihood method. The optimization procedure involved quite a few challenges. The problems originate from the fact that an approximation was used, which can be quite poor for some parameter values (see Section 6.1.2).

For processes with $\gamma = 1.1$, some simplifications were done to be able to run the optimization. When an approximation of a transition density was less then zero, resulting in an imaginary score function, that density was set to a small number, in order to continue the optimization.

One other challenge in the optimization was whether or not the best choice was a dynamic grid. The width of the grid is dependent on the unknown parameter θ. With a dynamic grid, the witdh of the grid is changed more dramatically when the optimization procedure run through different θ. When the grid width becomes larger the approximation gets worse. A potential problem with a dynamic grid is that the optimization routine might find a θ that overestimate the likelihood due to a bad approximation from a wide grid. To avoid this problem the position of the grid was based on the true parameter θ_0. With this choice, the width of the grid is dependent only of the value of the process.

The effect of the fixed grid on the likelihood function of the estimates were controlled. It showed that the approximate likelihood function was changed very little when the grid was based on the estimated θ, instead of on θ_0. Also, the effect of setting the density to a small number when the approximate density turned out to be negative was studied. This simplification did not seem influential.

6.2 Generalized method of moments and Martingale method

6.2.1 Optimization issues

The resulting vector in $\mathbb{R}^4$ from the functions (4.7) and (4.3) are squared and summed, resulting in a nonlinear least-squares problem. This was solved by the Matlab routine *lsqnonlin*. The objective function turned out to have many

different local minima for the Martingale method. In order to find the global minimum, many different start vectors were used.

Chapter 7

Evaluation

We simulated 12 processes in the time interval 0 to 25 for each parameter vector, and sampled 300 process values at distance $\frac{1}{12}$. The process was simulated with the time step $\frac{1}{12}10^{-3}$ between the sampled values.

For the optimization it is convenient to reparameterize the diffusion: Let $\Phi = (\eta, \kappa, \sigma, \gamma), \eta = -\alpha/\beta$ and $\kappa = -\beta$. Then the SDE (1.2) takes the form

$$dX_t = -\kappa(X_t - \eta)dt + \sigma X_t^{\gamma} dW_t.$$

Here η is the long term level of the process, and is easy to estimate. In the tables the estimates for Φ is presented. The numbers in parenthesis are the standard deviations of the estimates.

Φ_0 = [0.1 2 0.2 0.8] (θ_0 = [0.2 -2 0.2 0.8])

Method	η	κ	σ	γ
Generalized method of moments	0.0984	1.9893	0.1930	0.7504
	(0.0039)	(0.3558)	(0.1111)	(0.2782)
Martingale estimating function	0.0985	1.9351	0.1991	0.7535
	(0.0039)	(0.3018)	(0.1396)	(0.2824)
Approximate maximum likelihood	0.0986	2.1017	0.2754	0.8211
grid size: 100×8	(0.0039)	(0.3742)	(0.2212)	(0.3443)

Φ_0 = [0.1 2 0.2 1.1] (θ_0 = [0.2 -2 0.2 1.1])

Method	η	κ	σ	γ
Generalized method of moments	0.0991	1.9760	0.2875	1.0477
	(0.0020)	(0.3429)	(0.2855)	(0.5406)
Martingale estimating function	0.0991	1.9330	0.3508	1.0514
	(0.0020)	(0.3050)	(0.5056)	(0.5714)
Approximate maximum likelihood	0.0992	2.1068	0.5643	1.1260
grid size: 100×8	(0.0020)	(0.3823)	(0.8979)	(0.6762)

Chapter 8

Conclusions

We have evaluated some methods for inference of parameters for SDE:s. The methods seem to work quite well for some parameter values. Our estimates agree with other papers evaluating the methods for more restricted set of parameter values. However, the methods are not found to be reliable for general parameter values. Especially, for the martingale method and generalized method of moment, it is the vis effect that causes this problem. For the approximate likelihood method, it is instead poorer accuracy of the numerical solution that is the problem. The mean of the estimates are reasonably accurate, but the variation is for general parameter values very large.

Bibliography

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

[2] B.J. Christensen, R. Poulsen and M. Sørenensen (2001). Optimal inference in diffusion models of the short rate of interest. Preprint.

[3] B.M. Bibby, M. Jacobsen and M. Sørenensen (2004). Estimating functions for discretely sampled diffusion-type models. Preprint.

[4] R. Poulsen (1999). Approximate maximum likelihood estimation of discretely observed diffusion processes. Preprint.

[5] M. Richter (2004). A study of stochastic differential equations with volatility induced stationarity. Preprint.

[6] B.M. Bibby and M. Sørenensen (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17-39.

[7] A. Muszta, M. Richter, J.M.P. Albin and B.A. Jensen (2005). On volatility induced stationarity for stochastic differential equations. Preprint.

[8] O. Østerby (1998). The error of the Crank-Nicolson method for linear parabolic equations with a derivative boundary condition. Preprint.