CHALMERS | GÖTEBORG UNIVERSITY

MASTER THESIS

Parameter Estimation in Discretely Sampled Diffusions Experiencing Volatility Induced Stationarity

FREDRIK NIVEMAN

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY

Thesis for the Degree of Master of Science

Parameter Estimation in Discretely Sampled Diffusions Experiencing Volatility Induced Stationarity

Fredrik Niveman

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden
Göteborg, April 2010

Abstract

In this thesis we estimate parameters for discretely sampled stochastic differentials which experience Volatility induced stationarity (VIS). The specific process considered is of general CKLS-type where the volatility is large. The estimations were performed using general method of moments (GMM), Efficient method of moments (EMM) and approximate maximum likelihood based on Hermite polynomial expansion of the transition density. Because of problem finding appropriate moment conditions in case of VIS, estimation with GMM performed badly. In contrast both EMM approximate likelihood give accurate results although the computational burden increases significantly.

Acknowledgements

I would very much like to thank my advisor Patrik Albin for encouragement during the work and for always having an optimistic view on my progress and on things in general. Further I want to thank the people in the Thesis Work room during the spring of 2010 who have given me much input and advice and who have made this a time to remember. Finally I would like to thank my girlfriend Linnéa for her support and patience during this intense period of time.

Contents

1 Introduction **1**

2 Preliminaries on stochastic analysis and SDEs **3**
2.1 Existence and Uniqueness of Solutions to SDE 3
2.2 Explosion of time-homogenous diffusions 6
2.3 Stationarity of solutions to SDEs 8

3 Volatility Induced Stationarity - VIS **11**

4 CIR and CKLS processes **13**
4.1 Properties of solution 13
4.2 The VIS effect 14
4.3 Transition Densities 16

5 Parameter estimation methods **19**
5.1 Maximum likelihood approaches 20
5.1.1 Simple Maximum likelihood in SDE 21
5.1.2 Approximate Maximum likelihood 21
5.2 Method of Moments approaches 25
5.2.1 General method of moment 26
5.2.2 Efficient method of moment 27

6 Simulation **31**
6.1 The Euler scheme 32
6.2 Implicit approximation schemes 32

7 Implementation **35**
7.1 Finding initial estimates for parameters 35
7.2 Moment conditions for GMM 36
7.3 Explicit expression for the Hermite-polynomial expansion . . 37

8 Evaluation **39**

9 Conclusion **45**

Chapter 1

Introduction

The need for accurate and reliable parameter estimation methods for Stochastic Differential Equations (SDE) is strong in many fields of application. Independently whether we are modeling particle dynamics in physics or interest rate movement in finance we confront the problem of characterizing stochastic behavior in terms of SDEs. Once the stochastic behavior is characterized we could then use this model to explain physical properties or predict financial movements. In most cases the data is discretely sampled so that the full (continuous) movement of the process is not known. Still the data from the measurement is large enough to use statistical method for determine parametric values in the model. In this thesis we estimate parameters for the relatively general, one-dimensional, CKLS-model for the dynamic. This model is represented by the SDE given in Equation (1.1) and concludes a total of four unknown parameters. Especially it will be evaluated in cases for high γ in which the process experience Volatility Induced Stationarity (VIS).

$$X_t = X_s + \int_s^t \alpha + \beta X_u du + \int_s^t \sigma X_u^\gamma dW_u \tag{1.1}$$

The history of parameter estimation dates back to the introduction of the modern computer. In general the methods are computation demanding and therefore the research on the subject began first in the late 1980s. Since then many new approaches have been made. Roughly one can separate between methods which uses some kind of method of moment approach from methods based on maximum likelihood theory. To the former class we recognize the General Method of Moments (GMM) developed by Hansen [12] and Efficient Method of Moments (EMM) first presented by Gallant and Tauchen [14]. In the latter class we count the Simulated Maximum Likelihood (SML) by Pedersen and Transition Density approximation by Hermite polynomial by Ait-Shalia [16].

The idea behind Method of Moments approaches is generally to express the

parameters you want to estimate in terms on the moments. Estimates of the parameters are then given by approximating the moments using statistical inference of averages. In contrast the Maximum likelihood methods uses the transition densities of the processes and find estimates of parameters by maximizing the probability of the given data to occur. Given that the transition density for the process has a closed form expression the most efficient way to estimate parameters is by using direct Maximum Likelihood (ML). Unfortunately, closed form expressions are relatively rare and in the case of the process given in (1.1) it only exists for some specific choices on the parameters. In these cases one most often turn to GMM since it is still often relatively easy to find moments conditions (se e.g. [6]). Examples of application of GMM to the CKLS model in application to interest rate models could be seen in Chan et al [4]. Assume now that the moments conditions could be found. It then feels unnecessary to turn to more advanced methods for determining parameters in (1.1). Why should one not be satisfied with the estimation performed by GMM?

First, the GMM only requires specification of the moments rather than the full transition density. Although the implementation is simple, this causes a drawback in efficiency since it does not make use of all information in the sample. Second, as has been explained by Muszta and Albin [1] [11] the CKLS model shows extraordinary behaviour in cases when the process is stationary and parameter γ becomes high. In this case the second term in 1.1 is no longer a martingale but only a local martingale and will affect the mean reversion properties of the process and cause Volatility Induced Stationarity (VIS). More specifically the martingale approximation that is used for finding moment conditions in the GMM does not longer hold and the method does not give correct estimates of the parameters. It is this breakdown of the GMM that forces us to search for other estimation methods in the specific case with high γ. The purpose with this thesis is to identify the breakdown and conclude whether the use of other methods can overcome this problem. The alternative methods that are evaluated are EMM together with transition density approximation by Hermite polynomials.

The thesis is structured as follows: In Chapter 2 the preliminaries of stochastic analysis and SDE is given in order to understand basic properties. These are then used in Chapter 3 where the VIS effect is explained in greater detail. In Chapter 4 the theory is applied to the specific CKLS process and consequences of VIS are derived in terms of moment conditions. Chapter 5 then gives the background and theory of the estimation methods that will be used. Chapter 6 explains the schemes that are used for simulation of the processes that is estimated with special emphasises on the Implicit Euler scheme. In Chapter 7 discussion about implementation are made. Finally Chapter 8 and 9 discusses the results that was found from estimation.

Chapter 2

Preliminaries on stochastic analysis and SDEs

This section gives a quick introduction to stochastic analysis with emphasises on stochastic differential equations (SDE). The general form of a SDE that will be used is given by (2.1) with a standard notation for the drift $b(u, X_u)$ and diffusion $\sigma(u, X_u)$. Although both drift and diffusion are allowed to be time dependent we will primarily restrict our attention to the time homogenous case $b(u, X_u) = b(X_u)$ and $\sigma(u, X_u) = \sigma(X_u)$.

$$X_t = X_s + \int_s^t b(u, X_u)du + \int_s^t \sigma(u, X_u)dW_u \tag{2.1}$$

Special attention will be on the existence and uniqueness of strong solutions as well as conditions for stationary and absence of explosions. These properties will further on be essential in order to understand the concept of Volatility Induced Stationarity (VIS) in the subsequent chapter.

2.1 Existence and Uniqueness of Solutions to SDE

The first question to ask when studying properties of SDEs is the question of existence and uniqueness of solutions. Once existence and uniqueness is confirmed it is up to the user to choose an appropriate method of finding this solution. The methodology of proving existence is quite similar to the methodology used for ordinary differential equations where one use Picard-Lindelf successive approximation to integral equations and show that the problem converge. The existence and uniqueness is guaranteed locally once we assume integral operators to fulfil Lipschitz-continuous. Further for existence and uniqueness to hold globally we must guarantee that solutions not to explode at finite time. A sufficient condition for this to hold is to require linear growth condition on the integral operators. In this section we shall mention the corresponding theory of existence and uniqueness of solutions

to SDE given by (2.1). The non-deterministic character of the problem will make it necessary to distinguish between weak and strong solutions to the problem whether or not solutions exists for all probability spaces or just under special restrictions. The treatment of theory will be based on the one-dimensional case for simplicity. For higher dimensions the result follows analogously. If not, it will be stated explicitly.

Strong and Weak solutions

As the name suggest we always strive to find strong solutions to SDE since this enables more general treatment of properties for solutions. However in many cases the solution may only exists under some appropriate measurable space along with some probability measure.

Definition (*Strong solutions to SDE*): A strong solution to the SDE in (2.1) on a given probability space $(\Omega, \mathcal{F}, P)$ and w.r.t Brownian motion W is a process with continuous sample paths and following properties (Here x_0 refers to the initial condition of SDE):

1. X is adapted to the (*augmented*) filtration $\{\mathcal{F}_t\}$ generated by the Brownian motion W.
2. $P[X_0 = x_0] = 1$
3. $P[\int_s^t b(u, X_u) + \sigma^2(u, X_u)du < \infty] = 1, \forall 0 \leq s, t < \infty$
4. Equation (2.1) holds with probability one.

Conditions 2 and 4 assure that the solution fulfils SDE P-a.s. This is the only restrictions we demand in case of deterministic differential equations. For the solutions to make sense for SDE we also impose conditions that the solution should be independent of future values of the solution (causality) by condition 1.The condition 3 further restrict strong solutions to behave nicely by having finite drift and second moments on finite time. The only difference in the definition of a *weak solution* to a SDE is that the conditions in the above definition needs only to hold for a certain probability space. That is, if we were able to find a measure under which the conditions holds, we have a weak solution.

Definition (*Strong Uniqueness of solutions*): Assume X and $\hat{X}$ are two strong solutions to (2.1) with $P[X_t = \hat{X}_t; 0 \leq t < \infty] = 1$ whenever W is given Brownian motion with corresponding filtration $\mathcal{F}_t$ (*augmented*) on some probability space $(\Omega, \mathcal{F}, P)$ and with initial condition x_0. Then we say that the strong uniqueness holds for equation (2.1).

An even stronger uniqueness condition is obtained by strengthening the condition $P[X_t = \hat{X}_t; 0 \leq t < \infty] = 1$ to $P[X_t = \hat{X}_t; \forall 0 \leq t < \infty] =$ 1in which we obtain *pathwise uniqueness*. This condition is necessary to derive sufficient conditions for non exploding processes. We now present the necessary theorems that guarantee that a unique strong solution exists.

Existence of Strong solutions

From the previous argument it is possible to make an analogue to the deterministic case when deriving conditions on existence and uniqueness of solutions. The following theorem summarizes the argument.

Theorem 2.1.1 *Suppose that coefficients in $b(t,x)$, $\sigma(t,x)$ in Equation 2.1 satisfy global Lipschitz and linear growth condition:*

$$\| b(t,x) - b(t,y) \| + \| \sigma(t,x) - \sigma(t,y) \| \leq K \| x - y \| \tag{2.2}$$

$$\| b(t,x) \|^2 + \| \sigma(t,x) \|^2 \leq K^2(1 - \| x \|^2) \tag{2.3}$$

For every $0 \leq t < \infty$ and x,y $\in \Re^d$ and K is positive constant. If we have a d-dimensional random initial condition x_0 with bounded second moment, independent of an r-dimensinal Brownian motion $W = \{W_t, \mathcal{F}_t^W; 0 \leq t < \infty\}$, then there exists a continuous adapted process X which is a strong solution to (2.1) relative to W with initial condition x_0

In case of one dimensional SDEs it is possible to relax the condition somewhat in order to prove have existence and strong uniqueness of solutions. The modified conditions are given by Yamada and Watanabe theorem below. More specifically the Lipschitz condition on the diffusion term can be dropped. However the theorem does not include assumptions on the linear growth of the coefficients and hence it cannot guarantee that the solution will be infinite in finite time.

Theorem 2.1.2 (Yamada & Watanabe (1971)) *Assume that the coefficients $b(t, X_t)$ and $\sigma(t, X_t)$ in SDE given by (2.1) satisfy the following conditions:*

$$|b(t,x) - b(t,y)| \leq K|x - y| \tag{2.4}$$

$$|\sigma(t,x) - \sigma(t,y)| \leq h(|x - y|) \tag{2.5}$$

For every $0 \leq t < \infty$ and x,y $\in \Re$. Here K being a positive constant and $h : [0, \infty) \to [0, \infty)$ is a strictly increasing function with h(0)=0 and

$$\int_{(0,\epsilon)} h^{-2}(u)du = \infty, \forall \epsilon > 0 \tag{2.6}$$

Then strong uniqueness holds for (2.1)

Equations that fulfil Yamada and Watanabe can still go to infinity at finite time. Since this is a non desirable property, we need necessary and sufficient conditions for this to be avoided. In context with SDEs this is called Explosions and will be considered next.

2.2 Explosion of time-homogenous diffusions

In the subsequent chapters we will only be interested in time homogenous SDEs. That is continuous solutions to the time homogenous SDE given by 2.7 where the functions b and σ only depends on X.

$$X_t = X_s + \int_s^t b(X_u)du + \int_s^t \sigma(X_u)dW_u \tag{2.7}$$

We now seek conditions that guarantee that X behaves nicely in finite time. By nice behavior we mean that the process will not take on infinite values in finite time with positive probability. This property is known as explosion and is given by:

Definition (*Explosion of SDE*): Let τ_n be the time at which X hits level n, that is $\tau_n = \inf(t > 0; X_t \geq n)$. A solution to Equation 2.7 is said to explode if $P(\tau_\infty < \infty) > 0$.

To derive sufficient conditions X not to explode we will first consider the case without drift. The drift case is then only a matter of finding an appropriate transformation of the process and then using the already established results.

Case without drift

To prove conditions for the existence and uniqueness of a strong, non exploding solution to the time homogenous case without drift in 2.7 we will rely on the following theorem that could be found in Karatzas and Shreve [8]:

Theorem 2.2.1 *Pathwise uniqueness and the existence of a weak solution imply strong existence and uniqueness*

Hence it remains to find the conditions for pathwise uniqueness and existence of non-exploding weak solutions. The latter is given by the theorem proven by Engelbert & Smith in 1984.

Theorem 2.2.2 (Engelbert & Schmidt) *The SDE in 2.7 with b(X)=0 has a non-exploding weak solution for every initial distribution μ iff*

$$I(\sigma) \subseteq Z(\sigma) \tag{2.8}$$

where

$$I(\sigma) = \left\{ x \in \Re; \int_{-\epsilon}^{\epsilon} \frac{dy}{\sigma^2(x+y)} = \infty, \forall \epsilon > 0 \right\}$$

$$Z(\sigma) = \{ x \in \Re; \sigma(x) = 0 \}$$

The outline of the proof is built on the fact that a new brownian motion can always be defined in the case when the conditions are fulfilled. With this specific brownian motion, strong uniqueness holds. It is obvious from the definition of $I(\sigma)$ and $Z(\sigma)$ that the conditions are fulfilled once σ is continuous. In this case, referring to Theorem 2.2.1, it remains to fulfil conditions for pathwise uniqueness property in order to have sufficient conditions for existence and uniqueness of strong solution that does not explode. These conditions are given by the following theorem:

Theorem 2.2.3 *Suppose that there exists functions* $f : \Re \to [0, \infty)$ *and* $h : [0, \infty) \to [0, \infty)$ *such that at every* $x \in I(\sigma)^c$ *there exists* $\epsilon > 0$ *such that*

1. $$\int_{x-\epsilon}^{x+\epsilon} \left(\frac{f(y)}{\sigma(y)} \right)^2 dy < \infty \tag{2.9}$$

2. *the function* h *is strictly increasing,* $h(0) = 0$ *and fulfils 2.6*

3. *there exists a constant* $a > 0$ *such that* $|\sigma(x+y) - \sigma(x)| \leq f(x)h(|y|), \forall x \in \Re, y \in [-a, a]$

In this case pathwise uniqueness holds for the 2.7 without drift

Case with drift

To treat the general case when the SDE does have a non-zero diffusion we can remove the drift by using the removal of drift theorem given below. It is based upon the simple idea that we create a new process Y(X) from the original one and then use general Ito formula to obtain expression for Y(X). If we choose the function Y=Y(X) in an appropriate way we can eliminate drift and write Y(X) as a simple stochastic integral. To ensure existence of integrals we will assume the non-degeneracy and local integrability condition to hold:

$$\sigma^2(x) > 0, \forall x \in \Re \tag{2.10}$$

$$\int_{x-\epsilon}^{x+\epsilon} \frac{|b(y)|dy}{\sigma^2(y)} < \infty, \forall x \in \Re, \exists \epsilon > 0 \tag{2.11}$$

Theorem 2.2.4 (Removal of Drift) *Assume that conditions 2.10 and 2.11 hold and define the* **scale function** *by:*

$$p(x) = \int_{\epsilon}^{x} exp\left\{\int_{\epsilon}^{\xi} \frac{b(\eta)}{\sigma^2(\eta)} d\eta\right\} d\xi \tag{2.12}$$

A process $X = \{X_t, \mathcal{F}_t, 0 \leq t < \infty\}$ *is a strong solution to 2.7 iff process* $Y = \{Y_t = p(X_t), 0 \leq t < \infty\}$ *is a strong solution to:*

$$Y_t = Y_0 + \int_0^t \hat{\sigma}(Y_s) dW_s \tag{2.13}$$

$$\hat{\sigma}(y) = \begin{cases} p'(q(y))\sigma(q(y)) & p(-\infty) < y < p(\infty) \\ 0 & otherwise \end{cases} \tag{2.14}$$

$$p(-\infty) < Y_0 < p(\infty) \tag{2.15}$$

It is now possible to draw conclusions about sufficient conditions for non explosion in the case when the process has drift. Since Y given by 2.13 does not explode and $Y_t = p(X_t)$ we can guarantee that X_t does not explode by demanding the condition $p(\pm\infty) = \pm\infty$ to be fulfilled. Hence we have the sufficient condition that the process does not explode. This condition is however not necessary. Necessary and sufficient conditions are instead summarized in what is normally called the Fellers test for explosion.

Theorem 2.2.5 (Feller's test for explosions) *Given that the condition 2.10 and 2.11 holds and we let X be a strong solution to the SDE in 4.1 in a domain* $I = (l, r)$ *with some nonrandom initial condition* x_0*, then* $P[S := \inf\{t \geq: X_t \notin (l, r)\} = \infty] = 1$ *if we can guarantee that* $v(r-) = v(l+) = \infty$ *where:*

$$v(x) = \int_{x_0}^{x} p'(y) \left(\int_{x_0}^{y} dm(z)\right) dy = \int_{x_0}^{x} (p(x) - p(y)) dm(y) \tag{2.16}$$

And the measure m is the **speed measure** *of the process defined through:*

$$\frac{dm(x)}{dx} = \frac{2}{\sigma(x)^2 p'(x)} \tag{2.17}$$

2.3 Stationarity of solutions to SDEs

Later on we will look at processes that experience stationarity. By this we mean that there exists an invariant density $\pi(x)$ that satisfies:

$$\pi(y) = \int p(t, x, y)\pi(x) dx \tag{2.18}$$

here $p(t, x, y)$ being the transition density for the process to move from x to y in time t. By this equation the probability inflow to a point y is equal to the probability outflow. The stationary density is given by the solution to the *Fokker Planck*-equation.

$$-\frac{\partial \pi}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\sigma^2(x)\pi(x)\right) - \frac{\partial}{\partial x}\left(b(x)\pi(x)\right) = 0 \tag{2.19}$$

Since $\pi(x)$ is stationary the time derivative is zero. Using integrating factor it is easy to obtain the solution

$$\pi(x) = \frac{C}{\sigma^2(x)} \exp\left(\int_{x_0}^{x} \frac{2b(y)}{\sigma^2(y)} dy\right) \tag{2.20}$$

Using the relation in 2.17 we see that the stationary is, except for a normalisation constant, equal to the first derivative of the speed measure with respect to x. This will later on when finding the stationary distribution of the CKLS process.

Chapter 3

Volatility Induced Stationarity - VIS

In the previous section we summarised the theory behind existence and uniqueness of solution. Perhaps most useful results was Feller's test for explosions together with theorem 2.1.2. Together they provide us with the tools for determine if a specific time homogenous SDE has a unique strong solution and whether this solution will explode or not in finite time.

Most of the previous theory has been known for a decades and is mentioned in introductory books on stochastic calculus such as [5]. We now turn to a more recent field of research which deals with diffusions[1] exhibiting *Volatility induced stationarity* (VIS). These are strong solutions to SDE which fulfils the Feller's test for explosion but where the stationarity properties is causes by high volatility. Even though the term VIS has not been used explicitly in literature and research it has been studied by Conley et al [3] and a generalized definition is given in Muszta [1] and Muszta et al [11]. In this section the definition of VIS will be made and the general features will be described.

Normally when thinking of solutions to SDEs we can split up the problem into a deterministic and a stochastic part (diffusion term). The deterministic will then be an ordinary differential equation (ODE) and the corresponding solution gives a rough indication of how the solution will behave. The full solution is then obtained by adding the stochastic part to the deterministic. This way of thinking is especially valid in cases when the stochastic part is small compared to the solution of the ODE. In this case the stochastic part will have limited influence on e.g. the convergence and long time stationarity of the deterministic problem.

The concept of VIS was introduced for processes which opposed this tradi-

[1] By diffusions we mean solutions to SDE which has continuous sample paths

tional way of thinking. In these types of processes the stochastic part cannot longer be considered independently since it interacts with the deterministic part. As a direct consequence it is possible for it to induce properties of the solution that is not present in the absence of the stochastic term. The idea behind the definition of VIS is to characterize SDEs where the stationarity is partly or solely ensured by the structure of the diffusion function. Muszta [1] make the following definition of VIS:

Definition (VIS) The stationary solution to the time homogenous SDE given by 2.7 has *vis* at l (r) if $p'(l^+) \leq \infty$ ($p'(r^-) \leq \infty$). In this case we call l (r) a *vis*-boundary.

Starting from this definition it is possible to explain how volatility can induce stationarity. To do this we recall the definition of the speed measure:

$$m(dx) = \frac{2dx}{p'(x)\sigma(x)} \tag{3.1}$$

The speed measure is recognized as the (unnormalized) distribution for the process. For a process to be stationary we must have $m(I) < \infty$ where I is the domain in which the process takes its values. For a stationary process without drift ($p'(x) = 1$) we must have that $\sigma(x) \to \infty$ as $x \to \infty$ in order for stationarity to hold. Assume now that such process moves to high numeric values. In this case the $\sigma(x)$ will increase and hence the speed measure will decrease. In this case the process will tend to move back down to values where the process is locally unbiased. In other words it is possible for the process to take on large values, however the time spent at these values will be very short thus causing the speed-measure to create mean reversion. This mean reversion will result in the local martingale not being a martingale.

From the argument above it is obvious that there is a trade-off between the martingale property and stationarity of the process. In order for a strict diffusion process (no drift) to be stationary it is necessary one has a high volatility to create mean reversion. The price we have to pay for it is that the local martingale is not a martingale.

If we consider processes with drift a similar argument can be used by, once again, using the VIS definition. However we still want to separate the cases where the stationarity is caused by the volatility and not the mean reversion by the drift. But this follows directly from the definition of the scale function since the VIS condition $p'(r) < \infty$ imply that the $\sigma^2(x)$ must dominate over the drift near the VIS boundaries.

Chapter 4

CIR and CKLS processes

The specific types of processes that will be considered in this thesis are the CIR [10] and CKLS[4] processes. Both named after the authors which first presented the widespread application of the processed in terms of modelling the short term interest rates. In terms of SDEs in equation 4.1 the processes are characterized by a parameter set $\{\alpha, \beta, \sigma, \gamma\}$ and more specifically by:

$$X_t = X_s + \int_s^t \alpha + \beta X_u du + \int_s^t \sigma X_u^\gamma dW_u \tag{4.1}$$

The CIR process is the specific case of the CKLS process in the case when $\gamma = 0.5$. The existence of solution, non explosive property and the stationarity property is now analyzed. Also the existence of transition densities will be considered by using the theory outlined in the previous two chapters.

4.1 Properties of solution

Strong existence and uniqueness of solutions to the CKLS SDE is given by Theorem 2.1.2. By identifying $h = x^p$ where $p \geq 1/2$ we can guarantee that strong uniqueness holds in cases where $\gamma \geq 1/2$. The solution will however only exists locally since the linear growth condition of theorem 2.1.1 is not fulfilled. The strong solution hence only exists up to a possible explosion time.

To check whether the process explodes or not we simply check the conditions derived in Section 2.2 of Chapter 2 and ended up in the Fellers Test for Explosion. In cases when $\gamma \geq 1/2$ we have that the process moves in the domain $I = (0, \infty)$. Furthermore we have from definition of the speed measure in this case that $m(I) \leq \infty$ and $\lim_{x\to 0} p(x) = \lim_{x\to\infty} p(x) = \infty$. This latter condition ensures that $p'(x)$ will go to infinity at x=0 and be bounded away from zero at infinity. This ensures that the CKLS process does not explode in this case.

4.2 The VIS effect

By using the definition of VIS presented in the previous section it can easily be verified that the CKLS process experience VIS for the following four sets of parametric values: $\{1/2 < \gamma < 1, \alpha > 0, \beta = 0\}$, $\{\gamma = 1, \alpha > 0, 0 < \beta < 1/2\sigma^2\}$, $\{\gamma > 1, \alpha > 0\}$, $\{\gamma > 1, \alpha = 0, \beta > 0\}$

From the discussion in Chapter 3 the presence of VIS prevents the local martingale part of the SDE (second term in 4.1) from being a martingale and hence induces a mean reversion in addition to the one caused by the drift. The direct consequence of this is that traditional assumptions for the stationary mean made by for example Chan et al [4] does not hold. By not taking the VIS into account in parameter estimation methods may break down since moments conditions do not longer hold.

Moments conditions under VIS To derive valid moments conditions it is necessary to obtain the expression for the stationary density. This is simply given by the derivative of the speed measure as was pointed out in Section 2.3 of Chapter 2:

$$\frac{dm(x)}{dx} = 2x^{-2\gamma} \exp\left(\frac{2\alpha}{\sigma^2(1-2\gamma)}x^{1-2\gamma} - \frac{\beta}{\sigma^2(\gamma-1)}x^{2-2\gamma}\right) \tag{4.2}$$

From the stationary density it is then possible to derive the moment conditions:

$$\mathbf{E}[X_t^p] = \int_0^\infty x^p dm(x) = \int_0^\infty 2x^{p-2\gamma} \exp\left(-\frac{\beta}{\sigma^2(\gamma-1)}x^{2-2\gamma}\right) dx =$$

$$= \left(\frac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{-p/(2\gamma-1)} \frac{M(p)}{M(0)}$$

where

$$M(p) = \sum_{k=0}^{\infty} \frac{1}{k!} \left[\frac{-\beta}{\sigma^2(\gamma-1)} \left(\frac{\sigma^2(2\gamma-1)}{2\alpha}\right)^{-p/(2\gamma-1)}\right]^k \times$$
$$\Gamma\left(k\frac{2\gamma-2}{2\gamma-1} + 1 - \frac{p}{2\gamma-1}\right)$$

And Γ being the gamma function. In the specific case where p=1 we obtain the expected value $\mathbf{E}[X_t]$ of the process. Especially, performing the direct calculation, we obtain the following result.

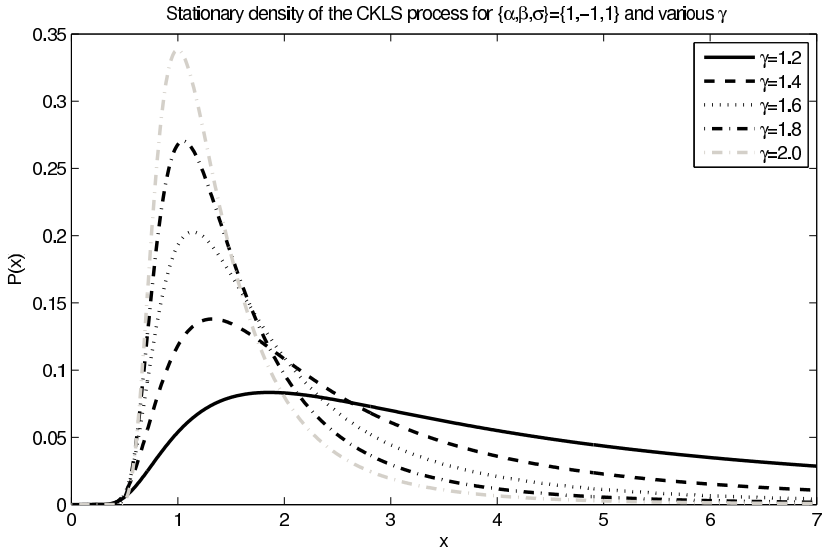

Figure 4.1: Stationary density of CKLS for $\{\alpha, \beta, \sigma, \gamma\} = \{1, -1, 1, \gamma\}$ for different values in γ. For high gamma it is obvious that the long time expected value of the process decreases since there is an increase in probability of the process taking small values. This is caused by the mean reversion from the volatility and the proposition given below.

Proposition In the case when $\gamma > 1$ we have that $\mathbf{E}[X_t] \in (0, -\alpha/\beta)$

What this proposition tells us is that if $\gamma > 1$ we will experience mean reversion from the local martingale part of the SDE. Consequently the expected value of the process will be lower than in absence of VIS in which $\mathbf{E}[X_t] = -\alpha/\beta$. It is possible to go even further and characterize the VIS effect in terms of the size of the mean reversion caused by the local martingale. This we do by taking expectation on both sides of the CKLS SDE and using Fubini's theorem. This yields the following proposition:

Proposition In case $\gamma > 1$ we have:

$$\mathbf{E}[\sigma \int_0^t X_s^\gamma dW_s] = -\left(\alpha + \beta \int_0^\infty x d\pi(x)\right) t := -d(\alpha, \beta, \sigma, \gamma)t \qquad (4.3)$$

Where $d(\alpha, \beta, \sigma, \gamma) > 0$ and π is the normalized stationary density given by the (normalized) differential of the speed measure (see above).

What this proposition tells us is that in case of VIS when $\mathbf{E}[X_t] \in (0, -\alpha/\beta)$ the Itô-intagral of the SDE will have drift. In other words it is not correct

to make a martingale approximation. This approximation is a standard approach in simple parameter estimation methods (see [4]) and will be considered below in relation to the General Method of Moment (GMM).

4.3 Transition Densities

As we will see later on the simplest way to estimate parameters for SDEs is to use the transition density defined by:

$$p(X_{t+\Delta}|X_t, \theta, \Delta t) \tag{4.4}$$

Given an analytic expression of this probability density it is only a question of choosing the right optimization algorithm in order to apply maximum likelihood and find specific values on the parameter vector θ that best agree with observed data. The problem for most SDEs are that such analytic expression does not exist. In this case one has to turn to more delicate approximation schemes for parameter estimation. In 1985 Cox Ingersoll and Ross derived the analytic expression for the transition density for the CIR process. It was shown to be a product of an axponential function and a modified Bessel function according to:

$$p(X_{t+\Delta}|X_t, \theta, \Delta t) = ce^{-(u+v)}\frac{v}{u}^{q/2} I_q(2\sqrt{uv}) \tag{4.5}$$

where

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2(1-e^{-\kappa\Delta t})} \\ u &= cX_t e^{-\kappa\Delta t} \\ v &= cX_{t+\Delta t} \\ q &= \frac{2\kappa\mu}{\sigma^2} - 1 \end{aligned} \tag{4.6}$$

and I_q is the modified Bessel function of order q. The more general CKLS process is however an example for which the transition density is unknown for most parameter sets. By simulating a large number of processes it is possible to get an empirical transition density from a histogram plot. Figure 4.2 show such histogram from 5000 simulation for the specific values on the parameters given by $\{\kappa, \mu, \sigma, \gamma\} = [0.2, 0.5, 0.1, \gamma]$ for $\gamma = 0.5, 1, 2$ respectively. The initial value X_0 was set to the long time average $\mu = 0.5$ while the time of evolution $\Delta t = 0.5$. It is obvious that a higher value on γ is associated with the higher mean reversion to the long time mean level. One thing that might not be equally obvious is that high values on γ is also associated heavier tails which is caused by the fact that the process might take high values for short periods of time.

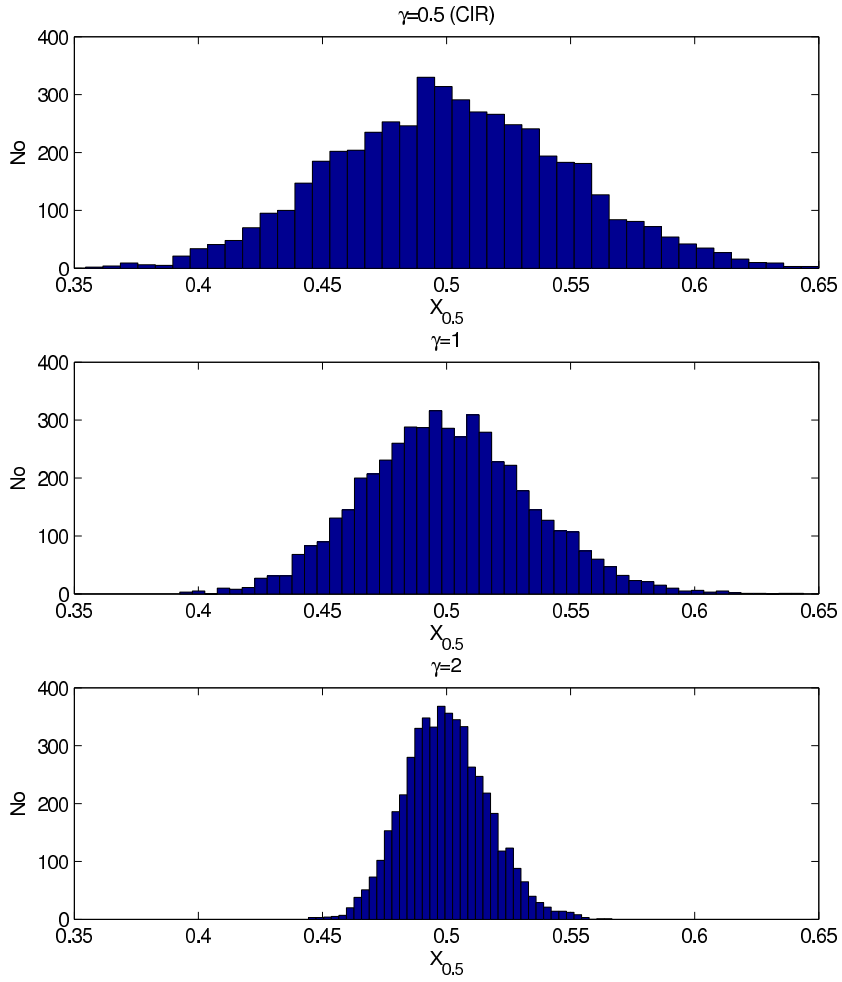

Figure 4.2: Histograms for transition density for different values on γ when $X_t = 0.5, \Delta t = 0.5$. It is obvious that the standard deviation of the transition probability decreases as the value on γ increases. Hence the γ value contribute to mean reversion to the long time stationary value of the process

Chapter 5

Parameter estimation methods

Given some arbitrary underlying model, there are basically two different approaches to estimate parameters from observed data. They both relies on the fact that estimated parameters will infer to the true parameters when the amount of data becomes large. The first class is Maximum likelihood methods where one estimate parameters through first order optimality conditions for the density function. In order to do this, the closed form expression for the transition density must be known. Even though a closed form expression is not known, it is possible to obtain an approximate closed form expression due to the recent methods based on Hermite polynomial expansion of the normal density. This method has proved to be efficient for transition densities in SDEs. The second class of methods is called Method of Moments in which one express the parameters in terms of the moments. Since, by the law of large numbers, the moments could be estimated from data, it is possible to find estimates of the parameters. Method of moments have been used extensively in relation to parameter estimation in SDEs since it is often relatively simple to find expression for the parameters in terms of first and second order moments. Examples of method based on a Method of Moment approaches is General Method of Moments (GMM), Simulated method of Moments (SMM) and Efficient Method of Moments (EMM). The differences between method comes from differences in initial assumptions. In the simple GMM model it is assumed that one have a known relationship between the parameters one want to estimate and the different moments. EMM and SMM do not assume that these relations are known and uses more delicate techniques in which help models are used.

In this chapter the theory behind Maximum Likelihood and Method of Moments will be explained in context of their application to parameter estimation in stochastic differential equation. As a general model for the SDE, a

continuous time parametric diffusion given by [5.1] will be used.

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \tag{5.1}$$

The functions $\mu(\cdot\,,\cdot\,)$ and $\sigma(\cdot\,,\cdot\,)$ are assumed to be known and the unknown parameter vector θ that we want to estimate belongs to some open bounded set $\Theta \subset R^K$. It should be pointed out that although the model is written in continuous time, the observed data are sampled in discrete time. Furthermore is should be elucidated that even though most of the theory will be applicable on multidimensional SDEs, the treatment will be restricted to one dimension.

5.1 Maximum likelihood approaches

In maximum likelihood theory it is assumed that one has a IID random sample $X_1, ..., X_n$ from a known k-variate distribution with density $f(\mathbf{x}|\theta_0)$ where $\theta_0 \subset \Theta$. Due to independence of the $X_i's$ it is possible to write the joint as a product of marginal densities $f(\mathbf{x}|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta_0)$. The likelihood function is obtained by replacing the nonrandom arguments x_i by the corresponding (k-dimensional) random vectors X_i and θ_0 by θ:

$$\hat{L}_n(\theta) = \prod_{i=1}^{n} f(Z_i|\theta) \tag{5.2}$$

The maximum likelihood (ML) estimator is then set to the value that maximize the likelihood function (or equivalently the log likelihood function).That is:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \ln \hat{L}_n(\theta) \tag{5.3}$$

The guarantee the uniqueness of $\hat{\theta}$ it is enough to prove that the log likelihood function is nowhere constant on Θ, something that must be proved from one case to another. For a proof of this statement turn to e.g Bierens pp 205-205 [9]

A very important property of the ML-estimator is that it is asymptotically normally distributed as $n \to \infty$. Also it is *asymptotically efficient* meaning that the likelihood-estimate defined by 5.3 results in the lowest possible variance of the estimator, compared to if we would minimize the argument by some other function different from 5.2. The efficiency property will be merely taken as a fact. However the asymptotically normality can be relatively easily understood by using the first order condition for optimality at $\theta = \hat{\theta} \in \Theta$ given by 5.4. Here we assume that the log likelihood is twice continuously differentiable in a neighbourhood of the true parameter θ_0 and that θ_0 is in the interior of Θ.

$$\nabla \hat{L}_n(\theta|X) = \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \ln f(X_i|\hat{\theta}) = \mathbf{0} \tag{5.4}$$

Using the first order condition and Taylor expand to the second order around θ_0 then gives an estimate for the difference $(\hat{\theta} - \theta_0)$ in the neighbourhood of θ_0:

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \ln f(X_i|\theta_0) + \frac{1}{n}\left[\sum_{i=1}^{n} \nabla_{\theta\theta} \ln f(X_i|\xi)\right](\hat{\theta} - \theta_0) = \mathbf{0} \Leftrightarrow$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = \left[-\frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta\theta} \ln f(X_i|\xi)\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nabla_\theta \ln f(X_i|\theta_0)$$

The first factor converges to $A = E[\nabla_{\theta\theta} \ln f(x|\theta_0)]^{-1}$ by the large numbers while the second factor converges, by the central limit theorem, in distribution to the multivariate normal with mean 0 and variance equal to the Fisher information Matrix defined by $I(\theta) = E[(\nabla_\theta \ln f(X_i|\theta_0))^2]$. From Slutsky's theorem ($Y_n^{-1}X_n \to c^{-1}X$ given that Y_n converges to c in probability and X_n converges to X in distribution) hence it is possible to draw the conclusion:

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, A^{-1}IA^{-1})$$

Especially if the model is correctly specified then A=I by the information equality. Hence one get that $\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, I^{-1})$

5.1.1 Simple Maximum likelihood in SDE

The theory on maximum likelihood is well suited for application to stochastic differential equation since the assumption on independence for the random sample $X_1, ..., X_n$ follows from the Markov property of the solutions to SDEs. If the transition density for SDE in 5.1 is denoted by $p(X_{t_{i+1}}|X_{t_i}; \theta, \Delta t)$ for an equidistant time discretization $\Delta t = t_{i+1} - t_i$, then the log likelihood function has the form:

$$\ln \hat{L}_n(\theta) = \sum_{i=1}^{n} \ln p(X_{t_{i+1}}|X_{t_i}; \theta, \Delta t) \tag{5.5}$$

5.1.2 Approximate Maximum likelihood

The major drawback of the Maximum Likelihood method in estimating parameters is that one must have a closed form expression for the density function for the random variable X above. Especially in the context of

stochastic differential equations this is an obvious disadvantage since the corresponding transition density for a process is almost never known. Only for the simplest types of SDEs such as the Ornstein and Uhlenbeck (OU) or Cox Ingersoll and Ross (CIR) one have succeeded in deriving closed form expressions. For the more general CKLS model considered the method of Maximum likelihood seems inapplicable.

An approach to overcome the problem of unknown closed form expression is to construct a sequence of closed form functions that converges to the true likelihood function. One such method has been developed by At-Sahalia [16] and which uses a Hermite series expansion around the N(0,1)-density for the transition density.

Although most SDEs' do not have transition densities that are close to a N(0,1) density, it is often possible to trans form it into another process which does. The transition density for the original process is then obtained by simple back transformation. This is explained below:

Transforming SDE to unitary diffusion

Suppose that one have a SDE on the form in 5.1. Assuming that the functions $\mu(x,\theta)$ and $\sigma(x,\theta)$ are infinitely differentiable in x and three times continuously differentiable in θ with $\sigma(x,\theta) > c > 0$, it is possible to make the transformation:

$$Y = \gamma(X;\theta) = \int^{X} \frac{du}{\sigma(y;\theta)}$$

Now by applying Itô's Lemma one get the following SDE for Y:

$$dY_t = \mu_Y(Y_t;\theta)dt + dW$$
$$\mu_Y(y,\theta) = \left(\frac{\mu(\gamma^{-1}(y;\theta))}{\sigma(\gamma^{-1}(y;\theta))} - \frac{1}{2}\sigma'(\gamma^{-1}(y;\theta))\right)$$

Making the transition from X_t to Y_t now makes it possible to construct an expansion of the transition density for Y_t. However since it is the transition density for X_t that we are interested it is necessary to find the relation between the two densities. This could be found by using Jacobian formula:

$$\begin{aligned} p_X(\Delta, x|x_0;\theta) &= \frac{\partial}{\partial x} P(X_{t+\Delta} \leq x | X_t = x_0;\theta) \\ &= \frac{\partial}{\partial x} P(Y_{t+\Delta} \leq \gamma(x;\theta) | Y_t = \gamma(x_0,\theta);\theta) \\ &= \frac{\partial}{\partial x}\left[\int^{\gamma(x;\theta)} p_Y(\Delta, y|\gamma(y_0;\theta))dy\right] \\ &= \frac{p_Y(\Delta,\gamma(x;\theta)|\gamma(x_0;\theta);\theta)}{\sigma(\gamma(x;\theta);\theta)} \end{aligned} \tag{5.6}$$

The transformed process Y_t has the desired process to be normally distributed. However in cases when the time increment Δ is small the transition density becomes close to a dirac shape. Therefor we use a second transform to get around this problem by defining the process:

$$Z = (Y - y_0)/\Delta^{1/2} \tag{5.7}$$

Using a similar Jacobian derivation as in 5.6 we obtain the relationship in transition density between Y and Z by:

$$p_Y(\Delta, y|y_0;\theta) = p_Z(\Delta, (y - y_0)/\Delta^{1/2}|y_0;\theta)/\Delta^{1/2} \tag{5.8}$$

Now that the relationship between densities for X Y and Z have been established we have taken the first step in approximating the transition density for X. If we are able to find an approximation for the transition density for Z, which is fairly close to N(0,1), we can find transition density for X by the above transformations.

Obtaining closed form expression for transition density

The fact that the transition density for Z is fairly close to the standard normal distribution makes it possible to expand the density in terms of Hermite polynomials through:

$$p_Z^J(\Delta, z|y_0;\theta) = \phi(z)\sum_{j=0}^{J}\eta_Z^j(\Delta, y_0;\theta)H_j(z) \tag{5.9}$$

Where $\phi(z)$ denotes the standard normal density, η_Z denotes the polynomial expansion coefficients and H_j is the j:th Hermite polynomial defined by:

$$H_j(z) = \sqrt{j}e^{z^2/2}\frac{d^j}{dz^j}\left[e^{-z^2/2}\right] \tag{5.10}$$

The reason to use the Hermite polynomial instead of some other polynomial is that they are orthogonal with respect to the $L^2(\phi)$ scalar product weighted by the Normal density. That is:

$$\langle H_j(z), H_i(z)\rangle_{L^2(\phi)} := \int_{-\infty}^{\infty} H_j(z)\phi(z)H_i(z)dz = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \tag{5.11}$$

The orthonormal property of the Hermite polynomials in $L^2(\phi)$ tells us that it constitute a base for $L^2(\phi)$ and hence any function in $L^2(\phi)$ (e.g a proper transition density) can be expressed as a linear combination of Hermite polynomials. Due to the orthonormal property it is possible to find

the coefficients in the expression for p_Z in 5.9 by multiplying by $H_j(z)$ and integrating giving:

$$\eta_Z^j(\Delta, y_0; \theta) = \frac{1}{\sqrt{j!}} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z|y_0; \theta) dz \tag{5.12}$$

$$\begin{aligned}
\eta_Z^j(\Delta, y_0; \theta) &= \frac{1}{\sqrt{j!}} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z|y_0; \theta) dz \\
&= \frac{1}{\sqrt{j!}} \int_{-\infty}^{\infty} H_j(z) \Delta^{1/2} p_Y(\Delta, \Delta^{1/2} z + y_0|y_0; \theta) dz \\
&= \frac{1}{\sqrt{j!}} \int_{-\infty}^{\infty} H_j(\Delta^{-1/2}(y - y_0)) p_Y(\Delta, y|y_0; \theta) dy \\
&= \frac{1}{\sqrt{j!}} E[H_j(\Delta^{-1/2}(Y_{t+\Delta} - y_0))|Y_t = y_0; \theta)]
\end{aligned} \tag{5.13}$$

For the case $J = \infty$ up until now there has just been a matter of reformulating the transition density p_Z of Z into a series of conditional moments. Conditional moments obtained in 5.13 is however a very convenient form in expression transition density since the mapping $s \mapsto E[f(Y_{t+s}, y_0)|Y_t = y_0]$ can be approximated by simple Taylor expansion.

In order to Taylor expand the function we may use the fact that the time differential in the expected value of f(X) (where X is governed by the SDE in 5.1) is given by its infinitesimal generator $\mathcal{A}$ through:

$$\lim_{t \downarrow 0} \frac{1}{t} \left[E[f(Y_{t+s})|Y_t = y_0] - f(y_0) \right] = \mu_Y(y_0; \theta) \frac{\partial f}{\partial y} + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} := \mathcal{A}_\theta f(y_0) \tag{5.14}$$

Denoting $A(\theta) = \mathcal{A}_\theta f(y_0)$ we can make the following Taylor expansion:

$$\begin{aligned}
E[f(Y_{t+s})|Y_t = y_0] = &\sum_{k=0}^{K} A^k(\theta) f(y_0, y_0) \frac{\Delta^k}{k!} \\
&+ E[A^{K+1}(\theta) f(Y_{t+\Delta}, y_0)| \frac{\Delta^{K+1}}{(K+1)!}
\end{aligned} \tag{5.15}$$

We restrict ourselves to the second order Taylor expansion (K=2) and derive the coefficients η in the expression for the transition density p_Z in 5.9. Here

$$\eta_Z^1 = \sum_{k=0}^{2} (A^k(\theta) f)(y_0) = \sum_{k=0}^{2} \left(\mu_Y(;\theta) \frac{\partial}{\partial y} + \frac{1}{2} \frac{\partial^2}{\partial y^2} \right)^k H_1(y_0) \frac{\Delta^k}{k!} =$$

$$= \left(\mu_Y(y_0; \theta) \frac{\partial H_1(s)}{\partial s} \frac{1}{\Delta^{1/2}} + \frac{\partial^2 H_1(s)}{\partial s^2} \frac{1}{\Delta} \right) \Delta$$

$$+\left[(2\mu_Y\mu'_Y+\mu''_Y)\frac{\partial H(s)}{\partial s}\frac{1}{\Delta^{1/2}}-(\mu_Y^2+2\mu'_Y)\frac{\partial^2 H(s)}{\partial s^2}\frac{1}{\Delta}+\right.$$
$$\left.+\frac{3}{2}\mu_Y\frac{\partial^3 H(s)}{\partial s^3}\frac{1}{\Delta^{3/2}}-\frac{1}{4}\frac{\partial^4 H_1(s)}{\partial s^4}\right]\frac{\Delta^2}{2!} \tag{5.16}$$

Here we have made a change of variables s=$(y-y_0)/\Delta^{1/2}$. Observing that H_1 is a polynomial of first degree all higher order derivatives than one will be zero and hence $\eta_Z^0=1$ and the first coefficients are given by:

$$\begin{aligned}
\eta_Z^1 &= -\mu_Y\Delta^{1/2}-(2\mu_Y\mu'_Y+\mu''_Y)\Delta^{3/2}/4\\
\eta_Z^2 &= (\mu_Y^2+\mu'_Y)\Delta^2+(6\mu_Y^2\mu'_Y+4\mu'^2_Y+7\mu_Y\mu''_Y+2\mu'''_Y)\Delta^2/12\\
\eta_Z^3 &= -(\mu_Y^3+3\mu_Y\mu'_Y+\mu''_Y)\Delta^{3/2}/6\\
&\quad -(12\mu_Y\mu_Y+28\mu_Y\mu'^2_Y+22\mu_Y^2\mu''_Y+24\mu'_Y\mu''_Y+12\mu_Y\mu'''_Y+2\mu_Y^{(4)})\Delta^{5/2}/48
\end{aligned} \tag{5.17}$$

Once the analytic expression for the coefficients has been determined we seek to rearrange the expression for the transition density given in [5.9] so that terms are arranged in terms of their powers of Δ and not in the order of the Hermite polynomials. At-Sahalia [16] derive an analytic formula for this. Using the same notation it is given by:

$$p_Y^K(\Delta,y|y_0;\theta)=\Delta^{-1/2}\phi(\frac{y-y_0}{\Delta^{1/2}})e^{\int_{y_0}^{y}\mu_Y(u,\theta)du}\sum_{k=0}^{K}c_k(y|y_0;\theta)\frac{\Delta^k}{k!} \tag{5.18}$$

where

$$\begin{aligned}
c_j(y|y_0;\theta) &= j(y-y_0)^{-j}\int_{y_0}^{y}(u-y_0)^{j-1}\\
&\quad\times\left[\lambda_Y(u;\theta)c_{j-1}(y|y_0;\theta)+\frac{\partial^2 c_j(y|y_0;\theta)}{\partial u^2}\right]du
\end{aligned} \tag{5.19}$$

and

$$\lambda_Y(y;\theta)=-\frac{1}{2}\left(\mu_Y^2(y;\theta)+\frac{\partial\mu_Y(y;\theta)}{\partial y}\right) \tag{5.20}$$

5.2 Method of Moments approaches

As in the previous case, assume that one has a collection if data from some random variable X with some distribution G, and that one want to determine the parameters for this distribution. The basic principle behind the Method of Moments is that there exists a direct relationship between the parameters

one want to estimate $\{\theta_1, \theta_2, \theta_3, ..\}$ and the moments $\{\mu_1, \mu_2, \mu_3, ..\}$ [1]. That is one have a vector of functions on the form $\theta_k = f_k(\mu_1, \mu_2, ...)$. Now, assuming that the number of data points $\{X_i\}$ are large, it is possible to make estimates of the moments using the Law of large numbers and find the kth sample moment by $\hat{\mu}_k = \sum_{i=1}^{n} X_i^k$. The direct relationship between $\{\theta_k\}$ and $\{\mu_i\}_{i=1}^n$ through f_k then gives an estimator $\hat{\theta}_k$. When applying this simple idea to SDEs it is rather straight forward and the only problem left is to specify the relations $\{f_k\}_{i=1}^n$ given the characteristic functions $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ for the SDE in [5.1]. This *General Method of Moment* (GMM) was first presented by Hansen [12] and often constitute the simplest way to estimate parameters in SDE. A generalization of the GMM method to cases where the relationships given by $\{f_k\}_{i=1}^n$ cannot be found have been developed by Gallant and Tauchen [14] and uses an "artificial" moment condition which approximates the true one. Below the GMM as well as Gallant and Tauchens Efficient Method of Moment (EMM) are described together with some important limit theorems for the estimates.

5.2.1 General method of moment

Assume that one have observed a series of data points $\mathbf{X}_n$ from an SDE with some specific drift $\mu(\cdot, \cdot)$ and diffusion $\sigma(\cdot, \cdot)$. In the General Method of Moment the functions $\mathbf{f}(\mathbf{X_n}, \theta) = [f_1(\mathbf{X_n}, \theta), ..., f_k(\mathbf{X_n}, \theta), ...]$ are assumed to be known and hence it is possible to construct vector functions $\mathbf{h}(\mathbf{X_n}, \theta)$ such that:

$$E[\mathbf{h}(\mathbf{X_n}, \theta)] = E[\mathbf{f}(\mathbf{X_n}, \theta) - \theta] = 0 \tag{5.21}$$

The intuition behind the formulation above is the following. Given the true parameter set $\theta_\mathbf{0}$ then the expectation of the vector $\mathbf{h}$ should equal zero. Now if we define $p = dim(\theta)$ and note that $\mathbf{h}(\mathbf{X_n}, \theta)$ is a r-dimensional vector function, that is $\mathbf{h}(\mathbf{X_n}, \theta)$:$\mathbb{R}^p \to \mathbb{R}^r$, it is necessary that $r \geq p$ in order for the equation system to be solvable. By applying the law of large numbers it is possible to approximate the expected value by:

$$\mathbf{g}(\mathbf{X_n}, \theta) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{h}(\mathbf{X_n}, \theta) \tag{5.22}$$

Since by [5.21] this expression should average to zero it is possible to obtain an estimate for θ by solving the equation $\mathbf{g}(\mathbf{X_n}, \theta) = \mathbf{0}$.

Even though it might be possible to obtain the estimete of θ by solving this (often nonlinear) equation system it gives no indication of the efficiency of the convergence of the estimator. Somehow it is necessary to develop a

[1] Here the usual definition of moment is used: $\mu_k = E[X^k]$

method such that the distribution of the estimator is proper in a way that it does not contain too heavy tails in a certain dimension. A very pleasant method for making the estimator efficient is to obtain the estimator by minimizing the quadratic function given by:

$$Q_N(\theta, \mathbf{X_n}) = [\mathbf{g}(\mathbf{X_n}, \theta)]^T \mathbf{\Sigma}_N [\mathbf{g}(\mathbf{X_n}, \theta)]^T \tag{5.23}$$

Here $\mathbf{\Sigma}_N$ is the positive semidefinite matrix obtained by inverting the covariance matrix obtained from the moment restrictions. Converting the estimation problem into a covariance-weighted minimization problem have two effects. First of all it is possible to guarantee that a solution exists (although the problem might contain several local minima). The second effect is that since the quadratic form is weighted by the inverse covariance matrix, more weight is put in dimensions where the variance is large end vice verse.

5.2.2 Efficient method of moment

The general method of moment presented above is readily applicable once the moments are known. However in many cases they are not known or the calculations are to cumbersome to perform. In either case it might be possible to turn to methods where explicit expressions for the moments are not necessary. One such method has been developed by Gallant and Tauchen [14] and is normally referred to as the Efficient Method of Moment (EMM).

The method uses two models for the underlying process - the structural model and the approximated help model. The structural model corresponds to the "real" model and is the one from which our observed data are generated from. It is the parameters in the structural model that we want to estimate. The help model is a more general model that we try to fit in order to replicate the structural model.

In order to estimate the structural model's parameter vector ρ one tries to minimize the difference between moments generated by the help model, where in one case the observed data has been used, and in the second case, simulated data from the structural model have been used. The estimate for ρ is the values that best agree with observed data. How to find an appropriate help model will be explained later and for now we will assume that we have the help model with transition density denoted by $f(y|x\theta)$. We will further denote that the observed data from the SDE by $\{\tilde{y}_t, \tilde{x}\}_{t=1}^n$. We assume that the observed data have been generated from the sequence of densities from the structural model $\{p_t(\tilde{y}_t, \tilde{x}; \rho^0)\}_{t=1}^n$.

Gallant and Tauchen now suggest that one take the expected values of the score vector $(\nabla_\theta \ln f_t(y_t|x_t;\theta))$ as the moment condition. That is if $m(\rho, \theta)$ denotes the moment condition we have:

$$m(\rho,\theta) = E\left[\nabla_\theta \ln(f(y_t|x_t;\theta))\right] = \int\int \nabla_\theta \ln(f(y_t|x_t;\theta))p(x,y|\rho)dydx \tag{5.24}$$

Since we are always free to chose the moment condition as we like this is a choice as good as any other. The condition given above has the property of equate to zero if the help model exactly replicates the behaviour of the structural model. In most cases the moment will deviate from zero since the help model is only an approximation. This is however not crucial since parameters will be found by minimizing the corresponding quadratic form in equation 5.25 where Σ is an appropriate weight matrix.

$$\rho = \arg\min_{\rho\in R} m(\rho,\theta)^T \Sigma m(\rho,\theta) \tag{5.25}$$

The above argument tells us that it is possible to find an estimate of ρ once we can approximate the integral in 5.24 in a suitable manner. The problem is however that neither the parameters θ in the help model or the transition density for the structural model is known. In addition it is necessary to approximate the expectation my Monte Carlo simulation in order to convert it into a numeric problem.

Approximating the moment condition

The fact that the transition density for the help model is known makes it possible to perform a Maximum Likelihood estimation of the parameter vector θ by using the first order optimality condition:

$$\nabla_\theta \ln f_t(y_t|x_t;\theta) = 0 \tag{5.26}$$

Now by using the observed values for the structural model $\{\tilde{y}_t,\tilde{x}\}_{t=1}^n$ it is possible to find the Quasi Maximum likelihood estimate of θ through:

$$\tilde{\theta} = \arg\max_{\theta\in\Theta} \ln f_t(\tilde{y}_t|\tilde{x}_t;\theta) \tag{5.27}$$

This estimate could then be inserted in the expression for the moment condition in 5.24. If the number of observed data are n we obtain the approximate moment by:

$$m_n(\rho,\tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\int \nabla_\theta \ln f(y|\tilde{x}_t,\tilde{\theta})p(y_t|\tilde{x}_t,\rho)dy \tag{5.28}$$

Still the transition density for the structural model $p(y|x,\theta)$ is not known. However it is still possible to simulate a large number of processes for a given parameter vector ρ and then use Monte Carlo Integration. Let N denote the

number of structural processes simulations in each time interval between observation, then the Monte Carlo estimation of the Moment is given by:

$$m_n(\rho, \tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{N}\sum_{\tau=1}^{N}\nabla_\theta \ln f(\hat{y}_{t\tau}|\tilde{x}_t, \tilde{\theta}_n) \tag{5.29}$$

Here $\hat{y}_{t\tau}$ denotes simulated values from the structural model with parameter vector ρ. The expression in 5.29 is now a function of ρ and by reformulating it as in 5.25 we have arrived at a solvable minimization problem with respect to ρ.

Convergence properties of the estimator

Assume now that we approximate the moment conditions $m(\rho, \theta)$ using a quasi maximum likelihood determination of θ and a Monte Carlo Integration over $p(y|x, \rho)$.To find an estimation of ρ through minimization of the equivalent problem given by 5.25 it then remains to specify the positive definite matrix Σ in a proper way. Proper in this context refers to the fact that the limiting distribution of the estimate as $n \to \infty$ is efficient in a way that the variance will be as small as possible. in order to obtain an efficient weight matrix it is necessary to find the limit distribution of the moment conditions since the efficient weight matrix is then only the inverse of the covariance matrix (Hessian) of the moment condition.

Observing that θ was estimated using Quasi maximum likelihood it possible to derive an analogue result to what was presented in Section 5.1. Especially if the help model has a transition density which fits the data well we have:

$$\sqrt{n}(\tilde{\theta}_n - \theta) \approx N(0, \tilde{I}_n) \tag{5.30}$$

Here $\tilde{I}_n$ is the approximation of the Fischer information matrix. If θ_0 represent the true value for the Quasi Maximum likelihood then the approximation has the form:

$$\begin{aligned}\tilde{I}_n &= \frac{1}{n}\sum_{i=1}^{n}\left[\nabla_\theta \ln f(\tilde{y}_t|\tilde{x}_t, \tilde{\theta}_n)\right]^T\left[\nabla_\theta \ln f(\tilde{y}_t|\tilde{x}_t, \tilde{\theta}_n)\right] \approx \\ &\approx \frac{1}{n}\sum_{i=1}^{n}\left[\nabla_\theta \ln f(\tilde{y}_t|\tilde{x}_t, \theta_0)\right]^T\left[\nabla_\theta \ln f(\tilde{y}_t|\tilde{x}_t, \theta_0)\right] = I_0\end{aligned}$$

Using that the moment condition $m_n(\rho, \tilde{\theta}_n)$ in our minimization problem is a direct function of $\tilde{\theta}_n$ it is possible to make a Taylor expansion around the true parameters (ρ_0, θ_0) to prove that:

$$\sqrt{n}m_n(\rho_0, \tilde{\theta}_n) \approx N(0, \tilde{I}_n) \tag{5.31}$$

Hence the the most efficient way to choose the positive semidefinite weight matrix to the minimization problem is to take $\Sigma = I_0^{-1} \approx \tilde{I}_n^{-1}$ and obtain:

$$\tilde{\rho}_n = \arg\min_{\rho \in R} m(\rho, \tilde{\theta}_n)^T \tilde{I}_n^{-1} m(\rho, \theta) \tag{5.32}$$

Once again the term efficient imply that it is not possible to find any other positive semidefinite weight matrix that gives a lower variance for the estimate of ρ when $n \to \infty$

Chapter 6

Simulation

Since it is our intention to estimate parameters for CKLS processes we must be able to construct solutions that is known to have given parameter values. That is we want to be able to simulate processes that obey the CKLS SDE. Without efficient simulation schemes we cannot tell whether the estimation errors comes from bad estimation methods or bad simulation schemes. The easiest way to find an closed form expression for X in 1.1 on the form $X = f(t, W_t)$ and then make realizations of the Brownian motion W. For the general CKLS such closed form expression does not exist and to solve such equations we must instead turn to appropriate simulation schemes that approximate the solution.

The simulation methods for SDE is very similar to the one that are used for numerical solutions to initial values problem in deterministic cases. The only differences are that in case of SDEs one must add the random noise to the solution. Since the existence of a unique, non exploding, strong solutions has already been proven in Chapter 4 it only remains to chose the scheme that best approximate the solution.

Unfortunately the theory in traditional textbooks such as Kloeden and Platen [13] does assume quite strong condition on the drift and diffusion coefficients. For SDEs which do not fulfil the conditions (which is often of globally Lipschitz-type) convergence of the numerical schemes cannot be guaranteed. Even though convergence of many schemes has been proved in latest years for less restricted conditions on drift and diffusion, (see for example Higham et al [15]) the convergence is often very slow and requires large amount of computational power. In this section we will present the elementary numerical schemes for simulation of solution together with more delicate methods appropriate for SDEs exhibiting *vis* characteristics.

6.1 The Euler scheme

The simplest way to simulate a solution to an SDE is to use an explicit scheme like the one used for solving initial value ODE in the deterministic case. The corresponding scheme for a general time homogenous SDE in 2.7 is the so called Euler Maruyama scheme given in equation 6.2. Here we have used abbreviation $\Delta_n = (t_{n+1} - t_n)$ and $\Delta W_n = (W_{t_{n+1}} - W_{t_n})$. In case of equidistance sampling points we write $\Delta_n = \Delta$.

$$X_{n+1} = X_n + b(X_n)\Delta_n + \sigma(X_n)\Delta W_n \tag{6.1}$$

Strong convergence of the Euler-Maruyama approximation to the true solution as $\Delta \to 0$ is guaranteed once we impose restriction on Lipschitz continuity and linear growth of the coefficients b and σ (see [5] section 9.6). Due to Gyong [7] we might relax the assumptions to the drift b(x) being one-sided Lipschitz and the volatility σ being locally Lipschitz and still have uniform a.s convergence.

6.2 Implicit approximation schemes

Just as in the deterministic case one could modify the explicit schemes to use information on future time steps. In this case we get a much more stable *implicit* schemes which have lower probability of collapse in finite time. This is especially desirable property when dealing with diffusions with *vis* since numerical approximations in general have high probability collapse. A general implicit computational scheme is given by 6.2.

$$Y_{n+1} = Y_n + (\theta_b \bar{b}(Y_{n+1}) + (1-\theta_b)\bar{b}(Y_n))\Delta + (\theta_\sigma \sigma(Y_{n+1}) + (1-\theta_\sigma)\sigma(Y_n))\Delta W_n \tag{6.2}$$

Here the parameters θ_b and θ_σ represents the degree of implicity in the for the drift and volatility coefficients respectively. We have three special cases: **1.** The Euler scheme (E) ($\theta_B = \theta_\sigma = 0$), **2.** The Stochastic theta model (ST), ($\theta_\sigma = 0$) and **3.** Fully implicit Euler model (FIE),($\theta_b = \theta_\sigma = 1$). The corresponding schemes for these models are given below in the above mentioned order:

$$Y_{n+1} = Y_n + b(Y_n)\Delta + \sigma(Y_n)\Delta W_n \tag{6.3}$$

$$Y_{n+1} = (\theta_b b(Y_{n+1}) + (1 - \theta_b b(Y_n)))\Delta + \sigma_(Y_n)\Delta W_n \tag{6.4}$$

$$Y_{n+1} = Y_n + \bar{b}(Y_{n+1})\Delta + \sigma(Y_{n+1})\Delta W_n \tag{6.5}$$

Explicit Euler method has already been mentioned above. The last two have been studied by Higham [2] and Muszta [1] respectively.

The general methodology for simulate implicit schemes is to divide the problem into one explicit part and one part containing the non-linear equation to solve for implicit step. For the CKLS model the problem comes down to solving $f(Y_{n+1}) = \hat{Z}_n$ where $\hat{Z}_n$ and $f(y)$ are given by the following:

$$\hat{Z}_n = Y_n + (1-\theta_b)(\alpha + \beta Y_n - \theta_\sigma \gamma \sigma^2 Y_n^{2\gamma-1})\Delta + (1-\theta_\sigma)\sigma Y_n^\gamma \Delta W_n \quad (6.6)$$

$$f(y) = y - \theta_b(\alpha + \beta y - \theta_\sigma \gamma \sigma^2 y^{2\gamma-1})\Delta - \theta_\sigma \sigma y^\gamma \Delta W_n \quad (6.7)$$

Simulations of the CKLS and CIR processes using the fully implicit Euler scheme are illustrated in Figure 6.1.

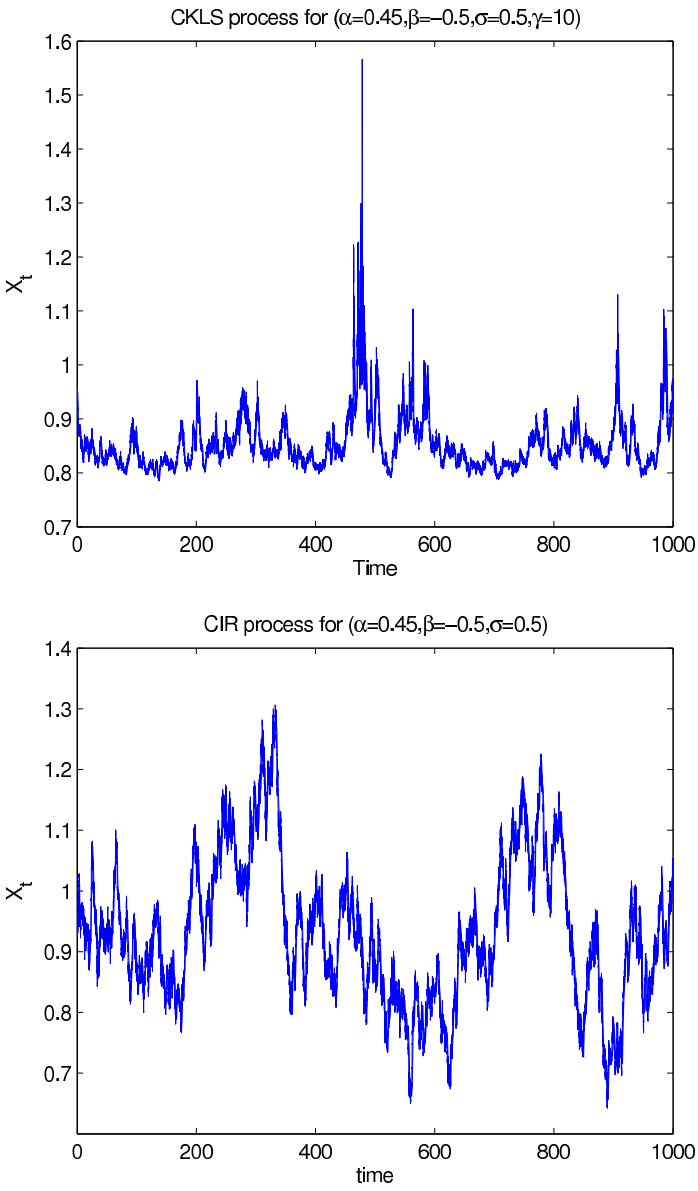

Figure 6.1: Simulations of CKLS and CIR processes when using the fully implicit Euler method. For the CKLS process the VIS effect is present and hence increase the mean reversion to the stationary mean 0.9. For the CIR process the mean reversion from VIS is not present and is only due to the drift term. Thus the process fluctuates more around 0.9. Also notice the high and short peaks present for the CKLS model.

Chapter 7

Implementation

The theory behind the estimation methods was outlined in Chapter 5. In order to implement them in the specific case of the CKLS model some computational aspects must be considered. These includes finding appropriate initial conditions for the algorithms, deriving moment conditions for GMM and obtaining explicit expression for the Taylor expansion of the Hermite polynomial expansion. It will turn out that implementation will be simpler if we rewrite the CKLS SDE given by 4.1 as the following:

$$X_t = X_s + \int_s^t \kappa(\mu - X_u)du + \int_s^t \sigma X_u^\gamma dW_u \tag{7.1}$$

From this representation it is easy to see that, in the case without VIS, the long time mean value of the process is μ. The parameter κ represent the speed of mean reversion caused by the drift term in cases when the process deviate from the mean. The diffusion term is unchanged.

7.1 Finding initial estimates for parameters

Whether we try to find parameters by making a maximum likelihood approach or a method of moment approach it is desirable to request some decent initial value for the parameters. By decent we here refer to an initial vector $\mathbf{x}_0$ sufficiently close to the actual vector so that we can guarantee that the optimization algorithms present in the estimation methods will converge to the global optima. In this thesis an Ordinary Least Square (OLS) method was used on the discretized versions of the CIR and CKLS processes. The discretized version of the two processes are given by (CIR case when $\gamma = 0.5$):

$$X_{t_\Delta t} - X_t = \kappa(\mu - X_t)\Delta t + \sigma X_t^\gamma \Delta W \tag{7.2}$$

Dividing by the X_t^γ on both sides then yield:

$$\frac{X_{t_\Delta t} - X_t}{X_t^\gamma} = \frac{\kappa(\mu - X_t)\Delta t}{X_t^\gamma} + \sigma \Delta W \tag{7.3}$$

Here we note that the random noise component ΔW is normally distributed with expectation zero. Hence we can find an estimation of the parameter triplet (μ, κ, γ) by inference and minimizing the OLS objective function:

$$(\hat{\mu}, \hat{\kappa}, \hat{\gamma}) = \arg\min_{\mu,\kappa,\gamma} \sum_{i=1}^{N-1} \left(\frac{X_{t_\Delta t} - X_t}{X_t^\gamma} - \frac{\kappa(\mu - X_t)\Delta t}{X_t^\gamma} \right)^2 \tag{7.4}$$

The estimator for the remaining parameter σ was then found by calculating the standard deviations of the residuals.

7.2 Moment conditions for GMM

When describing the GMM method in section 5.2.1 it is assumed that one can find a vector function $\mathbf{h}(\mathbf{X_n}, \theta)$ with expected value zero. To implement the theory of GMM to the CKLS or CIR SDE we need to specify these functions. In absence of VIS effects this can be done by approximating the Itô-integral representing the second term in the SDE as a martingale. Thus we can make the following approximations without worrying whether the GMM model giving the wrong result in the parameters. Here the time steps Δ are assumed to take small values and hence the drift, given by the first integral in Equation, can be approximated by:

$$\int_{t_i}^{t_{i+1}} \kappa(\mu - X_t)dt \approx \kappa(\mu - X_{t_i})\Delta \tag{7.5}$$

$$\int_{t_i}^{t_{i+1}} \sigma X_t^\gamma dW_t \approx \sigma X_{t_i}^\gamma \Delta W_t \tag{7.6}$$

Here ΔW_t is defined by $\Delta W_t = W_{t_i} - W_{t_{i-1}}$. Now the random error could be defined through $\epsilon_{t_{i+1}} = X_{t_{i+1}} - X_{t_i} - \Delta\kappa(\mu - X_{t_i})$. Hence one get the following expression for the first and second moments for $\epsilon_{t_{i+1}}$ under a time period of Δ:

$$\begin{aligned} \mathbf{E}(\epsilon_{t_{i+1}}|\mathcal{F}_{t_i}) &= 0 \\ \mathbf{E}(\epsilon_{t_{i+1}}^2|\mathcal{F}_{t_i}) &= \Delta\sigma^2 X_t^{2\gamma} \end{aligned} \tag{7.7}$$

Now using independence of increment following the Brownian motion we can seek inference for the parameter vector defined by

$$\mathbf{h}(\theta, \mathbf{X_{t_i}}) = \begin{pmatrix} \epsilon_{t_{i+1}} \\ \epsilon_{t_{i+1}} X_{t_i} \\ \epsilon_{t_{i+1}}^2 - \Delta\sigma^2 \\ (\epsilon_{t_{i+1}}^2 - \Delta\sigma^2 X_t^{2\gamma}) X_{t_i} \end{pmatrix} \tag{7.8}$$

Since the vector has expectation zero it is possible to use inference and the Law of large numbers telling us that the one can approximate expectation

$\mathbf{E}(h_{t_i}(\theta))$ by the average sum given by:

$$H_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} h_{t_i}(\theta) \tag{7.9}$$

To find Hansen [12] recommend to use optimization algorithms to find the parameter vector θ that solves the dual problem of minimizing the mean squared errors in the expression for $H_N(\theta)$. These optimization methods is readily available in Matlab. However one problem that remains is that the objective function contains many local minimum. To overcome this problem we used many initial values on the parameters. The initial values were randomly distributed around the estimated values from the previous section.

7.3 Explicit expression for the Hermite-polynomial expansion

Equation 5.19 in Section 5.1.2 gives a closed for expression for the coefficients in the Hermite polynomial expansion for the transition density. In order to apply this formula to the specific CIR and CKLS model it we need to find the explicit expression. Here we make the expansion to the second order to the transformed process Y with unitary diffusion. The transition density for X is then given by back transformation using the Relationship 5.6. We have the following expression for the expanded transition density:

$$p_Y^{(2)}(\Delta, y|y_0;\theta) = p_Y^{(0)}(\Delta, y|y_0,\theta)(1 + c_1(y|y_0,\theta) + c_2(y|y_0,\theta)) \tag{7.10}$$

It remains to estimate the specific structure of c_1 and c_2 for both the CIR and CKLS.

CIR

$$p_Y^{(0)}(\Delta, y|y_0,\theta) = \frac{1}{\sqrt{2\pi\Delta}} e^{-(y-y_0)^2/2\Delta - y^2\kappa/4 + \kappa y_0^2/4} y^{-0.5+2\mu\kappa/\sigma^2} y_0^{0.5-2\mu\kappa/\sigma^2}$$

$$\begin{aligned} c_1(y|y_0,\theta) = -\frac{1}{24yy_0\sigma^4}&(48\mu^2\kappa^2 - 48\mu\kappa\sigma^2 + 9\sigma^4 + y^2\kappa^2\sigma^4 y_0^2 \\ &+ y\kappa^2\sigma^4 y_0^3 + y\kappa^2\sigma^2(y^2\sigma^2 - 24\mu)) \end{aligned}$$

$$\begin{aligned} c_2(y|y_0,\theta) = &\frac{1}{576y^2y_0^2\sigma^8}\times \\ &(9(256\mu^4\kappa^4 - 512\mu^3\kappa^3\sigma^2 + 224\mu^2\kappa^2\sigma^4 + 32\mu\kappa\sigma^6 - 15\sigma^8) \\ &+ 6y\kappa^2\sigma^2(y^2\sigma^2 - 24\mu)(16\mu^2\kappa^2 - 16\mu\kappa\sigma^2 + 3\sigma^4)y_0 \\ &+ y^2\kappa^2\sigma^4(672\mu^2\kappa^2 - 48\mu\kappa(2 + y^2\kappa)\sigma^2 + (y^4\kappa^2 - 6)\sigma^4)y_o^2 \\ &+ 2y\kappa^2\sigma^4(48\mu^2\kappa^2 - 24\mu\kappa(2 + y^2\kappa)\sigma^2 + (9 + y^4\kappa^2)\sigma^4)y_0^3 \\ &+ 3y^2\kappa^4\sigma^6(y^2\sigma^2 - 16\mu)y_0^4 + 2y^3\kappa^4\sigma^8 y_0^5 + y^2\kappa^2\sigma^8 y_0^6) \end{aligned}$$

CKLS

Chapter 8

Evaluation

Although the estimation methods are readily applicable for all parameter sets we specifically turn to one where the CKLS experience VIS. Here the values $\{\mu, \kappa, \sigma, \gamma\} = \{0.2, 0.5, 0.1, \gamma\}$ was used for γ taking values between 0.5 (CIR) and 10. The reason for taking small values is that we are primarily interested in the relative size between the γ and other parameters. It is when this difference is large that we could expect the process to experience the extra mean reversion associated with VIS.

For each value on γ a total of 50 processes were used. The processes were simulated using the implicit Euler method presented in Chapter 6 to guarantee good approximations to solution. The time between simulation points $\Delta = 10^{-6}$ during a total time interval of T=10. A total of 1000 data points, with equal spacing between points, were then sampled for parameter estimation.

To see how well the estimation methods perform we consider the case where $\gamma = 0.5$ and we have the CIR-process. As pointed out in Chapter this process has a closed form expression for the transition density and hence it is possible to use simple Maximum likelihood to estimate parameters. The estimates from the Maximum likelihood thus can be used as a benchmark for measuring the relative strength of the other methods. This evaluation, for the parameters set $\{\mu, \kappa, \sigma, \gamma\} = \{0.2, 0.5, 0.1, 0.5\}$ is presented in Figure 8.1

For the general CKLS process the results from parameter estimation methods is given by Figures 8.2-8.5. For each parameter set and method the mean value of the estimate along with 0.25 and 0.75 quartiles are presented. In the estimate of γ in Figure 8.2 the residual $\hat{\gamma} - \gamma$ is plotted against γ.

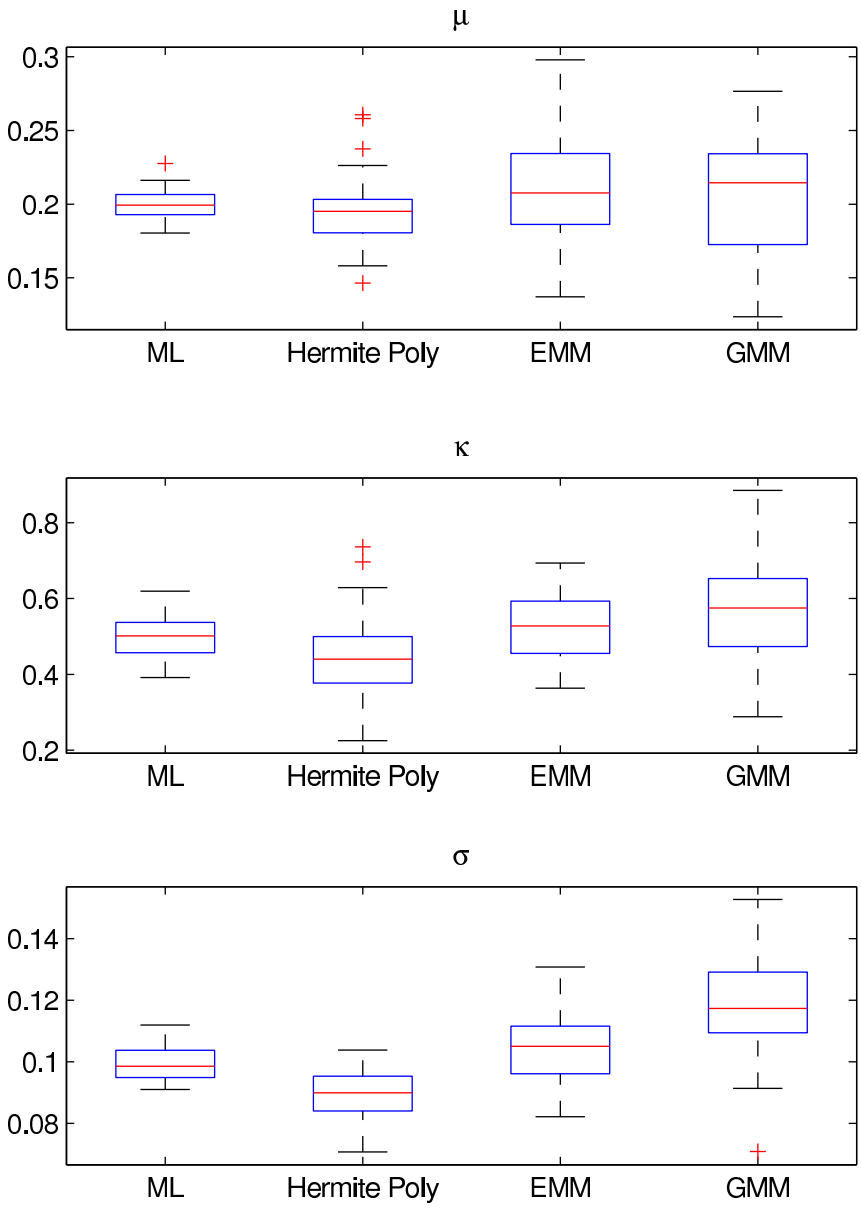

Figure 8.1: Boxplot of the obtained results from parameter estimation of 50 CIR Process. Boxes represents 25th and 75th quantiles along with median value. From this results it is obvious that the simple maximum likelihood (ML) out-performs other methods

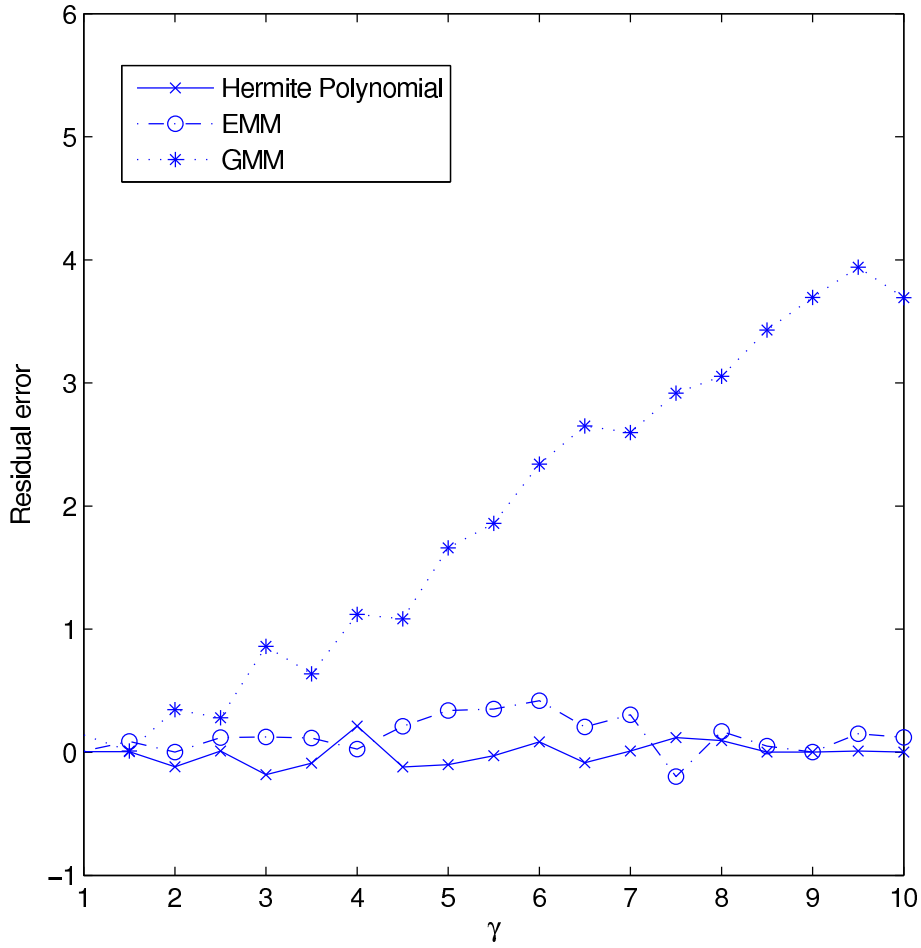

Figure 8.2: The residual error $\gamma - \hat{\gamma}$ between the actual gamma value and the mean of the estimated values obtained by different methods. While the methods based on Hermite polynomial expansion and the EMM performs well for high gamma, the GMM does not. This error comes from the VIS effects distorting the condition in the GMM

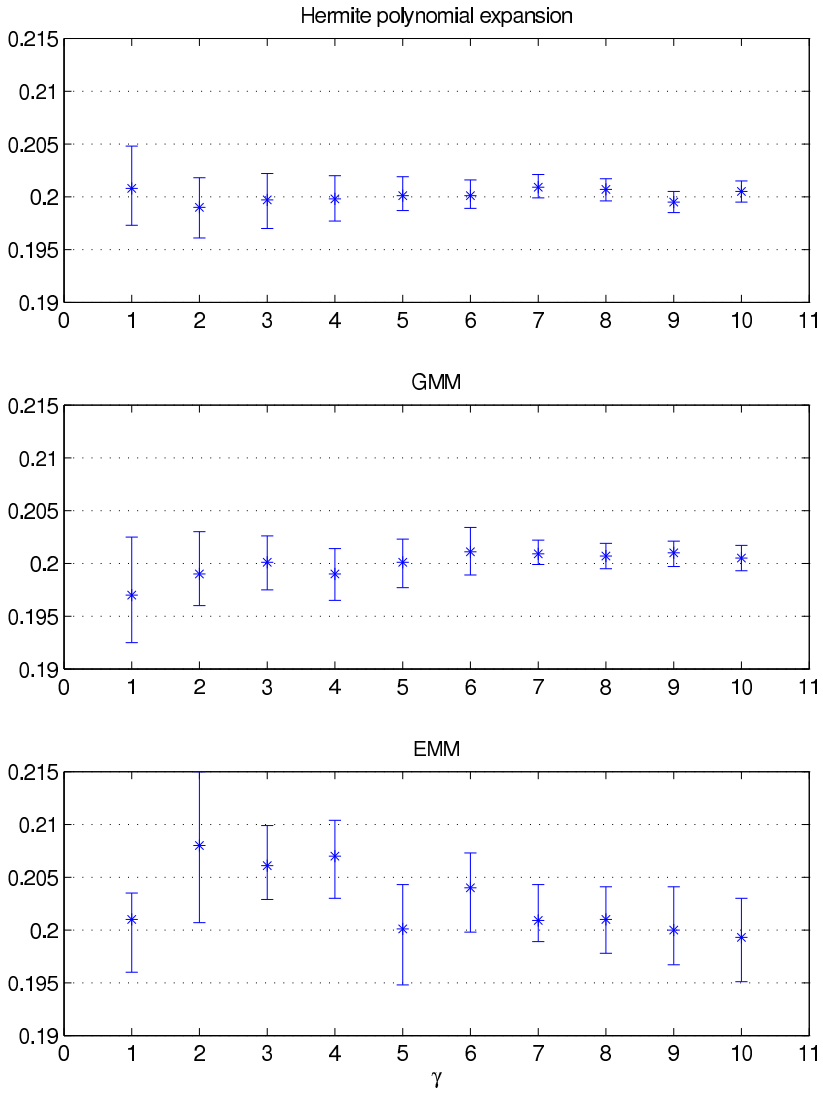

Figure 8.3: Estimation results for the μ-parameter in the CKLS model. Median values together with 25th and 75th quantiles are indicated for every Method and γ-value. Although every method performs well, the EMM gives the most uncertain estimation. For all methods the estimation seems to be more accurate for high gamma

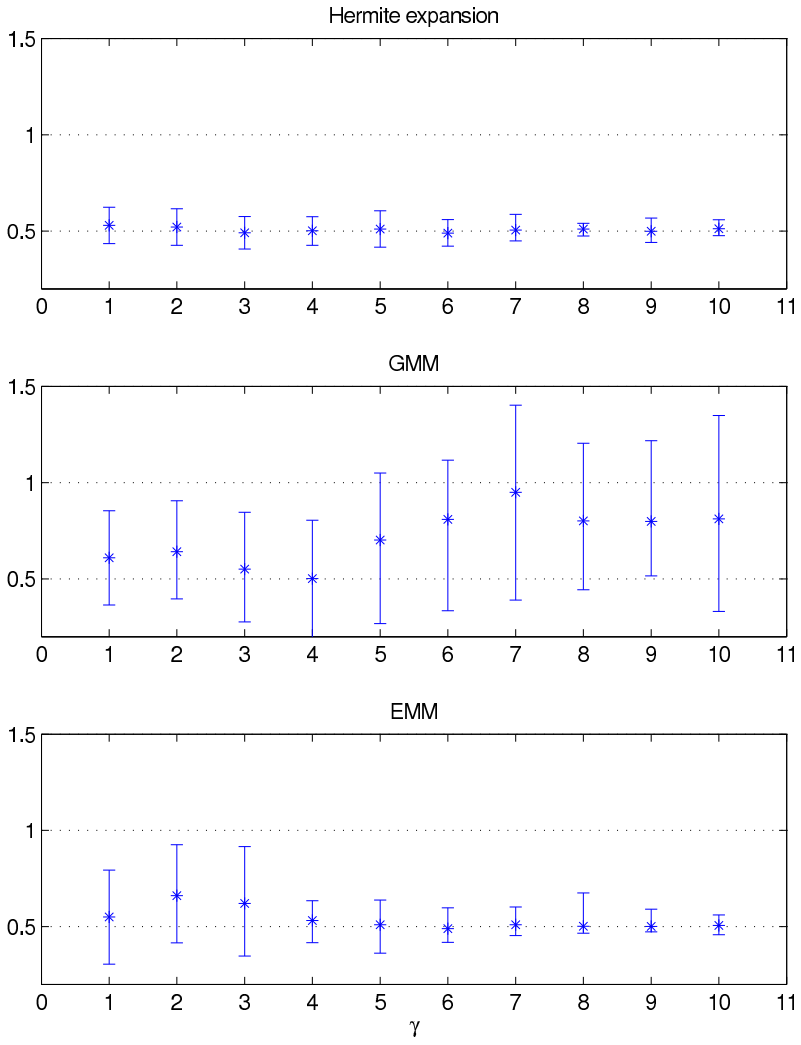

Figure 8.4: Estimation results for the κ-parameter in the CKLS model. GMM gives uncertain result which might be an effect of the VIS. Estimation with Hermite polynomial expansion performs very well

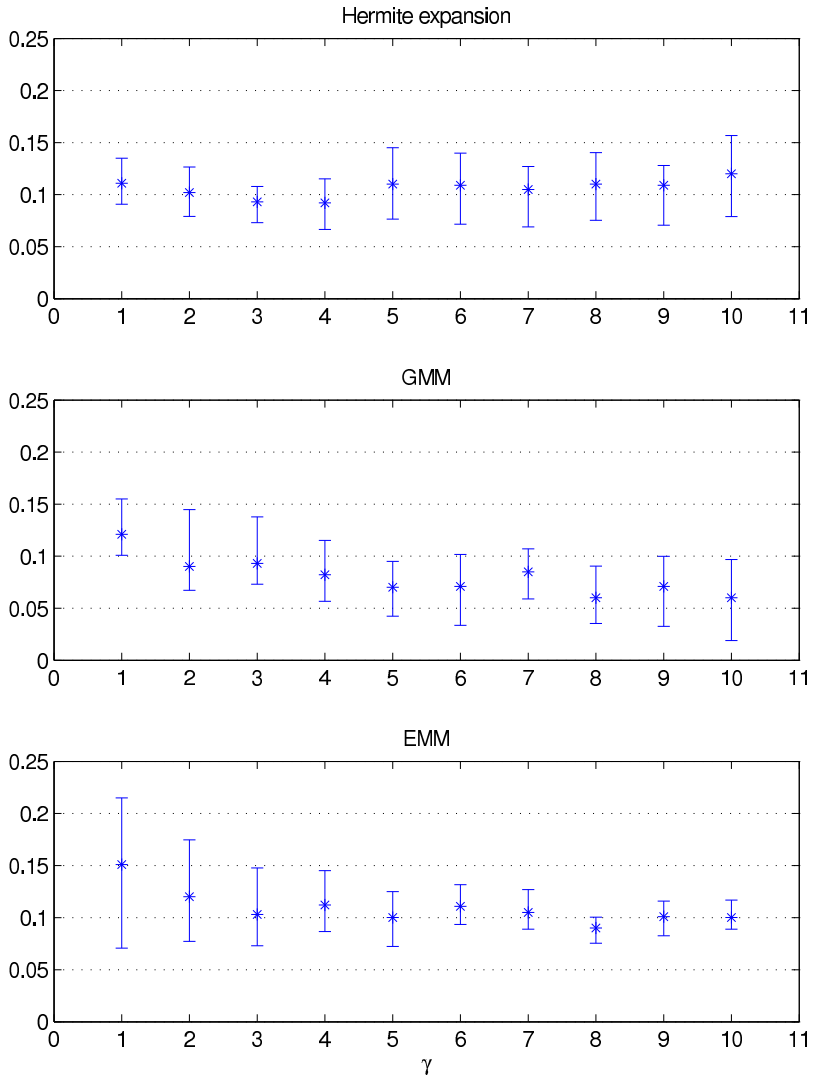

Figure 8.5: Estimation results for the σ-parameter in the CKLS model. The GMM indicate estimation bias. Effects that might be caused by VIS. For high gamma the EMM performs well

Chapter 9

Conclusion

This thesis was intended evaluate accuracy of parameter estimation methods for one dimensional CKLS processes that experience VIS. From theory we know that moments conditions are distorted in the presence of VIS which causes traditional GMM methods to fail. For the specific choice on the parameter set $\{\mu, \kappa, \sigma, \gamma\} = \{0.2, 0.5, 0.1, \gamma\}$ this failure in estimation is obvious from Figure 8.2 in case when γ is relatively large compared to other parameters. However for the absolute error to be significant the VIS must be strong. Most interestingly both the Efficient Method of Moments and Hermite polynomial expansion performs well in the VIS region. Figures 8.2-8.5 even indicate that the estimation becomes better in this case.

To get an idea of how well the different methods perform in general we compare them to the simple ML method. Since simple ML is only available in cases when we have a closed form expression for the transition density, we choose the specific CIR case of $\gamma = 1/2$ in which closed form are known. Evaluation in this case indicate good accuracy of alternative model in which the quantiles of the estimated values fell just outside the ones for the simple ML. Still, the GMM leaves us with the largest uncertainty in the estimation. That is, even in the simple cases it is an advantage to use more advanced methods for the estimation.

No emphasises has been made on computer efficiency. Although there are some ideas that must be highlighted when implementing the algorithms. There is an obvious trade off between applying computer intensive methods such as EMM and more direct methods of Hermite polynomials. EMM is a much more general method in a sense that the algorithm is the same independently of which type of process you analyze. This is in contrast to using Hermite polynomial expansion of the transition kernel where the expansion is directly related to the specific type of process. For every new type of process to be analyzed a new expansion must be performed. As the results

indicate, once the analytic expression is obtained, it outperformes the EMM method in accuracy.

The primary reason why EMM estimation is slow is that accurate simulations of VIS processes are computationally demanding in the case of VIS. Since EMM uses simulated paths of the processes when doing the Monte Carlo estimation of the moment conditions this prevents fast estimation. It is possible to reduce the accuracy in the simulation schemes but at the cost of less accurate estimation.

Finally, even though the convergence properties have been discussed, from the obtained results, we cannot tell how fast the methods will converge. The only thing that we know is how the estimator will be distributed once the number of sampled data points goes to infinity. From an application point of view the amount of data is limited and we don't know how valid the these limit assumptions are. From the estimation of the CIR process Figure 8.1 it looks like the estimators are normally distributed. Hence in the specific case considered the number of data points considered are enough to eliminate possible deviations from normality.

Bibliography

[1] Muszta. A. *Contributions to Numerical Solutions of Stochastic Differential Equations*. PhD thesis, Chalmers University of Technology, 2005.

[2] Higham D.J. Mean-square and asymptotic stability of the stochastic theta method. *SIAM J. Numerical analysis*, 38:753–769, 2000.

[3] Conley T.G. Hansen L.P. Luttmer E. and Scheinkman J.A. Short term interest rate and subordinated diffusions. *The Review o financial studies*, 10:525–577, 1997.

[4] Chan K. Koralyi G. Longstaff F and Sanders A. An empirical comparison of alternative models of the short term interest rate. *Journal of Finance*, 47:1209–1227, 1992.

[5] Klebaner F.C. *Introduction to Stochastic Calculus with Applications*. Imperial collage press, 2:nd edition, 2005.

[6] Nielsen J.N. Madsen H. and Young P.C. Parameter estimation in stochastic differential equations: An overview. *Annual Reviews in Control*, 24:83–94, 2000.

[7] Gyong I. A note on euler approximations. *Potential analysis*, 8:205–216, 1998.

[8] Karatzas I. and Shreve S.E. *Brownian Motion and Stochastic Calculus*. Graduate texts in Mathematics. Springer-Verlag, New York, 2:nd edition, 1988.

[9] Bierens. H. J. *Introduction to the Mathematical and Statistical Foundation of Econometrics*. Themes in modern econometrics. Cambridge University Press, 2004.

[10] Cox J.C. Ingersoll J.E and Ross S. A theory of the term structure of interest rate. *Econometrica*, 53:373–384, 1985.

[11] Muszta A. Richter M. Albin J.M.P and Jensen B.A. On volatility induced stationarity for stochastic differential equations. Applied Probability Trust, March 2005.

[12] Hansen L.P. Large sample properties of generalized methods of moments. *Econometrica*, 50:1029–1054, 1982.

[13] Kloeden P.E and Platen E. *Numerical Solutions to Stochastic Differential Equations*. Applications of Mathematics. Springer-Verlag, first edition, 1992.

[14] Gallant. A. R. and Tauchen G. Which moments to match? *Econometric Theory*, 12(4):657–681, 1996.

[15] Higham D.J. Mao. X. and Stuart A. Strong convergence of numerical methods for nonlinear stochastic differential equations. *SIAM Journal of numerical analysis*, 40:1041–1063, 2002.

[16] At-Sahalia. Y. Maximum likelihood estimation of discretely sampled diffusions: A closed form approximation approach. *Econometrica*, 70(1):223–262, 2002.