Parametric Inference for Stochastic Differential Equations

Angela Ciliberti

June 19th 2012

Abstract

Stochastic differential equations (SDEs) proved a fundamental mathematical tool to model dynamics subject to randomness and are nowadays a necessary instrument in e.g. financial mathematics, neuronal modelling, population growth and physiological modelling. In realistic applications SDEs parameters are unknown quantities that have to be estimated from available data. However inference for SDEs is non-trivial and a considerable amount of research effort has been devoted to such problem in the last 20 years. In this work we implement and compare several parameter estimation methods for SDEs based on (approximated) likelihood maximization using data collected at discrete times. The comparison has proved useful to select the most convenient likelihood approximation methodology for estimating the parameters of mixed-effects models based on SDEs. Such mixed-effect models are characterized by the introduction of random parameters into SDEs: this allow to model the inter-subjects variability characterising repeated-measurement experiments while simultaneously accounting for individual stochastic dynamics, thus providing a more precise estimation for population parameters. Finally a pharmacokinetic application considering real data from the time-course of theophilline concentrations when measured on several subjects is presented.

Contents

Chapter 1

Introduction

Parametric inference for stochastic differential equations are a rapidly expanding area of research. Stochastic differential equations (SDE) are a deterministic differential equations perturbed by a random disturbance that is not necessarily small. SDE models play an important role in a number of application areas, including biology, chemistry, epidemiology, mechanics, microelectronics, economics and finance. It is often convenient to model time evolution of dynamic phenomena in many fields by using a diffusion process which is characterized by a stochastic differential equation. Each SDE have different parameters, which are crucial for the characterization of dynamic phenomena considered. It is often the case that these parameters are not known accurately, while the data for the particular dynamic phenomena are available. Consequently, the estimation of the parameters of SDE from discretely - sampled data has received substantial attention. There are mainly two branches in the community of parameter estimation in stochastic differential equations (Hurn et al. (2007)): the branch adopting the Maximum Likelihood and the branch developing estimation techniques based on moment matching, which are not considered in this thesis. As for any parametric model, maximum likelihood is preferred method for estimating the parameters of the SDE. Unfortunately, exact maximum likelihood estimation is possible in a few cases, when the distribution of the discretely sampled data are known. However it possible to estimate the parameters by approximated maximum likelihood. The basic idea is construct consistent approximations of the transition densities of the diffusion and use these to evaluate

the likelihood function.

The aim of this thesis is investigate among some difference techniques of estimations, and choose one of this to estimate the parameters of a stochastic differential mixed-effect model whose transition desities are unknown. Several approaches to approximating the likelihood function have been suggest in literature. Lo (1998) proposes numerically solving the forward Kolmogorov partial differential equation, subject to the appropriate boundary conditions, to obtain the unknown transition densities of the diffusion. Ogawa (1994), Hurn and Lindsay (1999) and Nicolau (2000) apply nonparametric density estimation to simulated data from the Euler Maruyama discretizations to approximate the transition densities of the diffusion. In this work we describe and test another technique: the simulated maximum likelihood. Originally it was developed by Santa-Clara (1995) in a early version of the paper Brandt and Santa-Clara (2002) and independently by Pedersen (1995). It has since been implemented by Honoré (1997), Piazzesi (2000) and Durham (2000) to estimate a variety of continuous - time term structure models, including models with jumps and with stochastic volatility. This method has a theoretical appeal, in fact as we show, under some assumptions the approximated likelihood function converge to the exact function, but it have been computationally burdensome. To underline the limits of Pedersen (1995) and Brandt and Santa-Clara (2002) method, we study its extension of the simulated maximum likelihood method propose by Durham and Gallant (2002). The numerical study show that the Durham and Gallant (2002) proposal improve the estimation result, we made our numerical study using the Cox, Ingersoll and Ross model (CIR) of which the exact transition density is known. We focus on another technique to approximate the likelihood function the Hermite expansion propose by Aït-Sahalia (1999). The advantage of analytical expansions is that they are computationally less demanding then simulation. The disadvantage is that, for the expansions to converge, the diffusion must first be transformed to be sufficiently Gaussian. However our numerical study, made using the Vasicek model, shown that the Hermite expansion works better than the simulation likelihood estimation proposed by Brandt and

Santa-Clara (2002). Consequently we decide to apply the hermite expansion to approximate the probability density for a stochastic differential mixed-effect model.

Stochastic differential mixed-effect models are SDEs system in which one or more parameters are random variables. These models are useful in biomedical research, particularly on studies in which repeated measurements are taken on a series of individual or experimental animals. In this models it is assumed that all responses follow a similar function but the parameters vary among individuals. As such they are able to model the variation within-group and between-group. Pharmacokinetic and pharmacodynamic studies include random effects models, see Donnet and Samson (2008). We consider a SDE that mimic the theophyllin drug pharmacokinetic, in which we consider a parameter with a normal distribution. Our aim is estimate the parameters involved in the model, the mean and variance of the random parameter. We make the study using real data and simulated data.

The work is composed by three chapter. In the firs we introduce some important notions about stochastic calculus, the definition of the stochastic differential equations and other important knowledge which are recalled in the other chapters. In the second one we describe and compare the estimation techniques: we consider first the Brandt and Santa-Clara (2002) and Durham and Gallant (2002) methods and then we introduce the Aït-Sahalia (1999) Hermite expansion. The third chapter is focused on the stochastic differential mixed-effect models. In the appendix the MATLAB programs that we coded for our numerical study are reported.

Chapter 2

Stochastic Calculus Preliminaries

The development of stochastic differential equations theory is strongly connected with stochastic calculus. One stochastic process in particular, the Brownian motion, has been fundamental for the development of this field. The process takes the name from Robert Brown, a Scottish scientist that described as random the motion of pollen particles suspended in a liquid (1927). It seemed natural to use it as the noise component of a continuous time process in general and for a stochastic differential equation in particular. The work of the Japanese mathematician Kiyoshi Itô, was fundamental for the definition of a stochastic integral and a formula that can be used to solve some types of equations. This class of mathematically solvable Stochastic differential equations (SDEs) happens to be very narrow and often indirect or approximate techniques are needed. Researching for more accurate approximation and estimation methods has been one of the most interesting topic in the field as well as applications in different disciplines. In this chapter we introduce concepts and definition that will be used throughout the work.

2.1 Basic Definitions

In this section we introduce some basic concepts of stochastic process, Markov process and Brownian motion.

Stochastic processes are sequences of random variables generated by probabilis-

tic laws. The word "stochastic" comes from the Greek and means "random". Stochastic process is a family of random variables $\{X_t\}$ where t denotes a parameter running over a suitable index set $\mathfrak{T}$. The parameter t usually represents time, but different situations may be, for example, a distance from the origin in plane, in which case X_t may represent the number of points randomly scattering in the plane whose distances from the origin are less than t. However, in this text we refer to the parameter t as the time and call $\{X_t\}$ a $discrete-time$ process, if the index set is $\mathfrak{T}=\mathbb{Z}_+$; and a $continuous-time$ process if the index set is $\mathfrak{T}=\mathbb{R}_+$. See Capasso and Bakstein (2005)

Definition 2.1. Let $(\Omega,\mathcal{F},\mathrm{P})$ be a probability space on which a stochastic process $\{X_t\}$ is defined. For each $\omega\subset\Omega$, the function $X_t(\omega)$ with respect to t, denoted by $\{X_t(\omega),t\in\mathfrak{T}\}$, is called a sample path or realization of the process $\{X_t\}$.

Definition 2.2. The stochastic process $\{X_t\}$ on $(\Omega,\mathcal{F},\mathrm{P})$, is called a process with independent increments if for all $n\in\mathbb{N}$ an for all $t_1,t_2,...,t_n\in\mathbb{R}_+$, where $t_1<t_2<...<t_n$, the random variables $X_{t_1},X_{t_2}-X_{t_1},...,X_{t_n}-X_{t_{n-1}}$ are independent.

Definition 2.3. The realization of the process $\{X_t\}$ up to the time t is $\{X_s(\omega),s\leq t\}$.
A stochastic process is strictly stationary if it is invariant under time displacement.
We call a Gaussian process a stochastic process for which any joint distribution is Gaussian.

A Markov process is a stochastic process that is distinguished by the Markov property. Markov processes have many applications in operations research, biology, engineering, and economics.

If t is the present time, any time such that $s<t$ is called a *paste* time, while any time such that $s>t$ is a $future$ time. The following definitions are taken from Kijima (1997)

Definition 2.4. Let $\{X_t\}_{t\in\mathbb{R}}$ be a stochastic process on a probability space, valued in a measurable space $(\mathrm{E},\mathcal{B})$ and adapted to the increasing family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of σ-algebras of subsets of $\mathcal{F}$. $\{X_t\}$ is a Markov process with respect $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if the following condition is satisfied:

$$\forall B\in\mathcal{B},\forall(s,t)\in\mathbb{R}_+\times\mathbb{R}_+,s<t:$$

$$P(X_t \in \mathcal{B} \mid \mathcal{F}_s) = P(X_t \in \mathcal{B} \mid X_s).$$

Shortly we can define a Markov process as a stochastic process whose future behavior can be determined independently of the past. From the definition follow the properties.

Preposition 2.1.1. *Under the assumptions of the previous definition, the following two statements are equivalent:*

1. *for all $B \in \mathcal{B}$ and all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t$:*

$$P(X_t \in B \mid \mathcal{F}_s) = P(X_t \in B \mid X_s)$$

   almost surely;

2. *for all $g : E \to \mathbb{R}, and \mathcal{B}\ \mathcal{B}_{\mathbb{R}}$- measurable such that $g(X_t) \in L^1(\mathrm{P})$ for all t, for all $(s,t) \in \mathbb{R}_+^2, s < t$:*

$$\mathbb{E}\left[g(X_t) \mid \mathcal{F}_s\right] = \mathbb{E}\left[g(X_t) \mid X_s\right]$$

   almost surely.

Theorem 2.1.2. *Every real stochastic process $\{X_t\}_{t \in \mathbb{R}_+}$ with independent increments is a Markov process.*

The Markov property enables us to develop a rich system of concepts and theorems and to derive many results that are useful in applications.
Let $\mathfrak{T} = [0, +\infty)$ be the index set and consider a stochastic process $\{X_t, t \in \mathfrak{T}\}$ taking values on $\mathcal{N} = \{0, 1, 2, ...\}$. We say that the process $\{X(t)\}$ is a Markov chain if for each $t > 0$ and each set A,

$$P(X(t+s) \in A \mid X(u), 0 < u < s) = P(X(t+s) \in A \mid X(s)).$$

More precisely for each $s \geq 0$, $t > 0$, each $i, j \in \mathcal{N}$, and every history $x(u)$, $0 \leq u < s$,

$$P(X(t+s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s) =$$

$$= P(X(t+s) = j \mid X(s) = i),$$

then this process $\{X_t\}$ is called a Markov chain in continuous time. In other word, a continuous time Markov chain is a stochastic process having the Markov property.

Definition 2.5. Let $\{X_t\}$ be a Markov process and define

$$p_{ij}(n, n+m) = P(X_{n+m} = j \mid X_n = i),$$

with $n = 0, 1...$ and $m = 1, 2, ...$
The conditional probability $p_{ij}(n, n+1)$ is called the transition probability from state i to state j at time n.

Definition 2.6. A Markov process is homogeneous if all the transition probability depend only on time difference.

Now we can define the Brownian motion. The botanist Robert Brown in (1827) observed that a small particle suspended in a liquid is subject to infinitely collisions with atoms, therefore it was impossible to observe its exact trajectory. With the help of microscope it was only possible to confirm that the movement of the particle is entirely chaotic. This type of movement is called Brownian motion. Brown tried different materials and different solvents, and still the motion of these particles continued. This was a time when most scientists did not believe in atoms or molecules, so the underlying mechanism responsible remained a mystery for nearly a century. In the words of S. G. Brush *"three quarters of a century of experiments produced almost no useful results in the understanding of Brownian motion because no theorist had told the experimentalists what to measure"*. Its mathematical inventor Einstein already observed, it is necessary to make approximations, in order to describe the process. The first works on Brownian motion appeared in a paper by Einstein (1905) and on Bachelier 's thesis (1900). After the Brownian motion was rigorously formalized by Wiener (1923). Next definitions are taken from Øksendal (2005).

Definition 2.7. The real - valued process $\{W_t\}_{t \in \mathbb{R}_+}$ is a Brownian motion (or Wiener process) if it satisfies the following condition:

1. $W_0 = 0$ almost surely;
2. $\{W_t\}_{t \in \mathbb{R}_+}$ is a process with independent increments;
3. $W_t - W_s$ is normally distributed with $N(0, t-s), (0 \leq s < t)$.

Since the law of its increments is Gaussian, the Brownian motion is an example of Gaussian process.

Theorem 2.1.3. *Every Brownian motion* $\{W_t\}_{t\in\mathbb{R}_+}$ *is a Markov process.*

2.2 The Itô Calculus and Differential Stochastic Equations

In this section we introduce the Itô formula. In the same way that Lebesgue developed the ideas of set theory to provide a more general definition of Riemann's integral, Itô extended the ideas of Lebesgue to include integration with the Brownian motion, see van Handel (2007).

Then we will to present the stochastic differential equations. The notion of stochastic differential equation (SDE), defined as a deterministic differential equation perturbed by random disturbances that are not necessarily small, has been used profitably in a variety of disciplines(Jeisman Lindsay, 2007). SDEs are central to much of modern finance theory and have been widely used to model the behavior of key variables such as the instantaneous short-term interest rate, asset prices, asset returns and their volatility, see Sundaresan (2000). The SDE are a natural way to model population growth in a randomly variety environment (Population growth in random environments; Braumann (1983)). They are also used in neuronal modeling (Stochastic methods in neuroscience; Laing and Lord (2010)); computational systems biology (Stochastic modeling for systems biology; Wilkinson (2012)); and physiological models (Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations; Picchini et al. (2006)). They try to explain the oscillations of glycemia occurring in response to the hyperinsulinization and to the continuous glucose infusion at varying speeds, using a system of stochastic differential equations. Next definitions and lemmas are taken from Øksendal (2005)

Definition 2.8. Let $(\Omega, \mathcal{F}, \mathcal{P})$ a probability space, $\{W_t\}$ a Brownian motion, and $\{X_t\}$ a $\mathcal{F}_t$ adapted stochastic process with

$$P\left(\int_0^T X_t^2 dt < \infty\right) = 1$$

for all $T < \infty$. Then the Itô integral:

$$I_t(X_t) = \int_0^t X_s dW_s$$

is uniquely define.

Perhaps the most important topic in stochastic integration is the associated calculus, which gives us transparent tools to manipulate Itô integrals and stochastic differential equations (SDE).

Definition 2.9. The Itô SDE for a diffusion process X_t is:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t.$$

The process is determined by the deterministic scalar functions $\mu(\cdot)$ and $\sigma(\cdot)$, and the initial condition $X_0 = x_0$. In particular $\mu(X_t)$ is the *infinitesimal mean* of the Markovian process, defined as

$$\mu(x_t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbb{E}(X_{t+\delta t} - x_t);$$

$\sigma^2(X_t)$ is the *infinitesimal variance* of the process, defined as

$$\sigma^2(x_t) = \lim_{\delta t \to 0} \frac{1}{\delta t} Var(X_{t+\delta t} - x_t).$$

Note that in all of the above equality we are implicity conditioning throughout on $X_t = x_t$.

Lemma 2.2.1 (Itô's Lemma)**.** *Let $dX_t = \mu(t,\omega)dt + \sigma(t,\omega)dW_t$, be the SDE associated to the $n-$dimensional Itô process, where μ and σ are random functions with values respectively in $\mathbb{R}^n$ and $\mathbb{R}^{n\times p}$. Let*

$$f(x,t) : [0,\infty] \times \mathbb{R}^n \to \mathbb{R}^p,$$

a $\mathcal{C}^2$ function. Then the transformation process $Y_t = f(X_t, t)$ is also called a $n-$dimensional Itô process that, indicating with superscripts the component of the vectors, can be expressed as:

$$dY_t^k = \frac{\partial f^k}{\partial t}(X,t)dt + \sum_{i=1}^{n} \frac{\partial f^k}{\partial X^i}(X,t)dX^i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f^k}{\partial X^i \partial X^j}(X,t)dX^i dX^j \quad (2.1)$$

where $dW_t^i dW_t^i = \delta_{ij}dt$ and $dW_t^i dt = 0$.

One of the fundamental theorems of stochastic analysis is Girsanov's theorem, which tell us what happens to the Brownian motion under a change of measure.

Theorem 2.2.2. *Let $\{W_t\}$ be a $n-$dimensional $\mathcal{F}_t-$ Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and let $\{X_t\}$ be an Itô process of the form*

$$X_t = \int_0^t \mathcal{F}_s ds + W_t,$$

$t \in [0,T]$. Suppose furthermore that $\{\mathcal{F}_t\}$ is Itô integrable, and define:

$$\Lambda = exp\left(-\int_0^T (\mathcal{F}_s)^* dW_s - \frac{1}{2}\int_0^T \|\mathcal{F}_s\|^2 ds\right),$$

where $(\mathcal{F}_s)^* dW_s = \mathcal{F}_s^1 dW_s^1 + ... + \mathcal{F}_s^n dW_s^n$. If Novikov's condition

$$\mathbb{E}_P\left[exp\left(\frac{1}{2}\int_0^T \|\mathcal{F}_s\|^2 ds\right)\right] < \infty$$

is satisfied, then $\{X_t\}_{t\in[0,T]}$ is an $\mathcal{F}_t-$ Brownian motion under $Q(A) = \mathbb{E}_P(\Lambda I_A)$.

Now we can introduce the stochastic differential equations.

Let $(\Omega, \mathcal{F}, \mathcal{P})$, and let $\{X_t\}$ be a probability space, and $\{\mathcal{F}_t, t \geq 0\}$ a non decreasing family of $\sigma-$algebras in $\mathcal{F}$. Let us define a $n-$dimensional continuous and homogeneous in time Itô process which satisfies the following system of n differential equations governed by the $p-$dimensional Brownian motion W_t:

$$dX_t = \mu(X_t, t; \theta)dt + \Sigma(X_t, t; \theta)dW_t, \tag{2.2}$$

where $\mu(X_t, t; \theta) : (\mathbb{R}^n \times [0,T] \times \mathbb{R}^q) \to \mathbb{R}^n$ is the drift function and $\Sigma(X_t, t; \theta) : (\mathbb{R}^n \times [0,T] \times \mathbb{R}^q) \to \mathbb{R}^{n\times p}$ the diffusion function, both depending on an unknown parameter vector $\theta \in \Theta \subseteq \mathbb{R}^q$.

Theorem 2.2.3. *Suppose that*

1. *$X_0 \in \mathcal{L}^2$;*
2. *μ and Σ are Lipschitz and continuous uniformly on $[0,T]$;*
3. *$\|\mu(0,t)\|$ and $\|\Sigma(0,t)\|$ are bounded on $t \in [0,T]$.*

Then there exist a unique solution, $\{X_t\}$, $P-$almost everywhere to the associate stochastic differential equation, and moreover for its solution $\mu(X_t, t)$ and $\Sigma(X_t, t)$ are in $\mathcal{L}^2$.

Theorem 2.2.4. *The unique solution,$\{X_t\}$, of a stochastic differential equations is an $\mathcal{F}_t$ Markov process.*

It is helpful to introduce a transition density.

Definition 2.10. Let $t, t' \in \mathfrak{T}$ we define the transition density as

$$p(x, t, x', t') = \frac{\partial}{\partial x'} P(x, t, x', t'),$$

where $P(x, t, x', t') = P(X_{t+t'} \leq x' \mid X_t = x)$.

Now we can re-write the properties of a diffusion process in Definition 2.9 as integrals with respect to the transition density as

$$\mu(x) = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}} (x' - x) p(x, t, x', \delta t) dx';$$

$$\sigma^2(x) = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{\mathbb{R}} (x' - x)^2 p(x, t, x', \delta t) dx'.$$

These properties turn out to be useful for mathematical analysis to write down the Chapman-Kolmogorov equation for a diffusion process as

$$p(x, t, x', t' + t) = \int_{\mathbb{R}} p(z, t, x', t') p(x, t, z, t) dz.$$

Just as for the case of discrete state Markov chains in continuous time, we can use the Chapman-Kolmogorov equation in order to derive differential equations representing the Kolmogorov backward and forward equations for diffusion process.

Theorem 2.2.5. *Let $p(x, x_e, t)$ be the transition density which represents the density of an endpoint state, x_e, as function of an initial state x and the time prior to the endpoint, t. The Kolmogorov backward equation is:*

$$\frac{\partial}{\partial t} p(x, x_e, t) = \mu(x) \frac{\partial}{\partial x} p(x, x_e, t) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} p(x, x_e, t).$$

The backward equation can be useful in applications, but is slightly less useful than the forward equation, and also less intuitive.

Theorem 2.2.6. *The Kolmogorov forward equation for the transition density $p(x_0, x, t)$ of a univariate diffusion process governed by an Itô SDE defined in 2.9 is*

$$\frac{\partial}{\partial t} p(x_0, x, t) = -\frac{\partial}{\partial x} (\mu(x) p(x_0, x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\sigma^2(x) p(x_0, x, t)).$$

this equation is commonly referred to as the Fokker - Planck equation.

The Fokker - Planck has many important applications in mathematical analysis of diffusion process, unfortunately it is analytically intractable except in a few simple special cases.

Chapter 3

Likelihood Based Inference for SDE

As consequence of the importance of the stochastic differential equations, the estimation of the parameters of SDEs from discretely-sampled data has received substantial attention in financial econometrics literature, particularly in the last twenty years. Since a large number of competing estimation procedures have been proposed, Hurn et al. (2007) propose an evaluation of the various estimation techniques. The estimation procedures could be divided in two branches.

- Likelihood - based procedure to solve Fokker - Planck equation (Jensen and Pulsen (2002)), discrete maximum likelihood (Elerian (1998)), hermite polynomial expansion (Aït-Sahalia (2002b)), simulated maximum likelihood (Pedersen (1995), Brandt and Santa-Clara (2002)) and Markov chain Monte Carlo (Elerian et al. (2001)).

- A procedures obtained by aligning user-defined features of the model with those of the data, as general method of moments (Hansen (1982)), indirect estimation (Gallant and Tauchen (1996)), characteristic function (Singleton (2001)), estimating functions (Sørensen (2000)) and match to marginal density (Aït-Sahalia (1996a)).

In this chapter we describe some techniques to approximate the likelihood function. We first present the method proposed by Brandt and Santa-Clara (2002) and Pedersen (1995), and we compare it with the Hermite approximation described in Aït-Sahalia (1999), using the Vasicek's model. We try to improve the Brandt - Santa Clara's method using the variance reduction techniques propose by Durham and Gallant (2002). Some numerical results are also obtained using the Cox- Ingersoll - Ross model.

3.1 Monte Carlo Approximations

The continuous-time models has proved to be an immensely useful tool in finance and more generally in economics. Continuous-time models are widely used to study issues that include the decision to optimally consume, save, and invest, portfolio choice under a variety of constraints, contingent claim pricing, capital accumulation, resource extraction, game theory, and more recently contract theory. Many refinements and extensions are possible, but the basic dynamic model for the variable of interest X_t is a stochastic differential (2.2) where W_t is a standard Brownian motion and the drift μ and diffusion Σ are known functions except for an unknown parameter vector θ in a bounded set $\Theta \subset \mathbb{R}^d$.

One major impediment to both theoretical modeling and empirical work with continuous-time models of this type is the fact that in most cases little can be said about the implications of the dynamics in (2.2) for long time intervals. Though (2.2) fully describes the evolution of the variable X over each infinitesimal instant, one cannot in general characterize in closed form an object as simple as the conditional density of $X_{t+\Delta}$ given the current value X_t.
As for any parametric model, maximum likelihood is the preferred method for estimating the parameters of a diffusion. Maximum likelihood estimates of the parameters of stochastic differential equations are consistent and asymptotically efficient. Unfortunately, exact maximum likelihood estimation is only possible in a few special cases when the distribution of the discretely sampled data is known. In particular, the distribution is known explicitly for diffusions with lin-

ear mean and constant or proportional variance; in most cases, however, exact maximum likelihood estimation is impossible because the likelihood function of the model cannot be evaluated explicitly, and the alternative of approximating it has until recently proven difficult. Simulation of maximum likelihood (SML) method works as follows: first we construct consistent approximations to the transition densities of the diffusion and we use these approximations to evaluate the likelihood function. Then we maximize this approximated likelihood function. Since the approximations to the transition densities are consistent, the same is the approximation to the likelihood function. This implies that asymptotically the SML estimator behaves just like the unattainable exact maximum likelihood estimator. There are some different ways to approximate the transition probability, in the next paragraphs we present some of this ways.

3.1.1 Brandt - Santa Clara

We consider a continuous - time process $\{X_t\}$ described by the following system of stochastic differential equations:

$$dX_t = \mu(X_t, t; \theta)dt + \Sigma(X_t, t; \theta)dW_t, \tag{3.1}$$

where W_t denote a r-dimension vector of independent Brownian motions, defined in a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$; $\theta \in \Theta \subseteq \mathbb{R}^d$ is an unknown parameter; $\mu(\cdot, \cdot; \theta) : \mathbb{R}^k \times [0, \infty) \to \mathbb{R}^k$; and $\Sigma(\cdot, \cdot; \theta) : \mathbb{R} \times [0, \infty) \to M^{k \times r}$.
First of all we assume that the drift μ and the diffusion Σ are infinitely differentiable with continuous and bounded derivatives of all order. This assumption is stronger than the usual linear growth and uniform Lipschitz continuity conditions that are sufficient to guarantee the existence of a unique strong solution to the stochastic differential equations, see Theorem 2.2.3. The extreme degree of smoothness is sufficient, but most likely not necessary, to bound the asymptotic error of the approximations. We suppose, also, that $\theta \in \Theta \subset \mathbb{R}^d$, where Θ is a compact set that contain the true θ_0; and the covariance matrix $\Sigma\Sigma^T$ is positive defined. For practical reason, the continuous - time process is sampled only at

$N+1$ equally spaced points in time, denoted $t_0, t_1, ..., t_N$.

Let $p(X_{t_0}, X_{t_1}, ..., X_{t_N}; \theta)$ the density of the discrete-time data, generated by the continuous - time diffusion model. As a function of the parameters θ, this density represents the likelihood function:

$$\mathcal{L}(\theta) = p(X_{t_0}, X_{t_1}, ..., X_{t_N}; \theta) = p(X_{t_0}, t_0; \theta) \prod_{n=0}^{N-1} p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta). \tag{3.2}$$

The equality follows from the fact that $\{X_t\}$ is Markovian. It shows that, in order to evaluate the likelihood function, we require the initial unconditional density $p(X_{t_0}, t_0; \theta)$ and the N transition densities $p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$, for $n = 0, 1, ..., N-1$. The parameter vector that maximizes the log likelihood function $\mathcal{L}$ is the maximum likelihood estimator θ_{ML} of θ_0. We have to make the follow assumption. (Brandt and Santa-Clara (2002)) to guarantee the usual desirable asymptotic properties,for example consistency, asymptotical efficiently and asymptotical normality:

1. The likelihood function $\mathcal{L}$ is twice continuously differentiable in θ in a neighborhood of the true parameter vector θ_0.
   Furthermore, $\mathbb{E}\left[\left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta}\right]\left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta'}\right]\right]$ has full rank and is bounded for all parameters $\theta \in \Theta$.

2. For every vector $\lambda \in \mathbb{R}^k$, $\lambda' I(\theta) \lambda \to \infty$, where

$$I(\theta) = \tag{3.3}$$
$$\mathbb{E}\left[\sum_{n=0}^{N-1} \frac{\partial}{\partial \theta} \ln p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) \frac{\partial}{\partial \theta'} \ln p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)\right]$$

This assumption is required to establish that the maximum likelihood estimator θ_{ML} is consistent. For it to hold, it is sufficient that the gradients of the transition densities are bounded. The matrix I is called Fisher information matrix. The inverse of Fisher information matrix gives the Cramér-Rao lower bound on the covariance matrix of any consistent and unbased estimator of the parameter vector. The maximum likelihood estimator typically attains this lower bound. Now we construct an estimator based on a sequence of consistent approximations to the likelihood function, of (3.2). We first discretize

the process X_t between times t_n and t_{n+1} to construct a consistent approximation of $p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$ for two adjacent discrete time observations X_{t_n} and $X_{t_{n+1}}$. There exists an infinite number of discrete-time processes that approximate the diffusion process in this interval. We choose the Euler Maruyama scheme because it is computationally convenient. We divide the interval $[t_n, t_{n+1}]$ into M subintervals of length $h = \frac{t_{n+1}-t_n}{M}$. The Euler Maruyama discretization $\hat{X}_{t_n+mh}$ for $m = 0, 1, ..., M-1$, is the Gaussian process:

$$\hat{X}_{t_n+(m+1)h} = \tag{3.4}$$

$$\hat{X}_{t_n+mh} + \mu(\hat{X}_{t_n+mh}, t_n + mh; \theta)h + \sigma(\hat{X}_{t_n+mh}, t_n + mh; \theta)\sqrt{h}\varepsilon_{t_n+(m+1)h}$$

where ε_{t_n} has a standard normal distribution. The recursion starts at the initial condition $\hat{X}_{t_n} \equiv X_{t_n}$. With all this assumptions, the Euler Maruyama approximation converges weakly to the stochastic process X_t as $M \to \infty$. By definition (3.4), the one-step-ahead transition densities of the Euler Maruyama discretization are Gaussian. This means that the probability of $\hat{X}_{t_n+(m+1)h} = y$, conditional on $\hat{X}_{t_n+mh} = x$, is

$$q_M(y, t_n + (m+1)h \mid x, t_n + mh; \theta) =, \tag{3.5}$$

$$\phi(y; x + \mu(x, t_n + mh; \theta)h, V(x, t_n + mh; \theta)h)$$

where $\phi(y, mean, variance)$ denote a multivariate normal density; and $V = \Sigma\Sigma^T$. The density q_M is an approximation of $p(y, t_n + (m+1)h \mid x, t_n + mh; \theta)$. The accuracy of this approximation depends on how much time h elapses between the points x and y. In the limit, as $h \to 0$, the approximation is exact. The multi-step-ahead transition densities of the Euler Maruyama discretization are unknown in closed form. However, they can be evaluated through recursive integration. In particular, the probability that $\hat{X}_{t_n+(m+j)h} = y$, conditional on $\hat{X}_{t_n+mh} = x$, for $j = 2, 3, ..M-m$, is :

$$q_M(y, t_n + (m+j)h \mid x, t_n + mh; \theta) =$$

$$= \int_{\mathbb{R}} q_M(y, t_n + (m+j)h \mid z, t_n + (m+j-1)h; \theta) \times$$

$$\times q_M(z, t_n + (m+j-1)h \mid x, t_n + mh; \theta)dz$$

From (3.5), the first term in the integrand is a Gaussian density and is therefore known in closed form. The second term is itself a multi-step-ahead transition density that can be computed again recursively. With $y = X_{t_{n+1}}$, $x = X_{t_n}$, $j = M - m$, and the previous equations then yield an approximation of the continuous-time transition density $p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$. For the Euler Maruyama discetization, the probability density function of $\hat{X}_{t_{n+1}} = X_{t_{n+1}}$, conditional on $\hat{X}_{t_n} = X_{t_n}$, is

$$q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) = \tag{3.6}$$
$$= \int_{\mathbb{R}} \phi(X_{n+1}; z + \mu(z, t_n + (M-1)h; \theta)h, V(z, t_n + (M-1)h; \theta)h \times$$
$$\times q_M(z, t_n + (M-1)h \mid X_n, t_n)dz.$$

The approximate transition density $q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$ is still a convolution of M Gaussian densities that involves solving $M-1$ integrals:

$$q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) =$$
$$= \int_R \phi(X_{t_{n+1}}; z_{M-1} + \mu(z_{M-1}, t_n + (M-1)h; \theta)h; \sigma(z_{M-1}, t_n + (M-1)h; \theta)^2 h) \times$$

$$\times q_M(z_{M-1}, t_n + (M-1)h \mid X_{t_n}, t_n; \theta)dz_{M-1} =$$
$$= \int_{\mathbb{R}} \phi(X_{t_{n+1}}; z_{M-1} + \mu(z_{M-1}, t_n + (M-1)h; \theta)h; \sigma(z_{M-1}, t_n + (M-1)h; \theta)^2 h) \times$$
$$\times \int_{\mathbb{R}} \phi(z_{M-1}, t_n + (M-2)h + \mu(z_{M-2}, t_n + (M-2)h; \theta), \sigma(z_{M-2}, t_n + (M-2)h; \theta)^2) \times$$
$$q_M(z_{M-2}, t_n + (M-2)h \mid X_{t_n}, t_n; \theta)dz_{M-2}dz_{M-1};$$

and recursively, we obtain:

$$q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) =$$
$$= \int_{\mathbb{R}^{M-1}} \prod_{m=0}^{M-1} \phi(z_{m+1}, z_m + \mu(z_m, s + mh; \theta), \sigma(z_m, s + mh; \theta)^2)d\lambda(z_1 ..., z_{M-1}) \tag{3.7}$$

where $z_0 = X_{t_n}$, $z_M = X_{t_{n+1}}$ and λ denotes the Lebesgue measure. In general, these integrals cannot be computed analytically and quadrature-based

numerical integration techniques quickly become computationally infeasible as M increases. This means that the Euler Maruyama discretization by itself is not sufficient to facilitate maximum likelihood estimation. The innovation of the SML method is to interpret the integral in (3.6) as an expectation of the function ϕ of the random variable z: the distribution of this variable z is $f(z) = q_M(z, t_n + (M-1)h \mid X_{t_n}, t_n)$. Although we cannot easily evaluate the expectation, we can use the Euler Maruyama discretization to generate a large number of independent random variables z_s; for $s = 1, 2, ..., S$ from the distribution $f(z)$. Then, we approximate the expectation, and ultimately the corresponding continuous -time transition density p with a sample average of the function ϕ evaluated at these random draws of z. In more detail, the method works as follows. Starting at time t_n with $\hat{X}_{t_n} = X_{t_n}$, we iterate on the Euler Maruyama recursion (3.4) exactly $M-1$ times. This results in a single draw $z_s = \hat{X}_{t_n+(M-1)h}$ of the discrete-time process at time $t_n + (M-1)h$ from the distribution $f(z)$. We repeat this procedure S times. Finally, we average the function ϕ over this random sample of z to approximate the expectation. See Figure 3.1. The five lines represent incomplete ten-step discretizations of this diffusion, which connect X_0 and X_1. Each discretization is generated by starting the Euler Maruyama recursion. Formally, our approximation to the transition

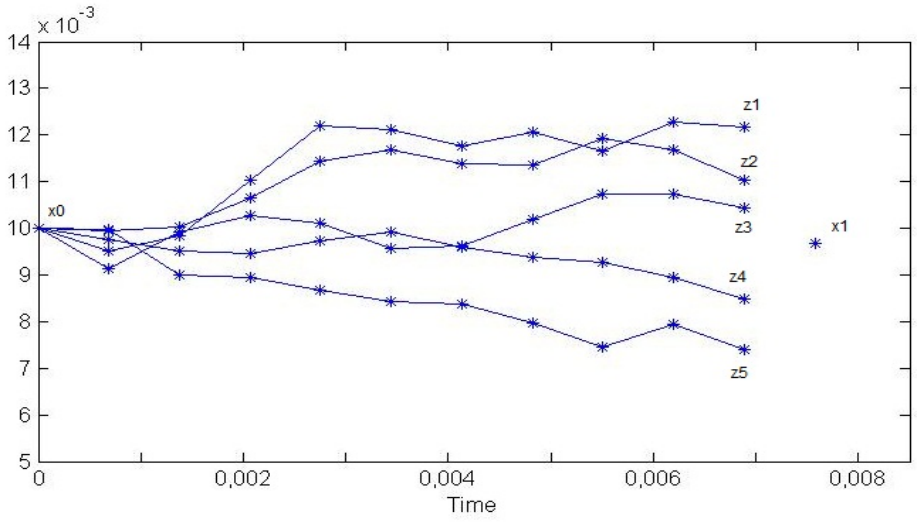


Figure 3.1: Approximating of tansition densities

density q_M of the Euler Maruyama discretization is:

$$\hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}; \theta) = \tag{3.8}$$

$$= \frac{1}{S} \sum_{s=1}^{S} \phi(X_{t_{n+1}}; z_s + \mu(z_s, t_n + (M-1)h\theta)h, V(z_s, t_n + (M-1)h\theta)h),$$

where the z_s, for $s = 1, 2, ..., S$, represent independent realizations of an M-step Euler Maruyama discretization after $M-1$ iterations, $\hat{X}_{t_n+(M-1)h}$. Each discretization starts at $\hat{X}_{t_n} = X_{t_n}$. The Strong Law of Large Numbers guarantees that the approximation $q_{M,S}$ converges to the transition density q_M of the Euler Maruyama discretization as $S \to \infty$. Since the transition density of the Euler Maruyama discretization converges to the transition density p of the continuous-time process $M \to \infty$, the approximation $\hat{q}_{M,S}$ also converges to the transition density of the continuous-time process as $S \to \infty$ and $M \to \infty$.

Lemma 3.1.1. *If μ and Σ are differentiable and $\Sigma\Sigma^T$ is positive define, as $M \to \infty$,*

$$q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) = O\left(\frac{1}{M}\right)$$

The Lemma 3.1.1 shows that as the accuracy of the Euler Maruyama discretization increases, or formally as $M \to \infty$ and thereby $h \to 0$, the transition density of the Euler Maruyama discretization converges to the corresponding transition density of the continuous-time process.

Lemma 3.1.2. *Under the hypothesis of Lemma 3.1.1, as $M \to \infty$ and $S \to \infty$,*

$$\hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) \to p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta),$$

almost surely.

Proof. Recall the equation (3.8) where we write $\mu(z_s) = \mu(z_s, t_n + (M-1)h; \theta)$ and $V(z_s) = V(z_s, +t_n + (M+1)h; \theta)$. The elements of the sum are i.i.d. with finite expectation :

$$\mathbb{E}\left[\phi(X_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h)\right] = q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta).$$

Hence, the Strong Law of Large Numbers applies, and as $S \to \infty$,

$$\hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) = q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$$

almost surely. Applying the previous Lemma we get thesis. □

Lemma 3.1.3. *Under the same hypothesis of Lemma 3.1.1, as $M \to \infty$ and $S \to \infty$, with $\frac{S^{\frac{1}{2}}}{M} \to 0$,*

$$S^{\frac{1}{2}} \left[\hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)\right] \sim$$

$$\sim N(0, var\left[\phi(X_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h)\right]).$$

Proof. Write

$$S^{\frac{1}{2}} \left[\hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - P(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)\right] =$$

$$= \frac{1}{S^{\frac{1}{2}}} \sum_{s=1}^{S} \phi(X_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h) - q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) +$$

$$+ S^{\frac{1}{2}} \left[q_M(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - P(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)\right].$$

Lemma 1.2.1 and the condition $\frac{S^{\frac{1}{2}}}{M} \to 0$ ensure that as $M \to \infty$ the second term in the sum converge to zero. We complete the proof applying the central limit theorem to the first term. □

If the diffusions are stationary and ergodic, the unconditional density can also be evaluated with simulations. Under the assumption of stationarity and ergodicity, the unconditional density does not depend on time, or $p(x, t_0; \theta) = p(x; \theta)$ with $p(x; \theta) = \lim_{t \to \infty} p(x, t \mid y, 0; \theta)$. This implies that we can start with any initial x and use the Euler Maruyama discretization to simulate a long continuous sample path of the diffusion. Then, we can approximate the unconditional probability of $x = X_0$ from the simulated data using standard density estimation tools. If the diffusions are non stationary, we need to assume a deterministic X_0. Fortunately, this assumption has a negligible effect on the likelihood function for sufficiently large samples.

Given the above approximations of the transition densities and of the initial unconditional density, we construct a consistent approximation of the likelihood function $\mathcal{L}(\theta)$. We define the simulated maximum likelihood estimator $\hat{\theta}_{M,S}$ as the parameters that maximize:

$$\ln \hat{\mathcal{L}}_{M,S}(\theta) = \ln \hat{q}_{M,S}(X_0, t_0; \theta) + \sum_{n=0}^{N-1} \ln \hat{q}_{M,S}(X_{n+1}, t_{n+1} \mid X_n, t_n; \theta). \quad (3.9)$$

We call this method of approximation for a transition density Brandt - Santa Clara. Some application of this method are shown in section 3.1.3 and 3.3.

Since the approximations of the unconditional density and of the transition densities converge to their true counterparts, it follows that this approximate log likelihood function converges to the true log likelihood function.

Lemma 3.1.4. *Under the assumption considered in this section, as* $N \to \infty$, $M \to \infty$, *and* $S \to \infty$, *with* $\frac{S^{\frac{1}{2}}}{M} \to 0$,

$$ln\hat{\mathcal{L}}_{M,S}(\theta) - ln\mathcal{L}(\theta) = o\left(\frac{N}{S^{\frac{1}{2}}}\right).$$

Proof. Let x_n denote the errors of the simulated transition densities:

$$x_n \equiv \hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta).$$

Let $p_n = p(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta)$ and write:

1.
$$\ln \hat{q}_{M,S}(X_{t_{n+1}}, t_{n+1} \mid X_{t_n}, t_n; \theta) - \ln p_n =$$
$$= \ln(x_n + p_n) - \ln p_n = \ln\left(1 + \frac{x_n}{p_n}\right).$$

Expanding the last term around $x_n = 0$ for a fixed p_n implies that for a sufficiently small x_n,

$$\ln\left(1 + \frac{x_n}{p_n}\right) \approx \frac{x_n}{p_n} + o\left(\frac{1}{S^{\frac{1}{2}}}\right)$$

The last equality follows from Lemma 1.2.3. Substituting the expansion into the equation 1 and summing over the N sample points completes the proof. □

The asymptotics of the SML method are summarized in the follows theorems.

Theorem 3.1.5. *Under all the assumption that we had done in this section, as* $M \to \infty$ *and* $S \to \infty$, *with* $\frac{S^{\frac{1}{2}}}{M} \to 0$, *the estimator* $\hat{\theta}_{M,S}$ *converges to the maximum likelihood estimator* $\hat{\theta}$, *which in turn converge to the true parameter vector* θ_0 *as* $N \to \infty$.

To prove this theorem we need all the lemmas listed in this section and the follows.

Lemma 3.1.6. *Under the hypothesis of Theorem 3.1.5, as* $N \to \infty$, $M \to \infty$ *and* $S \to \infty$, *with* $\frac{S^{\frac{1}{2}}}{M} \to 0$,

$$\hat{\theta}_{M,S} - \hat{\theta} = o\left(\frac{N^{\frac{1}{2}}}{S^{\frac{1}{4}}}\right),$$

where $\hat{\theta}$ *is the parameter vector the maximizes* $\ln \mathcal{L}(\theta)$.

Lemma 3.1.7. *Under the hypothesis of Theorem 3.1.5, as* $N \to \infty$,

$$\hat{\theta} \to \theta_0.$$

Lemma 3.1.8. *Assume that the hypothesis of the Theorem 3.1.5 are satisfied moreover the gradient* $\frac{\partial p(X_{t_{n+1}}, t_{n+1}|X_{t_n}, t_n; \theta)}{\partial \theta}$ *converges as* $N \to \infty$ *or diverges at a rate slower than the rate of convergence of* $I(\theta_0)^{\frac{1}{2}}$ *to zero, see* (3.3). *Then as* $N \to \infty$, *we have*

$$I(\theta_0)^{\frac{1}{2}} \left[\hat{\theta} - \theta_0\right] \sim N(0,1)$$

The last lemma and the consistency of our estimator from Theorem 3.1.5 imply the following theorem.

Theorem 3.1.9. *Under the hypothesis of the previous lemma, as* $N \to \infty$, $M \to \infty$ *and* $S \to \infty$, *with* $\frac{S^{\frac{1}{2}}}{M} \to 0$ $\frac{N}{S^{\frac{1}{4}}} \to 0$ *the asymptotic distribution of the estimator* $\hat{\theta}_{M,S}$ *is:*

$$I(\theta_0)^{\frac{1}{2}} \left[\hat{\theta}_{M,S} - \theta_0\right] \sim N(0,1).$$

3.1.2 Durham and Gallant

The simulation approach suggested by Brandt and Santa-Clara (2002) has great theoretical appeal, but previously available implementations have been computationally costly. In this section we examine a numerical technique propose by Durham and Gallant (2002), which claimed to improve the performance of Brendt-Santa Clara approach. See section 3.1.3 for numerical comparisons.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ a probability space, and let $\{W_t\}_{t\geq 0}$ be a Brownian motion defined on it. Let $\{\mathcal{F}_t, t \geq 0\}$ be a filtration generated by $\{W_t\}$ and augmented by P- null sets of $\mathcal{F}$. Let Θ be a compact subset of $\mathbb{R}^d$. We are interested in the parameterized family of scalar diffusion process $\{W(t;\theta), \theta \in \Theta\}$ generated by time-homogeneous process SDE of the form:

$$\begin{cases} dX = \mu(X;\theta)dt + \sigma(X;\theta)dW; \\ X(t_0) = X_0. \end{cases} \tag{3.10}$$

Let $\{X_i = X(t_i), i = 0, ..., n\}$ to be a sample. According to Durham and Gallant (2002), we make some assumptions.

Assumption 1 For each $\theta \in \Theta$, (3.10) has a non-exploding, unique weak solution.

By non-exploding, we mean that there is zero probability that the process diverges to infinity over any fixed time interval. Explosiveness would preclude the existence of a transition density and is thus disallowed. The basic idea is quite simple. We consider $t, s \in [0, T]$ and we suppose $s < t$. We wish to obtain the transition density $p(X_t, t \mid X_s, s; \theta)$. We know that the the first-order approximation $p^{(1)}(X_t, t \mid X_s, s; \theta)$ defined by equation (3.11) will be accurate if the interval $[s, t]$ is sufficiently short.

$$X_{i+1} = X_i + \mu(X_i; \theta)\Delta_i + \sigma(X_i; \theta)\Delta_i^{\frac{1}{2}}\varepsilon_i \tag{3.11}$$

where $\Delta_i = t_{i+1} - t_i$, and $\varepsilon_i \sim N(0, 1)$.

So $p^{(1)} = \phi(X_t, X_s + \mu(X_s; \theta)\Delta, \sigma(X_s; \theta)^2\Delta)$, where ϕ is the Gaussian density $\phi(x, \mu, \sigma^2)$. Otherwise we may partition the interval in M subintervals of length $h = \frac{\Delta}{M}$, such that $s = \tau_1 < \tau_2 < ... < \tau_M = t$, so that the first-order approximation is sufficiently accurate on each subinterval. As in (3.7) we have

$$q_M(X_t, t \mid X_s, s; \theta) =$$

$$= \int_{\mathbb{R}^{M-1}} \prod_{m=0}^{M-1} p^{(1)}(z_{m+1}, \tau_{m+1} \mid z_m, \tau_m; \theta) d\lambda(z_1 ..., z_{M-1})$$

where $z_0 = X_s, z_M = X_t$ and λ denotes the Lebesgue measure. Multiplying and dividing by $q(z_1, ..., z_{M-1})$, a probability density on $\mathbb{R}^{M-1}$, in (3.7), we obtain:

$$q_M(X_t, t \mid X_s, s; \theta) =$$

$$= \int_{\mathbb{R}^{M-1}} \frac{\prod_{m=0}^{M-1} p^{(1)}(z_{m+1}, \tau_{m+1} \mid z_m, \tau_m; \theta)}{q(z_1, ..., z_{M-1})} q(z_1, ..., z_{M-1}) d\lambda(z_1 ..., z_{M-1}).$$

The difficulty is how to efficiently evaluate the integral. Monte Carlo integration is generally the only feasible approach. To perform Monte Carlo integration, we require an importance sampler. According to Durham and Gallant notation we let $\{u_k = (u_{k,1}, ..., u_{k,M-1}), k = 1, ..., S\}$ be independent draws from $q(u_1, ..., u_{M-1})$, by Monte Carlo integration we obtain:

$$\hat{q}_{M,S}(X_t, t \mid X_s, s; \theta) = \frac{1}{S} \sum_{k=1}^{S} \frac{\prod_{m=0}^{M-1} p^{(1)}(u_{k,m+1}, \tau_{m+1} \mid u_{k,m}, \tau_m; \theta)}{q(u_{k,1}, ..., u_{k,M-1})}, \tag{3.12}$$

where $u_{k,0} = X_s$ and $u_{k,M} = X_t$ for all k.

Assumption 2 Let $U_0 = x_s$, $U_M = x_t, \theta \in \Theta$, and q be fixed, an let $(U_1, ..., U_{M-1})$

be a random vector with density q. Then

$$\mathbb{E}\left[\frac{\prod_{m=0}^{M-1} p^{(1)}(U_{m+1},\tau_{m+1} \mid U_m,\tau_m;\theta)}{q(U_1,...,U_{M-1})}\right] < \infty.$$

Under the Assumption 2, the strong law of large numbers implies that:

$$\lim_{S\to\infty} |\hat{q}_{M,S}(X_t,t \mid X_s,s;\theta) - q_M(X_t,t \mid X_s,s;\theta)| = 0.$$

If we use Euler Maruyama scheme to generate the sampler, as in Brandt - Santa Clara, we obtain that

$$q(u_{k,1},...,u_{k,M-1}) = \prod_{m=0}^{M-2} p^{(1)}(u_{k,m+1},\tau_{m+1} \mid u_{k,m},\tau_m;\theta),$$

since the density of the important sampler q is identical to the first $M-1$ factor of numerator in (3.12), they cancel and we left with (3.8).

Durham and Gallant examine same approach to reducing the variance of Monte Carlo integration. A basic principle of Monte Carlo integration is that we should draw points with higher probability in regions where the integrand is large. The reason why Brandt and Santa Clara method performs so poorly is that most of the samples are drawn from regions where the integrand has little mass; as Durham and Gallant (2002) the samplers discussed in this section are designed to address this shortcoming. The first important sampler we consider is based on the Brownian bridge. A Brownian bridge is a Brownian motion started as X_s at time s and conditioned to terminate at X_t at time t. The sampler is constructed in a manner similar to Euler Maruyama scheme. In this case, the mapping

$$T^{(M)} : (W_1,...,W_{M-1};\theta) \to (u_1,...,u_{M-1})$$

is define by recursion

$$u_{m+1} = u_m + \tilde{\mu}(u_m,\tau_m)h + \sigma(u_m,\tau_m;\theta)h^{\frac{1}{2}}W_{m+1}, \tag{3.13}$$

where the drift is given by

$$\tilde{\mu}(x,\tau) = \frac{X_t - x}{t-\tau}.$$

This is a Brownian bridge if and only if σ is constant. Figure 3.2.

Although it is possible to compute the approximate density directly from (3.12),

there is an interesting interpretation of this sample based on Girsanov's Theorem. Using this sample we obtain that

$$q(u_1, ..., u_{M-1}) = \prod_{m=1}^{M-1} p(u_m, \tau_m \mid u_{m-1}, \tau_{m-1}) =$$

$$\prod_{m=1}^{M-1} \phi(u_m; u_{m-1} + \tilde{\mu}(u_{m-1}, \tau_{m-1})h; \sigma(u_{m-1}, \tau_{m-1}; \theta)^2 h).$$

The second important sampler which we consider draws u_{m+1} from a Gaussian density based on the first order approximation, conditioned on u_m and X_t. That is, treating u_m and $u_M = X_t$ as fixed values, one draws u_{m+1} from the density

$$p(u_{m+1} \mid u_m, u_M) = \frac{p(u_{m+1} \mid u_m)p(u_M \mid u_{m+1})}{p(u_M \mid u_m)} \approx$$

$$\approx \frac{\phi(u_{m+1}; u_m + \bar{\mu}h; \bar{\sigma}^2 h)\phi(u_M; u_{m+1} + \bar{\mu}h^*; \bar{\sigma}^2 h^*)}{\phi(u_M; u_m + \bar{\mu}h^+; \bar{\sigma}^2 h^+)} =$$

$$= \phi(u_{m+1}; u_m + \tilde{\mu}_m h, \tilde{\sigma}_m^2 h),$$

where $h = \frac{(t-s)}{M}$, $h^* = t - \tau_{m+1}$, $h^+ = t - \tau_m$, $\bar{\mu} = \mu(u_m)$, $\bar{\sigma} = \sigma(u_m)$, and

$$\tilde{\mu}_m = \frac{u_M - u_m}{t - \tau_m}, \qquad \tilde{\sigma}_m^2 = \left(\frac{M - m - 1}{M - m}\right)\bar{\sigma}^2.$$

Note that u_{m+1} is define by recursion

$$u_{m+1} = u_m + \tilde{\mu}_m h + \tilde{\sigma}_m h^{\frac{1}{2}} W_{m+1}, \tag{3.14}$$

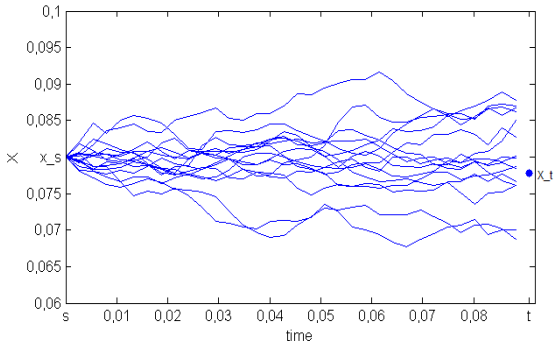
and

$$q(u_1, ..., u_{M-1}) = \prod_{m=1}^{M-1} \phi(u_m; u_{m-1} + \tilde{\mu}_m h; \tilde{\sigma}_m^2 h).$$

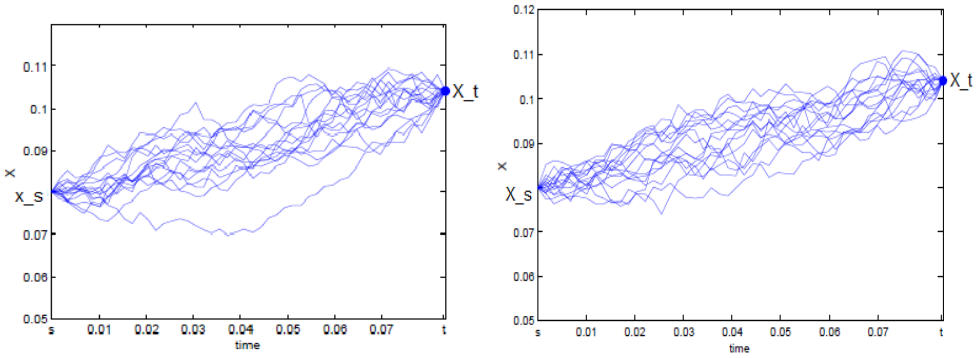
Note that this importance sampler turns out to be identical to the Brownian bridge sampler except the factor $\frac{M-m-1}{M-m}$ in the variance. According to Durham and Gallant (2002), we refer to this sampler as modified Brownian bridge, Figure3.2.

3.1.3 Durham and Gallant vs Brandt - Santa Clara

To compare the results of the different models described in Section 3.1.2 and 3.1.1, we use the Cox- Ingersoll-Ross model (CIR) (Cox, Ingersoll and Ross-1985). It was suggested as a model of the short interest rate, although the



(a) Sampler of Brandt - Santa Clara



(b) Modified bridge

(c) Brownian bridge

Figure 3.2: Simulated paths drawn using various importance samplers. We use the CIR model.

mathematical model was originally introduced by Feller (1952). Different parameterizations have been presented in the literature, and we used the following:

$$dX_t = k(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t \tag{3.15}$$

$$X_0 = x_0 > 0.$$

By limiting the parameter space $\Omega = \{(\alpha, k, \sigma) \mid (\alpha, k, \sigma) \in (0, \infty) \times (0, \infty) \times (0, \infty)\}$, the state space is given by $(X_t, t) \in [0, \infty) \times [0, T]$. The origin $(x = 0)$ is inaccessible if $2\alpha k \geq \sigma^2$, otherwise it is reflecting, see Feller (1951). Similarly, the Maximum Likelihood regularity conditions are valid if $2\alpha k \geq \sigma^2$, see Overbeck and Rydén (1997). The success of this model is due to the fact that (given the requirements on the parameters):

- the process is always non - negative;
- the mean converges towards the steady-state mean, α;
- closed form expressions can be derived for a large class of financial contracts due to the affine form of the drift term and the squared diffusion term.

Furthermore, the parameters in the model can be estimated using Maximum Likelihood estimator since the transition probabilities are explicitly known. It can be shown that the transition probability density is given by:

$$p(X_t, t \mid X_s, s; \theta) = \frac{\partial}{\partial x_t} P_\theta(X_t \leq x_t \mid X_s = x_s; \theta) =$$

$$= c \cdot e^{(-cx_t - c\delta x_s)} \left(\frac{x_t}{x_s \delta}\right)^{\frac{q}{2}} I_q(2c\sqrt{x_s x_t \delta}), \tag{3.16}$$

where $t, s \in [0, T]$,with $s < t$; $x_t, x_s \in [0, \infty)$; $I_q(z)$ is a modified Bessel function of the first kind of order q; and

$$\delta = e^{-k(t-s)}, \qquad c = \frac{2k}{\sigma^2(1-\delta)} \qquad q = \frac{2k\alpha}{\sigma^2} - 1.$$

We will use the Cox-Ingersoll-Ross model to measure the accuracy of the approximation of the transition probability density. We use time series of $N = 1000$ data simulated by Euler Maruyama scheme ($x_0 = 0.08$):

$$X_{t+1} = X_t + \mu(X_t)\Delta + \sigma(X_t)(W_{t+1} - W_t); \tag{3.17}$$

where $\{W_t\}_{t_0 \leq t \leq T}$ are stochastically independent and identically standard normally distributed random variables. We consider a uniform time discretization, so Δ is constant for all $t \in [0, T]$, in particular $\Delta = \frac{1}{12}$. Using the simulations data we approximate the transition density as Sections 3.1.1 and 3.1.2 describe. We used the MATLAB function `fminsearch` to minimize the function $-\ln(\mathcal{L})$, and obtain $\hat{\theta}$, the simulated maximum likelihood estimator. It is particulary important, during the maximization, to use the same vector $\{W_t\}_{t_0 \leq t \leq T}$. With the random generator re-initialized at the constant seed to ensure that the same integration base is re-simulated in each calculation of the approximate likelihood function, for this reason we use the MATLAB function `rng`. See Pedersen

	$\hat{\alpha}$	$\hat{k}$	$\hat{\sigma}$	Log-Likelihood
True density	0.0602	0.5484	0.1525	3.2120×10^3
Euler	0.0605	0.5200	0.1448	3.2599×10^3
Brandt - Santa Clara	0.056	0.6648	0.1666	7.1195×10^3
Durham and Gallant	0.0635	0.4167	0.1443	3.3126×10^3
Modified Brownian bridge	0.0618	0.4903	0.1503	3.3186×10^3

Table 3.1: Parameters estimate obtain by different approximation of the log-likelihood, using $\alpha = 0.06, k = 0.5$, and $\sigma = 0.15.,\Delta = \frac{1}{12}$ and $x_0 = 0.08$. For Durham and Gallant, modified Brownian bridge and Brant-Santa Clara simulation we use $M = 16$ and $S = 50$.

(1995). The quality of estimator depends on three quantities: the simple size N, the number the discratization steps M, and the simulation size S. The parameters used as starting value of the maximization are:$\alpha = 0.06, k = 0.5$, and $\sigma = 0.15$. We use also the Euler approximation to estimate the parameter,

$$p^{Euler}(X_{t+\delta}, \Delta \mid X_t; \theta) = (2\pi\Delta\sigma(X_t;\theta)^2)^{-\frac{1}{2}} \exp\left[\frac{-(X_{t+\delta} - X_t - \mu(X_t;\theta)\Delta)^2}{2\Delta\sigma^2(X_t;\theta)}\right]. \tag{3.18}$$

Since the performance of estimators is evaluated by comparing an Euler approximation, Durham and Gallant simulation, Brandt - Santa Clara method and Durham and Gallant approximation with a modified Brownian bridge. The estimates are summarized in Table 3.1.

Table 3.1 suggest that the methods proposed by Durham and Gallant (2002) lead to better results than Brandt Santa-Clara method. Now we try to test our approximations as in Lindström (2006). We told about the uniform convergence of the approximate likelihood to the true likelihood for all values of $\theta \in \Theta$. This condition is impossible to test numerically, but it is possible to test the approximate likelihood converges for a fixed θ. A conservative approximation of the distance between the approximate and true likelihood function is give by:

$$\left|\sum_{i=1}^{N} \log \hat{P}(X_{t_i}, t_i \mid X_{t_{i-1}}, t_{i-1}) - \log P(X_{t_i}, t_i \mid X_{t_{i-1}}, t_{i-1})\right| \leq$$

$$\leq \sum_{i=1}^{N} |\log \hat{P}(X_{t_i}, t_i \mid X_{t_{i-1}}, t_{i-1}) - \log P(X_{t_i}, t_i \mid X_{t_{i-1}}, t_{i-1})|.$$

By weighting the distance by $P(X_{t_i}, t_i \mid X_{t_{i-1}}, t_{i-1})$ and scaling by number of observations, we derive the mean absolute error (MAE) of the log-likelihood function.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\log \hat{P}(X_{t_i} \mid X_{t_{i-1}}) - \log P(X_{t_i} \mid X_{t_{i-1}})| \tag{3.19}$$

$$\approx \int |\log \hat{P}(X_{t_i} \mid X_{t_{i-1}}) - \log P(X_{t_i} \mid X_{t_{i-1}})| P(X_{t_i} \mid X_{t_{i-1}}) dX_i dX_{i-1}.$$

We can also measure the convergence as the root mean square error ($RMSE$) of the log-likelihood function.

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^{N} (\log \hat{P}(X_{t_i} \mid X_{t_{i-1}}) - \log P(X_{t_i} \mid X_{t_{i-1}}))^2 \right)^{\frac{1}{2}}. \tag{3.20}$$

In the following we only consider the RMSE.

The figure 3.3 show that we obtain the worst results applying the Brandt - Santa Clara method. The simulation-based approach suggest by Brandt-Santa Clara is appealing from a theoretical and intuitive viewpoint; however we find that it can be prohibitively costly to attain even the accuracy of the simple first-order approximation. Our results, according to Durham and Gallant (2002) and Lindström (2006), suggest that the best performance is obtained using the modified Brownian bridge sampler. Using the Brownian bridge largely solves the main problem associate with Brandt- Santa Clara's method. The modified Brownian bridge provides a futher dramatic reduction in variance. The number of subintervals, M, and sample paths, S, must be determined by experimentations.

3.2 Closed Form Approximations

In this section we would like to consider a different method to approximate the transition density, it is presented in Aït-Sahalia (1999). This method is based

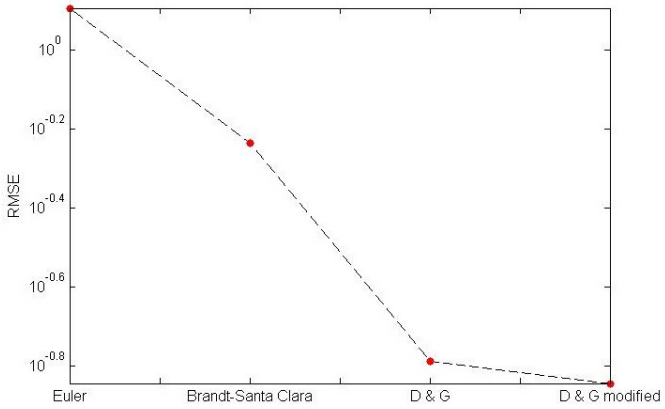


Figure 3.3: RMSE calculated using (3.20). $N = 1000$,$\alpha = 0.06, k = 0.5$, and $\sigma = 0.15$.,$\Delta = \frac{1}{12}$ and $x_0 = 0.08$. We compere Durham and Gallant, modified Brownian bridge, Brant-Santa Clara and Euler simulation we use $M = 16$ and $S = 50$.

on series of closed form approximations of the density. Numerical results show that this methodology improves the estimations and it possible to observe a decreased of computational time, see Section 3.3.

The firs step toward constructing the sequence of approximations to p_X consists of standardizing the diffusion function of X, transforming X into another diffusion Y defined as:

$$Y_t = \gamma(X_t; \theta) = \int^{X_t} \frac{du}{\sigma(u; \theta)}; \tag{3.21}$$

where any primitive of function $\frac{1}{\sigma(u;\theta)}$ may be selected. Let $D_X = (\underline{x}, \bar{x})$ denote the domain of the diffusion X. Because $\sigma > 0$ on the interior of the domain D_X, the function γ in (3.21) is increasing and thus invertible. It maps D_X into $D_Y = (\underline{y}, \bar{y})$, the domain of Y. For a given model under consideration, we will assume that the parameter space Θ is restricted in such a way that D_Y is independent of θ in Θ. This restriction on Θ is inessential, but it helps keep the notation simple. Note that in most of financial models we will have D_X and D_Y be either the whole real line, $(-\infty, +\infty)$ or the half line $(0, +\infty)$.

By applying formula (2.1), Y has unit diffusion as desired:

$$dY_t = \mu_Y(Y_t; \theta)dt + dW_t,$$

where

$$\mu_Y(Y_t;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta).$$

The motivation of the transformation from X to Y is that it is possible to construct an expansion for the transition density of Y. Of course it would be a little interest because we only observe X, not the artificially introduced Y, and the transformation depends on the unknown parameter vector θ. The main objective of the transformation was to provide a method of controlling the size of the tails of the transition density. The fact that Y has unit diffusion makes the tails of the density p_Y, in the limit where Δ goes to zero, similar in magnitude to those of Gaussian variable. So the tails of the density p_X are proportional to $exp\left(\frac{-\gamma(x;\theta)^2}{2\Delta}\right)$. In other words, while the leading term of expansion for p_Y is Gaussian, the expansion for p_X will start with a deformed Gaussian term, with the specific form of the deformation given by the function $\gamma(x;\theta)$. However the transformation is also useful because one can obtain the transition density p_X from p_Y through the Jacobian formula:

$$\begin{aligned} p_X(x,\Delta \mid x_0;\theta) &= \frac{\partial}{\partial x}P(X_{t+\Delta} \le x \mid X_t = x_0;\theta) = \\ &= \frac{\partial}{\partial x}P(Y_{t+\Delta} \le \gamma(x;\theta) \mid Y_t = \gamma(x_0;\theta);\theta) = \\ &= \frac{\partial}{\partial x}\left[\int_{\underline{y}}^{\gamma(x;\theta)} p_Y(y,\Delta \mid \gamma(x_0;\theta);\theta)dy\right] = \\ &= \frac{p_Y(\gamma(x,\theta),\Delta \mid \gamma(x_0;\theta);\theta)}{\sigma(\gamma(x;\theta);\theta)}; \end{aligned} \tag{3.22}$$

where $x, x_0 \in D_X$ and $y, y_0 \in D_Y$. Therefore, there is never any need to actually transform the data into observations on Y. Instead, the transformation is a simply a device to obtain an approximation for p_X from the approximation of p_Y.

As shown in Aït-Sahalia (1999), one can derive an explicit expansion for the transition density of the variable Y based on a Hermite expansion of its density, around a Normal density function. The analytic part of the expansion of p_Y up to order K is given by:

$$\hat{p}_Y^{(K)}(y,\Delta \mid y_0;\theta) = \Delta^{-\frac{1}{2}}\phi\left(\frac{y-y_0}{\Delta^{-\frac{1}{2}}}\right) exp\left(\int_{y_0}^{y}\mu_Y(\omega;\theta)d\omega\right)\sum_{k=0}^{K} c_k(y \mid y_0;\theta)\frac{\Delta^k}{k!}, \tag{3.23}$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$ denotes the $N(0,1)$ density function, $c_0(y \mid y_0; \theta) = 1$, and for all $j \geq 1$,

$$c_j(y \mid y_0; \theta) = j(y - y_0)^{-j} \int_{y0}^{y} (\omega - y_0)^{j-1} \times$$

$$\times \left[\lambda_Y(\omega) c_{j-1}(\omega \mid y_0; \theta) + \frac{\frac{\partial^2}{\partial \omega^2} c_{j-1}(\omega \mid y_0; \theta)}{2} \right] d\omega, \tag{3.24}$$

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

The leading term in expansion is the Gaussian, followed by a correction term that depend on the specification of the function $\lambda(y; \theta)$ and its successive derivatives. This correction term play two roles: they account for the non-normality of p_Y and they correct for the discretization bias implicit in starting the expansion with a gaussian term with no mean adjustment and variance Δ.

In general, the function p_Y is not analytic in time. Therefore (3.23) must be interpreted strictly as the analytic part, or Taylor series. In particular, for given (y, y_0, θ) it will generally have a finite convergence radius in Δ. The sequence of explicit function $\hat{p}_Y^{(K)}$ is designed to approximate p_Y. As discussed above, one can then approximate p_X by using the Jacobian formula for the inverted change of variable:

$$\hat{p}_X^{(K)} = \sigma(x; \theta)^{-1} \hat{p}_Y^{(K)}(\gamma(x; \theta), \Delta \mid \gamma(x_0; \theta); \theta). \tag{3.25}$$

Now we can obtain the approximation of the log-likelihood function:

$$\hat{\mathcal{L}}^{(K)}(\theta) = \sum_{i=1}^{N} \log(\hat{p}_X^{(K)}(X_{i\Delta}, \delta \mid X_{(i-1)\Delta}; \theta)).$$

Increasing the index K the accuracy of this method could be improved.

3.3 Comparison of Monte Carlo vs Closed Form

In this section we try to use and compare the approximation described in Section 3.1.1 and 3.2 . We consider the Ornstein - Uhlenbeck model proposed by Vasicek (1977) for the short - term interest rate:

$$dX_t = k(\alpha - X_t)dt + \sigma dW_t. \tag{3.26}$$

X is distributed on $D_X = (-\infty, +\infty)$ and has the Gaussian transition density:

$$P_X(\Delta, x_t \mid x_{t-1}; \theta) = \left(\frac{\pi\gamma^2}{k}\right)^{-\frac{1}{2}} exp\left[\frac{-(x_t - \alpha - (x_{t-1} - \alpha)e^{-k\Delta})^2}{\gamma^2}\right], \quad (3.27)$$

where $\theta = (\alpha, k, \sigma)$, $\Delta = t_{i+1} - t_i$, and $\gamma^2 = \sigma^2(1 - e^{-2k\Delta})$. As Aït-Sahalia (1999), we make the maximum likelihood estimation for the parameters of the model using the Fed found data, monthly from January 1963 to December 1998 ($N = 432$) with a Matlab following code.

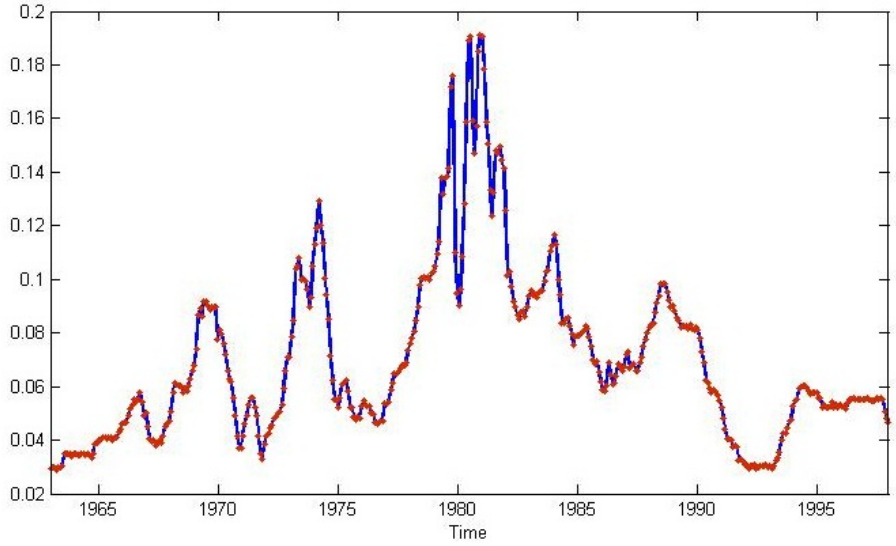


Figure 3.4: Fed Found data, monthly from January 1963 to December 1998

We apply the Brandt - Santa Clara method presented in Section 3.1.1 to the Vasicek model. As described in section 3.1.3 for each time interval $[t_i, t_{i+1}]$, we construct S trajectories iterating $M-1$ times,with the Euler-Maruyama scheme. We consider a uniform time discretization, so Δ is constant for all $t \in [0, T]$, in particular $\Delta = \frac{35}{N}$. Then we use the Hermite expansion with $K = 1, 2$ as shown in Section 3.2. From (3.23) we obtain:

$$\hat{p}_Y^{(1)}(y, \delta \mid y_0; \theta) = \hat{p}_Y^{(0)}(y, \delta \mid y_0; \theta)(1 + c_1(y \mid y_0; \theta)\Delta)$$

$$\hat{p}_Y^{(2)}(y, \delta \mid y_0; \theta) = \hat{p}_Y^{(0)}(y, \delta \mid y_0; \theta)(1 + c_1(y \mid y_0; \theta)\Delta + c_2(y \mid; \theta)\frac{\Delta^2}{2})$$

where

$$\hat{p}_Y^{(0)}(y,\delta \mid y_0;\theta) = \frac{1}{\sqrt{2\Delta\pi}} exp\left[\frac{(y-y_0)^2}{2\Delta} - \frac{y^2k}{2} + \frac{y_0^2k}{2} + \frac{y\alpha k}{\sigma} - \frac{y_0\alpha k}{\sigma}\right].$$

The term in expansion are evaluated by applying the formula (3.24).

$$c_0(y \mid y_0,\theta) = 1$$

$$c_1(y \mid y_0,\theta) = -\frac{k}{6\sigma^2}(3\alpha^2k - 3(y+y_0)\alpha k\sigma + (-3+y^2k+yy_0k+y_0^2k)\sigma^2);$$

$$c_2(y \mid y_0,\theta) = \frac{k^2}{36\sigma^4}(9\alpha^4k^2 - 18y\alpha^3k^2\sigma + 3\alpha^2k(-6+5y^2k)\sigma^2 -$$

$$-6y\alpha k(-3+y^2k)\sigma^3 + (3-6y^2k+y^4k^2)\sigma^4 +$$

$$+2k\sigma(-3\alpha+y\sigma)(3\alpha^2k - 3y\alpha k\sigma + (-3+y^2k)\sigma^2)y_0 +$$

$$+3k\sigma^2(5\alpha^2k - 4y\alpha k\sigma + (-2+y^2k)\sigma^2)y_0^2 +$$

$$+2k^2\sigma^3(-3\alpha+y\sigma)y_0^3 + k^2\sigma^4y_0^4).$$

In this case $Y_t = \gamma(X_t;\theta) = \sigma^{-1}X_t$, $\mu_Y(y,\theta) = k\alpha\sigma^{-1} - ky$, and $\lambda_Y(y;\theta) = \frac{k}{2} - \frac{k^2(\alpha-\sigma y)^2}{2\sigma^2}$.

Finally we use also the Euler approximation, (3.18) to estimate the parameter. Then we compere the results with the parameters estimate by the True density,see Table 3.2.

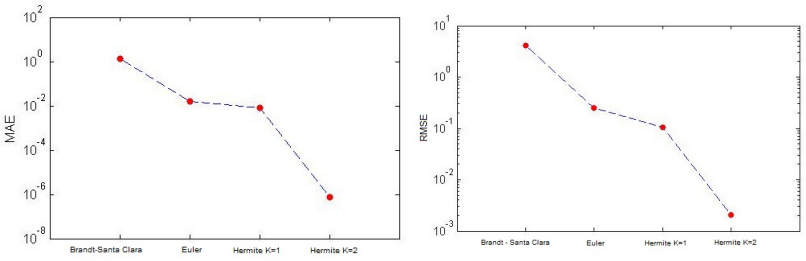
Numerical results in Table 3.2 shows that the first method, propose by Brandt and Santa-Clara (2002), does not work. One reason could be that most of the samples are draw from regions where the integrand has little mass. Furthermore the model propose a first order approximation, which holds when $t_{i+1} - t_i$ tends to zero.

Then, to test our models, we measure the convergence as the root mean square error ($RMSE$) explained in (3.20), see Section 3.1.3. We evaluate the MAE an RMSE on data series of $N = 432$ simulated by Euler-Maruyama scheme, using $\alpha = 0.0717$, $k = 0.258$, $\sigma = 0.02213$ the starting point $x_0 = 0.1$ in a time interval from 1963 to 1998.see Figure 3.5

Aït-Sahalia's Hermite polynomial expansion is clearly the best method in terms of the speed and accuracy. Small values of K already produce extremely

	Parameter estimate	Log-Likelihood
True Density	$\hat{\alpha} = 0.07170$ (0.0002) $\hat{k} = 0.226$ (0.015) $\hat{\sigma} = 0.0226$ (6.627×10^{-7})	1.5706×10^3
Expansion $K = 1$	$\hat{\alpha} = 0.07196$ (0.0002) $\hat{k} = 0.2637$ (0.136) $\hat{\sigma} = 0.0226$ (5.9×10^{-7})	1.5706×10^3
Expansion $K = 2$	$\hat{\alpha} = 0.07170$ (0.0002) $\hat{k} = 0.2676$ (0.015) $\hat{\sigma} = 0.0226$ (6.3×10^{-7})	1.5705×10^3
B and SC	$\hat{\alpha} = 0.0659$ (0.00005) $\hat{k} = 0.6975$ (0.166) $\hat{\sigma} = 0.0324$ (7.63×10^{-7})	3.86107×10^3
Euler	$\hat{\alpha} = 0.07169$ (0.0002) $\hat{k} = 0.2652$ (0.014) $\hat{\sigma} = 0.0224$ (5.56×10^{-7})	1.5706×10^3

Table 3.2: Parameters estimate (observed asymptotic standard errors,$I(\theta)$.) obtained by different approximation of the log-likelihood, using the Feunds data, monthly from January 1963 through December 1998. B and SC indicates the method propose by Brandt and Santa-Clara (2002), we use $S = 256$ and $M = 8$. The maximization start from the vector $[0.08, 0.3, 0.05]$



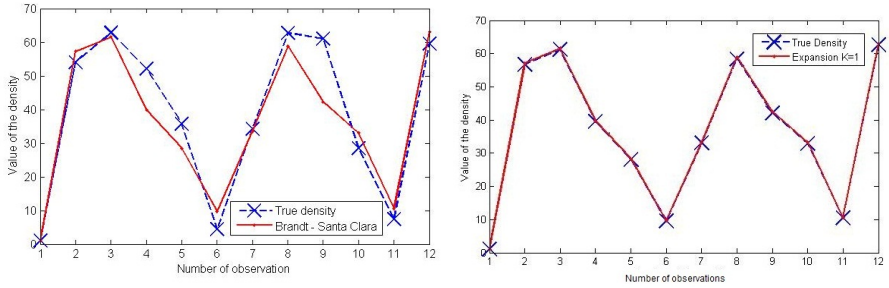
(a) MAE of the log-likelihood function

(b) RMSE of the log-likelihood function

Figure 3.5: MAE and RMSE are calculated on data series of $N = 432$ simulated by Euler Maruyama scheme, using $\alpha = 0.0717$, $k = 0.258$, $\sigma = 0.02213$, and $x_0 = 0.1$, in a time interval from 1963 to 1998.

precise approximations to the true density, and the approximation is even more precise if Δ is smaller. Of course, the exact density being Gaussian, in this case the expansion, whose leading term is Gaussian, has fairly little work to do to approximate the true density. In this case, the expansion involves no correction for non-normality, which is normally achieved through the change of variable X to Y; it reduces here to a linear transformation and therefore does not change the nature of the leading term in the expansion. Comparing the performance of the expansion to that of the Euler approximation in this model (where both have the correct Gaussian form for the density) reveals that the expansion is capable of correcting the discretization bias involved in a discrete approximation, whereas the Euler approximation is limited to a first-order bias correction. In this case, the Euler approximation can be refined by increasing the precision of the conditional mean and variance approximations. The worst way to approximate the density seems to be the simulation approach described first (Brandt-Santa Clara). An obvious extension of this study would be to apply the approximation techniques to model where we do not know the transition density, since we would use the approximations in the first place in this case. All the methods are so general that this is fairly easy to do. The Figure 3.6 and 3.7 show how the approximations are close to the true density. We use

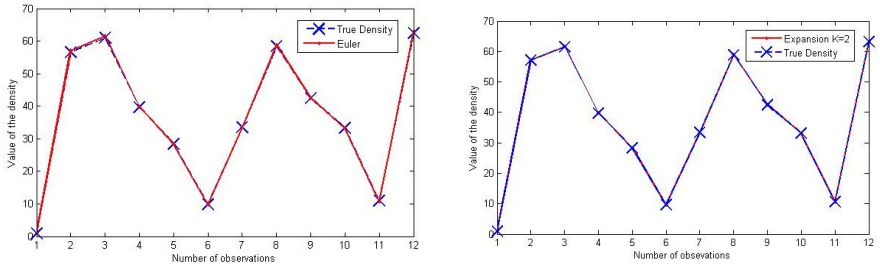
a sample of 12 simulate data in a unit time interval, which are obtain using Euler-Maruyama scheme.



(a) True density and Brandt - Santa Clara approximation

(b) True density and Hermite expansion $K = 1$

Figure 3.6: Comparison between True density and the different approximations. We use the Euler Maruyama scheme for the simulation of 12 data, then we calculate the value of the different approximation transition density in each point. We use $\alpha = 0.0717$, $k = 0.258$ and $\sigma = 0.02213$ The starting point is $x_0 = 0.1$. We use M=10 and S=50.



(a) True density and Euler approximation

(b) True density and Hermite expansion $K = 2$

Figure 3.7: Comparison between true density and the different approximations. We use the Euler Maruyama scheme for the simulation of 12 data, then we calculate the value of the different approximation transition density in each point. We use $\alpha = 0.0717$, $k = 0.258$ and $\sigma = 0.02213$ The starting point is $x_0 = 0.1$.

3.4 Conclusions

Despite the theoretical advantages of maximum likelihood estimations, the approximation it is hard to do, indeed often transition densities are not known. The numerical experiments showed that the simulated likelihood estimation lead to better results using the following importance sampling strategies: Brownian bridge and modifed Brownian bridge, both proposed by Durham and Gallant (2002), in particular the second performed best. For illustration we applied the CIR model to simulated data We used simulated data and apply to the CIR model. The results were compared to the estimations obtained using the true trnsitiond density.

A numerical comparison between the simulated, Vasicek model was performed using the Fed found data. Likelihood estimation proposed by Brandt and Santa-Clara (2002) and the Hermite expansion is made. The results were compared with the estimations obtained using the true transition density. They showed that also an Hermite expansion with $K = 1$ gave estimations more accurately. Better estimations are obtained with $K = 2$. The Hermite expansion seemed to be the best method in term of accuracy and speed.

Chapter 4

Stochastic Differential Mixed - Effects Models

In this chapter we present a class of models called Stochastic Differential Mixed Effect Models (SDMEM). In the context of biology, experimental studies often consist in repeated measurements of a biological criteria (drug concentration, viral concentration, etc) obtained from a population of subjects. Mixed effect models have the capacity to discriminate between the inter subjects variability by introduction of random parameters which vary among the individuals. This models are useful in pharmacokinetic/pharmacodynamic (PK/PD) models. PK is the study of time course of a drug and its metabolites following their introduction into the body. These study aim to provide an understanding of the pharmacokinetic using the estimation of population parameters, which is improved by introduction of mixed effect in the models.

Continuous biological process could be described by a system of ordinary differential equations (ODE), which do not consider the noise component often presents into biological system. A natural extension is given by systems of stochastic differential equations (SDE), where system noise is modeled by including a diffusion term of some suitable form in the driving equations. An

extension of SDE models are the SDMEM where the inter-individual variability is modeled with the random effect, and the intra-individual variability with an additive noise term. This approach which combine SDE and mixed effects is the results of recent research. As we have shown in the previous chapter, estimating parameters in SDE models is not simple to compute, except for a few cases. A natural approach would be likelihood inference, but the transition densities of the process are rarely known, and thus it is usually not possible to write the likelihood function explicitly.

The theory for mixed models is widely developed for deterministic models both linear and non linear (Lindstrom and Bates (1990), Breslow and Clayton (1993), Davidian and Giltinan (1995), Vonesh and Chinchilli (1997), McCulloch and Searle (2001), Diggle et al. (2002), Kuhn and Laville (2005), Guedj et al. (2007), Wang (2007)). In this context Ditlevsen and Gaetano (2005) proposed an estimation method adapted to linear mixed model defined by linear SDE, but their example is restricted to the case where the transition density has explicit expression. In Overgaard et al. (2007) and Tornøe et al. (2005) an SDE is introduced in non-linear mixed models, using an extended Kalman filter of the diffusion process, with linearization based estimation algorithm. The convergence of their algorithm is not proved. Donnet and Samson (2008) developed an estimation method based on a stochastic EM algorithm for fitting one-dimensional SDEs with mixed effects. In Donnet et al. (2010) a Bayesian approach is applied to a one-dimensional model for growth curve data.

In this chapter a computationally efficient method for estimating SDMEMs with random parameters following any sufficiently well-behaved continuous distribution is considered. See Picchini et al. (2010). First the conditional transition density of the diffusion process given the random effects is approximated in closed form by a Hermite expansion for time - inhomogeneous diffusion (Egorov et al. (2003)) , and then the conditional likelihood obtained is numerically integrated with respect to the random effects using Gaussian quadrature. The method turned out to be statistically accurate and computationally fast. However, in practice it was limited to one random effect only (Picchini et al. (2008)

for an application in neuroscience) since Gaussian quadrature is computationally inefficient when the number of random effects grows. Then we apply the method described using a pharmacokinetics model.

4.1 Formulation of Stochastic Differential Mixed - Effect Models

We consider a d-dimensional SDE model for some continuous process (X_t), involving M different experimental units randomly chosen from a theoretical population:

$$dX_t^i = \mu(X_t^i, t, \theta, b^i)dt + \sigma(X_t^i, t, \theta, b^i)dW_t^i \tag{4.1}$$

$$X_0^i = x_0^i, \quad i = 1, ..., M$$

where X_t^i is the value at time $t \geq t_0^i$ of the ith unit, with $X_0^i = X_{t_0^i}^i$; $\theta \in \Theta \subseteq \mathbb{R}^p$ is a p-dimensional fixed effect parameter (the same for the entire population), and $b^i \equiv b^i(\Psi) \in B \subset \mathbb{R}^q$ is a q-dimensional random effects parameter with components $(b_1^i, ..., b_q^i)$; each components may follow a different distribution. Let $p_B(b^i \mid \Psi)$ denote the joint distribution for b^i, parametrized by an r-dimensional parameter $\Psi \in \Upsilon \subset \mathbb{R}^r$. The W_t^i's are $d-$dimensional standard Brownian motions. Components of W_t^i and of b^i are assumed mutually independent. The initial condition X_0^i is assumed equal to a vector of constants $x_0^i \in \mathbb{R}^d$. The drift and the diffusion coefficient function $\mu(\cdot, t, \cdot, \cdot) : E \times \Theta \times B \to \mathbb{R}^d$ and $\sigma(\cdot, t, \cdot, \cdot) : E \times \Theta \times B \to \mathbb{S}$ are assumed known up to the parameters, and are assumed sufficiently regular to ensure a unique weak solution, where $E \in \mathbb{R}^d$ denotes the state space of X_t^i and $\mathbb{S}$ denote the set of $d \times d$ positive define matrices. The system of stochastic differential equations (4.1) describe the M different evolutions of the process X, we assume that the dynamics of X follow the same functional forms, and the differences are due to the Brownian motion and the introduction of a vector parameter randomly varying among units.
We assume that the distribution of X_t^i given (b^i, θ) and $X_s^i = x_s, s < t$, has a

strictly positive density w.r.t. the Lebesgue measure on E, which is denoted by

$$x \to p_X(x, t-s \mid X_s, b^i, \theta) > 0, \quad x \in E.$$

We assume moreover that unit i is observed at the same set of $n_i + 1$ discrete time points $\{t_0^i, t_1^i, ..., t_{n_i}^i\}$, for each coordinate of the process X_t^i. Let x^i be the $(n_i + 1) \times d$ matrix of responses for unit i, with the jth row given by $x^i(t_j^i) = (x_j^{(1)i}, ..., x_j^{(d)i})$, $N = \sum_{i=1}^{M}(n_i + 1)$. We write $\Delta_j^i = t_j^i - t_{j-1}^i$ for the time distance between x_{j-1}^i and x_j^i. Notice that this observation scheme implies that the matrix of data must not contain missing values.

The aim is to estimate (θ, Ψ) using simultaneously all the data in x. The specific value of the b^i's are not of interest, but only the identification of the vector parameter Ψ characterizing their distribution.

4.2 Maximum Likelihood Estimation

The marginal density of x^i is obtained by integrating the conditional density of the data given the non-observable random effect b^i with respect to the marginal density of the random effects, using that W_t^i and b^j are independent. This yelds the likelihood function:

$$\mathcal{L}(\theta, \Psi) = \prod_{i=1}^{M} p(x^i \mid \theta, \Psi) = \prod_{i=1}^{M} \int_B p_X(x^i \mid b^i, \theta) p_B(b^i \mid \Psi) db^i \tag{4.2}$$

where $p(\cdot)$, $p_X(\cdot)$ and $p_B(\cdot)$ are density functions. $p_X(x^i \mid \cdot)$ is the product of the transition densities for a given realization of the random effects and for a given θ :

$$p_X(x^i \mid b^i, \theta) = \prod_{j=1}^{n_i} p_X(x_j^i, \Delta_j^i \mid x_{j-1}^i, b^i, \theta). \tag{4.3}$$

The distribution of the random effects is often assumed to be (multi)normal, but $p_B(\cdot)$ could be any density function subject to mild regularity conditions. Solving the integral in (4.2) yields the marginal likelihood of the parameters, independent of the random effects b^i; by maximizing (4.2) with respect to θ and Ψ the corresponding maximum likelihood estimators (MLE) $\hat{\theta}$ and $\hat{\Psi}$ are obtained. Notice that it is possible to consider random effects having discrete

distributions: in that case the integral becomes a sum and can be easily computed when the transition density p_X is known. In simple cases the integral (4.2) can be solved, and explicit estimating equations for the MLE can be found. However, in general it is not possible to explicitly solve the integral, i.e. when:

- $p_X(x_j^i, \cdot \mid x_{j-1}^i, \cdot)$ is known but the integral cannot be solved analytically, the integral has to be numerically evaluated;
- $p_X(x_j^i, \cdot \mid x_{j-1}^i, \cdot)$ is unknown, we can approximate $p_X(x_j^i, \cdot \mid x_{j-1}^i, \cdot)$, then the integral is numerically solved.

In the second situation we propose to approximate the transition density in closed-form, using Hermite expansion suggest by Egorov et al. (2003).

4.2.1 Closed form transition density

In this section we try to approximate the transition density p_X in (4.2). According to Picchini et al. (2010), we consider an extension of the maximum estimation method of Aït-Sahalia (1999) described in the section 3.2. A closed-form approximation of likelihood function for discretely sampled time-inhomogeneous diffusions is then derived, following Egorov et al. (2003). While Aït-Sahalia (1999) considers only time-homogeneous diffusions, there are reasons to believe that the underlying data generating process for many biological and economic variables might change over time, the reason could be the changes in business cycles, monetary policy, and general macroeconomic conditions. One possible approach to capture the time-dependent behavior of asset prices given in the above examples is to model the drift and diffusion terms. In fact, time-inhomogeneous models of option pricing and term structure of interest rates have been developed in the finance literature. For example, to capture the "smiles" observed in the implied volatility from option prices, Rubinstein (1994), Derman and Kani (1994), and Dupire (1994) model stock return volatility as a deterministic function of stock price and time, and develop different techniques for pricing options on such assets. Black et al. (1990) and Black and Karasinski (1991) also develop

time-inhomogeneous term structure models.

Following Aït-Sahalia (1999) we built the approximation of the transition density p_X using the Hermite polynomials. Two transformations of the original process are needed before such an approximation can be obtained. The purpose of this transformation, as explained in Aït-Sahalia (1999), is to make transition density of the transformed process is close to a normal distribution, so that the standard Hermite expansion can be applied. The difference from Aït-Sahalia (1999) results is that we have to explicitly take into account the time-varying feature of the drift and diffusion coefficients of the process. Egorov et al. (2003) show that under certain regularity conditions, the method produces parameter estimates that converge to the true parameter values. We consider the model described by (4.1).

We need some assumptions.

Assumption 1 Functions $\mu(\cdot)$ and $\sigma(\cdot)$ are infinitely differentiable in $t \in [0, \infty)$ and X_t^i, and three times continuously differentiable in θ and b^i for all $X_t^i \in E$and $(\theta; b^i) \in \Theta \times B$.

Assumption 2 Let c be a positive constant such that $\sigma(X_t^i, t, \theta, b^i) > c > 0$ for all $X_t^i \in E$ and $(\theta; b^i) \in \Theta \times B$.

Weaker conditions on the diffusion coefficient close to the boundary of the state space can be considered, e.g. at 0 for positive diffusions so that also the Cox-Ingersoll- Ross model is covered; see Aït-Sahalia (2002b) for further details.

Two transformations of the original process X_t^i are needed before such an approximation can be obtained. The purpose of these two transformations, as explained in Aït-Sahalia (2002b), is to make the transition density of the transformed process close to a normal distribution, so that the standard Hermite expansion can be applied to such distributions.

For a generic SDE the first transformation of X_t standardizes the variance of the density so that it has unit variance. Using map:

$$Y_t \equiv \gamma(X_t) = \int^{X_t} \frac{du}{\sigma(u, t, \theta)},$$

where the lower bound of integration is arbitrary point interior of E, by Itô's lemma the result of the transformation Y_t is the solution of the SDE with unit

diffusion term and drift term given by:

$$\mu_Y(y,t,\theta) = \left[\frac{\partial\gamma}{\partial x}\mu(x,t,\theta) + \frac{\partial\gamma}{\partial t} + \frac{1}{2}\frac{\partial^2\gamma}{\partial x^2}\sigma^2(x,t,\theta)\right] = \tag{4.4}$$

$$= \frac{\mu(\gamma^{-1}(y,t,\theta),t,\theta)}{\sigma(\gamma^{-1}(y,t,\theta),t,\theta)} + \frac{\partial\gamma}{\partial t}(\gamma^{-1}(y,t,\theta),t,\theta) -$$

$$-\frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y,t,\theta),t,\theta).$$

The second transformation is a linear map that transforms Y_t into another process Z_t. It is defined by

$$Z_t \equiv \varphi(Y_t) = \frac{Y_t - y_s}{\sqrt{h}},$$

where h is the fixed sampling interval, let $t = s + h$. Let $p_Y(y,t \mid y_s,s,\theta)$ be the transition density of Y_t given $Y_s = y_s$, and $p_Z(z,t \mid y_s,s,\theta)$ be the transition density of Z_t given $Y_s = y_s$. The transition densities of p_X, p_Y and p_Z are related in the following ways:

$$p_Z(z,t \mid y_s,s,\theta) = \sqrt{h}p_Y(\sqrt{h}z + y_s \mid y_s,s,\theta),$$

$$p_Y(y,t \mid y_s,s,\theta) = \frac{1}{\sqrt{h}}p_Z\left(\frac{y - y_s}{\sqrt{h}} \mid y_s,s,\theta\right);$$

and

$$p_Y(y,t \mid y_s,s,\theta) = \sigma(\gamma^{-1}(y,t,\theta),t,\theta)p_X(\gamma^{-1}(y,t,\theta),t \mid \gamma^{-1}(y_s,s,\theta),s,\theta),$$

$$p_X(x,t \mid x_s,s,\theta) = \frac{1}{\sigma(x,t,\theta)}p_Y(\gamma(x,t,\theta),t \mid \gamma(x_s,s,\theta),s,\theta).$$

Thus, if the transition density p_Z or its approximation is known, then the approximation for p_X is obtained naturally. Next, we will show how to obtain such approximations. The Hermite expansion of transition density p_Z is:

$$p_Z(z,t \mid y_s,s,\theta) = \phi(z)\sum_{k=0}^{\infty}\beta_k(t,y_s,s,\theta)H_k(z), \tag{4.5}$$

where the coefficient β_k equal

$$\beta_k(t,y_s,s) = \frac{1}{k!}\int_{-\infty}^{+\infty}H_k(z)p_Z(z,t \mid y_s,s)dz, \tag{4.6}$$

and the Hermite polynomials H_k are easily computed using

$$H(w) = \phi(w)^{-1}\frac{d^k}{dw^k}\phi(w), \tag{4.7}$$

where ϕ denote the standard normal density. Let $p_Z^{(K)}$ denote the partial sum of integer order K of the Hermite expansion (4.5) of p_Z,

$$p_Z^{(K)} \equiv \phi(z) \sum_{k=0}^{K} \beta_k(t, y_s, s) H_k(z). \tag{4.8}$$

The corresponding approximations of p_Y and p_X are

$$p_Y^{(K)}(y, t \mid y_s, s, \theta) \equiv \frac{1}{\sqrt{h}} p_Z^{(K)} \left(\frac{y - y_s}{\sqrt{h}} \mid y_s, s, \theta \right),$$

$$p_X^{(K)}(x, t \mid x_s, s, \theta) \equiv \frac{1}{\sigma(x, t, \theta)} p_Y^{(K)}(\gamma(x, t, \theta), t \mid \gamma(x_s, s, \theta), s, \theta).$$

Using this approximation, we can write the transition density of X_t^i in the following way

$$p_X^K(x_j^i, \Delta_j^i \mid x_{j-1}^i, b^i, \theta) =$$

$$\frac{1}{\sigma(x_j^i, t_j, \theta, b^i)\sqrt{\Delta_j^i}} \phi \left(\frac{\gamma(x_j^i, t_j, \theta) - y_s}{\sqrt{\Delta_j^i}} \right) \times$$

$$\times \sum_{k=0}^{K} \beta_k(t_j, y_s, s, \theta, b^i) H_k \left(\frac{\gamma(x_j^i, t_j, \theta) - y_s}{\sqrt{\Delta_j^i}} \right).$$

Despite the fact that approximation (4.8) has nice theoretical properties, (Egorov et al. (2003)), its use will be quite limited if the approximation cannot be evaluated easily in practice. To carry out the approximation, we need a method to evaluate the coefficients β_k . Fortunately, β_k can be evaluated in a closed form with arbitrary precision.

Denote $\mathbb{E}_s[\cdot] = \mathbb{E}[\cdot \mid Y_s = y_s, s, \theta]$. Definition of β_k given in (4.6) and the properties of Hermite polynomials imply that

$$\beta_k(s + h, y_s, s, \theta) = \frac{1}{k!} \int_{-\infty}^{\infty} H_k(z) p_Z(z, s + h \mid y_s, s, \theta) dz = \frac{1}{k!} \mathbb{E}_s [H_k(z)].$$

Where the expectation can be evaluated using a variant of Taylor expansion given below.

Definition 4.1. Let U_t be a time - inhomogeneous diffusion in $\mathbb{R}$. The infinitesimal generator $\mathfrak{L}$ of U_t is defined by

$$(\mathfrak{L} \circ \chi)(u, t) = \lim_{\tau \searrow 0} \frac{\mathbb{E}[\chi(U_{t+\tau}, t + \tau) \mid U_t = u] - \chi(u, t)}{\tau}, \quad u \in \mathbb{R}.$$

The set of the functions $\chi : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ such that the limit exists at all $(u,t) \in \mathbb{R} \times [0,\infty)$ is denoted by $\mathcal{D}(\mathfrak{L})$ and called the domain of infinitesimal generator $\mathfrak{L}$.
Denote $\mathcal{D}(\mathfrak{L}^i)$ the domain of operator $\mathfrak{L}^i = \mathfrak{L} \circ \mathfrak{L} \circ ... \circ \mathfrak{L}$ (i times).

Preposition 4.2.1. *(A Variant of Taylor Expansion) Let $\mathfrak{A}_{\theta,\tilde{y},h}$ be the infinitesimal generator of the process Z_t for any fixed $(\theta,\tilde{y},h) \in \Theta \times D_Y$ $(0,\infty)$. Let $f(z) \in C_0^\infty(\mathbb{R})$. then for any $i = 1,2,...,$ $f \in \mathcal{D}(\mathfrak{A}^i_{\theta,\tilde{y},h})$ and for all $(z,t) \in D_Z \times [0,\infty)$,*

$$(\mathfrak{A}^i_{\theta,\tilde{y},h} \circ f)(z,t) = \frac{\partial(\mathfrak{A}^{i-1}_{\theta,\tilde{y},h} \circ f)}{\partial z}\mu_Z + \frac{\sigma_Z^2}{2}\frac{\partial^2(\mathfrak{A}^{i-1}_{\theta,\tilde{y},h} \circ f)}{\partial z^2} + \frac{\partial(\mathfrak{A}^{i-1}_{\theta,\tilde{y},h} \circ f)}{\partial t}, \quad (4.9)$$

where $\mu_Z(z,t;h,y_s,\theta) \equiv \frac{\mu_Y(\sqrt{h}z+\tilde{y},t;\theta)}{\sqrt{h}}$ and $\sigma_Z^2 \equiv \frac{1}{h}$.

Note that $H_k \notin C_0^\infty(\mathbb{R})$. Let $\{H_{k,j}(z) \equiv \frac{1}{2}e^j(coshj + coshz)^{-1}H_k(z), z \in \mathbb{R}\}_{k,j=0}^\infty$. Since $H_{k,j} \in C_0^\infty(\mathbb{R})$, Preposition 4.2.1 applies to $H_{k,j}$. Moreover, as $j \to \infty$, $H_{k,j}(z) \to H_k(z)$ uniformly on any compact subset of $\mathbb{R}$. The same is true for any derivative of $H_{k,j}$. Then taking large j, we get the approximation

$$\beta_k(s+h,y_s,s,\theta) \approx \frac{1}{k!}\mathbb{E}_s\left[H_{k,j}(Z_{s+h})\right] \approx \frac{1}{k!}\sum_{i=0}^{I}(\mathfrak{A}^i_{\theta,y_s,h} \circ H_{k,j})(0,s;h,y_s,\theta)\frac{h^i}{i!} \approx$$

$$\approx \frac{1}{k!}\sum_{i=0}^{I}(\mathfrak{A}^i_{\theta,y_s,h} \circ H_k)(0,s;h,y_s,\theta)\frac{h^i}{i!}, \quad (4.10)$$

The coefficients β_k for a PK model are given in the Section 4.3.1.

4.2.2 A Random Effect Following a continuous distribution

In the last section we have discussed about the approximation of the transition density p_X in (4.2). In this section we try to compute the integral (4.2), using a numerical integration. Following Picchini et al. (2010) we consider the general case of a random effect b^i having density p_B (not necessarily Gaussian), with certain conditions on existence of moments. In Golub and Welsch (1969) a Gaussian quadrature integration method for any non-negative measure is suggested: in particular, Fernandes and Atchley (2006) report explicit formulae for the cases of Normal, Gamma, log-Normal, Student's t, inverse Gamma, Beta

and Fisher's F distributions, covering a large class of problems commonly encountered in e.g. biomathematics/biostatistics. Consider the following integral

$$\int_B h(u)\omega(u)du$$

where $h(\cdot) \in \mathcal{C}^{2R}(B)$ for some chosen positive integer R and $\omega(\cdot)$ is a density function with support B fulfilling

$$\mathbb{E}(U^{2R}) < \infty \tag{4.11}$$

for $U \sim \omega(u)$. Then

$$\int_B h(u)\omega(u)du \simeq \sum_{r=1}^{R} h(z_r)\omega_r$$

using R evaluation points z_r (nodes) and weights ω_r, with approximation error E_R given by

$$E_R = \frac{1}{(2R)!}\frac{d^{2R}}{du^{2R}}h(u) \mid_{u=c} \int_B \omega(y)[\pi(y)]^2 dy$$

for some $c \in B$, where $\pi(y) = \prod_{r=1}^{R}(y - z_r)$. The last integral is finite under (4.11) and $E_R \to 0$ when $R \to \infty$ if B is bounded. The z_r's are the eigenvalues of a tridiagonal matrix J, define by:

$$J = \begin{pmatrix} \alpha_0 & \sqrt{\beta_1} & & 0 \\ \sqrt{\beta_1} & \ddots & \ddots & \\ & \ddots & \ddots & \sqrt{\beta_{R-1}} \\ 0 & & \sqrt{\beta_{R-1}} & \alpha_{R-1} \end{pmatrix}.$$

where the α_r's and β_r's are specific to the distribution $\omega(\cdot)$, and $\omega_r = q_{r,1}^2$, where $q_{r,1}$ is the first component of normalized eigenvector q_r of J. In Fernandes and Atchley (2006) the α_r's and β_r's are explicitly given for some important distributions $\omega(\cdot)$. If $\omega(\cdot) \equiv N(\mu, \sigma^2)$, than $\alpha_r = \mu$ and $\beta_r = r \cdot \sigma^2$ for all $r = 1, ..., R-1$. The approximation is exact whenever h is a polynomial of degree $2R-1$ or less. It follows how we can apply this numerical method for a one- dimensional integral to solve the integration problem in (4.2).

We consider a one-dimensional ($q = 1$) random effect b^i, define $\omega(b^i) = p_B(b^i \mid \Psi)$ and

$$h_K^i(b^i) = \prod_{j=1}^{n_i} p_X^{(K)}(x_j^i, \Delta_j^i \mid x_{j-1}^i, b^i, \theta).$$

Assuming that

$$h_K^i(b^i) \in \mathcal{C}^{2R}(B) \quad and \quad \mathbb{E}(b^{i^{2R}}) < \infty$$

the likelihood in (4.2) is approximated by

$$\hat{\mathcal{L}}^{K,R}(\theta, \Psi) = \prod_{i=1}^{M} \sum_{r=1}^{R} h_K^i(z_r)\omega_r, \tag{4.12}$$

and $(\hat{\theta}^{K,R}, \hat{\Psi}^{K,R}) = argmin_{\theta,\Psi}(-\ln \hat{\mathcal{L}}^{K,L})$ is an approximated MLE of (θ, Ψ). For some applications choosing using $K = 2$ and $R = 40$ seemd sufficiently accurate.

4.3 Theophillin Pharmacokinetic Example

In this section we use the estimation method developed in the previous section to a pharmacokinetic example proposed by Donnet and Samson (2008).

Pharmacokinetics (PK) studies the time course of drug substances in the organism. This can be described through dynamic systems, the human body being assimilated to a set of compartments within which the drug evolves with time. In general, these systems are considered in their deterministic version. However Krishna (2004) claims that the fluctuations around the theoretical pharmacokinetic dynamic model may be appropriately modeled by using SDEs rather than ODEs. Overgaard et al. (2005) suggest the introduction of SDEs to consider serial correlated residual errors due for example to erroneous dosing, sampling history or structural model misspecification. In the PK context, non-linear mixed-effects models are classically considered with a Gaussian distribution for the individual parameters: $b^i \sim N(\mu, \eta^2)$ for $i = 1, ..., M$. In this case, the parameter Ψ to estimate is $\Psi = (\mu, \eta^2)$. In the following, the hypothesis $t_j^i = t_j$ for all i, is not assumed and the observation times t_j^i may differ between subjects. We consider a classic one compartment PK model. The body acts as if it is a series of compartments. In many cases, the drug distributes from the blood into the tissues quickly, and a pseudo-equilibrium of drug movement between blood and tissues is established rapidly. When this occurs, a one-compartment model can be used to describe the serum concentrations of the drug. In particular we

consider a PK model with first order absorption and first order elimination, it is described by the following dynamic equation:

$$\begin{cases} \frac{dX_t}{dt} = \frac{(Dose \cdot K_a \cdot K_e)}{Cl} e^{-K_a t} - K_e X_t; \\ X_0 = 0. \end{cases} \tag{4.13}$$

X represents the drug concentration in blood, $Dose$ is the know drug oral dose received by subject, K_e is the elimination rate constant, K_a is the absorption rate constant and Cl is the clearance of the drug. See Figure 4.1. Drug Clearance is define as the volume of plasma in the vascular compartment cleared of drug per unit time by the processes of metabolism end excretion. Clearance for drug is constant if the drug is eliminated by first-order kinetics. Drug can be cleared by renal excretion or by metabolism or both. Mathematically, clearance is the product of the first-order elimination rate constant, and the apparent volume of distribution (V_d), $Cl = K_e \times V_d$. The volume of distribution has no direct physiological meaning; it is not a real volume. It is defined as that volume of plasma in which the total amount of drug in the body would be require to be dissolved in order to reflect the drug concentration attained in plasma. Populations PK studies consider the pharmacokinetics of a number

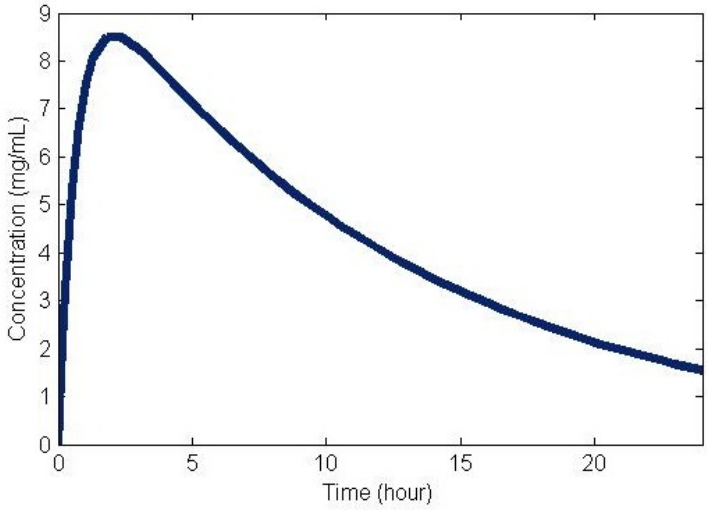


Figure 4.1: Concentration of drug versus time obtained as the solution of the ODE (4.13), using $K_e = 0.08$, $K_a = 1.49$, $Cl = 0.04$ and $Dose = 5mg$.

of individuals. The data from such studies typically consist of dose histories,

drug concentrations with associated sampling times, and often covariate measurements such as the age and weight of each subject. PK models are generally nonlinear functions of a set of PK parameters. Consequently we expect each individual to have their own set of PK parameters. Therefore we can extend the ODE (4.13) using a SDE which can consider the noise, and introduce a mixed effects to model random differences among individuals. A stochastic differential system can be deduced:

$$dX_t(b^i) = \left(\frac{Dose \cdot K_e \cdot K_a}{Cl} e^{-K_a t} - K_e X_t(b^i) \right) dt + \sigma dW_t^i \tag{4.14}$$

where W_t^i is a Brownian motion $\forall i = 1, ..., M$; σ is the volatility coefficient of the SDE. This SDE is linear and the law of the diffusion X is analytically known. However, this diffusion is nonlinear with respect to the individual parameter. Consequently, the likelihood of the corresponding non-linear mixed model has no analytical form.

4.3.1 The Parameter Estimation Methodology

In this section we use the PK model described in the Section 4.3 to mimic the Theophyllin drug pharmacokinetic to test the algorithm developed in Section 4.2 .

We consider a one-dimensional random parameter b^i. Following Pinheiro and Bates (1995), we make our estimations using first Cl^i and then K_a^i ($\ln K_a^i = \ln K_a + b^i$) as the random parameter ($\ln Cl^i = \ln Cl + b^i$). Pinheiro and Bates (1995) observe that analysis of Theophylline data, using (4.13) indicated that only Cl and K_a needed random effects to account for the variability among patients. In each case b^i follows a Gaussian distribution $N(0, \eta^2)$. Since $e^{\rho + b^i} \sim LN(\rho, \eta^2)$ according to the nature of the parameters which are positive. We have $\theta = (K_a, K_e, \rho, \sigma)$ and $\Psi = \eta^2$, where $\rho = \ln Cl$ first and then $\rho = \ln K_a$. We consider two cases:

1. Cl^i is the random parameter and $\ln Cl = \rho$. The SDE (4.14) become:

$$dX_t^i = \left(\frac{Dose \cdot K_e \cdot K_a}{e^{\rho + b^i}} e^{-K_a t} - K_e X_t^i \right) dt + \sigma dW_t^i.$$

Then

$$\mu_Y = \frac{Dose \cdot K_a \cdot K_e}{e^{\rho+b^i}\sigma} e^{-K_a t} - K_e Y_t^i.$$

2. K_a^i is random parameter, and $\ln K_a = \rho$. The SDE (4.14) become:

$$dX_t^i = \left(\frac{Dose \cdot K_e \cdot e^{\rho+b^i}}{Cl} e^{-e^{\rho+b^i} t} - K_e X_t^i \right) dt + \sigma dW_t^i.$$

Then

$$\mu_Y = \frac{Dose \cdot e^{\rho+b^i} \cdot K_e}{Cl\sigma} e^{-e^{\rho+b^i} t} - K_e Y_t^i.$$

As ordinarily observed in this context, the concentration profiles have a similar shape for all subjects; however, peak concentration achieved, rise, and decay vary substantially. See Figure 4.2. These differences are believed to be attributable to inter-subject variation in the underlying pharmacokinetic processes, understanding of which is critical for developing dosing guidelines. To compute the estimation of θ and Ψ, we maximize the approximation of the log-likelihood.

In the two cases we have supposed that b^i is a one-dimensional random parameter normally distributed with mean zero and variance equal to η^2. Our aim is solve the follow integral:

$$\mathcal{L}(\theta, \Psi) = \prod_{i=1}^{M} \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} p_X(x_j^i, \Delta_j^i \mid x_{j-1}^i, b^i, \theta) \times \frac{1}{\sqrt{2\pi\eta^2}} e^{\frac{-b^{i2}}{2\eta^2}} db^i \qquad (4.15)$$

If we define $u^i = \frac{b^i}{\sqrt{2}\eta}$ (4.15) becomes

$$\mathcal{L}(\theta, \Psi) = \prod_{i=1}^{M} \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} p_X(x_j^i, \Delta_j^i \mid x_{j-1}^i, u^i\sqrt{2}\eta, \theta) \frac{e^{-u^{i2}}}{\sqrt{\pi}} du^i =$$

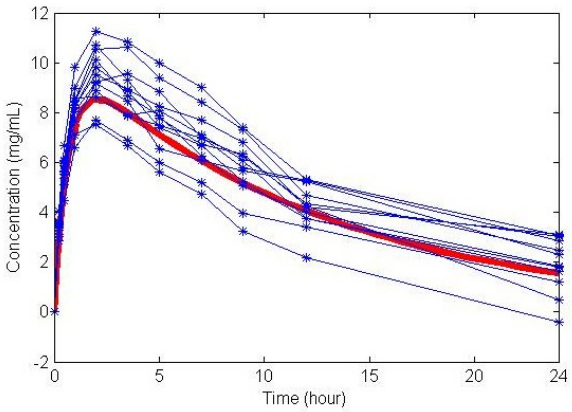
$$= \prod_{i=1}^{M} \int_{-\infty}^{+\infty} h^i(u^i) e^{-u^{i2}} du^i, \qquad (4.16)$$

where

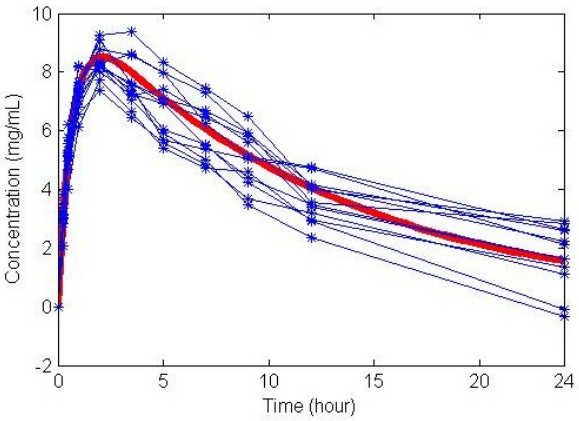
$$h^i(u^i) = \prod_{j=1}^{n_i} \frac{p_X(x_j^i, \Delta_j^i \mid x_{j-1}^i, u^i\sqrt{2}\eta, \theta)}{\sqrt{\pi}}.$$

Integral into (4.16) can be solved using Gaussian Hermite quadrature (see Fröberg (1985)), which is Gaussian quadrature formula approximating (4.16) as:

$$\int_{-\infty}^{+\infty} h_K^i(u^i) e^{-u^{i2}} du^i \simeq \sum_{r=1}^{R} h_K^i(z_r)\omega_r$$



(a) Trajectories referred to the first case (Cl^i random variable), using $K_e = 0.08$, $K_a = 1.49$, $\rho = -3.22$, $\sigma = 0.447$, $\eta^2 = 0.01$ and $Dose = 5mg$.



(b) Trajectories referred to the second case (K_a^i random variable), using $K_e = 0.08$, $Cl = 0.04, \rho = 0.4$, $\sigma = 0.447$, $\eta^2 = 0.01$ and $Dose = 5mg$.

Figure 4.2: simulated individual concentrations of the drug for 12 subjects. We used the Euler-Maruyama scheme for the simulation. The red line represents the solution of the ODE (4.13) using the same parameters.

where R is a positive integer, the "nodes" z_r and the "weights" ω_r are defined by

$$z_r = rth \quad zero \quad of \quad H_R(u),$$
$$\omega_r = \frac{2^{R-1}R!\sqrt{\pi}}{R^2\left[H_{R-1}(z_r)\right]^2},$$

with an approximation error

$$E_R = \frac{R!\sqrt{\pi}}{2^R(2R)!}\frac{d^{2R}}{du^{2R}}h(u)\mid_{u=c}$$

for some $c \in \mathbb{R}$. $H_R(\cdot)$ is the Hermite polynomial of degree R.

Then for the approximation of p_X we follow Aït-Sahalia (2002b) and Jensen and Poulsen (2000) in taking approximation $\beta_k^{[m]}$ for β_k as follow: for any non-negative m, we take $I = 2m$ and leave in (4.10) only terms up to h^m; $\beta_k^{[m]} = 0$ for all $k > 2m$. Since the $\mathfrak{A}^i \circ H_k$ can be compute iteratively using (4.9), approximation $\beta_k^{[m]}$ can be obtained in mechanical fashion. In particular, choosing $m = 3$ and omitting the null derivatives, for our model we obtain

$$\beta_1^{[3]} = -h^{\frac{1}{2}}\zeta - \frac{1}{4}h^{\frac{3}{2}}(2\zeta_{0,1} + 2\zeta\zeta_{1,0}) - \frac{1}{24}h^{\frac{5}{2}}(4\zeta_{0,2} + 4\zeta_{0,1}\zeta_{1,0} + 4\zeta\zeta_{1,0}^2),$$

$$\beta_2^{[3]} = \frac{1}{2}h^{\frac{1}{2}}h(\zeta^2 + \zeta_{1,0}) + \frac{1}{12}h^2(6\zeta\zeta_{0,1} + 6\zeta^2\zeta_{1,0} + \zeta_{1,0}^2) + \frac{1}{96}h^3(12\zeta_{0,1}^2 +$$
$$+16\zeta\zeta_{0,2} + 40\zeta\zeta_{0,1}\zeta_{1,0} + 28\zeta^2\zeta_{1,0}^2 + 16\zeta_{1,0}^3),$$

$$\beta_3^{[3]} = -\frac{1}{6}h^{\frac{3}{2}}(\zeta^3 + 3\zeta\zeta_{0,1}) - \frac{1}{48}h^{\frac{5}{2}}(12\zeta^2\zeta_{0,1} + 12\zeta^3\zeta_{1,0} + 12\zeta_{0,1}\zeta_{1,0} + 28\zeta\zeta_{1,0}^2),$$

$$\beta_4^{[3]} = \frac{1}{24}h^2(\zeta^4 + 6\zeta^2\zeta_{1,0} + 3\zeta_{1,0}^2) + \frac{1}{240}h^3(20\zeta^3\zeta_{0,1} + 20\zeta^4\zeta_{1,0} +$$
$$+60\zeta\zeta_{0,1}\zeta_{1,0} + 100\zeta^2\zeta_{1,0}^2 + 40\zeta_{1,0}^3),$$

$$\beta_5^{[3]} = -\frac{1}{120}h^{\frac{5}{2}}(\zeta^5 + 10\zeta^3\zeta_{1,0} + 15\zeta\zeta_{1,0}^2),$$

$$\beta_6^{[3]} = \frac{1}{720}h^3(\zeta^6 + 15\zeta^4\zeta_{1,0} + 15\zeta_{1,0} + 45\zeta^2\zeta_{1,0}^2),$$

$$\beta_0^{[3]} = 1; \quad \beta_k^{[3]} = 0 \quad k > 6;$$

where $\zeta \equiv \mu_Y(y_s, s)$ and $\zeta_{i,j} \equiv \frac{\partial^{i+j}\mu_Y(y,s)}{\partial y^i \partial s^j}\mid_{y=y_s}$ for all i and j. Since for our model the only non null derivatives are:

- $\zeta_{1,0} = -K_e$;

- $\zeta_{0,1} = -\frac{Dose \cdot K_e \cdot K_a^2}{Cl\sigma} e^{-K_a t}$;
- $\zeta_{0,2} = \frac{Dose \cdot K_e \cdot K_a^3}{Cl\sigma} e^{-K_a t}$.

The Hermite polynomials are computed using their definition (4.7). In particular the first seven Hermite polynomials are $H_0(z) = 1$, $H_1(z) = -z$, $H_2(z) = z^2 - 1$, $H_3(z) = -z^3 + 3z$, $H_4(z) = z^4 - 6z + 3$, $H_5(z) = -z^5 + 10z^3 - 15z$, $H_6(z) = z^6 + 15z^4 + 45z^2 - 15$.

4.3.2 A real application

We use data from a study by Dr. Robert Upton of the kinetics of the anti-asthmatic drug theophylline. Data can be obtained from the R "datasets" package by invoking the "Theoph" dataset. Twelve subjects were given oral doses of theophylline then serum concentrations were measured at 11 time points over the next 25 hours, see Figure 4.3.

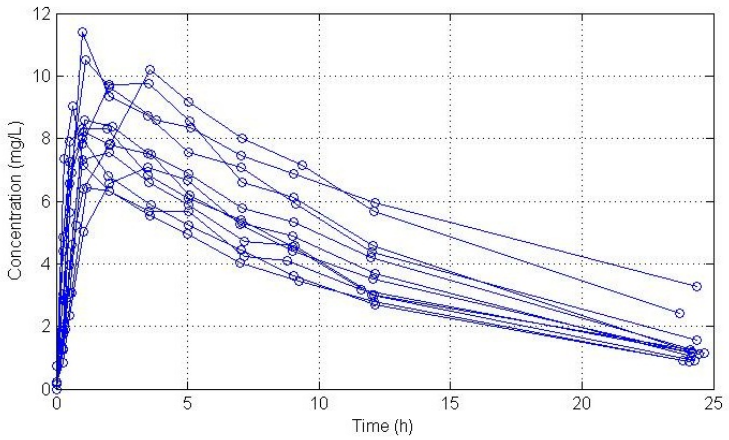


Figure 4.3: Individual concentrations for the pharmacokinetics of theophylline for 12 subjects.

The drug oral dose ($Dose$) received by the subjects is between 3 and 6 mg. We consider a Gaussian - Hermite integration approach with $R = 110$ (Pinheiro and Bates (1995) suggest $R > 100$) . In each case the likelihood was approximate

using (4.12), using $K = 6$ according to the coefficients given above. We use the estimate parameters of Pinheiro and Bates (1995), as staring values of the maximization, they use a ODE model for their estimations, for this reason the starting value of σ is taken from Donnet and Samson (2008). Estimation results are shown in the tables 4.1 and 4.2. To understand the results we compare the simulations obtained using the estimate parameters with the data. In the each cases, the concentration profiles have a similar shape. See Figure 4.4. Then we use the parametric bootstrap with 100 iteration, to obtain 95% confidence intervals of the parameter estimate.

In each cases we observe a growth of σ, it was predictable, indeed our model does not consider the measurement error.

We propose an hypothesis test for the variance η^2. The Hypothesis $H_0 : \eta = 0$ is tested against $H_1 : \eta > 0$. We denote $\hat{\theta}$ the estimate of all the parameters and $\hat{\theta}_0$ the estimate of all the parameters under the restriction that $\eta = 0$. The likelihood ratio statistic Λ is

$$\Lambda = \frac{\mathcal{L}(\hat{\theta}_0, \eta = 0)}{\mathcal{L}(\hat{\theta})},$$

where $\mathcal{L}$ is given by (4.12). The large sample distribution of $-2\log\Lambda$ under the null hypothesis and some some mild regularity conditions tend to a χ^2_1 distribution. If we consider the critical value $\alpha = 0.05$ we have

$$P(\Lambda \leq c \mid H_0) = \alpha$$

so the critical region is defined by

$$C = \{X \mid \Lambda \leq c\} \approx \{X \mid \log\Lambda \leq -\frac{d}{2}\},$$

where $d = 3.841$ is given by $\chi^2_{1,1-\alpha}$.
We obtain:

- $\log\Lambda = -1450$ using Cl as random parameters;
- $\log\Lambda = -871.7481$ using K_a as random parameters;

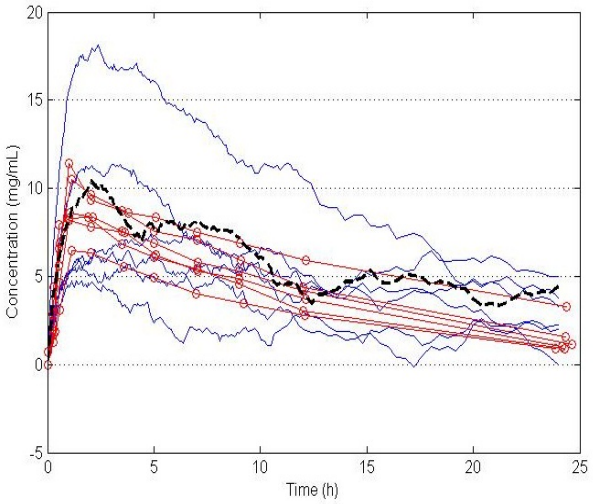
In each case the hypothesis H_0 is rejected, since is a good assumption consider Cl or K_a as random parameters.

	Estimate values
K_e	0.0597 ([0.0582, 0.3])
K_a	1.5987 ([0.7, 1.6697])
σ	0.5387 ([0.3752, 0.799])
η^2	0.8908 ([0.8376, 0.899])
μ	-3.822 ([−3.99, −2.8855])

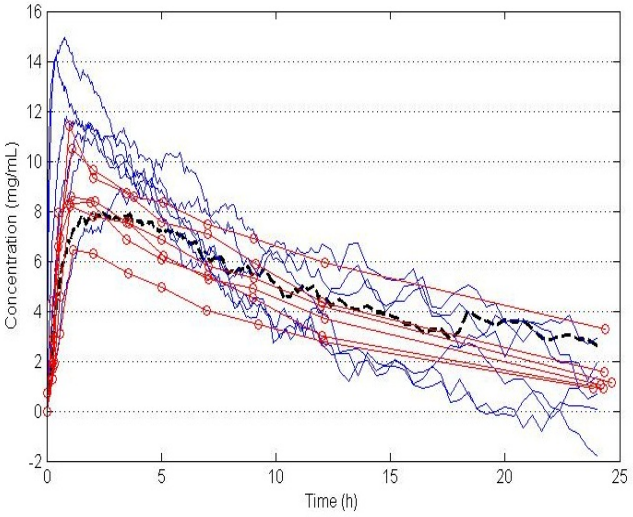
Table 4.1: Parameters estimate (95% confidence intervals) obtain maximizing the log-likelihood of equation (4.12).We consider the SDMEM whose random effect is the parameter Cl. The starting values for the maximization provided by Pinheiro and Bates (1995) and Donnet and Samson (2008). We start from the parameter vector [0.08; 1.8; 0.45; 0.03;-3.22].

	Estimate values
K_e	0.098 ([0.084, 0.1033])
Cl	0.0299 ([0.0299, 0.0302])
σ	0.699 ([0.6982, 0.7])
η^2	0.6226 ([0.6037, 0.8419])
μ	0.2715 ([0.1351, 0.4179])

Table 4.2: Parameters estimate (95% confidence intervals) obtain maximizing the log-likelihood of equation (4.12). We consider the SDMEM whose random effect is the parameter K_a. The starting values for the maximization provided by Pinheiro and Bates (1995) and Donnet and Samson (2008). We start from the parameter vector [0.08; 0.04; 0.45; 0.4;0.5]



(a) Trajectories referred to the SDMEM whose random effect is the parameter Cl.



(b) Trajectories referred to the SDMEM whose random effect is the parameter K_a.

Figure 4.4: Individual concentration of the drug for different subjects.In each figures the blu lines represent the solution of the simulation of the SDMEM using the Eulero-Maruyama scheme and the estimate parameters; the red lines show the real, the black dashed line represents the simulation using estimate parameters of Pinheiro and Bates (1995) and Donnet and Samson (2008) (4.14) without a random effect.

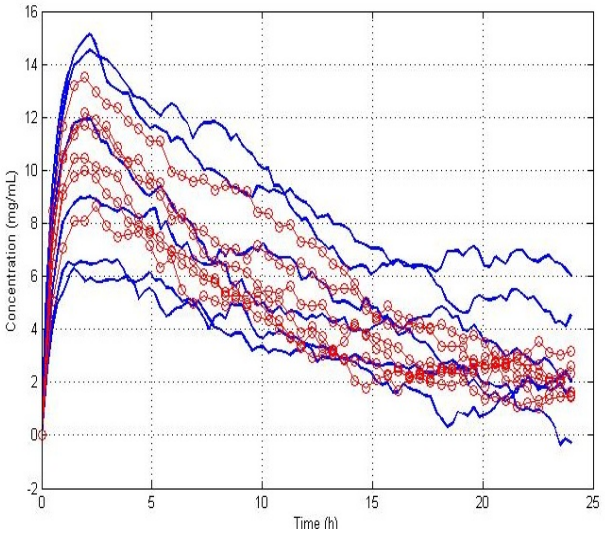
4.3.3 Simulation Study

In this section we try to estimate the parameter using simulated data. Reducing the time points distance end inceasing the number of subjects, we expect that the estimation using simulated data will perform better. We try to improve our results considering thirty-six subjects were given oral doses, $Dose = 5mg$, of theophylline then serum concentrations were measured at 50 equidistant time points over the next 25 hours. Under the setup specified above (M=36 and n=50 for each subject) simulated data are generated using the estimates obtained by Pinheiro and Bates (1995) and Donnet and Samson (2008) for σ ($[0.08; 1.6; 0.45; 0.03; -3.22]$ and $|\,[0.08; 0.04; 0.45; 0.4; 0.5]$), and the estimation is computed using these data. To understand the results we compare the simulations obtained using our estimate parameters with others obtained using the estimated parameters of Pinheiro and Bates (1995) and Donnet and Samson (2008) for σ, see Figure 4.5. Then we use the parametric bootstrap with 100 iteration, to obtain 95% confidence intervals of the parameter estimate.

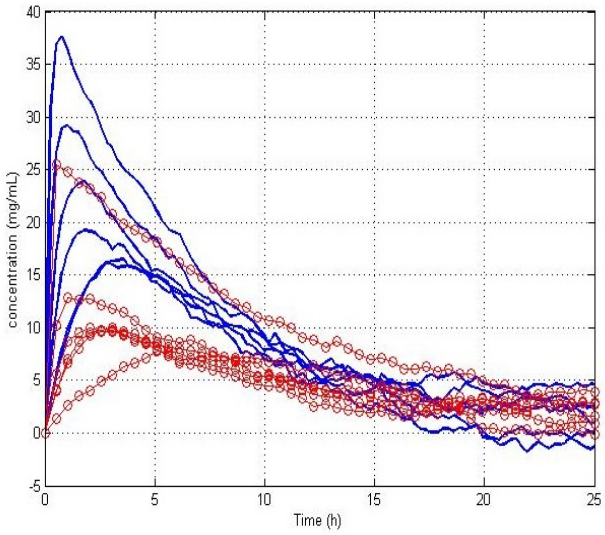
The results are shown in Table 4.3. Figure 4.5 show that the trajectories obtained using estimated parameters fit the data simulated, since the growt of the number of the subjects and the growt of the time point could lend to a better results.

Cl random parameters	K_a random parameters
$\hat{K_e} = 0.07$ ($[0.03608, 0.4809]$))	$\hat{K_e} = 0.1327$ ($[0.777, 0.1096]$)
$\hat{K_a} = 1.6333$ ($[1.448, 1.8038]$)	$\hat{Cl} = 0.0299$ ($[0.0289, 0.0312]$)
$\hat{\sigma} = 0.4517$ ($[0.36080.4809]$)	$\hat{\sigma} = 0.699$ ($[0.6104, 0.7]$)
$\hat{\eta}^2 = 0.4732$ ($[0.42, 0.99]$)	$\hat{\eta}^2 = 0.7437$ ($[0.7128, 0.99]$)
$\hat{\mu} = -3.1295$ ($[-3.99, -1.0156]$)	$\hat{\mu} = 0.3729$ ($[0.1055, 0.5928]$)

Table 4.3: Parameters (95% confidence intervals) obtained simulated data used for $M = 36$ subjects. We start the maximization from the parameter vectors [0.08; 1.6; 0.45; 0.03 ;-3.22] for the model with Cl as random parameter, and we start from parameter vector [0.08; 0.04; 0.45; 0.4;0.5] for the model with K_a as random parameter, the same parameters vectors are used to generated the simulated data.



(a) Trajectories referred to the SDMEM whose random effect is the parameter Cl.



(b) Trajectories referred to the SDMEM whose random effect is the parameter K_a.

Figure 4.5: Individual concentration of the drug for different subjects. In each figures the blu lines represent the solution of the simulation of the SDMEM using the Eulero-Maruyama scheme and the estimate parameters; the red lines show the simulated data.

4.4 Conclusions

The chapter had shown the usefulness of stochastic differential mixed-effects model. We proposed an approximated maximum likelihood estimation method

for the parameters of mixed-effects models defined by stochastic differential equations. We constructed a sequence of approximations of the transition densities using the Hermite expansion for time inhomogeneous diffusion process. We described the Gaussian quadrature scheme to compute the integral. We focused on a PK model and we applied the approximation technique described to the estimation of parameters using real data and simulated data. According to Pinheiro and Bates (1995) we assumed before the clearance and then the absorption rate as random parameter, our choice was supported by an hypothesis test. Satisfactory result were obtained in both cases using $R = 110$ and $K = 6$. We observed an improved of the estimations using simulated data, in fact the simulation have been done increasing the number of subjects and the time points, so reducing the time - distance between the data. In conclusion, we propose a parameter estimation method for SDE including random effects which al least seem to be able to estimate good parameters for the PK model considered.

Conclusions

The work proposed an introduction of some basic definitions about stochastic calculus. In particular we underlined the importance of the stochastic differential equations which are able to model time evolution of dynamic phenomena in many fields. We focused on the problem of parameters estimations which characterize each diffusion process. Among the estimations techniques linked to the maximization of the likelihood function we studied the simulated maximum likelihood proposed by Brandt and Santa-Clara (2002), its extensions proposed by Durham and Gallant (2002) and a closed form approximation using the Hermite expansion by Aït-Sahalia (1999). The results showed that the simulated maximum likelihood proposed by Brandt and Santa-Clara (2002) gave bad results compared to the simulated maximum likelihood using the two particular samplers proposed by Durham and Gallant (2002). An Hermite expansion stopped to the order $K = 2$ was sufficient to have better estimations then Brandt Santa-Clara method. Furthermore Hermite expansion was the fasted method, we decided to use it ($K = 6$) for the approximation densities in the stochastic differential mixed - effect model proposed in the second chapter.

Stochastic differential mixed - effect models, characterize by introduction of a random parameters in a diffusion process, are able to model the variations within-group and between-group. Our estimation study was made using a PK model. The results were satisfactory, the trajectories obtained using the estimated parameters fit in a good way the real data, despite our model do not consider the measurement error. The results were improved simulating the data, increasing the number of subjects and decreasing the distance between

time points. Concluding the Hermite approximations had a good performance at list in the stochastic differential mixed - effect model that we considered.

Bibliography

T. Aït-Sahalia. Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64:527–560, 1996a.

Y. Aït-Sahalia. Transition densities for interest rate and other non linear diffusions. *The journal of finance*, 4, 1999.

Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70:223–262, 2002b.

F. Black and P. Karasinski. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47:52–59, 1991.

F. Black, E. Derman, and W. Toy. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, 46:33–39, 1990.

M.W. Brandt and P. Santa-Clara. Simulated likelihood espimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 63:161–210, 2002.

C.A. Braumann. Population growth in random environments. *Bulletin of Mathematical Biology*, 4:635–641, 1983.

N. Breslow and D. Clayton. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.*, 88(421):9–25, 1993.

V. Capasso and D. Bakstein. *An Introduction to Continuous-Time Stochastic Processes*. Birkhäuser, 2005.

M. Davidian and D. Giltinan. Nonlinear models for repeated measurement data. *Chapman and Hall*, 1995.

E. Derman and I. Kani. *The Volatility Smile and its Implied Tree, Quantitative Strategies Research Notes*. Goldman Sachs, New York, 1994.

P. Diggle, P. Heagerty, K. Liang, and S. Zeger. *Analysis of longitudinal data*. Oxford University Press, 2002.

S. Ditlevsen and A. De Gaetano. Mixed effects in stochastic differential equations models. *REVSTAT Statistical Journal*, 3(2):137–153, 2005.

S. Donnet and A. Samson. Parametric inference for mixed models defined by stochastic differential equations. *ESAIM: probability and statisctics*, 12:196–218, 2008.

S. Donnet, J. Foulley, and A. Samson. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, 63(3): 733–741, 2010.

B. Dupire. Pricing with a smile. *Risk*, 7:18–20, 1994.

G. Durham. Likelihood-based speci?cation analysis of models of the short-term interest rate. *Unpublished workingpaper, University of North Carolina*, 2000.

G.B. Durham and A.R. Gallant. Numerical techniques for maximum likelihood espimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics*, 2002.

A.V. Egorov, H. Li, and Y. Xu. Maximum likelihood estimation of time-inhomogeneous diffusion. *Journal of Econometrics*, 114:107–139, 2003.

O. Elerian. A note on the existence of a closed form conditional transition density for the milstein scheme. *Working Paper, Nuśeld College, Oxford University*, 1998.

O. Elerian, S. Chib, and N. Shephard. A note on the existence of a closed form conditional transition density for the milstein scheme. *Econometrica*, 69:959–993, 2001.

W. Feller. The parabolic differential equations and the associated semi - grups of transformations. *Annals of Mathematics*, 55:468–519, 1952.

A.D. Fernandes and W.R. Atchley. Gaussian quadrature formulae for arbitrary positive measures. *Evolutionary Bioinformatics online*, 2006.

C. Fröberg. *Numerical mathematics.* The Benjamin/Cummings Publishing Company, 1985.

A.R. Gallant and G. Tauchen. Which moments to match? *Econometric Theory*, 12:657–681, 1996.

G. Golub and J. Welsch. Calculation of gauss quadrature rules. *Math. Comp.*, 23:221–230, 1969.

J. Guedj, R. Thiébaut, and D. Commenges. Maximum likelihood estimation in dynamical models of HIV. *Biometrics*, 63(4):1198–1206, 2007.

L.P. Hansen. Large sample properties of generalised method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

P. Honoré. Maximum likelihood estimation of non-linear continuous-time term-structure models. *Unpublished workingpaper, Aarhus School of Business*, 1997.

A.S. Hurn, J.I. Jeisman, and K.A. Lindsay. Seeing the wood for trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Finantial Econometrics*, 5:390–455, 2007.

S. Hurn and K. Lindsay. Estimatingthe parameters of stochastic differential equations. *Mathematics and Computers in Simulations*, 48:373–384, 1999.

B. Jensen and R. Poulsen. *A comparison of approximation techniques for transition densities of diffusion processes. Manuscript.* Aarhus University, 2000.

B. Jensen and R. Pulsen. Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives*, 9:18–32, 2002.

M. Kijima. *Markov Processes for Stochastic Modeling.* Chapman and Hall, 1997.

R. Krishna. *Applications of Pharmacokinetic principles in drug development.* Kluwer Academic/Plenum Publishers, 2004.

E. Kuhn and M. Laville. Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data*, 49(4):1020–1038, 2005.

C. Laing and G.J. Lord. *Stochastic methods in neuroscience.* Oxford University Press, 2010.

E. Lindström. Estimating parameters in diffusion processes using an approximate maximum likelihood approach. *Published online*, 2006.

M. Lindstrom and D. Bates. Nonlinear mixed-effects models for repeated measures data. *Biometrics*, 46:673–687, 1990.

A. Lo. Maximum likelihood estimation of generalized itô processes with discretely sampled data. *Econometric Theory*, 4:231–247, 1998.

C. McCulloch and S. Searle. Generalized, linear and mixed models. *Wiley Series in Probability and Statistics, John Wiley and Sons, Inc*, 2001.

J. Nicolau. New techniques for simulatingthe likelihood of stochastic differential equation. *Unpublished workingpaper, ISEG, Universidade Tecnica de Lisboa*, 2000.

S. Ogawa. Some problems in the simulation of nonlinear diffusion processes. *Mathematics and Computers in Simulations*, 28:217–223, 1994.

B. Øksendal. *Stochastic Differential Equations.* Springer, 2005.

R.V. Overgaard, N. Jonsson, C.W. Tornøe, and H.Madsen. Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *Journal Pharmacokinet. Pharmacodyn*, 32:85–107, 2005.

R.V. Overgaard, N. Holford, K.A. Rytved, and H. Madsen. PKPD model of interleukin-21 effects on thermoregulation in monkeys. *Pharmaceutical Research*, 24(2):298–309, 2007.

A. Pedersen. A new approach to mazimum likelihood estimation for stochastic differential equation based on discrete observations. *Scandinavian Journal of Statistics*, 22:55–71, 1995.

M Piazzesi. An econometric model of the yield curve with macroeconomic jump effects. *Unpublished workingpaper, University of California, Los Angeles*, 2000.

U. Picchini, S. Ditlevsen, and A. De Gaetano. Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *Journal of Mathematical Biology*, 5:771–796, 2006.

U. Picchini, S. Ditlevsen, and A. De Gaetano. Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology*, 25:141–155, 2008.

U. Picchini, A. De Gaetano, and S. Ditlevsen. Stochastic differential mixed-effects models. *Scandinavian Jurnal of Statistics*, 30:67–90, 2010.

J.C. Pinheiro and D.M. Bates. Approximations to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical statistics*, 4:12–35, 1995.

M. Rubinstein. Implied binomial trees. *Journal of Financial*, 49:771–818, 1994.

P. Santa-Clara. *Simulated likelihood estimation of diffusions with an application to the short terminterest rate*. PhD thesis, INSEAD, 1995.

K.J. Singleton. Estimation of aśne asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102:111–141, 2001.

M. Sørensen. Prediction based estimating functions. *Econometrics Journal*, 3: 123–147, 2000.

S.M. Sundaresan. Continuous-time methods in finance: A review and an assessment. *Journal of Finance*, 55:1569–1622, 2000.

C. Tornøe, R.V. Overgaard, H.A Nielsen, H. Madsen, and E.N. Jonsson. Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharmaceutical Research*, 22(8):1247–1258, 2005.

R. van Handel. Stochastic calculus, filtering, and stocastic control, 2007.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

E. Vonesh and V. Chinchilli. Linear and nonlinear models for the analysis of repeated measurements. *Marcel Dekker, New York*, 1997.

J. Wang. EM algorithms for nonlinear mixed effects models. *Comput. Stat. Data An.*, 51(6):3244–3256, 2007.

D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, 2012.

Appendix A

Matlab programs

Follow the main Matlab codes used for the parameters estimations.

```
function [stima_alpha,stima_k,stima_sigma,Starting_values,LOGL,var_I_fis
    ]=Brandt_Santa_Clara(alpha_0,k_0,sigma_0,l,X)
%Estimation using BRANDT-SANTA CLARA METHOD
%Parameters value
% True_sigma = 0.02213;
% True_alpha = 0.0717;
% True_k = 0.258;
Starting_values = [alpha_0;k_0;sigma_0];
if l==1
    x0=0.08;
    M = 120;
    interpyes = 1;
    n=12;
    %the data used for the estimation are simulated
    rng(100);
    [delta,X_interp,t_interp]=eulero_maruyama(sigma_0,alpha_0,k_0,x0,M
        ,0,1/12,n,interpyes);

end
if l==0
    %X=Fed Found data monthly from January 1963 to Decembrer 1998
    X_interp=X/100;

    n=432;
    t_interp=linspace(1963,1998,n);
    delta=t_interp(2)-t_interp(1);
    plot(t_interp,X_interp)
end
%S=number of trajectories
S=2500;
true_logL=ME_loglike_int(True_alpha,True_k,True_sigma)
[stima_MLE1,fval,exitflag,output,lambda,grad,hessian]= fmincon(@(x)
    ME_loglike_int(x(1),x(2),x(3)) , Starting_values
    ,[],[],[],[],[0.01,0.05,0.005],[0.3, 0.8, 0.1], [], optimset('
    MaxFunEvals',10000,'MaxIter',1e4,'display','iter','Algorithm','
    active-set','Hessian','bfgs'));
stima_alpha=stima_MLE1(1);
stima_k=stima_MLE1(2);
stima_sigma=stima_MLE1(3);

    function [LOGLIKE]=ME_loglike_int(alpha,k,sigma)
        if (alpha<=1e-7 || sigma<0 || k<0)
            LOGLIKE=9999999999;
            return;
        end
        M=10; % Number of auxiliary points
        interpyes = 0;
        z=zeros(n-1,S);
        rng(100)
```

```
            step=zeros(n-1,1);
            for j=1:n-1

                for s=1:S

                    [r,Y,d]=eulero_maruyama(sigma,alpha,k,X_interp(j),M,
                        t_interp(j),t_interp(j+1),M,interpyes);
                    z(j,s)=Y(end-1);
                    step(j)=r;
                end
            end
            for i=1:n-1
                q_Ms(i)=sum(normpdf(X_interp(i+1),z(i,:)+(k*(alpha-z(i,:)))*
                    step(i),(sigma^2)*step(i)));
            end

            LOGLIKE=log(S)-(sum(log(q_Ms)));

        end

    var_I_fis=1./(diag(hessian));
    LOGL=ME_loglike_int(stima_alpha,stima_k,stima_sigma);
    end
```

```
    function [stima_MLE,Starting_values,LOGLIKE,var_I_fis]=hermite_1(alpha_0
        ,k_0,sigma_0,l,X)
    %ESTIMATION USING THE HERMITE EXPANSION K=1
    Starting_values = [alpha_0;k_0;sigma_0];

    if l==1%we use the data simulated
        x0=0.10;
        M = 4320;
        interpyes = 1;
        n=432;%number of data
        rng(100);
        [delta,X_interp,t_interp]=eulero_maruyama(sigma_0,alpha_0,k_0,x0,M
            ,1963,1998,n,interpyes);

    end
    if l==0
        %X=Fed Found data monthly from January 1963 to Decembrer 1998
        X_interp=X/100;
        n=432;
        t=linspace(1963,1998,n);
        delta=t(2)-t(1);
        plot(t,X_interp)
    end
    [stima_MLE1,fval,exitflag,output,lambda,grad,hessian]= fmincon(@(x)
        ME_loglike_int(x(1),x(2),x(3)) , Starting_values
        ,[],[],[],[],[0.001,0.05,0.005],[0.3, 0.9, 0.1], [], optimset('
        MaxFunEvals',10000,'MaxIter',1e4,'display','iter','Algorithm','
        active-set','Hessian','bfgs'));
        function [LOGLIKE]=ME_loglike_int(alpha,k,sigma)
            if (alpha<=1e-4 || sigma<=1e-5 || k<=1e-5)
                LOGLIKE=9999999999;
                return;
            end
            a=0;B=0;
            Y=X_interp/sigma;
            for i=2:n

                B=B+(-((Y(i)-Y(i-1))^2/(2*delta))-(Y(i)^2*k)/2+(Y(i-1)^2*k)
                    /2+(Y(i)*alpha*k)/sigma-(Y(i-1)*alpha*k)/sigma);
                argomento=1-((1/(6*sigma^2))*k*(3*alpha^2*k-3*(Y(i)+Y(i-1))*
                    alpha*k*sigma+(-3+Y(i)^2*k+Y(i)*Y(i-1)*k+Y(i-1)^2*k)*
                    sigma^2))*delta;
                a=a+log(argomento);
            end
            LOGLIKE=(n/2*log(sigma^2)+(n/2)*log(delta*2*pi)-a-B);
        end

    stima_MLE(1)=stima_MLE1(1);
    stima_MLE(2)=stima_MLE1(2);
    stima_MLE(3)=stima_MLE1(3);
```

```
LOGLIKE=(ME_loglike_int(stima_MLE(1),stima_MLE(2),stima_MLE(3)));
true_logL=ME_loglike_int(True_alpha,True_k,True_sigma);
var_I_fis=1./(diag(hessian));
end
```

```
function [stima_MLE,Starting_values,LOGLIKE,var_I_fis]=hermite_2(alpha_0
    ,k_0,sigma_0,l,X)
%ESTIMATION USING THE HERMITE EXPANSION K=2
Starting_values = [alpha_0;k_0;sigma_0];

if l==1%we use the data simulated
    x0=0.10;
    M = 4320;
    interpyes = 1;
    n=432;%number of data
    rng(100);
    [delta,X_interp,t_interp]=eulero_maruyama(sigma_0,alpha_0,k_0,x0,M
        ,1963,1998,n,interpyes);

end
if l==0
    %X=Fed Found data monthly from January 1963 to Decembrer 1998
    X_interp=X/100;

    n=432;
    t=linspace(1963,1998,n);
    delta=t(2)-t(1);
    plot(t,X_interp,'*-')
end
[stima_MLE1,fval,exitflag,output,lambda,grad,hessian]= fmincon(@(x)
    ME_loglike_int(x(1),x(2),x(3)) , Starting_values
    ,[],[],[],[],[0.01,0.05,0.005],[0.3, 0.8, 0.1], [], optimset('
    MaxFunEvals',10000,'MaxIter',1e4,'display','iter','Algorithm','
    active-set','Hessian','bfgs'));
    function [LOGLIKE]=ME_loglike_int(alpha,k,sigma)
        if (alpha<=1e-4 || sigma<=1e-5 || k<=1e-5)
            LOGLIKE=9999999999;
            return;
        end
        a=0;B=0;
        Y=X_interp/sigma;
        for i=2:n

            B=B+(-((Y(i)-Y(i-1))^2/(2*delta))-(Y(i)^2*k)/2+(Y(i-1)^2*k)
                /2+(Y(i)*alpha*k)/sigma-(Y(i-1)*alpha*k)/sigma);
            c1=1-((1/(6*sigma^2))*k*(3*alpha^2*k-3*(Y(i)+Y(i-1))*alpha*k
                *sigma+(-3+Y(i)^2*k+Y(i)*Y(i-1)*k+Y(i-1)^2*k)*sigma
                ^2))*delta;
            c2=((1/(36*sigma^4))*k^2*(9*alpha^4*k^2-18*Y(i)*alpha^3*k^2*
                sigma+3*alpha^2*k*(-6+5*Y(i)^2*k)*sigma^2-6*Y(i)*
                alpha*k*(-3+Y(i)^2*k)*sigma^3+...
                (3-6*Y(i)^2*k+Y(i)^4*k^2)*sigma^4 +2*k*sigma*(-3*alpha+Y
                    (i)*sigma)*(3*alpha^2*k-3*Y(i)*alpha*k*sigma+(-3+
                    Y(i)^2*k)*sigma^2)*Y(i-1)+...
                3*k*sigma^2*(5*alpha^2*k-4*Y(i)*alpha*k*sigma+(-2+Y(i)
                    ^2*k)*sigma^2)*Y(i-1)^2+2*k^2*sigma^3*(-3*alpha+Y
                    (i)*sigma)*Y(i-1)^3 +k^2*sigma^4*Y(i-1)^4))*(
                    delta^2/2);
            a=a+log(c1+c2);
        end
        LOGLIKE=(n/2*log(sigma^2)+(n/2)*log(delta*2*pi)-a-B);
    end

stima_MLE(1)=stima_MLE1(1);
stima_MLE(2)=stima_MLE1(2);
stima_MLE(3)=stima_MLE1(3);

LOGLIKE=ME_loglike_int(stima_MLE(1),stima_MLE(2),stima_MLE(3));
true_logL=ME_loglike_int(True_alpha,True_k,True_sigma);
var_I_fis=1./(diag(hessian));
end
```

```
function [stima_alpha,stima_k,stima_sigma,Starting_values,LOGL]=
    Durham_Gallant_mod(alpha_0,k_0,sigma_0)
%Parameters value taken from Lindstom
% True_sigma = 0.15;
% True_alpha = 0.06;
% True_k = 0.5;

MM = 43200;
n=2000;%number of the data
T=(0.2)*n; v
Starting_values = [alpha_0; k_0;sigma_0];
%We use data simulated using Eulero Maruyama scheme
x0=0.08;
rng(1985)

interpyes=1;
[delta,X_interp,t_interp]=eulero_maruyama_CIR(sigma_0,alpha_0,k_0,x0,MM
    ,0,T,n,interpyes);

stima_MLE= fminsearchbnd(@(x) ME_loglike_int(x(1),x(2),x(3)) ,
    Starting_values,[0.02,0.05,0.005],[10, 3, 3], optimset('
    MaxFunEvals',1000,'MaxIter',1e4,'Display','iter'));%

stima_alpha=stima_MLE(1);
stima_k=stima_MLE(2);
stima_sigma=stima_MLE(3);

    function [LOGLIKE]=ME_loglike_int(alpha,k,sigma)
        if (alpha<=1e-7 || sigma<=0 || k<=0)
            LOGLIKE=9999999999;
            return;
        end

        S=50;%number of trajectories
        M=16;%number of auxiliary points

        u=zeros(S,M);
        rng(100)

        q_Ms=zeros(1,n-1);
        for j=1:n-1

            %we use the modifie Brownian Bridge
            X0=repmat(X_interp(j),1,S);
            XT=repmat(X_interp(j+1),1,S);
            interpyes = 0;

            [step,u,d]=brownian_bridge_modCIR(sigma,X0,XT,M,t_interp(j),
                t_interp(j+1));


            fi_num=ones(S,1);
            fi_den=ones(S,1);



            for m=1:M-2

                fi_num=fi_num.*normpdf(u(:,m+1),u(:,m)+k*(alpha-u(:,m))*
                    step,sigma*sqrt(u(:,m)*step));
                fi_den=fi_den.*normpdf(u(:,m+1),u(:,m)+((XT-u(:,m))./(
                    t_interp(j+1)-d(m)))*step,sqrt(((M-m-1)/(M-m)))*
                    sigma*sqrt(u(:,m)*step));

            end
            fi_num=fi_num.*normpdf(u(:,M),u(:,M-1)+k*(alpha-u(:,M))*step
                ,sigma*sqrt(u(:,M-1)*step));

            q_Ms(j)=(1/S)*sum(fi_num./fi_den);
        end
        LOGLIKE=-(sum(log(q_Ms)));

        if (fi_num<=1e-323)
            LOGLIKE=1e+200;
```

```
            return;
        end
        if  fi_den<=1e-300
            LOGLIKE=1e+250;
            return;
        end
    end


LOGL=ME_loglike_int(stima_alpha,stima_k,stima_sigma);

end
```

```
function [stima_alpha,stima_k,stima_sigma,Starting_values,LOGL]=
    Durham_Gallant(alpha_0,k_0,sigma_0)
%ESTIMATION USING DURHAM AND GALLANT METHOD
%Parameters value taken from Lindstom
% True_sigma = 0.15;
% True_alpha = 0.06;
% True_k = 0.5;

MM = 4320;
n=1000;%number of the data
T=(1/12)*n; %Delta=1/12, T=final time
Starting_values = [alpha_0;k_0;sigma_0];
x0=0.08;
%Simulation data using Eulero Maruyama scheme
interpyes=1;
[delta,X_interp,t_interp]=eulero_maruyama_CIR(sigma_0,alpha_0,k_0,x0,MM
    ,0,T,n,interpyes);
stima_MLE= fminsearchbnd(@(x) ME_loglike_int(x(1),x(2),x(3)) , [alpha_0;
    k_0;sigma_0],[0.02,0.05,0.005],[3, 10, 2], optimset('MaxFunEvals'
    ,1000,'MaxIter',1e4,'Display','iter'));

stima_alpha=stima_MLE(1);
stima_k=stima_MLE(2);
stima_sigma=stima_MLE(3);

    function [LOGLIKE]=ME_loglike_int(alpha,k,sigma)
        if (alpha<=1e-7 || sigma<=0 || k<=0)
            LOGLIKE=9999999999;
            return;
        end

        S=50;%number of trajectories
        M=16;%number of auxiliary points
        u=zeros(S,M);
        rng(100)

        q_Ms=zeros(1,n-1);
        for j=1:n-1

            %We use the Brownian Bridge
            X0=repmat(X_interp(j),1,S);
            XT=repmat(X_interp(j+1),1,S);
            interpyes = 0;

            [r,u,d]=brownian_bridge_CIR(sigma,X0,XT,M,t_interp(j),
                t_interp(j+1));
            step=r;
            fi_num=ones(S,1);
            fi_den=ones(S,1);

            for m=1:M-2

                fi_num=fi_num.*normpdf(u(:,m+1),u(:,m)+k*(alpha-u(:,m))*
                    step,sigma*sqrt(u(:,m)*step));


                fi_den=fi_den.*normpdf(u(:,m+1),u(:,m)+((XT'-u(:,m))./(
                    t_interp(j+1)-d(m)))*step,sigma*sqrt(u(:,m)*step)
                    );
            end
```

```
                fi_num=fi_num.*normpdf(u(:,M),u(:,M-1)+k*(alpha-u(:,M-1))*
                    step,sigma*sqrt(u(:,M-1)*step));
                q_Ms(j)=(1/S)*sum(fi_num./fi_den);

            end
            LOGLIKE=-(sum(log(q_Ms)));

            if (fi_num<=1e-323)
                LOGLIKE=1e+200;
                return;
            end
            if  fi_den<=1e-300
                LOGLIKE=1e+250;
                return;
            end
        end
    LOGL=ME_loglike_int(stima_alpha,stima_k,stima_sigma);

    end
```

```
    function [stima_Ke,stima_Ka,stima_sigma,stima_eta,stima_mu,
            starting_values]=hermite_random_effect_CL(Ke_0,Ka_0,sigma_0,
            eta_0,mu_0,l,theo,Time,dose)
    %WE CONSIDER THE MODEL USING Cl AS RANDOM PARAMETER
    starting_values = [Ke_0,Ka_0,sigma_0,eta_0,mu_0 ];
    x0=0;
    if l==0
        M=12; %number of subjects
        [d,X,time]=eulero_maruyama_random_CL(sigma_0,Ke_0,Ka_0,dose,mu_0,
                eta_0,x0,M);
        n=length(time);
        t=zeros(M,n);
        dose=zeros(M,1);
        for p=1:M
            t(p,:)=time;
            dose(p)=5;
            delta(p,:)=d;
        end
    end
    if l==1 %we use data from a study by Dr. R. Upton
        X=theo;
        t=Time;
        M=12;%number of subjects
        n=11;
        for p=1:M
            delta(p,:)=diff(t(p,:));
        end
    end

    stima_MLE= fminsearchbnd(@(x) ME_loglike_int(x(1),x(2),x(3),x(4),x(5)) ,
            starting_values
            ,[0.0001,0.01,0.01,0.001,-4.5],[0.3,2,0.799,1,-1], optimset('
            MaxFunEvals',10000,'MaxIter',1e4,'display','iter'));

        function [LOGLIKE]=ME_loglike_int(Ke,Ka,sigma,eta,mu)
            if (sigma<=1e-8 || eta<=1e-8 || Ke <=1e-8 || Ka<=1e-8)
                LOGLIKE=9999999999;
                return;
            end
            Y=X./sigma;
            R=110;
            [z,w]=GaussHermite(R);
            LogL=zeros(1,M);
            for i=1:M
                L=0;
                for r=1:R
                    px=zeros(1,n-1);
                    for j=2:n
                        psi10=-Ke;
                        den=(exp(z(r)*sqrt(2*eta)+mu)*sigma);
                        psi01=(dose(i)*Ka*Ke)/den*(-Ka)*exp(-Ka*t(i,j-1));
                        psi02=Ka^2*(dose(i)*Ka*Ke)/den*exp(-Ka*t(i,j-1));
                        psi=(dose(i)*Ka*Ke)/den*exp(-Ka*t(i,j-1))-Ke*Y(i,j
                            -1);
```

```
                        beta0=1;
                        beta1=-delta(i,j-1)^(1/2)*psi-1/4*delta(i,j-1)^(3/2)
                            *(2*psi01+2*psi*psi10)-1/24*(delta(i,j-1)
                            ^(5/2))*(4*psi02+4*psi01*psi10+4*psi*psi10^2)
                            ;
50                      beta2=1/2*delta(i,j-1)*(psi^2+psi10)+1/12*delta(i,j
                            -1)^2*(6*psi*psi01+6*psi^2*psi10+4*psi10^2)
                            +1/96*delta(i,j-1)^3*(12*psi01^2+16*psi*psi02
                            +40*psi*psi01*psi10+28*psi^2*psi10^2+16*psi10
                            ^3);
                        beta3=-1/6*delta(i,j-1)^(3/2)*(psi^3+3*psi*psi10)
                            -1/48*delta(i,j-1)^(5/2)*(12*psi^2*psi01+12*
                            psi^3*psi10+12*psi01*psi10+28*psi*psi10^2);
                        beta4=1/24*delta(i,j-1)^2*(psi^4+6*psi^2*psi10+3*
                            psi10^2)+1/240*delta(i,j-1)^3*(20*psi^3*psi01
                            +20*psi^4*psi10+60*psi*psi01*psi10+100*psi^2*
                            psi10^2+40*psi10^3);
                        beta5=-1/120*delta(i,j-1)^(5/2)*(psi^5+10*psi^3*
                            psi10+15*psi*psi10^2);
                        beta6=1/720*delta(i,j-1)^3*(psi^6+15*psi^4*psi10+15*
                            psi10^3+45*psi^2*psi10^2);
55                      zeta=(Y(i,j)-Y(i,j-1))/(sqrt(delta(i,j-1)));

                        px(j-1)=1/(sigma*sqrt(delta(i,j-1)*pi))*normpdf(zeta
                            ,0,1)*((beta0*1)+beta1*(-zeta)+beta2*(zeta
                            ^2-1)+beta3*(-zeta^3+3*zeta)+beta4*(zeta^4-6*
                            zeta^2+3)+beta5*(-zeta^5+10*zeta^3-15*zeta)+
                            beta6*(zeta^6-15*zeta^4+45*zeta^2-15));
                        if px(j-1)<=10^(-100)
                            px(j-1)=10^(-16);
60                      end
                    end


                    L=prod(px)*w(r)+L;
65              end
                LogL(i)=log(L);
            end
            LOGLIKE=-sum(LogL);
        end
70
    stima_Ke=stima_MLE(1);
    stima_Ka=stima_MLE(2);
    stima_sigma=stima_MLE(3);
    stima_eta=stima_MLE(4);
75  stima_mu=stima_MLE(5);

    end
```

```
    function [stima_Ke,stima_CL,stima_sigma,stima_eta,stima_mu,True_pars]=
            hermite_random_effect_Ka(Ke_0,CL_0,sigma_0,eta_0,mu_0,l,theo,Time
            ,dose)
    %WE CONSIDER THE MODEL USING Ka AS RANDOM PARAMETER
    starting_values = [Ke_0,Ka_0,sigma_0,eta_0,mu_0 ];
    x0=0;
 5  if l==1
        X=theo;
        t=Time;
        M=12;%number of subjects
        n=11;
10      for p=1:M
            delta(p,:)=diff(t(p,:));
        end
    end
    if l==0
15      M=36;%number of subjects
        dose=5;
        [d,X,time]=eulero_maruyama_random_Ka(sigma_0,Ke_0,CL_0,dose,mu_0,
                eta_0,x0,M);
        n=length(time);
        t=zeros(M,n);
20      dose=zeros(M,1);
        for p=1:M
            t(p,:)=time;
```

```
            dose(p)=5;
            delta(p,:)=d;
    end
end

stima_MLE= fminsearchbnd(@(x) ME_loglike_int(x(1),x(2),x(3),x(4),x(5)) ,
        starting_values
       ,[0.001,0.001,0.005,0.00001,0.001],[0.9,1,0.9,1,1], optimset('
       MaxFunEvals',10000,'MaxIter',1e4,'display','iter'));

    function [LOGLIKE]=ME_loglike_int(Ke,CL,sigma,eta,mu)
        if (sigma<=1e-8 || eta<=1e-8 || Ke<=1e-8 ||CL<=1e-8|| mu<=1e-8)
            LOGLIKE=9999999999;
            return;
        end
        Y=X./sigma;

        LogL=zeros(1,M);
        R=110;
        [z,w] = GaussHermite(R);
        for i=1:M
            L=0;
            for r=1:R
                px=zeros(1,n-1);
                for j=2:n
                    psi10=-Ke;
                    psi01=dose(i)*exp(z(r)*sqrt(2*eta)+mu)*Ke/(CL*sigma)
                        *(-exp(z(r)*sqrt(2*eta)+mu))*exp(-exp(z(r)*
                        sqrt(2*eta)+mu)*t(i,j-1));
                    psi02=exp(z(r)*sqrt(2*eta)+mu)^2*(dose(i)*exp(z(r)*
                        sqrt(2*eta)+mu)*Ke/(CL*sigma))*exp(-exp(z(r)*
                        sqrt(2*eta)+mu)*t(i,j-1));
                    psi=(dose(i)*exp(z(r)*sqrt(2*eta)+mu)*Ke)/(CL*sigma)
                        *exp(-exp(z(r)*sqrt(2*eta)+mu)*t(i,j-1))-Ke*Y
                        (i,j-1);
                    beta0=1;
                    beta1=-delta(i,j-1)^(1/2)*psi-1/4*delta(i,j-1)^(3/2)
                        *(2*psi01+2*psi*psi10)-1/24*(delta(i,j-1)
                        ^(5/2))*(4*psi02+4*psi01*psi10+4*psi*psi10^2)
                        ;
                    beta2=1/2*delta(i,j-1)*(psi^2+psi10)+1/12*delta(i,j
                        -1)^2*(6*psi*psi01+6*psi^2*psi10+4*psi10^2)
                        +1/96*delta(i,j-1)^3*(12*psi01^2+16*psi*psi02
                        +40*psi*psi01*psi10+28*psi^2*psi10^2+16*psi10
                        ^3);
                    beta3=-1/6*delta(i,j-1)^(3/2)*(psi^3+3*psi*psi10)
                        -1/48*delta(i,j-1)^(5/2)*(12*psi^2*psi01+12*
                        psi^3*psi10+12*psi01*psi10+28*psi*psi10^2);
                    beta4=1/24*delta(i,j-1)^2*(psi^4+6*psi^2*psi10+3*
                        psi10^2)+1/240*delta(i,j-1)^3*(20*psi^3*psi01
                        +20*psi^4*psi10+60*psi*psi01*psi10+100*psi^2*
                        psi10^2+40*psi10^3);
                    beta5=-1/120*delta(i,j-1)^(5/2)*(psi^5+10*psi^3*
                        psi10+15*psi*psi10^2);
                    beta6=1/720*delta(i,j-1)^3*(psi^6+15*psi^4*psi10+15*
                        psi10^3+45*psi^2*psi10^2);
                    zeta=(Y(i,j)-Y(i,j-1))/sqrt(delta(i,j-1));
                    px(j-1)=1/(sigma*sqrt(delta(i,j-1)*pi))*normpdf(zeta
                        ,0,1)*((beta0*1)+beta1*(-zeta)+beta2*(zeta
                        ^2-1)+beta3*(-zeta^3+3*zeta)+beta4*(zeta^4-6*
                        zeta^2+3)+beta5*(-zeta^5+10*zeta^3-15*zeta)+
                        beta6*(zeta^6-15*zeta^4+45*zeta^2-15));
                    if px(j-1)<=10^(-100)
                        px(j-1)=10^(-16);
                    end
                end
                L=prod(px)*w(r)+L;
            end
            LogL(i)=log(L);
        end
        LOGLIKE=-sum(LogL);

    end
stima_Ke=stima_MLE(1);
```

```
70 stima_CL=stima_MLE(2);
   stima_sigma=stima_MLE(3);
   stima_eta=stima_MLE(4);
   stima_mu=stima_MLE(5);
   end
```