Simulation and Estimation of Diffusion Processes

Applications in Finance

Åkerlindh, Carl

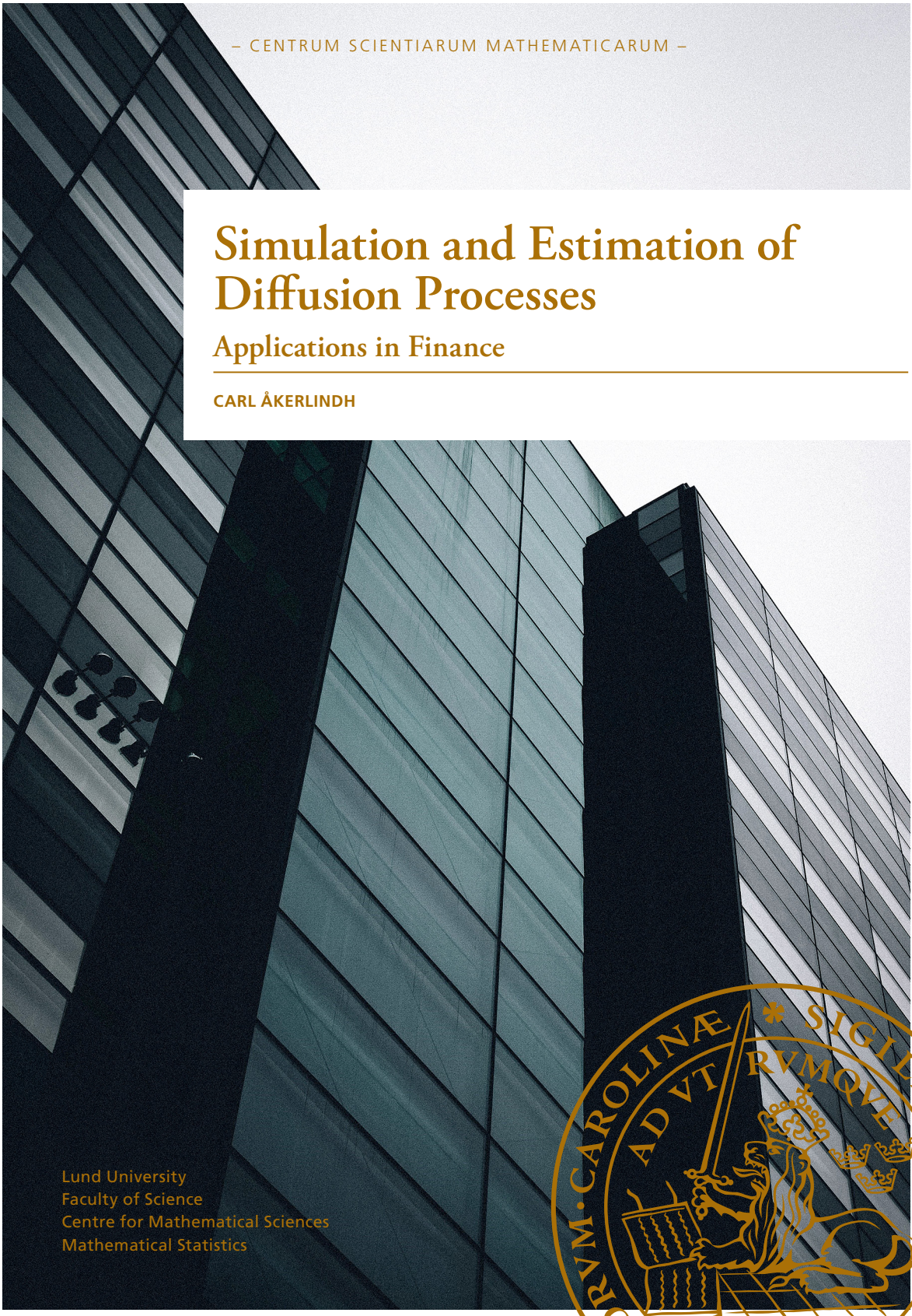
2019

– CENTRUM SCIENTIARUM MATHEMATICARUM –

Simulation and Estimation of Diffusion Processes

Applications in Finance

CARL ÅKERLINDH



Lund University
Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Simulation and Estimation of Diffusion Processes

Simulation and Estimation of Diffusion Processes

Applications in Finance

Carl Åkerlindh

LUND
UNIVERSITY

Thesis for the degree of Doctor of Philosophy in Engineering

To be, with permission of the Faculty of Science of Lund University, publicly defended on Friday 27th of September 2019, at 13:15 in lecture hall MH:Riesz at the Centre for Mathematical Sciences, Sölvegatan 18A, Lund.

Faculty opponent: Associate Professor Jörgen Blomvall, Linköping University

Organization
LUND UNIVERSITY

Mathematical Statistics
Centre for Mathematical Sciences
Box 118
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
2019-09-27

Sponsoring organization

Author(s)
Carl Åkerlindh

Title and subtitle
Simulation and Estimation of Diffusion Processes: Applications in Finance

Abstract
Diffusion processes are the most commonly used models in mathematical finance, and are used extensively not only by academics but also practitioners. Nowadays a wide range of models, that can capture many of the effects observed in financial markets, are available. A very important task is to calibrate the models to observed market data and to achieve a good fit, since a slight misspecification can have large monetary consequences. The focus of this thesis is to investigate both theoretical and computational aspects of parameter estimation for diffusion processes.

In the first paper we consider adaptive calibration where the model parameters are considered to be part of a hidden dynamic state. We then use filtering techniques to estimate the parameter paths. An optimal method for tuning the hyperparameters using the expectation maximization algorithm is presented. The method is evaluated on both simulated and real data, where it is shown to be robust.

The second and third paper cover simulation-based methods for density estimation of diffusion processes using multilevel Monte Carlo estimation. This is a technique that uses simulation on a hierarchy of discretization levels in order to reduce computational complexity. In the second paper we provide an improvement to existing multilevel kernel density estimation by proposing a bandwidth choice that takes model-specific information into account. The third paper extends a simulated maximum likelihood algorithm to the multilevel Monte Carlo framework. Both methods are evaluated on simulated data, where they are shown to provide improvements to the compared methods.

The fourth paper introduces a software package for high-performance simulation of diffusion processes in the Julia programming language. Specific features of Julia are utilized in order to create a simulation library that performs significantly better in terms of computational speed compared to other available libraries, while allowing models to be defined using mathematical notation instead of code.

Keywords
Diffusion processes, Kalman filter, Uncented Kalman filter, EM algorithm, Kernel estimation, Bandwidth selection, Multilevel Monte Carlo, Simulated maximum likelihood estimation, Julia language

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
1404-0034 Doctoral theses in mathematical sciences 2019:4

ISBN
978-91-7895-239-7 (print)
978-91-7895-240-3 (pdf)

Recipient's notes

Number of pages
16 + 129

Price

Security classification

Signature 

Date 2019-08-19

Simulation and Estimation of Diffusion Processes

Applications in Finance

Carl Åkerlindh

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration: Photograph by Oleg Magni.

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118
SE-221 00 Lund
Sweden

Doctoral theses in mathematical sciences 2019:4

ISSN: 1404-0034
ISRN: LUNFMS-1026-2019
ISBN: 978-91-7895-239-7 (print)
ISBN: 978-91-7895-240-3 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2019

Acknowledgments

With the completing of this doctoral thesis, I would like to express my gratitude to anyone that has helped me during the past five years. First, I would like to thank my supervisor Professor Erik Lindström for guiding me through the journey of becoming an independent researcher. I would also like to thank my co-supervisor Associate Professor Magnus Wiktorsson for providing answers to many theoretical questions. Additionally, thank you to my fellow Ph.D. students, especially Maria Juhlin, Unn Dahlén, Rachele Anderson and Philip Kennerberg, for making the workplace more fun. Thank you to my parents Katrin and Göran as well as the rest of my family for always encouraging me to continue learning. Most of all, thank you to my fiancée for all her love, support and many hours of proofreading.

Lund, August 2019 *Carl Åkerlindh*

Populärvetenskaplig sammanfattning

Kvantitativ analys är en teknik som syftar till att förstå komplexa system genom att använda matematisk och statistisk modellering. Det är en viktig del av dagens finansiella system och innebär bland annat att modellera den slumpmässiga utvecklingen av finansiella tillgångar, och för att förutspå verkliga händelser som till exempel förändring av riksbankens styrränta. En mycket vanlig metod som används för detta är så kallad Monte Carlo-simulering. Enkelt uttryckt innebär detta att ett stort antal slumpmässiga utfall från en matematisk modell simuleras för att sedan användas för att beräkna ett förväntad värde av de eftersökta kvantiteterna.

Det finns idag ett överflöd av matematiska modeller som kan fånga många av de egenskaper som observeras i verkligheten. Problemet ligger i att alla modeller styrs av parametrar som måste anpassas till historisk data för att modellerna ska vara praktiskt användbara. Även om en komplex modell i teorin är mer kapabel än en enkel modell, kan den i praktiken prestera sämre på grund av att den är svårare att kalibrera.

Denna avhandling syftar till att utveckla och förbättra metoder för att kalibrera diffusionsprocesser, som är den vanligaste typen av modeller som används inom finansiell matematik. I den första artikeln studeras en metod som tillåter parametrarna att fluktuera med tiden. Artikel två och tre studerar simuleringsbaserade metoder för att skatta fördelningen av observationer från diffusionsprocesser. Den fjärde artikeln beskriver ett mjukvarupaket för att definiera och simulera diffusionsprocesser mycket snabbt i programmeringsspråket Julia.

Abstract

Diffusion processes are the most commonly used models in mathematical finance, and are used extensively not only by academics but also practitioners. Nowadays a wide range of models, that can capture many of the effects observed in financial markets, are available. A very important task is to calibrate the models to observed market data and to achieve a good fit, since a slight misspecification can have large monetary consequences. The focus of this thesis is to investigate both theoretical and computational aspects of parameter estimation for diffusion processes.

In the first paper we consider adaptive calibration where the model parameters are considered to be part of a hidden dynamic state. We then use filtering techniques to estimate the parameter paths. An optimal method for tuning the hyperparameters using the expectation maximization algorithm is presented. The method is evaluated on both simulated and real data, where it is shown to be robust.

The second and third paper cover simulation-based methods for density estimation of diffusion processes using multilevel Monte Carlo estimation. This is a technique that uses simulation on a hierarchy of discretization levels in order to reduce computational complexity. In the second paper we provide an improvement to existing multilevel kernel density estimation by proposing a bandwidth choice that takes model-specific information into account. The third paper extends a simulated maximum likelihood algorithm to the multilevel Monte Carlo framework. Both methods are evaluated on simulated data, where they are shown to provide improvements to the compared methods.

The fourth paper introduces a software package for high-performance simulation of diffusion processes in the Julia programming language. Specific features of Julia are utilized in order to create a simulation library that performs significantly better in terms of computational speed compared to other available libraries, while allowing models to be defined using mathematical notation instead of code.

Contents

List of papers

This thesis consists of the following papers:

Paper A Erik Lindström and Carl Åkerlindh. Optimal adaptive sequential calibration of option models. In Giorgio Consigli, Silvana Stefani, and Giovanni Zambruno, editors, *Handbook of Recent Advances in Commodity and Financial Modeling: Quantitative Methods in Banking, Finance, Insurance, Energy and Commodity Markets*, volume 257 of *International Series in Operations Research & Management Science*, chapter 8, pages 165–181. Springer International Publishing, 2018.

Paper B Carl Åkerlindh and Erik Lindström. A reference bandwidth for multilevel kernel estimation of densities and distributions. 2019. Manuscript submitted for publication.

Paper C Erik Lindström and Carl Åkerlindh. Multilevel simulated maximum likelihood estimation of diffusion processes. 2019. Manuscript in preparation.

Paper D Carl Åkerlindh. High performance simulation of diffusion processes with SDEModels.jl. 2019. Manuscript in preparation.

Introduction

This doctorate thesis in mathematical statistics aims to investigate both theoretical and computational aspects of parameter estimation for statistical models used in finance. This introductory chapter will give a brief overview of the theory used in the enclosed papers.

While diffusion processes play an important role in many fields, their biggest use amongst academics as well as practitioners is likely within the study of mathematical finance. Diffusion processes were introduced to the field of mathematical finance by Merton (1969), and have played a very important role ever since Black and Scholes (1973) published their famous formula for option pricing. In 1997 Robert C. Merton and Myron S. Scholes were awarded the *Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel* for their work.

Nowadays there is an abundance of models based on diffusion processes available, but to be useful in practice it is necessary to calibrate them to market data. The first three papers of this thesis investigate how the parameter estimation of diffusion processes can be improved, by using theory from e.g. time series analysis, non-parametric estimation and Monte Carlo simulation. The fourth paper documents an open source software package written in Julia, which provides a flexible and highly efficient implementation of simulation of diffusion processes, that was developed alongside the other projects.

The rest of the introductory chapter of this thesis is organized as follows: Section 1 introduces the concept of diffusion processes and gives an overview of some of the theory related to these processes. Section 2 gives an overview of the different parameter techniques for diffusion processes that are used in the enclosed papers.

1 Diffusion processes

This section will introduce the theory required for a basic understanding of diffusion processes and stochastic differential equations. We will also cover some important

properties of diffusions, as well as methods for approximating the solution of stochastic differential equations.

1.1 Ordinary differential equations

Differential equations are used to describe the relation between a function and one or more of its derivatives. This type of relation appears almost everywhere around us in nature and society, and it is therefore easy to understand why differential equations play such an important role in so many fields. They are used to model dynamic systems in a deterministic way in physics, engineering and economics to name a few.

When a differential equation only includes functions and derivatives of a single independent variable it is called an ordinary differential equation, in contrast to partial differential equation. A simplified example is the evolution of funds in a bank account $B(t)$ when a deterministic and continuous interest rate r is applied,

$$\frac{\mathrm{d}B(t)}{\mathrm{d}t} = rB(t). \tag{1}$$

In reality though, not many systems in finance evolve in a deterministic manner, certainly not prices of financial instruments and assets. It is therefore necessary to consider a stochastic generalization of ordinary differential equations. Before this can be introduced however, some additional concepts are required.

1.2 Wiener processes

The Wiener process, also commonly called Brownian motion, is a stochastic process which is typically the driving process of stochastic differential equations.

Definition 1.1 (Wiener process) *A stochastic process W which satisfies the following conditions is called a Wiener process.*

1. $W(0) = 0$ *a.s.*
2. *W has independent increments, i.e. for all $t_0 < t_1 \leq t_2 < t_3$ it holds that $W(t_1) - W(t_0)$ and $W(t_3) - W(t_2)$ are independent stochastic variables.*
3. *W has Gaussian increments with $W(t_1) - W(t_0) \sim \mathcal{N}(0, t_1 - t_0)$ for all $t_0 < t_1$.*

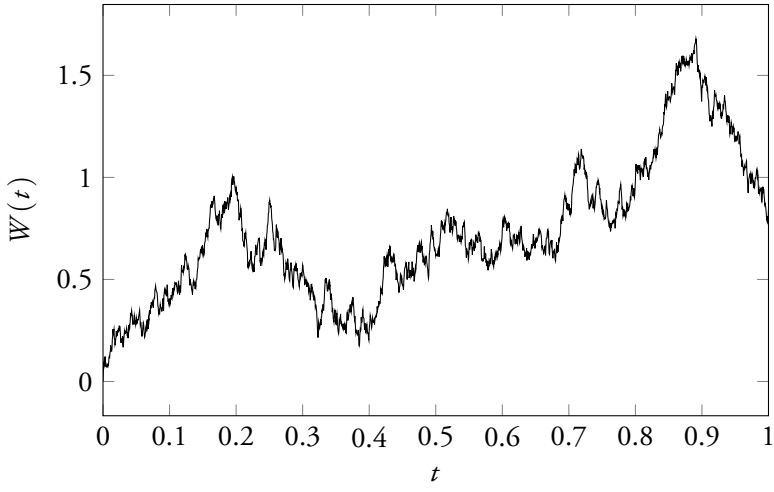


Figure 1: A simulated sample path from a Wiener process.

4. W has continuous paths a.s.

Difference equations can be used as discrete approximations of differential equations, so before considering a stochastic differential equation, we look at the stochastic difference equation given by

$$X(t+h) - X(t) = \mu\big(t, X(t)\big)h + \sigma\big(t, X(t)\big)\big(W(t+h) - W(t)\big). \quad (2)$$

The natural way to go from this difference equation to a differential equation would be to divide both sides with the step size h, and let h go to zero. However, this is not a viable strategy since it can be showed that the Wiener process is nowhere differentiable with probability 1 even though it is continuous, see e.g. Karatzas and Shreve (1998). This is illustrated in Figure 1, where an example of a sample path simulated from a Wiener process is shown. Even if one was to study an infinitely small subpart of the path, similar fractal-like behaviour would be exhibited.

Another option is to let h go to zero without dividing by h first, as a measure to circumvent the problem of non-differentiability. Consider the telescoping sum

$$X(t) - X(0) = \sum_{i=0}^{n-1} X(t_{i+1}) - X(t_i), \quad (3)$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of the time interval such that the mesh size tends to zero as n goes to infinity. Using the difference equation in (2) we can write this as

$$\begin{aligned} X(t) - X(0) = &\sum_{i=0}^{n-1} \mu\big(t_i, X(t_i)\big)(t_{i+1} - t_i) \\ &+ \sum_{i=0}^{n-1} \sigma\big(t_i, X(t_i)\big)\big(W(t_{i+1}) - W(t_i)\big). \end{aligned} \tag{4}$$

Sums and integrals are closely related, and with some intuition we can let n go to infinity to obtain the integral equation

$$X(t) - X(0) = \int_0^t \mu\big(s, X(s)\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, X(s)\big)\,\mathrm{d}W(s). \tag{5}$$

Here the first integral is interpreted as a normal Riemann–Stieltjes integral, while the second is non-standard since the integrator is a stochastic process. In the next section we will look more closely on how such integrals are defined.

1.3 Stochastic integrals

A stochastic integral is an integral where the integrator is a stochastic process. Like with normal integrals, we will need to pose some conditions on the integrand for integrals such as

$$\int_0^t Z(s)\,\mathrm{d}W(s) \tag{6}$$

to be defined. First, assume that Z is adapted to the filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ generated by the Wiener process W. Loosely speaking, it means that the process Z can be completely determined given the observed information generated by the process W. Additionally, the stochastic process Z needs to satisfy

$$\int_0^t \mathrm{E}[Z(s)^2]\,\mathrm{d}s < \infty. \tag{7}$$

We start by only considering integrands that are simple processes, i.e. processes which are constant over all intervals on the form $[t_i, t_{i+1})$ in the partition $0 = t_0 < t_1 < \cdots < t_n = t$. In this case the stochastic integral is simply defined as

$$I(t) = \int_0^t Z(s)\,\mathrm{d}W(s) = \sum_{i=0}^{n-1} Z(t_i)\big(W(t_{i+1}) - W(t_i)\big). \tag{8}$$

For a general integrand which is not a simple process, the process Z is approximated using a sequence Z_n of simple processes which converges as

$$\lim_{n\to\infty} \int_0^t \mathrm{E}\left[\left(Z_n(s) - Z(s)\right)^2\right] \mathrm{d}s = 0. \tag{9}$$

The stochastic integral for general integrands is then defined as a limit,

$$I(t) = \int_0^t Z(s)\, \mathrm{d}W(s) = \lim_{n\to\infty} \int_0^t Z_n(s)\, \mathrm{d}W(s). \tag{10}$$

Given the initial conditions, it can be shown that this limit is a well defined stochastic variable, see e.g Shreve (2004).

Note that the stochastic integral defined in this section has forward-increments which are independent of the integrator. This type of integral is called the Itô integral and is only one of many types of stochastic integrals. However, it has many nice properties particularly suited for financial applications and we will therefore assume that every stochastic integral in this thesis is an Itô integral.

In general, stochastic integrals have no analytic solution, however a special case is

$$\int_0^t \mathrm{d}W(t) = W(0) - W(t) = W(t), \tag{11}$$

which is easy to see from the definition for simple integrands. Although we cannot solve general Itô integrals, we can still characterize them. An important property of the Itô stochastic integral is $\mathrm{E}[I(t)] = 0$, which is easy to verify in the case of simple integrands. The variance is then given by the Itô isometry which is stated as follows.

Theorem 1.1 (Itô isometry) *The Itô integral defined in this section satisfies the relation*

$$\mathrm{E}\left[\left(\int_0^t Z(s)\, \mathrm{d}W(s)\right)^2\right] = \int_0^t \mathrm{E}[Z(s)^2]\, \mathrm{d}s. \tag{12}$$

In addition, assuming the right hand side is finite, we have that $I(t)$ is a $\mathcal{F}_t^W$-martingale. This means that $I(t)$ is adapted to the filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ with $\mathrm{E}[|I(t)|] < \infty$, and for all $s \leq t$ the relation

$$\mathrm{E}\left[I(s) \,\middle|\, \mathcal{F}_t^W\right] = I(s) \tag{13}$$

is satisfied.

1.4 Stochastic differential equations

We are now ready to define the concept of stochastic differential equations. The general form is given by the expression

$$\mathrm{d}X(t) = \mu\big(t, X(t)\big)\,\mathrm{d}t + \sigma\big(t, X(t)\big)\,\mathrm{d}W(t), \tag{14}$$

where μ is called the drift function, and σ is called the diffusion function. This differential should however just be viewed as a short form representation of an integral equation, and should be interpreted as such

$$X(t) - X(0) = \int_0^t \mu\big(s, X(s)\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, X(s)\big)\,\mathrm{d}W(s). \tag{15}$$

One of the most fundamental theorems in stochastic calculus is Itô's lemma, which can be viewed as a stochastic calculus analogue to the ordinary chain rule.

Theorem 1.2 (Itô's lemma) *Let $f(t,x)$ be a $C^{1,2}$-function, and assume the process X has a stochastic differential given by*

$$\mathrm{d}X(t) = \mu\big(t, X(t)\big)\,\mathrm{d}t + \sigma\big(t, X(t)\big)\,\mathrm{d}W(t). \tag{16}$$

Then the process Z defined as $Z(t) = f\big(t, X(t)\big)$ has the stochastic differential

$$\mathrm{d}Z(t) = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x^2}\right)\mathrm{d}t + \sigma\frac{\partial f}{\partial x}\,\mathrm{d}W(t), \tag{17}$$

where the function arguments for μ, σ and the partial derivatives of $f(t,x)$ are left out for readability.

This result can be used to transform a stochastic differential equation into a simpler representation. For example, consider the geometric Brownian motion given by

$$\mathrm{d}X(t) = aX(t)\,\mathrm{d}t + bX(t)\,\mathrm{d}W(t). \tag{18}$$

This expression cannot be solved directly, but if we use the transformation $Z(t) = \ln\big(X(t)\big)$ together with Theorem 1.2 we instead get an arithmetic Brownian motion given by

$$\mathrm{d}Z(t) = \left(a - \frac{b^2}{2}\right)\mathrm{d}t + b\,\mathrm{d}W(t). \tag{19}$$

The solution to this equation is easily found by writing the full integral representation. Together with the inverse transform $X(t) = \exp\big(Z(t)\big)$, the solution to the original equation is given by

$$X(t) = X(0) \exp\left(\left(a - \frac{b^2}{2}\right)t + b\,W(t)\right). \tag{20}$$

Theorem 1.2 can also be used to transform expressions into representations with better numerical properties.

1.5 Discrete-time approximations

General stochastic differential equations rarely have explicit solutions, instead we have to rely on numerical approximations, cf. Fuchs (2013). This is analogous to using various difference approximation schemes for solving deterministic differential equations. The most commonly used approximations are based on discretization in time, whilst state dimension in kept continuous.

These time-discrete approximations are commonly characterized by two convergence properties, strong and weak order of convergence, which are defined as follows.

Definition 1.2 (Strong convergence) *A general discrete-time approximation X^h is said to converge strongly of order γ to the continuous-time process X at time t, if there exists a constant C that does not depend on the time step h, and a $h_0 > 0$ such that*

$$\mathrm{E}\big[\big|X(t) - X^h(t)\big|\big] \leq C\,h^\gamma \tag{21}$$

for all $h \in (0, h_0)$.

Definition 1.3 (Weak convergence) *A general discrete-time approximation X^h is said to converge weakly of order β to the continuous-time process X at time t, if there exists a constant C that does not depend on the time step h, and a $h_0 > 0$ such that*

$$\Big|\mathrm{E}\Big[g\big(X(t)\big) - g\big(X^h(t)\big)\Big]\Big| \leq C\,h^\beta \tag{22}$$

for all $h \in (0, h_0)$, and for all functions $g \in C^{2(\beta+1)}$ with polynomial growth.

The strong order of convergence provides a measure of the pathwise approximation error, while the weak order of convergence provides a measure of the expected approximation error. Note that if g is globally Lipschitz continuous and a scheme has strong order of convergence γ, then weak order of convergence $\beta \geq \gamma$ is guaranteed. This is easy to verify using Jensen's inequality. For many applications in finance, it is enough to only consider the weak order of convergence. We will see later though, that the strong convergence property plays an important role when using a technique called multilevel Monte Carlo.

We will now look at a few commonly used discrete-time approximations for the stochastic differential equations on the form

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \tag{23}$$

where arguments of X and W are changed to subscripts for readability. We partition the interval, over which we are interested in the solution, in n parts as $0 = \tau_0 < \tau_1 < \cdots < \tau_n = t$. The schemes are then defined as a recursive relation, where we define

$$h_k = \tau_{k+1} - \tau_k \tag{24}$$

and

$$\xi_k = W_{k+1} - W_k \tag{25}$$

for convenience. In addition to the the ones presented here, a plethora of other numerical schemes can be found in Kloeden and Platen (1992).

Definition 1.4 (Euler-Maruyama scheme) *The Euler-Maruyama scheme defined recursively by*

$$X_{\tau_{k+1}} = X_{\tau_k} + \mu\big(\tau_k, X_{\tau_k}\big)h_k + \sigma\big(\tau_k, X_{\tau_k}\big)\xi_k \tag{26}$$

is the simplest possible scheme for stochastic differential equations. The scheme has strong order of convergence $\gamma = 0.5$ *and weak order of convergence* $\beta = 1$.

Definition 1.5 (Milstein scheme) *The Milstein scheme in one dimension defined by*

$$\begin{aligned} X_{\tau_{k+1}} = X_{\tau_k} &+ \mu\big(\tau_k, X_{\tau_k}\big)h_k + \sigma\big(\tau_k, X_{\tau_k}\big)\xi_k \\ &+ \frac{1}{2}\sigma\big(\tau_k, X_{\tau_k}\big)\sigma_x\big(\tau_k, X_{\tau_k}\big)(\xi_k^2 - h_k) \end{aligned} \tag{27}$$

adds one additional term to the Euler-Maruyama scheme. This results in a scheme with strong order of convergence $\gamma = 1$ *while the weak order of convergence still is* $\beta = 1$.

When applying the Milstein scheme on equations with state-independent diffusion term, the simple Euler-Maruyama scheme is obtained. This implies that the Euler-Maruyama scheme also has strong order of convergence $\gamma = 1$ for those cases.

Definition 1.6 (Implicit Euler-Maruyama scheme) *A fully implicit variant of the Euler-Maruyama scheme which is weakly consistent can be constructed as*

$$X_{\tau_{k+1}} = X_{\tau_k} + \bar{\mu}\big(\tau_{k+1}, X_{\tau_{k+1}}\big)h_k + \sigma\big(\tau_{k+1}, X_{\tau_{k+1}}\big)\xi_k, \tag{28}$$

where the corrected drift function in the one-dimensional case is defined by

$$\bar{\mu}(t,x) = \mu(t,x) - \sigma(t,x)\sigma_x(t,x). \tag{29}$$

It can be shown that this scheme has a weak order of convergence $\beta = 1$. This scheme has a strong order of convergence $\gamma = 0.5$ when the diffusion term is constant, which is the same as for the explicit Euler-Maruyama scheme.

Deriving numerical schemes with strong order of convergence $\gamma > 0.5$ generally requires simulation of additional random variables originating from iterated Itô integrals. It is not always possible to sample these exactly, however see Wiktorsson (2001) for an algorithm in the case of two-times iterated Itô integrals.

An example on numerical solutions for the geometric Brownian motion in (18), when using the Euler-Maruyama and Milstein schemes defined in Definition 1.4 and Definition 1.5, is shown in Figure 2.

1.6 Numerical integration

With many applications in finance we are interested in evaluating expected values like

$$\mathrm{E}\big[g(X_T) \,\big|\, \mathcal{F}_0\big], \tag{30}$$

where X_T is the solution of some stochastic differential equation at time T and g is a function evaluated at the solution. We will however for simplicity leave out the explicit conditioning in the rest of this section.

In the rare case that the solution X_T is explicitly known, and independent and identically distributed samples x_i where $i = 1, 2, \ldots$ can be simulated from the distribution of the solution, we know from the strong law of large numbers that

$$\hat{g}_N = \frac{1}{N}\sum_{i=1}^{N} g(x_i) \xrightarrow{\text{a.s.}} \mathrm{E}[g(X_T)] = \mu. \tag{31}$$

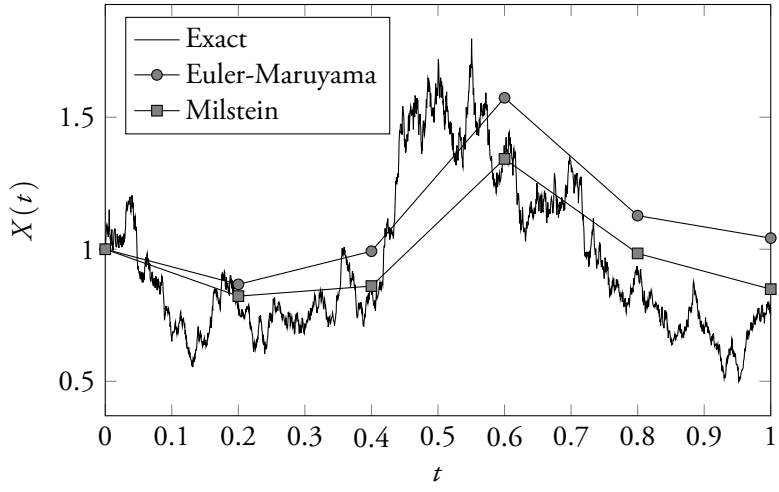


Figure 2: An exact solution together with Euler-Maruyama and Milstein approximate solutions for the geometric Brownian motion defined in (18) with parameters $a = b = 1$ and initial value $X(0) = 1$.

The cost of computing the estimate $\hat{g}_N$, in terms of number of random variables that need to be simulated, is simply N. From the central limit theorem we have

$$\sqrt{N}\,(\hat{g}_N - \mu) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2). \tag{32}$$

Since $\hat{g}_N$ is an unbiased estimator, it is clear that if we want to have an estimator with a mean squared error ϵ^2 we should choose the number of samples and thereby the computational cost as $\mathcal{O}(\epsilon^{-2})$.

When the exact distribution of the solution to a stochastic differential equation is unknown, we instead have to rely on approximations. If a numerical scheme with step size h is used to generate samples x_i^h, we can define a new estimate of the expectation in (30) as

$$\hat{g}_N^h = \frac{1}{N}\sum_{i=1}^{N} g(x_i^h) \xrightarrow{\text{a.s.}} \mathrm{E}[g(X_T^h)]. \tag{33}$$

It follows from the weak order of converge stated in Definition 1.3 that the bias of this estimate tends to zero as the step size goes to zero. To analyze the computational

cost of the estimator $\hat{g}_N^h$ we decompose the mean squared error as

$$\mathrm{MSE}(\hat{g}_N^h) = \frac{1}{N}\mathrm{V}[g(X_T^h)] + \mathrm{E}[g(X_T) - g(X_T^h)]^2. \tag{34}$$

Assuming $\mathrm{V}[g(X_T^h)] < \infty$, we can balance the two terms of the error decomposition. If we seek to obtain an estimate with a mean squared error ϵ^2 it can be shown that the optimal balance between the step size and number of Monte Carlo samples is obtained by choosing $h = \mathcal{O}(\epsilon^{1/\beta})$ and $N = \mathcal{O}(\epsilon^{-2})$, where β is the weak order of convergence of the numerical scheme. This will lead to a computational cost in terms of number of random numbers given by $NT/h = \mathcal{O}(\epsilon^{-2-\frac{1}{\beta}})$. In particular, the computational complexity when using the Euler-Maruyama scheme will be of $\mathcal{O}(\epsilon^{-3})$. In the next section we will introduce a brilliant technique that can reduce this cost significantly to $\mathcal{O}\big(\epsilon^{-2}(\log\epsilon)^2\big)$.

1.7 Multilevel Monte Carlo

Multigrid methods are standard tools in numerical analysis that allow for significantly reduced computational complexity by introducing a hierarchy of discretizations. Multilevel Monte Carlo provides a class of methods inspired by the same ideas, that can be used for reducing the computational cost when estimating expectations using Monte Carlo simulations, see e.g Giles (2015) for an overview.

Assume we can generate approximations of the diffusion X_T on a series of grids with step size $h_l = T/M^l$, where $l = 0, 1, \ldots, L$ and $M \geq 2$ is some integer, as illustrated in Figure 3. We can then construct an approximation of the expectation $\mathrm{E}[g(X_T)]$ as

$$\mathrm{E}[g(X_T^{h_L})] = \mathrm{E}[g(X_T^{h_0})] + \sum_{i=1}^{L}\mathrm{E}[g(X_T^{h_l}) - g(X_T^{h_{l-1}})]. \tag{35}$$

This can be seen as adding a series of gradually refined correction terms to an approximation on a very coarse discretization level. This would be a superfluous rewrite if all random variables were independent, but by letting the individual correction terms $g(X_T^{h_l}) - g(X_T^{h_{l-1}})$ be based on the same underlying Wiener process the variance is given by

$$\mathrm{V}[g(X_T^{h_L})] = \mathrm{V}[g(X_T^{h_0})] + \sum_{i=1}^{L}\mathrm{V}[g(X_T^{h_l}) - g(X_T^{h_{l-1}})]. \tag{36}$$

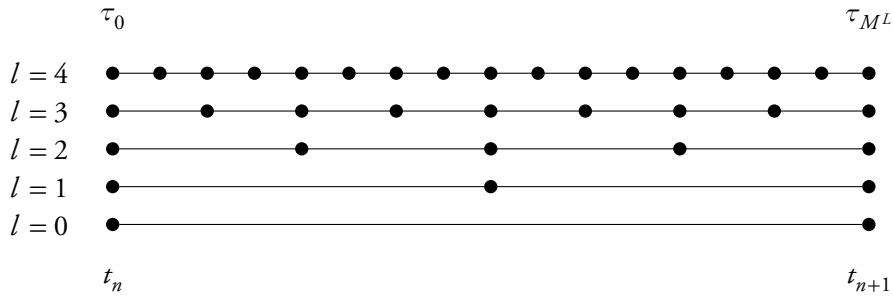


Figure 3: An example of the hierarchy of discretization levels for multilevel Monte Carlo simulations. Here the step size at each level is divided by $M = 2$ compared to the previous level.

This is where the strong order of convergence in Definition 1.2 is important. Given some minor extra conditions, see Kloeden and Platen (1992), it is possible to show that there exists a constant C such that the variance of each correction term can be bounded as

$$\mathrm{V}[g(X_T^{h_l}) - g(X_T^{h_{l-1}})] \leq C h_l^{2\gamma}. \tag{37}$$

Additionally, this has the benefit of allowing simulation of both $g(X_T^{h_l})$ and $g(X_T^{h_{l-1}})$ at the cost of simulating $g(X_T^{h_l})$ since only one Wiener process have to be simulated.

A Monte Carlo estimate of the approximated expectation $\mathrm{E}[g(X_T^{h_L})]$ is constructed as

$$\tilde{g}^L = \frac{1}{N_0} \sum_{i=1}^{N_0} g(x_i^{h_0}) + \sum_{l=1}^{L} \frac{1}{N_l} \sum_{i=1}^{N_l} g(x_i^{h_l}) - g(x_i^{h_{l-1}}), \tag{38}$$

where $N_l, l = 0, \ldots, L$ is the number of samples at each level. The variance of this estimator is then given by

$$\mathrm{V}[\tilde{g}^L] = \frac{1}{N_0} \mathrm{V}[g(X_T^{h_0})] + \sum_{l=1}^{L} \frac{1}{N_l} \mathrm{V}[g(X_T^{h_l}) - g(X_T^{h_{l-1}})] \tag{39}$$

and the computational cost in terms of number of random variables that need to be generated is given by $\sum_{l=0}^{L} N_l M^l$. By simulating more Monte Carlo samples on coarser levels and fewer on finer levels, we typically achieve the estimation bias

determined by the finer level but at a significantly reduced computational cost. By balancing the number of levels and the numbers of Monte Carlo samples at each level Giles (2008) showed that the computational complexity of obtaining an estimate with mean squared error ϵ^2 can be bounded by

$$\text{Cost}(\tilde{g}^L) \leq C \cdot \begin{cases} \epsilon^{-2} & \gamma > 1/2, \\ \epsilon^{-2}\log(\epsilon)^2 & \gamma = 1/2, \\ \epsilon^{-2-\frac{1-2\gamma}{\beta}} & 0 < \gamma < 1/2, \end{cases} \tag{40}$$

where C is some constant. This is a significant reduction compared to the cost ϵ^{-3} of the naive Monte Carlo estimate in Section 1.6.

In Paper B and Paper C we apply the multilevel Monte Carlo technique on two methods for density estimation of diffusion processes.

2 Parameter estimation

In this section we are going to briefly cover the different parameter estimation techniques for diffusion processes that are used in the enclosed papers. We will assume our diffusion process is observed at known discrete time points, and that the process is governed by an unknown parameter vector θ_{true} in which we are interested to estimate.

2.1 Maximum likelihood estimation

One of the most popular parameter estimation methods in statistics is the maximum likelihood estimator, which is based on the idea that the observed data should be most probable under the assumed model. The likelihood function is defined as the joint probability density for all observations under the given set of parameters θ,

$$\begin{aligned} L(\theta) &= p_\theta(x_{0:n}) \\ &= p_\theta(x_0)\prod_{i=1}^{n} p_\theta(x_i \mid x_{0:i-1}), \end{aligned} \tag{41}$$

where we have used the short hand notation $x_{0:n} = \{x_0, \ldots, x_n\}$ for the set of observations. The maximum likelihood estimator is then defined as the set of

parameters that globally maximize the likelihood function given the observed data, i.e.

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta), \tag{42}$$

where Θ is the possibly constrained parameter space. It is usually more convenient to work with the log-likelihood function $\ell(\theta) = \log L(\theta)$ which has the same maximizing argument. The maximum likelihood estimator may therefore equivalently be defined as

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta), \tag{43}$$

where

$$\ell(\theta) = \log p_\theta(x_0) + \sum_{i=1}^{n} \log p_\theta(x_i \mid x_{0:i-1}). \tag{44}$$

In practice the first term is usually disregarded when working with time series models. Since diffusions are Markov processes, all densities on the form $p_\theta(x_i \mid x_{0:i-1})$ reduce to a simple transition density $p_\theta(x_i \mid x_{i-1})$. Therefore we can simplify the definition of log-likelihood function for diffusion processes further to

$$\ell(\theta) = \sum_{i=1}^{n} \log p_\theta(x_i \mid x_{i-1}). \tag{45}$$

The maximum likelihood estimator has numerous desirable properties under rather general assumptions, see e.g Lehmann (1999). For example it is both consistent and asymptotically efficient, as well as asymptotically Gaussian with convergence given by

$$\sqrt{N}(\hat{\theta} - \theta_{\text{true}}) \xrightarrow{\text{d}} \mathcal{N}(0, I_F^{-1}). \tag{46}$$

Here I_F denotes the so-called Fisher information matrix defined by

$$I_F = -\text{E}\big[H\big(\ell(\theta_{\text{true}})\big)\big], \tag{47}$$

where $H\big(\ell(\theta_{\text{true}})\big)$ is the Hessian of the log-likelihood function given a single observation evaluated in the true parameter. By approximating the Fisher information it is straightforward to construct confidence intervals for the parameter estimates.

So far we have assumed that the correct distribution of the observations is known. This is however rarely the case when working with diffusion processes. For example, by applying a discretization scheme from Section 1.5 such as the

Euler-Maruyama scheme we can approximate the unknown likelihood function with a Gaussian likelihood. This will result in an estimator which is still consistent, but no longer efficient. Maximizing this approximated likelihood is often referred to as *naive* or *quasi* maximum likelihood estimation. This approach generally works well when the discretizing step size is small, but might lead to large estimation bias when the step size is too big.

In the rest of this section we will cover alternative techniques for approximating the likelihood when the density of the observed process is unknown.

2.2 EM algorithm

The expectation-maximization (EM) algorithm formally introduced by Dempster et al. (1977) provides an iterative procedure for finding maximum likelihood estimates of the parameters in a statistical model with some form of hidden or latent state. An example when this is particularly useful is when we are interested in estimating parameters in a stochastic volatility process, where the volatility is not directly observed, or when we observe our dynamical process x through a proxy y such as option- or bond prices.

The idea of the EM algorithm is to extend the original maximum likelihood problem into two closely related problems which individually are easier to solve than the original problem. After providing an initial set of parameters, the algorithm proceeds by alternating between two steps until convergence is reached, e.g. when the change in parameters is small enough. In iteration k of the EM algorithm, the steps are defined as follows:

E-step The first step of the algorithm is to obtain the objective function defined as the expectation

$$Q(\theta, \theta_k) = \mathrm{E}\left[\log p_\theta(x_{1:n}, y_{1:n}) \mid y_{1:n}, \theta_k\right]. \tag{48}$$

M-step The second step then involves finding the parameters that maximizes this objective function,

$$\theta_{k+1} = \underset{\theta \in \Theta}{\arg\max}\, Q(\theta, \theta_k). \tag{49}$$

The convergence properties of the EM algorithm are covered in McLachlan and Krishnan (2008). For example it can be shown that the log-likelihood given

by $\ell(\theta_k)$ is guaranteed to increase with each iteration of the algorithm. The EM algorithm however does not guarantee convergence to the global maximum. Despite this, the EM algorithm is typically considered to be an easy and robust method for obtaining maximum likelihood estimates, see e.g. Rydén (2008). For many distributions of the exponential family the EM algorithm is particularly well suited, since the M-step can be solved analytically without the need for costly numerical optimization.

In Paper A we introduce an adaptive sequential estimation algorithm for diffusion models based on the EM algorithm. The diffusion process is augmented with a latent state describing the evolution of the model parameters. This hidden state is allowed to randomly evolve according to some model dynamics governed by a set of tuning parameters, which in turn are estimated from data using the EM algorithm. In the next section we will cover a method that is used to compute quantities that appear in the EM update equations derived in Paper A.

2.3 Filtering and smoothing

Filtering is a powerful tool used for state estimation and forecasting of state space models. In finance filtering techniques are commonly used to estimate volatility in stochastic volatility models.

As an example, consider that we are interested in recovering the short rate $r(t)$, i.e. the instantaneous spot interest rate, which we want to model using the Vasicek model defined as

$$\mathrm{d}r(t) = \kappa\big(\theta - r(t)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{50}$$

Since the short rate is not a traded asset we have no direct observations. Instead we have to rely on e.g. observed bond prices $p(t,T)$, which for a large class of interest rate models, including the Vasicek model, can be written on an affine form

$$p(t,T) = e^{\alpha(t,T)-\beta(t,T)r(t)}. \tag{51}$$

The coefficients $\alpha(t,T)$ and $\beta(t,T)$ depend on the choice of model, see e.g. Björk (2009). In this case where there is no observation noise it is easy to solve this expression to recover the short rate. In finance however, there is rarely such a thing as a single correct price. There is a difference in what price people are willing to sell at and what people are willing to buy at, which is called the bid-ask spread. The uncertainty can be modeled as a form of measurement noise. In this case filtering

can provide an alternative approach to recover the short rate from observed bond prices.

2.3.1 Kalman filter

Continuing with the example, by discretizing the Vasicek model with the Euler-Maruyama scheme described in Definition 1.4 and assuming $\log p(t,T)$ has additive Gaussian measurement noise with zero mean, we get a system described by

$$r(t) = r(t-h) + \kappa\big(\theta - r(t-h)\big)h + \eta_t, \tag{52a}$$

$$\log P(t,T) = \alpha(t,T) - \beta(t,T)r(t) + \epsilon_t, \tag{52b}$$

where $\eta_t \sim \mathcal{N}(0, \sigma^2 h)$. We see that both the state equation and measurement equation are linear.

For general systems on the form

$$x_k = Ax_{k-1} + Bu_k + \eta_k \tag{53a}$$

$$y_k = Cx_k + Du_k + \epsilon_k \tag{53b}$$

with hidden state x_k, measurements y_k and deterministic input signal u_k we can use the Kalman filter. In each step the filter equations provide an estimated mean and covariance, which in the case of Gaussian system are optimal. Using the notation $x_{k|j} = \mathrm{E}[x_k \mid y_{1:j}]$ and $P_{k|j} = \mathrm{V}[x_k \mid y_{1:j}]$ the Kalman filter can be described as follows. The first step is the prediction step where we make predictions of the mean and covariances of the hidden state and the next measurement,

$$x_{k|k-1} = Ax_{k-1|k-1} + Bu_k, \tag{54a}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^\mathsf{T} + Q, \tag{54b}$$

$$y_{k|k-1} = Cx_{k|k-1} + Du_k, \tag{54c}$$

$$P^y_{k|k-1} = CP_{k|k-1}C^\mathsf{T} + R, \tag{54d}$$

$$P^{x,y}_{k|k-1} = P_{k|k-1}C^\mathsf{T}. \tag{54e}$$

In the second step we use the predictions and the next measurement to update the filter mean and covariances,

$$K_k = P^{x,y}_{k|k-1}(P^y_{k|k-1})^{-1}, \tag{55a}$$

$$x_{k|k} = x_{k|k-1} + K_k(y_k - y_{k|k-1}), \tag{55b}$$

$$P_{k|k} = P_{k|k-1} - K_k P^y_{k|k-1} K_k^\mathsf{T}. \tag{55c}$$

These steps are then repeated for each $k = 1, \ldots, N$, where N is the total number of measurements.

2.3.2 Kalman smoother

The Kalman smoother is closely related to the Kalman filter, but while the filter provides estimates such as

$$x_{k|k} = \mathrm{E}[x_k \mid y_{1:k}], \tag{56}$$

the smoother conditions on all the available data to get estimates such as

$$x_{k|N} = \mathrm{E}[x_k \mid y_{1:N}]. \tag{57}$$

This is achieved by running the filter backwards in time after the initial forward filter. The recursive smoothing equations are defined as follows,

$$L_k = P_{k|k} A^{\mathsf{T}} P_{k+1|k}^{-1}, \tag{58a}$$
$$x_{k|N} = x_{k|k} + L_k (x_{k+1|N} - x_{k+1|k}), \tag{58b}$$
$$P_{k|N} = P_{k|k} + L_k (P_{k+1|N} - P_{k+1|k}) L_k^{\mathsf{T}}. \tag{58c}$$

In Figure 4 an example of the Kalman filter and smoother estimates for the Vasicek model is shown. Looking carefully we can see that the Kalman smoother indeed provides a smoother estimate.

2.3.3 Unscented Kalman filter

A major limitation to the standard Kalman filter is that it only works for linear models. There are however many extensions to the Kalman filter which can be used for general models on the form

$$x_k = f(x_{k-1}) + \eta_k, \tag{59a}$$
$$y_k = g(x_k) + \epsilon_k. \tag{59b}$$

One such non-linear filter is the unscented Kalman filter. The basic idea is to cleverly select a number of points, called sigma points, in order to represent all prior

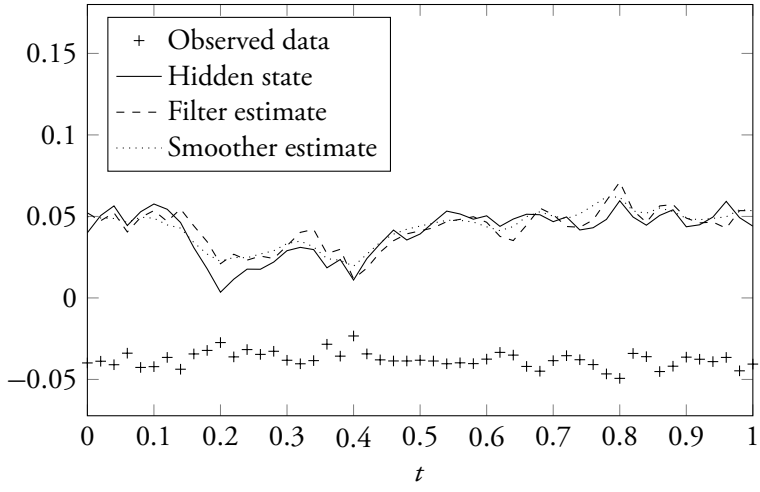


Figure 4: An example of a Kalman filter estimate and a Kalman smoother estimate for the Vasicek short rate model where the observations consist of the log price of a single zero coupon bond.

central moments. In each iteration of the filter the sigma points are defined as

$$\mathcal{X}_{0,k|k} = x_{k|k}, \tag{60a}$$

$$\mathcal{X}_{i,k|k} = x_{k|k} + \left(\sqrt{(d+1)P_{k|k}}\right)_i \qquad i = 1,\dots,d, \tag{60b}$$

$$\mathcal{X}_{i,k|k} = x_{k|k} - \left(\sqrt{(d+1)P_{k|k}}\right)_{i-d} \qquad i = d+1,\dots,2d, \tag{60c}$$

where $\left(\sqrt{(d+1)P_{k|k}}\right)_i$ is the ith matrix column. Alongside the sigma points a number of weights for estimating mean and covariance are defined as

$$w_0^m = \lambda/(d+\lambda), \tag{61a}$$

$$w_0^c = \lambda/(d+\lambda) + (1-\alpha^2+\beta), \tag{61b}$$

$$w_i^m = w_i^c = 1/\big(2(d+\lambda)\big) \qquad i = 1,\dots,2d, \tag{61c}$$

where some tuning parameters are introduced. A common choice for these are $\kappa = 10^{-3}$, $\alpha = 0$ and $\beta = 2$, cf. Wan and van der Merwe (2000).

The sigma points are then propagated through the non-linear state function f and predictions are computed using weighted average. Predictions for the measure-

ments are obtained in a similar way. The full prediction step for the unscented Kalman filter can be defined as

$$\mathcal{X}_{i,k|k-1} = f(\mathcal{X}_{i,k-1|k-1}), \tag{62a}$$

$$x_{k|k-1} = \sum_{i=0}^{2d} w_i^m \mathcal{X}_{i,k|k-1}, \tag{62b}$$

$$P_{k|k-1} = \sum_{i=0}^{2d} w_i^c (\mathcal{X}_{i,k|k-1} - x_{k|k-1})(\mathcal{X}_{i,k|k-1} - x_{k|k-1})^\mathsf{T} + Q, \tag{62c}$$

$$\mathcal{Y}_{i,k|k-1} = g(\mathcal{X}_{i,k|k-1}), \tag{62d}$$

$$y_{k|k-1} = \sum_{i=0}^{2d} w_i^m \mathcal{Y}_{i,k|k-1}, \tag{62e}$$

$$P^y_{k|k-1} = \sum_{i=0}^{2d} w_i^c (\mathcal{Y}_{i,k|k-1} - y_{k|k-1})(\mathcal{Y}_{i,k|k-1} - y_{k|k-1})^\mathsf{T} + R, \tag{62f}$$

$$P^y_{k|k-1} = \sum_{i=0}^{2d} w_i^c (\mathcal{X}_{i,k|k-1} - x_{k|k-1})(\mathcal{Y}_{i,k|k-1} - y_{k|k-1})^\mathsf{T}. \tag{62g}$$

The update step is then exactly the same as for the standard Kalman filter,

$$K_k = P^{x,y}_{k|k-1}(P^y_{k|k-1})^{-1}, \tag{63a}$$

$$x_{k|k} = x_{k|k-1} + K_k(y_k - y_{k|k-1}), \tag{63b}$$

$$P_{k|k} = P_{k|k-1} - K_k P^y_{k|k-1} K_k^\mathsf{T}. \tag{63c}$$

2.4 Kernel estimation

When little is known about the distribution of the observations we can either make some strong assumptions or we can rely on non-parametric estimation methods.

The kernel density estimator is probably one of the most well-recognized techniques for non-parametric estimation of probability density functions, see e.g. Silverman (1986) and Wand and Jones (1995) for an overview of the topic. The kernel density estimator is defined as

$$\hat{p}(x) = \frac{1}{\delta N} \sum_{i=1}^{N} k\Big(\frac{x - x^i}{\delta}\Big), \tag{64}$$

where $x^1, \ldots, x^N$ are independent samples from our unknown distribution.

The foundation of kernel density estimation is the kernel function $k\colon \mathbf{R} \to \mathbf{R}$ which can be any function as long as $\int_{-\infty}^{\infty} k(u)\,\mathrm{d}u = 1$. The two most common choices of kernel function are the Epanechnikov kernel introduced in Epanechnikov (1969) which is optimal in the sense of mean integrated squared error defined as

$$k(u) = \begin{cases} \frac{3}{4}(1-u^2) & \text{if } |u| \leq 1, \\ 0 & \text{if } |u| > 1 \end{cases} \tag{65}$$

and the Gaussian kernel defined as

$$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \tag{66}$$

The rth moment of a kernel function k is defined as $\mu_r(k) = \int_{-\infty}^{\infty} u^r k(u)\,\mathrm{d}u$ and if $\mu_r(k)$ is the first non-zero moment we say that the kernel is of order r. Both the Epanechnikov kernel and the Gaussian kernel defined above are of order $r = 2$. It can in fact be shown that all non-negative and symmetric kernels are of order $r = 2$. These kernels will guarantee that the resulting estimate satisfies the requirements of a probability density function such as non-negativity and integration to 1. While this is reassuring, it might be beneficial to choose a kernel of higher order since these result in a lower estimator bias. For this reason kernels of order $r > 2$ are sometimes called bias reducing kernels. It is possible to construct higher order variants of the Epanechnikov and Gaussian kernels by multiplying the original kernel with a polynomial factor, see Hansen (2005) for general formulas for a larger polynomial family of kernels and Wand and Schucany (1990) for the Gaussian family. A comparison of kernels of different orders is presented in Figure 5.

Although the Epanechnikov kernel is optimal, many other kernel functions perform equally well in practice. Much of the literature on kernel density estimation instead focuses on methods for selecting the smoothing parameter or bandwidth δ. If the bandwidth is too large the resulting estimate will be oversmoothed with low variance but high bias, and if the bandwidth is too small the estimate will be excessively noisy with high variance but low bias. It is necessary to balance the variance and the bias of the kernel density estimator and this has lead to many approaches being developed for solving the bandwidth selection problem, see e.g. Gramacki (2018). The most widely used techniques include rule of thumb, plug-in and cross validation methods.

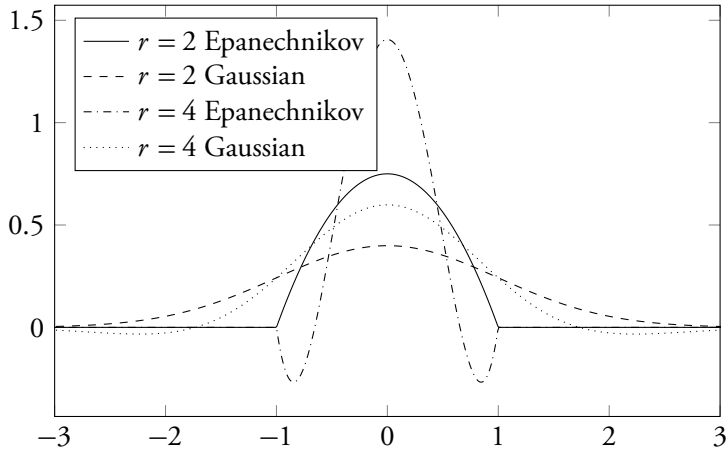


Figure 5: The Epanechnikov kernel and the Gaussian kernel of order $r = 2$ are the most commonly used kernel functions, but both can be extended to higher order. Here they are compared to their 4th order counterparts. Note that all kernels of order $r > 2$ have regions where the function is negative.

It can be shown that the optimal bandwidth that minimizes the asymptotic mean integrated squared error (AMISE) can be defined as

$$\delta_{\text{opt}} = \left(\frac{(r!)^2 R(k)}{2N r \mu_r^2(k) R\big(p^{(r)}\big)} \right)^{1/(2r+1)}, \tag{67}$$

where $R(f) = \int_{-\infty}^{\infty} f(u)^2 \, \mathrm{d}u$. While many other bandwidth selection methods are data-driven, the rule of thumb method introduced in Silverman (1986) proposes to replace the unknown target density p in the expression above with the density of a Gaussian reference distribution in order to be able to compute $R(p^{(r)})$.

In the context of diffusion processes we can estimate unknown transition densities using simulation. By discretizing the process using a numerical scheme from Section 1.5 we can generate the samples from a known distribution in order to estimate an unknown density. Desired accuracy can be achieved by adjusting the number of samples and the step size for the simulation. This idea was further developed by Giles et al. (2015) to incorporate multilevel Monte Carlo simulation of the diffusion process as a mean to reduce the computational complexity. In Paper B we further develop this lower-complexity method to give more robust

estimates at unchanged cost by proposing a reference distribution bandwidth which is balanced to the rest of the multilevel algorithm parameters.

2.5 Simulated maximum likelihood estimation

Another method for estimating probability density functions related to kernel estimation was introduced in Pedersen (1995) and in an earlier version of Brandt and Santa-Clara (2002). By combining the law of total probability with the Markov property of diffusion processes we can reformulate the transition density by imputing an intermediate unknown sample x_s as

$$\begin{aligned} p(x_t|x_0) &= \mathrm{E}\left[p(x_t|x_s) \;\middle|\; x_0\right] \\ &= \int p(x_t|x_s)p(x_s|x_0)\,\mathrm{d}x_s. \end{aligned} \tag{68}$$

We can then generate an empirical version of $p(x_s|x_0)$ using Monte Carlo simulation as

$$p_N(x_s|x_0) = \frac{1}{N}\sum_{i=1}^{N} \delta(x_s - x_s^i), \tag{69}$$

where δ denotes the Dirac delta function. Replacing $p(x_s|x_0)$ with $p_N(x_s|x_0)$ in the integral above we obtain the density estimate given by

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} p(x_t|x_s^i), \tag{70}$$

where the samples x_s^i can be generated using a numerical discretization scheme such as the Euler-Maruyama scheme defined in Section 1.5. However, the expression in (70) is still intractable since the transition density $p(x_t|x_s)$ is unknown. The final Pedersen estimator is thus obtained by also replacing this transition density with a known density via numerical discretization of the diffusion process. An illustration of the Pedersen estimator is shown in Figure 6.

The asymptotic properties of the Pedersen likelihood estimator are discussed in Stramer and Yan (2007), where an optimal balance between the number of Monte Carlo samples and the simulation step size is presented. In order to obtain a root mean squared error of $\mathcal{O}(\epsilon)$ the Pedersen estimator has a computational cost of $\mathcal{O}(\epsilon^{-3})$ when the Euler-Maruyama discretization scheme is used. In Paper C we adapt the Pedersen likelihood estimator to use the multilevel Monte Carlo techniques from Giles (2008) in order to reduce the computational cost significantly to $\mathcal{O}\left(\epsilon^{-2}\log(\epsilon)^2\right)$.

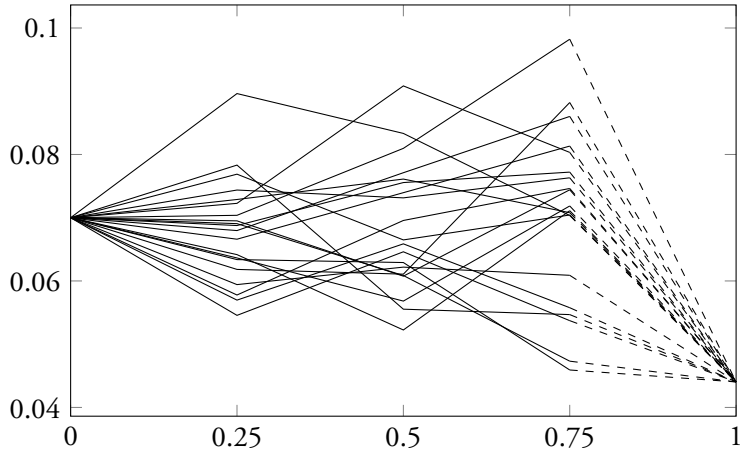


Figure 6: In the Pedersen likelihood estimator an intermediate data point x_s is imputed in order to estimate transition densities as $p(x_t|x_0) = \mathrm{E}\big[p(x_t|x_s) \;\big|\; x_0\big]$. The diffusion process is discretized using a numerical scheme, and samples of the intermediate point x_s are simulated using Monte Carlo (solid lines). The transition density is then finally evaluated by replacing the expectation with a simple average (dashed lines).

3 Outline of the papers

This section gives a very brief overview of the papers included in this thesis. Additionally, for each paper the individual contributions from Carl Åkerlindh (CÅ) and Erik Lindström (EL) are specified.

Paper A

Optimal adaptive sequential calibration of option models

In this paper we consider the problem of calibrating diffusion processes with adaptive parameters using discrete noisy observations. By assuming the diffusion parameters follow one of three different linear dynamic models and to be part of a hidden state, the problem is formulated on state space form with a nonlinear observation equation. The unscented Kalman filter is then used to recover the trajectories for the diffusion parameters. This approach, however, introduces additional unknown variables in the form of tuning parameters describing the evolution of the diffusion parameters. We therefore propose to use the EM algorithm for estimating the hyper parameters, and provide optimal recursive update equations for the considered model dynamics. The method is evaluated on two models commonly used in finance using simulated data as well as options on the S&P 500 index, where it is shown to be robust.

My contribution The initial idea for this project was conceived by EL. Theoretical derivations were performed jointly by CÅ and EL with equal contribution. The implementations and collecting of numerical results were done by CÅ. The paper was written jointly.

Paper B

A reference bandwidth for multilevel kernel estimation of densities and distributions

In this paper we introduce a method for pointwise estimation densities and distributions of diffusion processes, using multilevel Monte Carlo techniques for kernel estimation. It is known that choice of bandwith is crucial, and it is important to take model specific information into account. We improve on existing theoretical derivations of multilevel kernel estimation by proposing a model-dependent bandwith using the rule of thumb method. Simulations show that this approach results

in estimators that are significantly more robust and accurate than the compared multilevel kernel estimation methods.

My contribution The idea for this project was developed by CÅ. Theoretical derivations were performed mainly by CÅ, with the help of EL. The implementations and collecting of numerical results were done by CÅ. The paper was written mainly by CÅ, with the help of EL.

Paper C

Multilevel simulated maximum likelihood estimation of diffusion processes

In this paper we take a different approach on density estimation for diffusion processes, by deriving a multilevel Monte Carlo algorithm for simulated maximum likelihood. It is known that this technique is robust and does not rely on choosing a smoothing parameter. Additionally, we propose a technique for reducing estimator variance by computing the first term in the multilevel sum analytically. This is done by combining a fully explicit Gaussian numerical scheme with a fully implicit Gaussian numerical scheme. The proposed multilevel estimator is tested on three different models. It is seen that the estimator variance is reduced compared to the non-multilevel method, when using the same bias and computational budget.

My contribution The initial idea for this project was conceived by EL. Theoretical derivations were performed jointly by CÅ and EL with equal contribution. The implementations and collecting of numerical results were done by EL. The paper was written jointly.

Paper D

High performance simulation of diffusion processes with SDEModels.jl

This paper introduces a software package for simulating multivariate diffusion processes in the Julia programming language. A novel approach is taken by allowing a model to be defined using the same notation as the mathematical definition. The expression is then automatically transformed into code using the powerful metaprogramming tools available in Julia. By utilizing features such as automatic differentiation and statically sized arrays, a highly performant software framework is

created, without the need for multiple programming languages and thereby avoiding the two-language problem. The simulation performance of the provided package is compared to other available libraries, where it is shown to have significantly better performance.

My contribution All the work for this project was performed by CÅ.

4 References

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Brandt, M. W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 63(2):161–210.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158.

Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media.

Giles, M. B. (2008). Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617.

Giles, M. B. (2015). Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328.

Giles, M. B., Nagapetyan, T., and Ritter, K. (2015). Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):267–295.

Gramacki, A. (2018). *Nonparametric Kernel Density Estimation and its Computational Aspects*, volume 37 of *Studies in Big Data*. Springer.

Hansen, B. E. (2005). Exact mean integrated squared error of higher order kernel estimators. *Econometric Theory*, 21(6):1031–1057.

Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer.

Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer.

Lindström, E. and Åkerlindh, C. (2018). Optimal adaptive sequential calibration of option models. In Consigli, G., Stefani, S., and Zambruno, G., editors, *Handbook of Recent Advances in Commodity and Financial Modeling: Quantitative Methods in Banking, Finance, Insurance, Energy and Commodity Markets*, volume 257 of *International Series in Operations Research & Management Science*, chapter 8, pages 165–181. Springer International Publishing.

Lindström, E. and Åkerlindh, C. (2019). Multilevel simulated maximum likelihood estimation of diffusion processes. Manuscript in preparation.

McLachlan, G. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. John Wiley & Sons.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

Rydén, T. (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, 3(4):659–688.

Shreve, S. E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer finance. Springer.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC Press.

Stramer, O. and Yan, J. (2007). Asymptotics of an efficient Monte Carlo estimation for the transition density of diffusion processes. *Methodology and Computing in Applied Probability*, 9(4):483–496.

Wan, E. A. and van der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapmann & Hall.

Wand, M. P. and Schucany, W. R. (1990). Gaussian-based kernels. *The Canadian Journal of Statistics*, 18(3):197–204.

Wiktorsson, M. (2001). Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *The Annals of Applied Probability*, 11(2):470–487.

Åkerlindh, C. and Lindström, E. (2019). A reference bandwidth for multilevel kernel estimation of densities and distributions. Manuscript submitted for publication.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

Doctoral Theses in Mathematical Sciences 2019:4

ISBN 978-91-7623-096-1
ISSN 1404-0034