Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

A Monte Carlo approach to Fractional Brownian Motion

Supervisor:
Prof. Giacomo Bormetti

Submitted by:
Cristian-Alexandru Cioriia

Academic Year 2022/2023

Abstract

Used widely in physics, from the study of crystals to activity of solar flares, Fractional Brownian Motion (fBM) has witnessed an outburst of research in the last decade in the field of financial mathematics, giving birth to the field of rough volatility models. These models have the ability to account for a range of stylized facts of financial markets which Brownian motion based alternatives simply fail to replicate. However, rough volatility models are challenging to implement due to the non-Markovian and non-semimartingale nature of fBM.In this thesis we are leveraging Monte Carlo methods to investigate the efficiency of the Euler discretization scheme for the Rough Heston model. Our analysis reveals that the Euler implementation is not as effective as moment matching schemes, such as the Quadratic Exponential approach, in the classical Heston context, particularly when the Feller condition is not met. Despite this, it is capable of efficiently pricing European, Asian and Lookback options in the short term in the Rough Heston model. We also investigate convergence of the Euler scheme to the Rough Heston characteristic function and comment on its range of usability.

Contents

Introduction

Unanimously regarded as one of the key phenomena behind the establishment of the atomic theory of matter, Brownian motion has been used far and wide both within physics as well as outside of it in fields such as chemistry, biology, finance, information theory etc. Brownian motion gives rise to diffusion processes [1] [2] which imply movement of a generic substance from regions of high concentration to regions of lower concentrations, which typically results in an equilibrium state. As we know from thermodynamics, diffusion is responsible for loss of information, particularly because the diffusion equation is not deterministic (time - reversible). This means that systems undergoing diffusion lose memory of their previous states over time. On a more local level, particles experiencing Brownian motion are driven by forces which are intrinsically random, and hence their evolution is governed by a probability density function (pdf). The pdf associated to Brownian motion is, unsurprisingly, the time independent Gaussian distribution, which also makes it obvious why Brownian motion became so widespread. Notably, time independence of the pdf implies the system is in an equilibrium state (stationary), which is an assumption that constrains the range of phenomena we can study using it.

In physics, the thermal diffusion equation has been used since the middle of the nineteenth century. However interest in dynamics governed by pdfs different from the Gaussian distribution became mainstream only later. The mathematicians Levy, Kolmogorov, Markov, Wiener, Ito as well as others have set up the rigorous foundations for the study of random or stochastic processes, especially those ruled by diffusion type equations, in the first half of the twentieth century [3]. Another important advancement made in that period coming from physics is the Feynman-Kac formula [4] which mediates the connection between parabolic partial differential equations (PDE) and stochastic processes. The Feynman-Kac formula essentially allows to compute solutions of a class of PDEs by simulating paths of random processes and computing ensemble averages. This is where Monte Carlo methods start to shine. Indeed, the advent of computers was an important driver in the research of stochastic processes [5]. Nevertheless, up until the 1950s research focused on memoryless or Markov processes, whose signature is that the future of the system is dependant on the present state only, with past states being irrelevant.

Surprisingly, Brownian motion was adopted for use in finance even before the famous work of Albert Einstein in 1905. Louis Bachelier, the pioneer of financial mathematics, published his Phd thesis entitled "The theory of speculation" in 1900 [6]. He anticipated the importance of applying mathematics to the financial markets by studying the pricing of stock options. In his thesis he laid the groundwork of what is now called the Efficient

Market Hypothesis (EMH) and succeeded in deriving the formula for pricing European Options and making the link between random movements of stock prices and the diffusion equation. His insights were neglected by the mathematics community at the time because his arguments were not solid enough from a mathematical standpoint. For a long time, his ideas received sparse attention, although they were credited by the founders of the field of stochastic processes. Fortunately, his work reached a wider audience with the introduction of the Black-Scholes-Merton (BSM) [7] [8] model (1973) which assumes the following process for the price of an asset:

$$dS = \mu S dt + \sigma S dW, \tag{1}$$

where S is the asset price, μ is the drift of the process, σ is the volatility (standard deviation) of the price movement and dW stands for an increment of a standard Brownian motion. This process is also called the Geometric Brownian motion. BSM model acts as a standard for determining option prices. It assumes the market is efficient, meaning it is in a state of equilibrium and present asset prices reflect all available information (past price changes being irrelevant), and is thus a Markov process. In this regime, price evolution is driven by the familiar Brownian motion, plus a drift term (usually an interest rate).

BSM model offers analytical solutions for the prices of European call and put options, given specified parameters such as the underlying asset price ($S(t)$), the strike price (K), the risk-free rate (r), the market volatility (σ), the current time (t), and the expiration date or maturity (T). By fixing all parameters except for the volatility σ and varying it, we can obtain the price of an option at time t with that specific volatility. At the same time, we can specify an option price and all the other parameters apart from volatility σ and back track through the BSM formula to obtain the Implied Volatility (IV) labelled σ_{BS}. Thus the implied volatility allows one to gauge how far from the efficient market the option prices are traded at any particular time. By aggregating market data for various strikes K and maturities T we can create an Implied Volatility (IV) surface. By looking at the IV surface at a fixed T we can see what is called a volatility smile, an example of which can be seen in figure 1. If the market were efficient we would see a flat surface. However, in reality, we tend to see a convex shape resembling a smile as volatility tends to be higher for options with strike K different than the spot asset price. Obviously, the complex behaviour of people acting on financial markets cannot be explained away using just Brownian motion in one of its most simple forms. In order to make amends to our financial models we need to account for the behaviour of the implied volatility. This is where stochastic volatility (SV) models come in, a great example of which is the Heston model (1993). The Heston model [9] is described via two stochastic differential equations, one for the asset price S_t and one for the instantaneous variance v_t like so

$$dS_t = \mu_t S_t dt + \sqrt{v_t} S_t dW_1, \tag{2}$$

$$dv_t = -\lambda \left(v_t - \hat{v}\right) dt + \eta \sqrt{v_t} dW_2, \tag{3}$$

where μ_t is the (deterministic) instantaneous drift of stock price returns, v_t is the price variance parameter, η is the volatility of volatility, λ controls the speed of reversion to

historical average $\hat{v}$, and ρ is the correlation between random stock price returns and changes in v_t. dW_1 and dW_2 are standard Brownian motion processes.

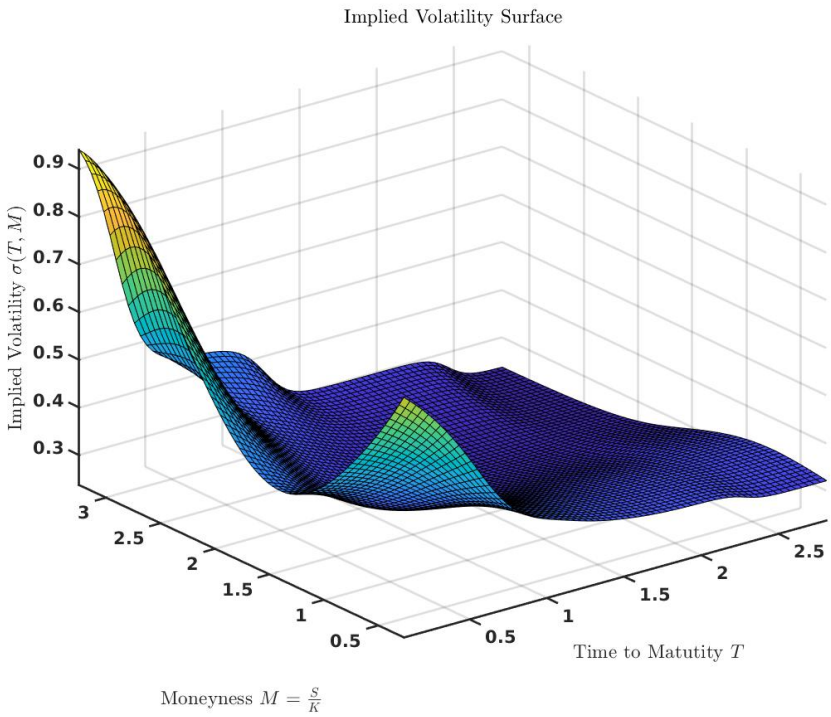


Figure 1: Implied volatility surface displaying the smile feature. For fixed time to expiry, we notice the log strike, $\log K$, has a distinct smile profile which is less noticeable as we increase time to maturity, Source: https://www.stonybrook.edu/commcms/ams/graduate/qf/

Stochastic volatility models such as the Heston model are designed to allows us to calibrate our parameters to market data in a way that is able to reproduce the implied volatility surface. The Heston model does this in an exceptionally simple way [10], namely the η controls the convexity of the smile, ρ skews the smile and the other parameters give the overall level of the surface. The Heston model owes its popularity to having closed form solutions to most quantities of interest such as the characteristic function, which encapsulates information about all the moments of distribution of the stock prices. The Heston model reproduces the IV surface remarkably well for large maturities, however it is unable to display realistic behaviour for short term maturities. Other more intricate stochastic volatility models have been developed to account for more features of financial markets, although there is no one size fits all in terms of SV models. Despite its evident simplicity, it is quite impressive to notice that Brownian motion based models are able to recreate so much of the complexity of financial markets.

So far we have discussed models which use Brownian motion as a fundamental building block. This means that we assume market behaviour can be ultimately described by diffusion type equations. Although attractive for analytical solutions and fast calibrations, these models miss the mark especially in times of market turbulence, such as crises. Other market features such as volatility clustering and fat tails are frequently misrepresented by Brownian motion based models. Even though the outlook of equilibrium states and

stationary Gaussian probability distribution is very enticing, it presents an incomplete picture of market behaviour. A more general class of stochastic processes, namely non stationary processes which exhibit memory (long range dependence), have been considered by mathematicians like Kolmogorov and were observed in turbulent media such as the atmosphere. Through an interesting turn of events a modified version of Brownian motion can account for a considerable amount of the complexity of turbulent flows and non linear diffusion.

Fractional Brownian Motion (fBM) [11] became well known following a publication by Madlebrot and van Ness in 1968. fBM sees itself as the center of many fields such as fractal theory, turbulent hydrodynamics, chaos theory, finance, engineering and others because it encompasses two very useful properties of pdfs, namely stationary Gaussian increments and long range dependence (LRD) on previous states. fBM is a pivotal development because even though it has some of the familiar properties of diffusion processes it cannot be described through diffusion type equations. As such, a different set of tools has to be used, an important example of which is fractional calculus. fBM, just like financial time series, displays rough trajectories, meaning that it exhibits persistent and irregular patterns over time. This property has been observed across a range of financial assets, including stock prices, interest rates, and exchange rates. This has important implications for risk management, as classical Brownian motion based models underestimate the likelihood of extreme events. This is usually referred to as non-Gaussianity of asset returns, which often deviates from the normal distribution, with more extreme events occurring more frequently than would be expected under efficient market assumptions.

In the last decade, a new generation of models has spurred from leveraging the properties of fBM [12]. These are the so called rough volatility models, of which the Rough Heston model [13] is a splendid example. The Rough Heston model is a recent extension of the Heston model that aims to address some of the previously mentioned features of financial markets, making it more accurate for pricing of financial derivatives and better suited for risk management. The Rough Heston model considers the asset price S_t follows a Geometric Brownian motion while the instantaneous variance process v_t follows a fractional mean reverting process

$$dS_t = S_t\sqrt{v_t}dW_t \tag{4}$$

$$v_t = v_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\lambda\left(\theta - v_s\right)ds + \frac{\eta}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\sqrt{v_s}dW_s \tag{5}$$

$$\langle dW_s, dW_t\rangle = \rho dt \quad and \quad H \in (0, 1/2) \tag{6}$$

where λ, θ, η, S_t, ρ are identically defined as previously in the classical Heston model and dW_t, dW_s are two correlated standard Brownian motions. The weighting function $\frac{1}{\Gamma(H+1/2)}(t-s)^{H-1/2}$ is called the kernel, and is the term responsible for long range dependence of rough volatility. H is called the Hurst index and dictates the smoothness of trajectories of the variance process.

As can be hinted from the integral representation of the volatility process, computations can become easily unwieldy. Despite its many advantages, non Markovianity of fBM

makes it difficult to compute quantities of interest because they typically have no closed form solutions. Thus, the search of fast and accurate methods of simulating fBM based processes has taken a considerable amount of the attention of the financial mathematics community [14].

This thesis concerns itself with the implementation of one such algorithm for simulating the Rough Heston model, specifically the Euler discretization scheme. Given that the Rough Heston model in the limit of $H \to \frac{1}{2}$ turns to the classical Heston model, we begin by first implementing the classical counterpart so as to have a benchmark to compare it to. We then show the behaviour of the Euler scheme in the context of the Heston model and how moment-matching schemes such as the Quadratic Exponential (QE) [15] provide a more efficient way of simulating the Heston process. Lastly, we concern ourselves with the short time behaviour of the Euler discretization scheme for the Rough Heston model. We show how the easy to implement Euler scheme gives satisfactory results when pricing derivatives such as European and Asian option in the short term and display the convergence behaviour to higher moments of distribution using the Rough Heston Characteristic function. Thus we show that the Euler scheme is a valuable tool for risk management professionals and portfolio managers which leverage money market options to protect against extreme market events which more traditional models are oblivious to. The code related to this project can be found at *https://github.com/CioriiaCristian/roughvolpy*.

The structure of this dissertation is as follows: Chapters 1 gives an account of Brownian motion, from historical foundation to mathematical formulation. We give a brief introduction to Stochastic process and focus particularly on the class of Markov processes, taking the time to familiarize ourselves with Stochastic differential equations. Ultimately, in this chapter we describe the stochastic processes used later in the thesis, namely Arithmetic and Geometric Brownian motion, the Ohrnstein-Uhlenbeck and the Mean reversion process. Chapter 2 concerns itself with the introduction of Fractional Brownian Motion (fBM) from three different angles, physics, mathematics and finance. Chapter 3 considers the application of stochastic processes to financial markets. Here we describe the Black-Scholes-Merton model, the classical Heston and the rough Heston model. Nevertheless, we also mention the forward variance models such as Bergomi and Rough Bergomi. Chapter 4 describes the different Monte Carlo implementations schemes for the Heston and Rough Heston models. Finally, Chapter 5 and 6 are left for a discussion on results, considerations on possible extensions of the current work, conclusion and acknowledgements.

Chapter 1

Brownian Motion

1.1 Historical introduction

Undoubtedly, one of the most important breakthroughs of the past century has been the transition from a deterministic view of physical processes to a probabilistic view. This happened because of the need to describe both atomic processes and chaotic processes, the latter of which we are concerned with in this paper. One of the first chaotic processes to be studied was Brownian motion. The first widely adopted formulation of Brownian motion was made by Einstein in 1905, in his famous paper [16] where he explains that the movement of suspended fine particles in water can only be described probabilistically. This paper essentially started the field of stochastic modelling [17] [1]. As Brownian motion serves as a basis for all other processes in this paper, a brief description is necessary.

1.1.1 Einstein derivation

In what follows the authors will give an account of the use of Brownian motion as captured in the famous 1905 paper of Einstein [16].

Consider a liquid on the surface of which there are n suspended particles. Take the Cartesian coordinate system to describe the particle position. After time interval τ, the X-coordinate of each particle will change by an amount Δ, positive or negative, different for each particle. Let dn be the number of particles whose x coordinates are displaced by an amount between Δ and $\Delta + d\Delta$. Also, let $\phi(\Delta)$ be the probability describing the observation of a displacement between Δ and $\Delta + d\Delta$ after time τ. We can then express dn as

$$dn = n\phi(\Delta)d\Delta, \tag{1.1}$$

where

$$\int_{-\infty}^{\infty} \phi(\Delta)d\Delta = 1. \tag{1.2}$$

Take ϕ to be non zero only for small Δ, and constrain it via $\phi(\Delta) = \phi(-\Delta)$. Next, take $f(x,t)$ to be the number of particles per unit volume. Using the displacement probability law assumed before, let us express the distribution of suspended particles found at position x at time $t+\tau$ as

$$f(x, t+\tau)dx = dx \int_{-\infty}^{\infty} f(x+\Delta, t)\phi(\Delta)d\Delta \tag{1.3}$$

τ is very small and so

$$f(x, t+\tau) = f(x,t) + \tau\frac{\partial f}{\partial t}. \tag{1.4}$$

Now, expand $f(x+\Delta, t)$ as a Taylor series in Δ :

$$f(x+\Delta, t) = f(x,t) + \Delta\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 f(x,t)}{\partial x^2} + \cdots \tag{1.5}$$

Substitute in the integral equation of 1.3 to get

$$f + \frac{\partial f}{\partial \tau}\tau = f\int_{-\infty}^{\infty}\phi(\Delta)d\Delta + \frac{\partial f}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta + \frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta. \tag{1.6}$$

Because $\phi(x) = \phi(-x)$, the odd Δ power integral terms vanish. We know

$$\int_{-\infty}^{\infty}\phi(\Delta)d\Delta = 1 \tag{1.7}$$

and setting

$$\frac{1}{\tau}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta = D, \tag{1.8}$$

allow us to finally write

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}. \tag{1.9}$$

We thus have arrived at the heat diffusion equation. The solution to this equation is

$$f(x,t) = \frac{n}{\sqrt{4\pi D}}\frac{e^{-x^2/4Dt}}{\sqrt{t}}. \tag{1.10}$$

A well known result is that the mean displacement as a function of time from the starting position is

$$\langle x^2\rangle = 2Dt. \tag{1.11}$$

To note, Einstein makes the assumption that the particle position changes only at times $0, \tau, 2\tau, 3\tau \ldots$, and τ is considered so small as to take t to be continuous. Even though rational and intuitive at first glance, this derivation is not foolproof mathematically. Importantly, all the first derivations and uses of Brownian motion were not fully general and they were given a much more rigorous treatment in later decades [18].

1.1.2 Langevin derivation

The later derivation of Langevin acts as a first instance of a stochastic differential equation [1].

From the foundational work of Boltzmann on statistical mechanics, we know that the kinetic energy of Brownian particles should have an equilibrium value of

$$\left\langle \frac{1}{2} m v^2 \right\rangle = \frac{1}{2} kT, \tag{1.12}$$

where T is the absolute temperature and K is Boltzmann's constant. The Brownian particle should also experience 2 other forces:

1. a drag force which, borrowed from macroscopic hydrodynamics, is $-6\pi\eta a dx/dt$, where η is the viscosity and a the particle diameter.

2. a fluctuating force X which stands for the random impacts of molecules on the Brownian particle. A critical assumption on this force is that it has a uniform angular distribution at all times, and that it has no correlation to the particle position $x(t)$. Thus, by virtue of Newton's laws we can write

$$m \frac{d^2 x}{dt^2} = -6\pi\eta a \frac{dx}{dt} + X. \tag{1.13}$$

Multiply by x to get

$$\frac{m}{2} \frac{d^2}{dt^2} \left(x^2\right) - m v^2 = -3\pi\eta a \frac{d\left(x^2\right)}{dt} + Xx, \tag{1.14}$$

where $v = dx/dt$. By computing an expectation value of the current equation we obtain

$$\frac{m}{2} \frac{d^2 \left\langle x^2 \right\rangle}{dt^2} + 3\pi\eta a \frac{d \left\langle x^2 \right\rangle}{dt} = kT, \tag{1.15}$$

where $\langle xX \rangle$ cancels out because of the assumptions we made on X. The solution to this differential equation is

$$\frac{d \left\langle x^2 \right\rangle}{dt} = kT/(3\pi\eta a) + C \exp(-6\pi\eta a t/m), \tag{1.16}$$

where C is the arbitrary constant. The exponential factor becomes negligible on the timescale of 10^{-8} in Langevin's estimation, which then allows us to write

$$\left\langle x^2 \right\rangle - \left\langle x_0^2 \right\rangle = [kT/(3\pi\eta a)]t. \tag{1.17}$$

This corresponds to a random walk $s(t)$ described by

$$s(t) = \sqrt{\langle x^2 \rangle - \langle x_0^2 \rangle} = \sqrt{2D}\sqrt{t}. \tag{1.18}$$

This result can be derived from a *Heat diffusion* equation

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}, \tag{1.19}$$

where $D = kT/(6\pi\eta a)$ is the diffusion coefficient and f is the particle number density function [19]. The *Heat diffusion* equation is thus closely connected to stochastic processes. For inhomogeneous processes [20] the *Diffusion Equation* is:

$$\frac{\partial f(\mathbf{r},t)}{\partial t} = \nabla \cdot [D(\mathbf{r},t) \cdot \nabla f(\mathbf{r},t)] = \nabla D(\mathbf{r},t) \cdot \nabla f(\mathbf{r},t) + D(\mathbf{r},t) \cdot \nabla^2 f(\mathbf{r},t) =$$

$$= \mathbf{a}(\mathbf{r},t) \cdot \nabla f(\mathbf{r},t) + b(\mathbf{r},t) \cdot \nabla^2 f(\mathbf{r},t). \tag{1.20}$$

This has the same shape as the Fokker-Planck equation which will be introduced later. In this format, the diffusion equation is useful for continuously differentiable functions, and for isotropic diffusion coefficient. However, stochastic processes may have neither of these properties. Hence, for a start, we need a way to compute diffusion processes which are both discretized and non differentiable and/or non continuous.

It is clear to see that Brownian motion is closely related to the diffusion equation because of the thermal diffusion of water particles. Langevin expressed this insight by assuming the force on the Brownian particle $X(t)$ and the the Brownian particle position $x(t)$ are independent of each other and that $X(t)$ is extremely irregular. Einstein expressed the same insight, but instead of using a stochastic variable in a differential equation, he used a probability distribution for the displacement of Brownian particles and further assumed the continuity and differentiability of the Brownian particle number density. Thus we notice that in talking about stochastic processes multiple important concepts come together, namely continuity and discreteness, differentiability, probability distributions, statistical independence of stochastic variables, homogeneity, isotropy and diffusion. In order to have any confidence on our results, these concepts need to be disentangled one by one.

1.1.3 Bachelier derivation

In what follows the author will describe the derivation performed by L.Bachelier in his 1900 Phd thesis [6].

Let us focus on the time evolution of a stock price $s(t)$. Let us consider some key probabilities. Call the probability of the stock price being between x and $x + dx$ at time t_1

$$p(x, t_1)dx. \tag{1.21}$$

Next, consider the probability of the stock price being between z and $z + dz$ at time $t_1 + t_2$, given the price was x at time t_1.

$$p(z, t_1 + t_2|x, t_1)dz. \tag{1.22}$$

In his Phd thesis Bachelier then makes multiple implicit assumptions about the process driving the stock price. First, Bachelier assumes the time translational invariance of the probability distribution of increments, stated mathematically as

$$p(z, t_1 + t_2|x, t_1) = p(z, t_2|x, 0). \tag{1.23}$$

This is motivated by the assumption that we analyze the market prices only during a stable period. Next, he assumes the process is Levy stable which, in other words, also states that the difference between the stock prices at different times $s(t_2) - s(t_1)$ follows the same process as both $s(t_1)$ and $s(t_2)$. This allows us to use the same symbol p for the increment probability distribution as well as for the stock price probability distribution. Among other arguments, Bachelier also states that during a stable period there is no consensus of whether the price will go up or down, half the market agents being in favour for each. This also means, by extension, that each new price increment is independent of the last, because we still assume a stable market period. In other words, Bachelier also assumes spatial translation invariance which can be stated as

$$p(z, t_2|x, 0) = p(z - x, t_2|0, 0) = p(z - x, t_2). \tag{1.24}$$

Statistical independence implies the compound probability is simply the product between probabilities. We can then write the probability of the stock price being z at time $t_1 + t_2$ as

$$p(z, t_1 + t_2)\, dz = \left\{ \int_{-\infty}^{\infty} p(x, t_1)p(z - x, t_2)dx \right\} dz. \tag{1.25}$$

Bachelier then uses the ansatz

$$p(x,t) = A(t)e^{-B(t)^2x^2}. \tag{1.26}$$

By imposing the normalization condition

$$\int_{-\infty}^{\infty} p(x,t)dx = 1, \tag{1.27}$$

we find $B(t) = A(t)\sqrt{\pi}$ and so

$$p(x,t) = A(t)e^{-\pi A(t)^2x^2}. \tag{1.28}$$

We can clearly see this already has the form of the well known Gaussian distribution. Plugging it in equation 1.25 the result will be

$$A(t_1+t_2)^2 = \frac{A(t_1)^2A(t_2)^2}{A(t_1)^2 + A(t_2)^2}. \tag{1.29}$$

By differentiating this equality with both t_1 and t_2 we notice

$$\frac{A'(t_1}{A(t_1)^3} = \frac{A'(t_2)}{A(t_2)^3} = const, \tag{1.30}$$

which then means

$$A(t) = \frac{a}{\sqrt{t}}, \tag{1.31}$$

where a is a constant. Thus we see that we have arrived at the Gaussian distribution of mean 0 and standard deviation $\frac{\sqrt{t}}{a\sqrt{2\pi}}$.

L.Bachelier also makes the connection between this stock price process and the diffusion equation. By imposing on the stock price to change by Δx for any new time increment Δt with equal probability of up or down, he was able to show that if we know the stock price at time $t = 0$ then the probability of finding it again at the same value at time t is governed by the law

$$c^2\frac{dp(x,t)}{dt} - \frac{d^2p(x,t)}{dx^2} = 0, \tag{1.32}$$

where c is a constant. Remarkably, Bachelier then goes on to use these properties to find out the relation between the premium of European options and the expected value of the stock price at expiration date. These results will find their deserving recognition 70 years later when restated in the Black-Scholes-Merton formulation, which leverages the mathematical advancements by Ito, Kolmogorov, Markov, Wiener and others [18].

1.2 Introduction to Stochastic Processes

A word with Greek origin, "stokhastikos" means enabling one to guess or conjecture. The field of stochastic processes concentrates on a family of phenomena about which we cannot make arbitrarily precise predictions, and thus we are left with only guessing the outcomes. In other words, when performing a stochastic or random experiment, we cannot expect any outcome to succeed with probability 1. Stochastic processes concern the dynamics of the probabilities underlying such a random experiment. As opposed to a simple coin toss game, where the probability of a coin falling heads or tails remains constant over time, stochastic processes expand the area of interest to include the concept of time in more complex ways than just labeling consecutive trials of an experiment. Given that experiments are at the very foundation of the scientific method, notation is bound to vary from one field to another. In the next section, some of the notation pertaining to stochastic processes will be defined.

1.2.1 Probability distribution functions

Via a proof by Kolmogorov, a specific stochastic process can be known precisely if and only if we know the complete hierarchy of joint probability distributions [2].

We can introduce the hierarchy of probability densities functions f_n via

$$dP_n\left(x_1,t_1;\ldots;x_n,t_n\right)=f_n\left(x_1,t_1;\ldots;x_n,t_n\right)dx_1\ldots dx_n, \tag{1.33}$$

where P_n represents the joint probability of events $x_1,\ldots,x_n$ happening at times $t_1,\ldots,t_n$. f_n is also called the n-point density.

Consider a time series $x(t)$ representing one run of a stochastic process. Histograms can be built based on the data set which, under certain conditions, should converge to probability densities $f_1(x,t)$, $f_2(x_1,t_1;x_2,t_2)$ *etc.* We note that

$$f_{n-1}\left(x_1,t_1;\ldots;x_{k-1},t_{k-1};x_{k+1},t_{k+1};\ldots;x_nt_n\right)=\int dx_k f_n\left(x_1,t_1;\ldots;x_n,t_n\right), \tag{1.34}$$

so

$$f_1(x,t)=\int dy\ f_2(y,s;x,t). \tag{1.35}$$

Densities of all orders are normalized to unity if $\int dx\ f_1(x,t)=1$. We also need to look into the specifics of when a histograms can be trusted to converge to a probability density. Given we only have a finite set of observations for any experiment we can only hope there are certain distributions that converge fast enough. We will look more closely at this issue later.

1.2.2 Averages and statistics

This subsection is based on Chapter 1 of [2]. Both averages and expectation values can be computed based on the 1-point density. The unconditioned expectation, ensemble average or, simply, average is given by

$$E[x(t)] = \langle x(t) \rangle = \int x f_1(x,t) dx. \tag{1.36}$$

The conditional expectation is defined by

$$E[x(t+T)|x(t)] = \langle x(t+T) \rangle_{cond} = \int y p_{1|1}(y, t+T|x,t) dy. \tag{1.37}$$

The average of a dynamical variable A(x) is defined by

$$\langle A(x) \rangle = \int A(x) f_1(x,t) dx. \tag{1.38}$$

Moments of the distribution are computed via

$$\langle x^n \rangle = \int x^n f_1(x,t) dx. \tag{1.39}$$

The variance is given by

$$Var[x(t)] = \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \int (x - \langle x \rangle)^2 f_1(x,t) dx. \tag{1.40}$$

"Scaling" with Hurst exponent H means that the probability distributions of $x(t)$ and $t^H x(0)$ are the same, which is usually written

$$x(t) = t^H x(0), \tag{1.41}$$

or more formally

$$\langle x^n(t) \rangle = t^{nH} \langle x^n(0) \rangle . \tag{1.42}$$

This can only hold if

$$f_1(x,t) = t^{-H} f_1 \left(x/t^H, 0 \right) = t^{-H} F_1(u), \tag{1.43}$$

with $u = x/t^H$, where

$$\langle x^n(0) \rangle = \int u^n F_1(u) du. \tag{1.44}$$

However, when analyzing a time series the 1-point density is not enough to get a clear view of the dynamics of the process. For this we also need pair correlations, also called

autocorrelation when the same process is sampled twice. To calculate this we need the 2 point density, $f_2(y,t;x,s)$ and thus define the autocorrelation to be

$$\langle x(t)x(s)\rangle = \int x\ y\ f_2(x,t;y,s)dxdy. \tag{1.45}$$

Another useful quantity is the covariance of two random variables $x(t)$ and $y(s)$

$$Cov(x(t)y(s)) = \langle\,(\,x(t) - \langle x(t)\rangle\,)\,(\,y(s) - \langle y(s)\rangle\,)\,\rangle = \langle x(t)y(s)\rangle - \langle x(t)\rangle\,\langle y(s)\rangle\,. \tag{1.46}$$

It can be noticed that the correlation and covariance are equivalent when the mean of one of the stochastic variables vanishes. Next, statistical independence of the variables at different times means that

$$f_n\,(x_n,t_n;\ldots;x_1,t_1) = f_1\,(x_n,t_n)\ldots f_1\,(x_2,t_2)\,f_1\,(x_1,t_1)\,. \tag{1.47}$$

Consider a displacement, or increment $x(t,T) = x(t+T) - x(t)$, where $x(t,-T) = x(t) - x(t-T)$. Calculation of increment correlations

$$\langle x(t,T)x(t,-T)\rangle = \int (x-y)\ (y-z)\ f_3(x,t+T;y,t;z,t-T)\ dxdydz. \tag{1.48}$$

reduces to a sum of pair correlations,

$$\langle x(t,T)x(t,-T)\rangle = \langle x(t+T)x(t)\rangle - \sigma^2(t) - \langle x(t+T)x(t-T)\rangle + \langle x(t)x(t-T)\rangle. \tag{1.49}$$

If the $x(t)$ at different times are statistically independent, then so are the increments

$$\langle x(t,T)x(t,-T)\rangle = \langle x(t,T)\rangle\langle x(t,-T)\rangle. \tag{1.50}$$

1.2.3 Conditional Probability densities

For a random variable $x(t)$ defined as a stochastic process, let $t_1 < \ldots < t_n$. Two point conditional probability densities $p_{1|1}(x|y)$, or transition probability densities, are defined by

$$f_2\,(x_2,t_2;x_1,t_1) = p_{1|1}\,(x_2,t_2\mid x_1,t_1)\,f_1\,(x_1,t_1)\,, \tag{1.51}$$

which can be rewritten to give Bayes's theorem

$$p_{1|1}(x_2,t_2|x_1,t_1) = \frac{f_2(x_2,t_2;x_1,t_1)}{f_1(x_1,t_1)}. \tag{1.52}$$

This can be generalized for higher order

$$f_3\left(x_3,t_3;x_2,t_2;x_1,t_1\right) = p_{1|2}\left(x_3,t_3 \mid x_2,t_2;x_1,t_1\right) f_2\left(x_2,t_2;x_1,t_1\right) = \\ = p_{1|2}\left(x_3,t_3 \mid x_2,t_2;x_1,t_1\right) p_{1|1}\left(x_2,t_2 \mid x_1,t_1\right) f_1\left(x_1,t_1\right), \tag{1.53}$$

and similarly for larger sample space.

Normalization of conditional densities follows easily from

$$\int f_2(y,s;x,t)dydx = \int f_1(x,t)dy = 1, \tag{1.54}$$

and so we obtain

$$\int p_{1|1}(y,s \mid x,t)dy = 1. \tag{1.55}$$

Nevertheless, the question still stands whether it is possible to estimate these densities just by observation.

1.2.4 Characteristic Function

Next we define the moment generating function or the characteristic function, following section 2.6 of [1].

Let u be the vector $(u_1, u_2, \ldots, u_n)$, and $\boldsymbol{X}$ the vector of random variables $(X_1, X_2, \ldots, X_n)$. The characteristic function ϕ is defined by

$$\phi(u) = \langle \exp(iu \cdot \boldsymbol{X}) \rangle = \int d\boldsymbol{x} p(\boldsymbol{x}) e^{iu \cdot \boldsymbol{x})}. \tag{1.56}$$

The characteristic function obeys

$$\phi(0) = 1 \quad , \quad |\phi(u)| \leq 1, \tag{1.57}$$

$\phi(u)$ is a uniformly continuous function of its arguments for all finite real u. If the moments $\langle \prod_t X_i^{m_i} \rangle$ exist, then

$$\left\langle \prod_i X_t^{m_i} \right\rangle = \left[\prod_i \left(-\mathrm{i} \frac{\partial}{\partial u_i} \right)^{m_t} \phi(u) \right]_{u=0}. \tag{1.58}$$

In physics, the characteristic function is often encountered in quantum field theory classes under the name of the correlation function. The correlation function has a central role in QFT, as it is can be used to quantify the interaction between different quantum fields essentially enabling the computation of physical observables. Next, via fourier inversion formula

$$p(x) = (2\pi)^{-n} \int du \phi(u) \exp(-ix \cdot u) \tag{1.59}$$

If the random variables $\boldsymbol{X}$ were independent, in which case

$$p\left(x_1, x_2, \ldots, x_n\right) = p_1\left(x_1\right) p_2\left(x_2\right) \ldots p_n\left(x_n\right), \tag{1.60}$$

then

$$\phi\left(u_1, u_2, \ldots, u_n\right) = \phi_1\left(u_1\right) \phi_2\left(u_2\right) \ldots \phi_n\left(u_n\right). \tag{1.61}$$

Finally, if we define Y as

$$Y = \sum_{i=1}^{n} X_i, \tag{1.62}$$

then the characteristic function of Y is

$$\phi_y(u) = \langle \exp(iuY) \rangle \tag{1.63}$$

and so

$$\phi_y(u) = \prod_{i=1}^{n} \phi_i(u). \tag{1.64}$$

We can see from the previous properties of the characteristic function it essentially encapsulates all the information about the probability distribution. Thus it is capable of computing moments of distribution of any order. Many times the characteristic function is preferred over the conditional probability distribution as it allows for faster proofs. Nonetheless, it is a very useful tool when computing measurable quantities relating to stochastic variables.

1.2.5 Statistical ensembles

A crucial question in studying financial data is whether it is even possible to pin point any part of the process dynamics by using a singular time-series. Obviously, we cannot repeat 100 years of financial history in order to build probability densities describing various financial parameters.

Given that in financial mathematics we always start with the observed market data, then we need to make sure we can confidently infer certain properties of the underlying dynamics. Importantly, this data analysis requires us to use statistical ensembles for model parameter estimation, which can be defined both for stationary and non stationary processes. In order to specify a stochastic model class both the 1-point density and the pair correlations are needed [2] [21].

In [22] it has been shown that time averages can produce fat tails and scaling when we perform the computation on non stationary processes. This is a consequence of the non ergodicity of a non stationary process. Importantly, for analysis on singular time series, the mean square fluctuation, the pair correlation, and scaling parameters are amongst the few measures that converge fast enough (compared to the quantity of data we hold) so as to be confident in our estimates [2]. This is true for both stationary and non stationary processes. Notice, even though the 1-point density is arguably the most important factor

in defining a stochastic process, it cannot be measured in practice as it converges too slow. The intuition behind slow and fast in this context rests on what we can consider roughly separate experiments within the same time series. We know that exchange houses are open 5 days a week and work for roughly 9 hours per day. This gives a regularity of the behavior of market agents. Thus if we take a data set for each day for 5 consecutive years that would roughly give 1500 data points for a statistical ensemble. Without this regularity to human behavior there would be no chance of estimating any parameter with confidence [2].

In our analysis we should avoid assuming that the time averages we can produce are meaningful for the process at hand, unless we prove the process is stationary and ergodic. But, in order to prove stationarity and ergodicity we need to perform an ensemble analysis. This is where regularity in trader behaviour is the key aspect allowing us to build an ensemble.

Importantly, a distinction needs to be made between a stationary process and a process with stationary increments. In physics what we think of as a stationary process is a dynamical system in equilibrium. The equilibrium here stands generally for time translational invariance which leads to the concept of energy conservation. The same cannot be said for a stationary stochastic process where the energy cannot be defined so neatly, and where everything is changing over time. The equivalent stationarity for stochastic processes is encapsulated in the time translational invariance of the transition probabilities from one state to the next. This can be neatly defined in terms of n-point probability densities functions and transition probabilities.

$$f_n\left(x_1, t_1 + T; \ldots; x_n, t_n + T\right) = f_n\left(x_1, t_1; \ldots; x_n, t_n\right), \tag{1.65}$$

$$p_{1|1}\left(x_n, t_n \mid x_{n-1}, t_{n-1}\right) = p_{1|1}\left(x_n, t_n - t_{n-1} \mid x_{n-1}, 0\right), \tag{1.66}$$

as well as for higher order joint probability distributions.

Because of the normalization condition at equation 1.54, the key probability density function to be constrained is the 1-point density function. This needs to be time translational invariant and normalizable. As a result of these properties both the mean and variance and all higher distribution moments are similarly time translation invariant [2]. Thus, a stochastic stationary process is one for which the mean and all higher distribution do not change with time, on average. As useful as process stationarity may be however, it leads to long memory in the autocorellation function which is an indicative of a non efficient market where profit can be made consistently.

A condition which allows for more freedom in defining the process is increment stationarity [2] [21] [23]. For this to happen, only the 1-point increment probability density function needs to be constrained. In contrast, process stationarity implies time translational invariance to probability density functions and transition densities at all orders, a much more strict constraint. Increment stationarity can thus be stated as $x(t+T) - x(t) = x(t, T)$ has a probability density function which is independent of starting time t.

$$-2\langle x(t+T)x(t)\rangle = \left\langle (x(t+T) - x(t))^2 \right\rangle - \left\langle \left(x^2(t+T)\right\rangle - \left\langle x^2(t)\right\rangle \right. \tag{1.67}$$

and then use increment stationarity $x(t,T) = x(0,\ T)$, meaning these two variables have the same probability distribution, to obtain

$$-2\langle x(t+T)x(t)\rangle = \left\langle \left(x^2(0,T)\right\rangle - \left\langle \left(x^2(t+T)\right\rangle - \left\langle x^2(t)\right\rangle \right.\right. . \tag{1.68}$$

This pair correlation generally does not vanish [2]. For increments with nonoverlapping time intervals the increment autocorrelation function is

$$\begin{aligned} 2\langle (x(t,T)x(t,-T)\rangle \quad &= \left\langle (x(t+T)-x(t-T))^2\right\rangle - \left\langle (x(t)-x(t-T))^2\right\rangle - \\ - \left\langle (x(t+T)-x(t))^2\right\rangle \quad &= \left\langle x^2(0,2\ T)\right\rangle - 2\left\langle x^2(0,\ T)\right\rangle . \end{aligned} \tag{1.69}$$

If $x(0) = 0$ we get

$$\langle x(t,T)x(t,-T)\rangle = \sigma^2(2\ T) - 2\sigma^2(\ T). \tag{1.70}$$

Notice that if the variance is linear with respect to time t then the increment autocorellation would vanish. This is true for Brownian motion, however, any departure from that would create slow decay of the autocorellation function. This means again that profit can be made based on the history of past returns, if the process is nonstationary. Again, we are hit with the realisation that even something as simple as increment stationarity leads to an ineffective market. Surprisingly however, the most general representation of an efficient market is not based on a stationary Brownian process, but on processes where increments are nonstationary martingales [24], which illuminates us on the ubiquity of martingales in financial mathematics. Importantly, martingale processes can be described by diffusion type equations [2]. Thus, the connection between diffusion, stationary processes and efficient markets is brought to light. Next, we shall look at some of the simplest stochastic processes.

1.2.6 Coin Toss Game

Before people became interested in quantifying errors related to experimental outcomes they used probabilistic concepts to help their gambling pursuits. Many enlightenment figures such as Fermat, Pascal, Bernoulli considered probabilistic problems in terms of *fair games* under different conditions. One of the simplest such fair games is a *Coin Toss* game. Let us assume the probability of a coin toss resulting in heads is p, while a tails outcome happens with probability $1-p$. The game which we are going to focus on is a simple betting game. If the coin toss results in heads then the player wins one wealth unit, otherwise the player loses one unit of his wealth. We are going to look at the wealth $X(n)$ after n coin tosses. Given the outcomes of the coin tosses are all independent of one another we can use the binomial or Bernoulli distribution [2]. The probability of x successes out of n total coin tosses is then

$$p_{X(n)}(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \quad x = 0,1,2,\ldots,n \tag{1.71}$$

We can then easily compute the expectation for successful trials after n tosses to be

$$E[X(n)] = np. \tag{1.72}$$

Taking into account expected losses, the change in wealth after n tosses has the expectation of $n(2p-1)$. We can clearly see if the coin is fair the wealth change is nil, and thus, in this case, the game is a martingale. We can also compute the variance of successes to be

$$Var\left[X(n)\right] = np(1-p). \tag{1.73}$$

If we rescale the process such that we consider $Z_n = \frac{X_n}{\sqrt{Var[X_n]}}$ then the asymptotic behaviour of the distribution will tend to the Gaussian distribution in the limit $n \to \infty$ [25]. To note, the interested reader may find it interesting that the term martingale comes from a very simple betting strategy related to coin tosses. The player starts with a bet value, and bets on heads. If the coin falls heads up then the player wins his bet, else he loses. The strategy implies the player doubles his bet after every loss such that upon the first upcoming win, the player would regain all loses.

1.3 Markov processes

In its simplest expression, a system undergoing a Markov process is one for which the future state can only be predicted based on the current state, all past states being irrelevant. Take $\mathbf{X}(t)$ to be a time dependent random variable and $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$ measured values at different times. Assume a set of joint probability densities exist $f_n(\mathbf{x_1}, t_1; \mathbf{x_2}, t_2...)$, which describe the system completely.

Assuming $t_1 < t_2 < t_3$ then

$$f_3(\mathbf{x_1}, t_1; \mathbf{x_2}, t_2; \mathbf{x_3}, t_3) = p_{1|2}(\mathbf{x_3}, t_3|\mathbf{x_2}, t_2; \mathbf{x_1}, t_1) f_2(\mathbf{x_2}, t_2; \mathbf{x_1}, t_1), \tag{1.74}$$

by using equation 1.53, further removing long term dependence other than the last event on $p_{1|2}(\mathbf{x_3}, t_3|\mathbf{x_2}, t_2; \mathbf{x_1}, t_1)$ and using Bayes's theorem on $f_2(\mathbf{x_2}, t_2; \mathbf{x_1}, t_1)$ we get

$$f_3(\mathbf{x_1}, t_1; \mathbf{x_2}, t_2; \mathbf{x_3}, t_3) = p_{1|1}(\mathbf{x_3}, t_3|\mathbf{x_2}, t_2) p_{1|1}(\mathbf{x_2}, t_2|\mathbf{x_1}, t_1) f_1(\mathbf{x_1}, t_1), \tag{1.75}$$

which is the defining property of a Markov process. This can be generalized for any number of states $t_1 < t_2 < ...t_{n-1} < t_n$ such that

$$\begin{aligned} &f_n(\mathbf{x_1}, t_1; \mathbf{x_2}, t_2; ...; \mathbf{x_{n-1}}, t_{n-1}; \mathbf{x_n}, t_n) = \\ &= p(\mathbf{x_n}, t_n|\mathbf{x_{n-1}}, t_{n-1})....p(\mathbf{x_3}, t_3|\mathbf{x_2}, t_2) p(\mathbf{x_2}, t_2|\mathbf{x_1}, t_1) f_1(\mathbf{x_1}, t_1) \end{aligned} \tag{1.76}$$

1.3.1 Continuity conditions

This section is based on section 3.3 of [1]. Continuity of a stochastic variable X(t) can be looked at in both the continuity of the domain and the continuity of the sample path. To contrast these 2 concepts, take the model of a perfect gas. The molecules themselves can potentially have any speed, but because the gas molecules have instantaneous impacts the speed of the molecule cannot change continuously. However, in the same model we can consider the location of the molecule to vary continuously. This problem could be alleviated if we looked into the details of the impact. In this scenario, the recent history of the entire system would be required for making any sort of prediction. However, this would certainly remove the underlying property of a Markov process. As a result, a Markov process is based on the assumption that there is a certain memory time, beyond which past history becomes irrelevant. So observation of the system needs to happen on the scale where past history is too far behind to have any impact on future predictions.

Nevertheless, Markov processes which display continuous sample paths exist, and are more useful mathematically. More precisely, the sample paths are continuous functions of time t, if for any $\epsilon > 0$:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} d\mathbf{x} p_{1|1}(\mathbf{x}, t+\Delta t \mid \mathbf{z}, t) = 0 \tag{1.77}$$

uniformly in $\mathbf{z}$, t and dt. This very condition means that the probability for the final position $\mathbf{x}$ to be different to $\mathbf{z}$ goes to 0 faster than Δt. Define

$$\lim_{\Delta t \to 0} \frac{p_{1|1}(\mathbf{x}, t+\Delta t|\mathbf{z}, t)}{\Delta t} = W(\mathbf{x}|\mathbf{z}, t). \tag{1.78}$$

The $W(\mathbf{x}|\mathbf{z}, t)$ function represents the probability distribution of a discontinuous jump from position $\mathbf{x}$ to position $\mathbf{z}$. Without jumps the stochastic process is continuous. Nevertheless, jumps are useful for a variety of applications and can be implemented into stochastic processes, but for this section we will not consider them. Next, by introducing further assumptions on the probability density of the process we can write an evolution equation which takes into account both continuity and differentiability and splits the motion into continuous and discontinuous parts. We require the following conditions for all $\varepsilon > 0$:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}|<\epsilon} d\mathbf{x} \, (x_i - z_i) \, p_{1|1}(\mathbf{x}, t+\Delta t \mid \mathbf{z}, t) = A_i(\mathbf{z}, t) + O(\epsilon); \tag{1.79}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} d\mathbf{x} \, (x_i - z_i)(x_j - z_j) \, p_{1|1}(\mathbf{x}, t+\Delta t \mid \mathbf{z}, t) = B_{ij}(\mathbf{z}, t) + O(\varepsilon) \tag{1.80}$$

uniformly in $\mathbf{x}, \mathbf{z}$ and t for $|\mathbf{x}-\mathbf{z}| \geq \epsilon$. All integrals in higher powers of $\mathbf{x}$ vanish. For the process to be fully continuous, then $W(\mathbf{x}|\mathbf{z}, t)$ vanishes for all $\mathbf{x}$ different from $\mathbf{z}$. Thus, this function encapsulated the discontinuous part of the stochastic process, whilst A_i and B_{ij} are related to the continuous part.

1.3.2 Chapman-Kolmogorov equation

The continuity conditions, together with the Markov process condition on joint probability distributions can be used to obtain the forward time Chapman-Kolmogorov Equation:

$$\partial_t p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right) = -\sum_t \frac{\partial}{\partial z_i}\left[A_i(\mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right] + \sum_{t,j}\frac{1}{2}\frac{\partial^2}{\partial z_i \partial z_j}\left[B_{ij}(\mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right]$$

$$+\int d\mathbf{x}\left[W(\mathbf{z} \mid \mathbf{x},t)p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right) - W(\mathbf{x} \mid \mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right]. \tag{1.81}$$

For a derivation of the Chapman-Kolmogorov equation please refer to Section 3.4 of [1]. There is also a backward time Chapman-Kolmogorov equation. These two are equivalent, although they are used for different purposes. To get the backward time version we need to first consider the first derivative of the joint probability distribution

$$\lim_{\Delta t' \to 0}\frac{1}{\Delta t'}\left[p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'+\Delta t'\right) - p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right)\right] =$$

$$\lim_{\Delta t' \to 0}\frac{1}{\Delta t'}\int dz p_{1|1}\left(\mathbf{z},t'+\Delta t' \mid \mathbf{y},t'\right)\left[p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'+\Delta t'\right) - p_{1|1}\left(\mathbf{x},t \mid \mathbf{z},t'+\Delta t'\right)\right] \tag{1.82}$$

by use of the Chapman-Kolmogrov equation in the second term and by noting that the first term gives $1 \times p_{1|1}(x,t \mid y,t'+\Delta t')$. The assumptions that are necessary are now the existence of all relevant derivatives, and that $p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},\boldsymbol{t}'\right)$ is continuous and bounded in $\mathbf{x},t,t'$ for some range $t-t'>\delta>0$. We may then write

$$= \lim_{\Delta t' \to 0}\frac{1}{\Delta t'}\int dz p_{1|1}\left(\mathbf{z},t'+\Delta t' \mid \mathbf{y},t'\right)\left[p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right) - p_{1|1}\left(\mathbf{x},t \mid \mathbf{z},t'\right)\right] \tag{1.83}$$

By similar manipulation as those used to arrive at the forward time Chapman-Kolmogorov equation we arrive at

$$\frac{\partial p_{1|1}\left(\boldsymbol{x},t \mid \mathbf{y},t'\right)}{\partial t'} = -\sum_i A_r\left(\mathbf{y},t'\right)\frac{\partial p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right)}{\partial y_i} -$$

$$-\frac{1}{2}\sum_{ij}B_{ij}\left(\mathbf{y},t'\right)\frac{\partial^2 p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right)}{\partial y_l \partial y_j} + \int dz W\left(\mathbf{z} \mid \mathbf{y},t'\right)\left[p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right) - p_{1|1}\left(\mathbf{x},t \mid \mathbf{z},t'\right)\right] \tag{1.84}$$

which is called the backward differential Chapman-Kolmogorov equation. The appropriate initial condition for both Chapman-Kolmogorov equations is $p_{1|1}(\mathbf{x},t \mid \mathbf{y},t) = \delta(\mathbf{x}-\mathbf{y})$ for all t.

The two equations can turn into one another depending on the boundary conditions. The forward time equation is best suited for process that evolve in time, for which we provide an *initial condition*, whilst the backward time equation is most useful for *first passage time* or *exit problems* for which we provide a *final condition*.

1.3.3 Processes obeying Chapman-Kolmogorov dynamics

There are 3 terms on the r.h.s each which give rise to 3 distinct features known as jumps, drift and diffusion.

Jump Processes

Assume

$$A_l(\mathbf{z},t) = B_{ij}(\mathbf{z},t) = 0, \tag{1.85}$$

so we are left with

$$\partial_t p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right) = \int d\mathbf{x}\left[W(\mathbf{z} \mid \mathbf{x},t)p_{1|1}\left(\mathbf{x},t \mid \mathbf{y},t'\right) - W(\mathbf{x} \mid \mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right]. \tag{1.86}$$

The first order approximation of the derivative of the probability distribution with initial condition $p_{1|1}(\mathbf{z},t \mid \mathbf{y},t) = \delta(\mathbf{y}-\mathbf{z})$ is given by

$$p_{1|1}(\mathbf{z},t+\Delta t \mid \mathbf{y},t) = \delta(\mathbf{y}-\mathbf{z}) + W(\mathbf{z} \mid \mathbf{y},t)\Delta t + O(\Delta t^2), \tag{1.87}$$

which plugged in equation 1.86 gives

$$p_{1|1}(\mathbf{z},t+\Delta t \mid \mathbf{y},t) = \delta(\mathbf{y}-\mathbf{z})\left[1 - \int d\mathbf{x} W(\mathbf{x} \mid \mathbf{y},t)\Delta t\right] + W(\mathbf{z} \mid \mathbf{y},t)\Delta t. \tag{1.88}$$

The last expression shows there is a finite probability of the particle to stay at the same position and also a finite probability of the particle to make a discontinuous jump as dictated by $W(\mathbf{z} \mid \mathbf{y},t)$. Thus, the plot of the particle evolution will look like straight lines connected by discontinuous jumps.

Fokker-Planck equation

This time, we assume $W(\mathbf{z}|\mathbf{x},t) = 0$ to get the *Fokker-Planck* equation

$$\begin{aligned}\frac{\partial p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)}{\partial t} &= -\sum_i \frac{\partial}{\partial z_i}\left[A_i(\mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right] \\ &+\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial z_i \partial z_j}\left[B_{ij}(\mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right].\end{aligned} \tag{1.89}$$

This is called a *diffusion process* and it is strikingly similar to the equation 1.20, the main difference being that previously we considered the diffusion as isotropic. The vector $\mathbf{A}_i(\mathbf{z},t)$ is called the drift vector while $\mathbf{B}_{ij}(\mathbf{z},t)$ is called the diffusion matrix, which needs to be positive semidefinite and symmetric. Given we removed the discontinuity of the jump process, this process describes a continuous variable, however not differentiable. Consider the same initial condition as previously

$$p_{1|1}(\mathbf{z},t \mid \mathbf{y},t) = \delta(\mathbf{y}-\mathbf{z}) \tag{1.90}$$

The first order approximation of the derivative of the probability distribution vanishes, because $W(\mathbf{z}|\mathbf{x},t)=0$. Considering the removal of the time dependence of $\mathbf{A}_i(\mathbf{z},t)$ and $\mathbf{B}_{ij}(\mathbf{z},t)$ for small Δt we can then write

$$\frac{\partial p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)}{\partial t} = -\sum_i A_i(\mathbf{y},t)\frac{\partial p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)}{\partial z_i} + \sum_{i,j}\frac{1}{2}B_{ij}(\mathbf{y},t)\frac{\partial^2 p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)}{\partial z_i \partial z_j}, \tag{1.91}$$

which can be solved to give a Gaussian distribution with variance matrix $\mathbf{B}(\mathbf{y},t)$ and mean $\mathbf{y}+A(\mathbf{y},t)\Delta t$. This solution is continuous as $\Delta t \to 0, \mathbf{y}(t+\Delta t) \to \mathbf{y}(t)$ but nowhere differentiable because of a $\Delta t^{1/2}$ occuring in the solution.

Liouville's equation

Assume $W(\mathbf{z}|\mathbf{x},t)=0$ and $B_{ij}(\mathbf{z},t)=0$. Thus

$$\frac{\partial p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}\left[A_i(\mathbf{z},t)p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right)\right]. \tag{1.92}$$

which occurs in classical mechanics. This equation describes a completely deterministic motion i.e., if $\mathbf{x}(\mathbf{y},t)$ is the solution of the ordinary differential equation

$$\frac{d\mathbf{x}(t)}{dt} = A[\mathbf{x}(t),t] \tag{1.93}$$

with $\mathbf{x}\left(\mathbf{y},t'\right)=\mathbf{y}$, then the solution to 1.92 with initial condition

$$p_{1|1}\left(\mathbf{z},t' \mid \mathbf{y},t'\right) = \delta(\mathbf{z}-\mathbf{y}) \tag{1.94}$$

is

$$p_{1|1}\left(\mathbf{z},t \mid \mathbf{y},t'\right) = \delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]. \tag{1.95}$$

The proof of this assertion is best obtained by direct substitution. For

$$-\sum_t \frac{\partial}{\partial z_t}\left\{A_t(\mathbf{z},t)\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]\right\} = -\sum_t \frac{\partial}{\partial z_t}\left\{A_r[\mathbf{x}(\mathbf{y},t),t]\delta[\mathbf{z}-\mathbf{x}(\mathbf{y},t)]\right\}$$

$$= -\sum_t \left\{ A_t[\mathbf{x}(\mathbf{y},t),t] \frac{\partial}{\partial z_t} \delta[\mathbf{z} - \mathbf{x}(\mathbf{y},t)] \right\} \tag{1.96}$$

and

$$\frac{\partial}{\partial t}\delta[\mathbf{z} - \mathbf{x}(\mathbf{y},t)] = -\sum_t \frac{\partial}{\partial z_t}\delta[\mathbf{z} - \mathbf{x}(\mathbf{y},t)] \frac{dx_t(\mathbf{y},t)}{dt}, \tag{1.97}$$

and hence satisfies 1.92. Thus, if the particle is in a well-defined initial position $\mathbf{y}$ at time t', it stays on the trajectory obtained by solving the ordinary differential equation 1.92. Notice that for Liouville's equation the Gaussian part is 0, and thus completely deterministic motion is allowed. Notably, deterministic motion is an elementary type of Markov process.

Wiener Process

Assume there is no drift and no jumps and that the diffusion matrix is equal to unity and take the 1d case. As a result

$$\frac{\partial}{\partial t} p_{1|1}\left(z,t \mid x,t'\right) = \frac{1}{2}\frac{\partial^2}{\partial z^2} p_{1|1}\left(z,t \mid x,t'\right). \tag{1.98}$$

By taking the initial condition to be $p_{1|1}(z,t \mid y,t) = \delta(y-z)$ and using a Fourier transform we can easily arrive at the solution

$$p_{1|1}\left(z,t \mid x,t'\right) = \left[2\pi\left(t-t'\right)\right]^{-1/2} \exp\left[-\left(z-x\right)^2 / 2\left(t-t'\right)\right]. \tag{1.99}$$

This is a Gaussian distribution with mean $\langle W(t)\rangle = x$ and variance $\left\langle [W(t)-x]^2 \right\rangle = t - t'$. The characteristic function of the Wiener process is

$$\phi(u,t) = \exp\left[iux_0 - \frac{1}{2}u^2\left(t - t_0\right)\right], \tag{1.100}$$

when the Wiener process starts at x_0, t_0.

The one dimensional Wiener process is often called Brownian motion as it complies with the same ordinary differential equation Einstein derived to describe Brownian motion. Importantly, for a Wiener process the sample paths are everywhere continuous but nowhere differentiable and its variance tends to ∞ as $t \to \infty$.

1.3.4 Stochastic Differential Equations (SDE)

Given the simplicity of Langevin's formulation of Brownian motion, but the introduction of a stochastic force $X(t)$ poorly defined as highly irregular and independent to the Brownian particle position $x(t)$, we try to put it on a more stable footing. Langevin's equation can be written as

$$\frac{dx}{dt} = a(x,t) + b(x,t)\xi(t), \tag{1.101}$$

where $a(x,t)$ and $b(x,t)$ are known functions and $\xi(t)$ is the rapidly fluctuating term. A formulation Langevin's assumption on the stochastic force would be to impose statistical independence of $\xi(t)$ and $\xi(t')$ for any $t \neq t'$. Moreover, $\langle \xi(t) \rangle = 0$ as any nonzero can be absorbed into the other two known functions. Statistical independence can be expressed as

$$\langle \xi(t) \xi(t') \rangle = \delta(t - t'), \tag{1.102}$$

which complies with no correlation at different times and which unfortunately has the consequence of making the variance of $\xi(t)$ infinite. To alleviate this problem start with the differential equation 1.101 and integrate with respect to time to get

$$x(t) - x(0) = \int_0^t a[x(s), s]ds + \int_0^t b[x(s), s]\xi(s)ds. \tag{1.103}$$

This form of Langevin equation can be interpreted consistently. To see why the differential equation from is not fully sound let's have a look at the integral of the stochastic term

$$u(t) = \int_0^t dt' \xi(t'), \tag{1.104}$$

which we assume exists. We require $u(t)$ to be continuous. This property is necessary to establish what differential equation dictates the evolution of $u(t)$. To make progress notice

$$u(t') = \int_0^t ds\xi(s) + \int_t^{t'} ds\xi(s) = \lim_{\varepsilon \to 0} \left[\int_0^{t-\varepsilon} ds\xi(s) \right] + \int_t^{t'} ds\xi(s) \tag{1.105}$$

where $\epsilon > 0$. Remember $\xi(s)$ from the first and second integrals are statistically independent. As $u(t)$ is continuous then $u(t)$ and $u(t') - u(t)$ are statistically independent for $t' > t$ and also that $u(t') - u(t)$ is independent of $u(t'')$ for $t'' < t'$. This essentially states that $u(t)$ is a Markov process. A Markov process which is also continuous can be described via the Fokker-Planck equation. We can deduce the drift and diffusion coefficients through the assumptions made in equations 1.79 and 1.80.

$$\langle u(t + \Delta t) - u_0 \mid [u_0, t] \rangle = \left\langle \int_t^{t+\Delta t} \xi(s)ds \right\rangle = 0 \tag{1.106}$$

and

$$\begin{aligned} \left\langle [u(t + \Delta t) - u_0]^2 \mid [u_0, t] \right\rangle &= \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \langle \xi(s) \xi(s') \rangle \\ &= \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \delta(s - s') = \Delta t \end{aligned} \tag{1.107}$$

which leads to the following drift and diffusion coefficients

$$A(u_0, t) = \lim_{\Delta t \to 0} \frac{\langle u(t + \Delta t) - u_0 \mid [u_0, t] \rangle}{\Delta t} = 0 \tag{1.108}$$

$$B\left(u_0,t\right)=\lim_{\Delta t\to 0}\frac{\left\langle\left[u(t+\Delta t)-u_0\right]^2\mid[u_0,t]\right\rangle}{\Delta t}=1. \tag{1.109}$$

This is precisely the Fokker-Planck equation of the Wiener process and so

$$\int_0^t\xi\left(t'\right)dt'=u(t)=W(t). \tag{1.110}$$

The pathological result is that the derivative of $W(t)$ turns out to be $\xi(t)$. But $W(t)$ is not differentiable, so clearly the differential equation 1.80 is not well defined. Instead of a differential equation we can alternatively write the Wiener process as

$$dW(t)\equiv W(t+dt)-W(t)=\xi(t)dt \tag{1.111}$$

which flows from the integral version of the equation 1.80 rather than the differential one. The integral version of the Langevin equation is, however, self consistent.

1.3.5 Existence conditions

A random variable $x(t)$ is said to obey an Ito Stochastic Differential Equation (SDE)

$$dx(t)=a[x(t),t]dt+b[x(t),t]dW(t), \tag{1.112}$$

iff for all t,t_0 the variable $x(t)$ obeys

$$x(t)-x(t_0)=\int_0^t a[x(s),s]ds+\int_0^t b[x(s),s]dW(t). \tag{1.113}$$

Now, let's have a closer look into the properties of the parameters in the integral form of the SDE, namely $a[x(s),s]$ and $b[x(s),s]$. The existence of a solution to the SDE is guaranteed by the Lipschitz and growth conditions. The former states that a K exists such that

$$|a(x,t)-a(y,t)|+|b(x,t)-b(y,t)|\leq K|x-y| \tag{1.114}$$

for all x and y, and all t in the range $[t_0,T]$. The growth condition on the other hand states that a K exists such that for all t in the range $[t_0,T]$,

$$|a(x,t)|^2+|b(x,t)|^2\leq K^2\left(1+|x|^2\right). \tag{1.115}$$

The Lipschitz condition is met by most processes in practice. In contrast, the growth condition, which does not allow the variable $x(t)$ to grow arbitrarily large in a finite time is often violated. However, there are alternative conditions that ensure the solution to the SDE has a well behaved growth [26].

Rather remarkably, the solution of the quite general Ito SDE is a Markov process. This is true because the solution of $x(t)$ is determined only from the initial condition $x(t_0)$ and the particular instance of $W(t)$. Then $x(t)$ is independent from $x(t')$ for $t' < t$, or it is a Markov process [26].

1.3.6 Ito's lemma

Start with the Ito SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dW(t). \tag{1.116}$$

We now consider the time development of an arbitrary well-behaved $f(x(t))$. Let's look at a change of variable first

$$df[x(t)] = f[x(t) + dx(t)] - f[x(t)] = f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)]dx(t)^2 + \dots \tag{1.117}$$

Let us focus on the second differential term. Expanded this term looks like

$$dx(t)^2 = a^2[x(t), t]dt^2 + 2a[x(t), t]b[x(t), t]dW(t)dt + b^2[x(t), t](dW(t))^2. \tag{1.118}$$

We are only interested in terms linear in dt. Let us focus on the last term $(dW(t))^2$. To get some insight let us focus on

$$I = \int_0^t dW^2 \approx \sum_{k=1}^{N} \Delta W_k^2. \tag{1.119}$$

The variance of I is

$$\sigma_1^2 = N\left(\left\langle \Delta W^4 \right\rangle - \left\langle \Delta W^2 \right\rangle^2\right). \tag{1.120}$$

As a result of Gaussianity of the increments ΔW, we obtain $\langle \Delta W^4 \rangle = 3\Delta t^2$. By taking $\Delta T = t/N$ we get $\sigma_I^2 = 2t^2/N$ and thus $lim_{N\to\infty}\sigma_1^2 = 0$. The variance of ΔW^2 vanishes which means I is a delta function. This means the $(dW)^2$ is deterministic, or otherwise stated

$$(dW)^2 = dt. \tag{1.121}$$

An alternative, heuristic approach to this proof, would be to consider random walks in 1 dimension. Imagine a particle taking a step of either ± 1 with equal probability at time $t > 0$, where t is a natural number. If we take W_t to be the sum of the number of integer steps up until time t then the expectation is clearly $E[W_t] = 0$. However, consider W_t^2 and it is clear that $W_t^2 = t$, because t steps are made. At first sight, we could say

$|W_t| = O(\sqrt{t})$. So in Δt steps, $W_t = O(\sqrt{\Delta t})$. As $\Delta t \to 0, W_t^2 \approx \Delta t$, since increments are infinitesimal.

As a result

$$dx(t)^2 = b^2[x(t),t](dW(t))^2, \tag{1.122}$$

which leads to equation 1.117 to look like

$$\begin{aligned} &= f'[x(t)]\{a[x(t),t]dt + b[x(t),t]dW(t)\} + \frac{1}{2}f''[x(t)]b[x(t),t]^2[dW(t)]^2 + \dots \\ &= \left\{a[x(t),t]f'[x(t)] + \frac{1}{2}b[x(t),t]^2 f''[x(t)]\right\}dt + b[x(t),t]f'[x(t)]dW(t), \end{aligned} \tag{1.123}$$

This is the key tool of Ito calculus. As alternative formulations of the SDEs such as Stratonovich integral or differential formulations are too cumbersome to use when deriving results, the simplicity of Ito's formula becomes very useful.

1.3.7 Connection with Fokker Planck equation

Using equation 1.123, computing the expectation value of the differential gives

$$\langle df[x(t)]\rangle/dt = \left\langle \frac{df[x(t)]}{dt} \right\rangle = \frac{d}{dt}\langle f[x(t)]\rangle = \left\langle a[x(t),t]\partial_x f + \frac{1}{2}b[x(t),t]^2\partial_x^2 f \right\rangle, \tag{1.124}$$

as $\langle W(t)\rangle = 0$. Take $x(t)$ to have conditional probability density $p_{1|1}(x,t \mid x_0,t_0)$ which leads to

$$\begin{aligned} \frac{d}{dt}\langle f[x(t)]\rangle &= \int dx f(x)\partial_t p_{1|1}(x,t \mid x_0,t_0) = \\ &= \int dx \left[a(x,t)\partial_x f + \frac{1}{2}b(x,t)^2\partial_x^2 f\right] p_{1|1}(x,t \mid x_0,t_0). \end{aligned} \tag{1.125}$$

Performing an integration by parts and canceling boundary terms we get

$$\begin{aligned} &\int dx f(x)\partial_t p_{1|1}(x,t \mid x_0,t_0) = \\ &= \int dx f(x)\left\{-\partial_x[a(x,t)p_{1|1}(x,t \mid x_0,t_0)] + \frac{1}{2}\partial_x^2\left[b(x,t)^2 p_{1|1}(x,t \mid x_0,t_0)\right]\right\} \end{aligned} \tag{1.126}$$

and because we consider $f(x)$ as arbitrary this leads to

$$\partial_t p_{1|1}(x,t \mid x_0,t_0) = -\partial_x\left[a(x,t)p_{1|1}(x,t \mid x_0,t_0)\right] + \frac{1}{2}\partial_x^2\left[b(x,t)^2 p_{1|1}(x,t \mid x_0,t_0)\right] \tag{1.127}$$

Amazingly, we started with a Ito stochastic differential equation and we arrived at a diffusion process with drift coefficient $a(x,t)$ and a diffusion coefficient $b(x,t)^2$. We thus

see that all Markov processes, which obey a Fokker-Planck equation, can be stated as a Ito SDE and are essentially diffusion processes. Nevertheless, diffusion is obeyed by a larger family of processes than just Markov processes, although these will not have an associated Ito SDE.

1.3.8 Examples of SDE

Given its ubiquity, let us focus on some particular cases of the Ito SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dW(t). \tag{1.128}$$

Arithmetic Brownian Motion

We define Arithmetic Brownian motion [27] [28] to be a process $\{X(t), t \geq 0\}$ described by Brownian motion with drift coefficient μ and variance parameter σ^2 such that

1. $X(0) = 0$;
2. $\{X(t), t \geq 0\}$ has stationary and independent increments;
3. $X(t)$ is normally distributed with mean μt and variance $t\sigma^2$.

The SDE describing this process is

$$dX(t) = \mu dt + \sigma dW(t), \tag{1.129}$$

and the 1 point density function is given by

$$f_1(x, t) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left[-\frac{1}{2\sigma^2 t}(x - \mu t)^2\right]. \tag{1.130}$$

The characteristic function is

$$\phi(u, t) = exp(iu\,(x_0 + \mu t))e^{-\frac{1}{2}u^2\sigma^2 t}, \tag{1.131}$$

if the initial conditions are $(x_0, t_0 = 0)$. The autocorrelation is easily computed. For $t > s$ we have

$$\langle X(t)X(s)\rangle = \langle (X(t) - X(s))X(s)\rangle + \langle X(s)^2\rangle = \langle X(s)^2\rangle = \sigma^2 s. \tag{1.132}$$

Clearly, if we were to set the drift coefficient $\mu = 0$ and the standard deviation $\sigma = 1$ then we would get the standard Brownian motion or the Wiener process.

Geometric Brownian Motion

One property of the Arithmetic Brownian motion is that it allows generation of negative values. A geometric Brownian motion, on the other hand, does not [25] [29] . This is a clear advantage for financial modelling and will indeed see its application later on. A geometric Brownian motion $X(t)$ has the following SDE

$$dX(t) = \alpha X(t)dt + \sigma X(t)dW(t) \quad \text{or} \quad \frac{dX(t)}{X(t)} = \alpha dt + \sigma dW(t), \tag{1.133}$$

where both α and σ are constants. The solution to the differential equation is

$$X(t) = X(0)e^{\left(\alpha - \frac{\alpha^2}{2}\right)t + \sigma W(t)} \tag{1.134}$$

We can easily arrive back to the original SDE via Ito's lemma. Note that $ln(X(t))$ is a lognormal distribution with mean

$$E[\ln(X(t)] = \ln X(0) + \left(\alpha - \sigma^2/2\right)t, \tag{1.135}$$

and variance

$$Var[\ln X(t)] = \sigma^2 t \tag{1.136}$$

Note that

$$E[X(t)] = X(0)e^{\alpha t} \tag{1.137}$$

If $\{Y(t), t \geq 0\}$ is a Brownian motion process with drift coefficient μ and variance parameter σ^2, then the process $\{X(t), t \geq 0\}$ defined by

$$X(t) = e^{Y(t)}$$

is called geometric Brownian motion. For a geometric Brownian motion process $\{X(t)\}$, let us compute the expected value of the process at time t given the history of the process up to time s. That is, for $s < t$, consider $E[X(t) \mid X(u), 0 \leq u \leq s]$. Now,

$$\begin{aligned} E[X(t) \mid X(u), 0 \leq u \leq s] &= E\left[e^{Y(t)} \mid Y(u), 0 \leq u \leq s\right] = \\ &= E\left[e^{Y(s)+Y(t)-Y(s)} \mid Y(u), 0 \leq u \leq s\right] = \\ &= e^{Y(s)} E\left[e^{Y(t)-Y(s)} \mid Y(u), 0 \leq u \leq s\right] = \\ &= X(s) E\left[e^{Y(t)-Y(s)}\right] \end{aligned} \tag{1.138}$$

where the next to last equality follows from the fact that $Y(s)$ is given, and the last equality from the independent increment property of Brownian motion. Now, the moment generating function of a normal random variable W is given by

$$\phi(u,t) = E\left[e^{uW}\right] = e^{uE[W] + u^2 \operatorname{Var}(W)/2} \tag{1.139}$$

Hence, since $Y(t) - Y(s)$ is normal with mean $\mu(t-s)$ and variance $(t-s)\sigma^2$, it follows by setting $u = 1$ that

$$E\left[e^{Y(t)-Y(s)}\right] = e^{\mu(t-s)+(t-s)\sigma^2/2} \tag{1.140}$$

Thus, we obtain

$$E[X(t) \mid X(u), 0 \leq u \leq s] = X(s)e^{(t-s)\left(\mu+\sigma^2/2\right)} \tag{1.141}$$

Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process has the following SDE

$$dX(t) = -kX(t)dt + \sqrt{D}dW(t). \tag{1.142}$$

In order to find a solution to this SDE we use the following stochastic variable

$$y = X(t)e^{kt}. \tag{1.143}$$

Next, we look for the differential of y by using Ito's Lemma,

$$dy = (dX(t))d\left(e^{kt}\right) + (dX(t))e^{kt} + X(t)d\left(e^{kt}\right) =$$

$$= [-kX(t)dt + \sqrt{D}dW(t)]ke^{kt}dt + [-kX(t)dt + \sqrt{D}dW(t)]e^{kt} + kX(t)e^{kt}dt, \tag{1.144}$$

as a result,

$$dy = \sqrt{D}\mathrm{e}^{kt}dW(t). \tag{1.145}$$

By integrating this expression and using equation 1.143 we arrive at the solution

$$X(t) = X(0)\mathrm{e}^{-kt} + \sqrt{D}\int_0^t \mathrm{e}^{-k(t-t')}dW\left(t'\right). \tag{1.146}$$

If the initial condition is deterministic or Gaussian distributed [1], then $x(t)$ is Gaussian with mean and variance

$$\langle X(t)\rangle = \langle X(0)\rangle \mathrm{e}^{-kt}, \tag{1.147}$$

$$Var[X(t)] = \left\langle \left\{ [X(0) - \langle X(0)\rangle]\mathrm{e}^{-kt} + \sqrt{D}\int_0^t \mathrm{e}^{-k(t-t'')}dW\left(t'\right) \right\}^2 \right\rangle. \tag{1.148}$$

If the initial condition is independent of $dW(t)$ for $t > 0$ we get

$$Var[X(t)] = Var[X(0)]\mathrm{e}^{-2kt} + D\int_n^t \mathrm{e}^{-2k(t-t)}dt' =$$

$$= \left\{Var[X(0)] - D/2k\right\}\mathrm{e}^{-2kt} + D/2k \tag{1.149}$$

The time correlation function can also be calculated directly and is,

$$\langle X(t)X(s)\rangle = Var[X(0)]\mathrm{e}^{-k(t+s)} + D\left\langle \int_0^t \mathrm{e}^{-k(t-t)}dW\left(t'\right)\int_0^s \mathrm{e}^{-k(s-s')}dW\left(s'\right)\right\rangle =$$

$$= Var[X(0)]\mathrm{e}^{-k(t+s)} + D\int_0^{\min(t,s)} \mathrm{e}^{-k(t+s-2t\prime)}dt' =$$

$$= \left\{Var[X(0)] - \frac{D}{2k}\right\}\mathrm{e}^{-k(t+s)} + \frac{D}{2k}\mathrm{e}^{-k|t-s|}. \tag{1.150}$$

Notice that if $k > 0$, as $t, s \to \infty$ with finite $|t-s|$, the correlation function becomes stationary. Next, looking at the conditional probability density function, we find

$$\partial_t p_{1|1}\left(X(t), t \mid X(0), 0\right) = \partial_x(kX(t)p_{1|1}\left(X(t), t \mid X(0), 0\right)) + \frac{1}{2}D\partial_x^2 p_{1|1}\left(X(t), t \mid X(0), 0\right) \tag{1.151}$$

The normalized solution in the stationary regime, when $\partial_t p_{1|1}\left(X(t), t \mid X(0), 0\right) = 0$, is

$$p_{1|1}\left(X(t), t \mid X(0), 0\right) = (\pi D/k)^{-1/2}\exp\left(-kX(t)^2/D\right) \tag{1.152}$$

which is a Gaussian with mean 0 and variance $D/2k$. Finally, the characteristic function is

$$\phi(u,t) = \exp\left[-\frac{-Du^2}{4k}\left(1-e^{-2kr}\right) + iux_0e^{-kt}\right], \tag{1.153}$$

for the initial conditions $(x_0, 0)$. For more details relating to the derivations in this section please consult Section 3.8.4 of [1].

Mean-Reversion Process

Cox-Ingersoll-Ross model [30] is a process describing a timeseries which experiences higher inertia the further it is from the historical mean θ, which can be describe in an Ito SDE as

$$dX(t) = \kappa(\theta - X(t))dt + \sigma\sqrt{X(t)}dW(t), \tag{1.154}$$

where κ is a constant dictating the speed with which reversion happens and σ is a scaling constant relating to the variance of the process. We can immediately write the equation obeyed by the conditional probability distribution

$$\partial_t p_{1|1}\left(X(t), t \mid X(0), 0\right) =$$

$$= -\partial_x[\kappa(\theta - X(t))p_{1|1}\left(X(t), t \mid X(0), 0\right)] + \frac{1}{2}\sigma^2\partial_x^2[X(t)p_{1|1}\left(X(t), t \mid X(0), 0\right)]. \tag{1.155}$$

Next, the joint probability density is

$$f_2(X(s), s; X(t), t) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q\left(2(uv)^{1/2}\right), \tag{1.156}$$

where

$$c = \frac{2\kappa}{\sigma^2\left(1 - e^{-\kappa(s-t)}\right)}, \quad u \equiv cX(t)e^{-\kappa(s-t)}, \quad v \equiv cX(s), \quad q = \frac{2\kappa\theta}{\sigma^2} - 1$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. The distribution function is the noncentral chi-square, $\chi^2[2cX(s); 2q+2, 2u]$, with $2q+2$ degrees of freedom and parameter of noncentrality $2u$ and mean $2(q+u+1)$.

From this we can derive the expected value and variance of $X(s)$ as:

$$E[X(s) \mid X(t)] = X(t)e^{-\kappa(s-t)} + \theta\left(1 - e^{-\kappa(s-t)}\right), \tag{1.157}$$

$$Var[X(s) \mid X(t)] = X(t)\left(\frac{\sigma^2}{\kappa}\right)\left(e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}\right) + \theta\left(\frac{\sigma^2}{2\kappa}\right)\left(1 - e^{-\kappa(s-t)}\right)^2. \tag{1.158}$$

As k approaches ∞, the mean goes to θ and the variance to zero, while as κ approaches zero, the conditional mean goes to the current interest rate and the variance to $\sigma^2 X(t) \cdot (s-t)$.

If the interest rate does display mean reversion ($\kappa, \theta > 0$), then as s becomes large its distribution will approach a gamma distribution. The steady state density function is:

$$f_2[X(\infty), \infty; X(t), t] = \frac{\omega^\nu}{\Gamma(\nu)} X(t)^{\nu-t} e^{-\omega r} \tag{1.159}$$

where $\omega = 2\kappa/\sigma^2$ and $\nu \equiv 2\kappa\theta/\sigma^2$. The steady state mean and variance are θ and $\sigma^2\theta/2\kappa$, respectively. Finally, the characteristic function is

$$\phi(iu, t) = \exp\left\{C(iu, t) + D(iu, t)X_0\right\} \tag{1.160}$$

with

$$\phi(iu, t) = \left(1 - \frac{iu}{2\kappa}\sigma^2\left(1 - e^{-\kappa t}\right)\right)^{-\frac{2\kappa\theta}{\sigma^2}} \exp\left\{\frac{iuX_0 e^{-\kappa t}}{1 - \frac{iu}{2\kappa}\sigma^2\left(1 - e^{-\kappa t}\right)}\right\} \tag{1.161}$$

For more information on derivation please consult [30].

Chapter 2

Fractional Brownian Motion (*fBM*)

Fractional Brownian Motion, first considered by Kolmogorov (in 1940), Levy and others [31], reached the larger scientific community after the influential publication by Mandlebrot and V.Ness [11]. The reasons why this stochastic process gained traction is because of its many useful properties namely time and space translational invariance (stationarity) of increments which also imply increment Gaussianity and invariance under scaling with Hurst index H [2]. fBM is a non stationary process that can exhibit both long term memory ($\frac{1}{2} < H < 1$) and short term memory ($0 < H < \frac{1}{2}$), whose 1 point density satisfies the same pde as a scaling Gaussian Markov process but whose conditional probability density satisfies no pde at all. Furthermore, the fBM does not satisfy martingale pair correlation [32] [33]

$$\langle X(s)X(t)\rangle = \langle X(s)^2\rangle \quad s < t, \tag{2.1}$$

and is thus, not a martingale. Let us explore how this interesting mix of properties has found application in almost every scientific field e.g. physics, biology, computer vision, finance [31] by focusing on physics, mathematics and finance in particular.

2.1 *fBM* in Physics

As we already know, Brownian motion was first introduced as an example of a diffusion process. We have already established the link between Gaussian processes and diffusion in Section 1.3. However, nature breaks the Gaussian behaviour in many circumstances. fBM was first considered by Kolmogorov who was interested in noises which experience non linear mean squared displacement growth [34]. This non linearity was observed in anomalous diffusion processes present in polymers, crystals and others [35] in the form of a power law w.r.t time. The power law behaviour was found to have profound geometrical implications and led to the establishment of the field of fractals. One of the more recognizable first physical applications of this property is found by the hydrologist E.Hurst who studied cumulative water overflows of the Nile. But applications of fBM go even further. It successfully models a variety of natural processes such as: particle

trajectories in single-file diffusion, the integrated current in diffusive transport, polymer translocation through a narrow pore, solar flare activity, telecommunication networks, persistence of coal particulates in the atmosphere nearby mines [31] [36] [37]. Fractional Brownian motion also found applications in assessing statistical properties of interstellar gas [38], quantum chaos [39], turbulent flows [40]. Clearly, the combination of self-similarity and local Gaussianity make fBM a great candidate to model a large collection of physical processes.

2.2 *fBM* in Mathematics

Fractional Brownian Motion can be defined [2] [33] by the following two properties:

- stationary increments,
- process scaling with Hurst index H.

Let us formulate these two properties mathematically. Call the fBM process $B^{(H)}(t)$, where H is the Hurst index. Let $B^{(H)}(t,T) = B^{(H)}(t+T) - B^{(H)}(t)$ denote an increment and let $B^{(H)}(t,-T) = -(B^{(H)}(t) - B^{(H)}(t-T))$ denote the preceding increment. We obtain

$$-2\langle B^{(H)}(t,-T)B^{(H)}(t,T)\rangle =$$
$$= \left\langle(-B^{(H)}(t,-T) + B^{(H)}(t,T))^2\right\rangle - \left\langle(B^{(H)}(t,-T))^2\right\rangle - \left\langle(B^{(H)}(t,T))^2\right\rangle. \tag{2.2}$$

If the stochastic process $B^{(H)}(t)$ has stationary increments, meaning that $B^{(H)}(t,T) = B^{(H)}(t+T) - B^{(H)}(t) = B^{(H)}(T)$ is independent of t, then the mean-square fluctuation calculated from any starting point $B^{(H)}(t)$ is independent of starting time t,

$$\left\langle(B^{(H)}(t+T) - B^{(H)}(t))^2\right\rangle = \left\langle(B^{(H)}(0,\ T))^2\right\rangle. \tag{2.3}$$

Long-time increment autocorrelations follow,

$$-2\langle B^{(H)}(t,-T)B^{(H)}(t,T)\rangle = \left\langle(B^{(H)}(2\ T))^2\right\rangle - 2\left\langle(B^{(H)}(\ T))^2\right\rangle, \tag{2.4}$$

if the variance is not linear in the time, where we've written $B^{(H)}(0,T) = B^{(H)}(T)$ by taking $B^{(H)}(0) = 0$. When we combine these two assumptions we then obtain

$$\left\langle(B^{(H)}(t+T) - B^{(H)}(t))^2\right\rangle = \left\langle(B^{(H)}(\ T))^2\right\rangle = cT^{2H} \tag{2.5}$$

where $c = c(H)$ is a constant of proportionality whose structure is irrelevant for our purposes and also

$$-2\left\langle B^{(H)}(t,-T)B^{(H)}(t,T)\right\rangle / \left\langle(B^{(H)}(T))^2\right\rangle = 2^{2H-1} - 1. \tag{2.6}$$

To try to construct such a process, consider stochastic integrals of the form

$$B^{(H)}(t) = \int_{t_0}^{t} k(t,s)dW(s). \tag{2.7}$$

With $k(t,s)$ dependent on the later time t the stochastic integrals cannot be generated by an Ito sde. Stationary increments occur if $t_0 = -\infty$ in equation 2.3 [2]. Satisfying the stationarity condition

$$B^{(H)}(t+T) - B^{(H)}(t) = \int_{-\infty}^{T} k(T,s)dW(s) = B^{(H)}(T) \tag{2.8}$$

requires a very special class of kernels k. B. Mandelbrot and J. Van Ness have provided in their 1968 paper an example of a scaling kernel that generates stationary increments, hence describes fBM,

$$B^{(H)}(t) = \int_{-\infty}^{0} \left[(t-s)^{H-1/2} - (-s)^{H-1/2}\right] dW(s) + \int_{0}^{t} \left[(t-s)^{H-1/2}\, dW(s)\right. \tag{2.9}$$

The autocorrelation structure of the fBM is the following

$$\left\langle B^{(H)}(t)B^{(H)}(s)\right\rangle = 1/2\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{2.10}$$

For $H = 1/2$, the fBM is then a standard Brownian motion.

A standard fBM, $B^{(H)}$, has the following properties:

- $B^{(H)}(0) = 0$ and $\left\langle B^{(H)}(t)\right\rangle = 0$ for all $t \geq 0$.
- $B^{(H)}$ has homogeneous increments, i.e., $B^{(H)}(t+s) - B^{(H)}(s)$ has the same law of $B^{(H)}(t)$ for $s, t \geq 0$.
- $B^{(H)}$ is a Gaussian process and $\left\langle (B^{(H)}(t))^2\right\rangle = t^{2H}, t \geq 0$, for all $H \in (0,1)$
- $B^{(H)}$ has continuous trajectories.

In equation 2.5 we notice the variance of the increment vanishes as the time interval T goes to 0. For $\frac{1}{2} < H < 1$, this fact combined with a proof by Kolmogorov is enough to prove continuity of sample paths [11]. The extension for $0 < H < \frac{1}{2}$ can also be made but it is more subtle mathematically. Furthermore, fBM, similar to standard Brownian motion, is not differentiable.

Nevertheless, *fBM* exhibits long/short range dependence depending on the Hurst index H. More formally this property is expressed as

$$\lim_{T\to\infty} \frac{\rho(T)}{cT^{-\alpha}} = 1, \tag{2.11}$$

where $\rho(T)$ is the autocovariance function between two increments separated by time interval T, the c term is a constant and $\alpha \in (0,1)$.

Let us prove the long range dependence. Let $t_1 - T_1 < t_1 < 0 < t_2 < t_2 + T_2$ designate points on the time axis. With the autocorrelation function defined by

$$2\left\langle \left(B^{(H)}(t_2+T_2) - B^{(H)}(t_2)\right)\left(B^{(H)}(t_1) - B^{(H)}(t_1-T_1)\right)\right\rangle =$$

$$\left\langle \left(B^{(H)}(t_2+T_2) - B^{(H)}(t_1-T_1)\right)^2\right\rangle + \left\langle \left(B^{(H)}(t_2) - B^{(H)}(t_1)\right)^2\right\rangle -$$

$$-\left\langle \left(B^{(H)}\left(t_2+T_2\right)-B^{(H)}\left(t_1\right)\right)^2\right\rangle -\left\langle \left(B^{(H)}\left(t_2\right)-B^{(H)}\left(t_1-T_1\right)\right)^2\right\rangle \tag{2.12}$$

where $2(a-c)(d-b)=(a-b)^2+(c-d)^2-(a-d)^2-(c-b)^2$ was used. Using stationarity of the increments, and also dividing by a product of the variances at times T_1 and T_2, taking $t_2=t/2=-t_1$, we can evaluate

$$2Cov\left(S_1,\ S_2\right) \quad =\frac{\left\langle \left(B^{(H)}\left(t/2+T_2\right)-B^{(H)}(t/2)\right)\left(B^{(H)}(-t/2)-B^{(H)}\left(-t/2-T_1\right)\right)\right\rangle}{\sqrt{\left(\left\langle (B^{(H)}\left(\ T_1\right))^2\right\rangle \left\langle (B^{(H)}\left(\ T_2\right))^2\right\rangle\right)}}, \tag{2.13}$$

where $S_1=T_1/t, S_2=T_2/t$, to obtain

$$Cov\left(S_1,\ S_2\right)=\left[(1+S_1+S_2)^{2H}+1-(1+S_1)^{2H}-(1+S_2)^{2H}\right]/2\left(\ S_1\ S_2\right)^H. \tag{2.14}$$

Thus, if we make $T_1=T_2=T$ and consider the limit $t\to\infty$ then

$$\lim_{t\to\infty} Cov(S)\sim H(2H-1)\left(\frac{t}{T}\right)^{2H-2}, \tag{2.15}$$

which tends to 0 as t tends to ∞ as $2H-2<0$.

At a first reading, fBM may seem rather abstract and difficult to put into practice because of the ∞ in the lower integral limit. As such let us look at how to simulate it via the Cholesky method [41] [42].

Instances of fBM can be produced via any method for producing stationary Gaussian processes with a known covariance function. Cholesky method is the simplest, but it has $O\left(n^3\right)$ complexity. We note that more efficient algorithms such as the circulant embedding method of Dietrich & Newsam (1997) have a $O\left(n\log n\right)$ complexity.

First, consider n reference points on the time axis $t_1,\ldots,t_n$, in ascending order. Construct the matrix

1. $\Gamma_{n\ by\ n}=\left(R\left(t_i,t_j\right),i,j=1,\ldots,n\right)$ where $R(t,s)=\left(s^{2H}+t^{2H}-|t-s|^{2H}\right)/2$
2. Compute the square root matrix of Γ which we call $L(n)$, i.e. $L(n)L(n)^*=\Gamma$. This is the equivalent of the one dimensional standard deviation
3. Pick n numbers from a standard Gaussian distribution and assemble them into a vector v
4. Define $u=L(n)v$, which represent a valid run of a standard fBM process.

The crucial part of this algorithm is computing the $L(n)$ function via the so-called Cholesky decomposition. Knowing $\Gamma(n)$ is a positive definite matrix we can write $\Gamma(n)=L(n)L(n)^*$, where $L(n)^*$ is the conjugate transpose of $L(n)$. In this case, there exists a $L(n)$ a lower triangle matrix for which this decomposition is possible.

Next, call the (i,j)-th element of $L(n)$ as l_{ij} and likewise call $\gamma(i,j)$ the (i,j)-th of $\Gamma(n)$ for $i,j=0,\ldots,n$. Thus we can write

$$\gamma(i,j) = \sum_{k=0}^{j} l_{ik} l_{jk}, \quad j \leq i \quad . \tag{2.16}$$

We know then for $\gamma(0,0) = l_{00}^2$. For $i = 1$ we get the two equations

$$\gamma(1,0) = l_{10} l_{00}, \quad \gamma(1,1) = l_{10}^2 + l_{11}^2, \tag{2.17}$$

which determine l_{10} and l_{11}. We can similarly compute all l_{ij} terms from $\Gamma(n)$. Notice there is no n dependence in the l_{ij} terms and as a result we can compute $L(n+1)$ by padding a row with the following values

$$l_{n+1,0} = \frac{\gamma(n+1)}{l_{00}} \tag{2.18}$$

$$l_{n+1,j} = \frac{1}{l_{jj}} \left(\gamma(n+1-j) - \sum_{k=0}^{j-1} l_{n+1,k} l_{jk} \right), 0 < j \leq n \tag{2.19}$$

$$l_{n+1,n+1}^2 = \gamma(0) - \sum_{k=0}^{n} l_{n+1,k^*}^2 \tag{2.20}$$

$L(n)$ is uniquely determined if we constrain the diagonal elements to be strictly positive. $\Gamma(n)$ is a positive definite and as a result $L(n)$ is real.

Now that $L(n)$ is computed we can easily perform step 4. We can recursively compute $u(n+1)$ by using the padding prescription of $L(n+1)$ described above.

2.3 *fBM* in Finance

When we look at the normalized price increments of any asset on the market, the untrained eye will jump to the conclusion that these are random and uncorrelated, in other words they follow a Gaussian distribution around 0, which is the defining property of a Wiener / Brownian process. Of course, this is the simplest assumption one could make about the market, which can be seen as a linear, first order approximation if we borrow some physics terminology. Nevertheless, physics shows us time and again that all these linear models break down at some point, and more so, that if you add some specific type of complexity they become non linear and many times non predictive.

The Efficient Market model is the anecdotal "spherical cow in a vacuum" of financial modeling. The key assumptions are price increments follow a Brownian process (price history does not matter), people are rational actors who try to maximize profits while diminishing risks (homogeneity in trading approach), new information is absorbed into the price changes instantly (no latency) which leads to no arbitrage (risk free profit) [43]. Although a good starting point to study the market formally, these assumptions fall short of a realistic assessment of asset pricing. The tools originating from this model, among

which there are the Black Scholes Merton option pricing model and Capital Asset Pricing Model for investment portfolios, heavily used by institutions for a long time, have even facilitated some market crises because of too little risk protection. Risk in this case was based on a static notion of market volatility (price change variance) which is part and parcel of the gaussianity assumption of price changes.

A qualitative jump in financial modeling was made through the introduction of Fractional Brownian motion as the asset variance process [44]. Before going into the mathematics some comments are in order about its suitability for market modeling. If we were to take a picture of a price history of a common asset and remove the axes labels, just by watching the graph we would not be able to say if the plot ticks should be days, months, years etc. This particular property is called scaling and states that if we zoom in on a process history the evolution looks similar at all scales. Furthermore, especially when we observe market crises, we notice big price changes tend to follow other big price changes. This indicates short term dependence or stated differently market turbulence clusters together. In addition to this, we observe price changes tends to linger after a market blow or boom which indicates long term dependence. The short term and long term dependence are included in the assumption that market history influences future price change. Equally important is that the fBM is a moving average process where each past prices changes gets weighed by a kernel function. Moreover, prices follow jumps rather than the mild Brownian movement [23]. This indicates that extreme events are much more likely than a Gaussian assumption would make us think.

Specifically for finance, fBM finds its application through the MMAR (Multifractal Model of Asset Returns) model [45]. MMAR is a notable model specifically through the usage of multiple advancements in the understanding of stochastic processes. MMAR leverages long memory, a feature of *fBM*, long tails in comparison to Gaussian distributions, specific to Levy stable distributions, and trading time, a concept introduced by B. Mandlebrot and Taylor to represent the way people actually place their buy/sell orders on the exchange. These three feature make the brunt of the regularities of financial time series. To get a good grasp of what the model entails we need to introduce a handful of important concepts.

Self affinity

Between economists there is consensus that the distribution of returns should look the same under time rescaling. The invariance of the distributions under summing of independent increments is called L-stability. This class of processes can be generalized to include dependent increments. This is specifically called process self-affinity. In the literature, this property is called, at times, self-similarity, although there is a clear difference between them. Self-similiarity implies the distributions are identical, except for a rescaling factor, if we isotropically rescale all dimensions of the process. Self affinity is a more general feature which preserves the distribution even though rescaling is different along different directions of the basis. More formally, for a random process $\{X(t)\}$ with $X(0) = 0$ self-affinity is

$$\{X(ct_1), \ldots, X(ct_k)\} \stackrel{d}{=} \left\{c^H X(t_1), \ldots, c^H X(t_k)\right\} \tag{2.21}$$

for some $H > 0$ and all $c, k, t_1, \ldots, t_k \geq 0$.

The reason we do not stop at L-stable distributions is that there are no periods of high volatility versus low volatility, where past price changes positively or negatively influence future changes. For this reason *fBM* is attractive as a model component. Nevertheless, *fBM* does not have fat tails and fails to decouple volatility from future returns.

Multifractals

Multifractals play an important role in MMAR and will thus desribe it here. Take for a random process $\{X(t)\}$ with $X(0) = 0$. A multifractal process obeys

$$X(ct) \equiv M(c)X(t), \tag{2.22}$$

where X and M are independent random functions. Self-similar processes obey

$$M(c) = c^H \tag{2.23}$$

To bring this general property to a more familiar shape we define $H(c) = \log_c(M(c))$ such that $X(ct) \equiv c^{H(c)}X(t)$. For multifractal processes $H(c)$ is a random function of c.

We impose the following property: $M(ab) \equiv M_1(a)M_2(b)$, where M_1 and M_2 are distinct instances of M. This condition implies

$$E\left[|X(t)|^q\right] = c(q)t^{\tau(q)+1} \tag{2.24}$$

where $\tau(q)$ and $c(q)$ are both deterministic functions of q. We can alternatively define multifractal processes as processes which obey this equation. This is because for many cases it is advantageous to define processes in terms of moments as it makes it more accessible for statistical tests. Using this definition, we note that multifractal processes are globally scaling with scaling function $\tau(q)$. By taking $q = 0$ in equation 2.24 , we notice that it is always true for $\tau(0) = -1$.

Turning back to self-affine processes $\{X(t), t \geq 0\}$ with affinity index H we know $X(t) \equiv t^H X(1)$, and consequently

$$E\left[|X(t)|^q\right] = t^{Hq}E\left[|X(1)|^q\right]. \tag{2.25}$$

Notice $\tau(q)$ and $c(q)$ look like

$$\tau(q) = Hq - 1 \quad \textit{and} \quad c(q) = E\left(|X(1)|^q\right). \tag{2.26}$$

Multifractal Measure

Before we advance in describing MMAR more precisely, we need to introduce the concept of multifractal measure which is the key to producing the trading time with our desired properties. Put simply, a probability measure on the interval $[0, 1]$ is a prescription of how to slice the interval in a set of subintervals, such that their union is equal to $[0, 1]$. A very similar procedure is the partitioning of the domain when calculating a Riemann integral, without the additional step of creating the set of subintervals. The simplest example of a multifractal measure is the Binomial/Bernoulli measure. This measure is defined on the compact interval $[0, 1]$. The Bernoulli measure is the limit of an iterative algorithm called a *multiplicative cascade*.

Take the *mass* of the interval to be 1. The *mass* in this context stands for the integration of the measure over the entire domain. Choose two positive numbers m_0,m_1 such that $m_0 + m_1 = 1$. At stage $k = 0$ we define the Bernoulli measure μ_0 to be the uniform probability distribution over the interval $[0, 1]$. At stage $k = 1$ we slice the interval $[0, 1]$ in two equally sized subintervals $\{[0, 1/2], [1/2, 1]\}$ which receive masses m_0 to the left interval, and m_1 to the right interval. The distribution over each fundamental subinterval is the uniform distribution. Integrating the Bernoulli measure μ_1 over the left subinterval gives the mass m_0, while integrating on the right subinterval gives m_1. At stage $k = 2$, the subintervals from the $k = 1$ stage are split equally. We thus have the splitting $\{[0, 1/4], [1/4, 1/2], [1/2, 3/4], [3/4, 1]\}$ underpinned by the Bernoulli measure μ_2. The weights of the subintervals are the following

$$\mu_2([0, 1/4]) = m_0^2, \quad \mu_2([1/4, 1/2]) = m_0 m_1, \tag{2.27}$$

$$\mu_2([1/2, 3/4]) = m_1 m_0, \quad \mu_2([3/4, 1]) = m_1^2. \tag{2.28}$$

By iterating this procedure, in the limit $k \to \infty$, we find the Bernoulli measure. To be noted, summing the measure over all subintervals gives 1 by definition. Now, the way we can make the multifractal measure stochastic is by assigning the mass at each step of the procedure in a prescribed random fashion. This can either be that we randomly assign m_0,m_1 (in the case of the Bernoulli measure) to either left or right, or alternatively that we choose between different mass splittings based on a certain distribution function.

Now, let us show that this algortihm produces multifractals by deriving the result in the definition 2.24.

Let us start by imposing mass conservation at each stage of the multiplicative cascade, and that the mass splittings is made via a random distribution function. Next, we choose to split the interval into b parts at each step of the algorithm. We can then use the notation of b-adic numbers to write each unique interval by naming the lower interval limit via

$$0.\eta_1\eta_2 \dots \eta_k = \eta_1 \frac{1}{b^1} + \eta_2 \frac{1}{b^2} + \dots \eta_k \frac{1}{b^k} = \sum \eta_i b^{-i}, \tag{2.29}$$

where $\eta_1 \dots \eta_k = 0, \dots b - 1$.

We assign each cell of length $\Delta t = b^{-k}$ a random mass

$$\mu_k(\Delta t) = M(\eta_1)M(\eta_1, \eta_2)\dots M(\eta_1, \dots, \eta_k) \tag{2.30}$$

and as a result

$$[\mu_k(\Delta t)]^q = M(\eta_1)^q M(\eta_1, \eta_2)^q \dots M(\eta_1, \dots, \eta_k)^q \tag{2.31}$$

for all $q \geq 0$. Taking the expectation of the equation above we obtain

$$E[\mu_k(\Delta t)^q] = [E[M^q]]^k. \tag{2.32}$$

as the different mass multipliers are statistically independent. To bring the measure to a more general multifractal form we now choose the multipliers M_β be picked from the different random distribution that preserve mass "on average" in the sense that $E[\sum M_\beta] = 1$. This specific measure is called canonical. Its total mass, denoted Ω, is generally random, and the mass of a b-adic cell takes the form:

$$\mu_k(\Delta t) = \Omega(\eta_1, \dots, \eta_k) M(\eta_1) M(\eta_1, \eta_2) \dots M(\eta_1, \dots, \eta_k). \tag{2.33}$$

Importantly $\Omega(\eta_1, \dots, \eta_k)$ has the same distribution as Ω. By taking the expectation of a moment of canonical measure we then see that

$$E[\mu_k(\Delta t)^q] = E[\Omega^q][E[M^q]]^k. \tag{2.34}$$

This is exactly the definition of a multifractal process as captured in equation 2.24.

Finally, we can describe the MMAR model. In this model the volatility of a financial asset is a multiscaling process with long memory and fat tails whose roughness is generated by a random trading time, modelled via a cumulative distribution function of a random multifractal measure. Take the price of a financial asset to be $\{S(t); 0 \leq t \leq T\}$. Let

$$X(t) = \ln S(t) - \ln S(0). \tag{2.35}$$

MMAR make the following assumptions. Take $X(t)$ to be a compound process. Then

1. $X(t) \equiv B_H[\theta(t)]$ where $B_H(t)$ is a fBM with self-affinity or Hurst index H, and $\theta(t)$ is the multifractal stochastic trading time.
2. $\theta(t)$ is the cumulative ditribution function of a multifractal measure [46] defined on $[0, T]$. That is, $\theta(t)$ is a multifractal process with continuous, non-decreasing paths, and stationary increments.
3. $\{B_H(t)\}$ and $\{\theta(t)\}$ are independent.

Thus multifractality becomes part of MMAR through the trading time. The long memory and fat tails will thus be part of the price evolution by compounding. The original MMAR [45] [47] [48] uses the standard Brownian motion to create a compound process. In this case, the model has the rather fortunate property of being a martingale, which we know already to be deeply connected to an efficient market. However, the extended MMAR model takes $H \neq 1/2$ but does not satisfy the martingale property [49]. This means there is room for arbitrage [43] [50] [51].

Multiple extensions have been made to compound multiscaling processes such as MMAR. We mention *mBm* [52], or otherwise multifractional Brownian motion, which takes into account a time varying Hurst index. We also mention the Lux extension to MMAR [53] which concentrates on the causality of MMAR. Causality in this context refers to distinction between past and future when predicting future behaviour. MMAR does not have this distinction built in because the multiplicative cascade algorithm generates the full timeline of the process.

Spectrum of MMAR

If $S(t)$ is the price of an asset and $X(t)$ is defined as the log-price as before, then it was proven in [45] [47] [48] that

1. $f_X(\alpha) = f_\theta(H\alpha)$, where f stands for the spectra of a process
2. $\tau_X(q) = \tau_\theta(Hq)$, where τ is the scaling function of the process
3. $\tau_X(1/H) = \tau_\theta(1) = 0$, which enables estimation of H from data.

Next, let us see how to estimate the parameters of the multiplicative cascade of the multifractal trading time. Take the mass assignment given to a cell, of the multiplicative cascade, with length $\Delta t = b^{-k}$ to be

$$\mu_k([t, t + \Delta t]) = M(\eta_1)M(\eta_1, \eta_2)(\eta_1, \ldots, \eta_k).$$

Remember that all $M(\eta_1)(\eta_1, \ldots \eta_k)$ belong to the same distribution function and are statistically independent. Next, consider the coarse Hölder exponents

$$\alpha_k(t) = \frac{\ln \mu_k[t, t + \Delta t]}{\ln \Delta t}$$

$$= -\frac{1}{k} \left[\log_b M(\eta_1) + \ldots + \log_b M(\eta_1, \ldots, \eta_k)\right] \tag{2.36}$$

If k is large enough, and we make the assumption of convergence to a most probable exponent we can use the tools of probability theory.These are the Strong Law of Large

Numbers (SLLN), the Central Limit Theorem (CLT). These give insight for the most likely estimate of α_k, the width of the bell curve and the tail of the distribution.

By SLLN, α_k converges almost surely to

$$\alpha_0 = -E[\log_b M] \tag{2.37}$$

But $E[M] = 1/b$, and using Jensen's inequality [45] implies that $\alpha_0 > 1$. As $k \to \infty$ in equation 2.36, by our assumption, the Hölder exponents should be all contained in region around α_0. Using CLN, we infer the shape of the bell around α_0. Assuming $\log_b M$ has finite variance σ^2 we can write

$$\sqrt{k}\,(\alpha_k - \alpha_0) \to N\left(0, \sigma^2\right), \tag{2.38}$$

where $N\left(0, \sigma^2\right)$ is the normal distribution with mean 0 and variance σ^2. The asymptotic behaviour of the distribution should be

$$\frac{N_k}{N} \sim \frac{1}{\sqrt{2\pi\sigma^2/k}} \exp\left[-\frac{1}{2}\left(\frac{\alpha - \alpha_0}{\sigma/\sqrt{k}}\right)^2\right]. \tag{2.39}$$

Thus, the spectrum of the asset price should be

$$f(\alpha) \sim 1 - \frac{1}{2\ln b}\left(\frac{\alpha - \alpha_0}{\sigma}\right)^2. \tag{2.40}$$

The CLT thus shows that the multifractal spectrum is locally quadratic around the most probable exponent α_0. To create the multiplicative cascade it was proven in [45] that the mass at each splitting should be chosen via a Gaussian distribution with mean λ and variance σ^2 according to

$$\hat{\lambda} = \frac{\widehat{\alpha_0}}{\widehat{H}}, \tag{2.41}$$

$$\hat{\sigma}^2 = \frac{2(\hat{\lambda} - 1)}{\ln[b]}, \tag{2.42}$$

where $\widehat{\alpha_0}$ is the previous estimate of the Holder exponent, and H is computed via statement number 3 of this subsection.

Chapter 3

Stochastic Processes in Finance

3.1 Definitions: Stocks, Money Market, Options

So far we have talked rather loosely about terms such as money, stocks, options etc. without pin pointing the meaning of any of them. Without further ado, most of the financial lingo that is useful for the remainder of this article will be defined in what follows, based on chapters 2 and 3 of [54].

Money

Not surprisingly, we cannot introduce any financial term without reference to the word *money*. To start with, money can have different physical representations. Most people take, for instance, coins or dollar/euro bills to be money. Although valid, it is not the only way money can exist physically. Economists like to define money as anything acceptable for payment of goods or services or in repayments of debts [54]. Coins and paper money is what is called *currency*. Apart from currency, we consider money to be *bank cheques*, *gold* or *silver coins* and many others. The reason money can take so many forms is because it has to satisfy three very important functions.

1. Money is a **medium of exchange**. It is used to pay for goods and services. Without money as a medium of exchange, people revert to a barter economy, where standardization of prices is not possible and transaction costs (the time spent finding a person that has the good or service you want and who wants a good or service you offer) are very high. Money allows people to trade with each other efficiently. Because of its usefulness, all but the most primitive civilizations did not invent it. For something to be considered money it has to be easily standardized, widely accepted, must be divisible into smaller units, must be easy to carry and hard to deteriorate.
2. Money provides a **unit of account**. This means money can be used to ascertain the value of things. We use money to price goods and services in the same way we use to measure distances in terms of kilometers. In contrast to a barter economy, money enables a standard according to which everything is valued. Without money,

the value of any good or service would need to be quoted in terms of the number of units of any good it is exchanged with (e.g. 5 peaches, 10 bottles of water, 1 haircut etc.).

3. Money is a **store of value**. This means it represents purchasing power. Nevertheless, there are many alternative stores of value such as stocks, bonds, land, real estate, cars and art to name a few. Many of these alternatives have useful properties that money does not have such as price appreciation or enabling someone to travel at high speeds on the highway. What money has over all others stores of value is liquidity. Liquidity is what makes the economy efficient and money is the most liquid asset because it represents the *medium of exchange*.

Now that we grappled with the definition of money let us look at the definition of a financial market.

Financial markets

Financial markets have the function of allowing the transfer of funds from a surplus of funds to a shortage of funds. An instance of this can happen when a wealthy retiree invests in a start up or when a bank gives a loan to a private consumer. Yet again, to get a grasp of why markets can look so different we need to take a quick look into their functions and structure.

More specifically, financial markets are a means of transferring funds from households, firms and government who have a surplus of funds (e.g by saving their income) to those who have a shortage because they need to spend more than they make.

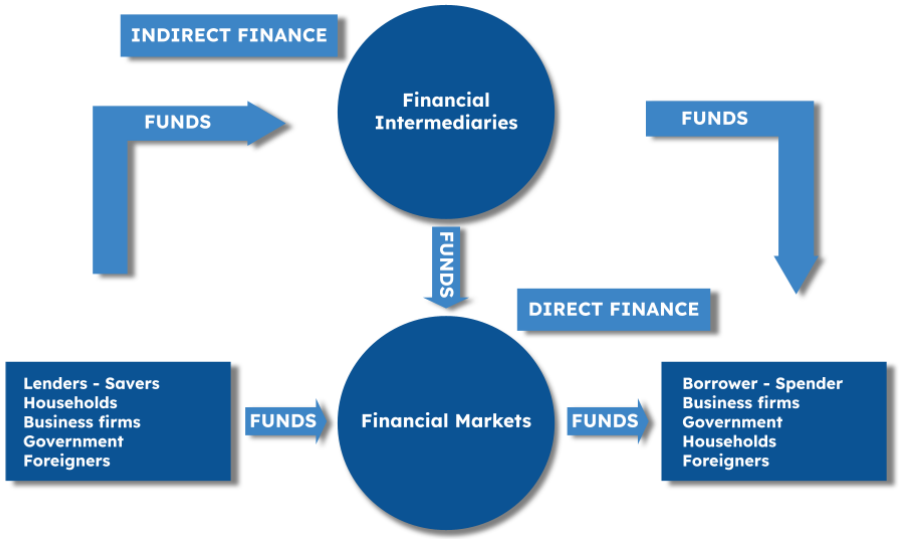


Figure 3.1: Flowchart of how funds travel in an economy. The direct path is through Financial Markets while the indirect path is via Financial Intermediaries such as bank, government etc.

In the absence of financial markets, we end up in a situation similar to a barter economy in that transaction costs are high. By transaction costs we mean the time people with

surplus of funds would need to find the people who need funds so that they can borrow it at an interest.

In terms of structure, there are multiple types of financial markets, of which we mention the ones most relevant to this article. Namely there are *Debt and Equity Markets*, *Exchange markets* and *Money and Capital Markets*.

First of all, *The Debt and Equity Market* is the most common channel for obtaining funds. In the Debt Market, a lender issues a debt instrument, such as a mortgage, such that the borrower has a contractual obligation to fulfill his debt with a specified interest, until a specified date. In the Equity Market, the transfer of funds is made by issuing stocks, whereby the seller of the stock offers ownership and a part of a company's profits for a specified price. There are pros and cons for using any financial instrument, although we will not go into it here.

Second of all, there are *The Exchange markets*. The Exchanges are places where buyers and sellers of financial instruments meet to trade. An example of such an exchange are Commodity Exchanges where people buy securities based on wheat, barley, gold etc.

Finally, *The Money and Capital markets* are distinguished from the other types based on the maturity of securities traded. Maturity means the period over which the contractual obligations underpinned by the security must be fulfilled. The Money Market comprises of only short-term debt instruments (maturity of less than one year) whilst the Capital Market comprises all debt and equity trades of maturity longer than one year. It thus stand to reason the Money Market is more liquid than the Capital Market.

Financial instruments : Stocks, Bonds, Options

As specified in Figure 3.1, in direct finance, borrowers borrow money from lenders by selling the lender *financial instruments* or *securities*. These *securities* act as claims on the future income or assets of the borrower. For instance, if a company wants to build a factory, it will borrow money from savers in the form of either *stocks* (which give ownership rights to the company and a slice of profits and assets via dividends) or *bonds*, debt securities which promise periodic payments for a specified period of time.

On the Money Market the most commonly traded securities is the Treasury Bill (T-Bill), which is a short-term bond (either 4,8,13,26,52 weeks) that gives a small interest to the T-Bill owner. On the Capital Market, the most traded securities are Corporate Stocks, Residential Mortgages, Corporate Bonds, Bank commercial loans and Consumer loans.

Options are contracts which give the buyer the *right* but not the obligation to buy or sell a specified financial security, called the *underlying*, for a specified price, called *exercise or strike price*, by a specified date, called *expiration date*. The seller of the option (also called *writer*) is then obligated to buy or sell the underlying asset based on whether the option buyer decides to buy or sell the underlying. The reader is advised to pay attention to the distinction between the right of the option buyer to exercise the option and the obligation of the writer to fulfill his contractual agreement. As can be noticed, the option seller exposes himself at a risk because of his obligation. In exchange for the additional

risk the writer sells the option at a price called *premium*. There are multiple types of options, of which we focus on the European Option. This type of option enables the option buyer to exercise his right to buy or sell the underlying for the strike price only on the expiration date, or strike date. We make the additional distinction between a Call Option, which allows the option buyer to only buy the underlying, in contrast to a Put Option, which only gives the right of the option buyer to sell the underlying on the expiration date.

What we focus on specifically are the time evolution of stocks and options and the various models built to deal with them.

3.2 L. Bachelier

Louis Bachelier (1870-1946) is one of the most important figures in mathematical finance. The following summary of his work is based on [55] [56] [3].

Bachelier is considered the first to apply trajectories of Brownian motion to describe price dynamics and calculating option prices, having done this in his Phd thesis in 1900. His assumptions make up what is now called the*Efficient Market Hypothesis*. For very short time frames during stable market periods his model fits market data very well. However, Norbert Wiener is the first to define Brownian Motion rigorously, in 1923, using the *Measure theory* of Borel and Lebesgue, conationals and contemporary with Bachelier but who did not pay attention to his work. Brownian motion has some interesting properties that make it applicable to a wide range of phenomena. It is continuous, but nowhere differentiable, it is self-similar it is a Levy process and also a martingale.

Bachelier was, for many years, unknown in the mathematics community of his time. He did all his thesis work without making any reference to Boltzmann, as his scientific articles were not translated to French until later in 1902. His thesis was evaluated by a committee of very important mathematicians, namely Poincare, Appell and Boussinesq. Out of them, Poincare is the only one who realised the key insights brought to light by the thesis and who encouraged Bachelier to continue pursuing this research. Indeed Poincare was the main driving force that enabled the publication of all later work of Bachelier into scientific journals. Before the time of Bachelier there were not really many people concentrating on probabilistic processes and with respect to application on the stock market there is only one book written before him, in 1863, by Jules Regnault called *Calcul des chances et philosophie de la bourse*. This book was more philosophical than mathematical although it contains the first known reference of a square root law of deviation applied on a price movement. For many years Bachelier, very aware of the importance of his contributions, continued writing articles and books about probability, even though the french mathematical community did not consider it as a serious enough endeavour. This was both because of the attitude of the french mathematics community towards more "creative" avenues of research and also because Bachelier left gaps in his proofs and they were thus not up to the usual standard of rigor. It is very telling to know that Bachelier could not secure a permanent academic position until very late in his career, at the age of 57, when he took a teaching position at The University of Besancon. Up to

that point he only taught as a guest lecturer and was not paid for a long period of time for teaching his courses at The University of Paris. This is due to the fact that his doctorate thesis was awarded the "honorable" mark (which was the highest mark given to theses on applied topics) and his arguments strayed from the well established rules of axiomatic proofs. Furthermore, he had conflicts with Levy and Gevrey based on his application to the chair of mathematics at The University of Dijon, which was left vacant in the early 1920s. M. Gevrey took a page out of Bachelier's thesis, which, out of context, could be easily interpreted as flawed (although a closer look into the asymptotic behaviour of those statements would reveal no mistake) and forwarded it to Levy. Many years later they resolved this conflict when Levy read more of his work, after he found Bachelier referenced in works by Kolmogorov. Nevertheless, all results derived by Bachelier were indeed correct and he touched upon many topics of probability and grasped many concepts about random processes long before his predecessors. Bachelier's work on random processes predates influential mathematicians and physicists such as Markov, Kolmogorov, Wiener, Einstein, Levy. His work came to the attention of the wider community after Kolmogorov credited him (in 1931) for his intention to develop the analytic theory of Markov processes. Bachelier's thesis was translated in english only in 1964 but his name became widespread earlier. Kiyosi Ito, of Ito's Lemma, stated he was more influenced by Bachelier than Wiener. Other international mathematicians that recognized his work before the 1950s were Paul Erdos, Mark Kac and Stanislaw Ulam. In the 1950s Paul Samuelson introduced Bachelier to economists and very soon the importance of his insights was established in the mathematics community. His main insights now find themselves crystalized into the Black-Scholes-Merton model and Corporate Asset Pricing model which we will go into next.

3.3 Capital Asset Pricing Model (CAPM)

CAPM [57] was built as an investment portfolio tool. Put simply, it allows the investor to choose his protfolio based on his risk or volatility (variance) tolerance. To get an idea of how it works, let us start with two assets H and L. Let us also assume we can borrow and lend at the risk-free rate. M is a high risk asset, and L is a riskless asset, such as a Treasury Bond. The audacious investor will choose to invest all his wealth into H, which promises a higher return for a higher risk exposure. The risk avert investor will invest everything in L, but with not such a promising return.

Now, let x be the fraction of the wealth invested in asset H. Then the investor puts $1-x$ into the riskless asset. The expected return of this portfolio is

$$(1-x)r_f + xE_H = r_f + x\left(E_H - r_f\right), \tag{3.1}$$

where E_H is the percentage return of the asset H and r_f is the risk free rate.

The risk associated with this portfolio is thus $x\sigma_H$, where σ_H is the standard deviation of asset H. The risk is proportional to how much of asset H we hold because we have no other risk involved. The Sharpe Ration of this portfolio is defined as

$$Sharpe\,Ratio = (E_H - r_f)/\sigma_H, \tag{3.2}$$

where $E_H - r_f$ is called the risk premium of asset H and σ_H is called the risk of asset H. The higher the Sharpe Ratio is the better the investment. Then, CAPM makes the assumption that during market equilibrium the price of each risky asset is fair. Then all investors hold the same proportions of the risky asset. In this case the portfolio with the highest Sharpe Ratio must be the market portfolio. Consequently the risk premium of each asset should be chosen such that

$$E_S - r_f = \beta\,(E_M - r_f)\,, \tag{3.3}$$

where E_S is the expected return of any given asset and E_M is the expected return of the market, and β is the correlation of the asset's return to market return. This is the defining prescription of the Capital Asset Pricing model. If we assume the Market portfolio does not have the best Sharpe Ratio then investors could outperform the market up until the CAPM becomes true. This is done by including new stocks with better Sharpe Ratio to the market value. If we assume rational investors which try to maximize Sharpe Ratio, this process will happen until the price equalize in such a way to make the market Sharpe Ratio the highest.

3.4 Black-Scholes-Merton model

The topics introduced so far have focused on the mathematical definitions of stochastic processes. This section started with analysis of the first stochastic process studied in physics, and in the same spirit in this subsection we look at one of the first stochastic model used in finance. Black-Scholes-Merton model for options pricing, published in 1973, is one of the first mathematical attempts to protect investment portfolios from incurring losses [27] [8] [58] . In order to define the model, we use the Merton approach. Start with a portfolio consisting of options, stocks and cash which has 0 value at all times and the property of self-financing. Suppose that stock price S follows a geometric Brownian motion

$$dS = \mu S dt + \sigma S dW, \tag{3.4}$$

where μ is the instantaneous drift of stock and σ the instantaneous variance of the Brownian motion. Now consider a European option with value $V(S,t)$ based on this stock. From Ito's lemma at equation 1.123 we can write

$$dV = \left(\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)dt + \sigma S\frac{\partial V}{\partial S}dW \tag{3.5}$$

The goal of the BSM model is to define a portfolio of stocks and options that is risk free. To make progress, some definitions are in place.

- Q_S is number of stocks, each valued at S
- Q_V is number of options, each valued at V

- B is the cash on the account continuously compounded using the risk-free rate r
- dQ_S is the change in the number of stocks
- dQ_V is the change in the number of options
- δB is the change in the cash, caused by buying/selling stocks and options

Now we are able to define the zero value of the portfolio as

$$SQ_S + VQ_V + B = 0 \tag{3.6}$$

and the self-financing property as

$$SdQ_S + VdQ_V + \delta B = 0. \tag{3.7}$$

Change in the cash is

$$dB = rBdt + \delta B \tag{3.8}$$

Differentiating equation 3.6 we get

$$0 = d\left(SQ_S + VQ_V + B\right) = d\left(SQ_S + VQ_V\right) + \overbrace{dB}^{rBdt+\delta B}$$

$$0 = \overbrace{SdQ_S + VdQ_V + \delta B}^{=0} + Q_S dS + Q_V dV + rBdt$$

$$0 = Q_S dS + Q_V dV \overbrace{-r\left(SQ_S + VQ_V\right)}^{rB} dt. \tag{3.9}$$

We divide by Q_V and denote $\Delta = -\frac{Q_S}{Q_V}$. This allows us to write

$$dV - rVdt - \Delta(dS - rSdt) = 0 \tag{3.10}$$

dS is defined in equation 3.4 and dV is defined in equation 3.5.We then choose Δ so that it eliminates the randomness (the Brownian coefficient of dW will be zero). As a result, we obtain

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \tag{3.11}$$

which holds for $S > 0, t \in [0, T)$. This partial differential equation holds for any derivative that pays a payoff at time T depending on the stock price at this time. Notice the drift term μ does not appear in this equation. Thus, the PDE 3.11 works even if we take $\mu = r$. If we do this, then the option seller can remove the part of the premium related to holding the stock until the option buyer exercises his right to buy the stock. This scenario would only be true if all investors are risk neutral, in other words, whose wealth would accrue interest equal to the risk free rate. This assumption leads to *risk neutral pricing* method which we now present shortly.

A marketplace is risk-neutral [59] if for all assets A traded and all times t, the value of the asset $C(A, 0)$ at time $t = 0$ is the expected value of the asset at time t discounted to its present value using the risk-free rate r according to

$$C(A, 0) = e^{-rt}E[C(A, t)]. \tag{3.12}$$

In reality, most people are not risk neutral and for some risky asset A and time period t, they will value the asset at $C(A,0) < e^{-rt}E[C(A,t)]$. This is because an investor would like to believe the risk he takes for holding the stock and losing the interest is less. Nevertheless, in the context of BSM, risk neutral pricing is a often used tool.

To solve equation 3.11 we need to impose boundary conditions. We take the particular case of a call option which, for convenience we call $C(S,T)$ instead of the $V(S,T)$ use so far. The boundary conditions are $C(S,T) = \max(S(T) - K, 0) = (S(T) - K)^+$ and $C(0,t) = 0$ for all t and $C(S,t) \to S$ as $S \to \infty$. With this particular choice of boundary conditions the solution for the call option price equation is

$$C(S,t) = S_t N(d_1) - e^{-r(T-t)} K N(d_2), \tag{3.13}$$

where

$$d_1 = \frac{\log\left(\frac{S_t}{K}\right) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \tag{3.14}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{3.15}$$

and the function N is the cumulative distribution function of the standard normal distribution.

Feynman Kac formula

The BSM formula 3.11 can also be solved using the Feynman-Kač theorem [4]. This theorem allows us to write the solution to partial differential equations in terms of an expectation value. The theorem states that for the partial differential equation

$$\frac{\partial F}{\partial t} + \mu(x,t)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 F}{\partial x^2} = rF, \tag{3.16}$$

with $F = F(x,t)$ a solution, with boundary condition

$$F(T,x) = \Phi(x), \tag{3.17}$$

then, if ϕ is well behaved enough, $F(x,t)$ can be written as

$$F(x,t) = E_{x,t}\left[\Phi\left(X_T\right)\right], \tag{3.18}$$

where $X(t)$ is a stochastic process satisfying the stochastic differential equation

$$dX(t) = \mu\left(X(t),t\right)d\tau + \sigma\left(X(t),t\right)dW(t) \tag{3.19}$$

with initial condition $X(t) = x$, which was written as a subscript in equation 3.18.

Interestingly, this theorem is a consequence of how the Ito integral is defined. At the same time, from a physical standpoint, this formulation is allowed because the pde in equation 3.16 can be recast into the heat equation and so can Brownian processes. We can readily apply Feynman-Kac formula to the Black-Scholes equation for a call option $C(S,T)$ to yield

$$C\left(S_0,0\right) = e^{-rT} E\left[\max(S(T)-K,0)\right], \tag{3.20}$$

which is consistent with the expectation taken under a risk-neutral measure. This solution tells that we could equally work with a stochastic process for the stock price $S(t)$ described by

$$dS = rSdt + \sigma SdW. \tag{3.21}$$

Note that, whereas the drift is changed to the risk neutral rate, the volatility is unchanged. Importantly, the expectation value written above is computed in practice using Monte Carlo methods. The key concept here is that if we can produce M independent random samples $S_T^{(j)}, j = 1,\ldots,M$, of the stock price, under the risk neutral probability measure, then by the law of large numbers we could estimate the expected value by the sample mean

$$\hat{C} = \frac{1}{M}\sum_{j=1}^{M} e^{-rT} \max\left\{S_T^{(j)} - K, 0\right\}. \tag{3.22}$$

This procedure represents the core of the Monte Carlo simulations.

Option Greeks

To make a short detour, another tool often used with Black Scholes model is called option greeks. The "Greeks" are simply (calculus) derivatives. Take the call option price $C = C(S,T,K,\sigma,r)$ and perform a Taylor expansion

$$dC = \frac{\partial C}{\partial S}dS + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}dS^2 + \frac{\partial C}{\partial T}dT + \frac{\partial C}{\partial \sigma}d\sigma + \frac{\partial C}{\partial r}dr + \ldots \tag{3.23}$$

The following naming convention is in order. The Δ (delta) is the derivative of the call option price with respect to the stock price $\Delta = \frac{\partial C}{\partial S}$. The Γ (Gamma) is the second order derivative of C with respect to S coming from Ito's Lemma $\Gamma = \frac{\partial^2 C}{\partial S^2}$. The Θ (Theta) is the derivative of C with respect to T, or $\Theta = \frac{\partial C}{\partial T}$. The ν (Vega), regularly used to signify volatility in finance literature, is the derivative of C with respect to volatility σ, or $\nu = \frac{\partial C}{\partial \sigma}$. Finally ρ (Rho) is the derivative of C with respect to interest rate changes, namely $\rho = \frac{\partial C}{\partial r}$.

Price, delta and gamma are related just as position, velocity and acceleration are in physics. Thus, using delta to estimate how much the price will change only works if there is no acceleration ($\Gamma = 0$). The greeks are regularly obtained directly from the BSM solution for the call option price or, additionally, with a variable implied volatility surface which will describe next.

Remarks on Black-Scholes-Merton model

BSM is the reason the financial derivatives market developed so much since its creation in 1973. Although simple, many assumptions miss the mark in terms of empirical accuracy. For a start, if BSM were correct, then there would be no market for derivatives. This is because in a complete market, adding new derivatives is redundant. In practice the market is not complete and perfect hedging is also not possible because of transaction costs, jumps in stock prices, non Gaussian probability distribution for stock price increments, volatility clustering [60] etc.

In addition to this, market incompleteness combined with our no arbitrage assumption means there is no unique risk neutral rate of interest. In fact, there should be a range of risk neutral measures for which no arbitrage is still satisfied. The interested reader can check [4] for more information. Nevertheless, BSM is still useful because it provides a benchmark for our expectation of what market prices should be for derivatives. Accordingly, trading desks prefer to look up the prices on the market and then rely on the previously described machinery to generate new prices for derivatives such that they are consistent with what is observed. The procedure of looking up prices and cross checking with BSM reveals a very important insight about volatlity which we will analyze next.

Implied Volatility Surface

If the assumptions behind the BSM model were true we would expect volatility of stock prices to be constant over time or, stated otherwise, the implied volatility (IV) surface would be flat [7] [10]. For general scenarios, this assumption is inadequate. Stock prices jump, their observed distributions have fat tails compared to those of a geometric Brownian motion. The volatility surface is a function of strike price K, and time to maturity, T, and is defined in terms of the call option partial differential equation as

$$C(S, K, T) \equiv BS(S, T, r, K, \sigma(K, T)) \tag{3.24}$$

where $C(S, K, T)$ are the observed call option prices on the market with time-to-maturity T and strike K.

Thus, we define the volatility of the stock price as a function that when substituted in the BSM formula it will produce the market price $C(S, K, T)$. This may seem circular but it is a way to price assets consistently. Importantly, the volatility appearing in the BSM formula, called implied volatility, is different to the realized (historical) volatility of the stock price. Implied volatility is a measure of the expectation of the market participants of how realized volatility will act like in the future. IV is not a directly observable quantity and has to be inferred from market data. Next, the solution for the IV surface is guaranteed because the BSM formula is monotonously increasing as a function of σ. A flat volatility surface would have the form $\sigma(K, T) = \sigma$ for all K and T. In reality though, the surface is full of detail and varies over time. We mention that, $\sigma(K, T)$ has to satisfy several arbitrage constraints ([7] for more information).

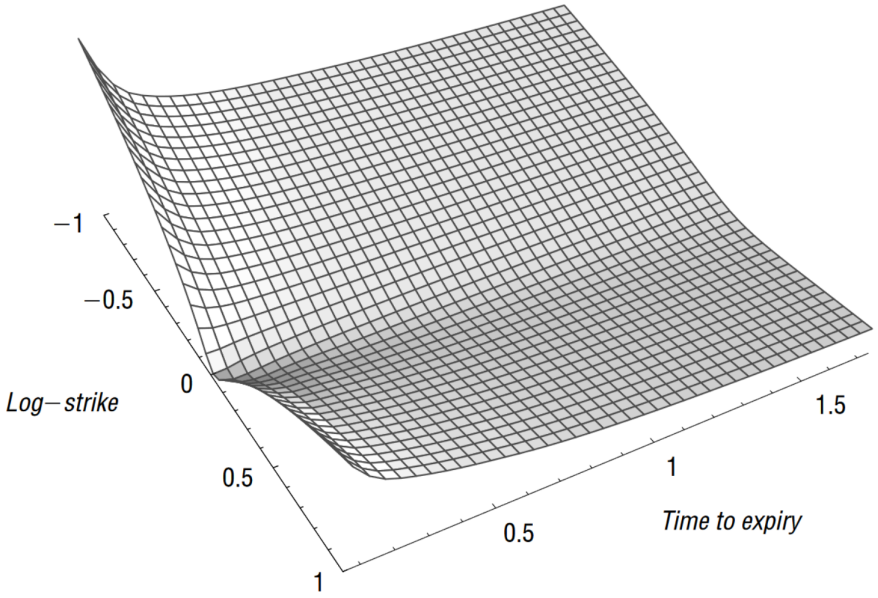


Figure 3.2: Implied volatility surface displaying the smile feature. For fixed time to expiry, we notice the log strike, $\log K$, has a distinct smile profile which is less noticeable as we increase time to maturity, Source: [61]

Take a slice of the volatility surface at a fixed time to expiry. The plot obtained is called volatility *smile*. By comparison, if we take a slice of the surface at fixed strike price K, the implied volatility can either increase or decrease with time to expiry. In general, for $T \to \infty$, the surface $\sigma(K, T)$ tends to a constant. For a T commonly used in practice, volatility tends to be higher with short term options than long term options.

These features are robust with different implementations of BSM. Notably, the traditional explanation behind the skew of the Implied Volatility surface are

- stock price changes happen in jumps, the downward jumps being generally larger and more frequent
- when market prices go down, volatility tends to rise
- investors usually engage in protecting their portfolios with put options with strike prices lower than the market price (out of the money puts).
- The Leverage effect [7] which, in short, means that the volatility of the equity of a company (stocks) should increase as the equity decreases. This happens because the value of the company, equity plus debt, is a much better candidate for geometric Brownian motion than just equity.
- There is a risk of default of the issuer of the stocks in which case the stocks become worthless

The multiple features of the IV surface have motivated the development of stochastic volatility models (SV). In these models, not only does the stock price follow a stochastic process, but the volatility of the stock process follows its own stochastic process.

There is a generous literature providing evidence that volatility is roughly log-normally distributed [10]. For these models in particular, the implied volatility function $\sigma(K, T)$

is statistically independent to $\frac{\partial\sigma(K,T)}{\partial K}$, and this in return can help quantify skew easily. Even though options are traded across a wide range of strikes and times to expiry, most of the complexity of the IV surface has been shown to be quantified by a few factors, usually three [10] called level, term structure and skew.

So far, because of the assumptions of BSM we may have gotten the impression that once we protect (hedge) an investment portfolio against the volatility of the underlying stock price with standard European (vanilla) options then no more protection is needed. This reasoning is flawed because sources of volatility come from both underlying and the option itself [62]. To ensure a risk free portfolio we assume that we constantly buy and sell options and stock to offset losses coming from any particular stock price trajectory. This removes the stock price volatility (realized) but exposes us to the risk (implied volatility) of the option prices which does indeed affect the volatility of the profits and loss associated with an investment portfolio. By contrast, a portfolio consisting of only stock is exposed only to realized volatility. This is where stochastic volatility models do indeed find their key application. Their objective does not have anything to do with realized volatility, but with the dynamics of implied volatility.

3.5 Heston model and Stochastic Volatility

This section is based on Chapter 2 of [61]. Suppose that the stock price S and its variance v satisfy the following stochastic differential equations (SDEs):

$$dS_t = \mu_t S_t dt + \sqrt{v_t} S_t dW_1 \tag{3.25}$$

$$dv_t = -\lambda\left(v_t - \theta\right) dt + \eta\sqrt{v_t} dW_2 \tag{3.26}$$

where μ_t is the (deterministic) instantaneous drift of stock price returns, v_t is the price variance parameter, η is the volatility of volatility, λ controls the speed of reversion to historical average θ, and ρ is the correlation between random stock price returns and changes in v_t. dW_1 and dW_2 are Wiener processes. Heston model takes the volatility to obey a mean reversion process, or CIR process (section 1.3.8).

The standard time-dependent volatility version of the Black Scholes formula may be retrieved in the limit $\eta \to 0$. This is an important fact, as we know that BSM matches market data in stable market periods.

The option valuation formula in the Heston case will be

$$\begin{aligned}\frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \rho\eta v S \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\eta^2 v \frac{\partial^2 V}{\partial v^2} + rS\frac{\partial V}{\partial S} - rV \\ = \lambda(v-\theta)\frac{\partial V}{\partial v}\end{aligned} \tag{3.27}$$

where V is the option value and the other constants have been defined previously.

The Heston model owes its popularity to its capacity to output most quantities of interest for European options in closed form solution, most notably its characteristic function

(Fourier transform of probability density) and vanilla option prices. Importantly, this makes it amenable to efficient computational implementation. Moreover, the Heston process is a semimartingale and Markovian. This means the no arbitrage condition is satisfied.

Heston Characteristic function

In what follows we quote the Heston characteristic function

$$\phi(u,T) = E[exp(iu \log S_T)] = \exp\left[A(u) + B(u)v_0\right] \tag{3.28}$$

$$A(u) = \frac{\lambda\theta}{\eta^2}\left[(\beta - d)T - 2\ln\left(\frac{ge^{-dT}-1}{g-1}\right)\right] \tag{3.29}$$

$$B(u) = \frac{\beta - d}{\eta^2}\left(\frac{1-e^{-dT}}{1-ge^{-dT}}\right) \tag{3.30}$$

$$d = \sqrt{\beta^2 - 4\hat{\alpha}\gamma}, \tag{3.31}$$

$$g = \frac{\beta - d}{\beta + d}, \tag{3.32}$$

$$\hat{\alpha} = -\frac{1}{2}u(u+i), \quad \beta = \lambda - iu\eta\rho, \quad \gamma = \frac{1}{2}\eta^2 \tag{3.33}$$

For more information on derivation please consult [10]. Fortunately, there is a simple way to get from the characteristic function to the European call price. For this we use the formula from [63] which gives the European call prike for strike K and maturity T via

$$C(K,T) = S_0 e^{-qT} - \frac{Ke^{-rT}}{\pi}\int_0^\infty \Re e\left[e^{\left(iu+\frac{1}{2}\right)k}\phi\left(u - \frac{i}{2}, T\right)\right]\frac{du}{u^2 + \frac{1}{4}}, \tag{3.34}$$

where $k = \ln(S_0/K) + (r-q)T$, and r, q are the interest and dividend rates specifically.

Comments on the Heston model

The Heston model, upon appropriate calibration, was shown to reproduce not only the volatility smile and thus replicates the IV shape quite well, but also replicates fat tails and leverage effect. Furthermore, the Heston parameters have an easily understandable interpretation. The η parameter (volatility of volatility) changes the convexity of the smile, the correlation ρ rotates the smile, and $\nu_0, \hat{\nu}, \lambda$ dictate the overall level of the IV surface over time. Interestingly, the Heston model is incomplete, in the sense that there is no unique PDE for option price valuation. This is a consequence of there being two sources of uncertainty, the two Brownian motions, and only one risky asset to trade [64]. In addition to this, a downside to the Heston model is that it gives rives to unrealistic dynamics especially for short time to maturity options [61]. In particular, as shown in

Figure 3.3, the volatility surface of S&P 500 has the $\log k = 0$ (strike price is equal to stock price, also called at-the-money) term structure

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \sigma_{BS}(k, \tau) \right|_{k=0} \tag{3.35}$$

well-approximated by a power-law function of time to expiry τ. In comparison, Heston generate a term structure of at-the-money skew that is constant for small τ.

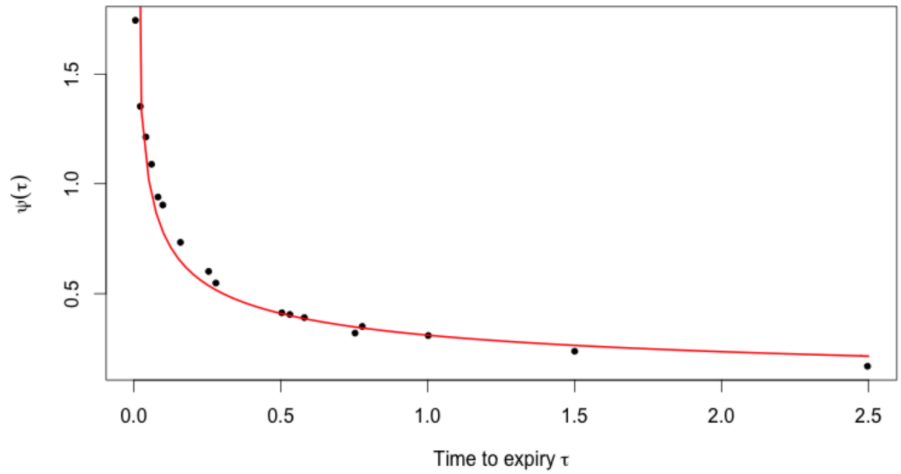


Figure 3.3: Plot of volatility skew of at the money (strike K is equal to stock price) SPX volatility surface. The fitting function is a decaying exponential of exponent close to -0.4 Source: [12]

Moreover, it was shown [65], by analysis of historical SPX stocks prices, that both daily and monthly calibrations of all the Heston parameters $(\hat{\nu}, \rho, \theta, \lambda)$ reveal they are actually time varying, which is not accounted for in the Heston formulation. In addition to this, the Heston model significantly overestimates the realized variance in the sense that averages of the implied variance over the same time period are even twice as high, depending on the time windows chosen. If we also look at the forward variance generated by the Heston model (the conditional probability that variance will take a certain value in the future given the variance today, also called spot variance), the forward volatility smile will diverge from empirical observation, especially for strikes larger than the spot price (the current stock price).

The mentioned shortcoming of the Heston model have been addressed in many other models of financial time series such as SABR, ARCH, GARCH etc but these all suffer from the curse of dimensionality (difficult to calibrate models with many parameter). All these alternatives, even though successful in replicating many stylized facts of financial markets, have not distanced themselves from Brownian motion as a fundamental building block. What we know for a fact about trajectories built using process based on Brownian motion is that they display a certain level of regularity and smoothness. However, it has been shown [12] that log volatility models display rough paths similar to ones generated by fBM, introduced in chapter 2, with a Hurst index of about 0.1. Furthermore, there is significant evidence that historical volatility is much rougher than that generated by

Markov processes [13]. Given this evidence, the flawed short time to maturity at the money skew of the Heston model along with the much sought after long term memory of fBM which stochastic volatility models generally fail to account for, we will look into the application of fBM for modelling volatility [66].

3.6 Rough Heston model

We can consider the volatility to follow a rough (that exhibits characteristics of fBM) process which leads to the Rough Heston model [12] [13]

$$dS_t = S_t\sqrt{v_t}dW_t \tag{3.36}$$

$$v = v_0 + \frac{1}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\lambda\left(\theta - v_s\right)ds + \frac{\eta}{\Gamma(H+1/2)}\int_0^t (t-s)^{H-1/2}\sqrt{v_s}dW_s \tag{3.37}$$

$$\langle dW_t, dW_s\rangle = \rho dt \quad and \quad H \in (0, 1/2) \tag{3.38}$$

The distinction between volatility models is based on the dynamics of σ_t. For Black Scholes model, σ_t is either a constant or a function of time t. Stochastic volatility models take σ_t to be a continuous Brownian semi martingale. The alternative to stochastic volatility models (Dupire) [12] are the local volatility models that are made to fit the market data exactly, although these frequently generate wrong volatility profiles which do not match future data points. Apart from dynamics of volatility another important trait of financial time series is roughness of sample paths. Volatility modelled on fBM provides a solution to this issue, while also having the benefit of very few parameters. Moreover, the persistence of the skew of the IV surface for different maturity points us towards volatility processes which are time homogenous (with time and price independent parameters), of which fBM is an example. fBM volatility has been shown to recreate the term structure (skew) of implied volatlity in accordance to the observed data for both short and long maturities.

Now, a critical aspect of fractional volatility models is they make for inefficient simulations. An alternative to Heston, called the Lifted Heston model, considers volatility to be generated by the superposition of n CIR processes and still mananges to be faster than the Rough Heston simulation for $n = 20$.

Rough Heston Characteristic function

The derivation of the characteristic function for the Rough Heston model is quite involved and such we will only quote the main results of [13], namely

$$E\left[e^{iuX_t}\right] = \exp\left(g_1(u,t) + V_0 g_2(u,t)\right) \tag{3.39}$$

where

$$g_1(a,t) = \theta\lambda\int_0^t h(a,s)ds, \quad g_2(a,t) = I^{1-\alpha}h(a,t) \tag{3.40}$$

and h is a solution of the following fractional Riccati equation:

$$D^{\alpha}h = \frac{1}{2}\left(-a^2 - ia\right) + (ia\rho\eta - \lambda)h(a,s) + \frac{\eta^2}{2}h^2(a,s), \quad I^{1-\alpha}h(a,0) = 0, \tag{3.41}$$

with D^{α} and $I^{1-\alpha}$ the fractional derivative and integral operators.

$$I^r f(t) = \frac{1}{\Gamma(r)}\int_0^t (t-s)^{r-1} f(s)ds, \quad r \in (0,1], \tag{3.42}$$

$$D^r f(t) = \frac{1}{\Gamma(1-r)}\frac{d}{dt}\int_0^t (t-s)^{-r} f(s)ds, \quad r \in [0,1). \tag{3.43}$$

To note, for $\alpha < 1$, the solutions of such Riccati equations are no longer explicit. For $\alpha = 1$ we actually transition into the classical Heston regime. Fortunately, this integral can be solved using numerical integration techniques, which we will describe in the next chapter.

3.7 Forward variance

With the development of the financial derivatives market, the introduction of exotic options with more complex dynamics became a certainty. In particular, the prices of Napoleons and reverse cliquets [65] depend not only on the behaviour of the underlying but also on the dynamics of implied volatilities. Namely, it became important to have models that match empirical data for the spot/volatility correlation, the dynamics of the term structure (smile) of the volatility of volatility and the forward skew, which we already know that models like Heston are not suitable for. Notably, the introduction of such options allows for hedging against dynamics of volatility, which we already now is not possible with implied volatilities as they are unobservable. Let us first look into the first tradable asset based on volatility, called a Variance Swap (VS).

A variance swap is a financial derivative of maturity T, observed at time t, whereby the realized variance is paid in exchange for a fixed strike called the "variance swap (VS) variance" V_t^T. V_t^T is chosen such that the VS has a value of 0 at time t.

Realized variance is defined as the sum of squares of daily log-returns of a stock price $(S_t)_{t\geq 0}$ for the period $[t,T]$ with business days $t_0 = t < \ldots < t_n = T$ such that

$$RV^{t,T} \equiv \frac{1}{T-t}\sum_{t_i \in [t,T]}^{n}\left(\ln\frac{S_{t_1}}{S_{t_{i-1}}}\right)^2, \tag{3.44}$$

where $T-t$ is measured in years. Note, there are alternative conventions for the realized variance but we prefer this one as it is unbiased. The market convention for the payoff of a VS is

$$RV^{t,T} - V_t^T, \tag{3.45}$$

where V_t^T is called the VS volatility for maturity T and is chosen such that the initial value at time t of the payoff is 0. Now, let us define forward variances.

Let $T_1 < T_2$ be two maturities and $V_t^{T_1}, V_t^{T_2}$ the corresponding implied VS variances at time $t < T_1, T_2$. Then the forward VS variance $V_t^{T_1,T_2}$ is defined as

$$V_t^{T_1,T_2} := \frac{(T_2 - t)\, V_t^{T_2} - (T_1 - t)\, V_t^{T_1}}{T_2 - T_1}. \tag{3.46}$$

It can easily be shown that the forward VS variance has no drift. Moreover, by considering $T_1 = T$ and $T_2 = T + dT$ and taking the limit $dT \to 0$ we can easily see

$$V_t^{T,T} = \partial_T[(T - t)V_t^T] \equiv \xi_t^T, \tag{3.47}$$

where ξ_t^T is called the variance swap curve or instantaneous forward variance. Based on this definition we can easily reformulate the VS variance and the forward VS variance as

$$V_t^T = \frac{1}{T - t} \int_t^T \xi_t^u du, \quad t < T, \tag{3.48}$$

$$V_t^{T_1,T_2} = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_t^u du, \quad t < T_1 < T_2. \tag{3.49}$$

Because of the useful property of no drift, forward variance modelling concentrates on the forward variance curve ξ_t^T. A model which addresses all the issues mentioned so far is the Bergomi model which we introduce in the following.

3.8 Bergomi and Rough Bergomi models

So far we know that the forward variance curve ξ_t^T has to be driftless, nonnegative and bounded and otherwise we are free to choose the process that drives it. Naturally, we prefer the ξ_t^T to have a small number of factors. As a start, pick ξ_t^T to be lognormally distributed and to have its volatility a function of $T - t$ (time translation invariant).

$$d\xi_t^T = \omega(T - t)\xi_t^T dW(t), \tag{3.50}$$

where dW is a Wiener process. This shape of the forward variance curve has the disadvantage that for a general function $\omega(.)$ we would need to know the full history of $dW(t)$ so as to compute all possible ξ_t^T. This breaks would break the simple Markovian assumption. Based on this, the one factor Bergomi model is defined such that

$$\omega(T - t) = \omega \cdot e^{-k_1(T-t)}, \tag{3.51}$$

and thus

$$d\xi_t^T = \omega e^{-k_1(T-t)} \xi_t^T dW(t). \tag{3.52}$$

Fortunately we can use Ito lemma to get this expression to an alternative shape

$$\xi_t^T = \xi_0^T \exp\left(\int_0^t \omega e^{-k_1(T-u)} dW(u) - 0.5 \int_0^t \omega^2 e^{-2k_1(T-u)} du\right)$$
$$= \xi_0^T \exp\left(\omega e^{-k_1(T-t)} \int_0^t e^{-k_1(t-u)} dW(u) - \frac{\omega^2}{2} e^{-2k_1(T-t)} \int_0^t \omega^2 e^{-2k_1(t-u)} du\right) \quad (3.53)$$

Defining

$$X(t) \equiv \int_0^t e^{-k_1(t-u)} dU(u) \quad (3.54)$$

we get:

$$\xi_t^T = \xi_0^T \exp\left(\omega e^{-k_1(T-t)} X(t) - \frac{\omega^2}{2} e^{-2k_1(T-t)} \int_0^t e^{-2k_1(t-u)} du\right) \quad (3.55)$$

$X(t)$ is actually an Ornstein-Uhlenbeck process (section 1.3.8), since it is the solution of the following SDE:

$$dX(t) = -k_1 X(t) dt + dU(t) \quad with \quad X(0) = 0. \quad (3.56)$$

Moreover

$$Var\left[X(t)^2\right] = \int_0^t e^{-2k_1(t-u)} du \quad (3.57)$$

and so this brings ξ_t^T into its final shape

$$\xi_t^T = \xi_0^T \exp\left(\omega e^{-k_1(T-t)} X(t) - \frac{\omega^2}{2} e^{-2k_1(T-t)} Var\left[X(t)^2\right]\right). \quad (3.58)$$

As the one factor Bergomi model has its only depence upon a Markovian process, it is itself Markovian. However, it is not flexible enough to be empirically accurate. Luckily, there is a way to generalize the Bergomi model to have n factors so as to account for the full complexity of the forward variance.

Subsequently, we can use the properties of the fBM to also simulate forward variance processes. Consider two dimensional correlated Brownian motions W and B.The asset price S is supposed to follow a Geometric Brownian motion and thus

$$dX_t = -\frac{1}{2} v_t dt + \sqrt{v_t} dW_t, \quad (3.59)$$

where $X = \log S$, $v_t \geq 0$ is the instantaneous spot variance process. Let $\xi_t^u, u \geq t$ be the instantaneous forward variance for date u observed at time t; in particular $\xi_t^t = V_t$ corresponds to the spot variance. The rough Bergomi model [67] where the forward variance follows

$$d\xi_t^u = \xi_t^u \eta \sqrt{2\alpha + 1} (u-t)^\alpha dB_t, u \geq t \quad (3.60)$$

where W and B have correlation $\rho, \alpha H - \frac{1}{2} \in \left(-\frac{1}{2}, 0\right)$ is a negative exponent depending on the Hurst exponent $H \in \left(0, \frac{1}{2}\right)$ of the underlying fractional Brownian motion, and η is a positive parameter depending on H.

The rBergomi stochastic volatility model takes the form

$$dX_t = -\frac{1}{2}v_t dt + \sqrt{v_t}dW_t \tag{3.61}$$

$$d\xi_t^u = \xi_t^u \eta\sqrt{2\alpha+1}(u-t)^\alpha dB_t \tag{3.62}$$

where $\alpha = H - \frac{1}{2} \in \left(-\frac{1}{2}, 0\right)$, and $d\langle W, B\rangle_t = \rho dt$

Rough Forward Variance models replicate the implied volatility surface to a greater degree of accuracy than the classical counterparts and it achieves this using only the three parameters ρ, η, H. In particular, rBergomi reproduces very well the power law for at the money volatility skew and is consistent with the time invariance of the IV surface features [68] [67] [69].

Chapter 4

Monte Carlo implementations

In this chapter we focus on the implementation of two stochastic volatility processes in particular, the Heston model and the Rough Heston model. Simulating both these processes presents several difficulties. Once discretized the Heston model can produce negative volatility. Conditions can be built around it, although they do introduce bias. Furthermore, for an efficient simulations additional conditions needs to included such as those prescribed by the Quadratic Exponential scheme. Thus in what follows we describe a few Monte Carlo approaches that address the previous concerns.

4.1 Simulating the Heston model

We start by outlining the Euler discretization scheme. This section follows closely section 3 and 4 from [15].

4.1.1 The Euler Scheme

First, denote $\log S = X$ as a proxy for the asset price. This is considered as an appropriate step as log returns of an asset are relatively close to a Geometric Brownian Motion [15], and we will work with X instead of S when simulating Heston. Next, we use $\hat{X}$ and $\hat{v}$ to mark the discrete version of each process. Then the Euler scheme can neatly be written as

$$\hat{X}(t+\Delta) = \hat{X}(t) - \frac{1}{2}\hat{v}(t)\Delta + \sqrt{\hat{v}(t)}Z_X\sqrt{\Delta}, \tag{4.1}$$

$$\hat{v}(t+\Delta) = v(t) + \lambda(\theta - \hat{v}(t))\Delta + \eta\sqrt{\hat{v}(t)}Z_V\sqrt{\Delta}, \tag{4.2}$$

where Z_X and Z_V are standard Gaussian variables with correlation ρ, and $\Delta = T/N$ represents a time step of N total time steps. A run of two correlated random variables can easily be produced via

$$Z_V = W_1 \quad , \quad Z_X = \rho Z_V + \sqrt{1-\rho^2}W_2, \tag{4.3}$$

where W_1, W_2 are two uncorrelated runs of a normal distribution with mean 0 and variance 1. As pointed out previously, the volatility process can become negative and, because of the square root, we cannot allow this to happen. An easy fix is to introduce the max function, $x^+ = \max(x, 0)$, in order to create a barrier at 0. Thus

$$\hat{X}(t+\Delta) = \hat{X}(t) - \frac{1}{2}\hat{v}(t)^+\Delta + \sqrt{\hat{v}(t)^+}Z_X\sqrt{\Delta}, \tag{4.4}$$

$$\hat{v}(t+\Delta) = \hat{v}(t) + \lambda\left(\theta - \hat{v}(t)^+\right)\Delta + \eta\sqrt{\hat{v}(t)^+}Z_V\sqrt{\Delta}, \tag{4.5}$$

This is called the full truncation Euler scheme. Its main feature is that when the process reaches 0 then it get a deterministic upward drift, that does introduce bias in the estimation. As we will see, for realistic Heston parameters the Euler scheme will not perform so well, and thus we shall look more carefully at ways to improve the efficiency of our simulation.

4.1.2 Quadratic exponential method

Given that there are two SDE driving the Heston process, we need to the address them in order. The stock price is dependent on the volatility process and thus we need to simulate this first. As was described before, the volatility process for the Heston model is the CIR process, outlined in section 1.3.8. Namely, using the previous notation

$$dv_t = -\lambda\left(v_t - \theta\right)dt + \eta\sqrt{v_t}dW_2, \tag{4.6}$$

where ν_t is the volatility of the stock price. Now, we can quote the conditional expectation of volatility and variance of volatility from section 1.3.8

$$E[\nu_s \mid \nu_t] = \nu_t e^{-\lambda(s-t)} + \theta\left(1 - e^{-\lambda(s-t)}\right), \tag{4.7}$$

$$Var[\nu_s \mid \nu_t] = \nu_t\left(\frac{\eta^2}{\lambda}\right)\left(e^{-\lambda(s-t)} - e^{-2\lambda(s-t)}\right) + \theta\left(\frac{\eta^2}{2\lambda}\right)\left(1 - e^{-\lambda(s-t)}\right)^2. \tag{4.8}$$

More than this, the joint probability distribution function is

$$f_2(\nu_s, s; \nu_t, t) = ce^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q\left(2(uv)^{1/2}\right), \tag{4.9}$$

where

$$c = \frac{2\lambda}{\eta^2\left(1 - e^{-\lambda(s-t)}\right)}, \quad u \equiv c\nu_t e^{-\lambda(s-t)}, \quad v \equiv c\nu_s, \quad q = \frac{2\lambda\theta}{\eta^2} - 1$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. The joint pdf function is the noncentral chi-square, $\chi^2[2cv_s; 2q+2, 2u]$, with $2q+2$ degrees of freedom, parameter of noncentrality $2u$ and mean $2(q+u+1)$.

For small volatility ν_t the $u \to 0$, and thus the distribution becomes proportional to the central chi squared distribution of $4\frac{\lambda\hat{\nu}}{\eta}$ degrees of freedom. Note the pdf of a central chi-square distribution with v degrees of freedom is

$$f_{\chi^2}(x; v) = \frac{1}{2^{v/2}\Gamma(v/2)} e^{-x/2} x^{v/2-1} \tag{4.10}$$

In practice, $4\frac{\lambda\theta}{\eta} \ll 2$, so this makes distribution peak around 0. However, generally approximating a chi squared distribution with a Gaussian is bad when ν is small. The jist of the lower bound approximation is to have the central chi squared density approximation outlined above have a portion its probability mass below zero concentrated into a delta function at the origin plus an exponential tail proportional to $e^{-x^2/2}$ as described in equation to 4.10. For this, we consider

$$Pr(\hat{\nu}(t+\Delta) \in [x, x+dx]) \approx \left(p\delta(0) + \beta(1-p)e^{-\beta x}\right) dx, \quad x \geq 0 \tag{4.11}$$

where δ is a Dirac delta-function, and p and β are non-negative constants to be determined. Sampling according to the probability above is straightforward and efficient. To see this, first we integrate 4.11 to generate a cumulative distribution function

$$\Psi(x) = Pr(\hat{\nu}(t+\Delta) \leq x) = p + (1-p)\left(1 - e^{-\beta x}\right), \quad x \geq 0. \tag{4.12}$$

The inverse of Ψ is then

$$\Psi^{-1}(u) = \Psi^{-1}(u; p, \beta) = \begin{cases} 0, & 0 \leq u \leq p, \\ \beta^{-1} \ln\left(\frac{1-p}{1-u}\right), & p < u \leq 1. \end{cases} \tag{4.13}$$

As a result QE implies

$$\hat{\nu}(t+\Delta) = \Psi^{-1}\left(U_V; p, \beta\right) \tag{4.14}$$

where U_V is a draw from a standard uniform distribution.

Importantly, the non-central chi-square distribution approaches a Gaussian distribution as the non-centrality parameter, $2u \to \infty$ [15]. This means that for large enough u a good approximation to v_t would be a Gaussian distribution that enjoys the expectation and variance in equations 4.7 and 4.8. An even better approximation in the moderate to high v_t regime, the non central chi distribution can be well approximated by a power function applied to a Gaussian variable [15]. For large enough $\hat{\nu}_t$ the QE scheme considers

$$\hat{\nu}(t+\Delta) = a\left(b + W_t\right)^2 \tag{4.15}$$

where W_t is a standard Gaussian random variable, and a and b are certain constants, to be determined by moment-matching. a and b will depend on the the time-step Δ and $\hat{\nu}_t$, as well as the parameters in the SDE for ν.

Now, to determine a, b, p, and β we perform moment-matching. Thus, let m and s be as defined by

$$E[\hat{\nu}(t+\Delta)] = m = \theta + (\hat{\nu}_t - \theta)e^{-\lambda\Delta}, \tag{4.16}$$

$$Var[\hat{\nu}(t+\Delta)] = s^2 = \frac{\hat{\nu}_t\eta^2 e^{-\lambda\Delta}}{\lambda}\left(1-e^{-\lambda\Delta}\right) + \frac{\theta\eta^2}{2\lambda}\left(1-e^{-\lambda\Delta}\right)^2, \tag{4.17}$$

and

$$\psi = \frac{s^2}{m^2}. \tag{4.18}$$

Provided that $\psi \leq 2$, set

$$b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1}-1} \geq 0, \quad a = \frac{m}{1+b^2}. \tag{4.19}$$

Similarly, to compute p and β assume that $\psi \geq 1$ and set

$$p = \frac{\psi - 1}{\psi + 1} \in [0, 1), \tag{4.20}$$

and

$$\beta = \frac{1-p}{m} = \frac{2}{m(\psi+1)} > 0. \tag{4.21}$$

Note that switching between the lower bound approximation to higher bound approximation happens for $\psi \in [1, 2]$. Generally, we can choose a critical ψ_c to switch between the approximations without too much impact on the accuracy of our computation. For proofs related to the computation of the a,b,β,p parameters as well as the accuracy of the bounds of ψ please consult [15].

To summarize, assume that some arbitrary level $\psi_c \in [1, 2]$ has been selected. The QE scheme then implies

- Given $\hat{\nu}(t)$, compute m and s^2 from equations 4.16 and 4.17
- Compute $\psi = s^2/m^2$
- If $\psi \leq \psi_c$ then compute a and b from equations 4.19 and set $\hat{\nu}(t+\Delta) = a\left(b + W_t\right)^2$
- Otherwise, if $\psi > \psi_c$, compute β and p according to equations 4.20, 4.21 and set $\hat{\nu}(t+\Delta) = \Psi^{-1}\left(U_V; p, \beta\right)$, where Ψ^{-1} is given in 4.13.

Let us now focus on the discretization scheme for X. First, it has to satisfy the right correlation with the volatility process. We then write

$$X(t+\Delta) = X(t) + \frac{\rho}{\eta}(v(t+\Delta) - v(t) - \lambda\theta\Delta)$$

$$+ \left(\frac{\lambda\rho}{\eta} - \frac{1}{2}\right)\int_t^{t+\Delta} v(u)du + \sqrt{1-\rho^2}\int_t^{t+\Delta}\sqrt{v(u)}dW \tag{4.22}$$

We recognize the term $\frac{\rho}{\eta}v(t+\Delta)$ as the key driver of correlation between $X(t+\Delta)$ and $v(t+\Delta)$. Next, let us concentrate on the time-integral of v. We simply write

$$\int_t^{t+\Delta} v(u)du \approx \Delta\left[\gamma_1 v(t) + \gamma_2 v(t+\Delta)\right], \tag{4.23}$$

for constants γ_1 and γ_2. The simplest choice is $\gamma_1 = 1, \gamma_2 = 0$, similar to the Euler scheme. A central discretization would set $\gamma_1 = \gamma_2 = \frac{1}{2}$. In our simulation we make this choice. The random variable W is independent of v, conditional on $v(t)$ and $\int_t^{t+\Delta} v(u)du$. Thus

$$\int_t^{t+\Delta} \sqrt{V(u)} dW(u) \tag{4.24}$$

is Gaussian with mean zero and variance $\int_t^{t+\Delta} V(u)du$. Using again the approximation 4.23 we can finally write

$$\hat{X}(t+\Delta) = \hat{X}(t) + \frac{\rho}{\eta}(\hat{v}(t+\Delta) - \hat{v}(t) - \lambda\theta\Delta) +$$

$$+\Delta\left(\frac{\lambda\rho}{\eta} - \frac{1}{2}\right)(\gamma_1\hat{v}(t) + \gamma_2\hat{v}(t+\Delta)) + \sqrt{\Delta}\sqrt{1-\rho^2}\sqrt{\gamma_1\hat{v}(t) + \gamma_2\hat{v}(t+\Delta)} \cdot Z$$

$$= \hat{X}(t) + K_0 + K_1\hat{v}(t) + K_2\hat{v}(t+\Delta) + \sqrt{K_3\hat{v}(t) + K_4\hat{v}(t+\Delta)} \cdot Z \tag{4.25}$$

where Z is a standard Gaussian random variable, independent of $\hat{V}$, and $K_0, \ldots, K_4$ are given by

$$K_0 = -\frac{\rho\lambda\theta}{\eta}\Delta, \quad K_1 = \gamma_1\Delta\left(\frac{\lambda\rho}{\eta} - \frac{1}{2}\right) - \frac{\rho}{\eta}, \tag{4.26}$$

$$K_2 = \gamma_2\Delta\left(\frac{\lambda\rho}{\eta} - \frac{1}{2}\right) + \frac{\rho}{\eta}, \quad K_3 = \gamma_1\Delta\left(1-\rho^2\right), \quad K_4 = \gamma_2\Delta\left(1-\rho^2\right), \tag{4.27}$$

and $K_i, i = 0, \ldots, 4$ depend on the time-step as well as on the constants γ_1 and γ_2. As a final step, the QE scheme suggests the previous X discretization in conjunction with the moment matching volatility scheme described above. In short, the QE prescription is given by

1. Using $\hat{v}(t)$, compute $\hat{v}(t+\Delta)$ using the algorithm above
2. Given $\hat{X}(t), \hat{v}(t)$, and the value for $\hat{v}(t+\Delta)$ then compute $\hat{X}(t+\Delta)$.

4.2 Simulating the Rough Heston model

4.2.1 The Euler Scheme

This section is based on [70]. Choose a splitting of the $[0,T]$ interval marked as $\pi_n = \{0 = t_0^n < t_1^n < t_2^n < \ldots < t_n^n = T\}$ for each $n \geq 1$, with $\Delta t_{k+1}^n \equiv t_{k+1}^n - t_k^n$. We will simulate $(X = \log S, v)$ on the discrete-time grid π_n. For the discrete grid π_n, let us write t_k^n as t_k for simplicity, and denote by $(X^n, v^n) = \left(X_{t_k}^n, v_{t_k}^n\right)_{k=0,1,\cdots,n}$ the corresponding numerical solution, which is given as follows

$$X_{t_k}^n = X_0 + \sum_{i=0}^{k-1}\left(-\frac{1}{2}\left(v_{t_i}^n\right)_+ \Delta t_{i+1}^n + \rho\sqrt{\left(v_{t_i}^n\right)_+}\left(W_{t_{i+1}} - W_{t_i}\right)\right)$$

$$+\sum_{i=0}^{k-1}\left(+\sqrt{1-\rho^2}\sqrt{\left(v_{t_i}^n\right)_+}\left(W_{t_{i+1}}^\perp - W_{t_i}^\perp\right)\right), \tag{4.28}$$

$$v_{t_k}^n = v_0 + \sum_{i=0}^{k-1} \left(K\left(t_k - t_i\right) \lambda \left(\theta - \left(v_{t_i}^n\right)_+ \right) \Delta t_{i+1}^n + K\left(t_k - t_i\right) \eta \sqrt{\left(v_{t_i}^n\right)_+} \left(W_{t_{i+1}} - W_{t_i} \right) \right), \tag{4.29}$$

where the kernel function $K(t) := \Gamma\left(H + \frac{1}{2}\right)^{-1} t^{H-\frac{1}{2}}$. In [70] a weak convergence scheme is provided.

4.2.2 Rough Heston Characteristic Function

Previously, in section 3.6 we brought to the readers attention that the core of Rough Heston characteristic function [13] of equation 3.39 is the function h which follows Volterra equation

$$h(a,t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} F(a, h(a,s)) ds, \tag{4.30}$$

$$F(a,x) = \frac{1}{2}\left(-a^2 - ia\right) + (ia\rho\eta - \lambda)x + \frac{\eta^2}{2}x^2. \tag{4.31}$$

To solve it, we choose the well-known fractional Adams method investigated in [70] [71] [72]. Consider $g(a,t) = F(a, h(a,t))$. Over a regular discrete time-grid $(t_k)_{k\in N}$ with mesh $\Delta\,(t_k = k\Delta)$ approximate

$$h\left(a, t_{k+1}\right) = \frac{1}{\Gamma(\alpha)} \int_0^{t_{k+1}} \left(t_{k+1} - s\right)^{\alpha-1} g(a,s) ds \tag{4.32}$$

by

$$\frac{1}{\Gamma(\alpha)} \int_0^{t_{k+1}} \left(t_{k+1} - s\right)^{\alpha-1} \hat{g}(a,s) ds \tag{4.33}$$

where

$$\hat{g}(a,t) = \frac{t_{j+1} - t}{t_{j+1} - t_j} \hat{g}\left(a, t_j\right) + \frac{t - t_j}{t_{j+1} - t_j} \hat{g}\left(a, t_{j+1}\right), \quad t \in \left[t_j, t_{j+1}\right), \quad 0 \le j \le k. \tag{4.34}$$

As a result we can summarize the previous steps in

$$\hat{h}\left(a, t_{k+1}\right) = \sum_{0\le j\le k} a_{j,k+1} F\left(a, \hat{h}\left(a, t_j\right)\right) + a_{k+1,k+1} F\left(a, \hat{h}\left(a, t_{k+1}\right)\right), \tag{4.35}$$

with

$$a_{0,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)} \left(k^{\alpha+1} - (k-\alpha)(k+1)^\alpha \right), \tag{4.36}$$

$$a_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)} \left((k-j+2)^{\alpha+1} + (k-j)^{\alpha+1} - 2(k-j+1)^{\alpha+1} \right), \quad 1 \le j \le k, \tag{4.37}$$

$$a_{k+1,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+2)}. \tag{4.38}$$

Notice $\hat{h}\left(a, t_{k+1}\right)$ is on both sides of the equation 4.35. We then need to compute a pre-estimation of $\hat{h}\left(a, t_{k+1}\right)$ to have a functional scheme. For this we follow [73], and use the so called predictor denoted by $\hat{h}^{P}\left(a, t_{k+1}\right)$ is defined by

$$\hat{h}^{P}\left(a, t_{k+1}\right) = \frac{1}{\Gamma(\alpha)} \int_{0}^{t} (t-s)^{\alpha-1} \tilde{g}(a, s) ds, \tag{4.39}$$

with

$$\tilde{g}(a, t) = \hat{g}\left(a, t_{j}\right), \quad t \in \left[t_{j}, t_{j+1}\right), \quad 0 \leq j \leq k \tag{4.40}$$

Therefore,

$$\hat{h}^{P}\left(a, t_{k+1}\right) = \sum_{0 \leq j \leq k} b_{j,k+1} F\left(a, \hat{h}\left(a, t_{j}\right)\right), \tag{4.41}$$

where

$$b_{j,k+1} = \frac{\Delta^{\alpha}}{\Gamma(\alpha+1)} \left((k-j+1)^{\alpha} - (k-j)^{\alpha}\right), \quad 0 \leq j \leq k \tag{4.42}$$

To conclude, the scheme for computation of h can be succintly written as

$$\hat{h}\left(a, t_{k+1}\right) = a_{k+1,k+1} F\left(a, \hat{h}^{P}\left(a, t_{k+1}\right)\right) + \sum_{0 \leq j \leq k} a_{j,k+1} F\left(a, \hat{h}\left(a, t_{j}\right)\right), \quad \hat{h}(a, 0) = 0 \tag{4.43}$$

Chapter 5

Results & Discussion

In this chapter we outline the investigation on the Euler discretization scheme in the context of classical Heston model and the Rough Heston Model. We begin by analyzing its performance in the classical Heston model and we highlight its shortcomings. We then review the Quadratic Exponential scheme and emphasize how it can be used to alleviate some of the weaknesses of the Euler scheme. Finally, we examine the efficiency of the Euler scheme in the Rough Heston model and make comments on its range of applicability.

5.1 Classical Heston

For the reader's convenience we rewrite the equations underpinning the Heston model.

$$dS_t = \mu_t S_t dt + \sqrt{v_t} S_t dW_1 \tag{5.1}$$

$$dv_t = -\lambda \left(v_t - \theta\right) dt + \eta \sqrt{v_t} dW_2 \tag{5.2}$$

all the parameters being defined the same as previously.

An important feature resulting from the discretization of the mean reverting process is that volatility can turn negative. This, of course, is unacceptable for our purposes. In the limit of $dt \to 0$, a way to keep the CIR process always positive is to have the parameters satisfy the Feller condition, which is simply stated as $2\lambda\theta > \eta^2$. Unfortunately, this condition is not always satisfied by real markets. As pointed in [15], calibrating the Heston parameters process using data for the market of long dated FX options, reveals they are non compliant with the Feller condition. As such, we will be using the set of parameters suggested in [15] paper, more specifically Table 1, Case 1. Thus, $\lambda = .5$, $\eta = 1$, $\theta = .04$, $\rho = -.9$, $v(0) = .04$, $S(0) = 100$, $r = 0$, $q = 0$ and maturity $T = 10$ years. Notably, we perform $N = 10^6$ simulations for a range of discretization steps per year $n = 1, 2, 4, 8, 16, 32$. As we pointed out previously, the Heston model fits the IV surface very well in the long term and, as a result, this is the range we test it on.

For simulation, we use the methods outlined in Chapter 4, that is the Euler and Quadratic Exponential schemes. In figure 5.1 we show how both these schemes conform to the analytical expectation for the volatility process in the Heston model, captured in equation 4.7. Notice that for the choice $v(0) = \theta$ we should simply get a straight line. This choice of parameters essentially means the volatility no longer has a decaying exponential component back to the historical volatility, and instead the price variance process simply gravitates close to it. Evidently from figure 5.1, the Euler scheme overestimates the analytical value even for small maturities ($T < 1$ years). This is because, by not setting any extra conditions in the Euler scheme the volatility process will reach 0 many times. As a result of this there will be a deterministic upward shift pushing the expectation estimate into larger numbers. This can be alleviated by increasing the number of discretization steps n, however it might also make it prohibitively time consuming depending on the accuracy needed. The Quadratic Exponential scheme, on the other hand, does exactly what it is advertised to do. Given that the first two moments of distribution, the expectation and variance, of the volatility process are built into the scheme, the behaviour is very good at any maturity.

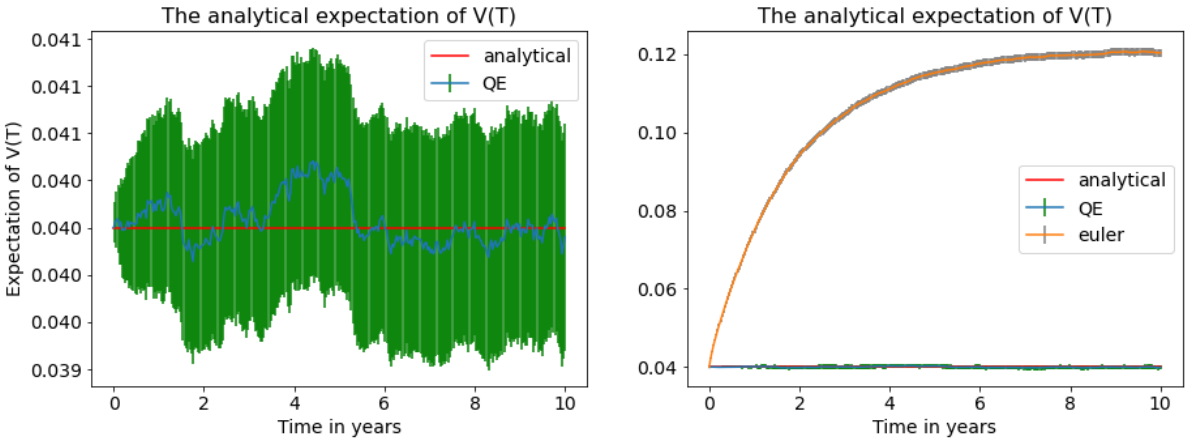


Figure 5.1: The Euler and Quadratic exponential schemes are compared against the analytical conditional expectation of the volatility process. The error bars represent 3 standard deviations around each discretization point. Note, $N = 10^6$ paths were used to infer the standard deviation for any point.

Next, let us have a look at the variance of the volatility process. Similarly to the previous case, the Euler scheme diverges from the analytical value, although it does align with it in the very short term. We also observe that, in contrast to the QE scheme, the Euler scheme diverges more and more with increase in maturity. Even though convergence to analytical behaviour is assured in the limit $dt \to 0$, we see that for a practical number of simulations and discretization steps per year we get much better results with the Quadratic Exponential scheme.

Now, the behaviour of the expectation of $\log S$ process is shown in figure 5.3. As we can suppose from the behaviour of the volatility process, the asset price behaviour of the Euler scheme diverges from analytical behaviour. The very short term behaviour conforms to the analytical formula, only to quickly diverge after the 1 year mark. This

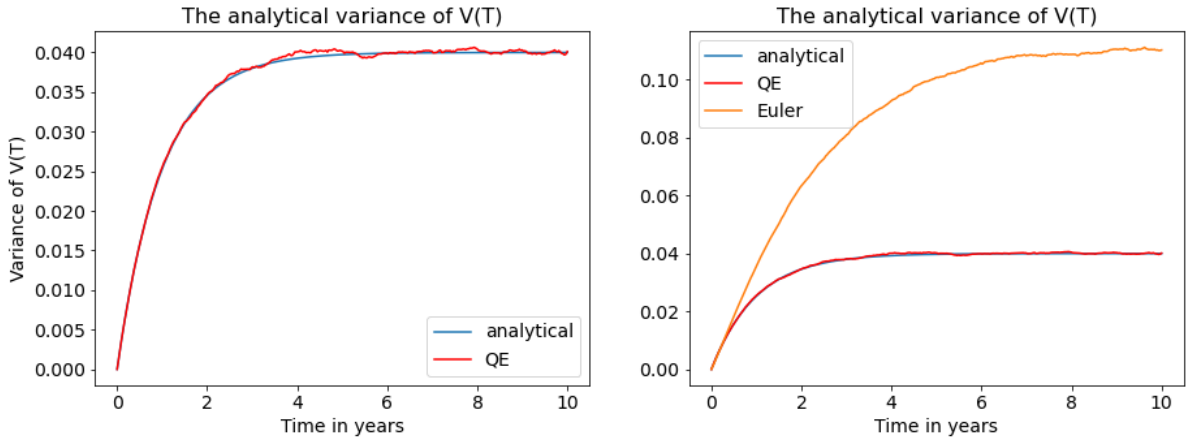


Figure 5.2: Comparison between the analytical conditional variance of volatility in the Heston model and the Euler and QE schemes. Notice how the QE scheme essentially overlaps with the analytical formula.

is to be expected because we set the correlation between volatility and price increments to be negative, and as such, an overestimation into the volatility leads to an underestimation into the $\log S$ process. We see that the QE scheme aligns perfectly with the analytical formula. This is because we applied moment matching also to the log asset price increment.

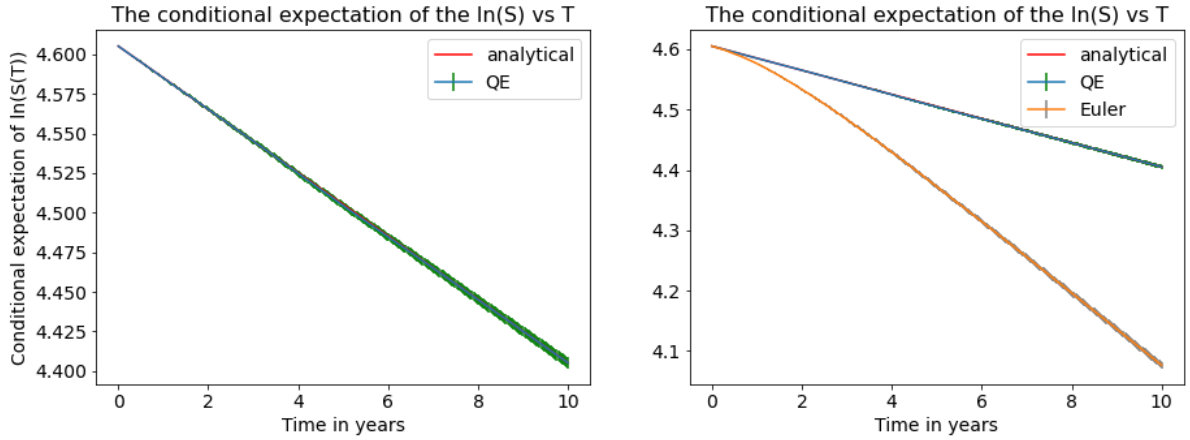


Figure 5.3: The analytical conditional expectation of the $\log S$ process versus the Euler and QE schemes. Again, the error bars signify 3 standard deviations around each discretization point.

Before we explore convergence to the characteristic function we perform one last check on the variance of the $\log S$ process in figure 5.4. The QE scheme outperforms the Euler scheme as expected, although before the 1 year mark both schemes show a very good match to analytical $\log S$ variance. The divergence of the Euler scheme happens similarly quickly to the other moments of distribution.

At this point, we are ready to look at the behaviour of the discretization schemes with respect to the Heston characteristic function. Based on our results so far, we are interested

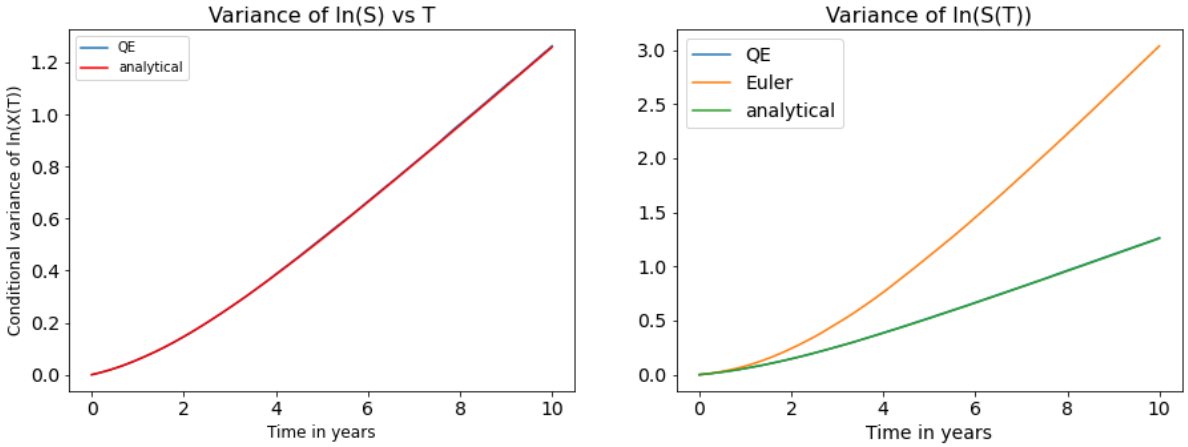


Figure 5.4: Analytical variance of the $\log S$ process versus the QE and Euler implied values. We observe again the QE and analytical values are perfectly overlapping.

in testing the bias of the QE scheme in particular. We prefer not to complete the same test for the Euler scheme as we already know it is biased for maturities larger than 1 year, when we use only up to $n = 32$ discretization steps per year. Surprisingly, from figure 5.5 we see the percentage bias is as low as 1% for maturities up to 10 years. Importantly, the mismatch of 1% shown for large values of U and T is of not concern because this is a region where the Heston characteristic function is very close to 0, and thus relative errors can be easily caused by a lack of attention to round off errors behind numerical integration. We are then very confident that the QE scheme is very well equipped to compute any derivatives based on the Heston model, as it is able to capture all moments of distributions of the $\log S$ process.

Next, we compute the bias e of European call option prices using the Euler and QE schemes for a maturity of $T = 10$ years and strike prices of $K = 70, 100, 140$ in Table 5.1. To estimate the call price $\hat{C}(0)$ we do N independent simulations using both discretization schemes. We label the asset price S at maturity T for separate runs as $\hat{S}^{(1)}(T), \hat{S}^{(2)}(T), \ldots, \hat{S}^{(N)}(T)$. $\hat{C}(0)$ is then

$$\hat{C}(0) \approx \frac{1}{N} \sum_{i=1}^{N} \left(\hat{S}^{(i)}(T) - K \right)^{+}, \tag{5.3}$$

$$e = \hat{C}(0) - C(0) \tag{5.4}$$

where $C(0)$ is the analytical call option price. The standard deviation associated to the call price is then $\sigma = \sqrt{Var(S)}/\sqrt{N}$, where $N = 10^6$ is the sample number. The Quadratic Exponential schemes is always within three standard deviations of the analytical Call price for the Heston model when we use $n = 32$ discretization steps per year or more. The reason why this is important is that we can use the QE scheme to price any derivatives, even those without closed form solutions, in the long term with a high degree of confidence. This means we do not need to refer to an analytical formula, but only simulate the process $N = 10^6$ times for $n = 32$ steps per year resolution or more and we will get an accurate answer. This can routinely be done on a low end consumer laptop within a couple of minutes.

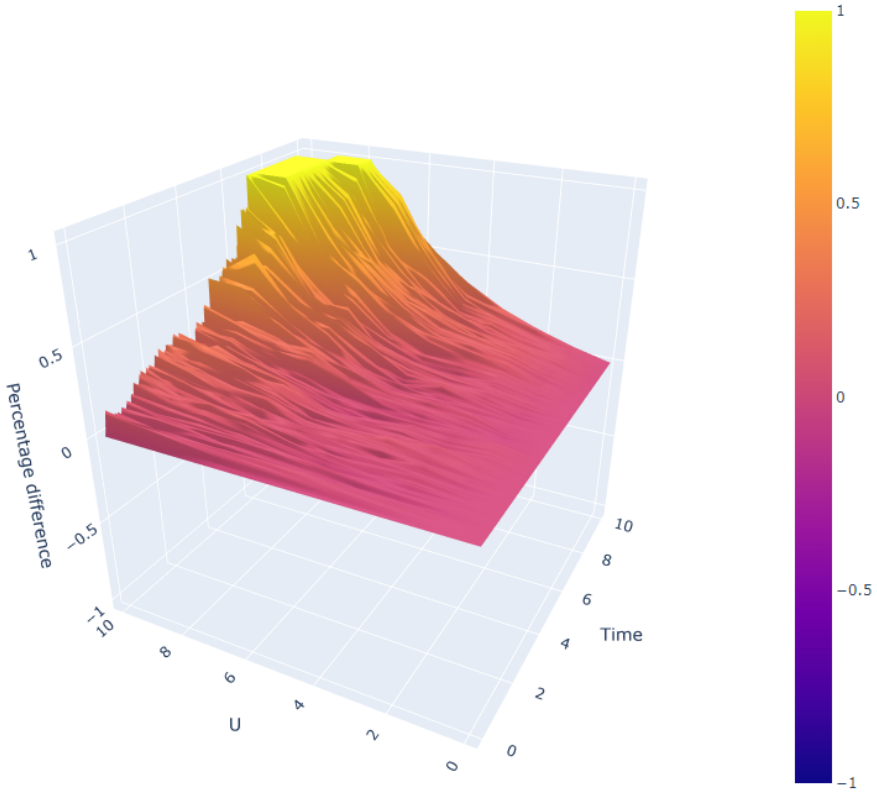


Figure 5.5: Percentage bias between the analytical formula for the Heston characteristic function and the estimated value for $N = 10^6$ simulations based on the QE scheme with $n = 32$ discretization steps per year.

Finally, we compute the IV bias for $K = 100$ at a maturity of $T = 10$ years and we notice that from $n = 10$ steps per year (100 for the whole 10 year process) we are within 1% bias relative bias to the analytical IV surface. This is to be expected as the IV surface is based on the first moment of distribution, namely the expectation value of the asset price. The very good agreement of both the analytical expectation of $\log S$ and the characteristic function indicate a very low IV bias. In conclusion, a moment matching scheme such as QE proves to be a very reliable scheme for pricing any type of Heston derivatives for any long term maturity. Nevertheless, we point out that despite its slower convergence in the long term, the Euler scheme shows signs of alignment with analytical expectation in the short term even for a low resolution discretization. Motivated by its simplicity and the fact that the Rough Heston is best used in the short term, we then go on to use the Euler scheme for its discretization.

	Euler	QE	Euler	QE
	K = 140 (0.2957)		K = 100 (13.0846)	
n = 1	13.4032 (0.0288)	-0.1387 (0.002)	11.7381 (0.0359)	0.0825 (0.012)
n = 2	13.3707 (0.0282)	-0.0585 (0.0023)	12.4697 (0.036)	-0.0674 (0.0127)
n = 4	12.7859 (0.0269)	-0.0256 (0.0024)	12.6717 (0.0356)	-0.155 (0.013)
n = 8	12.0127 (0.0253)	-0.0121 (0.0025)	12.7769 (0.035)	-0.1099 (0.0132)
n = 16	10.994 (0.0234)	-0.0072 (0.0025)*	12.5084 (0.0341)	-0.0533 (0.0132)
n = 32	9.9449 (0.0214)	-0.0033 (0.0025)*	12.0410 (0.0330)	-0.0359 (0.0133)*

	Euler	QE
	K = 70 (35.8497)	
n = 1	0.3277 (0.0401)	0.2677 (0.0216)
n = 2	1.4492 (0.0403)	-0.0941 (0.0221)
n = 4	2.2457 (0.0402)	-0.1612 (0.0222)
n = 8	3.0856 (0.0399)	-0.1108 (0.0223)
n = 16	3.6012 (0.0395)	-0.0299 (0.0223)*
n = 32	3.9032 (0.0390)	-0.0409 (0.0223)*

Table 5.1: The bias of European Option prices computed using the Euler and QE schemes with maturity $T = 10$ years for different strike prices K. The asterisk values are within 3 standard deviations of the analytical value of the call price. All the results are obtained using $N = 10^6$ simulations for various numbers of discretization steps n. The analytical reference values for the option price are indicated near each K.

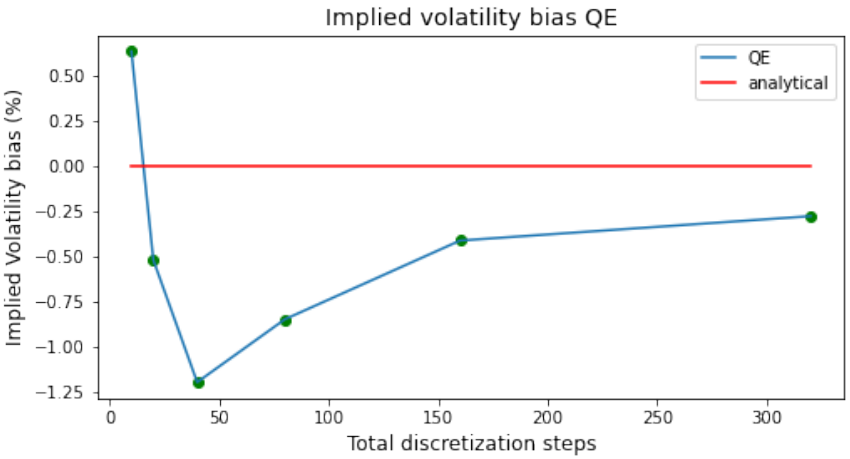


Figure 5.6: Implied volatility bias for $T = 10$ years and $K = 100$ for the QE scheme for various numbers of discretization steps for the entire process.

5.2 Rough Heston

Firstly, to facilitate our discussion we restate the Rough Heston equations

$$dS_t = S_t\sqrt{v_t}dW_t, \tag{5.5}$$

$$v_t = v_0 + \frac{1}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} \lambda \left(\theta - v_s\right) ds + \frac{\eta}{\Gamma(H+1/2)} \int_0^t (t-s)^{H-1/2} \sqrt{v_s} dW_s, \tag{5.6}$$

$$\langle dW_s, dW_t \rangle = \rho dt \quad and \quad H \in (0, 1/2), \tag{5.7}$$

where all the quantities have been defined previously. We will be using the following values for the Rough Heston parameters $\lambda = 0.3, \eta = 0.3, v(0) = 0.02$, $\theta = 1/15, \rho = -0.7, S(0) = 1$, $r = 0$, $q = 0$, $H = 0.1$, based on the choice made in [70]. This is mainly because we will use their results to perform sanity checks to our implementation.

One of the most important differences between the Rough Heston and classical Heston model is that there are no closed form solutions to quantities that we use to asses the performance of our simulations. In the Heston case, because of the analytical solution to the characteristic function it was very easy to find moments of distribution of $\log S$ and to compute option prices. In the Rough Heston case, to make sure our implementation is correct we can address this issue in two ways.

The first is to test Rough Heston in the regime where $H \to 1/2$, where it turns into classical Heston. It can be easily seen from equation 5.6 that in this limit the kernel function approaches 1 and thus we have a very smooth transition from the classical to the rough case. Although this step was performed for benchmarking our implementation we will not focus on it because we are already familiar with the behaviour of the Euler scheme in the classical regime.

The second class of tests focuses on the rough regime, when $H < 1/2$. Here we need to rely on a handful of semi-closed analytical formulas that enables us to see whether our implementation misses the mark. One of the few such formulas is the Forward Variance curve of the Rough Heston model [74] which can be written as

$$d\xi_t(T) = \kappa(T-s)\sqrt{V_t} dW_t, \tag{5.8}$$

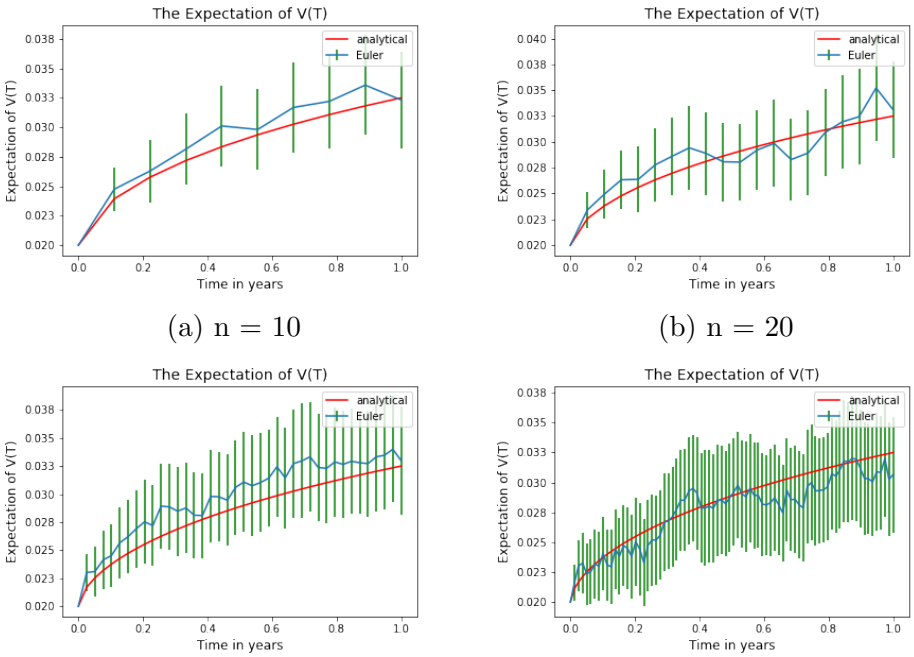
with the kernel

$$\kappa(x) = x^{\alpha-1} E_{\alpha,\alpha}\left(-\lambda x^{\alpha}\right) \tag{5.9}$$

and where $E_{\alpha,\beta}(x)$ denotes the generalized Mittag-Leffler function. Thus

$$\xi_0(T) = V_0 + \left(\theta - V_0\right) \lambda \int_0^T \kappa(s) ds. \tag{5.10}$$

We remind the reader the forward variance curve is exactly the expectation for the volatility process of equation 5.6, namely $E[v_T | F_t] = \xi_t^T$. Luckily, this integral can be handled using inbuilt numerical integration techniques provided by the *numpy* python package which also allows to set arbitrary constraints to the accuracy of the integration. In figure 5.7 we can see the behaviour of the Euler implementation at different numbers of discretization steps $n = 10, 20, 40, 80$. There is good agreement with analytical expectation for as few as $n = 10$ discretization points per year. Note, in order to obtain the standard



(a) n = 10 (b) n = 20

(c) n =40 (d) n=80

Figure 5.7: The Euler expectation value of the volatility process for various numbers of discretization points per year n. The error bars all represent 3 standard deviations for each point. Note, only $N = 10^4$ paths were used to infer the standard deviations.

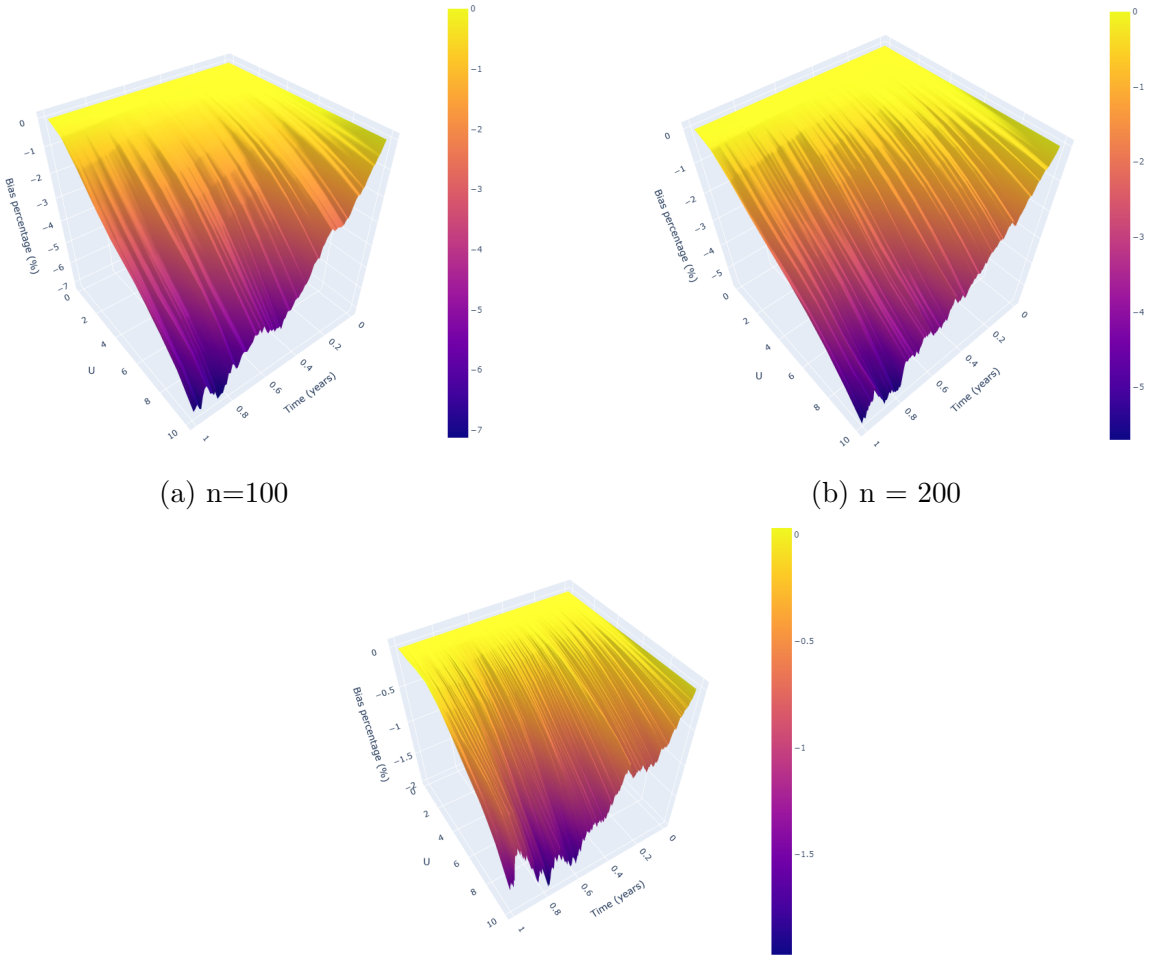
deviation displayed in this figure we used $N = 10^4$ simulations for each resolution. We will stick to using this number of simulations for the remainder of all the other checks.

Next, we mention that we cannot mirror the analysis performed for the classical Heston. Unfortunately there are no easy to implement analytical formulas to cross check higher moments of distribution of the volatility process, although there is formula for the characteristic function of volatility. We then choose to focus on the convergence of Euler S process to the rough Heston characteristic function.

In figure 5.8 we notice we can achieve an acceptable degree of accuracy starting from $n = 500$ steps per year discretization. Unfortunately, because of the time complexity of the Euler scheme, performing the standard $N = 10^6$ simulations is not possible in a matter of minutes on a regular consumer laptop for $n = 500$ discretization points. Looking at the Taylor expansion of the characteristic function

$$\begin{aligned} E[exp(iu \log S_t)] &= 1 + iuE[logS_t] + (iu)^2 E[logS_t]^2 + \ldots = \\ &= 1 + iuE[logS_t] - u^2[Var(\log S_t) + E[logS_t]] + \ldots \end{aligned} \tag{5.11}$$

we can see that, at first order, the bias comes from both the expectation and the variance of the $\log S$ process. Based on the fact that we consistently have a negative bias to the analytical characteristic function we can assume that either one or both are larger than the analytical values, although without further analysis we cannot pronounce ourselves which distribution moment is more biased. To do this, we decide to cross check our



(a) n=100

(b) n = 200

(c) n = 500

Figure 5.8: The Euler scheme percentage bias to the characteristic function of the Rough Heston model. The bias is inferred for various numbers of discretization steps per year n. This bias is inferred using $N = 10^4$ simulated paths. The reference value for the analytical characteristic function was obtained using the Adams scheme described in the earlier chapter with $n = 5000$ discretization steps.

	$E[(S_T - K)_+]$	$E[(A_T - K)_+]$	$E[(M_T - K)_+]$
	0.056832	0.032479	0.092904
n = 4	0.0578 (0.005)	0.0318 (0.0008)	0.074 (0.0008)
n = 10	0.0631 (0.0049)	0.033 (0.0009)	0.0815 (0.0008)
n = 20	0.0551 (0.005)	0.0334 (0.0009)	0.0862 (0.0007)
n = 40	0.0588 (0.005)	0.0325 (0.0009)	0.0879 (0.0007)
n = 80	0.0561 (0.0049)	0.0322 (0.0009)	0.0901 (0.0007)

Table 5.2: Estimation of European, Asian and Lookback options for maturity of $T = 1$ year and strike $K = 1$. The reference values are displayed in the top row.

implementation with option prices based on the first moment of the distribution, the expectation value of the asset price. For this we choose European, Asian ($A_T := \int_0^T S_t dt$) and Lookback ($M_T := \max_{0 \leq t \leq T} S_t$) options to address the bias of the expectation of $logS$ process.

Interestingly, there is a very good agreement starting from simulations with $n = 4$ discretization steps a year. This means that most of the bias in the Rough Heston Characteristic function comes from the variance of the $logS$ process. As a result we trust that the Euler discretization scheme performs well on derivatives based on the expectation of the asset price, but would however be deficient for derivatives based on higher order moments of distribution.

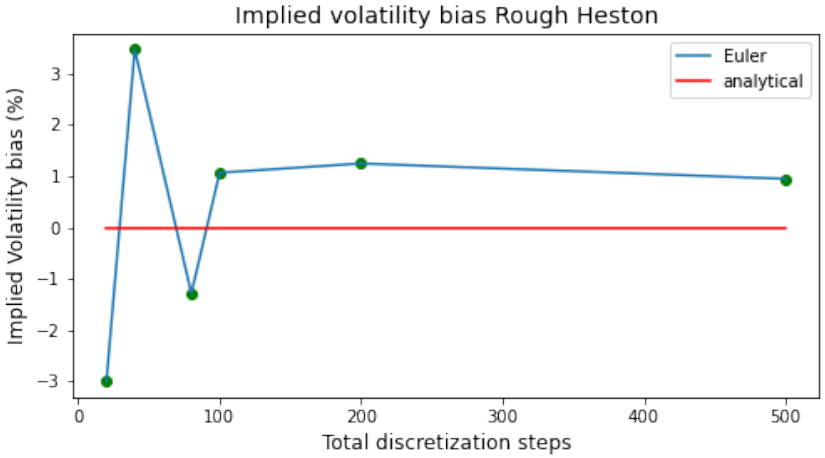


Figure 5.9: Implied volatility bias for $K = 1$ and $T = 1$. The bias was computed via $N = 10^4$ simulation of the Euler scheme for various numbers of discretization steps.

We notice the same behaviour also for the IV surface. Of course, this is becuase the IV surface relies on the European call price about which we already know we have good agreement on. This is encouraging because the Rough Heston model is most useful for short to midterm maturities ($T < 5$ years). For the options reliant on the money markets thus the Euler scheme is a quick and easy algorithm for accurate option pricing if $T < 1$ years. Also, in figure 5.10 we can see that the explosion of the at the money skew of the IV is well replicated by the Euler scheme with only $n = 100$ discretization points. Relying on our intuition from the classical Heston model and the convergence to the Rough

Characteristic function it is believed that for maturities larger than 1 year we need to use moment matching schemes such as the Quadratic Exponential method. Importantly, Gatheral [75] has come up with a QE scheme for the Rough Heston model inspired from the work of [15]. This would represent a natural extension to the current body of work, with interest in investigating the behaviour of the QE scheme for maturities between 1 to 5 years. Based on the results obtained for classical Heston by leveraging the QE scheme we believe it would enable for accurate pricing of a more diverse range of options such as VIX options and Variance Swaps.

Lastly, the reason why this current work is relevant to the issues discussed thus far is that there are only a handful of semi closed analytic formulas for price computation of derivatives in the Rough Heston. This means we need to rely on simulations if we were to price more exotic derivatives that have not been studied as rigorously as European options. Performing $N = 10^6$ simulations with $n = 10$ discretization steps per year using the Euler discretization scheme for Rough Heston can be done within a few minutes on a low end consumer laptop. As a result, as we can then use the Euler scheme to price Rough Heston options as long as they rely on moment of distribution no higher than the expectation of the asset price.

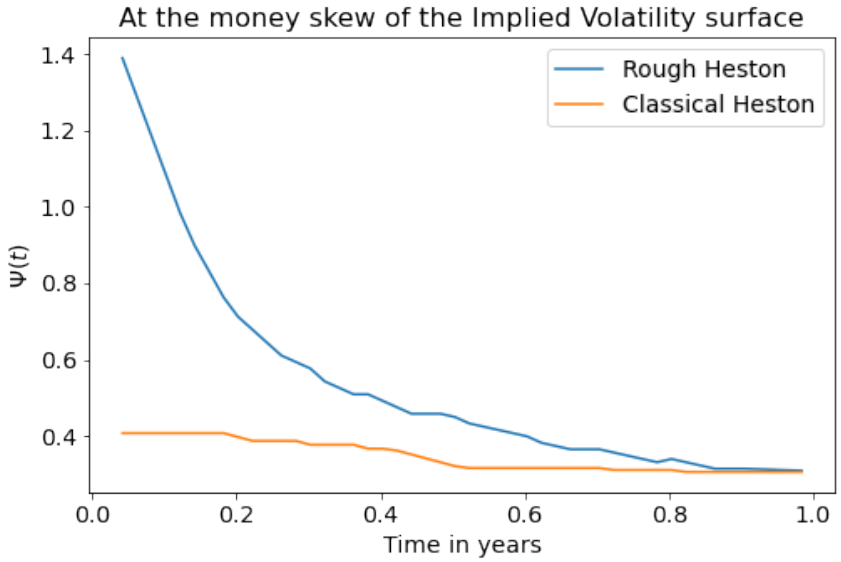


Figure 5.10: At the money IV skew computed via $\psi(t) := \left|\frac{\partial}{\partial k}\sigma_{BS}(k,\tau)\right|_{k=0}$ for $k = \log K$, where K is the strike price. The Rough Heston skew was computed using 10^4 simulated paths of the Euler scheme with $n = 100$ steps. Classical Heston skew was computed through analytical pricing. The step like behaviour of the IV skew is caused by a low tolerance in the root finding algorithm used to compute the IV. The parameters used to compute the IV skew were the Rough Heston ones.

Chapter 6

Conclusion

In this thesis we showed how the easy to implement Euler discretization scheme enables efficient pricing of European, Asian and Lookback options in the context of the Rough Heston model in the short term $T < 1$ year. In our analyses we started from benchmarking its performance in the classical Heston context and realized its slower convergence rate in the long term regime. However, its mild convergence in the short term regime of classical Heston motivated us to try its efficacy for pricing Rough Heston options. This was deemed as an appropriate step because we know that the Rough Heston model gives rise to more realistic implied volatility surface for short maturities than its classical counterpart. In addition to this, testing the Euler scheme against the Rough Heston characteristic function indicates the domain of applicability is constrained to only pricing derivatives based on the expectation of the price of the underlying. This is because convergence to higher order moments of distribution is rather slow and becomes prohibitive for larger maturities. Nevertheless, moment matching based scheme such as the Quadratic Exponential scheme are likely to allow for a more efficient convergence behaviour, especially for larger maturities. Our results indicate that for maturities less than 1 year, appropriate for money market options, the Euler scheme is a reliable tool for pricing derivatives. Lastly this shows that Affine Forward Variance models [74] such as the Rough Bergomi can use the same Euler scheme to compute instantaneous volatility, similarly in the short term.

Acknowledgements

I would like to thank my supervisor, Prof. Giacomo Bormetti, for making this project possible. The insights he continuously provided were invaluable in the making of this documents. I would also like to thank Fabio Baschetti for his very thoughtful comments during thesis meetings which allowed me to make progress on multiple accounts. Nevertheless, I would like to thank my family for their unwavering support and encouragement throughout my academic journey.

Bibliography

[1] C. W. Gardiner, *Handbook of stochastic methods*. Springer Series in Synergetics, Springer, 2nd ed., 1996.

[2] J. L. McCauley, *Stochastic Calculus and Differential Equations for Physics and Finance*. Cambridge University Press, 2013.

[3] E. Akyıldırım and H. M. Soner, "A brief history of mathematics in finance," *Borsa Istanbul Review*, vol. 14, no. 1, pp. 57–63, 2014.

[4] P. Brandimarte, *Numerical Methods in Finance and Economics: A MATLAB-Based Introduction*. Statistics in Practice, Wiley, 2013.

[5] G. B. G. C. P. Greco., *Computer Meets Theoretical Physics: The New Frontier of Molecular Simulation.* (The Frontiers Collection.) Springer, 2020.

[6] L. Bachelier, "Louis bachelier's theory of speculation: The origins of modern finance," pp. 1–188, 01 2011.

[7] M. Haugh, "Ieor e4706: Foundations of financial engineering course. the black-scholes model," *University of Columbia*, 2016.

[8] B. Stehlíková, "V. black-scholes model: Derivation and solution," *Financial derivatives, winter term 2014/2015*.

[9] S. S. Mikhailov and U. Nögel, "Heston ' s stochastic volatility model implementation , calibration and some extensions," 2003.

[10] J. Gatheral, "Implied volatility surface," *City University of New York. Baruch College*, 2010.

[11] B. B. Mandelbrot and J. W. V. Ness, "Fractional brownian motions, fractional noises and applications," *Siam Review*, vol. 10, pp. 422–437, 1968.

[12] J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility is rough," 2014.

[13] O. E. Euch and M. Rosenbaum, "The characteristic function of rough heston models," 2016.

[14] F. Baschetti, G. Bormetti, S. Romagnoli, and P. Rossi, "The sinc way: A fast and accurate approach to fourier pricing," 2020.

[15] L. Andersen, "Efficient simulation of the heston stochastic volatility model," *J. Computat. Finance*, vol. 11, 01 2007.

[16] A. Einstein, *Investigations on the Theory of the Brownian Movement*. Dover Publications, 1926.

[17] E. Mackinnon, "Einstein's 1905 brownian motion paper," *CSI Communications*, vol. 29, pp. 6–8, 01 2005.

[18] M. Davis and A. Etheridge, "Louis bachelier's theory of speculation," 12 2011.

[19] J. Crank and E. Crank, *The Mathematics of Diffusion*. Oxford science publications, Clarendon Press, 1979.

[20] S. Regev, N. Grønbech-Jensen, and O. Farago, "Isothermal langevin dynamics in systems with power-law spatially dependent friction," *Physical Review E*, vol. 94, jul 2016.

[21] H. M. T. Samuel Karlin, *A first course in stochastic processes*. Academic Press, 2d ed ed., 1975.

[22] J. McCauley, *Dynamics of Markets: Econophysics and Finance*. Dynamics of Markets: Econophysics and Finance from a Physicist's Standpoint, Cambridge University Press, 2004.

[23] A. Gloria, "When are increment-stationary random point sets stationary?," *Electronic Communications in Probability*, vol. 19, jan 2014.

[24] A. P. (auth.), *PDE and Martingale Methods in Option Pricing*. Bocconi and Springer Series, Springer, 1 ed., 2011.

[25] S. Ross, "Introduction to probability models (eleventh edition)," p. iv, Boston: Academic Press, 2014.

[26] L.Arnold, *Stochastic differential equations: theory and applications.* Wiley, 1974.

[27] R. McDonald, *Derivatives Markets*. Addison-Wesley series in finance, Pearson Higher Education & Professional Group, 2013.

[28] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, vol. 82. 01 2000.

[29] S. S. Vempala, "Geometric random walks: a survey," 2007.

[30] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," *Econometrica*, vol. 53, no. 2, pp. 385–407, 1985.

[31] T. Sadhu and K. J. Wiese, "Functionals of fractional brownian motion and the three arcsine laws," *Physical Review E*, vol. 104, nov 2021.

[32] S. Cohen and J. Istas, *Fractional Fields and Applications*, vol. 73. 01 2013.

[33] G. Shevchenko, "Fractional brownian motion in a nutshell," *International Journal of Modern Physics: Conference Series*, vol. 36, 06 2014.

[34] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications*. Probability and Its Applications, Springer London, 2008.

[35] R. Metzler and J. Klafter, "The random walk's guide to anomalous diffusion: a fractional dynamics approach.," *Physics Reports*, vol. 339, 01 2000.

[36] L. Decreusefond and A. Üstünel, "Fractional brownian motion: theory and applications," *http://dx.doi.org/10.1051/proc:1998014*, vol. 5, 01 1999.

[37] J. Ortega, A. Acevedo, D. Prada, and Y. Herrera, "Physical applications: fractional brownian movement applied to the particle dispersion," *Journal of Physics Conference Series*, vol. 1702, p. 1, 12 2020.

[38] M.-A. Miville-Deschenes, F. Levrier, and E. Falgarone, "On the use of fractional brownian motion simulations to determine the three-dimensional statistical properties of interstellar gas," *The Astrophysical Journal*, vol. 593, pp. 831–847, aug 2003.

[39] F. Evers and A. D. Mirlin, "Anderson transitions," *Reviews of Modern Physics*, vol. 80, pp. 1355–1417, oct 2008.

[40] C. Barton and P. Pointe, *Fractals in Petroleum Geology and Earth Processes*. Springer US, 2012.

[41] T. Dieker, "Simulation of fractional brownian motion," 01 2013.

[42] R. Sibirtsev, "How to avoid bankruptcy?: Monte carlo simulation of three financial markets, using the multifractal model of asset returns," p. 104, 2019.

[43] C. Bender, T. Sottinen, and E. Valkeila, "Arbitrage with fractional brownian motion?," *Theory of Stochastic Processes*, vol. 13, 01 2007.

[44] B. Mandelbrot and R. Hudson, *The Misbehavior of Markets: A Fractal View of Financial Turbulence*. Basic Books, 2007.

[45] L. Calvet and A. Fisher, "A multifractal model of assets returns," 11 1999.

[46] J. Peyriere, *Multifractal measures*. Dordrecht: Springer Netherlands, 1992.

[47] L. Calvet, A. Fisher, and B. Mandelbrot, "Large deviations and the distribution of price changes," 10 1997.

[48] A. Fisher, L. Calvet, and B. Mandelbrot, "Multifractality of deutschemark/us dollar exchange rates," 10 1997.

[49] P. Oswiecimka, J. Kwapien, S. Drozdz, A. Z. Gorski, and R. Rak, "Multifractal model of asset returns versus real stock market dynamics," 2006.

[50] S. Drożdż, M. Forczek, J. Kwapień, P. Oświcimka, and R. Rak, "Stock market return distributions: From past to present," *Physica A-statistical Mechanics and Its Applications*, vol. 383, pp. 59–64, 2007.

[51] Z. Eisler and J. Kertész, "Multifractal model of asset returns with leverage effect," *Physica A-statistical Mechanics and Its Applications*, vol. 343, pp. 603–622, 2004.

[52] S. Bianchi, A. Pantanella, and A. Pianese, "Modeling stock prices by multifractional brownian motion: An improved estimation of the pointwise regularity," *Quantitative Finance*, vol. 13, 07 2011.

[53] T. Lux, "The multi-fractal model of asset returns: Its estimation via gmm and its use for volatility forecasting," 2003. urn:nbn:de:101:1-200911022352.

[54] F. Mishkin, *The Economics of Money, Banking, and Financial Markets*. Addison-Wesley series in economics, Pearson/Addison Wesley, 2007.

[55] J.-M. Courtault, Y. Kabanov, B. Bru, P. Crépel, I. Lebon, and A. Marchand, "Louis bachelier on the centenary of "théorie de la spéculation"," *Mathematical Finance*, vol. 10, pp. 339–353, 07 2000.

[56] M. Taqqu, "Bachelier and his times: A conversation with bernard bru," *Finance and Stochastics*, vol. 5, pp. 3–32, 01 2001.

[57] A. F. Perold, "The capital asset pricing model," *Journal of Economic Perspectives*, vol. 18, pp. 3–24, September 2004.

[58] D. W. Edwards, "Risk management in trading (techniques to drive profitability of hedge funds and trading desks) ——," vol. 10.1002/9781118819692, jun 2014.

[59] S. Lalley, "Mathematical finance 345 course, lecture 10: Change of measure and the girsanov theorem," *University of Chicago*, 2001.

[60] Y. Ding, "Conditional Heteroskedasticity in the Volatility of Asset Returns," Nov. 2021.

[61] J. Gatheral, "The volatility surface (a practitioner's guide)," vol. 10.1002/9781119202073, jan 2012.

[62] L. Bergomi, *Stochastic Volatility Modeling*. Chapman and Hall/CRC Financial Mathematics Series, CRC Press, 2015.

[63] A. Lewis, *Option Valuation Under Stochastic Volatility*. 01 2000.

[64] H. Chinwenyi, "The-non-uniqueness-of-partial-differential-equations-options-price-valuation-formula-for-heston-stochastic-volatility-model," 12 2020.

[65] S. Desmettre, "Four generations of asset pricing models and volatility dynamics," 2007.

[66] F. Comte and E. Renault, "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, vol. 8, no. 4, pp. 291–323, 1998.

[67] Q. Zhu, G. Loeper, W. Chen, and N. Langrené, "Markovian approximation of the rough bergomi model for monte carlo option pricing," *Mathematics*, vol. 9, p. 528, mar 2021.

[68] R. McCrickerd and M. S. Pakkanen, "Turbocharging monte carlo pricing for the rough bergomi model," *Quantitative Finance*, vol. 18, pp. 1877–1886, apr 2018.

[69] C. Bayer, P. K. Friz, and J. Gatheral, "Pricing under rough volatility," *Quantitative Finance*, vol. 16, pp. 887 – 904, 2015.

[70] A. Richard, X. Tan, and F. Yang, "On the discrete-time simulation of the rough heston model," 2021.

[71] S. W. Jeng and A. Kiliçman, "Spx calibration of option approximations under rough heston model," *Mathematics*, vol. 9, no. 21, 2021.

[72] S. W. Jeng and A. Kilicman, "Fractional riccati equation and its applications to rough heston model using numerical methods," *Symmetry*, vol. 12, no. 6, 2020.

[73] C. Li and C. Tao, "On the fractional adams method," *Computers Mathematics with Applications*, vol. 58, no. 8, pp. 1573–1588, 2009.

[74] J. Gatheral and M. Keller-Ressel, "Affine forward variance models," 2018.

[75] J. Gatheral, "Efficient simulation of affine forward variance models," 2021.

List of Figures

List of Tables