Exotic Equity Derivatives: A Comparison of Pricing Models and Methods with both Stochastic Volatility and Interest Rates

By Jaundré Scheltema

Submitted in fulfilment of the requirements in respect of the Master's Degree

M.Sc. Actuarial Science

in the Department of Mathematical Statistics and Actuarial Science in the Faculty of Natural and Agricultural Science at the University of the Free State.

BLOEMFONTEIN

May 2017

Supervisor: Mr. Jan-Paul Venter

Declaration

I, Jaundre Scheltema, declare that the Master's Degree research dissertation that I herewith submit for the Master's Degree qualification M.Sc. Actuarial Science at the University of the Free State is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education.

Jaundre Scheltema

Acknowledgement

I would firstly like to thank my supervisor, Jan-Paul Venter, for his guidance and support. The various stylistic comments and guidance on the material that he has provided has given critical support to completing this thesis.

I would also like to thank dr. Lech A. Grzelak, who invented the H2CIR model used in this thesis, for his suggestions on how to implement a vectorised version of the H2CIR model in MATLAB®.

Lastly, I would like to thank my family for a tremendous amount of support both in completing this thesis and otherwise this past year. I want to thank my brother for allowing me to use his two laptops to produce this thesis. Without this support I would not have been able to complete this thesis in a timely manner. I would also like to thank my mother, Dr. S Van Zyl, for her caring support and scholarly guide as a researcher in producing this thesis. Her suggestions and "shortcuts" proved to be invaluable

Abstract

The traditional Black Scholes methodology for exotic equity option pricing fails to capture the features of latent stochastic volatility and observed stochastic interest rate factors exhibited in financial markets today. The detailed study presented here shows how these shortcomings of the Black Scholes methodology have been addressed in literature by examining some of the developments of stochastic volatility models with constant and stochastic interest rates.

A subset of these models, notably with models developed within the last two years, are then compared in a simulated study design against a complex Market Model. Each of the select models were chosen as "best" representatives of their respective model class. The Market Model, which is specified through a system of Stochastic Differential Equations, is taken as a proxy for real world market dynamics. All of the select models are calibrated against the Market Model using a technique known as Differential Evolution, which is a globally convergent stochastic optimiser, and then used to price exotic equity options.

The end results show that the Heston-Hull-CIR Model **(**H2CIR) outperforms the alternative Double Heston and 4/2 Models respectively in producing exotic equity option prices closest to the Market Model. Various other commentaries are also given to assess each of the select models with respect to parameter stability, computational run times and robustness in implementation, with the final conclusions supporting the H2CIR Model in preference over the other models.

Additionally a second research question is also investigated that relates to Monte Carlo pricing methods. Here the Monte Carlo pricing schemes used under the Black Scholes and other pricing methodologies is extended to present a semi-exact simulation scheme built on the results from literature. This new scheme is termed the Brownian Motion Reconstruction scheme and is shown to outperform the Euler scheme when pricing exotic equity derivatives with relatively few monitoring or option exercise dates.

Finally, a minor result in this study involves a new alternative numerical method to recover transition density functions from their respective characteristic functions and is shown to be competitive against the popular Fast Fourier Transform method.

It is hoped that the results in this thesis will assist investment and banking practitioners to obtain better clarity when assessing and vetting different models for use in the industry, and extend the current range of techniques that are used to price options.

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Appendix C:

Appendix D:

Appendix E:

List of Tables

Chapter 2:

Chapter 3:

Chapter 4:

Appendix B:

Appendix D:

Appendix E:

Appendix F:

Chapter 0 – Notes To The Reader

The purpose of this chapter is to clarify a few aspects of this thesis before the reader engages with the material. These aspects include the Mathematical conventions used, the style of the literature review, a guide to reading the material and lastly a personal note from myself.

Firstly, the material presented in this thesis is highly mathematical and draws upon many sources. As a consequence I have tried to reconcile the notation between the various authors and to derive a style that is broadly consistent across each of the themes and extracts reproduced from various journal articles. Unfortunately, ambiguities do arise and the context of the material should make it clear what precisely is meant. Additionally, in some cases, the notation reflects that of the articles from which the segments are reproduced. This is done so that the reader who wishes to follow up on the given references can do so with relative ease and avoid confusion by having otherwise two different sets of notations for the same underlying ideas and equations.

Some practitioners and academics have a preference for certain notations used which will likely differ from the conventions assumed here. Some examples of this include zero-coupon prices which are denoted here by $Z(t,T)$ while others have understood this in the notation $P(t,T)$. The notation used for risk-neutral dynamics is denoted here by $\mathbb{Q}$or a *. The notation for a characteristic function is given as ϕ instead of ψ, although the former can also be interpreted as the density function of a normal distribution, etc.

In addition, most standard mathematical conventions regarding the notation of vectors in bold face, e.g. $\mathbf{y}$, and matrices as capital letters, e.g. $\mathcal{M}$, are given. At times there will be symbols that have no universal notation and the reader is expected to assimilate this quickly while reading through the material. Again, ambiguities do arise, particularly when one has over 300 equations, although the context should make it clear what is meant with a particular symbol.

A few words about the literature review. I have stayed clear of a "cryptic" literature review style, meaning that a topic is discussed with four or five obscure references added into a sentence and a whole paragraph being difficult to understand when the work of other authors is mentioned. This style is typically found in articles where a paragraph indicating what has been done in the past is filled with many references.

I have aimed, in contrast to the above style, to present ideas (old and brand new) here from articles, and textbooks, with the material reworked and additional (original) explanations given. The end result is that the reader can almost read each idea as if it were presented in a textbook style. Consequently this thesis is longer than what would be expected on average, although my view is to present the audience with a clear grasp of what is going on so that even non-experts in the field can read the material. Throughout some comments about one or two other sources are given that relates to the same referenced material, without burdening the reader with having to go on a reference hunt to understand a sentence.

In spite of keeping a clear style, the reader is cautioned that the material relates to a wide range of mathematical ideas. Hence in order to cope with the mathematical demands and keep this thesis self-

contained, as far as possible, a significant amount of material is included in Appendix B and referenced as ideas are introduced. It is thus beneficial to the reader to skip to Appendix B whenever given the cue to pursue the necessary background material, if needed.

The entire process of pricing derivatives is a complicated matter and simply giving a straightforward literature review is possible, although I suspect to the lay person this would make the material unreadable. For instance, in pricing derivatives, one has to:

- develop a model for the asset price process, including stochastic volatility and sometimes stochastic interest rates
- develop some pricing method, be it through Partial Differential Equations (PDEs) or characteristic functions or numerically (Monte Carlo Simulations)
- calibrate the models
- implement the models correctly and,
- finally derive a price for the derivative (and repeating the process to calculate the associated Greeks!).

This entire process is easier to understand when broken down thematically and each issue is, although not always possible, handled separately. Consequently the literature review, largely in Chapter 3, will be interspersed/mixed with original work from my side, rather than separating it into its own chapter. The original work relates to the numerical techniques involved in calibrating and simulating some of the models that are presented in Chapter 2. In any case, the majority of the original work is done separately when answering the research questions in Chapter 4.

Appendix A gives details about the code and all accompanying figures used to produce this thesis. The code is stored on an online cloud facility as printing all the code would constitute well over a 100 pages of additional material, almost all of which is original work. The MATLAB® and C++ constitute a major part of this thesis and thus many references to the supporting material is given, although nothing is lost when reading through the thesis without having the supporting material.

On a personal note, I have really enjoyed producing this thesis, especially the programming aspect. It took many hours of hard work, a couple of computer crashes and many rewrites to produce the final document as it stands. I have had many ideas and in one particular case I have spent 4 weeks pursuing an idea that did not lead anywhere, wasting essentially an entire month. Frustration is always part of the process but luckily this phase of hard work has led to some of the ideas underpinning this thesis.

On another note, this thesis was originally around 270 pages with many additional stochastic volatility models that have been removed from the content. There are many other ideas that have also been cut out, including some ideas related to Malliavin Calculus. Nonetheless, this process has been repeated and the material streamlined to produce this final document.

I am proud of the work that is presented here and I hope that reading this thesis will be worthwhile. I have strived hard to eliminate all possible errors, whether mathematical, typographical and in the supporting computer code. But, if any remain then they are ultimately my responsibility, and I give my apology in advance.

Finally, I hope that when you read this thesis you will discover some new ideas and hopefully take these as inspiration for further research in the future or to implement in practice.

- Jaundre Scheltema
May 2017

Chapter 1- Introduction

1.1 Introduction to derivative pricing

Contemporary derivative pricing is a very diverse and rich field that has gone through several developments in the last two decades. A large number of financial derivatives have been invented, some of which pose considerable challenges in deriving an arbitrage-free price for derivative contracts. Similarly, advancements in models that describe the evolution of some asset price have led to challenges in pricing even basic derivative contracts, such as Call and Put options.

We will consider some exotic equity derivative contracts, and with the emphasis on the underlying models used. In particular, models based on stochastic volatility and stochastic interest rate processes are presented, and thus Chapter 2 is dedicated to expanding on this idea by drawing upon the body of literature available.

In this introductory chapter the Black-Scholes method is reviewed, and evidence given from actual market data that the crucial assumption of the Black-Scholes method of constant volatility is incorrect. Following this a brief introduction is given as to what stochastic volatility is as well as a discussion of some of its uses. In the final subsection a return is made to the Black-Scholes pricing methodology, with some examples given, mirroring the layout of this thesis.

1.1.1. Review of the Black-Scholes method

The Black-Scholes Model (Black and Scholes, 1973), denoted by BS from now on, depends up the following key assumptions:

- The asset price process follows a Geometric Brownian motion path, implying that the returns are lognormally distributed.
- The interest rate and volatility parameters are constant for all durations.
- The asset price process follows a Stochastic Differential Equation (SDE) of the following form:

$$dS(t) = S(t)\mu dt + \sigma S(t)dW(t). \tag{1.1}$$

Note that by assuming a constant volatility parameter, σ, one automatically has, by construction, a constant implied volatility surface in the BS Model which is now further considered below.

1.1.2. Implied volatility surface

The implied volatility parameter σ is the constant parameter value that solves the BS equation for a given observed (market) Call option price[1], denoted as $C(S, K, T, t, \sigma)$:

$$C(S, K, T, t, \sigma) = S \times \mathcal{N}\left(d_+\left(\frac{S}{k}, T-t\right)\right) - K \times e^{-r(T-t)} \times \mathcal{N}\left(d_-\left(\frac{S}{k}, T-t\right)\right), \tag{1.2}$$

[1] Note that in the demonstration of the notion of implied volatility we could have also used the market prices of Put options or other derivatives with the correct corresponding BS equation for the price of the derivative.

$$where\ d_-\left(\frac{S}{k}, T-t\right) = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\times\sigma^2\right)\times(T-t)}{\sigma\times\sqrt{T-t}};$$

$$d_+\left(\frac{S}{k}, T-t\right) = d_-\left(\frac{S}{k}, T-t\right) + \sigma\times\sqrt{T-t}$$

The observed values, including the underlying $S(t) = S$, should have constant values for any time to maturity and moneyness of the option. (The "moneyness" of an option will be defined here by the convention of the strike price divided by the current asset price. This convention is used in order to compare option prices by varying strike rates for a fixed current asset price in order to be consistent with the market practice of quoting option prices.) This constant parameter then leads to the implied volatility surface which is illustrated as:

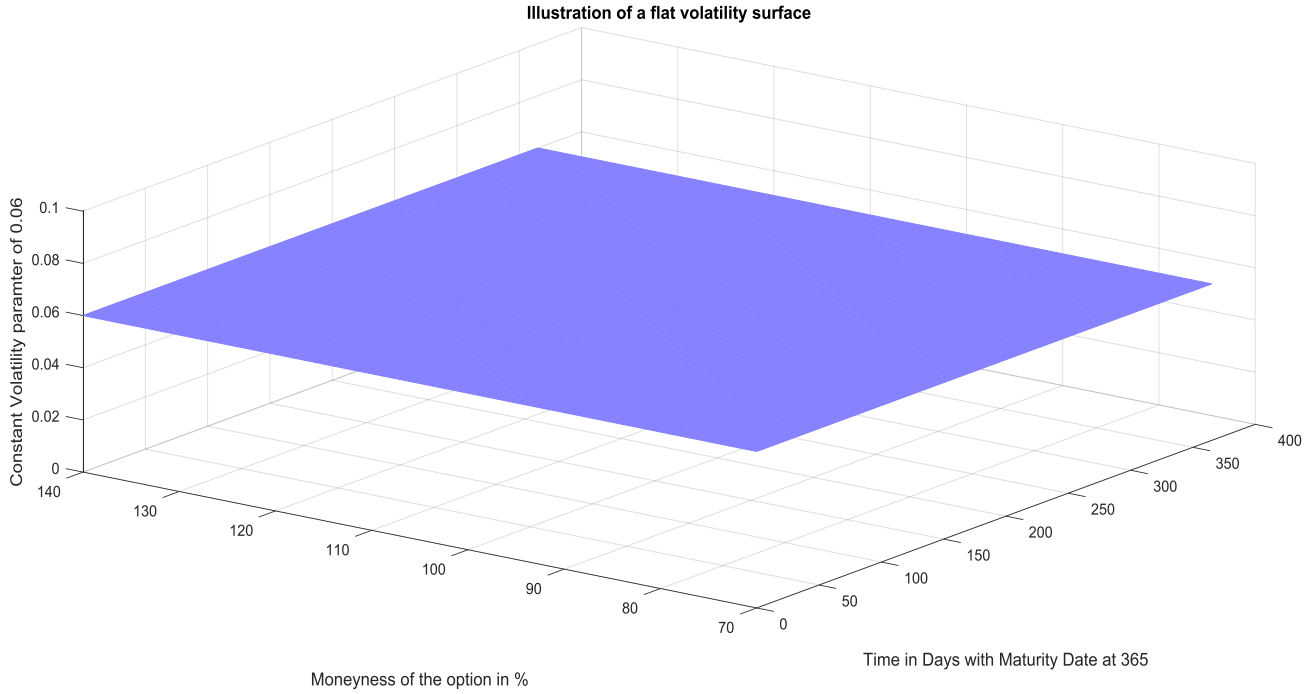

Figure 1.1: Flat Implied Volatility Surface, where the constant volatility parameter is 0.06

Naturally this implies that when viewed in two dimensions the graph should have a straight line ($y = 0.06$) for all values on the moneyness axis (or otherwise time to maturity axis).

1.1.3. **Implied volatility skew or smile**

Prior to 1987, the implied volatility surface for index options was observed to be relatively flat, similar to that shown in Figure 1.1. After the general stock market crash on 19 October 1987, popularly known as 'Black Monday', the implied volatility surface on index options was observed to follow a volatility "smile" (Gatheral, 2006).

In the markets the observed shape of the implied volatility as a function of the strike price K, denoted as $\sigma(K)$, vs the strike price (K) for a fixed time to maturity and an underlying asset process $S(t) = S$, is generally observed to follow either a "U-shape" or a downwards slope. We classify the former as the volatility smile: High (implied) volatility for high and low strikes. The latter is classified as the volatility skew: High volatility for low strikes and low volatility for high strikes. The volatility skew and smile stand to contradict the notion that the implied volatility surface is constant (by assumption).

The change in the implied volatility surface can be explained in terms of "market fear". The stock market crash of October 1987 indicated that the market could fall by 20% in any given day. Consequently, given that there is a non-zero probability that such a movement could happen in the future, we see that low-strike Put options are worth more than high-strike Call options. In this situation there is a "fear" that the market could drop significantly and hence leads to low-strike Put options being more expensive. Thus the implied volatility for low-strike Put options is higher than at-the-money or high-strike Put options and can lead to either an implied volatility smile (more common for single stock options) or a downwards skew (more common for market indices such as the S&P500).

The minimum of the graph of the volatility smile can be obtained as $Se^{r(T-t)}$. This result holds under the assumption that the market is described by an uncorrelated stochastic volatility model, which is defined in detail in Chapter 2, section 2.2.5.1. The argument and assumptions are given in detail in Appendix B. Note that when working on the moneyness axis we would divide by S to obtain the minimum strike corresponding to the minimum of the implied volatility surface, $e^{r(T-t)}$.

To illustrate the concepts of the volatility smile and skew consider delayed data from YAHOO! FINANCE. The underlying represents the S&P500 Index and series of Call and Put options written on the index with expiry date on the 26 February 2016 are considered. The data can be found in the supporting material, see Appendix A.

The option specifics (Call and Put) are:

- Expiry: 26 February 2016
- Time at calculation: 23 February 2016, 10:24 AM EST
- Current (Delayed) Price of the Underlying: 1931
- Current Risk-free rate on a 1yr Treasury Bill on 22 February 2016 is 0.55%

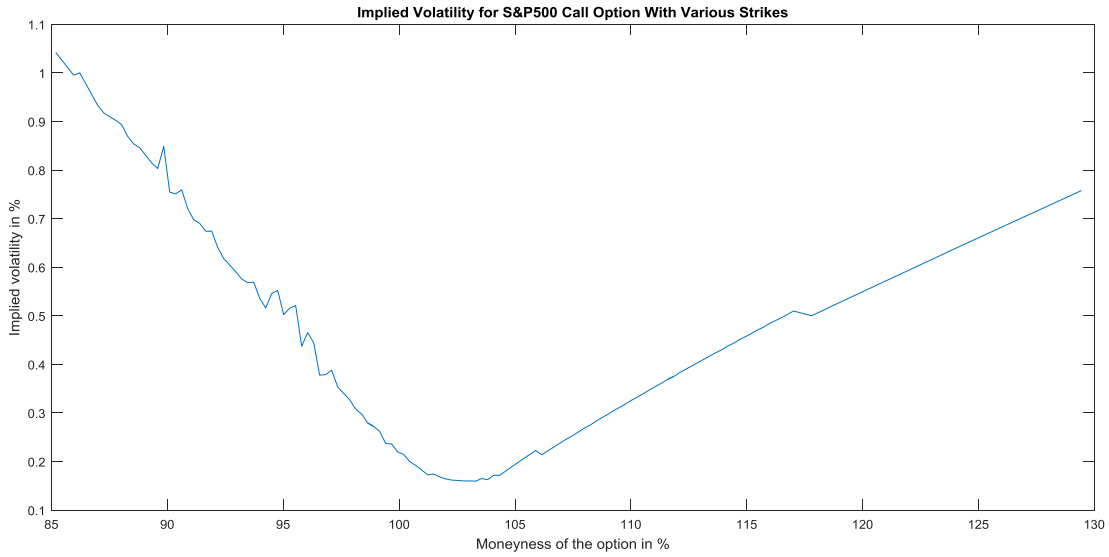

Figure 1.2: Illustration of Volatility Smile for S&P500 Call Options

Commentary: It is clear that the constant volatility assumption does not always hold in practice. And based on the extensive work done in Fouque et al (2011) we assume in general that it does not. The "U-shape", as mentioned above, can clearly be seen from the graph. The minimum of the graph does not correspond well with the simple formula given above, i.e. $e^{r(T-t)} = 100.0045\%$. The explanation for this given in Appendix B, following the proof of the Renault-Touzi theorem.

It is further mentioned by Fouque et al (2011) that the minimum of the implied volatility graph is usually shifted to the right[2] of $e^{r(T-t)}$. An interpretation of this phenomenon is that the market follows a correlated stochastic volatility model with the correlation coefficient being negative, i.e. $\rho < 0$ (see Chapter 2, section 2.2.5.2 for a detailed definition). With $\rho < 0$ it is argued that there is a greater change for the asset's price to drop following a change in the asset's volatility. Subsequently low-strike Put options will have higher implied volatilities and high-strike Put options will have lower implied volatilities. This introduces a skew effect that effectively shifts the point where the minimum of the implied volatility curve shifts upwards along the moneyness axis.

The previous explanation also holds for Call options. In an environment where there is a risk of a significant market drop (via the negative correlation with the asset's volatility) it is better for investors to move from owning the underlying asset and instead purchase Call options. In this manner the downside potential is capped and any upwards market movements can be leveraged through Call options to give a better ROI (Return-On-Investment) than by otherwise owning the underlying asset. This demand for Call options increases the demand for low-strike and at-the-money Call options and hence increases the price for those options. Consequently any price increases will increase the implied volatility observed and introduce a skew effect that shifts the minimum to the right.

Continuing our discussion, below follows an illustration for the volatility skew, for a Put option, using the same data as above:

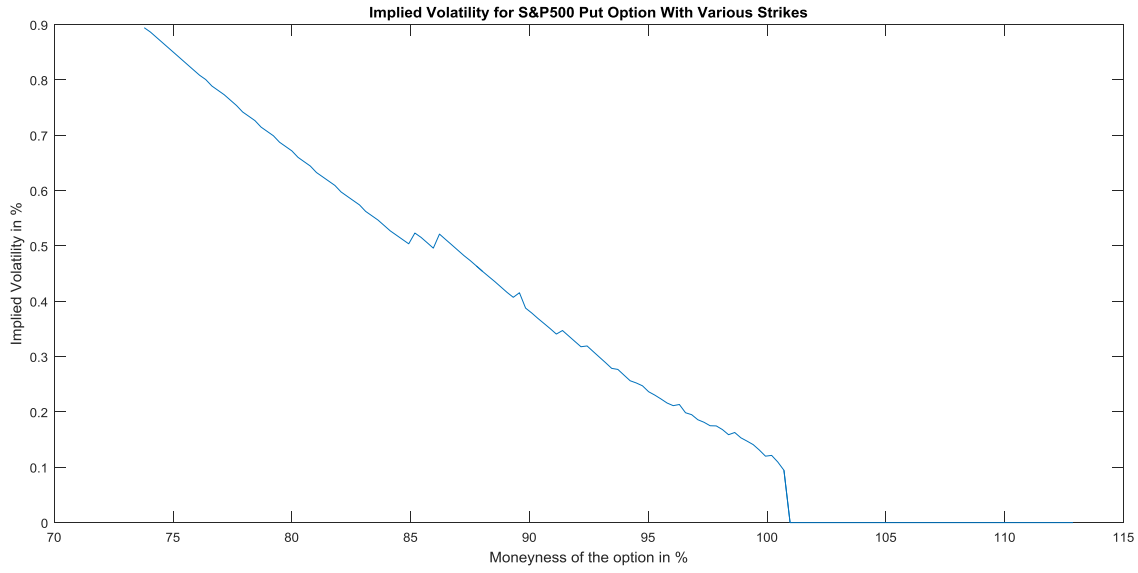

Figure 1.3: Illustration of Volatility Skew for S&P500 Put Options.

[2] In commodity markets the opposite typically holds, i.e. shift to the left. Following the explanation above, the correlation coefficient between the underlying and its stochastic volatility factor is positive, $\rho > 0$. This leads to a forward skew where high-strike options have higher implied volatilities than low-strike options.

Commentary: The graph is clearly downwards sloping. At some point the Put options return an implied volatility of 0%[3] for higher strike values. For these Put options, where $\sigma = 0\%$, an elementary no-arbitrage argument can be used to show that the value of the Put option at any time $t < T$ must then be the $\max(Ke^{-r(T-t)} - S, 0)$. For example, for a strike of 2135 (yielding a moneyness >100%, large enough so that volatility is 0%) the option value would be $2135 \times e^{-0.05\% \times \frac{3}{365}} - 1931.28 = 203.62$. Interestingly the observed bid/offer spread is $192.4/204$, and thus the value of the option (which is independent of the BS framework) is consistent with the observed market price (the no arbitrage price lies between bid/offer spread).

Additionally, we see that the ratio of the Call volatility to the Put volatility is greater than 100%. Thus the market would price Calls higher than Puts and hence the market shows that it has an upside (bullish) bias towards the underlying. This is consistent with observing the skew seen with the Put options. Put options function as a type of insurance against falling prices. Hence when the market has a bullish bias there is effectively less demand for Puts. Having the option price reflect this (i.e. the lower price) would automatically adjust the implied volatility downwards.

The last paragraph may seem to be in conflict with a basic consequence of the Put-Call parity relationship and the BS Model. Generally a Call and a Put option with the same time to maturity, same underlying, same borrowing costs and same strike should have the same implied volatilities (by the Put-Call parity relationship) or else an arbitrage opportunity exits. However, note that the market data exhibits the following features which contradicts most of the key BS assumptions that we mentioned in the beginning of the chapter, as well as some others:

- There are transaction costs.
- Markets are not perfectly liquid.
- The S&P 500 does not follow a lognormal model nor does it have constant volatility.

Additionally, the underlying, the S&P 500, cannot be bought or sold like a stock (it can be tracked however via an exchange traded fund - ETF). This provides problems for a replicating strategy used in the formulation of BS options pricing theory.
Thus with most of the BS assumptions not fully met, it may be reasonable to accept that the implied volatility is different for the Put and Call options of the same duration and same strike price.

In closing this subsection, an illustration of the volatility surface for an S&P500 Call option follows below. Essentially by taking a three-dimensional view of the volatility surface we can also see how the previously mentioned problems regarding the BS assumptions are carried over across time (or the time-to-maturity dimension). The details behind this calculation are:

- Expiry Times: 29 February, 31 March, 29 April, 31 May, 30 June, 29 July 2016.
- Time at calculation: 24 February 2016, 4:32 PM EST.
- Current (Delayed) Price of the Underlying: 1921
- Current Risk-free rate on a 1yr Treasury Bill on 23 February 2016 is 0.55%.
- Strike Prices: Starting at 1700 ending 2225, with increments of 25. Note that around 11 values

[3] An implied volatility of 0% shows that there is no time value embedded in the option price. This implies that there is no demand for these options prices, e.g. the options are not actively traded and their prices are equal to their no-arbitrage lower bounds. We can use this to infer what the market expectation is for the underlying asset movements, e.g. bullish.

had to be interpolated since there was no data available, this happened only at later durations.

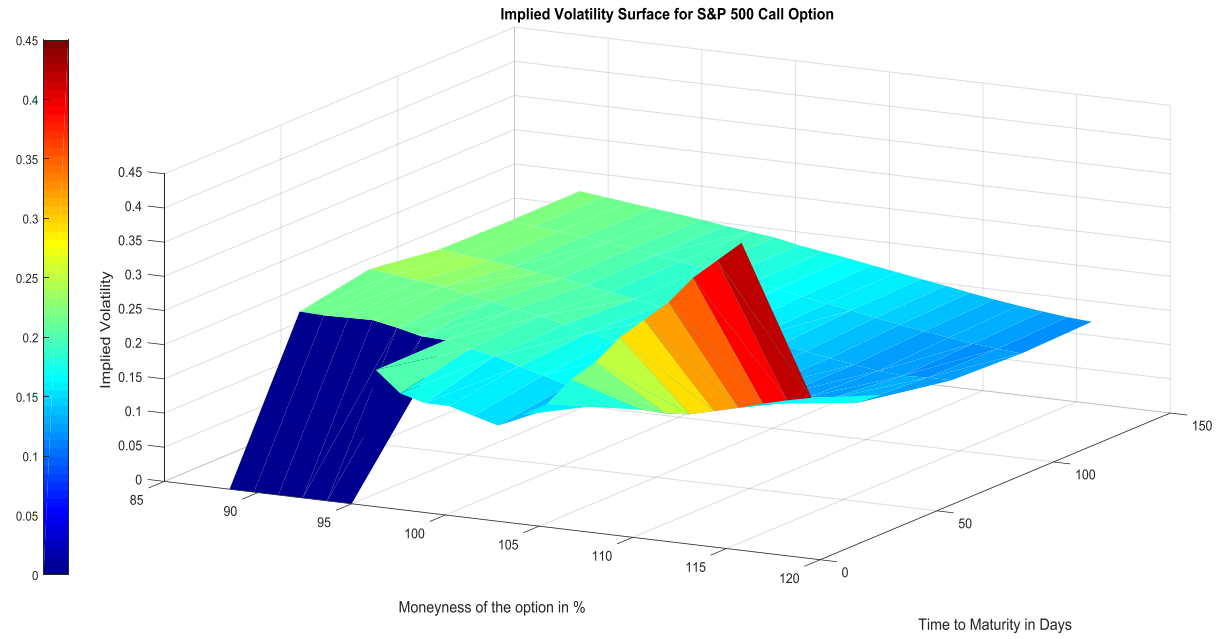

Figure 1.4: Implied volatility surface of S&P500 Call Option

Commentary: Clearly Figure 1.4 shows that the volatility surface is not constant. At later durations ($T > 50$) one can clearly see a steady downwards slope as the moneyness of the option increases. Thus at later durations we have an almost perfect example of the volatility skew. The convergence of the implied volatility surface to volatility skew can be explained as part of the mean-reversion of the asset's volatility, which is justified in Chapter 2, section 2.2.5.2. Essentially the realised long run volatility will be mean-reverting to some level θ and hence when applying a BS Model we infer that $\sigma \approx \theta$ after a sufficient period of time. The slight skew accounts for the differences in the demand for long-dated options. In the longer term we expect asset prices to increase and hence low-strike Call options will be more attractive (increased returns through leverage) than high-strike Call options. The differences in demand will adjust the option prices and hence the implied volatility surface.

1.1.4. **Additional problems with the Black-Scholes pricing methodology**

(Hagan et al, 2002) discuss the following problems with the BS Model. The constant volatility parameter will differ when attempting to match market prices, since the implied volatility surface is not constant. Thus every time a different value for σ is chosen a parametrisation of the BS Model needs to be applied. The implications thereof are summarised briefly:

1.) Pricing of exotic options where the Call option depends on more than one strike price becomes ambiguous. Which of the implied volatilities should be used to price the option when the entire volatility surface is not constant?

2.) Delta and Vega risks for hedging are not necessarily consistent across strikes as different implied volatilities for various strikes imply different parametrisations of the BS Model being used. Additionally the aggregate of such risks for a portfolio of various options might leave residual risks where otherwise the portfolio should be theoretically delta-neutral[4].

[4] Following the example of (Hagan et al, 2002), consider a portfolio that has x high strike options with total delta positive at $0.5x$ and that has x low strike options with total delta at $-0.5x$. The total delta of the portfolio should be

Note that in later chapters we will not be interested in modelling the implied volatility surface. This will follow naturally from the stochastic volatility models (as described in Chapter 2). We are mainly interested in modelling the volatility of the asset process, deriving appropriate pricing formulas and checking whether they are consistent with the observed prices. Thus the volatility skew and smile effects will serve as a check on the models. However we will often refer to the volatility skew (or smile) in assessing and calibrating different models.

1.2. **Addressing the shortcomings of the Black-Scholes method using stochastic volatility**

A "solution" via stochastic volatility to the shortcoming of the BS Model is to define a class of models where the volatility is modelled stochastically:

$$dS(t) = S(t)\mu(t)dt + f\big(V(t)\big)S(t)dW_1(t). \tag{1.3}$$

Here $V(t)$, the volatility process, is stochastic itself and may be driven by its own SDE process with a Brownian motion which can be correlated to $dW_1(t)$. The function f is assumed to be a smooth function. A list of stochastic models are covered in detail in chapter 2.

To illustrate the concept of stochastic volatility, we will "jump the gun" and give a simulated path for the Heston model (introduced in Chapter 2) in figure 1.5 below. Note that with the Heston Model the index (asset price) is modelled according to (1.3) above, and the stochastic volatility component is modelled according to equation (1.4) below, with $f(x) = \sqrt{x}$:

$$dV(t) = \kappa\big(\theta - V(t)\big)dt + \xi\sqrt{V(t)}\,dW_2(t), \tag{1.4}$$

where $dW_2(t)$ is a second source of randomness that may be correlated with $dW_1(t)$. Here κ denotes the mean reversion parameter, θ the long term mean of the volatility process and ξ is a constant parameter value that indicates the volatility of the volatility process.

Details of the parameters and the code for the simulation below may be found in the supporting material (details given in Appendix A).

Some notable features of the given graphs:

- The volatility is not constant.
- The index graph contains features one would expect from observing stock market dynamics, i.e. abrupt jumps, direction of price movement is opposite to the volatility movement and volatility clustering. Technical discussions of these features will follow in later chapters.
- There are remarkable jumps over very short periods. Around time 500 and 800 we can clearly see that a higher volatility has led to a great move in the index price from 1000 to 2000 and back down from 2400 to 1300 approximately.
- The volatility process follows a mean reverting process with long term average around 0.15.
- The volatility process itself has a constant volatility of 0.04.

Generally for equity assets it is the case that there is a negative correlation between the movement

zero but since different parameterisations of the BS model is applied to different strikes it is not evident that these risks offset each other since changes in the underlying asset will likely yield different deltas at any future time.

between the volatility movement, although some co-movement can take place especially if the correlation coefficient is closer to 0 than to -1.

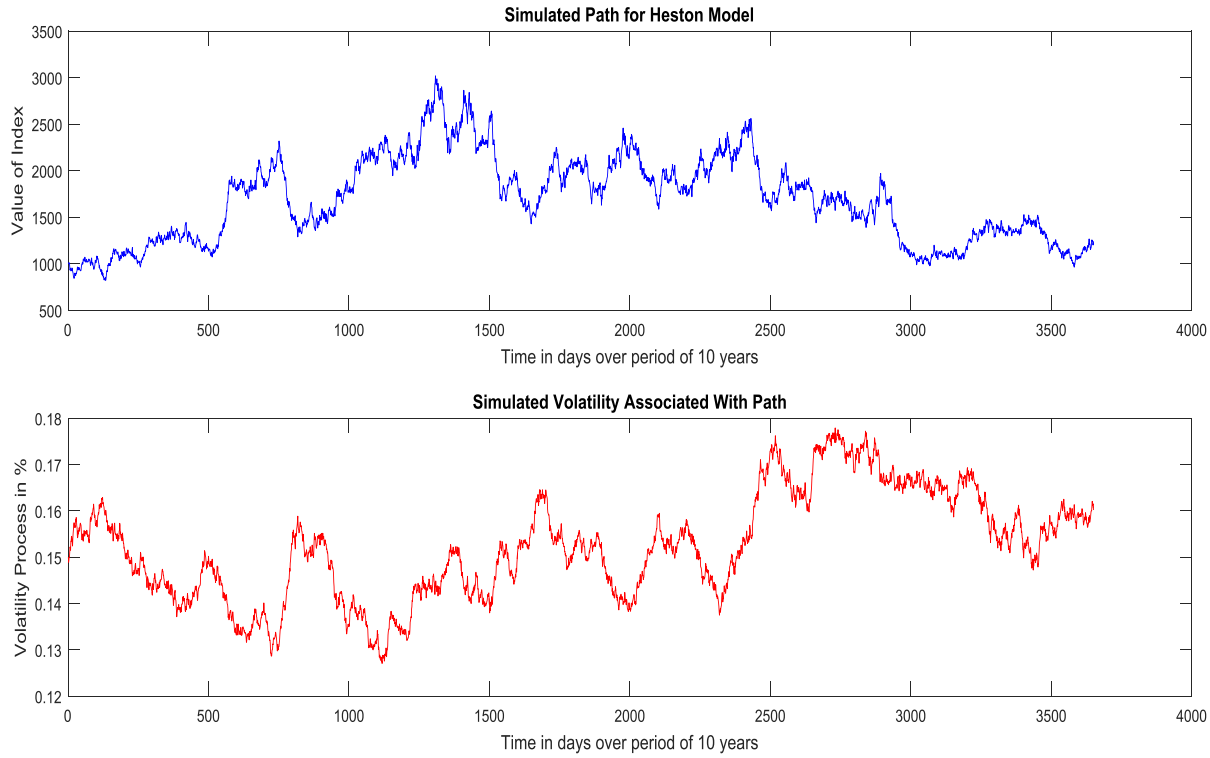

Figure 1.5: Simulated stock path and associated volatility path under the Heston Model.

1.2.1. **Modelling the risk-free rate stochastically**

Another improvement that can be made to the BS methodology is to model the risk-free interest rate stochastically. The motivation for this change stems from the fact that the fixed income markets have recognised the stochastic nature of the short rate process and the correlation interest rates have with the overall market movements (see (Alam and Uddin, 2009) for a thorough discussion). Additionally, there has been an introduction in recent years of equity-interest rate hybrid derivatives which cannot be valued without using stochastic interest rates (Overhaus et al, 2007). (Equity-interest rate hybrid derivatives are derivatives where the payoff profile depend on both the evolution of the interest rates and the underlying asset, e.g. conditional trigger swaps[5]).

The inclusion of stochastic interest rates, based on models such as the Hull-White short rate model, makes derivative pricing far more complicated than, for instance, including two stochastic volatility components. Indeed, without some simplifying assumptions and complicated approximations, the pricing methods are reduced to resorting to Monte Carlo techniques to price derivatives. Later in

[5] A standard interest rate swap is defined as a agreement between two parties where the parties exchange interest rate payments based on some notional amount. Party A typically pays a fixed interest rate to party B and party B pays a variable interest rate to party A. In a conditional trigger swap the fixed interest rate payments depend on the performance of some equity index at some future time T. If the equity index rises above a barrier level B then the fixed coupons are set at c_1 and otherwise at some level c_2 for duration of the swap agreement.

Chapter 2 we will look at how semi-analytical formulae can be developed to incorporate stochastic interest rates. A general solution, subject to certain assumptions, is then also presented in Chapter 3.

1.3. **Conclusion**

In this introductory chapter the notion of stochastic volatility was introduced and its necessity was supported by examples. The BS framework was also shown to be inconsistent with market observations. This was accomplished by examining implied volatility graphs and surfaces from real market data.

The shortcomings of the BS Model were addressed in a brief manner by introducing the notion of stochastic volatility and giving reasons why the risk-free interest rate should be a stochastic process. These ideas will be expanded on in Chapter 2, with some of the finer theoretical detail being discussed in Chapter 3.

Mention of the endeavours we will undertake in Chapter 2 merit discussion here: The existence of many stochastic volatility models with constant and stochastic interest rates for pricing exotic equity derivatives. One might ask whether some of the large amount of models found in literature are redundant. In other words, is there a single model that will trump all others, or should each model be considered as a suitable candidate that might apply in certain situations? The answer is provided in Chapter 4 together with our final conclusions.

Lastly, here we present a rough guide on how the rest of this thesis is laid out:

- A model is chosen to reflect the underlying and the volatility and interest rate processes are specified (Chapter 2).
- The pricing techniques under a selection of these models are either given in closed form or a formula for numerical implementation is given (Chapter 3).
- Next the models are calibrated to market data and implementation is considered (Appendix F).
- Finally through various means we simulate option prices and compare our results to simulated market data to determine the best stochastic volatility model (Chapter 4).

While this gives an accurate picture of the coming chapters, it should be noted that a number of additional topics will be considered as well.

Chapter 2- Literature Review: Stochastic Interest Rate and Volatility Models

2.1 **Introduction**

In the previous chapter stochastic volatility was introduced. It was showed, using market data, that the volatility smile and skew contradict the traditional Black Scholes (BS) assumption of constant volatility. Then following this, two possible (but not mutually exclusive) solutions to overcome the failures of the BS Model was mentioned, i.e. using stochastic volatility and stochastic interest rates. Now building on this, our aim in this chapter is to investigate in detail how principally these two solutions can improve on the BS methodology.

First, some background knowledge is given regarding Stochastic Differential Equations (SDE). This mostly provides a very general theoretical foundation for the discussion of local volatility models and the important CIR (Cox -Ingersoll-Ross) process. Much of the theory will not be explicitly used since the traditional pricing methods via partial differential equations, for example by using Itô's Lemma, are difficult to apply numerically when pricing exotic equity derivatives in a stochastic volatility context. The very brief highlight on characteristic functions, considered in detail in Chapter 3, serves essentially as motivation for why, for the majority of models, the characteristic functions for the majority of the models are given.

Second, after the background knowledge section, a discussion into local volatility models is given which highlights some theoretical developments. Thereafter two important interest rate models are discussed and each of these (effectively local volatility) models will be used as part of the building blocks to model stochastic volatility and stochastic interest rates later on. The CEV (Constant Elasticity of Variance) Model is also briefly mentioned before a technical explanation is given why local volatility models, on their own, are not a valid solution to the shortcomings of the BS methodology.

Thereafter, the remaining, and largest, portion of this chapter is devoted to stochastic volatility models. Each of the sections are split according to models based on Brownian motion diffusions, jump diffusions and advance models. The section on Brownian motion is split into two parts respectively covering basic models, such as the Heston Model and then extensions of the basic models. The section on including jumps in the stochastic processes is mainly devoted to sequentially developing a single sophisticated model found in literature. Thereafter an advance section is given for the latest models that have appeared in literature in the last year (from end of 2015 to end 2016).

Finally, the reader should note that not all the models will be considered in depth and neither are all models found in literature given here. Many technical issues and proofs will be omitted. The discussion of all the models will be incomplete in the sense that some technical detail (e.g. exact simulation schemes, calibration of models, etc.) is deferred until Chapters 3. However, important models will be given considerable attention with a wide variety of mathematical techniques employed to analyse their dynamics. Where possible, a brief overview of technical material is included to improve readability.

2.2 **Local and stochastic volatility models**

In this section we will present diffusion models only. These models are based on Stochastic Differential Equations with Brownian Motion (BM) as the source of randomness. Note that the choice of a particular stochastic volatility model is motivated by its prominence in literature as well as the popular models used in practice.

Later on in section 2.3 we will also consider Lévy Processes - a more general class of models that extends the BM framework.

2.2.1 **Background knowledge**

Equation (1.1) in Chapter 1 gave the classical Geometric Brownian Motion (GBM) model's Stochastic Differential Equation (SDE). Here a general framework regarding stochastic diffusion models is presented as preparation for the concepts that follow later on.

According to (Campolieti and Makarov, 2014), the following n-dimensional diffusion process $\{\boldsymbol{X}(t)\}_{t\geq 0}$ generalises to a system of SDE:

$$dX_i(t) = \mu_i\big(t, \boldsymbol{X}(t)\big)dt + \sum_{j=1}^{d} \sigma_{ij}\big(t, \boldsymbol{X}(t)\big) \cdot d\boldsymbol{W}(t), \qquad (2.1)$$

where drift function $\mu(t, \boldsymbol{x})$ *and volatility* $\sigma(t, \boldsymbol{x})$ *are known functions of time and the unknown vector asset price process* $\boldsymbol{X}(t) = \big(X_1(t), X_2(t), \dots, X_n(t)\big)$ *in n dimensions.*
Similarly we define $\boldsymbol{W}(t) = \big(W_1(t), W_2(t), \dots, W_d(t)\big)$ *in d dimensions.*
Equivalently we can write the equation as:

$$X_i(t) = \mu_i\big(t, \boldsymbol{X}(t)\big)dt + \boldsymbol{\sigma}_i\big(t, \boldsymbol{X}(t)\big) \cdot d\boldsymbol{W}(t),$$

where $\boldsymbol{\sigma_i}(t) = \big(\sigma_{i1}(t),\ \sigma_{i2}(t), \dots, \sigma_{id}(t)\big)$ *for* $i = 1, \dots, n$.

In other words we have an n-dimensional diffusion process driven by d-dimensional Brownian motions. In literature it is well known that if the values of n and d are equal then the market has a unique risk-neutral measure $\mathbb{P}$ and the market model admits no arbitrage opportunities. Importantly if $n < d$ then there exists many risk-neutral measures. The discussion on risk-neutral measures is carried on in Chapter 3 where more general cases are considered.

Denote the drift vector by $\boldsymbol{\mu}(t, \boldsymbol{x}) = (\mu_1(t, \boldsymbol{x}), \dots, \mu_n(t, \boldsymbol{x}))$. Recall that the norm of a vector is defined as $\|\boldsymbol{a}\| = \sqrt{\sum_{i=1}^{n} a_i^2}$. Now, under global Lipschitz conditions there exists a unique strong solution to (2.1) with continuous sample paths. The conditions for this are (Campolieti & Makarov 2014):

1.) The functions $\boldsymbol{\mu}(t, \boldsymbol{x})$ *and* $\boldsymbol{\sigma}(t, \boldsymbol{x})$ are locally Lipschitz in $\boldsymbol{x}$ and uniformly in t, in other words

we have that for some constants T and N there exists a constant K:

$$\|\boldsymbol{\mu}(t,\boldsymbol{x}) - \boldsymbol{\mu}(t,\boldsymbol{y})\| + \|\boldsymbol{\sigma}(t,\boldsymbol{x}) - \boldsymbol{\sigma}(t,\boldsymbol{y})\| < K\|\boldsymbol{x} - \boldsymbol{y}\|.$$

$where\ \|\boldsymbol{x}\|, \|\boldsymbol{y}\| \leq N\ and\ 0 \leq t \leq T$

2.) The functions $\boldsymbol{\mu}(t,\boldsymbol{x})\ and\ \boldsymbol{\sigma}(t,\boldsymbol{x})$ satisfy the following linear growth condition in the variable x:

$$\|\boldsymbol{\mu}(t,\boldsymbol{x})\| + \|\boldsymbol{\sigma}(t,\boldsymbol{x})\| \leq K\,(1 + \|\boldsymbol{x}\|).$$

In other words we do not want the process to explode to infinity in finite time.

3.) The starting value of the process $\boldsymbol{X}(0)$ is independent of $F_T^{\boldsymbol{W}} \equiv \sigma(\boldsymbol{W}(t); 0 \leq t \leq T)$ and $E\big(\boldsymbol{X}^2(0)\big) < \infty$. Here $\sigma(\boldsymbol{W}(t))$ refers to the Sigma-Algebra generated by the process $\boldsymbol{W}(t)$.

The above conditions are sufficient, but not necessary, to guarantee a strong unique solution. Additionally, if these conditions are not met then a "weak" solution may also exist. Here a "weak" solutions means the SDE can be described in probabilistic terms. The above conditions are only applicable to local volatility models. For stochastic volatility models it is generally impossible to find an analytical solution.

To give an example of a strong solution we consider the following one-dimensional SDE:

$$dX(t) = \big(\alpha(t) + \beta(t)X(t)\big)dt + \big(\gamma(t) + \delta(t)X(t)\big)dW(t), \tag{2.2}$$

$where\ \alpha(t), \beta(t), \gamma(t), \delta(t)\ are\ all\ adapted\ processes, assumed\ continuous\ functions\ of\ time. The\ functions\ may\ be\ random, however\ they\ may\ not\ be\ functions\ of\ X(t).$
$Note\ that\ \mu(t,x) = \alpha(t) + \beta(t)x,\ \sigma(t,x) = \gamma(t) + \delta(t)x\ (see\ equation\ (2.1)).$

The equation is known as a linear SDE and equation (2.2) can be found in (Campolieti and Makarov, 2014). They provide the solution as:

$$X(t) = U(t)\left(X(0) + \int_0^t [\alpha(s) - \gamma(s)\delta(s)]U^{-1}(s)ds + \int_0^t \gamma(s)U^{-1}(s)dW(s)\right), \tag{2.3}$$

$where\ U(t) = e^{\int_0^t \left(\beta(s) - \frac{1}{2}\sigma^2(s)\right)ds + \int_0^t \sigma(s)dW(s)}\ and\ U^{-1}(t) = \big(U(t)\big)^{-1}.$

Now, pricing of derivatives often involves discounting at a constant risk-free rate of interest r. This assumption can be relaxed by modelling r stochastically as mentioned in Chapter 1. Several models for a stochastic risk-free (short) rate process $\{r(t)\}_{t\geq 0}$ are of a similar form to (2.2):

$$dr(t) = a\big(t, r(t)\big)dt + b\big(t, r(t)\big)dW(t). \tag{2.4}$$

Typically we assume the functions a and b are smooth and obey the criteria mentioned for (2.1) to have a strong solution. Additionally some short rate models are classified as *affine term structure models*, in other words any short-rate model that gives zero-coupon bond prices of the form:

$$Z(t,T,r) = \exp(A(t,T) - C(t,T)r(t)). \tag{2.5}$$

Here t is the current time, T the maturity date and furthermore we have that $A(t,T)$ and $C(t,T)$ are independent of $r(t)$. In a risk neutral framework equation (2.4) be can re-written into the following form:

$$dr(t) = a(t, r(t)) - \gamma(t, r(t))b(t, r(t))dt + b\big(t, r(t)\big)dW_Q(t). \tag{2.6}$$

The quantity $\gamma(t, r(t))$ is known as the market price of risk. It is defined via Girsanov's Theorem[6] when changing probability measures from $\mathbb{P}$ in (2.4) to $\mathbb{Q}$ in (2.6).

As is shown by (Campolieti and Makarov, 2014) the following pricing function

$$Z\big(t, T, r(t)\big) = E_Q\left[e^{-\int_t^T r(u)du} \,|\mathcal{F}_t\right].$$

satisfies the following PDE via the Discounted Feynman-Kac Theorem (Theorem 2.3, as stated below Theorems 2.1 and 2.2):

$$\frac{\partial Z}{\partial t} + \frac{b^2(t,r)}{2}\frac{\partial^2 Z}{\partial r^2} + a(t,r) - \gamma(t,r)b(t,r)\frac{\partial Z}{\partial r} - rZ = 0. \tag{2.7}$$

Here the notation is shortened for readability, e.g. $r(t) = r$.

The theory given so far on the stochastic interest rates can be generalised by considering Itô's Formula for a Function of Several Processes. This is stated as:

Theorem 2.1

Let the vector-valued process $\boldsymbol{X}(t) = \{X_1(t), X_2(t), \dots, X_n(t)\}_{t\geq 0}$ satisfy the SDE in equation (2.2) and assume that $f(t, \boldsymbol{x}) = f(t, x_1, \dots, x_n)$ is a real-valued function on $\mathbb{R}_+ \times \mathbb{R}^n$ that is continuously diffenentiable w.r.t time and twice continuously differentiable w.r.t to the n-variable ables $x_1, \dots, x_n$. Then the process defined by $F(t) \coloneqq f\big(t, \boldsymbol{X}(t)\big)$ is an Itô process with stochastic differential $dF(t)$ given by

$$dF(t) = \frac{\partial f}{\partial t}\big(t, \boldsymbol{X}(t)\big)dt + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\big(t, \boldsymbol{X}(t)\big)dX_t(t) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}\big(t, \boldsymbol{X}(t)\big)\cdot d\big[X_i, X_j\big](t). \tag{2.8}$$

Recall that the operator $d\big[X_i, X_j\big] = dX_i(t)\cdot dX_j(t) = \boldsymbol{\sigma_i}(t)\cdot\boldsymbol{\sigma_j}(t)dt = C_{ij}(t)dt$where we define $C_{ij}(t) = \boldsymbol{\sigma_i}(t)\cdot\boldsymbol{\sigma_j}(t) = \sum_{k=1}^{n}\sigma_{ik}(t)\times\sigma_{jk}(t)$.

By inserting equation (2.1) into equation (2.8) we arrive at the following general formula:

$$df\big(t, \boldsymbol{X}(t)\big) = \left[\frac{\partial f}{\partial t}\big(t, \boldsymbol{X}(t)\big) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}(t))\frac{\partial^2 f}{\partial x_i \partial x_j}\big(t, \boldsymbol{X}(t)\big) + \sum_{i=1}^{n}\mu_i(t)\frac{\partial f}{\partial x_i}(t, \boldsymbol{X}(t))\right]dt + \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\big(t, \boldsymbol{X}(t)\big)\boldsymbol{\sigma_i}(t)\cdot d\boldsymbol{W}(t). \tag{2.9}$$

Later in section 2.3 this formula is extended to include the effects of adding in a jump diffusion.

Next we consider the issue of a density function for some arbitrary process $X(t)$ satisfying equation (2.1). Often in pricing derivatives we have to deal with partial differential equations and the

[6] Girsanov's Theorem informally states that there exists a change of measure generated by the process $\gamma(t)$ such that $W_{\mathbb{Q}}(t) = W_{\mathbb{P}}(t) + \int_0^t \gamma(s)ds$

expectation of the payoff function of a random variable. The link between the two is given naturally by the Feynman-Kac theorem (theorem 2.2 below) or the "Discounted" Feynman-Kac theorem (theorem 2.3 below). These are respectively:

Theorem 2.2

Let $\{\boldsymbol{X}(t) \coloneqq (X_1(t), \dots, X_n(t))\}_{0 \le t < T}$ *solve the system of SDE in equation* (2.1) *and let* $\phi: \mathbb{R}^n \to \mathbb{R}$ *be a Borel function. Additionally assume that the following holds as well*:

$$\int_0^T E\left[\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(s, \boldsymbol{X}(s))\sigma_{ij}(s, \boldsymbol{X}(s))\right)^2\right] ds < \infty, j = 1, \dots, n.$$

Suppose the function $f(t, \boldsymbol{x})$ *is a solution of the following PDE*:

$$\frac{\partial f}{\partial t}(t, \boldsymbol{X}(t)) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}(t, \boldsymbol{x}))\frac{\partial^2 f}{\partial x_i \partial x_j}(t, \boldsymbol{X}(t)) + \sum_{i=1}^{n} \mu_i(t)\frac{\partial f}{\partial x_i}(t, \boldsymbol{X}(t)) = 0,$$

$\forall\, \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, t < T$, *subject to the terminal condition* $f(T, \boldsymbol{x}) = \phi(\boldsymbol{x})$. *Then, assuming* $E_{t,\boldsymbol{x}}[\phi(\boldsymbol{X}(T))] < \infty$, $f(t, \boldsymbol{x})$ *can be solved as*:

$$f(t, \boldsymbol{x}) = E_{t,\boldsymbol{x}}[\phi(\boldsymbol{X}(T))] \equiv E[\phi(\boldsymbol{X}(T))|\boldsymbol{X}(t) = \boldsymbol{x}],$$

for all $\boldsymbol{x} \in \mathbb{R}^n, 0 \le t \le T$.

The function $\phi(\boldsymbol{x})$ represents the payoff profile of a derivative at time T. Note that the first condition in Theorem 2.2. is a technical term known as the square-integrability condition. Intuitively this ensures that the stochastic process $\sigma_{ij}(s, \boldsymbol{X}(s))$ is well defined (does not grow to infinity in finite time) otherwise the process $f(t, \boldsymbol{X}(t))$ is not a martingale and we can not obtain the price of derivative through the given the expectation $E[\phi(\boldsymbol{X}(T))|\boldsymbol{X}(t) = \boldsymbol{x}]$.

Theorem 2.3

Let $\{\boldsymbol{X}(t) \coloneqq (X_1(t), \dots, X_n(t))\}_{0 \le t < T}$ *solve the system of SDE in equation* (2.1) *and let* $\phi: \mathbb{R}^n \to \mathbb{R}$ *be a Borel function and let* $r(t, \boldsymbol{x}): [0, T] \times \mathbb{R}^n \to \mathbb{R}$ *be a lower bounded continuous function. Additionally assume that following holds as well*:

$$\int_0^T E\left[\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(s, \boldsymbol{X}(s))\sigma_{ij}(s, \boldsymbol{X}(s))\right)^2\right] ds < \infty, j = 1, \dots, n.$$

Then, the function defined by the following conditional expectation

$$f(t, \boldsymbol{x}) = E_{t,\boldsymbol{x}}\left[e^{-\int_t^T r(u, \boldsymbol{X}(u))du}\phi(\boldsymbol{X}(T))\right] \equiv E\left[e^{-\int_t^T r(u, \boldsymbol{X}(u))du}\phi(\boldsymbol{X}(T))|\boldsymbol{X}(t) = \boldsymbol{x}\right],$$

for all $\boldsymbol{x}, 0 < t < T$, *subject to the terminal condition* $f(T, \boldsymbol{x}) = \phi(\boldsymbol{x})$.

Note that $r(u, \boldsymbol{X}(u))$ can be interpreted to represent a stochastic interest rate process.

These theorems are the versions found in (Campolieti and Makarov, 2014), with slight alterations for notational consistency.

One needs the risk neutral density of the process $\boldsymbol{X}(t)$ in order to implement some numerical procedure for solving the conditional expectations given in theorems 2.2 and 2.3. We face a problem however in that the risk neutral density is not always given or often cannot be found at all in closed form. We could solve the partial differential equation directly, although we will use the characteristic function, defined below, that works better for option pricing.

In mathematical statistics it is well established that the risk neutral density can be recovered from the characteristic function of a Random Variable. Using this approach in option pricing has been championed in the works of (Duffie et al, 2000). We will look at their methods in greater detail in Chapter 3, subsection 3.1.1, but the core idea is given here.

Let $i = \sqrt{-1}$, a $characteristic\ function\ \varphi(u)$ of a Random Variable X is defined as:

$$\varphi(u) = E[\exp(iuX)] = \int_{-\infty}^{\infty} \exp(iux)\, dF(x). \tag{2.10}$$

Now, in order to determine the density function from the characteristic function we resort to inverting (2.10) using the following relationships, which are proved in Chapter 3:

$$Cumulative\ distribution\text{:}\ \ F(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty} \frac{e^{iux}\varphi(-u) - e^{-iux}\varphi(u)}{iu}\, du.$$

$$Density\ function\text{:}\ F'(x) = f(x) = \frac{1}{\pi}\int_0^{\infty} e^{-iux}\varphi(u) du. \tag{2.11}$$

For most of the models considered here the characteristic function is available, although to determine the density function we would have to use some numerical scheme. In this chapter the characteristic functions will be given for most of the models although for pricing purposes some of the technical detail will be omitted.

Now that most of the background knowledge has been given we can move onto the local and stochastic models which are the focus of this chapter.

2.2.2 **Time-dependent volatility models**

Consider the following Stochastic Differential Equation (SDE):

$$dX_t = rX_t dt + \sigma(t)X_t dW_t^*,$$

$where\ W_t^*\ is\ a\ \mathbb{P}\text{-}Brownian\ motion\ (under\ the\ unique\ risk\ neutral\ measure\mathbb{P})$
$and\ r\ is\ the\ risk\text{-}free\ rate\ of\ interest.$

This is a time-dependent volatility model for some arbitrary deterministic function $\sigma(t)$. Using techniques in stochastic calculus shows that the following is the solution for $X_T\,, for\ 0 < t < T$:

$$X_T = X_t \times \exp\left(r(T-t) - \frac{1}{2}\int_t^T \sigma^2(s)ds + \int_t^T \sigma(s)dW_s^*\right)$$

If we define $\overline{\sigma^2} = \frac{1}{T-t}\int_t^T \sigma^2(s)ds$ then we can still use the BS formula in Chapter 1 exactly as it is with σ replaced by $\overline{\sigma^2}$. Since $\overline{\sigma^2}$ differs for different maturity dates we have a term structure emerging for the implied volatility. However the implied volatility surface is still flat (level) across all strike prices as we are still using a constant, albeit averaged, value for the volatility.

Time dependent volatility is only useful for long dated options and only then as an approximation since the slope of the implied volatility surface still shows a skew, as was seen in Figure 1.4. We end our discussion here since there are better approaches mentioned in the coming subsections.

2.2.3 **Local volatility models**

In the following subsections characteristics of local volatility models are discussed. Following that the Constant Elasticity of Variance Model is discussed briefly. Lastly, a set of proofs are given showing why local volatility models, under the given assumptions, are inaccurate and can lead to underperform the BS Model.

The discussion given here underpins some of the characteristics of the stochastic volatility and stochastic interest rate models that appear in subsection 2.2.4 and the rest of this chapter. Additionally, one has to recognise that Local Volatility models are also an attempt to improve on the BS methodology. However, as will be shown, this attempt in literature has not worked out well. Nonetheless, the theory is still important as some of the work done on stochastic volatility and stochastic interest rate models rely on the concepts given here.

2.2.3.1 **Characteristics of local volatility models**

A discussion of general characteristics of Local Volatility (LV) models merit attention. The work of (Dupire, 1994) as well as (Derman and Kani, 1994), to whom the original development of these models are accredited, is presented here with alterations made only to assist the logical flow of concepts and adjusting the notation to be consistent with the preceding notation used.

Consider the following asset price SDE:

$$\frac{dS(t)}{S(t)} = (r-q)dt + \sigma\big(t,S(t)\big)dW_Q(t), t \geq 0, S(0) = S_0.$$

Here r is the risk-free interest rate, q is constant dividend yield and $W_Q(t)$ is a Brownian motion under the probability measure Q. The volatility parameter is dependent both on the current time and state of the process and is called the $local\ volatility\ function$.

Now consider a European-style derivative written on the underlying asset $S(t)$. As is well known, the no-arbitrage value of the contract, $V(t,S)\ at\ time\ t < T$ is given by:

$$V(t,S) = e^{-r(T-t)}E_Q[V(T,S(T)]. \tag{2.12}$$

By Theorem 2.3 in subsection 2.2.1 we have that $V(t,S)$ satisfies the famous Black-Scholes partial differential equation:

$$\frac{\partial V(t,S)}{\partial t} + \frac{1}{2}\sigma^2(t,S)S^2\frac{\partial^2 V(t,S)}{\partial S^2} + (r-q)S\frac{\partial V(t,S)}{\partial S} - rV(t,S) = 0. \tag{2.13}$$

Here the price of a European call option $C(t,S,T,K)$ satisfies (2.12). Now, if the European call

option price is taken to be a function of T and K only, i.e. $C(t,S,T,K) \coloneqq C(T,K)$, i.e. t,S held fixed, then the price satisfies $Dupire's\ Equation$[7]:

$$\frac{\partial C(T,K)}{\partial T} = \frac{1}{2}\sigma^2(T,K)K^2\frac{\partial^2 C(T,K)}{\partial K^2} + (q-r)K\frac{\partial C(T,K)}{\partial K} - qC(T,K). \qquad (2.14)$$

This equation can easily be proved via equation (2.12) by taking suitable derivatives.

Next for this local volatility model there is a relationship between the local volatility $\sigma^2(T,K)$ and Call option prices:

$$\sigma^2(T,K) = \frac{2}{K^2}\frac{\frac{\partial C(K,T)}{\partial T} + rK\frac{\partial C}{\partial K}}{\frac{\partial^2 C(K,T)}{\partial K^2}}.$$

This gives a semi-explicit formula for calculating the volatility directly from market prices. Once $\sigma(T,K)$ is determined from Call option prices then this function can be used to calibrate the local volatility function $\sigma\big(t,S(t)\big)$. As an example, consider the following specification $\sigma\big(t,S(t)\big) = \alpha\big(S(t)\big)^{\beta+1}$. Using $\sigma(T,K)$ allows us to determine the parameters α and β via a calibration procedure.

However for this to be possible we would ideally need a substantial number of call option prices and a continuous number of strike prices in order to determine $\sigma^2(T,K)$. This is not practical as option data is limited and there are too few strike prices around to make any sort of meaningful numerical implementation. Furthermore (Jiang et al, 2003) suggest that evaluating the derivatives numerically, especially with reference to the strike K might lead to unstable procedures. Nonetheless, (Jiang et al, 2003) found a procedure using an "optimal control framework" to recover the local volatility function assumed to be independent of time.

Additionally (Rodrigo and Mamon, 2007) show how to recover time dependent parameters using an inverse Stieltjies Moment approach in equation (2.10), i.e. where the risk-free rate of interest and the dividend rate and volatility function are all functions of time. They also show how to adopt this method to calculating the implied volatility curve.

Lastly, a flexible choice of the function $\sigma(t,S)$ is given by $\sigma(t,x) = \big(x + a(t)\big)^{\beta(t)}$,see (Kienitz and Wetterau, 2012), and this choice of the local volatility function is mentioned to provide great flexibility in modelling, especially w.r.t time, but at the cost of more involved valuation formulae.

2.2.3.2 **Constant elasticity of variance (CEV) Model**

The CEV Model will not play an important role in the coming chapters, however its inclusion provides insights into the β parameter of the SARB Model in section 2.2.5.3.

Now, (Cheung and Ng, 1992) cite and explore a phenomenon known as the "leverage effect" regarding stock prices. Essentially this means that there is a tendency for the (stock) price to be negatively correlated with its volatility. A common explanation for this is simple: As the stock price decreases the company becomes more leveraged as its debt/equity ratio increases. This leads to additional volatility in the stock price. Note that this is not the only explanation.

[7] We are essentially saying that the local volatility function can be expressed in terms of strike prices and the final time at maturity T when pricing Call options under this model.

(Cox, 1975) developed the following model which is essentially an extension of the BS Model to allow for the leverage effect via an additional parameter β (which is often negative for stocks):

$$dS(t) = \nu S(t)dt + \alpha \times \big(S(t)\big)^{\beta+1} dW(t). \qquad (2.15)$$

The parameters ν, α, β are all assumed to be constant. In relation to the discussion, later, in section 2.2.4.2 regarding the Squared Bessel Process (SQB) process[8], (Schroder, 1989) suggests that to transform (2.15) into an SQB process, (2.29) below, the following formulas and steps are used (using the notation and parametrisation given in (Campolieti and Makarov, 2014)):

Change (2.15) into the process $\{F(t)\}_{t\geq 0}$:

$$S(t) = e^{\nu t} F(\tau_t), where\ \tau_t = \begin{cases} \frac{1}{2\nu\beta}\big(e^{2\nu\beta t} - 1\big) & if\ \mu \neq 0 \\ t & if\ \mu = 0 \end{cases}$$

Next, a change of variables yields the following SQB process:

$$X(t) = \frac{F(t)^{-2\beta}}{\alpha^2\beta^2}.$$

Hence, if one were to apply Monte Carlo technique by simulating $X(t)$, which is a SQB process, the corresponding simulated process for the CEV can be obtained by a simple transformation:

$$S(t) = e^{\nu t}\big(\alpha^2\beta^2 X(\tau_t)\big)^{-\frac{1}{2\beta}}.$$

We will discuss how the SQB process can be simulated in detail in Chapter 3, section 3.2.4. This concludes our discussion on the CEV Model.

2.2.3.3 **Technical discussion on general local volatility models**

Below a technical discussion is given as to why local volatility models are generally inadequate and to serve as motivation for stochastic volatility models should be considered instead for derivative pricing. The argument is mathematical and is attributed to (Hagan et al, 2002) who also introduced the Stochastic-Alpha-Rho-Beta (SARB) Model considered in subsection 2.2.5.3 below. Two claims are given with proofs provided. This section then ends our discussion of Local Volatility models and they will not be considered again.

Claim 1: The market (implied volatility) smile predicted by Local Volatility models are inconsistent with observed market behaviour (Hagan et al, 2002).

Proof:

Suppose that a European Call option is written on some asset S. The option is exercised at time t_{ex} and on the settlement date, t_{set}, the holder of the options receives the asset S in exchange for the payment of K (strike). Valuing the option can be done via a forward contract: let $F(t)$ denote the

[8] The SQB process is explored in detail later, but for clarity it is defined by the following SDE: $dX(t) = \lambda_0 dt + \nu\sqrt{X(t)}\, dW(t)$.

forward price of the asset S for a forward contract on asset S that matures at time t_{set}. The current forward price is defined by $f = F(0)$. Finally define discount factor $D(t)$ which is simply the present value at time $t = 0$ of one monetary unit (e.g. R1, $1, etc) at time t.

Under some appropriate forward measure F, the value of the above option (at time $t = 0$) is given

$$V_{call} = D(t_{set})E_{\mathrm{F}}[(F(t_{ex}) - K)^{+}|\mathcal{F}_0]. \tag{2.16}$$

Here $\mathcal{F}_0$ denotes the Sigma-Algebra associated with the forward price process $\{F(t)\}_{t\geq 0}$.

The value of the Put option is given by:

$$V_{put} = V_{call} + D(t_{set})\big(K - F(t_{ex})\big). \tag{2.17}$$

Martingale Pricing theory demonstrates that the process $\{F(t)\}_{t\geq 0}$ is a martingale under measure F, and this implies the following SDE, related to $F(t)$:

$$dF(t) = C(t,F)dW(t), \qquad F(0) = F_{0.} \tag{2.18}$$

Here the coefficient $C(t,F)$ must be postulated. The following are famous choices for $C(t,F)$:

$Black\ Scholes$: $C(t,F) = \sigma_B F(t)$.

$Local\ Volatility$: $C(t,F) = \sigma_{LV}(t,F)F(t)$.

In the work of (Dupire, 1994) and shown earlier, the Local Volatility (LV) component is constructed by direct calibration to market prices (instead of specifying $\sigma_{LV}(t,F)$).

The process involves starting with some LV function, say $\sigma_{LV}\big(t,\hat{F}\big)$ (where the "hat" notation signifies an estimate) and evaluating (2.16) & (2.17). Consequently $\sigma_{LV}\big(t,\hat{F}\big)$ is varied until the model and market prices agree for every strike and exercise date. In practice there are several discrete $(j = 1,2,\dots,J)$ exercise dates $\{t_{ex}^{j}:\ t_{ex}^{K} < t_{ex}^{L}\ for\ K < L\}_{j\geq 1}$ and the various LV's $\sigma_{LV}\big(t,\hat{F}\big)$ are taken to be piecewise constant with respect to time:

$$\sigma_{LV}\big(t,\hat{F}\big) = \sigma_{LV}^{1}\big(\hat{F}\big), \qquad for\ t < t_{ex}^{1}.$$
$$\sigma_{LV}\big(t,\hat{F}\big) = \sigma_{LV}^{j}\big(\hat{F}\big), \qquad for\ t_{ex}^{j-1} < t < t_{ex}^{j}.$$
$$\sigma_{LV}\big(t,\hat{F}\big) = \sigma_{LV}^{J}\big(\hat{F}\big), \qquad for\ t > t_{ex}^{J}.$$

Thus the process starts by calibrating $\sigma_{LV}\big(t,\hat{F}\big)$ to market prices until they are reproduced at time t_{ex}^{1} for all strikes K and the process carries on until all set of exercise dates are exhausted for all strikes. Hence a model is constructed that can consistently value exotic options for various strikes. The problem with this comes from examining the SDE:

$$d\hat{F} = \sigma_{LV}\big(\hat{F}\big)\hat{F}dW, \qquad \hat{F}(0) = f. \tag{2.19}$$

Using perturbation techniques[9] for the model in (2.19) it is found that European Call and Put prices can be priced via a modified form for the BS volatility (in other words contingent claims are priced according to BS formula with volatility in (2.20)):

[9] Perturbation techniques refers to class of mathematical techniques where the solution to a difficult problem, say $\mathcal{S}$, is approximated by a series of solutions: $\mathcal{S} = \mathcal{S}_0 + \epsilon\mathcal{S}_1 + \epsilon^2\mathcal{S}_2 + \cdots$, here ϵ is some small quantity and $\mathcal{S}_0$ represents a very accurate approximation to $\mathcal{S}$. The terms $\mathcal{S}_1, \mathcal{S}_2, \dots$ represent further solutions so that $\mathcal{S}_0$ is adjusted by small amounts until it approximates $\mathcal{S}$ very well.

$$\sigma_{BS}(K,f) = \sigma_{LV}\left(\frac{1}{2}[f+K]\right)\left\{1 + \frac{1}{24}\frac{\sigma''_{LV}\left(\frac{1}{2}[f+K]\right)}{\sigma_{LV}\left(\frac{1}{2}[f+K]\right)}(f-K)^2 + \cdots\right\}. \quad (2.20)$$

The remaining error terms are very insignificant and only contribute at most 1% to the left-hand term. Suppose that on a given day the following values are observed: Forward price f_0 and volatility formula $\sigma^0_{BS}(K)$. (Here $\sigma^0_{BS}(K)$ is the implied volatility curve in two-dimensions observed by using equation (1.2) for various strikes).

Calibrating the LV model to this (market) implied volatility requires the following formula:

$$\sigma_{LV}(\hat{F}) = \sigma^0_{BS}(2\hat{F} - f_0)\{1 + \cdots\}. \quad (2.21)$$

Assuming that the LV model is calibrated by (2.21), we examine what happens when the forward price changes from f_0 to f. Using both (2.20) & (2.21) the new implied volatility curve is predicted (given that forward changed) to be:

$$\sigma_{BS}(K,f) = \sigma^0_{BS}(K + f - f_0)\{1 + \cdots\}.$$

Clearly if f_0 increases to f the curve (in two-dimensions, i.e. implied volatility vs K) shifts left and conversely if f_0 decreases. This is in contrast to market behaviour where if f_0 to f increases the curve shifts right. This point is further examined by (Hagan et al, 2002) who additionally mention that the implied volatility curve also shifts up regardless of whether f_0 increases or decreases (under the assumption of a perfect smile for the implied volatility smile).

Claim 2: The Delta hedging performance of Local Volatility models may perform worse than the BS Model.

Formally the value of a European Call option under LV models is given by:

$$V_{call} = BS(f, K, \sigma_{BS}(K,f), t_{ex}). \quad (2.22)$$

Here $\sigma_{BS}(K,f)$ is given by (2.19). Computing the Delta of this option using (2.22):

$$\Delta = \frac{\partial V_{call}}{\partial f} = \frac{\partial BS}{\partial f} + \frac{\partial BS}{\partial \sigma_{BS}} \times \frac{\partial \sigma_{BS}(K,f)}{\partial f}.$$

Informally this expression can be understood to mean that the original BS Delta is corrected by adding on the second term which is simply the BS Vega multiplied by the estimated change in volatility due to changes in market movements. This last factor, $\frac{\partial \sigma_{BS}(K,f)}{\partial f}$, has the wrong sign as the change in volatility $\sigma_{BS}(K,f)$ was shown under claim 1 to be in the wrong direction. This simply means that the sign of $\partial \sigma_{BS}(K,f)$ at any instantenous change of f is the opposite of what it should be. Consequently the naïve BS model has better or more accurate Delta hedges for Call options.

This ends the discussion of LV models. Hence the discussions from here on and in subsequent chapters will focus only on Stochastic Volatility models. However the background knowledge in the previous sections will still be of use in the coming sections. For further discussion on the technical details of LV models the reader is recommended to consult the work of (Rebonato, 2004).

2.2.4. **Models for stochastic interest rates**

There are many stochastic interest rate models to be found in literature. Since our primary concern is in pricing equity derivatives it makes little sense to develop overly complicated stochastic interest rate models. The two models presented here are not capable of generating interest rate smiles but serve nonetheless as useful models for the short rate process. The Cox-Ingersoll-Ross Model in particular is very important as it serves as the primary stochastic process that drives the volatility

process of many of the stochastic volatility models and stochastic interest rates models later in this chapter.

2.2.4.1 **Hull-White Model**

Regarding equation (2.4) for the short rate process $r(t)$, (Hull and White, 1990) proposed the following functions:
$a\big(t,r(t)\big) = \big(\alpha(t) - \beta(t)r(t)\big)$.
$b\big(t,r(t)\big) = \sigma(t)$.

Additionally we assume a known $r(0) = r_0$, as a starting value for the short rate process.
The discussion of the Hull-White Model is based on (Grzelak et al, 2012) who give a different parametrisation of the model which is more useful for the presentation of this model:

$$dr(t) = \lambda\big(\theta(t) - r(t)\big)dt + \sigma dW(t). \tag{2.23}$$

Note here that sigma is now time independent. There are substantial differences in literature between parameters that are to be taken as time dependent, with most sources agreeing on θ to be time dependent. An example of a time dependent $\theta(t)$ is where the parameter is used to match the initial term structure of interest rates as follows:

$$\theta(t) = f(0,t) + \frac{1}{\lambda}\frac{\partial}{\partial t}f(0,t) + \frac{\sigma^2}{2\lambda^2}\big(1 - e^{-2\lambda t}\big),$$

$$where\ f^M(0,t) = -\frac{\partial \log\big(Z^M(0,t)\big)}{\partial t}.$$

Here $f^M(0,t)$ is the forward rate calculated from market (observed) zero coupon bonds $Z^M(0,t)$.

The parameter λ is referred to as the mean reversion parameter of the process and serves as an indication of how quickly the short rate is pulled back towards $\theta(t)$. Until recently, negative interest rates were not a feature of real world markets and traditionally any short rate model that admitted negative rates were seen to be inadequate as the dynamics contradicted what was observed in the markets, i.e. only positive rates.

However, accurate at 15 February 2016, the following countries have negative interest rates Denmark, Sweden, Switzerland and Japan. Here we mean that the rate at which commercial banks' excess funds are held with the respective central banks earns a negative return. This means that negative interest rates are a reality and thus some models, which allow for this, may be used for modelling.

Nonetheless, the model given in (2.23) can become negative with probability:

$$P(r(t) < 0) = \mathcal{N}\left(-\frac{f^M(0,t) + \frac{\sigma^2}{2\lambda^2}\big(1 - e^{-\lambda t}\big)^2}{\sqrt{\frac{\sigma^2}{2\lambda}\big[1 - e^{-2\lambda t}\big]}}\right).$$

With $N(x)$ denoting the normal distribution CDF. Some statistical facts of the model are provided which we will use later on for modelling and simulation purposes (e.g. see section 2.2.6.3 where the Hull White Model is used to specify stochastic interest rates).
Using equation (2.3) it immediately follows that

$$r(t) = \exp\big(-\lambda(t-s)\big)\, r(s) + \lambda \int_s^t \exp\big(-\lambda(t-u)\big)\, \theta(u)du + \sigma \int_s^t \exp\big(-\lambda(t-u)\big)\, dW_u. \tag{2.24}$$

Here we have that $s < t$. Furthermore we define the sigma algebra $\mathcal{F}_s^W := \sigma(W_u): 0 \leq u \leq s$

Calculating the conditional (on the sigma algebra $\mathcal{F}_s$) expected value and variance gives:

$$E[r(t)|\mathcal{F}_s] = \exp\big(-\lambda(t-s)\big)\, r(s) + \lambda \int_s^t \exp\big(-\lambda(t-u)\big)\, \theta(u) du.$$

$$\mathbb{V}[r(t)|\mathcal{F}_s] = \frac{\sigma^2}{2\lambda}\big(1 - \exp\big(-2\lambda(t-s)\big)\big).$$

Where the calculation of the variance is based on the well-known result:

$$\int_s^t f(u) dW_u \overset{d}{=} \mathcal{N}\left(0, \int_s^t f^2(u) du\right).$$

Equation (2.24) can be re-written into the following form which is known as the Hull-White Decomposition:

$$r(t) = \tilde{r}(t) + \psi(t). \tag{2.25}$$

Here (2.25) is split between a deterministic function $\psi(t)$ and a stochastic function with:

$$d\tilde{r}(t) = -\lambda \tilde{r}(t) dt + \sigma dW(t). \tag{2.26}$$

Using (2.26) the characteristic function of the Hull-White Model is given by (Kienitz and Wetterau, 2012):

$$\varphi_{HW}(u,t,T) = \exp\left(-\int_t^T \psi(s) ds + iu\psi(T)\right) \exp\big(A(u,\tau) + B(u,\tau)\tilde{r}(t)\big). \tag{2.27}$$

Here $\tau = T - t$ and the functions $A(u,\tau)$, $B(u,\tau)$ and ψ are given by:

$$A(u,\tau) = \frac{\sigma^2}{2\lambda^3}\left(\lambda\tau - 2\big(1 - e^{-\lambda\tau}\big) + \frac{1}{2}\big(1 - e^{-2\lambda\tau}\big)\right) - iu\frac{\sigma^2}{2\lambda^2}\big(1 - e^{-\lambda\tau}\big)^2 - \frac{u^2\sigma^2}{4\lambda}\big(1 - e^{-2\lambda\tau}\big).$$

$$B(u,\tau) = iue^{-\lambda\tau} - \frac{\big(1 - e^{-\lambda\tau}\big)}{\lambda}.$$

$$\psi(t) = \begin{cases} e^{-\lambda\tau} r(0) + \lambda \int_0^t e^{-\lambda(t-s)} \theta(s) ds \,; for\ 0 \leq t < T \\ f^M(0,T) + \frac{\sigma^2}{2\lambda^2}\big(1 - e^{-\lambda T}\big)^2 ; for\ t = T \end{cases}$$

The Hull-White Model's density function can be obtained from the characteristic function. The point of presenting $\theta(t)$ in the manner given above is to give a practical example of how time dependent parameters can be used in the modelling process where until now we have defined such parameters with a parametric form.

2.2.4.2 **Cox-Ingersoll-Ross Model**

In the article "A Theory of the Term Structure of Interest Rates" (Ross et al, 1985) proposed the following model for the short rate process $r(t)$:

$$dr(t) = \left(\alpha - \tilde{\beta} r(t)\right) dt + \sigma\sqrt{r(t)}\, dW_Q(t). \tag{2.28}$$

This model is widely known as the CIR Model and is contrasted from the Hull-White with the additional $\sqrt{r(t)}$ that appears with the σ term. The form given is in terms of risk neutral dynamics. Note that $\tilde{\beta} = \beta + \gamma\sigma$ where γ is known as the market price of risk (see Chapter 3, section 3.1.3 for a

detailed explanation and equation (2.6) above.). Additionally the model can never go negative as long as $r(0) = r_0 > 0$ and the process is mean-reverting with long-term mean given by $\alpha/\tilde{\beta}$.

The CIR Model is an affine model in that zero-coupon bond prices can be expressed in the form as in equation (2.5), i.e. $Z(t,T,r) = \exp(A(t,T) - C(t,T)r(t))$. Thus in order to derive $A(t,T)$ and $C(t,T)$ in that equation we begin with the following ODEs (ordinary differential equations) which specify $A(t,T)$ and $C(t,T)$ for the CIR Model:

$$\frac{\partial C(t,T)}{\partial t} - \tilde{\beta}C(t,T) - \frac{\sigma^2}{2}C^2(t,T) + 1 = 0. \tag{2.29}$$

$$\frac{\partial A(t,T)}{\partial t} - \alpha C(t,T) = 0. \tag{2.30}$$

with terminal conditions $A(T,T) = C(T,T) = 0$. In order to solve these equations a few standard techniques are used:

Use the transformation

$$\psi(t) = \exp\left(\frac{\sigma^2}{2}\int_t^T C(s,T)ds\right).$$

Taking derivatives:

$$\frac{\psi'(t)}{\psi(t)} = \frac{d}{dt}\ln\big(\psi(t)\big) = -\frac{\sigma^2}{2}C(t,T) \Rightarrow C(t,T) = \frac{-2}{\sigma^2}\frac{\psi'(t)}{\psi(t)}. \tag{2.31}$$

Similarly computing $\frac{\psi''(t)}{\psi(t)}$ yields the following:

$$C'(t,T) = \frac{-2}{\sigma^2}\frac{\psi''(t)}{\psi(t)} + \frac{\sigma^2}{2}C^2(t,T). \tag{2.32}$$

Substituting the derived equations into (2.29) and (2.30) above yields a second order linear ODE:

$$\psi''(t) - \tilde{\beta}\psi'(t) - \frac{\sigma^2}{2}\psi(t) = 0.$$

Recall the solution to this equation is given by

$$\psi(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}, \tag{2.33}$$

where c_1 & c_2 are constants to be determined by solving the appropriate boundary conditions and λ_1 & λ_2 are given by solving the appropriate characteristic equation. The solutions are:

$$\lambda_1 = \frac{\tilde{\beta} + \sqrt{\tilde{\beta}^2 + 2\sigma^2}}{2}, \lambda_2 = \frac{\tilde{\beta} - \sqrt{\tilde{\beta}^2 + 2\sigma^2}}{2}.$$

Denoting $\omega = \sqrt{\tilde{\beta}^2 + 2\sigma^2}$ the solution to (2.33) can be written:

$$\psi(t) = c_1 e^{\frac{1}{2}(\tilde{\beta}+\omega)t} + c_2 e^{\frac{1}{2}(\tilde{\beta}-\omega)t}. \tag{2.34}$$

Solving the coefficients involves constructing a linear system of equations based on the boundary conditions $\psi(T) = 1$ and $\psi'(T) = 0$ and hence the solutions are given by:

$$c_1 = \left(\frac{1}{2} - \frac{\tilde{\beta}}{2\omega}\right)e^{-\frac{1}{2}(\tilde{\beta}+\omega)T}, c_2 = \left(\frac{1}{2} + \frac{\tilde{\beta}}{2\omega}\right)e^{-\frac{1}{2}(\tilde{\beta}-\omega)T}.$$

Now recall the definitions $\cosh(x) = \frac{e^x + e^{-x}}{2}$ and $\sinh(x) = \frac{e^x - e^{-x}}{2}$. Using these functions (2.34) can

be rewritten as

$$\psi(t) = e^{-\frac{1}{2}\tilde{\beta}(T-t)}\left[\cosh\frac{\omega(T-t)}{2} + \frac{\tilde{\beta}}{2\omega}\sinh\frac{\omega(T-t)}{2}\right].$$

Equation (2.31) can now be used to finally solve for $C(t,T)$:

$$C(t,T) = \frac{2\sinh\frac{\omega(T-t)}{2}}{\omega\cosh\frac{\omega(T-t)}{2} + \tilde{\beta}\sinh\frac{\omega(T-t)}{2}}.$$

and using (2.30) $A(t,T)$ can be computed by integrating over $C(t,T)$. However using (2.31) this can be done more easily:

$$A(t,T) = \frac{2\alpha}{\sigma^2}\int_t^T \frac{d}{du}\ln\big(\psi(u)\big)\,du = \frac{2\alpha}{\sigma^2}\ln\left(\frac{1}{\psi(t)}\right); \psi(T) = 1,$$

$$\therefore A(t,T) = \frac{2\alpha}{\sigma^2}\ln\left(\frac{\omega e^{\frac{1}{2}\tilde{\beta}(T-t)}}{\omega\,cosh\frac{\omega(T-t)}{2} + \tilde{\beta}\,sinh\frac{\omega(T-t)}{2}}\right).$$

Hence using equation (2.5) we can use $A(t,T)$ & $C(t,T)$ to derive a closed form formula for zero-coupon bond prices. More importantly we will discuss the use these closed form expressions in Chapter 3, section 3.3.1.4 and use them in the numerical work in Chapter 4.

Naturally to use equation (2.5) the process $r(t)$ needs to be simulated. In order to accomplish this some theory is introduced here which is important and will be of use later in Chapter 3 when considering Monte Carlo simulations. The algorithms used to simulate the CIR process are based on the article by (Makarov and Glew, 2009) in which they discuss simulation of the Squared Bessel process.

Consider the λ_0 – dimensional Squared Bessel (SQB) process:

$$dX(t) = \lambda_0 dt + \nu\sqrt{X(t)}\,dW(t). \tag{2.35}$$

To gain an understanding of how this process looks, a few sample paths are generated and displayed in Figure 2.1 below.

The graph displays the SQB process for various values of $\mu = \frac{2\lambda_0}{\nu^2} - 1$. if $\mu \geq 0$ then the left-hand endpoint 0 acts as an entrance boundary. If $-1 < \mu < 0$ then 0 is a regular boundary and finally for $\mu \leq -1$ we have that 0 is an exit boundary for the process (Makarov and Glew, 2009). The significance of these boundary conditions is that they will form restrictions on the relationship between various parameters defined for some of the models that we will use later, e.g. the Heston Model in section 2.2.5.2.

The discussion on how the SQB process relates to the CIR Model mirrors that of (Campolieti and Makarov, 2014) and the sampling algorithm follows that of (Makarov and Glew, 2009).

A special case of (2.35) is given by:

$$dX(t) = (2\mu + 2)dt + 2\sqrt{X(t)}dW(t). \tag{2.36}$$

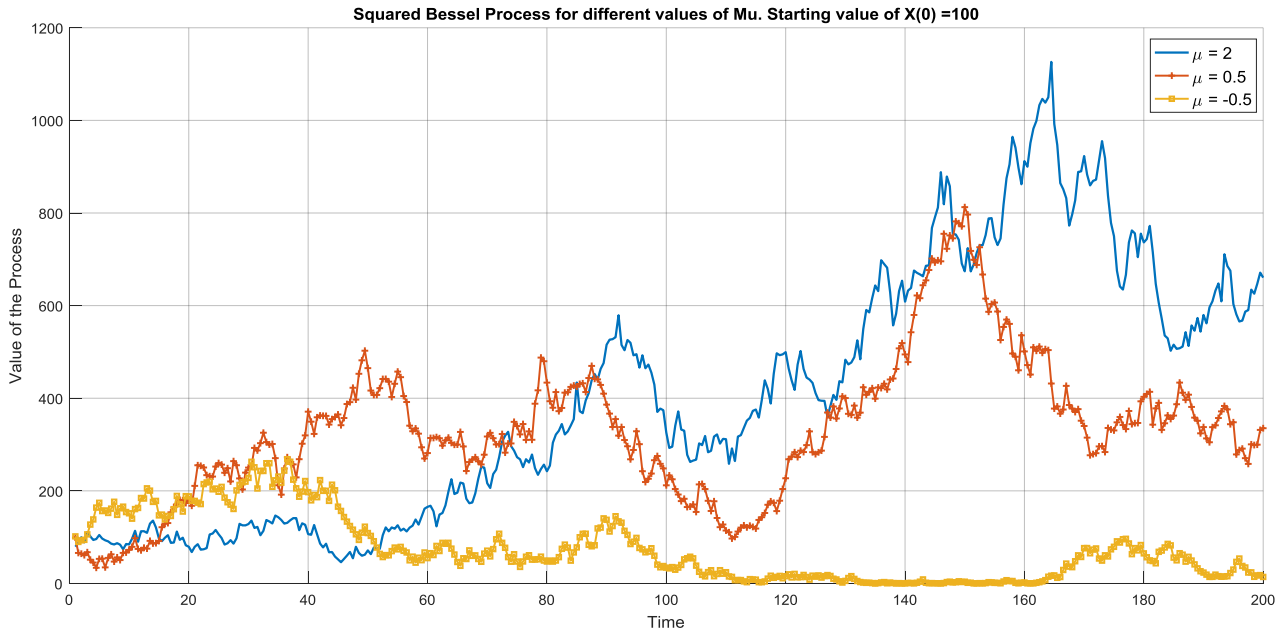

Figure 2.1: Sample paths under the Square Bessel Process for different values of μ.

The transition probability density function of the process (2.36) is known and is given by:

$$p_\mu\,(t;x_0,x) = \left(\frac{x}{x_0}\right)^{\frac{\mu}{2}} \frac{\exp\left(-\frac{x+x_0}{2t}\right)}{2t} I_{\tilde{\mu}}\left(\frac{\sqrt{xx_0}}{t}\right). \tag{2.37}$$

where $I_u(z)$ denotes the modified Bessel function of the first kind:

$$I_u(z) = \left(\frac{z}{2}\right)^u \sum_{j=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^j}{j!\,\Gamma(u+j+1)}.$$

Note that we have changed the notation of the probability function from $f(x)$ to $p_\mu(t;x_0,x)$ in order to be consistent with the notation of (Makarov and Glew, 2009).

Finally the relationship between the CIR process and SQB process (2.36) is given by the transformation (where $\tilde{\beta}$ and σ are from equation (2.28)):

$$r(t) = \exp\left(-\tilde{\beta}t\right)\frac{\sigma^2}{4}X(\tau_t).$$

and the time transformation τ_t is given by:

$$\tau_t = \begin{cases} t & if\ \tilde{\beta} = 0 \\ \dfrac{\exp\left(\tilde{\beta}t\right)-1}{\tilde{\beta}} & if\ \tilde{\beta} \neq 0 \end{cases}$$

Hence (2.37) for the CIR process can be expressed as

$$\tilde{p}(t;r_0,r) = c_t \exp\left(\tilde{\beta}t\right)\left(\frac{r\exp\left(\tilde{\beta}t\right)}{r_0}\right)^{\frac{\mu}{2}} \exp\left(-c_t\left(r\exp\left(\tilde{\beta}t\right)+r_0\right)\right) I_\mu\left(2c_t\sqrt{rr_0\exp\left(\tilde{\beta}t\right)}\right). \tag{2.38}$$

with $c_t = \frac{2}{\sigma^2 \tau_t}$ and $\mu = \frac{2\alpha}{\sigma^2} - 1$. Equations (2.37) and (2.38) hold for the specific parameterisation of the CIR process as given in (2.28). The relationship between (2.37) and (2.38) can be seen through the following relationship:

$$p^{(CIR)}(t, r_0, r) = \exp(\tilde{\beta} t)\, p^{(SQB)}(\tau_t, r_0, \exp(\tilde{\beta} t)\, r).$$

Note that for any probability distributions (e.g. (2.37)) that contain Bessel or hypergeometric functions it is difficult to generate (pseudo) random numbers or calculate the density function quickly. Hence, in the spirit of the article by (Makarov and Glew, 2009) three Randomised Gamma Distributions are now introduced that will assist us to calculate the density function quickly.

Equation (2.38) is a non central chi-square distribution and the general formula for such a distribution is:

$$f(x; \kappa, \lambda) = \frac{1}{2} e^{-\frac{(x+\lambda)}{2}} \left(\frac{x}{\lambda}\right)^{\frac{\kappa}{4} - \frac{1}{2}} I_{\frac{\kappa}{2}-1}\left(\sqrt{\lambda x}\right), \qquad x > 0. \tag{2.39}$$

Where $\kappa > 0$ is the degrees of freedom and $\lambda > 0$ is the noncentrality parameter.

Observe that if we expand the modified Bessel function $I_{\frac{\kappa}{2}-1}$ and collect similar terms then (2.39) can be written as:

$$f(x; \kappa, \lambda) = \frac{1}{2} e^{-\frac{(x+\lambda)}{2}} \left(\frac{x}{\lambda}\right)^{\frac{\kappa}{4} - \frac{1}{2}} \left(\frac{\sqrt{\lambda x}}{2}\right)^{\frac{\kappa}{2} - 1} \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda x}{4}\right)^j}{j!\, \Gamma\left(\frac{\kappa}{2} + j\right)}$$

$$= \sum_{j=0}^{\infty} \underbrace{e^{-\frac{\lambda}{2}} \frac{\left(\lambda/2\right)^j}{j!}}_{\text{Poisson probability}} \left[\frac{1}{2} \left(\frac{x}{2}\right)^{\frac{\kappa}{2} + j - 1} \frac{e^{-x/2}}{\Gamma\left(\frac{\kappa}{2} + j\right)} \right] \text{Gamma density function.}$$

$$= \sum_{j=0}^{\infty} p_j f_j(x)\,; \textit{where } p_j \textit{ are poisson mass probabilities and } f_j(x) \textit{is the gamma PDF.}$$

Hence such a distribution is called a $randomised\ gamma\ distribution$ which is denoted as $Gamma\left(Y_1 + \frac{\kappa}{2}, \frac{1}{2}\right), Y_1 \sim Poiss\left(\frac{\lambda}{2}\right)$.

Generally, consider a mixture of densities of the form: $Gamma(Y + \nu + 1, \theta)$ with $\nu > -1$ and $\theta > 0$. Y is a $randomiser$ and is taken to be a discrete distribution. The resulting distributions are classified as $randomised\ gamma\ distributions$ of the first, second and third types. For a general discrete distribution Y the PDF of $f(y)$ (the $randomised\ gamma\ distribution$) takes the form:

$$f(y) = \sum_{j=0}^{\infty} p_j \frac{\theta}{\Gamma(\nu + j + 1)} (\theta y)^{\nu + j} e^{\theta y}, \qquad y > 0. \tag{2.40}$$

The first type has already been dealt with, see the equation after (2.39) above, although with respect to the general formula note that when $Y_1 \sim Poiss(\alpha)$ equation (2.40) becomes

$$f_1(y) = \theta \left(\frac{\theta y}{\alpha}\right)^{\frac{\nu}{2}} e^{-\alpha - \theta y} I_\nu\left(2\sqrt{\alpha \theta y}\right), \qquad y > 0.$$

The second type involves the Bessel probability distribution, i.e. $Y_2 \sim Bess(\nu, b)$ with PMF:

$$\mathbb{P}(Y_2 = j) = \frac{\left(\frac{b}{2}\right)^{2j+\nu}}{I_\nu(b) j!\, \Gamma(j+\nu+1)}, \qquad j = 0,1,2 \ldots$$

Hence the second type is defined by the mixture distribution $Gamma(Y_1 + 2Y_2 + \nu + 1, \theta)$ where $Y_1 \sim Poiss\left(\frac{a+b}{4\theta}\right)$ and $Y_2 \sim Bess\left(\nu, \frac{\sqrt{ab}}{2\theta}\right)$ and both Y_1 and Y_2 are assumed independent. Hence for the parameters $a, b, \theta > 0$ and $\nu > -1$ equation (2.40) becomes

$$f_2(y) = \theta e^{-\theta y - \frac{a+b}{4\theta}} \frac{I_\nu\left(\sqrt{ay}\right) I_\nu\left(\sqrt{by}\right)}{I_\nu\left(\frac{\sqrt{ab}}{2\theta}\right)}, \qquad y > 0.$$

The third type involves the $incomplete\ gamma\ probability\ distribution$ which is denoted by $Y_3 \sim \mathrm{I}\Gamma(\nu, \lambda)$ with both $\nu, \lambda > 0$. The probability mass function (PMF) is given by

$$\mathbb{P}(Y_3 = j) = \frac{e^{-\lambda}\lambda^{j+\nu}}{\Gamma(j+\nu+1)} \frac{\Gamma(\nu)}{\gamma(\nu, \lambda)}, \qquad y > 0.$$

where $\gamma(a, x) = \int_0^x t^{-a-1} e^{-t}\, dt$ is known as the lower incomplete gamma function. Thus a mixture probability distribution is of the third type where $Gamma(Y_3 + 1, \theta)$ with PDF

$$f_3(y) = \frac{\theta\Gamma(\nu)}{\gamma(\nu, \lambda)} \left(\frac{\theta y}{\lambda}\right)^{-\frac{\nu}{2}} e^{-\lambda - \theta y} I_\nu\left(\sqrt{4\lambda\theta y}\right), \qquad y > 0.$$

Finally, the transition density (2.38) for the CIR Model is plotted with given parameters and varying parameters α & $\tilde{\beta}$ in Figure 2.2. below. The methodology to simulate a CIR with absorption at the origin (once the process reaches zeros it stays there) uses the second type of the $randomised\ gamma\ distribution$. Note that without absorption at the origin[10] one can sample via using the first type, as was done to produce the sample for the SQB process in Figure 2.1 above.

In Chapter 3 explicit algorithms to simulate the CIR process are given and implemented in MATLAB®. These algorithms form the backbone of many of the Monte Carlo simulations performed later on.

Technical Note: Computing the Bessel function in (2.38) can be done numerically by solving the differential equation: $z^2 \frac{d^2y}{dz^2} + z\frac{dy}{dz} - (z^2 + \mu^2)y = 0$. MATLAB® does however have this function as part of its built-in list of functions. The supporting material, see Appendix A for details, should be examined for the algorithms used to produce the Figures 2.1 and 2.2.

[10] Stochastic process either never reaches zero or reflects off of it.

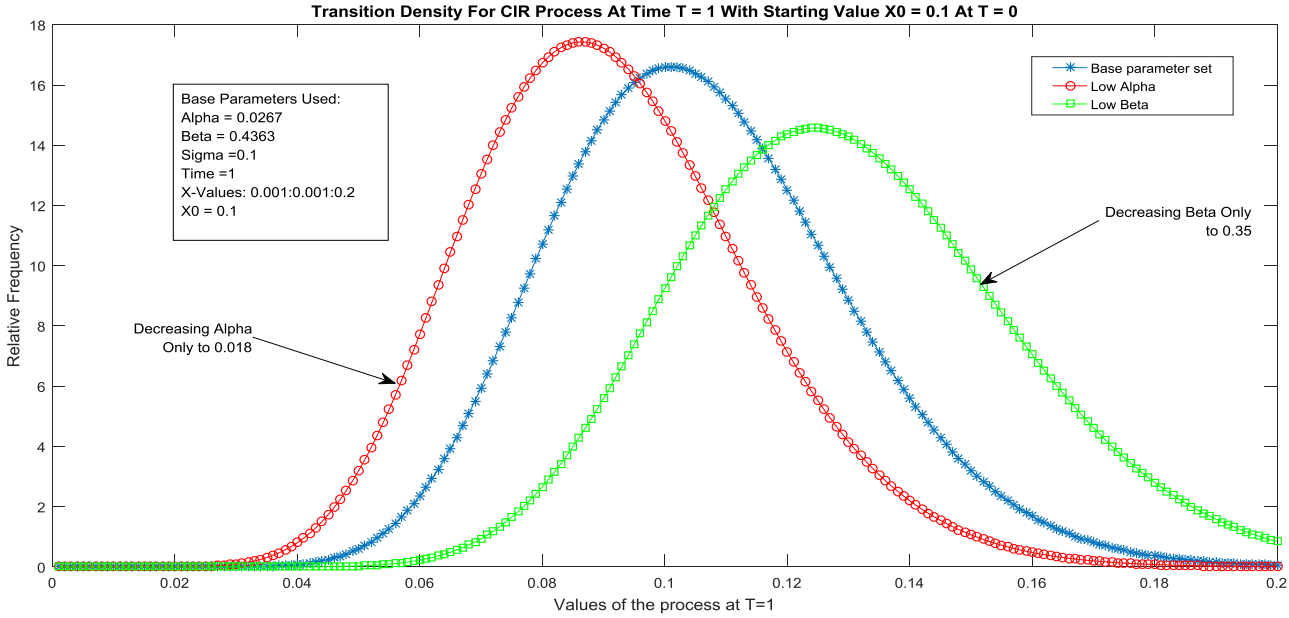

Figure 2.2: Transition Density Function for the CIR process with varying α and β parameters.

Having dealt with stochastic interest rate models, we are now in a position to consider stochastic volatility models in the next section.

2.2.5 **Stochastic volatility models part 1**

Our focus shifts now to Stochastic Volatility models and this section is presented in two parts. Part 1 considers basic models and Part 2 their extensions. However to appreciate the dynamics of the general class of Stochastic Volatility models some of their general features are discussed and these apply to models not considered here as well.

The models given here under Part 1 are basic stochastic volatility models. These models have been used extensively in pricing derivatives, and have been developed further. Although the eventual goal is to compare stochastic volatility models in this thesis, it is felt that the models here are special cases of those that follow and thus will not be considered in the analysis later on.

2.2.5.1 **Brief discussion on stochastic volatility models**

Given the discussions above, we may infer the general form of basic stochastic volatility models as:

$$\begin{aligned} dS(t) &= \mu_t S(t)\,dt + \sigma_t S(t) dW_1(t) \\ dV(t) &= \alpha\big(V(t)\big)dt + \beta\big(V(t)\big)dW_2(t) \\ \langle dW_1(t), dW_2(t)\rangle &= \rho dt \\ S(0) &= S_0 \\ v(0) &= v_0. \end{aligned} \tag{2.41}$$

Here (2.41) indicates the system of equations given. The volatility component can be denoted $\sigma_t \coloneqq f\big(V(t)\big)$ where f is a smooth, increasing function (e.g. $f(x) = \sqrt{x}$). Additionally, $u_t \coloneqq \mu(t)$, α and β are assumed to be smooth functions. The initial starting points of the asset process and the volatility process is given by the S_0 and V_0 values respectively. The <> notation indicates the quadratic covariation between each of the Brownian Motion processes. Any "co-movement" between the processes is captures by the ρ parameter, where $|\rho| \leq 1$.

Focussing on the volatility's SDE, it is natural to ask which forms of α and β should be taken? This depends of the specification of whether one wants $V(t)$ to follow either one of the following processes:

- Lognormal: $\alpha\big(V(t)\big) \equiv \alpha \times V(t)$, $\beta\big(V(t)\big) \equiv \beta \times V(t)$. Here α and β on the right-hand sides are both constant parameters.
- OU (Ornstein-Uhlenbeck): $\alpha\big(V(t)\big) \equiv \alpha\big(m - V(t)\big)$, with both α and m being constant parameters.$\beta\big(V(t)\big) = \beta$, i.e. a constant value.
- CIR: See equation (2.28) as reference.

In early literature many authors take $\rho = 0$ and a described $V(t)$ as either a lognormal, CIR or Ornstein-Uhlenbeck (OU). This assumption is unrealistic and is further explained when considering the Heston Model in the next section

Examples of these early literature models, which are not considered here, are:

- (Stein and Stein, 1991): $V(t)$ follows an OU process with $f(x) = |x|$.
- (Ball and Roma, 1994): $V(t)$ follows a CIR process with $f(x) = \sqrt{x}$.
- (Hull and White, 1987): $V(t)$ follows a Lognormal process $f(x) = \sqrt{x}$.

An important remark is that the class of models in (2.41) are single or one-factor stochastic volatility models in that there is only one stochastic volatility source. This is extended later to multi-factors.

In financial markets it is desirable to use mean-reverting models for the volatility process. In particular a simple argument demonstrates this: If volatility where strictly increasing then asset prices would either explode, oscillate wildly or reach zero. Each of these are unlikely in developed markets. Additionally if volatility reaches zero then prices would stop moving. This is unrealistic as well as after-market trading does occur on most markets and trading typically picks up as each new week begins.

The question of which of the OU or CIR processes are adequate to use in stochastic volatility modelling is largely answered by what is considered popular in literature. Undoubtingly the CIR process is mostly used, as will become apparent later on, however literature in support of the OU process can also be found. For instance, (Fouque et al, 2003) propose various fast and slow mean reversion scales for volatility modelling purposes using an OU process.

A question remains as to why the CIR process, which was introduced as a stochastic process to model interest rates is used here to model volatility? The answer lies with the fact that both stochastic interest rates and stochastic volatility processes are largely understood to be positive mean-reverting processes. Given this similarity it is then natural to use the CIR Model to model both, although we do expect different parameter sets when calibrating the CIR Model to each of the processes.

2.2.5.2 **Heston Model**

The very famous Heston Model is presented here. It can be viewed essentially as an extension of the model by (Ball and Roma, 1994) where $\rho \neq 0$. Here ρ denotes the correlation coefficient between the two Brownian Motions. (Heston, 1993) proposed the following form for the set of equations (2.41):

$$dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)dW_1(t)$$
$$dV(t) = \kappa\big(\Theta - V(t)\big)dt + \nu\sqrt{V(t)}dW_2(t)$$
$$\langle dW_1(t), dW_2(t)\rangle = \rho dt$$
$$S(0) = S_0$$

$$V(0) = v_0, \tag{2.42}$$

were κ denotes the mean reversion parameter, Θ the long term mean of the volatility process and ξ is a constant parameter value that indicates the volatility of the volatility process.

Here ρ denotes the correlation coefficient between the two Brownian Motions

Interestingly, when comparing the second equation in (2.42) to the equations of the SQB process (equations (2.35) & (2.36)) it is worthwhile to note that if $2\kappa\Theta \geq v^2$ (which is known as the Feller condition/property) the process $v(t)$ can never reach zero. Otherwise it can reach zero and zero can either be a reflecting or absorbing boundary condition. Note that (2.42) is given in terms of the real world measure and the issue of risk neutral dynamics will be addressed in Chapter 3.

The parameters in (2.42) can be seen as improving on the classical BS Model as follows:

- Volatility of the process is stochastic and mean-reverting. This is consistent with most market stocks, i.e. the volatility is mean-reverting otherwise stocks would either not move or explode towards infinity. Naturally there are periods when stocks move more (e.g. following a news announcement, market crash, etc.) and move less (e.g. after-hours trading, during weekends, etc.) and both these correspond to a latent stochastic volatility process. Thus the volatility follows a CIR-process (see equation (2.28)).
- ν controls the volatility of the variance process and consequently controls the level of kurtosis for the distribution of the asset return process $S(t)$.
- ρ controls the asymmetry of the distribution of the asset return process $S(t)$. Additionally ρ allows for an important feature of asset data: the direction of price movement is correlated to the underlying volatility (typically $\rho < 0$), i.e. prices going down (up) implies higher (lower) volatility. (The opposite typically happens in commodity markets, i.e. $\rho > 0$).

The characteristic function of the Heston Model is given by (Kientiz and Wetterau, 2012) (with minor alterations and corrections by the Author as per the work of (Christoffersen, Heston and Jacobs, 2009)):

$$\varphi_H(u, X(t), t, T) = \exp(A_H(u,t,T) + B_H(u,t,T)V(t) + iuX(t)) \tag{2.43}$$

where $X(t) = \ln(S(t))$

$$A_H(u,t,T) = \frac{\kappa\Theta}{\nu^2}\left((\kappa - \rho\nu ui - D)(T-t) - 2\log\left(\frac{G\,exp\big(-D(T-t)\big)}{G-1} - 1\right)\right)$$

$$B_H(u,t,T) = \frac{\kappa - \rho\nu ui - D}{\nu^2}\left(\frac{1 - \exp\big(-D(T-t)\big)}{1 - G\exp\big(-D(T-t)\big)}\right)$$

$$G = \frac{\kappa - \rho\nu ui - D}{\kappa - \rho\nu ui + D}$$

$$D = \sqrt{(\kappa - \rho\nu ui)^2 + u(i+u)\nu^2}.$$

A sample path for the Heston Model was shown in Chapter 1. The Heston Model, while important, is not always able to capture all of the complicated dynamics of the implied volatility surface (Christoffersen, Heston and Jacobs, 2009). Hence an improvement on this model is suggested in subsection 2.2.6.1.

2.2.5.3 **SABR Model**:

The SABR Model, introduced by (Hagar et al, 2002), is specified by the following dynamics for the

forward price $\hat{F}$:

$$\begin{aligned} d\hat{F}_t &= \hat{\alpha}_t \hat{F}_t^{\beta} dW_1(t), \qquad \hat{F}(0) = f \\ d\hat{\alpha}_t &= v\hat{\alpha}_t dW_2(t), \qquad \hat{\alpha}(0) = \alpha \\ \langle dW_1(t), dW_2(t) \rangle &= \rho dt. \end{aligned} \tag{2.44}$$

Thus the $C(t,F)$ coefficient in (2.18) is chosen as $\hat{\alpha}\hat{F}^{\beta}$ and consequently SABR stands for "stochastic-$\alpha\beta\rho$" and the probability measure underlying (2.44) is the usual T-forward martingale measure[11], denoted as $\mathbb{Q}^{\mathbb{T}}$. The SABR Model is investigated in quite some depth in (Hagar et al, 2002) and some of the more technical results are mentioned here.

The SABR Model produces skew and smile dynamics for the underlying asset as will be demonstrated. However, the SABR Model will not be used for further analysis as our focus will tend to be on more recent models in literature. Hence a brief, but technical, summary of the literature regarding this model is given here.

As a starting point, note that the SABR Model can be used to price Call options (and related derivatives) using the BS formula in (2.40) below, but with the implied volatility replaced by

$$\begin{aligned} \sigma_{SABR}(K,f) \approx & \frac{\alpha}{(fK)^{\frac{1-\beta}{2}}\left\{1 + \left(\frac{(1-\beta)^2}{24}\log^2\frac{f}{K}\right) + \left(\frac{(1-\beta)^4}{1920}\log^4\frac{f}{K}\right) + \cdots\right\}} \times \left(\frac{z}{x(z)}\right) \\ & \times \left\{1 + \left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(fK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(fK)^{\frac{1-\beta}{2}}} + \frac{2-3\rho^2}{24}\nu^2\right] t_{ex} + \cdots\right\}, \end{aligned} \tag{2.45}$$

$$where\ z = \frac{\nu}{\alpha}(fK)^{\frac{1-\beta}{2}}\log\frac{f}{K}\ and\ x(z) = \log\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}.$$

Note that K represents the strike price of the Call option in (2.45).
The corresponding BS formula for a call option is:

$$V_{call} = D(t_{set})\{f\mathcal{N}(d_1) - K\mathcal{N}(d_2)\}. \tag{2.46}$$

Here $d_1 = \frac{\log\frac{f}{K} + 0.5\sigma_{SABR}^2 t_{ex}}{\sigma_{SABR}\sqrt{t_{ex}}}$ and $d_2 = \frac{\log\frac{f}{K} - 0.5\sigma_{SABR}^2 t_{ex}}{\sigma_{SABR}\sqrt{t_{ex}}}$ and recall from section section 2.2.3.3 that t_{ex} denotes the time at which the Call option is exercised with settlement date t_{set}.

Thus a characteristic function for this model is not needed in order to price derivatives. An approximate joint density function for (2.45) is given here and is due to (Wu, 2010). This joint density is particularly useful in valuing both exotic equity and interest rate derivatives, particularly forward starting call options (e.g. on Libor rates). Only the final results are given since the derivation is too lengthy to summarise here.

Summary of Wu's Results:
Denote by F and A the time T-forward values of $\hat{\alpha}_t$ and $\hat{F}_t$, i.e. $\hat{\alpha}_T = A$ and $\hat{F}_T = F$. Additionally let $s = T - t$ then the joint transition density function of $\hat{\alpha}_t$ and $\hat{F}_t$, given by $p(T-t, f, \alpha, F, A)$, admits the series expansion:

[11] The risk neutral measure we typically use here for derivative pricing uses the risk-free rate as a unit to discount cashflows. The T-forward measure is simply used to refer to using zero-coupon bonds as unit to discount cashflows. The two measures are related by the Radon-Nykodim derivative.

$$p_n(T-t,f,\alpha;F,A)=\frac{1}{\nu TF^{\beta}A^2}\sum_{i=1}^{n}\left(\nu\sqrt{T}\right)^k\hat{p}_k\left(\frac{T-t}{T},\frac{f^{1-\beta}-F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}},\frac{\ln\left(\frac{\alpha}{A}\right)}{\nu\sqrt{T}}\right). \quad (2.47)$$

An explicit formula for this expansion up to a second order is given here and will be used from here on:

$$p_2(\tau,f,\alpha;F,A)=\frac{1}{\nu TF^{\beta}A^2}\left(1+\frac{\nu\sqrt{T}}{2(\rho^2-1)}\left(a_{11}+\frac{a_{10}}{\tau}\right)+\frac{\nu^2T}{24(1-\rho^2)^2}\left(a_{23}\tau+a_{22}+\frac{a_{21}}{\tau}+\frac{a_{20}}{\tau^2}\right)\right)$$
$$\times\frac{\exp\left(-\frac{u^2-2\rho uV+V^2}{2\tau(1-\rho^2)}\right)}{\left(2\pi\tau\sqrt{1-\rho^2}\right)}. \quad (2.48)$$

The parameters in (2.48) are given by:

$\tau=T-t$

$u=\frac{\left(f^{1-\beta}-F^{1-\beta}\right)}{\alpha(1-\beta)\sqrt{T}}$

$V=\frac{\log\left(\frac{\alpha}{A}\right)}{\nu\sqrt{T}}$ (Note that this differs from Wu's notation where it is easy to confuse ν (nu) with υ $(upsilon)$.)

$a_{11}=-\beta F^{\beta-1}A\nu^{-1}(u-\rho V)$

$a_{10}=u^2V-u\rho V^2$

$a_{23}=\rho^4(20-6b)+\rho^3(-12a)+\rho^2(-28+3a^2+12b)+\rho(12a)+(8-3a^2-6b)$

$a_{22}=u^2[3a^2-12a\rho+6b(\rho^2-1)+2\rho^2+10]$
$-2uV[\rho^3(2+3b)+\rho^2(-9a+3c)+\rho(10+3a^2-3b)-(3a+3c)]$
$+V^2[(2+3(a-2\rho)^2)\rho^2+6c\rho(\rho^2-1)-2]$

$a_{21}=u^4+V^4\rho^2+u^3V(8\rho-6a)+uV^3(8\rho^3-6a\rho^2)+u^2V^2(-14\rho^2+12a\rho-4)$

$a_{20}=3u^4V^2-6u^3V^3\rho+3u^2V^4\rho^2$

$a=2\rho+\beta F^{\beta-1}A\nu^{-1}$

$b=2+\beta(1-\beta)F^{2(\beta-1)}A^2\nu^{-2}$

$c=\beta F^{\beta-1}A\nu^{-1}$.

Finally, we give a brief discussion of the underlying dynamics of the SABR Model. Equation (2.45) can be approximated by the following, (Hagar et al, 2002):

$$\sigma_{SABR}(K,f)\approx\frac{\alpha}{f^{1-\beta}}\left\{1-\frac{1}{2}(1-\beta-\rho\lambda)\log\left(\frac{K}{f}\right)+\frac{1}{12}\left[(1-\beta)^2+(2-3\rho^2)\lambda^2\right]\log^2\frac{K}{f}\right\}. \quad (2.49)$$

Here $\lambda=\frac{\nu}{\alpha}f^{1-\beta}$ and this is a measure of the relative ratio of the volatility of the volatility (volvol from now on), ν, to the local volatility $\frac{\alpha}{f^{1-\beta}}$. Consider the implied volatility for at-the-money $(K=f)$ options in (2.49). This implies that the approximation takes the form $\frac{\alpha}{f^{1-\beta}}$ and in the special case that $\beta=0$ the implied volatility slopes downwards (if we trace it across all strikes $K=f$) and is flat at level α if $\beta=1$.

In fact, the parameter β in the system of equations (2.44) corresponds to the same parameter as in the CEV Model (i.e. $\beta_{SABR}=\beta_{\text{CEV}}+1$ under the parametrisation given by (2.15)). Thus if $\beta_{SABR}=1$ (which implies $\beta_{CEV}=0$) then from (2.15) we have the BS Model which has a flat implied volatility slope (constant). If $\beta_{SABR}=0$ then from (2.15) we have the equivalent Bachelier Model (which assumes stock prices follow a Gaussian process) with its downwards sloping implied volatility. This is consistent with the discussion of previous paragraph where mentioned that a $\beta=1$ gives a flat

implied volatility surface. Clearly the SABR Model is a substantial improvement on the BS[12] (and its predecessor, i.e. Bachelier framework) model as the implied volatility surface is modelled very well. We explain this in more detail below.

The function $\sigma_{SABR}(f,f) = \frac{\alpha}{f^{1-\beta}}$ is referred to as the "backbone" of the implied volatility. In order to generate nuanced features of the volatility skew or the smile the additional factors of (2.49) (for $k \neq f$) can be interpreted as:

Downwards sloping skew:$-\frac{1}{2}(1-\beta-\rho\lambda)\log\left(\frac{K}{f}\right)$, which can be split into the *beta* skew $-\frac{1}{2}(1-\beta)\log\left(\frac{K}{f}\right)$and the *vanna* skew $\frac{1}{2}(\rho\lambda)\log\left(\frac{K}{f}\right)$. The former is downwards sloping due to $0 \leq \beta \leq 1$ and the latter represents the skew caused by the negative correlation between the underlying and its volatility.

Smile: $\frac{1}{12}[(1-\beta)^2]\log^2\frac{K}{f}$ is a quadratic term generating a smile.

Taken together, it is easy to see that the SARB Model describes the implied volatility surface quite well and how the generated surface is related to its underlying parameters. In practice there are unfortunately other factors that the SARB Model does not describe and these will be addressed by considering the models in subsections 2.2.6.1-2.2.6.4.

2.2.5.4 **3/2 Model**

The 3/2 Model was introduced by (Heston, 1997) and its dynamics are similar to (2.42). Here the version presented in (Baldeaux, 2012) is given (risk neutral dynamics):

$$\frac{dS(t)}{S(t)} = rdt + \sqrt{V(t)}\rho dW_1(t) + \sqrt{V(t)}\sqrt{1-\rho^2}dW_2(t)$$
$$dV(t) = \kappa V(t)\big(\theta - V(t)\big)dt + \epsilon\big(V(t)\big)^{3/2}dW_1(t). \qquad (2.50)$$

Note that (2.50) is nothing other than the inverted variance process given in the Heston Model. In other words if $V^*(t)$ is some CIR process then $V(t)$ in (2.50) is defined by $V(t) = \frac{1}{V^*(t)}$. It should be mentioned that the inverted CIR process captures more extreme volatility events than the CIR process. (Note that parameters in (2.50) under the CIR process (for $V(t)$) would be different, e.g. $\alpha = \kappa + \frac{1}{2}\epsilon^2$).

An alternative version of (2.50) is given by (Goard, 2011), who considers pricing derivative contracts for VIX[13] options, variance swaps and volatility swaps. Goard's version, which is (2.51) below, admits an analytical pricing formula. This version builds on the 3/2 Model as follows:

$$dV_t = \Big(c(t)V(t) + c_3\big(V(t)\big)^2\Big)dt + kV^{\frac{3}{2}}(t)dW_1(t). \qquad (2.51)$$

[12] The BS Model can be improved upon using either or both stochastic volatility and stochastic interest rates. Models are presented here that show exactly how these improvements have been undertaken in literature. The SARB model is attractive in that we obtain very detailed dynamics of the implied volatility surface. The models following the SARB model are presented as being attractive in modelling real world asset dynamics and fitting the implied volatility surface well as a consequence. Therefore we will not use the SARB model any further.

[13] VIX options are options written on the implied volatility index (VIX) that is constructed from regular options written on the S&P500.

Here c_3 and k are constants and $c(t)$ is some arbitrary function of time.

The volatility diffusion process (2.51) is an extension of (2.50) and can allow for greater flexibility in modelling and the implementation of (2.51) is a trivial extension of (2.50). The 3/2 Model as given in (2.50) is a special case of the 4/2 Model, introduced in subsection 2.4.1, under the list of advanced models. Thus we end this subsection and part 1 of the stochastic volatility models and move onto part 2 below.

2.2.6 **Stochastic volatility models part 2**

In this section further models are examined. They are essentially generalisations or extensions of the previous models and two of these models will be considered for further analysis in Chapters 3 and 4.

2.2.6.1 **Double Heston Model**

There are many improvements on the Heston Model. Here the Double Heston Model by (Christoffersen, Heston and Jacobs 2009) is chosen as a natural extension. One other notable model that the reader can pursue is based on the Wishart process that replaces the CIR process in the Heston Model (see for instance (Bru, 1991) and also (Gourieroux et al, 2009)). In the course of our considerations here the Wishart process will not be used.

The argument offered by (Christoffersen, Heston and Jacobs 2009) is simply that single factor models (such as the Heston and SABR models above) do not accurately capture the fluctuating level and slope of the implied volatility smile over time. The model is specified exactly as is given in the article by the authors mentioned (under risk neutral dynamics $\mathbb{Q}$):

$$\begin{aligned} dS(t) &= rS(t)dt + \sqrt{V_1(t)}S(t)dW_Q^1(t) + \sqrt{V_2(t)}S(t)dW_Q^2(t) \\ dV_1(t) &= \big(a_1 - b_1V_1(t)\big)dt + \sigma_1\sqrt{V_1(t)}dW_Q^3(t) \\ dV_2(t) &= \big(a_2 - b_2V_2(t)\big)dt + \sigma_2\sqrt{V_2(t)}dW_Q^4(t). \end{aligned} \tag{2.52}$$

The correlation structure between each of the four Brownian motions can be summarised by the following correlation matrix (each column and row represents the superscript of the respective Brownian Motion, i.e. column 2 and row 2 represent $dW_Q^2(t)$) :

$$\begin{bmatrix} 1 & 0 & \rho_1 & 0 \\ 0 & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & 0 \\ 0 & \rho_2 & 0 & 1 \end{bmatrix}$$

Additionally note that $Var\left(\frac{dS(t)}{S(t)}\right) = \big(V_1(t) + V_2(t)\big)dt = V(t)dt$ since the $dW_Q^3(t)$ and $dW_Q^4(t)$ are uncorrelated. In other words the variance of the stock return is the sum of the two variance factors. Following the Heston Model, the characteristic function of (2.52) is given similarly as (remembering $X(t) = \ln(S_t)$)

$$\begin{aligned} &E_t[\exp(iuX(t+\tau)] = \big(S(t)\big)^{iu} f(V_1(t), V_2(t), \tau, u) \\ &f(V_1(t), V_2(t), \tau, u) = \exp\big(A(\tau,u) + B_1(\tau,u)V_1(t) + B_2(\tau,u)V_2(t)\big) \\ &A(\tau,u) = ruri\tau + \frac{a_1}{\sigma_1^2}\left[(b_1 - \rho_1\sigma_1 ui + d_1)\tau - 2\ln\left[\frac{1 - g_1\exp(d_1\tau)}{1 - g_1}\right]\right] \end{aligned} \tag{2.53}$$

$$+\frac{a_2}{\sigma_2^2}\left[(b_2-\rho_2\sigma_2 ui+d_2)\tau-2\ln\left[\frac{1-g_2\exp(d_2\tau)}{1-g_2}\right]\right]$$

$$B_j(\tau,u)=\frac{b_j-\rho_j\sigma_j ui+d_j}{\sigma_j^2}\left[\frac{1-\exp(d_j\tau)}{1-g_j\exp(d_j\tau)}\right]$$

$$g_j=\frac{b_j-\rho_j\sigma_j ui+d_j}{b_j-\rho_j\sigma_j ui-d_j}$$

$$d_j=\sqrt{\left(\rho_j\sigma_j ui-b_j\right)^2+\sigma_j^2(ui+u^2)}.$$

Here $j=1\ or\ 2$.

The correlation between the asset returns and both variances processes are given by[14]:

$$corr\left[\frac{dS(t)}{S(t)},dV(t)\right]=\frac{\sigma_1\rho_1V_1(t)+\sigma_2\rho_2V_2(t)}{\sqrt{\sigma_1^2V_1(t)+\sigma_2^2V_2(t)}\sqrt{V_1^2(t)+V_2^2(t)}}dt. \tag{2.54}$$

Equation (2.54) essentially gives a stochastic correlation structure between the asset price process and the variance processes. It is argued that this enables fluctuations in the implied volatility's skewness and level over time and thus presents a genuine improvement over the Heston Model.

The above characteristic function in (2.53) has an equivalent form when one inverts all the plus and minus signs in front of the d_js. At first this may seem harmless, but as pointed out by (Albrecher et al, 2006), in the important article titled "The Little Heston Trap", this slight change makes a considerable difference when considering numerical stability. The reason is that the $d_j s$ represent complex roots and that the version in (2.53) crosses the negative axis (branch cut) in the complex plane causing numerical instabilities. Essentially for long times to maturity this subtle effect will give inaccurate results. This is avoided by inverting all the plus and minus signs in front of the $d_j s$.

Although the comments apply strictly to the Heston Model, it is just as valid for the Double Heston Model, seeing as the underlying mathematics is similar, and thus the reader should note that in numerical work (2.53) should be used with the signs changed in front of the d_js. This consideration has been applied in the numerical work done in Chapters 3 and 4.

2.2.6.2 **SABR-Hull-White Model**

The SABR-Hull-White Model (denoted SABR-HW from now on) is an extension of the SABR Model by combining both the SABR and Hull-White models. This approach was first formulated by (Chen and Grzelak, 2012) and part of the motivation for this model is to allow for equity-interest rate hybrid derivative products that have been introduced in the market. The risk neutral dynamics for this model are:

$$\frac{dS(t)}{S(t)}=r(t)dt+\Sigma(t)S^{\beta-1}(t)dW_x(t),\qquad S(0)>0$$

$$dr(t)=\lambda\big(\theta(t)-r(t)\big)dt+\ \eta dW_r(t),\qquad r(0)>0$$

$$\Sigma(t)=g(t)k(t)$$

[14] If the second volatility factor $V_2(t)$ is taken out then then the expression reduces to ρ_1 which is simply the correlation between the asset price and the volatility parameter. For any stochastic volatility model that does not have more than one stochastic volatility factor it is immediately obvious that the correlation between asset price and volatility will be constant, e.g. equal to ρ.

$$
\begin{aligned}
&dk(t) = h(t)k(t)dW_{\Sigma}(t), \qquad k(0) = 1\\
&dW_x(t)dW_{\Sigma}(t) = \rho_{x,\sigma}dt\\
&dW_x(t)dW_r(t) = \rho_{x,r}dt\\
&dW_r(t)dW_{\Sigma}(t) = 0.
\end{aligned} \tag{2.55}
$$

Here a mix of constant parameters and appropriate time dependent functions are present. As with the Hull-White Model in (2.23), $\theta(t)$ can be chosen to match the initial term structure of interest rates. Following the work of (Brigo and Mercurio, 2007), the functions $h(t)$ and $g(t)$ can be given by:

$$h(t) = (a_1 + b_1 t)e^{-c_1 t} + d_1. \tag{2.56}$$

$$g(t) = (a_2 + b_2 t)e^{-c_2 t} + d_2. \tag{2.57}$$

The determination of the coefficients, i.e. recommended values, in (2.56) and (2.57) are given in the article by (Chen and Grzelak, 2012).

The keen reader may ask why the Hull-White Model is chosen to model $r(t)$ instead of the CIR Model given in (2.28). In the same article by (Chen and Grzelak, 2012) it is argued that the CIR Model becomes intractable when there is a nonzero correlation present in the equity process (SABR driven part) with the interest rate process. However, the CIR Model admits non-negative interest rates whereas the Hull-White can admit negative interest rates.

Thus in choosing between these processes there is a trade-off between better analytical tractability vs having non-negative interest rates. The Author's opinion is that the interest rate environment should dictate the choice of interest rate model. For instance in South Africa, accurate as at 19 May 2016, the prevailing benchmark repo rate is 7% and for the last 20 years has reached a minimum of 5% (July 2012) and a maximum of 23.99% (June 1998).

In contrast in the United States the Federal Funds rate is currently 0.5% (19 May 2016) and the interest rates have been depressed since 2008. Additionally past 1990 the Federal Funds rate has never been higher than 10%. Hence the Hull-White Model may be suitable for valuing derivative products in the US, but not in South Africa.

2.2.6.3 **Heston-Hull-White models**

Consider the following general model stochastic volatility with stochastic interest rate model:

$$dS(t) = r(t)S(t) + \big(V(t)\big)^p S(t)dW_S(t). \tag{2.58}$$

$$dV(t) = \kappa\big(\Theta - V(t)\big)dt + \gamma V^{1-p}(t)dW_V(t). \tag{2.59}$$

$$dr(t) = \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta dW_r(t). \tag{2.60}$$

The covariance matrix between all the Brownian Motions is given by:

$$\Sigma = \begin{bmatrix} dW_S(t)dW_S(t) & dW_S(t)dW_V(t) & dW_S(t)dW_r(t) \\ dW_V(t)dW_S(t) & dW_V(t)dW_V(t) & dW_V(t)dW_r(t) \\ dW_r(t)dW_S(t) & dW_r(t)dW_V(t) & dW_r(t)dW_r(t) \end{bmatrix} = \begin{bmatrix} 1 & \rho_{S,V} & \rho_{S,r} \\ \rho_{V,S} & 1 & \rho_{V,r} \\ \rho_{r,S} & \rho_{r,V} & 1 \end{bmatrix} dt.$$

This general model can be found in the article by (Grzelak et al, 2012) (with minor differences in notation). It should be noted that if $r(t)$ is kept constant and if $p = 0.5$ then the Heston Model is recovered. There are models that can be recovered by varying the parameter for p, keeping $r(t)$ or $V(t)$ constant or stochastic, dropping correlations between Brownian motions, etc. Some of the other stochastic volatility models in literature, such as the Schöbel-Zhu, are not considered here but do fall under the framework given here. Finally the famous BS Model is simply given by $p = 0.5$ and with both $r(t)$ or $V(t)$ held constant and the covariance matrix collapsing.

The Heston-Hull-White (HHW) Model is a special case of the above and is given by (in the notation of (Kienitz and Wetterau, 2012)):

$$dS(t) = r(t)S(t) + \sqrt{V(t)}S(t)dW_S(t). \tag{2.61}$$

$$dV(t) = \kappa\big(\Theta_V - V(t)\big)dt + \nu\sqrt{V(t)}dW_V(t). \tag{2.62}$$

$$dr(t) = \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta dW_r(t). \tag{2.63}$$

The covariance matrix is given as above. Note that this model is essentially a mix of the Hull-White and Heston models given earlier. There are problems with this model however. The characteristic function of the HHW is not known to have a closed form (Grzelak et al, 2011) and it is not clear how to infer the correlation coefficient between the variance and the interest rate process.

The first solution to these problems is given in (Grzelak et al, 2012) and (Grzelak et al, 2011) gives two other more involved solutions. In Chapter 3, subsection 3.1.1, the detail of the existence of a closed form characteristic function for a system of stochastic differential equations, such as (2.61) to (2.63), is explored in more depth. Essentially the solutions presented below show how to obtain the characteristic function when the general methods for guaranteeing a closed form solution, as presented in Chapter 3, fail.

In all the cases the notation has been reworked to achieve consistency across the presentation of the various models.

Solution 1:

Let the correlations $\rho_{r,S}$ and $\rho_{r,V}$ both be zero and hence movements in the interest rate and stochastic volatility processes are uncorrelated with the underlying asset's movement. Additionally this also implies that the volatility and interest rate processes run independently. The rest of the model is specified by (2.61-2.63). Additionally let $x(t) = \log\big(S(t)\big)$, and denote current time by t with $t < T$ and hence the HHW model admits the following (discounted) characteristic function:

$$\phi_{HHW}(u,t,T) = \varphi_H(u,x(t),t,T)\exp\big(B_r(u,\tau)r(t) + A_{HW}(u,\tau)\big). \tag{2.64}$$

Here the auxiliary functions are defined by:

$$\varphi_H(u,t,T) = \exp(A_H(u,t,T) + B_v(u,t,T)v(t) + iux(t))$$

$$A_H(u,t,T) = \frac{\kappa\Theta_V}{\nu^2}\left(\big(\kappa - \rho_{x,V}\nu ui - D\big)(\tau) - 2\log\left(\frac{G\,exp\big(-D(\tau)\big)}{G-1} - 1\right)\right)$$

$$B_H(u,t,T) = \frac{\kappa - \rho_{x,V}\nu ui - D}{\nu^2}\left(\frac{1-\exp\big(-D(\tau)\big)}{1-G\exp\big(-D(\tau)\big)}\right)$$

$$A_{HW}(u,t,T) = \frac{-1}{2}\left(\frac{\eta^2\tau}{\lambda^2}(iu-1) - \eta^2 B_r(u,\tau)\right) + \Theta_r(u,\tau)$$

$$B_r(u,\tau) = (iu-1)\lambda^{-1}\big(1-e^{-\lambda\tau}\big)$$

$$G = \frac{\kappa - \rho\nu ui - D}{\kappa - \rho\nu ui + D}$$

$$D = \sqrt{(\kappa - \rho\nu ui)^2 + u(i+u)\nu^2}$$

$$\Theta_{\mathrm{r}}(u,\tau) = (1-iu)\left(\log\left(\frac{P(0,t)}{P(0,T)}\right) + \frac{\eta^2}{2\lambda^2}\left(\tau + \frac{2\left(e^{-\lambda T} - e^{-\lambda t}\right)}{\lambda} - \frac{e^{-2\lambda T} - e^{-2\lambda t}}{2\lambda}\right)\right)$$

$$\tau = T - t.$$

Note: Θ_V (constant) is the long term mean of the variance process and $\Theta_r(u,\tau)$ is used to match the term structure of interest rates. The terminal condition for the characteristic function is given by $\phi_{HHW}(u,T,T) = \exp\big(iux(T)\big)$.

Solution 2: Deterministic Approach (H1-HW)

It has been proposed that for many hybrid derivatives the correlation coefficient between the asset process and the interest rate process should be non-zero (Hunter, 2006). Hence solution 1 given may not be appropriate. The following is a summary of the methodology used in (Grzelak et al, 2011) to obtain an approximation for the characteristic function. This model is called the H1-HW Model.

Consider the following system of equations:

$$\frac{dS(t)}{S(t)} = r(t)dt + \sqrt{V(t)}dW_S(t) + \left(\rho_{S,V} - \bar{\rho}_{S,V}\right)\sqrt{V(t)}dW_V(t) + \rho_{S,r}\sqrt{V(t)}dW_r(t). \tag{2.65}$$

$$dV(t) = \kappa\big(\Theta_{\mathrm{V}} - V(t)\big)dt + \nu\sqrt{V(t)}dW_V(t). \tag{2.66}$$

$$dr(t) = \lambda\big(\Theta_{\mathrm{r}}(t) - r(t)\big)dt + \eta dW_r(t). \tag{2.67}$$

$$\langle dW_S(t), dW_V(t)\rangle = \bar{\rho}_{S,V}$$

$$\bar{\rho}_{S,V} = \rho_{S,V}^2 + \rho_{S,r}^2.$$

The correlation between the asset process and the interest rate process is incorporated in $\bar{\rho}_{S,V}$ and hence $\langle dW_S(t), dW_r(t)\rangle = 0$ and also $\langle dW_V(t), dW_r(t)\rangle = 0$.

The above modified model still does not admit a characteristic function (i.e. the model is not affine, see subsection 3.1.1 in Chapter 3) and hence the following (delta-method) approximation (see (Oehlert, 1992)) is made to the third term in (2.65):

$$\rho_{S,r}\sqrt{V(t)} \approx \rho_{S,r}E\left[\sqrt{V(t)}\right]. \tag{2.68}$$

$E\left[\sqrt{V(t)}\right]$ is now then further approximated by the following function:

$$\widetilde{\Lambda}(t) = a + be^{-ct}. \tag{2.69}$$

$$a = \sqrt{\Theta_V - \frac{\nu^2}{8\kappa}}, \qquad b = \sqrt{V(0)} - a, \qquad c = -\log\big(b^{-1}(\Lambda(1) - a)\big).$$

The function $\Lambda(t)$ is given by:

$$\Lambda(t) = \sqrt{c(t)(\lambda(t)-1) + c(t)d + \frac{c(t)d}{2\big(d+\lambda(t)\big)}}. \tag{2.70}$$

and the extra functions in (2.70) are given by:

$$c(t) = \frac{1}{4\kappa}\nu^2(1-e^{-\kappa t}), \qquad d = \frac{4\kappa\Theta_v}{\nu^2}, \qquad \lambda(t) = \frac{4\kappa V(0)e^{-\kappa t}}{\nu^2(1-e^{\kappa t})}.$$

Hence the characteristic function can be approximated by the following, remembering that $x(t) =$

$\ln\big(S(t)\big)$:

$$\phi_{H1-HW}(u,t,T) \approx \exp\big(A(u,\tau) + B_x(u,\tau)x(t) + B_V(u,\tau)V(t) + B_r(u,\tau)r(t)\big) \tag{2.71}$$

$$B_x(u,\tau) = iu$$

$$B_r(u,\tau) = \frac{iu-1}{\lambda}\big(1-e^{-\lambda\tau}\big)$$

$$B_V(u,\tau) = \frac{(1-e^{-d\tau})}{v^2(1-ge^{-d\tau})}\big(\kappa - v\bar{\rho}_{X,V}iu - d\big)$$

$$A(u,\tau) = \lambda\Theta_r I_1(\tau) + \kappa\Theta_V I_2(\tau) + \frac{1}{2}\eta^2 I_3(\tau) + \eta\rho_{X,r}I_4(\tau).$$

Here the additional functions are given:

$$d = \sqrt{\big(v\rho_{X,V}iu-\kappa\big)^2 - v^2\big(iu(iu-1)\big)}$$

$$g = \frac{\kappa - v\rho_{X,V}iu - d}{\kappa - v\rho_{X,V}iu + d}$$

$$I_1(\tau) = \frac{1}{\lambda}(iu-1)\left(\tau + \frac{e^{-\lambda\tau}-1}{\lambda}\right)$$

$$I_2(\tau) = \frac{\tau}{v^2}\big(\kappa - v\rho_{X,V}iu - d\big) - \frac{2}{v^2}\log\left(\frac{(1-ge^{-d\tau})}{1-g}\right)$$

$$I_3(\tau) = \frac{1}{2\lambda^3}(i+u)^2\big(3 + e^{-2\lambda\tau} - 4e^{-\lambda\tau} - 2\lambda\tau\big)$$

Before giving I_4, it should be noted that the theory given in Grzelak et al (2011) shows that the general form of I_4 is given by:

$$I_4(\tau) = iu\int_0^\tau E\left[\sqrt{V(T-s)}\right]B_r(u,s)ds \tag{2.72}$$

$$= -\frac{1}{\lambda}(iu+u^2)\int_0^\tau E\left[\sqrt{V(T-s)}\right]\big(1-e^{-\lambda s}\big)ds.$$

The approximations in (2.69-2.70) determine a particular form of $I_4(\tau)$. This is called a deterministic approximation of (2.72):

$$I_4(\tau) = -\frac{1}{\lambda}(iu+u^2)\left[\frac{b}{c}(e^{-ct}-e^{-cT}) + a\tau + \frac{a}{\lambda}\big(e^{-\lambda\tau}-1\big) + \frac{b}{c-\lambda}e^{-cT}\big(1-e^{\tau(\lambda-c)}\big)\right]. \tag{2.73}$$

Here the formulae from (2.68-2.70) are used to approximate the quantity $E\left[\sqrt{V(T-s)}\right]$. Finally note that the terminal condition for the characteristic function is given by $\phi_{H1-HW}(u,T,T) = \exp\big(iux(T)\big)$.

Solution 3: Stochastic Approximation (the H2-HW Model)

The stochastic approximation is an improvement over the deterministic one although the price to pay is more expensive computational time. The resulting model is called the H2-HW Model and is more involved than the first and so some preliminary steps are required:

Let $V(t)$ be a CIR process given in (2.65), i.e. the variance process follows a CIR process. (Dufresne, 2001) investigated the CIR process in great detail and some of the results are given here:
For any $t > 0$:

$$E\left[\sqrt{V(t)}\right] = \sqrt{2c(t)}\, e^{-\lambda(t)/2} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\lambda(t)}{2}\right)^k \frac{\Gamma\left(\frac{1+d}{2}+k\right)}{\Gamma\left(\frac{d}{2}+k\right)}. \tag{2.74}$$

$$\mathbb{V}\left[\sqrt{V(t)}\right] = c(t)\big(d+\lambda(t)\big) - 2c(t)e^{-\lambda(t)}\left(\sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\lambda(t)}{2}\right)^k \frac{\Gamma\left(\frac{1+d}{2}+k\right)}{\Gamma\left(\frac{d}{2}+k\right)}\right)^2. \tag{2.75}$$

$$c(t) = \frac{1}{4\kappa}\nu^2(1-e^{-\kappa t}), \qquad d = \frac{4\kappa\Theta_V}{\nu^2}, \qquad \lambda(t) = \frac{4\kappa\lambda(0)e^{-\kappa t}}{\nu^2(1-e^{\kappa t})}. \tag{2.76}$$

For clarity, the Gamma function is given by:

$$\Gamma(k) = \int_0^{\infty} t^{k-1}e^{-t}dt.$$

Keeping these results in mind, it can be shown that the quantity $\sqrt{V(t)}$ in (2.68) can be approximated by a normal distribution. (This approximation is needed as Itô's lemma in (2.8) cannot be applied to $\sqrt{V(t)}$ since the square root function is not twice differentiable at the origin.). For clarity, the deterministic approximation implies the following distribution of $\sqrt{V(t)}$:

$$\sqrt{V(t)} \;\approx\; \mathcal{N}\left(\sqrt{c(t)(\lambda(t)-1) + c(t)d + \frac{c(t)d}{2(d+\lambda(t)}}, c(t) - \frac{c(t)d}{2\big(d+\lambda(t)\big)}\right). \tag{2.77}$$

However, in order to utilise create a stochastic (and hence more accurate) approximation, and to show that this model has affine elements in its instantaneous covariance matrix (so that the characteristic function can be constructed, see subsection 3.1.1 in Chapter 3), we need to specify the dynamics of $\sqrt{V(t)}$ in terms of an SDE.

Suppose a stochastic process $\xi(t)$ can be constructed such that this process is equal in distribution to (2.77). This implies that $\xi(t) \overset{d}{=} \sqrt{V(t)}$ and hence $\xi(t)$ has a normal distribution. Formally, construct $\xi(t)$ as:

$$d\xi(t) = \mu^{\xi}(t)dt + \psi^{\xi}(t)dW_V(t), \qquad \xi(0) = \sqrt{V(0)}. \tag{2.78}$$

Here the Brownian motion is the same as the Brownian motion driving the variance process. (2.78) is normally distributed and hence in order to find the mean and standard deviation a moment matching technique is used:

$$\mu^{\xi}(t) = \frac{d}{dt}E\left[\sqrt{V(t)}\right], \qquad \psi^{\xi}(t) = \sqrt{\frac{d}{dt}\;\mathbb{V}\left[\sqrt{V(t)}\right]}.$$

Now (2.74) and (2.75) can be used and thus explicit formulae are given:

$$\mu^{\xi}(t) = \frac{1}{2\sqrt{2}}\frac{\Gamma\left(\frac{1+d}{2}\right)}{\sqrt{c(t)}}\left[{}_1\tilde{F}_1\left(-\frac{1}{2},\frac{d}{2},-\frac{\lambda(t)}{2}\right)\frac{1}{2}\nu^2e^{-\kappa t} + {}_1\tilde{F}_1\left(\frac{1}{2},\frac{2+d}{2},-\frac{\lambda(t)}{2}\right)\frac{V(0)\kappa}{1-e^{\kappa t}}\right]. \tag{2.79}$$

$$\psi^{\xi}(t) = \left(\kappa\big(\Theta_V - V(0)\big)e^{-\kappa t} - 2\sqrt{2c(t)}\, e^{-\frac{\lambda(t)}{2}} \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\lambda(t)}{2}\right)^k \frac{\Gamma\left(\frac{1+d}{2}+k\right)}{\Gamma\left(\frac{d}{2}+k\right)} * \mu^{\xi}(t)\right)^{\frac{1}{2}}. \tag{2.80}$$

Here the (regularised) confluent hypergeometric function of the first kind is given by:

$$ {}_1\tilde{F}_1(a,b,z) = {}_1F(a,b,z)/\Gamma(b) = \sum_{j=0}^{\infty} \frac{(a)_j}{(b)_j}\frac{z^j}{j!} \times \frac{1}{\Gamma(b)}. $$

The notation $(c)_j$, known as the Pochhammer symbol, is defined as:

$$ (c)_j = c(c+1)\big((c+1)+1\big)\ldots(c+j-1), \qquad c_0 = 1. $$

**Technical Remark: Strict path wise equality between the processes $\xi(t)$ and $\sqrt{V(t)}$ is not needed when pricing European options which have a single payoff at some time $T, t < T$. Only equality in terminal distributions is needed, i.e. $t = T$. The effect of using this model for pricing non-European claims is left to a discussion in Chapter 3.

Finally, the H2-HW Model can now be fully specified. Incorporating $\xi(t)$ into the system of SDE's in (2.65-2.67) yields a new four-dimensional space $\boldsymbol{Y}(t) = [S(t), V(t), r(t), \xi(t)]^T$ given by:

$$ \left.\begin{aligned} \frac{dS(t)}{S(t)} &= r(t)dt + \sqrt{V(t)}dW_S(t), && S(0) > 0 \\ dV(t) &= \kappa\big(\Theta_V - V(t)\big)dt + \nu\sqrt{V(t)}dW_V(t), && V(0) > 0 \\ dr(t) &= \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta dW_r(t), && r(0) > 0 \\ d\xi(t) &= \mu^{\xi}(t)dt + \psi^{\xi}(t)dW_V(t), && \xi(0) = \sqrt{V(0)} \end{aligned}\right\} \tag{2.81} $$

The correlation structure, after taking $x(t) = \log\big(S(t)\big)$, is as follows:

$$ \begin{aligned} dW_x(t)dW_V(t) &= \rho_{x,V}dt \\ dW_x(t)dW_r(t) &= \rho_{x,r}dt \\ dW_V(t)dW_r(t) &= 0. \end{aligned} $$

For completeness, noting the transform, $x(t) = \log\big(S(t)\big)$, the instantaneous covariance matrix is given by:

$$ \tilde{\Sigma} = \begin{bmatrix} V(t) & \nu\rho_{x,v}V(t) & \rho_{x,r}\eta\xi(t) & \rho_{x,v}\psi^{\xi}(t)\xi(t) \\ * & \nu^2 V(t) & 0 & \nu\psi^{\xi}(t)\xi(t) \\ * & * & \eta^2 & 0 \\ * & * & * & \big(\psi^{\xi}(t)\big)^2 \end{bmatrix} \tag{2.82} $$

The model in (2.81) with covariance matrix (2.82) fulfils the requirements to construct a characteristic function and hence the characteristic function is given by (remembering $\tau = T - t, T > t$):

$$ \phi_{H2-HW}(u,t,T) = $$
$$ \exp\Big(A(u,\tau) + B_x(u,\tau)x(t) + B_r(u,\tau)r(t) + B_v(u,\tau)V(t) + B_{\xi}(u,\tau)\xi(t)\Big). \tag{2.83} $$

The terminal condition is given by $\phi_{H2-HW}(u,T,T) = \exp\big(iux(T)\big)$. Each of the functions can be found by solving the following ordinary differential equations:

$B'_x(u,\tau) = 0, \qquad B_x(u,0) = iu$

$B'_r(u,\tau) = -1 + B_x(u,\tau) - \lambda B_r(u,\tau), \qquad B_r(u,0) = 0$

$B'_V(u,\tau) = (B_x(u,\tau) - 1)B_x(u,\tau)/2 + \big(\nu\rho_{x,V}B_x(u,\tau) - \kappa\big)B_V(u,\tau) + \nu^2\big(B_V(u,\tau)\big)^2/2,$
$B_V(u,0) = 0$

$B'_{\xi}(u,\tau) = \rho_{x,r}\eta B_x(u,\tau)B_r(u,\tau) + \psi^{\xi}(t)\rho_{x,V}B_x(u,\tau)B_{\xi}(u,\tau) + \nu\psi^{\xi}(t)B_V(u,\tau)B_{\xi}(u,\tau),$
$B_{\xi}(u,0) = 0$

$$A'(u,\tau) = \kappa\Theta_V B_V(u,\tau) + \lambda\Theta_r(t)B_r(u,\tau) + \mu^{\xi}(t)B_{\xi}(u,\tau) + \eta^2 B_r(u,\tau)/2 + \left(\psi^{\xi}(t)\right)^2\left(B_{\xi}(u,\tau)\right)^2/2\,,$$
$$A(u,0) = 0.$$

Note that all derivatives are taken with respect to time t and are typically denoted as $B_x'(u,\tau) = \frac{d}{dt}B_x(u,\tau)$. Lastly note that the solutions to $B_x(u,\tau)$, $B_V(u,\tau)$ and $B_r(u,\tau)$ are given exactly as the solutions given in (2.71) under the H1-HW Model. The two remaining functions need to be solved numerically.

2.2.6.4 **Heston-Hull-CIR Model (H2-CIR)**

This model is essentially a continuation of the discussion above with the interest rate process assumed to follow a CIR process. The material is from the same article by (Grzelak et al, 2011). The dynamics are specified as:

$$dS(t) = r(t)S(t) + \sqrt{V(t)}S(t)dW_S(t). \tag{2.84}$$
$$dV(t) = \kappa\big(\Theta_{\mathrm{V}} - V(t)\big)dt + \nu\sqrt{V(t)}dW_V(t). \tag{2.85}$$
$$dr(t) = \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta\sqrt{r(t)}dW_r(t). \tag{2.86}$$

The correlation structure, after taking $x(t) = \log\big(S(t)\big)$, is as follows:

$$dW_x(t)dW_{\mathrm{V}}(t) = \rho_{x,V}dt$$
$$dW_x(t)dW_{\mathrm{r}}(t) = \rho_{x,r}dt$$
$$dW_V(t)dW_{\mathrm{r}}(t) = 0.$$

The instantaneous covariance matrix of the system (2.84)-(2.86), with $x(t) = \log\big(S(t)\big)$, is given as:

$$\Sigma = \begin{bmatrix} V(t) & \rho_{x,V}\nu V(t) & \rho_{x,r}\eta\sqrt{r(t)}\sqrt{V(t)} \\ * & \nu^2 V(t) & 0 \\ * & * & \eta^2 r(t) \end{bmatrix} \tag{2.87}$$

The problem is that the construction of a characteristic function for the log-price process is not possible as the $\Sigma_{1,3}$ term in (2.87) is not affine (see Chapter 3, subsection 3.1.1). A deterministic approximation for each of $\sqrt{r(t)}$ &$\sqrt{V(t)}$ can be done, however as in the previous subsection the focus will lie on a stochastic approximation. This approximation will transform the 3-dimensional SDE system in (2.84)-(2.86) into a 6-dimensional system (although still driven only by 3 Brownian motion diffusion terms). The terms in $\Sigma_{1,3}$ are thus approximated by normally distributed processes,$\xi(t)$ and $R(t)$, as follows:

$$\Sigma_{1,3} \approx \tilde{\Sigma}_{1,3} = \rho_{x,r}\eta\xi(t)R(t), \qquad \xi(t) \approx \sqrt{V(t)}, R(t) \approx \sqrt{r(t)}\,. \tag{2.88}$$

As the reader would note, the product $\xi(t)R(t)$ is not affine. Luckily Itô's Lemma can be applied to specify the dynamics of $z(t) = \xi(t)R(t)$ in terms of a SDE:

$$\mathrm{d}z(t) = \left(\mu^R(t)\xi(t) + \mu^{\xi}(t)R(t)\right)dt + \psi^{\xi}(t)R(t)dW_V(t) + \psi^R(t)\xi(t)dW_r(t). \tag{2.89}$$

Recall that $\mathrm{d}\big(X(t)Y(t)\big) = Y(t)dX(t) + X(t)dY(t) + dX(t)dY(t)$ specifies Itô's Lemma for the

product of two Itô processes. The 6-dimensional system $\boldsymbol{X}(t) = [x(t), V(t), r(t), \xi(t), R(t), z(t)]^T$ can now be fully specified as (2.90):

$$
\begin{aligned}
dx(t) &= \left(r(t) - \frac{1}{2}V(t)\right)dt + \sqrt{V(t)}dW_x(t), \qquad x(0) = \log\big(S(0)\big) \\
dV(t) &= \kappa\big(\Theta_{\mathrm{V}} - V(t)\big)dt + \nu\sqrt{V(t)}dW_V(t), \qquad V(0) > 0 \\
dr(t) &= \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta\sqrt{r(t)}dW_r(t), \qquad r(0) > 0 \\
d\xi(t) &= \mu^{\xi}(t)dt + \psi^{\xi}(t)dW_V(t), \qquad \xi(0) = \sqrt{V(0)} \\
dR(t) &= \mu^{R}(t)dt + \psi^{R}(t)dW_r(t), \qquad R(0) = \sqrt{r(0)} \\
\mathrm{d}z(t) &= \left(\mu^{R}(t)\xi(t) + \mu^{\xi}(t)R(t)\right)dt + \psi^{\xi}(t)R(t)dW_V(t) + \psi^{R}(t)\xi(t)dW_r(t), \\
z(0) &= \sqrt{V(0)}\sqrt{r(0)}, \xi(t) \approx \sqrt{V(t)}, R(t) \approx \sqrt{r(t)}, z(t) \approx \sqrt{V(t)}\sqrt{r(t)}.
\end{aligned}
$$

As confirmation, the instantaneous covariance matrix admits the requirements to be in the form of equation (3.8), Chapter 3:

$$
\Sigma = \begin{bmatrix}
V(t) & \rho_{x,V}\nu V(t) & \rho_{x,r}\eta z(t) & \rho_{x,V}\psi^{\xi}(t)\xi(t) & \rho_{x,r}\psi^{R}(t)\xi(t) & s_1(t) \\
* & \nu^2 V(t) & 0 & \psi^{\xi}(t)\nu\xi(t) & 0 & \nu\psi^{\xi}(t)z(t) \\
* & * & \eta^2 r(t) & 0 & \psi^{R}(t)\eta R(t) & \eta\psi^{R}(t)z(t) \\
* & * & * & \left(\psi^{\xi}(t)\right)^2 & 0 & \left(\psi^{\xi}(t)\right)^2 R(t) \\
* & * & * & * & \left(\psi^{R}(t)\right)^2 & \left(\psi^{R}(t)\right)^2 \xi(t) \\
* & * & * & * & * & s_2(t)
\end{bmatrix} \tag{2.91}
$$

Here $s_1(t) = \rho_{x,V}\psi^{\xi}(t)z(t) + \rho_{x,r}\psi^{R}(t)V(t)$ and $s_2(t) = \left(\psi^{\xi}(t)\right)^2 r(t) + \left(\psi^{R}(t)\right)^2 V(t)$ and note that the functions $\psi^{\xi}(t)$ & $\psi^{R}(t)$ are deterministic functions. The set of functions $\{\mu^{\xi}(t), \mu^{R}(t), \psi^{\xi}(t),\ \psi^{R}(t)\}$ are given as in (2.74)-(2.80), with changes to compute $\mu^{R}(t)$ & ψ^{R} being made easily by taking the parameters in (2.85) and replacing them by the corresponding parameters in (2.86) (e.g. $\nu \to \eta$, etc.). Consequently, as specified by (3.10)-(3.12), in the next chapter, the characteristic function can be constructed:

$$
\begin{aligned}
\phi_{H2-CIR}(u, \boldsymbol{X}(t), \tau) = \exp\big(&A(u,\tau) + B(u,\tau)x(t) + C(u,\tau)r(t) + D(u,\tau)v(t) + \\
&E(u,\tau)\xi(t) + F(u,\tau)R(t) + G(u,\tau)z(t)\big).
\end{aligned} \tag{2.92}
$$

The function in (2.92) satisfy the following sets of ordinary differential equations, as (2.93), for $u \in \mathbb{R}, \tau > 0$, and all functions have a terminal condition equal to zero when $\tau = 0$ except $B(u,0) = iu$:

$$
\begin{aligned}
B'(u,\tau) &= 0, \\
C'(u,\tau) &= -1 + B(u,\tau) - \lambda C(u,\tau) + \frac{\eta^2 C^2(u,\tau)}{2} + \frac{\left(\psi^{\xi}(t)\right)^2 G^2(u,\tau)}{2}, \\
F'(u,\tau) &= \mu^{\xi}(t)G(u,\tau) + \psi^{R}(t)\eta C(u,\tau)G(u,\tau) + \left(\psi^{\xi}(t)\right)^2 E(u,\tau)G(u,\tau), \\
G'(u,\tau) &= \eta\rho_{x,r}B(u,\tau)C(u,\tau) + \rho_{x,V}\psi^{\xi}(t)B(u,\tau)G(u,\tau) + \nu\psi^{\xi}(t)D(u,\tau)G(u,\tau) \\
&\quad + \eta\psi^{R}(t)C(u,\tau)G(u,\tau)
\end{aligned}
$$

$$D'(u,\tau) = \frac{B(u,\tau)(B(u,\tau)-1)}{2} - \kappa D(u,\tau) + \nu\rho_{x,V}B(u,\tau)D(u,\tau) + \frac{\nu^2 D^2(u,\tau)}{2} + \rho_{x,r}\psi^R(t)B(u,\tau)G(u,\tau) + \frac{\left(\psi^R(t)\right)^2 G^2(t)}{2},$$

$$E'(u,\tau) = \mu^R(t)G(u,\tau) + \psi^{\xi}(t)\rho_{x,V}\,B(u,\tau)E(u,\tau) + \nu\psi^{\xi}(t)D(u,\tau)E(u,\tau) + \rho_{x,r}\psi^R(t)B(u,\tau\,)F(u,\tau) + \left(\psi^R(t)\right)^2 F(u,\tau)G(u,\tau),$$

$$A'(u,\tau) = k\Theta_V D(u,\tau) + \lambda\Theta_r C(u,\tau) + \mu^{\xi}(t)E(u,\tau) + \mu^R(t)F(u,\tau) + \frac{\left(\psi^{\xi}(t)\right)^2 E^2(u,\tau)}{2} + \frac{\left(\psi^R(t)\right)^2 F^2(u,\tau)}{2}.$$

The above is typically solved using numerical techniques. The procedures to solving the above are given in the supporting material, i.e. see Appendix A for details.

2.3 **General asset price dynamics and models**

Stochastic analysis of asset prices have often been associated with assuming that the source of randomness follows a Brownian motion (BM). This is of course within reason given the useful properties that BM possess, e.g. independence of non-overlapping increments, normality of increments, continuity of paths, etc.

One can however question whether BM should be used at all as the source of randomness. In the following subsections (2.3.1-2.3.5) we will expand the basic BM Model, by considering a few generalisations of the BM and in later subsections we introduce stochastic clocks. We will also look at models with jumps. Most of these models are popular in literature and, where possible, references to the original author(s) who introduced these models will be given.

2.3.1 **Fractional Brownian motion**

Here the BM Model is extended and so we define the Fractional Brownian Motion (FBM), as found in (Mandelbrot and van Ness, 1968), that has the following (unique) characteristics:

1.) $B_H(0) = 0, and\ with\ probability\ one\ B_H(t)\ is\ continuous.$
2.) $For\ any\ t \geq 0\ and\ any\ \delta t > 0\ the\ increment\ \left(B_H(t+\delta t) - B_H(t)\right)\ is\ normally\ distributed\ with\ mean\ zero\ and\ variance\ \delta t^{2H}.$

Alternatively (Dieker, 2004) discusses many advanced issues on the topic, gives us a formal definition of FBM that is also used widely in literature and this definition is originally due to (Mandelbrot and van Ness, 1968) :

$$B_H(t) \coloneqq \frac{1}{\Gamma(H+0.5)}\left(\int_{-\infty}^{0}((t-s)^{H-0.5} - (-s)^{H-0.5})dB(s) + \int_0^t (t-s)^{H-0.5}\,dB(s)\right).$$

The parameter H is known as the Hurst Index (or parameter), $0 < H < 1$, $\Gamma(\alpha)$ is the usual gamma function and $B(s)$ is the standard BM integrator.

Note that we can possibly have dependent increments (unless $H = 0.5$, which reduces to ordinary BM as can be seen from the equation given for $B_H(t)$).

Some properties of FBM are given below:

1.) $E[B_H(t) \times B_H(s)] = 0.5 \times (|t|^{2H} + |s|^{2H} - |t-s|^{2H})$
2.) Stationary increments: $B_H(t) - B_H(s) = B_H(t-s)$
3.) $B_H(c \times t) = c^H \times B_H(t)$ in other words the process is self-similar.

Recall that a random process $X(t)$ is called a Self-Similar process if it satisfies:

$$\{X(ct_1), \dots, X(ct_k)\} \overset{d}{=} \{c^H X(t_1), \dots, c^H X(t_k)\}.$$

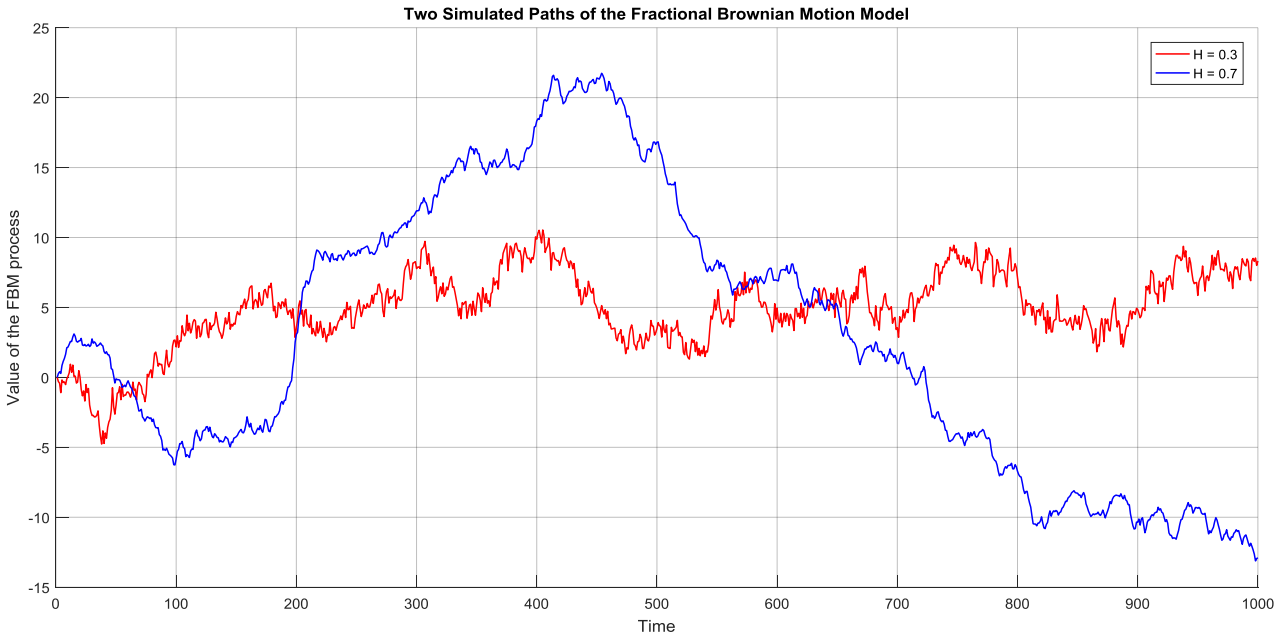

Figure 2.3: Fractional Brownian Motion paths with different Hurst parameters.

$for\ some\ number\ H > 0\ and\ all\ c, k, t_1, \dots, t_k \geq 0$. The Hurst parameter has two effects depending on whether $0 < H < 0.5$ (anti-persistence) or $0.5 < H < 1$ (persistence). Here we mean that if a process is persistent then a general upwards movement will tend to be followed by further up movements and similarly downward movements are followed up by further down movements. An illustration of FBM is given in Figure 2.3.

Clearly one can see that for $H = 0.7$ the graph shows a high level of persistence and conversely for $H = 0.3$ there is no apparent dependency on past increments. If the Hurst parameter is equal to 0.5 then we reduce to the case of ordinary BM.

Below we will estimate the Hurst Index for the S&P 500 data presented in Chapter one. The reader is encouraged to read the article by (Bayraktar et al, 2003) in which the authors have estimated the Hurst parameter of the S&P500 over 1-minute sampling intervals from January 1989 to May 2000. They found that the average Hurst parameter is 0.6156, which is not equal to the theoretical 0.5 of BM.

We will embark on a similar analysis (briefly), although we will be using daily data. Unfortunately our estimates will only be "rough" and the idea is to give the reader only an impression (rather than a rigorous investigation) of how the parameter H works in practice (again under the assumption that FBM fully describes the underlying process):

Code:
HEST = wfbmesti(SP)

Results:
HEST = [0.5288 0.5444 0.4306]

Working with the logarithm of the prices instead gives:

Code:
HEST = wfbmesti(log(SP))

Results:
HEST = [0.5028 0.5237 0.4664]

The MATLAB® (2016a) documentation describes this code and result sufficiently:

"*HEST = wfbmesti(X) returns a one-by-three vector HEST which contains three estimates of the fractal index H of the input signal X. The signal X is assumed to be a realization of fractional Brownian motion with Hurst index H.*

The first two elements of the vector are estimates based on the second derivative with the second computed in the wavelet domain.

The third estimate is based on the linear regression in log-log plot, of the variance of detail versus level."

Additionally the reader can find details relating to these methods by reading the survey by (Bardet et al, 2003).

For the purpose of this exposition no conclusions will be made as what the Hurst parameter for the S&P500 is although it is interesting to note that the values are not far from 0.5. In fact in the previously mentioned article by (Bayraktar et al, 2003) the authors noted that towards the end of the decade (1997-2000) the Hurst estimate approached 0.5 and conjectured that this might be due to the rise in internet trading, i.e. many market participants trading leads to a closer ideal of an efficient market. Naturally the S&P500 Index would then reflect this.

Before ending this section it is necessary to mention that irrespective of the value of the Hurst parameter, FBM and hence BM are ultimately not ideal models for asset prices. (Calvet and Fisher, 2008) mention that empirical evidence contradict the idea that time series are self-similar processes as described above. This is due to the fact that as the sampling interval increases (i.e. intra-daily to daily to yearly) the tails of the distribution become thinner and less peaked.

The BM framework is now further extended by considering Lêvy Processes.

2.3.2 **Models with jumps and Lêvy processes**

The general theory underlying Lêvy processes is given here. It is essentially a summary and is intended to make the presentation of the following three models clearer. Additionally, the theory underpins the theory behind Brownian motion as well. The details of the theory can be found in many books on stochastic analysis, here the material is reworked from the books of (Kyprianou, 2006) and (Kallenberg, 2010).

Formally, a Lêvy Process, $\{X(t)\}_{t\in[0,T]}$, is defined by the following properties:

1.) $\mathbb{P}[X_0 = 0] = 1$. (The process starts at zero with probability 1)

2.) For any time s, $0 < s \leq t$, we have that $X_t - X_s \overset{d}{=} X_{t-s}$. (Equal in distribution)

3.) For any time $u: u \leq s$, the process defined by $X_t - X_s$ is independent of X_u.

4.) Paths of the process X_t are right continuous (a.s.) and all left limits exists. (Cadlag Property).
The last property implies that the process can have jumps. Indeed, this notion can be made precise:

$$J(t) = X(t) - X(t^-) \coloneqq X(t) - \lim_{s \to t, s \leq t} X(s).$$

Denote U to be a subset of the Borel Algebra on $[0,T]$, i.e. $U \in \mathfrak{B}([0,T])$. Then the number of jumps N, in the interval $[0,t]$, each jump of size $J(s)$, is defined by:

$$N(t,U) \coloneqq \sum_{0 \leq s \leq t} I_U\big(J(s)\big).$$

The indicator function $I_U(x)$ is defined as being 1 if $x \in U$ and 0 if $x \notin U$. Formally N is called a Jump measure (in particular here the measure is known as a Poisson measure). It is assumed that $N(t,U) < \infty$ for all $U \in \mathfrak{B}(\mathbb{R}^+)$ and maps from the space and time states to the real numbers.

The Poisson measure as given above is a special case of the Lêvy measure. Let $\{X(t)\}_{t \geq 0}$ be a Lêvy process on $\mathbb{R}^d$, for some positive integer d (i.e. the number of dimensions). Define a (positive) measure v on $\mathbb{R}^d$ in the following manner:

$$v(A) = E[\#\{t \in [0,1] | J(t) \neq 0, \Delta X(t) \in A\}], \qquad A \in \mathfrak{B}(\mathbb{R}^d)$$

This simply reads as the expected number of jumps, whose jump sizes are in the set A, per time unit of the process $X(t)$. In other words the Lêvy measure v is simply a way of quantifying the "jump" behaviour of the process $X(t)$. Additionally, the Poisson measure is related to the Lêvy measure by

$$v(U) = E[N(1,U)].$$

A Lêvy measure must also satisfy the following properties:

1.) $\int_{|x| \leq 1} x^2 v(dx) < \infty$

2.) $\int_{|x| \geq 1} v(dx) < \infty$

Essentially there are three building blocks that make up a Lêvy process. This triplet, denoted by (Σ, b, v), uniquely specifies the process $\{X(t)\}_{t \geq 0}$. This leads to the following theorem which is given without proof see (Kallenberg, 1997).

Theorem 2.4

Let $X(t)$ be an $\mathbb{R}^d$ valued process with independent increments and continuous in probability. Then, there is a unique continous function $\mathbb{R}^d \times \mathbb{R}^+ \to \psi_t(a)$ with $\psi_0(a) = 0$ and

$$E\big[e^{ia(X(t)-X(0))}\big] = e^{i\psi_t(a)}.$$

$\forall\, a \in \mathbb{R}^n$ and $t \geq 0$. Additionally $\psi_t(a)$ can be written as:

$$\psi_t(a) = ia \cdot b_t - \frac{1}{2} a^T \Sigma_t a + \int_{\mathbb{R}^d \times [0,t]} \left(e^{ia \cdot x} - 1 - \frac{ia \cdot x}{1 + \|x\|} \right) dv(x,s).$$

where Σ_t, b_t and v are uniquely determined and satisfy the following:

1.) $t \to \Sigma_t$ is a continous function from $\mathbb{R}^+$ to $\mathbb{R}^{d^2}$ such that $\Sigma_0 = 0$ and $\Sigma_t - \Sigma_s$ is positive semidefinite for all $t \geq s$.

2. $t \to b_t$ is a continuous function from $\mathbb{R}^+$ to $\mathbb{R}^d$ with $b_0 = 0$.

3. v is a Borel measure on $\mathbb{R}^d \times \mathbb{R}^+$ with $v(\{0\} \times \mathbb{R}^+) = 0, v(\mathbb{R}^d \times \{t\}) = 0\ \forall\, t \geq 0$ and

$$\int_{\mathbb{R}^d \times [0,t]} \|x\|^2 \wedge 1 \; d\nu(x,s) < \infty.$$

Furthermore the triplet, (Σ, b, ν), uniquely determines all finite distributions of the process $X(t) - X(0)$ and conversely, if (Σ, b, ν) is any triplet satisfying the conditions $1-3$ given then there exists some process $X(t)$ with independent increments admitting the presentation given by $\psi_t(a)$.

**Technical note: The function given by $\frac{1}{1+\|x\|}$ in Theorem 2.4 can be written in the form $ia \cdot x\theta(x)$ where $\theta(x) \to 1$ as $\|x\| \to 0$. The particular form of $\theta(x)$ does not matter as changing it will simply lead to differences in the integral being absorbed into the drift term b. Commonly $\theta(x)$ is taken to be $\theta(x) = 1_{\{\|x\| \leq 1\}}$.

Now if $X(t)$ is a Lêvy process then Theorem 2.4 (with the additional condition of stationary increments on the process $X(t)$) yields a special case known as the Lêvy-Khintchine formula:

$$E\left[e^{ia(X(t)-X(0)}\right] = \exp\left(t\psi(a)\right). \tag{2.94}$$

$$\psi(a) = ia \cdot b - \frac{1}{2}a^T \Sigma a + \int_{\mathbb{R}^d} \left(e^{ia\cdot x} - 1 - \frac{ia \cdot x}{1 + \|x\|}\right) d\nu(x). \tag{2.95}$$

It is well known in mathematical statistics that all continuous processes with independent increments are Gaussian. As a consequence if $X(t)$ is a continuous d-dimensional process with independent increments then (2.95) can be given as:

$$\psi(t,a) = ia \cdot b_t - \frac{1}{2}a^T \Sigma_t a. \tag{2.96}$$

In the case of stationary increments this formula reduces to taking time out as a common factor, i.e. $b_t = t \times b, \Sigma_t = t \times \Sigma$. Comparing this to (2.96) and (2.95) shows the effect of including a jump process (the last term of (2.95)). Thus an extended version of Itô's Formula for a Function of Several Processes can be given which includes the effects of a jump process (see (Kienitz and Wetterau, 2012)). Note there are different versions of this formula and the one given is very general.

Theorem 2.5

Let $\bar{N}(dt, dz) = N(dt, dz) - v(dz)dt$, which is known as the Compensated Jump measure. Consider d_1- dimensional Brownian motions $\boldsymbol{W}(t) = \left(W_1(t), \dots, W_{d_1}(t)\right)$ and furthermore d_2- dimensional Compensated Jump measures $\boldsymbol{N}(dt, dz) = \left(N_1(dt, dz), \dots, N_{d_2}\left(dt, dz_{d_2}\right)\right)$. Let $z = \left(z_1, \dots, z_{d_2}\right)$ and consider d Lêvy processes of the form:

$$d\boldsymbol{X}(t) = \boldsymbol{\mu}(t)dt + \boldsymbol{\sigma}(t)d\boldsymbol{W}(t) + \int_{(\mathbb{R}^+)^d} \boldsymbol{\gamma}(t,z)\boldsymbol{N}(dt,dz).$$

Lastly, let $f: (0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function and define

$$Y(t) = f\left(t, \boldsymbol{X}(t)\right).$$

Hence the process $Y(t)$ admits the following SDE:

$$dY(t) = \frac{\partial f}{\partial t}\left(t, \boldsymbol{X}(t)\right)dt + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\mu_i(t)dt + \sum_{i=1}^{d}\sum_{j=1}^{d_1} \frac{\partial f}{\partial x_i}\left(t, \boldsymbol{X}(t)\right)\sigma_{i,j}(t)dW_j(t)$$

$$+\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d_1}\frac{\partial^2 f}{\partial x_i \partial x_j}\big(t,\boldsymbol{X}(t)\big)\,(\sigma\sigma^T)_{i,j}\,(t)dt$$
$$+\sum_{k=1}^{d_2}\int_{\mathbb{R}^+}\left[f\big(t,\boldsymbol{X}(t)+\gamma_k(t,z)\big)-Y(t)-\sum_{i=1}^{d}\frac{\partial f}{\partial x_i}\big(t,\boldsymbol{X}(t)\big)\gamma_{i,k}(t,z)\right]\nu_k\big(d_{z_k}\big)dt$$
$$+\sum_{k=1}^{d_2}\int_{\mathbb{R}^+}\big[f\big(t,\boldsymbol{X}(t^-)+\gamma_k(t,z)\big)-f\big(t,\boldsymbol{X}(t^-)\big)\big]N_k\big(dt,d_{z_k}\big).$$

The equation involving $d\boldsymbol{X}(t)$ is shorthand notation and is similar to (2.1). The function $\boldsymbol{\gamma}(t,z)$ is assumed to be a smooth function and adapted to the filtration generated by the process $\boldsymbol{X}(t)$.

The theory involved in pricing options using theorem 2.5 is useful if one wishes to specify the correct dynamics for functionals of multidimensional jump diffusions. Regarding the scope of our work here, we will not use theorem 2.5 directly and instead we will focus on the theory of Lêvy processes and characteristic functions in detail for pricing purposes. We now turn to the Merton Model.

2.3.3 **Merton Model**

The Merton Model (Merton, 1976) extends the basic diffusion process in equation (1.1) by adding in a jump component. Although this model will not be used to price derivatives it nevertheless gives us insights into how the mechanics of including a simple jump diffusion into a BM framework works.

Recall the compound Poisson process defined by a sequence of i.i.d. random variables Y_n with distribution F. Let $N(t)$ be a counting process with jump intensity $\lambda \in \mathbb{R}^+$, in other words

$$P[N(t)=k]=\frac{e^{-\lambda t}(\lambda t)^k}{k!},\qquad k\in\ \mathbb{N}$$

Hence the compound Poisson process is defined by

$$N_{Y(t)}=\sum_{n=0}^{N(t)}Y_n.$$

Consider the case where F admits a log-normal distribution. Using the representation

$$Y=\mu_J\exp\left(-\frac{1}{2}\sigma_J^2+\sigma_J Z\right),\qquad Z\sim\mathcal{N}(0,1)$$

allows us to define the Merton Model:

$$dS(t)=\mu S(t)dt+\sigma S(t)dW(t)+S(t)(Y-1)dN(t),\qquad S(0)=S_0$$
$$N(0)=0,\qquad \mu_J=\ln(1+a)-\frac{b^2}{2},\qquad \sigma_J=b,\qquad a>-1,b\geq 0.$$

The correlations between each the Brownian motion, the process Y and the counting process are all 0.

Note that letting the drift term be equal to $r-d-\frac{\sigma^2}{2}-\lambda\left(e^{-\frac{\sigma^2}{2}-\mu_J}-1\right)$ allows use to obtain risk

neutral dynamics for the model. The BS Model parameters together with the new constant parameters gives a model that is capable of capturing the volatility smile for short dated options.

Finally, for completeness the pricing formula (assuming no dividends) for a European call option is given by:

$$C(K,T) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} e^{rt + k\mu_J + \frac{k}{2}\sigma_J} \left(S(0)\mathcal{N}(d_{1,k}) - K\mathcal{N}(d_{2,k}) \right)$$

$$d_{1,k} = \frac{\log\left(\frac{S(0)}{K}\right) + \left(\left(r + \frac{\sigma^2}{2}\right)t + k\left(\mu_J + \sigma_J^2\right)\right)}{\sqrt{\sigma^2 t + k\sigma_J^2}}$$

$$d_{2,k} = \frac{\log\left(\frac{S(0)}{K}\right) + \left(\left(r - \frac{\sigma^2}{2}\right)t + k\left(\mu_J + \sigma_J^2\right)\right)}{\sqrt{\sigma^2 t + k\sigma_J^2}}.$$

Pricing of exotic options can be done using the (risk neutral) transition density and is explained in more general terms in the next chapter:

$$p(x,t;X,T) = \exp\left(-\lambda(T-t)\right) \sum_{k=0}^{\infty} \frac{\left(\lambda(T-t)\right)^k \exp\left(-\frac{\left((X-x) - (r-d)(T-t) - k\mu_J\right)}{2\left(\sigma^2(T-t) + k\sigma_J^2\right)} \right)}{k!\sqrt{2\pi\left(\sigma^2(T-t) + k\sigma_J^2\right)}}.$$

The formulas given above converge quickly and it is evident how the typical Poisson density is incorporated into both formulas. We now turn to the basic Variance Gamma Model and its extensions.

2.3.4 **Variance Gamma Model and its extensions**

Treatment of the basic model is taken from (Cariboni and Schoutens, 2009), (Campolieti and Makarov, 2014) and (Kienitz and Wetterau, 2012). The material has been reworked to better aid understanding.

The Variance Gamma Model (VG) was fully developed by (Madan et al, 1998). The model can be defined in two equivalent ways. The first is known as the CGM representation and the second the σ, ν, θ representation. The two representations are mathematically equivalent considering the characteristic function, but the basic formulation of each case is different.

For both cases consider a gamma process denoted by $\{G(t;\mu,\nu)\}_{t\geq 0}$. The process has mean rate μ and variance rate ν and the process begins at zero with stationary and independent gamma distributed increments. Let h be a sufficiently small positive number. The increment $\Delta G = G(t+h) - G(t)$ over the interval $[t, t+h]$ has a gamma distribution with probability density given:

$$p_G(g;\mu,\nu,h) = \left(\frac{\mu}{\nu}\right)^{\frac{\mu^2 h}{\nu}} \frac{g^{\frac{\mu^2 h}{\nu}-1}\exp\left(-\frac{\mu}{\nu}g\right)}{\Gamma\left(\frac{\mu^2 h}{\nu}\right)}, \qquad g > 0. \tag{2.97}$$

Additionally the following properties hold:

1.) $G(0) = 0\ (with\ probability\ 1)$

2.) $G(1) \sim \Gamma(\alpha, \beta)$

3.) $G(t) - G(s) \sim \Gamma(\alpha(t-s), \beta),\ 0 < s < t$

Here $\alpha = \frac{\mu^2}{\nu}, \beta = \frac{\mu}{\nu}$.

The CGM representation:

The CGM representation is formulated as the difference between two independent Gamma processes:

$$X_{VG}(t) = G_1(t; C, M) - G_2(t; C, G)$$

Evidently it can be seen that the mean rate is the same for both processes, i.e. C, but the variance rate for both processes differ. Decreasing the parameter C leads to a more peaked distribution (kurtosis) whereas G and M determine the asymmetry of the distribution.

Consider the Lêvy measure of $X_{VG}(t)$:

$$\nu_{VG}^{CGM}(dx) = \frac{C}{|x|}\left[e^{-M|x|}\mathbf{1}_{\{x>0\}}dx + e^{-G|x|}\mathbf{1}_{\{x<0\}}dx\right]. \tag{2.98}$$

Two technical remarks:

By calculating $\int_{|x|\leq 1} x^2 v(dx) < \infty$ implies that the process has finite variation. Furthermore, $\nu_{VG}^{CGM}(\mathbb{R}) = \infty$, which is due to the singularity at the origin. This implies that the process has infinite activity, i.e. on any compact interval almost all sample paths have an infinite number of jumps. The general proof of these facts can be found in (Sato, 1999).

The latter of the previous remarks has an interesting economic interpretation for Lêvy processes. Infinite activity can be interpreted as an *approximation* to the daily supply and demand activity generated by buying and selling shares almost continuously. Finite activity on the other hand can be interpreted as sudden shocks, such as unexpected news events leading to a sudden drop or rise in prices. Thus an improvement to the VG process is to somehow change the measure in order to ensure that the process admits finite activity with small jumps occurring more frequently than larger jumps.

Lastly, for the CGM representation the characteristic function is given by:

$$\varphi_{VG}^{CGM}(u) = \left(\frac{GM}{GM + (M-G)iu + u^2}\right)^{C*t}. \tag{2.99}$$

The σ, ν, θ representation:

In the case of the σ, ν, θ representation, the VG Model is obtained by evaluating the Brownian motion at a random time. The time evaluation is given by the gamma process. Let $\{B(t;\theta,\sigma) \coloneqq W^{\theta,\sigma}(t) = \theta t + \sigma W(t)\}_{t\geq 0}$ be a scaled Brownian motion with drift rate θ and σ be the corresponding scale parameter. Hence the VG Model is defined by:

$$X(t;\sigma,\nu,\theta) \coloneqq B(G(t;1,\nu);\theta,\sigma) = \theta G(t;1,\nu) + \sigma W\big(G(t;1,\nu)\big). \tag{2.100}$$

Here the gamma process has mean rate $\mu = 1$ and variance rate ν. The density function of the process in (2.100) can numerically be obtained by the integral:

$$\begin{aligned} f_{X(t)}(x) &= \int_0^\infty f_{X(t)|G(t)}(x|g) f_{G(t)}(g)\,dg \\ &= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}} \exp\left(-\frac{(x-\theta g)^2}{2\sigma^2 g}\right) \frac{g^{\frac{t}{\nu}-1}\exp\left(-\frac{g}{\nu}\right)}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)}\,dg. \end{aligned} \tag{2.101}$$

The use of characteristic functions to recover the density function is an alternative to directly calculating (2.101) and hence is given by

$$\varphi_{VG}^{\sigma,\nu,\theta}(u) = \left(1 - iu\theta\nu + \frac{\sigma^2}{2}\nu u^2\right)^{-\frac{t}{\nu}} \tag{2.102}$$

The corresponding Lêvy measure for (2.102) is:

$$\nu_{VG}^{\sigma,\nu,\theta}(dx) = \frac{\exp\left(\frac{\theta x}{\sigma^2}\right)}{\nu|x|}\exp\left(-\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}}|x|\right)dx$$

As shown in (Carr et al, 2002), a direct mapping between the CGM and VG representations exist:

$$\begin{aligned} C &= \frac{1}{\nu} \\ G &= \left(\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} - \frac{1}{2}\theta\nu\right)^{-1} \\ M &= \left(\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} + \frac{1}{2}\theta\nu\right)^{-1}. \end{aligned}$$

There is a generalized version of the CGM representation presented below with the material is reworked from (Carr et al, 2002) with additional explanations added by the Author. This model has and additional parameter "Y" which allows for more flexible modelling. Consider a more general form of the Lêvy measure in (2.98):

$$\nu_{VG}^{CGMY}(dx) = \frac{C}{|x|^{1+Y}}\left[e^{-M|x|}\mathbf{1}_{\{x>0\}}dx + e^{-G|x|}\mathbf{1}_{\{x<0\}}dx\right]. \tag{2.103}$$

Here $Y < 2$ in order to ensure that the integral $\int_{|x|\le 1} x^2 \nu(dx)$ is integrable (i.e. is less than ∞) in a small neighbourhood of zero ($x = 0$), in accordance to the properties that a Lêvy measure must satisfy. If $Y = 0$ then the CGM representation is recovered. A few remarks on the values of Y is given in Table 2.1 below, as can be found in (Carr et al, 2002).

Some explanation of the terms in Table 2.1 is needed. The term "Completely Monotone" gives structure to the arrival rates between large and small jump sizes and how these relate to each other, i.e. large jumps arrive less frequently than small jumps. The term finite variation is a term that relates

to the $p\text{-}variation$ of a function over some interval $[a,b]$ and is given by

$$V_{[a,b]}^{(p)}(f) = \limsup_{\delta_t^{(n)} \to 0} \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p.$$

Here the limit is taken over all possible partitions of the interval: $a = t_0 < t_1 < \cdots < t_n = b$ and where $\delta_t^{(n)} := \max_{i=1:n}(t_i - t_{i-1}) \to 0$ as $n \to \infty$. If $V_{[a,b]}^{(p)}(f) < \infty$ then the function is said to have a $bounded\ p\text{-}variation$ on $[a,b]$. (A workable method for calculating the $p\text{-}variation$ of a process such as Brownian motion involves taking a suitable expected value of the process and proving that the variance tends to zero. Then the value of $V_{[a,b]}^{(p)}(f)$ is equal to this expected value (a.s.)).

Values for Y	Properties on the Resulting Lêvy Process
$Y < -1$	Not Completely Monotone, finite activity.
$-1 < Y < 0$	Completely Monotone, finite activity.
$0 < Y < 1$	Completely Monotone, infinite activity, finite variation.
$1 < Y < 2$	Completely Monotone, infinite activity, finite quadratic variation.

Table 2.1: Characteristics of the Y parameter in the CGYM Model.

Now it is known that a differentiable function with a bounded derivative has a bounded first variation and that for such a function the quadratic ($p = 2$) variation is zero (Campolieti and Markarov, 2014). Consequently if the quadratic variation is nonzero then the sample paths of a process are not differentiable (a.s.). For example, if $1 < Y < 2$ then sample paths of the VG process (CGMY) are not differentiable (a.s.).

A remark regarding the term "infinite activity" has already been made in a previous paragraph. As anticipated, the desirable improvements to the VG Model were made here via the extension to the CGMY Model.

The corresponding characteristic function of the CGMY Model and its proof now follow:

$$\phi_{CGMY}(u,t;C,G,M,Y) = \exp(tC\Gamma(-Y)[(M-iu)^Y - M^Y + (G+iu)^Y - G^Y]). \qquad (2.104)$$

Proof:

The Lêvy-Khintchine formula in (2.95) yields (remembering that VG is a pure jump process, hence $b\ \&\ \Sigma\ are\ both\ zero$):

$$\phi(u,t) = \exp\left(t\int_{-\infty}^{\infty}\left(e^{iux}-1\right)\nu_{VG}^{CGM}(x)dx\right). \qquad (2.105)$$

Considering (2.105) and splitting up the integral yields*:

$$\int_{-\infty}^{\infty}\left(e^{iux}-1\right)\nu_{VG}^{CGM}(x)dx = \int_0^{\infty}\left(e^{iux}-1\right)\frac{Ce^{-Mx}}{x^{1+Y}}dx + \int_0^{\infty}\left(e^{-iux}-1\right)\frac{Ce^{-Gx}}{x^{1+Y}}dx.$$

As a preliminary step recall that the gamma function is defined by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx.$$

Hence the first of two integrals is given:

$$\int_0^\infty \left(e^{iux}-1\right)\frac{Ce^{-Mx}}{x^{1+Y}}dx = \int_0^\infty x^{-1-y}\,C\left(e^{-(M-iu)x}-\mathrm{e}^{-\mathrm{Mx}}\right)dx$$
$$= C\int_0^\infty (M-iu)^Y w^{-Y-1}\exp(-w)\,dw - C\int_0^\infty M^Y w^{-Y-1}\exp(-w)\,dw$$
$$= C\Gamma(-Y)[(M-iu)^Y - M^Y].$$
The substitution for each integral is simple: $w=(M-iu)x$ and $w=Mx$ respectively.
The second integral is calculated:
$$\int_0^\infty \left(e^{-iux}-1\right)\frac{Ce^{-Gx}}{x^{1+Y}}dx = \int_0^\infty x^{-1-Y}C\left(e^{-(G+iu)x}-e^{-Gx}\right)dx$$
$$= C\int_0^\infty (G+iu)^Y w^{-Y-1}\exp(-w)\,dw - C\int_0^\infty G^Y w^{-Y-1}\exp(-w)\,dw$$
$$= C\Gamma(-Y)[(G+iu)^Y - G^Y]^{**}\,.$$

Finally adding the two integrals yields (2.103). ■

A few technical remarks: *In the proof the sum of integrals were considered instead of the difference as would be expected. Essentially the absolute values in (2.103) yield $|x| = -x\ for\ x<0$. By making the substitution $y=-x$ which implies $dy=-dx$ yields the correct sign (+) for the second integral. **Y may take positive values $\{Y: 0<Y<2, Y\neq 1\}$ and this is no problem for the gamma function. For example $\Gamma\left(-\frac{3}{2}\right)=\frac{4}{3}\sqrt{\pi}$, but $\Gamma(-1)=(-2)!=\infty$.

Asset price dynamics under the CGMY Model can now be constructed as:
$$S(t) = S(0)\exp\big((\mu+\omega)t + X_{CGMY}(t;C,G,M,Y)\big). \quad (2.106)$$
Here μ is the mean rate of return for the stock and ω is a known as a "convexity" correction, the value of the convexity correction is given by (remembering that $i=\sqrt{-1}$):
$$\exp(-\omega t) = \phi_{CGYM}(-i,t;C,G,M,Y). \quad (2.107)$$
Equation (2.106) can be re-formulated to include a diffusion term as follows:
$$X_{CGMY_e}(t;C,G,M,Y) = X_{CGMY}(t;C,G,M,Y) + \eta W(t). \quad (2.108)$$
As expected, $W(t)$ is a standard Brownian motion independent of $X_{CGMY}(t;C,G,M,Y)$.
Thus (2.106) can be re-written as:
$$S(t) = S(0)\exp\left(\left(\mu+\omega-\frac{\eta^2}{2}\right)t + X_{CGMY_e}(t;C,G,M,Y)\right). \quad (2.109)$$
For most of the derivative pricing that will be done in Chapters 3 and 4, it is desirable to use the characteristic function of the logarithm of the stock price in (2.109) (i.e. $X(t)=\ln\big(S(t)\big)$. This is given by:
$$\phi_{X(t)}(u) = \exp\left(iu\left\{X(0)+\left(\mu+\omega-\frac{\eta^2}{2}\right)t\right\}\right)\times\phi_{CGMY}(u,t;C,G,M,Y)\exp\left(-\frac{\eta^2\mu^2}{2}\right). \quad (2.110)$$

There is a final improvement to the VG Model that can be made – to model time stochastically. The reason for this improvement is that Lêvy processes exhibit the Floating Smile Property (see (Kienitz and Wettarau, 2012)) which basically means that if $IV_0\left(\frac{K}{S(t)},T\right)$ denotes the implied volatility surface (moneyness and time to maturity) at time $t=0$ then $\forall\ t\in[0,T]$:
$$IV_t\left(\frac{K}{S(t)},T\right) = IV_0\left(\frac{K}{S(t)},T\right).$$
This simply means that the future shape of the implied volatility surface is the same as the shape of the surface today. As an example, the implied volatility surfaces (as a function of K) for the S&P500

call options, in Chapter 1, calculated on 26 Feb 2016 and 29 Feb 2016 look completely different.

The improvement is done via stochastic clocks. One implicit assumption behind all the models considered so far is that time moves according to calendar time. This assumption can be relaxed by modelling time stochastically according to business day activity. For instance during a normal trading day there are periods of high trading activity (e.g. following a major news event) and periods of low trading activity. Naturally this impacts the volatility of the underlying asset. In fact this improvement is equivalent to inducing stochastic volatility into the model. An easy way to think of this is that, instead of the usual process $X(t)$, time is modelled to give the process $X\big(Y(t)\big)$, where $Y(t)$ is the stochastic process modelling time.

Interestingly, the authors of the CGMY Model first introduced this improvement in the article, "Stochastic Volatility For Lêvy Processes", see (Carr et al, 2003). First, consider a slightly extended version of (2.107) which is given by (where C_p, Y_p are parameter values for $x > 0$ and C_n, Y_n for the case $x < 0$):

$$\phi_{CGMY}(u,t;C,G,M,Y)$$
$$= \exp\Big(t\big((C_p\Gamma\big(-Y_p\big)[(M-iu)^{Y_p} - M^{Y_p}] + C_n\Gamma(-Y_n)[(G+iu)^{Y_n} - G^{Y_n}]\big)\Big). \quad (2.111)$$

Later the following special case of (2.111) will be needed ($t = 1, \xi = C_n/C_p$), the log characteristic function:

$$\psi_{CGMY}\big(u,1;C_n,G,M,Y_p,Y_n,\xi\big)$$
$$= C_p\big(\Gamma\big(-Y_p\big)[(M-iu)^{Y_p} - M^{Y_p}] + \xi\Gamma(-Y_n)[(G+iu)^{Y_n} - G^{Y_n}]\big).$$

The stochastic clock considered here is based on the CIR process. Keeping the notation consistent with (Carr et al, 2003), a slightly different version of (2.28) is defined by

$$dy(t) = \kappa\big(\eta - y(t)\big)dt + \lambda\sqrt{y(t)}dW(t). \quad (2.112)$$

Time moves forward and is clearly not mean-reverting, this implies that an "integrated" version of (2.112) is needed, which is formally given by

$$Y(t) = \int_0^t y(u)du. \quad (2.113)$$

The characteristic function of $Y(t)$, the integrated process, is known:

$$\phi_Y(u,t,y(0);\kappa,\eta,\lambda) = A(t,u)\exp\big(B(t,u)y(0)\big), \quad (2.114)$$

$$A(t,u) = \frac{\exp\left(\frac{\kappa^2\eta t}{\lambda^2}\right)}{\left(\cosh\left(\frac{\gamma t}{2}\right) + \frac{\kappa}{\gamma}\sinh\left(\frac{\gamma t}{2}\right)\right)^{\frac{2\kappa\eta}{\lambda^2}}},$$

$$B(t,u) = \frac{2iu}{\kappa + \gamma\coth\left(\frac{\gamma t}{2}\right)},$$

$$\gamma = \sqrt{\kappa^2 - 2\lambda^2 iu}.$$

As mentioned, the objective is to incorporate a stochastic model of time into the Lêvy process. Generally, a stochastic volatility Lêvy process (SVLP) is defined by

$$Z(t) = X\big(Y(t)\big). \quad (2.115)$$

Here $X(t)$ is a Lêvy process and $Y(t)$ the integrated version of some mean-reverting process, commonly either an Ornstein-Uhlenbeck or CIR process. The characteristic function of (2.115) is available, although the precise construction details are omitted here. Generally,

$$E\big[\exp\big(iuZ(t)\big)\big] = E\big[\exp\big(Y(t)\psi_X(u)\big)\big]$$
$$= \phi_Y(-i\psi_X(u),t,y(0);\kappa,\eta,\lambda). \quad (2.116)$$

The CGMY Model with a CIR stochastic clock, the CGMYSV Model, effectively admits nine parameters: $C_p, G, M, Y_p, Y_n, \xi, \kappa, \eta, \lambda$, additionally we let $y(0) = C_p$ and thus the characteristic function is

$$E\left[\exp\left(iuZ_{CGMYSV}(t)\right)\right] = \phi_Y\left(-i\psi_{CGMY}\left(u, 1; G, M, Y_p, Y_n, \xi\right), t, C_p; \kappa, \eta, \lambda\right). \tag{2.117}$$

To obtain a visual reference on how the stochastic clock works the following graphs are provided below in Figure 2.4:

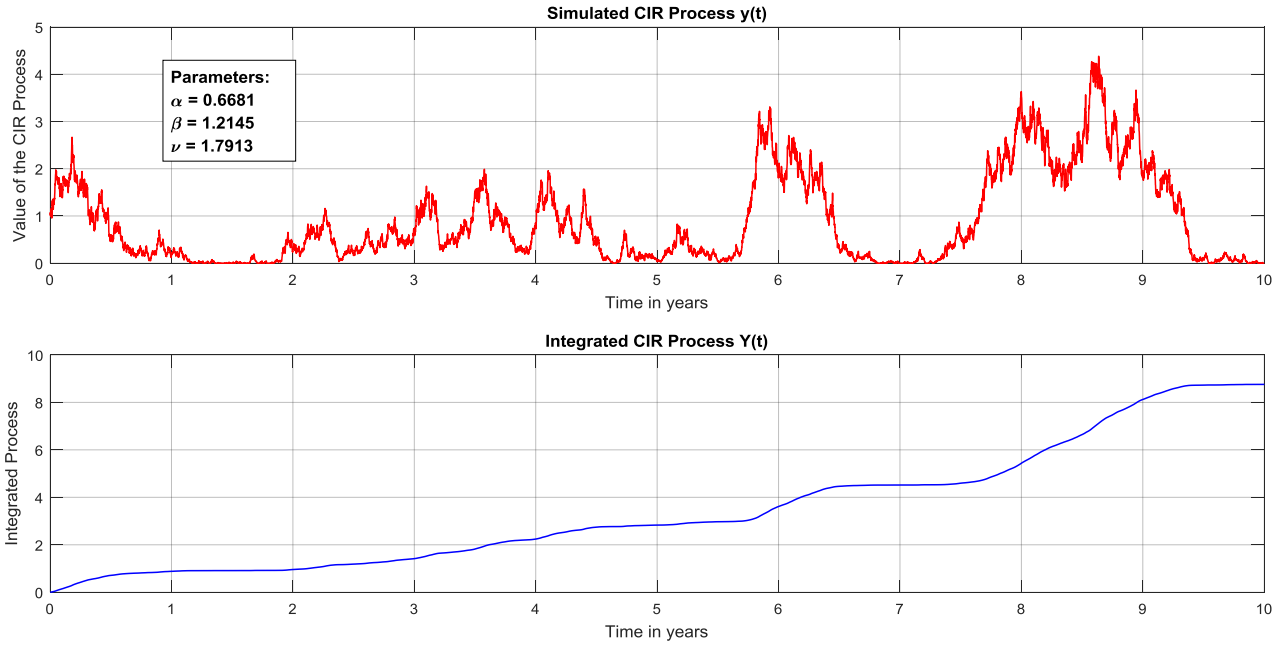

Figure 2.4: Sample path of the stochastic clock under the CIR process and the corresponding integrated path.

The top graph in Figure 2.4 gives an example of a simulated CIR process, simulated exactly according to Algorithm 3.9, which is given later in Chapter 3. The bottom graph in Figure 2.4 represents the integrated process, computed using the MATLAB® function *cumtrapz*. Implementing this version or a simple Riemann integration scheme does not lead to any significant differences in the above integrated path.

A final issue that needs to be addressed here is the question of asset price dynamics for the CGMYSV Model. This short piece is given to complete the discussion on the CGMYSV Model and to specify the form the characteristic function that can be used in practice.

Recall that for the CGMYSV Model the asset dynamics are specified by

$$S(t) = S(0)\exp\left(\left(r - q - \omega\right)t + X\left(Y(t)\right)\right). \tag{2.118}$$

One can take the convexity correction ω, which if we recall is defined by:

$$\exp(-\omega t) = \phi_{CGYM}(-i, t; C, G, Y, M),$$

and bring it into the denominator of (2.118) and hence, when adjusting for the stochastic clock, the asset price dynamics in (2.118) become, under the risk neutral measure (see (Carr et al, 2003)):

$$S(t) = S(0)\frac{\exp\left(\left(r - q\right)t + X\left(Y(t)\right)\right)}{E\left[\exp\left(X\left(Y(t)\right)\right)\right]} = S(0)\frac{\exp\left(\left(r - q\right)t + Z(t)\right)}{E\left[\exp\left(Z(t)\right)\right]}. \tag{2.119}$$

$$E\left[\exp\left(Z(t)\right)\right] = \phi_{CIR}(-i\psi_{CGYM}(-i), t, y(0); \alpha, \beta, \nu). \tag{2.120}$$

Consequently the characteristic function for the log-price at time t can be given as:

$$\varphi_{CGYMSV}(t, S(0), u) = \exp\left(iu\left(log\left(S(0)\right) + (r-q)t\right)\right) * \frac{\phi_{CIR}(-i\psi_{CGYM}(u), t, y(0); \alpha, \beta, \nu)}{\phi_{CIR}(-i\psi_{CGYM}(-i), t, y(0); \alpha, \beta, \nu)^{iu}}. \quad (2.121)$$

Other models with jumps are available to price options, i.e. NIG Model (see (Barndorff-Nielsen, 1995)), the Kou double-exponential Model (Kou and Wang, 2004) and the Generalized Hyperbolic Model (Prause, 1999). They will however not be treated here. Furthermore, the CGMYSV Model was shown in the article by (Carr et al, 2003) to be the best performing model out of all the other SVLP (Stochastic Volatility Lêvy Process) models, these other models include among them a NIG process with various stochastic clocks incorporated.

2.3.5 **A stochastic volatility model based on α-stable distributions**

There are even further improvements that can be made to the above CGMYSA Model (for example, the Lêvy processes can be turned into fractional Lêvy processes, see (Bender et al, 2011) for a technical discussion on some of the properties for finite fractional Lêvy processes). However, the CGMYSV Model presented in the previous section will not be improved on further. Instead, in this section we will discuss the literature regarding Stable Distributions, which takes a different direction to that of Lêvy processes in the previous subsection.

The material here is reworked from (Kleinert and Borbel, 2016), (Carr and Wu, 2003), and notably (Nolan, 2013). The reader should note the distinction between Lêvy Processes and α-Stable (Lêvy) distributions, the former is characterised by a jump measure and the latter admits a characteristic function with four parameters (see below).

The general form of an α-stable distribution (X) allows for a four-parameter family, α, β, c, μ which is expressed through the characteristic function as follows (see (Nolan, 2013)), parametrisation given as under definition 1.7):

$$\phi(\alpha, \beta, c, \mu)(p) = \begin{cases} \exp\left(-c^{\alpha}|p|^{\alpha}\left[1 + i\beta\left(tan\left(\frac{\pi\alpha}{2}\right)sign(p)\right)\left(|cp|^{1-\alpha} - 1\right)\right] + i\mu p\right) & if \quad \alpha \neq 1 \\ \exp\left(-c|p|\left[1 + i\beta\frac{2}{\pi}\left(sign(p)\right)log(c|p|)\right] + i\mu p\right) & if \quad \alpha = 1 \end{cases} \quad (2.122)$$

Recall the $sign(p)$ function which simply returns the sign of the input number:

$$sign(p) = \begin{cases} -1 \; if \; p < 0 \\ 0 \quad if \; p = 0 \\ 1 \quad if \; p > 0 \end{cases}$$

Each of the parameters in (2.122) may take values in the following ranges: $\alpha \in (0,2], \beta \in [-1,1]$, $c \geq 0$ and $\mu \in \mathbb{R}$. A quick remark on the interpretation of each of the parameters and their effect on the resulting distributions: μ is known as the location parameter, c is the scaling parameter of the distribution, α is known as the characteristic exponent and lastly β is the skewness parameter. Interestingly, if $\alpha > 1$ then μ is equal to the expected value of the distribution. The following formula defines the support for the density function:

$$support\ f(x|\alpha, \beta, c, \mu) = \begin{cases} \left[\mu - c \times tan\left(\frac{\pi\alpha}{2}\right), \infty\right) & if\ \alpha < 1\ and\ \beta = 1 \\ \left(-\infty, \mu + c \times \tan\left(\frac{\pi\alpha}{2}\right)\right] & if\ \alpha < 1\ and\ \beta = -1 \\ (-\infty, \infty) & otherwise \end{cases}$$

(2.123)

In what follows two parameters are fixed: c & μ. When referring to a specific distribution, the notation $L_{\alpha,\beta}(x)$ will be used, keeping in mind that the location and scale parameters have been fixed (this will be specified later). Otherwise the full parameter set is specified using the notation $L_{\alpha,\beta,c,\mu}(x)$ or the notational equivalent $L_{\alpha}(x;\beta,c,\mu)$. The distribution and/or density function for $L_{\alpha,\beta,c,\mu}(x)$ is generally unknown and thus formulas (2.10) and (2.11) are generally used to determine these numerically. The following table gives values for which $L_{\alpha,\beta,c,\mu}(x)$ admits a known density function.

Known Distributions:

Distribution	Alpha Value	Beta Value	C Value	Mu Value
Normal (0,1)	2	0	$1/\sqrt{2}$	0
Lêvy (0,1)	0.5	1	1	1
Cauchy(0,1)	1	0	1	0

Table 2.2: α-Stable distributions with parameter values that imply known density functions

For the distributions given in Table 2.2, the following density functions can be plotted for reference:

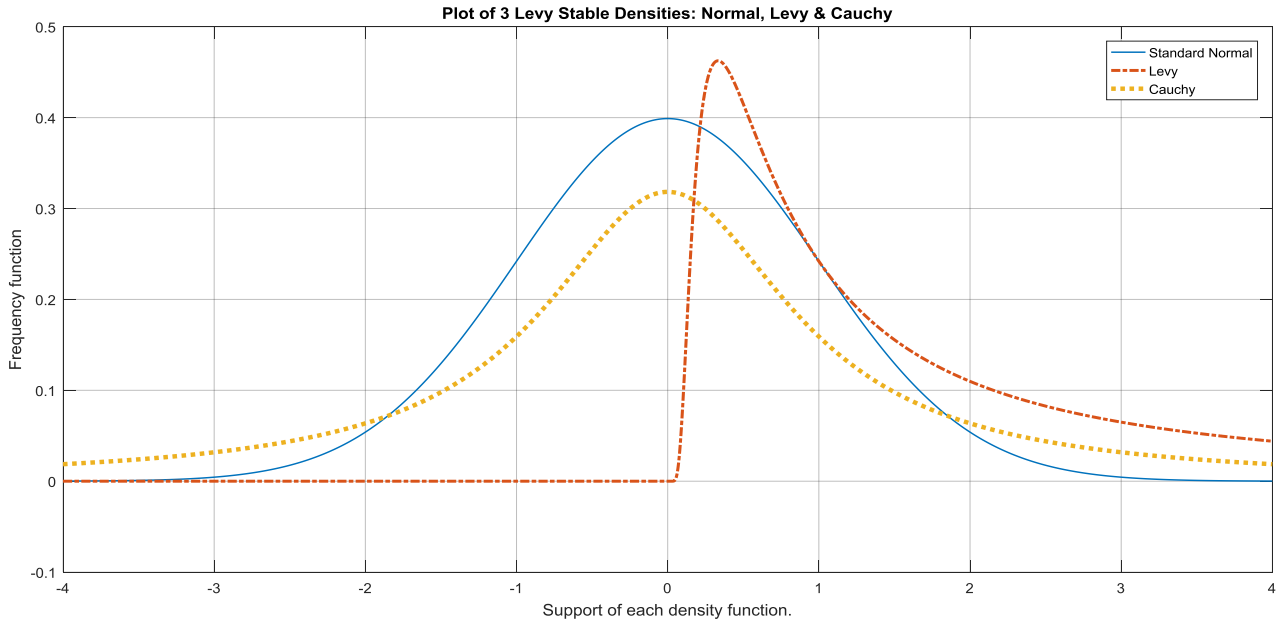

Figure 2.5: Plot of the density functions of three α-stable distributions.

**Some technical remarks: Strictly speaking (2.122) is a specific parametrization of the characteristic function of a stable distribution. The reader should consult (Nolan, 2013) for a comprehensive discussion of the various parametrizations. Following his recommendations, the form given by (2.122) is suitable for numerical work and continuous in the parameter α.

α-Stable distributions (without reference to characteristic functions) are generally defined as any distribution (non-degenerate) X for which if $n > 1$ there exists constants $c_n > 0$ & $d_n \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \overset{d}{=} c_n X + d_n.$$

here each of the $X_i's$ are i.i.d. random variables and each X_i has the same distribution as X. A familiar example is the sum of i.i.d. Normal random variables being again normally distributed. All distributions in the above table have this property. Some additional properties of α-stable distributions are given below:

1.) (Zolotarev, 1986) Suppose $\alpha < 2$, the tail behavior of an α-stable distribution X exhibits the following probabilistic behavior:

$$\lim_{\lambda\to\infty} \lambda^{\alpha} P(x > \lambda) = C_{\alpha}\frac{1+\beta}{2}c^{\alpha}, \qquad \lim_{\lambda\to\infty} \lambda^{\alpha} P(x < -\lambda) = C_{\alpha}\frac{1-\beta}{2}c^{\alpha}.$$

for some parameter C_{α}. In the following sections it is convenient to work with $\beta = -1$ in financial work (Carr and Wu, 2003) and hence the tail behavior on the right-side decays exponentially fast (faster than $\lambda^{-\alpha}$, resulting in a thin tail):

$$P(x > \lambda) \sim \frac{1}{\sqrt{2\pi\alpha(\alpha-1)}}\left(\frac{\lambda}{\alpha\hat{\sigma}_{\alpha}}\right)^{-\alpha/(2(\alpha-1))} \exp\left(-(\alpha-1)\left(\frac{\lambda}{\alpha\hat{\sigma}_{\alpha}}\right)^{-\frac{\alpha}{(\alpha-1)}}\right).$$

for some parameter $\hat{\sigma}_{\alpha}$.

2.) Let $\alpha \in (0,2)$, it follows that if $0 < p < \alpha$ then $E[|x|^P] < \infty$ and if $p > \alpha$ then $E[|x|^P] = \infty$. (Samorodnitsky and Taqqu, 1994)

**End of technical remarks.

A further example of a stable distribution (with an unknown density function) with $\alpha = \frac{\pi}{3.5}, \beta = \frac{1}{6}, c = 4, \mu = 10$ and support truncated into the range $x \in [-10,30]$ is given below:

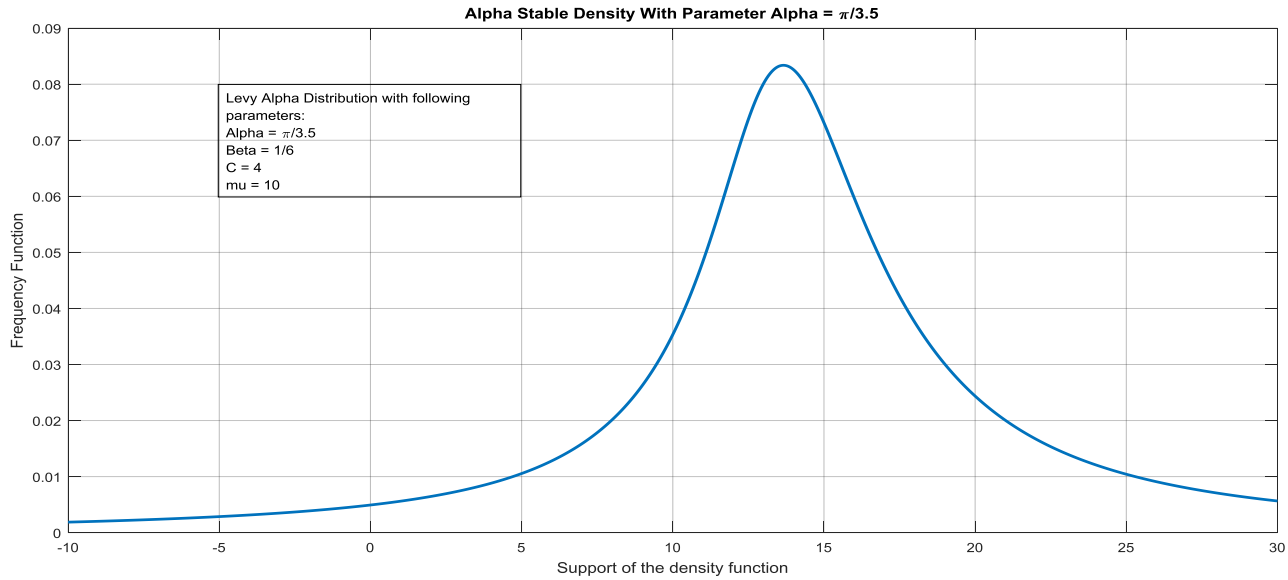

Figure 2.6: Plot of a custom α-stable density function.

Ideally one would like to use Stable distributions so that the Brownian motion term, $dW(t)$, which is normally distributed, can be replaced by a process, say $dL_{\alpha,\beta}(t)$ which would then be "$L_{\alpha,\beta}$" distributed (with fixed parameter c, μ). Formally (Carr and Wu, 2003), the α-stable process, which is a particular type of Lêvy process, is defined as a stochastic process with independent and stationary increments such that the increment $dL_{\alpha,\beta}(t)$ has an α-stable distribution with zero drift ($\mu = 0$), scale parameter of $c = dt^{\frac{1}{\alpha}}$, i.e. $L_{\alpha,\beta,t^{1/\alpha},0}$. The suggestion by (Carr and Wu, 2003) is that $\beta = -1$ and $\alpha \in (1,2)$.

Keeping these considerations in mind, let $S(t)$ be the underlying index with state space $\Omega = \mathbb{R}^{+}$ over some time interval $t \in (0,T]$, $\{F(t)\}_{t\geq 0}$ the natural filtration associated with process and let $\mathbb{Q}$ be the corresponding risk neutral probability measure such that the process $S(t)$ obeys the following SDE:

$$\frac{dS(t)}{S(t)} = (r-q)dt + \sigma dL_{\alpha,-1}(t). \qquad (2.124)$$

here the constant dividend rate q is incorporated into the process and may be zero. The model in (2.124) is known as the FMLS (Finite Moment Log Stable) Model. The following results are taken from (Carr and Wu, 2003), altered slightly for notational consistency:

1.) The log return $X(\tau) = \log\left(\frac{S_T}{S_t}\right)$ over some period $\tau = T - t$ has an α-stable distribution $L_\alpha\left(-1, \sigma\tau^{\frac{1}{\alpha}}, \mu + r - q\right)$ with $\mu = \sigma^\alpha \sec(\pi\alpha/2)$.

2.) For the asset process S_T the p^{th} conditional moments are well defined:

$$E_{\mathbb{Q}}[S^p(T)|\mathcal{F}(t)] = S^p(t) \times \exp\left(p(\mu + r - q) \times \tau - \tau(p\sigma)^\alpha \, sec\left(\frac{\pi\alpha}{2}\right)\right) < \infty, \forall\, p \in [0, \infty).$$

(Dufresne, 2009), amongst many others, suggest that since the moments of $X(t)$ are infinite (for $p > \alpha$), as was given a bit earlier, this approach of using an α-stable process is not justified as the resulting options formula may give infinite option prices. However, as is seen above, the asset process has finite moments and this is essential for generating finite option prices and the existence for equivalent martingale measures.

The problem is not that the underlying process generates moments of infinite magnitude, but whether the asset price process generates finite moments. (An example in literature where infinite higher moments are used is given by (Meintanis and Taufer, 2011) where they apply stable-Paretian stochastic volatility modelling techniques to exchange rates).

The model in (2.124) is not a stochastic volatility model, but essentially a new diffusion process that replaces the Brownian motion component in the BS Model. One can price options using this model by an appropriate characteristic function (2.122) and the formulas in Chapter 3, which essentially requires two numerical integration procedures.

The question of using stochastic volatility coupled with α-stable processes has been addressed by (Carr and Wu, 2003), and has not received much further attention in literature. They only give the characteristic function, which will be used later for our analysis, however pricing under this approach has not been popular in literature. We begin with equation (2.124) and incorporate a CIR process to drive the underlying volatility as follows:

$$\frac{dS(t)}{S(t)} = (r-q)dt + \big(V(t)\big)^{\frac{1}{\alpha}} dL_{\alpha,-1}(t)$$
$$dV(t) = \kappa\big(\theta - V(t)\big)dt + \sigma_V\sqrt{V(t)}dW(t).$$

The *log-return* characteristic function is can then be defined:

$$\phi_{X(\tau)}(u) = \exp\big(iu(r-q)\tau - b(\tau)V(t) - c(\tau)\big).$$

The auxiliary functions are:

$$b(\tau) = \frac{2\lambda(1 - e^{-\eta\tau})}{2\eta - (n-\kappa)(1-e^{-\eta\tau})}$$
$$c(\tau) = \frac{\kappa\theta}{\sigma_V^2}\left[2\ln\left(1 - \frac{\eta-\kappa}{2\eta}(1-e^{-\eta\tau})\right) + (\eta-\kappa)\tau\right]$$
$$\eta = \sqrt{(\kappa)^2 + 2\sigma_V^2\lambda}, \qquad \lambda = [(iu)^\alpha - iu]\sec\left(\frac{\pi\alpha}{2}\right).$$

It should be highlighted that in order to use this model we need to specify the log-return payoff function which is simply:

$$Payoff\ Call = S_0\left(e^{\log\frac{S(t)}{S(0)}} - e^{\log\frac{K}{S(0)}}\right).$$

This model will not be used in the work to come later however we give this model to complete the literature review on using alternative asset price dynamics to model asset price behaviour. The results in this section can thus be used by practitioners as a starting point for introducing models with alternative asset price dynamics.

2.3.6 **A double-fractional diffusion based Black-Scholes Model**

To put the contents of sections 2.3.1-2.3.5 into perspective, the goal here was to give a portion of the available literature regarding alternative stochastic processes that can drive the differential equations used to give asset prices and hence price derivatives. For some of these alternative processes one can build stochastic volatility models with interesting features.

Below a different pricing approach can be taken by considering the double-fractional diffusion process, which essentially further extends the model given in section 2.3.5 and completes our literature review on alternative asset price dynamics. However, before continuing, a few pieces of background material is provided in Appendix B, under section B.2.2. The material is essentially an introduction into Fractional Calculus and it is highly recommended that it be read before continuing. One reference is also made to the material in Appendix B.

Essentially the work of (Kleinert and Kobel 2016) asks whether the time derivative in the PDE (2.127), below, can be replaced by an appropriate fractional derivative. The motivation for this is that the resulting diffusion equation, the double-fractional diffusion equation (DFD), can allow for more realistic option pricing and represents an interesting advancement in the field of derivative pricing.

Now by continuing the discussion on α-stable Distributions, we have all the ingredients to obtain a fractional Black-Scholes equation for log-prices, which is given by (Kleinert and Kobel, 2016). Consider again the following model for asset prices in (2.124), slightly altered:

$$S(t) = S(0)\exp\left((r+\mu)t + \sigma L_{\alpha,-1}\right). \tag{2.125}$$

μ is defined as $\mu = \sigma^{\alpha}\sec(\pi\alpha/2)$. Using this model the price of a call option is

$$C(K,\tau) = e^{-r\tau}\int_{-\infty}^{\infty}\max(S(t)e^{r\tau+x},0)\int_{-\infty}^{\infty}\frac{e^{-ipx}}{2\pi}e^{(\tau)[ip\mu+\sigma^{\alpha}\times\phi(\alpha,-1,1,0)(p)]}dpdx. \tag{2.126}$$

As usual, $\tau = T - t$. Making a change of variables, $z = \log\big(S(t)\big)$ and $y = x + r\tau + z$, leads to:

$$\begin{aligned} C(z,\tau) &= \mathrm{e}^{-\mathrm{r}\tau}\int_{-\infty}^{\infty}\max(e^{y}-K,0)\int_{-\infty}^{\infty}\frac{e^{ip(z-y)}}{2\pi}e^{\tau[ip(r+\mu)-\mu(ip)^{\alpha}]}dpdy \\ &= \int_{-\infty}^{\infty}\max(e^{y}-K,0)\,\tilde{g}_{\alpha}(z,\tau|y)dy. \end{aligned}$$

Here $\tilde{g}_{\alpha}(z,\tau|y)$ is effectively a Green function and it is defined as the integral in the first line as can be seen. Next, define $g_{\alpha}(\xi,\tau) = e^{r\tau}\tilde{g}_{\alpha}(z,\tau|y)$, change variables $\xi = z - y + \tau(r+\mu)$ and take two types of derivatives in $g_{\alpha}(\xi,\tau)$, one ordinary partial derivative in τ and secondly apply the differential operator in (B.2.2.13), Appendix B, to obtain (with initial condition $g_{\alpha}(\xi,\tau) = \delta(\xi)$):

$$\frac{\partial g_{\alpha}(\xi,\tau)}{\partial\tau} = -\mu\big[{}^{\alpha-2}D_{\xi}^{\alpha}g_{\alpha}\big](\xi,\tau). \tag{2.127}$$

This equation is defined as the Fractional Black-Scholes equation and it can be shown by that if $\alpha = 2$, the classical BS diffusion equation is obtained.

The basic idea is that (2.127) has a time derivative that can either be taken as the Riesz-Feller or Caputo fractional derivative, however this means that $g_\alpha(\xi,\tau)$ has to be replaced in (2.127) by some other Green function as $g_\alpha(\xi,\tau)$ will then not be a solution to the fractional differential equation anymore. This new Green function is denoted by $g(\xi,\tau)$ and the fractional differential equation is given by (Kleinert and Kobel, 2016):

$$\left({}^{K}\partial_\tau^\gamma + \mu\left[{}^{\alpha-2}D_\xi^\alpha\right]\right)g(\xi,\tau) = 0. \tag{2.128}$$

The K can denote either of the two fractional derivatives, i.e. ${}^{C}\partial_\tau^\gamma = {}_0^*D_\tau^\gamma$ for the Caputo derivative or ${}^{RF}\partial_\tau^\gamma = D_\tau^\gamma$ for the Riesz-Feller derivative. The parameter γ corresponds to the parameter ν when taking fractional derivatives, i.e. the "γ^{th}" time derivative. The derivation of all the results is too involved to be introduced here, and hence a summary of the solution to (2.128) is given. The Green function $g(\xi,\tau)$ has the following representation depending on the fractional derivative taken in (2.128):

$$g^{RF}(\xi,\tau) = \int_0^\infty \left(\frac{\Gamma(\gamma)}{\tau^{\gamma-1}}\right)\frac{1}{l^{\frac{1}{\gamma}}}L_{\gamma,1}\left(\frac{\tau}{l^{\frac{1}{\gamma}}}\right)g_\alpha(\xi,\tau)dl. \tag{2.129}$$

$$g^{C}(\xi,\tau) = \int_0^\infty \left(\frac{\tau}{l\gamma}\right)\frac{1}{l^{\frac{1}{\gamma}}}L_{\gamma,1}\left(\frac{\tau}{l^{\frac{1}{\gamma}}}\right)g_\alpha(\xi,\tau)dl. \tag{2.130}$$

Here $L_{\gamma,1}$ is a α –stable distribution, i.e. the distribution function of $L_{\gamma,1,1,0}(x)$ (defined earlier). Unfortunately formulas (2.129) & (2.130) only apply when $0 < \gamma < 1$ and a more general formula can be found that holds even for $\gamma > 1$:

$$g^{DF}(\xi,\tau) = \frac{\Gamma(\kappa)}{2\alpha\pi i\xi}\int_{c-i\infty}^{c+i\infty}\frac{\Gamma\left(\frac{s}{\alpha}\right)\Gamma\left(1-\frac{s}{\alpha}\right)\Gamma(1-s)}{\Gamma\left(\kappa-\frac{\gamma}{\alpha}s\right)\Gamma\left(\frac{(\alpha-\theta)s}{2\alpha}\right)\Gamma\left(1-\frac{(\alpha-\theta)s}{2\alpha}\right)}\left[\frac{\xi}{(-\mu\tau^\gamma)^{\frac{1}{\alpha}}}\right]^s ds. \tag{2.131}$$

Some remarks about this formula:

- $\kappa = \gamma$ for the RF derivative and $\kappa = 1$ for the Caputo derivative.
- For all three formulas (2.129)-(2.131) we require $0 < \gamma < \alpha \leq 2$ if $\gamma > 1$ otherwise $\alpha \in (0,2]$.
- θ ,by the suggestion of (Kleinert and Kobel, 2016), is left to be left undetermined, although values of $\theta = 0$ or $\theta = \alpha - 2$ seem to be permissible.
- Lastly, the number c can be any number in the interval $(0,1)$.

Using (2.131), the call option formula in (2.118) can be re-expressed as (allowing for dividends):

$$C_{\alpha,\gamma,\kappa}(S(t),K,\tau) = e^{-r\tau}\int_{-\infty}^{\infty}\max\left(S(t)e^{\tau(r-q+\mu)+y} - K, 0\right)g^{DF}(y,\tau)dy\,. \tag{2.132}$$

Lastly, a stochastic volatility extension using the DFD process has not been considered so far in literature, given that the DFD process is relatively new. It is hoped that by presenting this material that further research can be taken into incorporating stochastic volatility and stochastic interest rates into the DFD pricing methodology.

It would be difficult to use mixes of stochastic processes in comparing stochastic volatility models as further considerations have to be given to simulation schemes and other pricing methods. This would take up too much space here and make the analysis much too complicated. Hence, going forward, we will work with BM as the standard stochastic process that drives the SDE's of some of the models. Thus our focus lies in using stochastic volatility models with BM diffusions.

This brings us to the next section, where recent advances in stochastic volatility models with BM

diffusions are considered.

2.4 **Advanced models**

This section considers the latest developments of models within the last 2 years. The first model, the 4/2 Model, lends itself to a bit of a technical discussion and will be used for further analysis later. The Jacobi Model is then given briefly and will not be used for further analysis, however a detailed discussion of this model can be found in Appendix B.

2.4.1 **4/2 Model**

This model is due to (Grasselli, 2015). Consider a non-dividend paying asset with the following asset price process:

$$\frac{dS(t)}{S(t)} = rdt + \left(a\sqrt{V(t)} + \frac{b}{\sqrt{V(t)}}\right)dW_S(t). \tag{2.133}$$

$$dV(t) = \kappa\big(\theta - V(t)\big) + \sigma\sqrt{V(t)}\,dW_V(t). \tag{2.134}$$

The constant parameters are $r, \kappa, \theta, \sigma \in \mathbb{R}^+$ and $a, b \in \mathbb{R}$ and it is assumed that $2\kappa\theta \geq \sigma^2$. Lastly the Brownian motions are assumed to be correlated, i.e. $dW_S(t)dW_V(t) = \rho dt$. The Heston Model is recovered by taking $b = 0$ and the 3/2 Model by taking $a = 0$, using the substitution $V'(t) = \frac{1}{V(t)}$ and substituting $\tilde{\kappa} = \kappa\theta - \sigma^2, \tilde{\theta} = \frac{\kappa}{\kappa\theta - \sigma^2}, \tilde{\sigma} = -\sigma$ for each parameter κ, θ, σ. Effectively a linear combination of a CIR and a 3/2 volatility component (1/2 + 3/2 = 4/2) drives the stochastic volatility in (2.133).

Deriving the appropriate log-price characteristic function for (2.133) is difficult and depends on results from Lie Algebra. (Grasselli, 2015) shows how this construction can be achieved and hence a summary of the end results is given below. The characteristic function given here is given in a slightly more general form. The generalised log-price characteristic function is:

$$\phi(t, T, u) = E_t\big[e^{uX(t)}\big], u \in \mathcal{D}_{t,T} \subseteq \mathbb{C}. \tag{2.135}$$

Here $\mathcal{D}_{t,T}$ is the domain for which the function is well-defined. An explicit formula for (2.135) is given by

$$\begin{aligned}\phi(0,t,u) = \exp\left\{uX(0) + \frac{\kappa^2\theta}{\sigma^2}t + u\left(r - ab - \frac{a\rho\kappa\theta}{\sigma} + \frac{b\rho\kappa}{\sigma}\right)t + u^2(1-\rho^2)abt\right\}\\ \times\left(\frac{\sqrt{A_u}}{\sigma^2\sinh\left(\frac{\sqrt{A_u}}{2}t\right)}\right)^{m_u+1}\quad (V(0))^{\frac{1}{2}+\frac{m_u}{2}-\frac{\kappa\theta}{\sigma^2}-\frac{ub\rho}{\sigma}}\left(K_u(t) - \frac{ua\rho}{\sigma}\right)^{-\left(\frac{1}{2}+\frac{m_u}{2}+\frac{\kappa\theta}{\sigma^2}+\frac{ub\rho}{\sigma}\right)}\\ \times\exp\left\{\frac{V(0)}{\sigma^2}\left(-\sqrt{A_u}\coth\left(\frac{\sqrt{A_u}}{2}t\right) + \kappa - ua\rho\sigma\right)\right\}\frac{\Gamma\left(\frac{1}{2}+\frac{m_u}{2}+\frac{\kappa\theta}{\sigma^2}+\frac{ub\rho}{\sigma}\right)}{\Gamma(m_u+1)}\\ \times {}_1F_1\left(\frac{1}{2}+\frac{m_u}{2}+\frac{\kappa\theta}{\sigma^2}+\frac{ub\rho}{\sigma}, m_u+1, \frac{A_uV(0)}{\sigma^4\sinh^2\left(\frac{\sqrt{A_u}}{2}t\right)\left(K_u(t)-\frac{ua\rho}{\sigma}\right)}\right).\end{aligned} \tag{2.136}$$

The last line in (2.136) has the confluent hypergeometric function, introduced earlier under the H2-HW Model. The auxiliary functions in (2.136) are as follow:

$$A_u = \kappa^2 - 2\sigma^2\left(u\left(\frac{a\rho\kappa}{\sigma} - \frac{1}{2}a^2\right) + \frac{1}{2}u^2(1-\rho^2)a^2\right). \tag{2.137}$$

$$m_u = \frac{2}{\sigma^2}\sqrt{\left(\kappa\theta - \frac{\sigma^2}{2}\right)^2 - 2\sigma^2\left(u\left(\frac{b\rho}{\sigma}\left(\frac{\sigma^2}{2} - \kappa\theta\right) - \frac{b^2}{2}\right) + \frac{u^2}{2}(1-\rho^2)b^2\right)}. \tag{2.138}$$

$$K_u(t) = \frac{1}{\sigma^2}\left(\sqrt{A_u}\coth\left(\frac{\sqrt{A_u}}{2}t\right) + \kappa\right). \tag{2.139}$$

There are some technical remarks and a few details behind the proof of (2.136) that need to be highlighted.

**A technical discussion now follows.

The function in (2.136) is well-defined for all $t \geq 0$ provided $u \in \mathcal{D}_{0,+\infty}$. This strip can be decomposed as $\mathcal{D}_{0,+\infty} = \mathcal{A}_{0,+\infty} + i\mathbb{R} \in \mathbb{C}$. Since our considerations are based on pricing options with finite time to expiry it follows that $u \in \mathcal{D}_{0,t}$ and hence the decomposition is instead given by $\mathcal{D}_{0,t} = \mathcal{A}_{0,t} + i\mathbb{R} \in \mathbb{C}$. Formally, the convergence set $\mathcal{A}_{0,t}$ is specified as:

$$\mathcal{A}_{0,t} = \{u \in \mathbb{R} : A_u \geq 0 \; and \; f_i(u) \; satisfy \; (2.140) - (2.143)\}$$

$$f_1(u) = u\left(\frac{a\rho\kappa}{\sigma} - \frac{1}{2}a^2\right) + \frac{1}{2}u^2(1-\rho^2)a^2 - \frac{\kappa^2}{2\sigma^2} \leq 0. \tag{2.140}$$

$$f_2(u) = \frac{\kappa\theta}{\sigma^2} + \frac{1}{2} + \frac{m_u}{2} + \frac{ub\rho}{\sigma} > 0. \tag{2.141}$$

$$f_3(u) = \left(\kappa\theta - \frac{\sigma^2}{2}\right)^2 - 2\sigma^2\left(u\left(\frac{b\rho}{\sigma}\left(\frac{\sigma^2}{2} - \kappa\theta\right) - \frac{b^2}{2}\right) + \frac{u^2}{2}(1-\rho^2)b^2\right) \geq 0. \tag{2.142}$$

$$f_4(u) = \sqrt{A_u} + \kappa - ua\rho\sigma \geq 0. \tag{2.143}$$

Additionally, care must be taken as the function in (2.136) is only well defined up to certain time t^* if one wishes to relax condition (2.143), i.e. allowing $f_4(u) < 0$:

$$t^* = \frac{1}{\sqrt{A_u}}\log\left(1 - \frac{2\sqrt{A_u}}{\kappa - ua\rho\sigma + \sqrt{A_u}}\right).$$

A sketch of proving the result in (2.136) is as follows:

Firstly, rewrite (2.133) in integral form, and noting that $X(t) = \log\big(S(t)\big)$:

$$X(t) = X(0) + rt - \frac{1}{2}\int_0^t\left(a\sqrt{V(s)} + \frac{b}{\sqrt{V(s)}}\right)^2 ds + \int_0^t a\sqrt{V(s)} + \frac{b}{\sqrt{V(s)}}dW_S(s).$$

This can be simplied as:

$$X(t) = X(0) + \left(r - ab + \frac{a\rho\kappa\theta}{\sigma} + \frac{b\rho\kappa}{\sigma}\right)t + \frac{a\rho}{\sigma}\big(V(t) - V(0)\big) + \frac{b\rho}{\sigma}\log\left(\frac{V(t)}{V(0)}\right)$$
$$+\left(\frac{a\rho\kappa}{\sigma} - \frac{1}{2}a^2\right)\int_0^t V(s)ds + \left(\frac{b\rho}{\sigma}\left(\frac{\sigma^2}{2} - \kappa\theta\right) - \frac{b^2}{2}\right)\int_0^t \frac{1}{V(s)}ds$$

$$+\sqrt{1-\rho^2}\int_0^t\left(a\sqrt{V(s)} + \frac{b}{\sqrt{V(s)}}\right)dW_S^{\perp}(s).$$

Here the Brownian motion in (2.133) was split into $W_V(t)$ and its orthogonal counterpart $W_S^{\perp}(s)$, in other words if $dW_S(t) = \rho dW_V(t) + \sqrt{1-\rho^2}W_S^{\perp}(s)$ then $dW_S(t)dW_V(t) = \rho dt$ since $dW_S(t)dW_S^{\perp}(s) = dW_V(t)dW_S^{\perp}(s) = 0$ and $\big(dW_V(t)\big)^2 = dt$. Additionally, the differential

$d\log V(t)$ is calculated as

$$d\log V(t) = \frac{\sigma}{\sqrt{V(t)}}dW_V(t) + \left(\frac{\kappa\theta}{V(t)} - \kappa\right)dt - \frac{\sigma^2}{2V(t)}dt.$$

Using the result above for $X(t)$ allows us to compute (2.135):

$$\begin{aligned}
\phi(0,t,u) &= E_t\left[e^{uX(t)}\right] \\
&= \exp\left\{uX(0) + u\left(r - ab - \frac{a\rho\kappa\theta}{\sigma} + \frac{b\rho\kappa}{\sigma}\right)t - \frac{ua\rho}{\sigma}V(0) - \frac{ub\rho}{\sigma}\log\big(V(0)\big)\right\} \\
&\quad \times E\left[\exp\left\{u\left(\frac{a\rho\kappa}{\sigma} - \frac{1}{2}a^2\right)\int_0^t V(s)ds + u\left(\frac{b\rho}{\sigma}\left(\frac{\sigma^2}{2} - \kappa\theta\right) - \frac{b^2}{2}\right)\int_0^t \frac{1}{V(s)}ds\right.\right. \\
&\qquad + \frac{1}{2}u^2(1-\rho^2)\left(a^2\int_a^t V(s)ds + b^2\int_0^t \frac{1}{V(s)}ds + 2abt\right) + \frac{ua\rho}{\sigma}V(t) \\
&\qquad \left.\left. + \frac{ub\rho}{\sigma}log\big(V(t)\big)\right\}\right] \\
&= \exp\left\{uX(0) + u\left(r - ab - \frac{a\rho\kappa\theta}{\sigma} + \frac{b\rho\kappa}{\sigma}\right)t + u(1-\rho^2)abt\right\} \\
&\quad \times \exp\left\{-\frac{ua\rho}{\sigma}V(0) - \frac{ub\rho}{\sigma}log\,V(0)\right\} \\
&\quad \times E\left[\big(V(t)\big)^{-\alpha}\exp\left\{-\lambda V(t) - \mu\int_0^t V(s)ds - \nu\int_0^t \frac{1}{V(s)}ds\right\}\right]. \qquad (2.144)
\end{aligned}$$

The additional constants are defined by

$$\begin{aligned}
\mu &= -u\left(\frac{a\rho\kappa}{\sigma} - \frac{1}{2}a^2\right) - \frac{1}{2}u^2(1-\rho^2)a^2 \\
\nu &= -u\left(\frac{b\rho}{\sigma}\left(\frac{\sigma^2}{2} - \kappa\theta\right) - \frac{b^2}{2}\right) - \frac{1}{2}u^2(1-\rho^2)b^2 \\
\alpha &= -\frac{ub\rho}{\sigma}, \qquad \lambda = -\frac{ua\rho}{\sigma}.
\end{aligned}$$

The last line in (2.144) is essentially where the results from Lie Algebra are needed. It can be shown that this expectation is a solution to the following PDE, see (Baldeaux and Platen, 2013):

$$u(t) = \frac{\sigma^2}{2}xu_{xx}(t) + (a - bx)u_x(t) - \left(\frac{\nu}{x} + \mu x\right)u(t), \qquad x > 0, t \geq 0$$

Solving this PDE and rearranging the terms will then yield the result in (2.136). ■

2.4.2 **Jacobi Model**

The Jacobi Model was introduced very recently (17 June 2016) by authors (Ackerer, Filipović and Pulido, 2016). The model itself is very impressive, the mathematics behind it departs from the usual characteristic function and Fourier pricing techniques and improves quite a bit on the traditional approaches. However, this requires specific knowledge related to techniques in Linear Algebra. In any case, many examples and supplementary explanations are given beyond what is found in the abovementioned article in Appendix B, should the reader wish to pursue these.

To start off, fix two parameters $0 \leq v_{min} < v_{max}$ and define the function $Q(v), v \in [v_{min}, v_{max}]$:

$$Q(v) = \frac{(v - v_{min})(v_{max} - v)}{\left(\sqrt{v_{max}} - \sqrt{v_{min}}\right)^2}.$$

Clearly $Q(v) \geq 0 \; \forall \; v \in [v_{min}, v_{max}]$, $Q(v) \leq v$ and $Q(v) = v \;\; iff \; v = \sqrt{v_{max} v_{min}}$. These inequalities imply that $0 \leq Q(v) \leq v_{max}$. The Author's *jacobiQdemo.m* script can be used to generate the following plot ($v_{min} = 0.01, v_{max} = 0.6$), which shows the function $Q(v)$ and its properties:

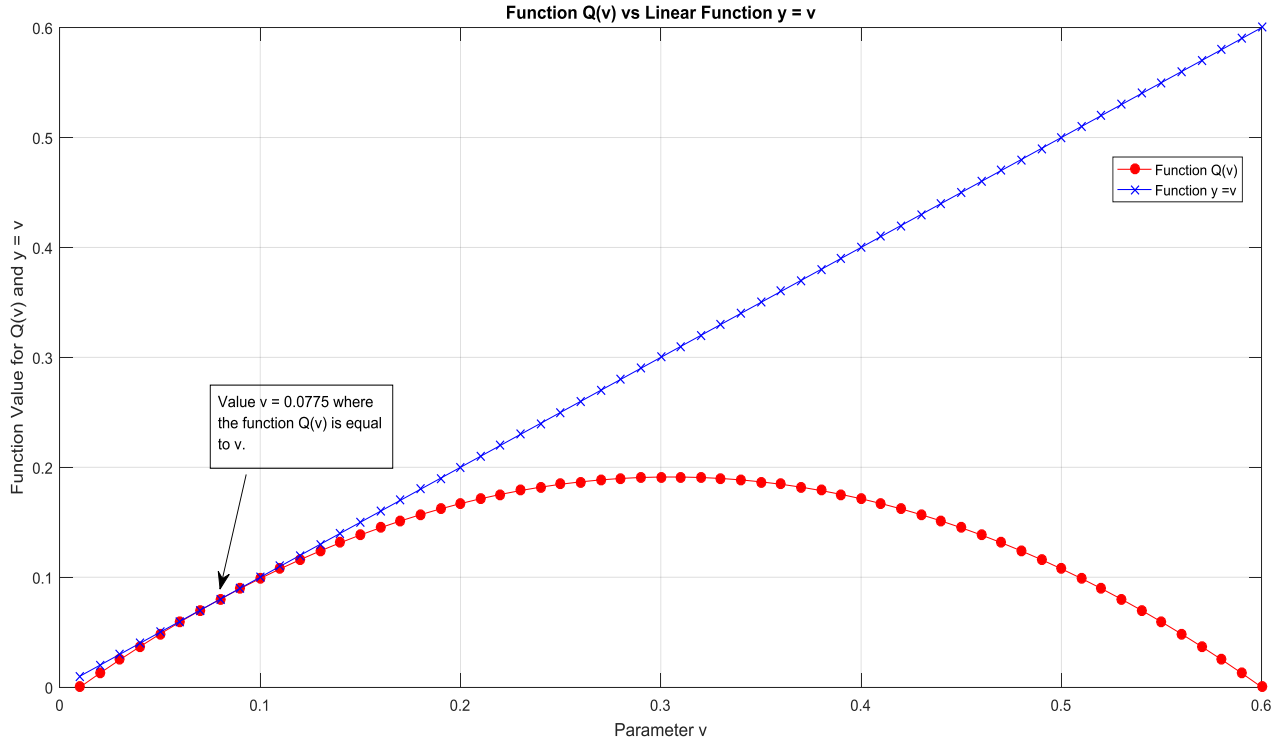

Figure 2.7: Illustration of the Q(y) function of the Jacobi Model.

Formally, the Jacobi Model is defined by the following set of SDE's:

$$dX(t) = \left(r - q - \frac{V(t)}{2}\right)dt + \rho\sqrt{Q\big(V(t)\big)}dW_V(t) + \sqrt{V(t) - \rho^2 Q\big(V(t)\big)}\, dW_S^{\perp}(t). \qquad (2.145)$$

$$dV(t) = \kappa\big(\theta - V(t)\big)dt + \sigma\sqrt{Q\big(V(t)\big)}dW_V(t). \qquad (2.146)$$

The Jacobi Model defined here is in accordance to the conventions assumed in the previous models. The constant parameters have the usual interpretation with the addition of q, constant dividend yield, which can be zero. Essentially the Jacobi transforms the instantaneous volatility $V(t)$. The reader would have noticed that by allowing $Q(v) = v \; \forall \; v \in [v_{min}, v_{max}]$ the Heston Model is recovered and that letting $V(0) = \theta = v_{max}$ recovers the BS Model.

The novel option pricing method under this model is rather complicated and is left to Appendix B, section B.2.3. Therein a full explanation of the dynamics is given by the Author which essentially completes the discussion on the Jacobi Model and all the stochastic volatility models.

2.5 **Concluding remarks**

Having worked through roughly 20 (similar and different) models in all the previous sections and subsections, the literature review on stochastic volatility models for derivative pricing is completed. The majority of the models presented were framed with respect to finding a closed form expression

for the characteristic function which serves as the basis for the pricing techniques in the next chapter. Other models that do not admit a characteristic function were presented with sufficient detail given, but without going too much in depth. This sets the stage where in the next chapter some of the models are further developed with regards to pricing derivatives.

The reader would have noticed that more emphasis was given to covering a range of models rather than going into detail. As such the focus was mathematical in nature, with many of the models being special cases of more complex models. Such a hierarchy immediately gives an implicit comparison between models and thus three models in this chapter will be considered for further development.

Lastly, in this chapter a large body of literature on stochastic volatility models has been presented. Regrettably many other models in literature, such as the Multiscale Stochastic Volatility Model, by (Fouque et al, 2011), have not represented in the interest of saving space. The Author believes however that the given classes of models do broadly represent the development of equity derivative pricing models and that these are sufficient for comparing against each other in the Chapter 4.

Chapter 3 – Literature Review: General Derivative Pricing Techniques For Stochastic Volatility Models

3.1. **Introduction to advanced derivative pricing.**

Derivative pricing has developed into a very broad and sophisticated field. There are many issues to consider, for example the form of the model (Chapter 2), suitable numerical pricing schemes, simulation, calibration and optimisation, etc. Thus the aim in this chapter is to focus primarily on suitable numerical pricing schemes and formulas for some of the models considered in the previous chapter as well as Monte Carlo simulations.

Formally, in literature there are three distinction pricing methods: Pricing via partial differential equations, Monte Carlo simulations and numerical integration methods (Fang and Oosterlee, 2008). The latter two method are our primary focus as the majority of literature is concentrated on developing techniques to either develop exact simulation schemes or fast and efficient numerical integration techniques.

It is impossible to cover all the models in Chapter 2 regarding their numerical or Monte Carlo techniques and hence the only Stochastic Volatility models considered from here on are given in the following table.

Model:	Stochastic or constant interest rate:
Double Heston	Constant
H2-CIR	Stochastic
4/2 Model	Constant

Table 3.1: Select stochastic volatility models that will be compared in Chapter 4

Each of these models covers adequate ground related to derivative pricing, the developments in literature in the past decade and the versatility of the models in capturing the implied volatility surface, as will be shown later. In particular note that each model contributes a unique feature in capturing the implied volatility surface.

The H2-CIR model's key features are using both stochastic interest rate and volatility components to describe the implied volatility surface. The 4/2 Model models the volatility component by considering a linear combination of a CIR and its inverted process. Lastly the Double Heston Model simply improves on the Heston Model by considering an additional stochastic volatility factor which allows for finer capturing of the implied volatility surface.

In the following subsections 3.1.1-3.1.4 additional background knowledge is given here that is essential to the work for the rest of this chapter. This background knowledge will also serve to fill in some of the obvious gaps or issues not addressed in the previous chapter and, e.g. risk neutral dynamics. The rest of the chapter then specifically focuses on derivative pricing and Monte Carlo techniques for the models given in the table above. Almost all the techniques reflect the

developments in literature in the past decade.

The reader should note that the techniques involved here are from a variety of mathematical disciplines. A brief overview is given although the reader should consult the references given for more detail.

3.1.1. **Advanced derivative pricing techniques part 1: Characteristic functions**

In the previous chapter reference was made to using characteristic functions to recover the risk neutral density function for a given model. All the models in our analysis admit characteristic functions and thus the methods here are valid. Indeed, a large part of the literature in derivative pricing is centred on numerical techniques for computing the risk neutral density (or transition density function) from the characteristic function.

The basic formula for a call option (at $t = 0$) under this general approach is (Heston, 1993):

$$V_{call} = S_0\Pi_1 - e^{rT}K\Pi_2. \tag{3.1}$$

Let ϕ_{X_t} denote the characteristic function of the log price X_t which is assumed known. Consequently the parameters Π_1 (which is also the Delta of the option) and Π_2 (the risk neutral probability $P(S_T > K)$) are given by:

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-iuln(K)}\phi_{X_t}(u-i)}{iu\phi_{X_t}(-i)}\right]du. \tag{3.2}$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-iuln(K)}\phi_{X_t}(u)}{iu}\right]du. \tag{3.3}$$

Often the density function is needed to price more complicated derivatives. Both the density and distribution function have general formulas that are given here, with the first formula's proof also given below:

$$Cumulative\ distribution\colon\ F(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)}{iu}du. \tag{3.4}$$

$$Density\ function\colon F'(x) = f(x) = \frac{1}{\pi}\int_0^\infty e^{-iux}\phi(u)du. \tag{3.5}$$

The proof of (3.4) is inspired by (Kendall et al, 1994) and is given below:

Let

$$I = \int_0^\infty \frac{e^{iux}\phi_X(-u) - e^{-iux}\phi_X(u)}{iu}du.$$

Using equation (3.4) we can replace $\phi_X(w)$'s with their integral forms:

$$\begin{aligned} I &= \int_0^\infty \frac{e^{iux}\int_{-\infty}^\infty e^{-iuz}dF(z) - e^{-iux}\int_{-\infty}^\infty e^{iuz}dF(z)}{iu}du \\ &= \int_0^\infty\int_{-\infty}^\infty \frac{e^{iux}e^{-iuz} - e^{-iux}e^{iuz}}{iu}dF(z)du \\ &= \int_0^\infty\int_{-\infty}^\infty \frac{e^{iu(x-z)} - e^{-iu(x-z)}}{iu}dF(z)du. \end{aligned}$$

Let $\theta = u(x-z)$ (note we are not changing variables) and use the identity $\sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}$ to simplify:

$$I = \int_0^\infty \int_{-\infty}^\infty \frac{2\sin\big(u(x-z)\big)}{u} dF(z)du$$
$$= \int_{-\infty}^\infty \int_0^\infty \frac{2\sin\big(u(x-z)\big)}{u} dudF(z)$$
$$= \int_{-\infty}^\infty \pi * sgn(x-z)dF(z).$$

where it is known that $\lim_{n\to\infty} \int_0^n \frac{\sin(yu)}{u} du = \frac{\pi}{2} sgn(y)$; $sgn(y) = \begin{cases} -1 \; if \; y < 0 \\ 0 \; if \; y = 0 \\ 1 \; if \; y > 0 \end{cases}$

$$\therefore I = \pi\left(\int_{-\infty}^x 1dF(z) + 0 + \int_x^\infty (-1)dF(z)\right)$$
$$= \pi\left(F(x) + 0 + (-1)\big(1 - F(x)\big)\right)$$
$$= \pi\,(2F(x) - 1) => F(x) = \frac{1}{2} + \frac{1}{2\pi} I. \qquad \blacksquare$$

The above formulas (3.1)-(3.5) are quite powerful and their numerical implementation is discussed a bit later. The Author modifies (3.4) later on which results in an alternative expression of (3.5) and consequently a proof of (3.5).

Theorem 2.4 in the previous chapter shows how to construct an appropriate characteristic function given a suitable Lêvy measure. The question remains as to how to construct a characteristic function if one wishes to specify a set of stochastic differential equations? For instance, under the 4/2 Model two stochastic volatility processes were added together, however the mathematics needed to construct the characteristic function relies on methods from Lie Algebra and the particular problem was only recently solved.

Following the approach in (Duffie et al, 2000) and (Grzelak and Oosterlee, 2011), suppose a system of SDE's can be specified as follows, with some state vector $\boldsymbol{X}(t)$:

$$d\boldsymbol{X}(t) = \mu\big(\boldsymbol{X}(t)\big)dt + \sigma\big(\boldsymbol{X}(t)\big)dW(t). \tag{3.6}$$

Furthermore suppose that the system in (3.6) can be decomposed into the following (affine) form:

$$\mu\big(\boldsymbol{X}(t)\big) = a_0 + a_1\boldsymbol{X}(t), \qquad where\ (a_0, a_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n} \tag{3.7}$$

$$\Big(\sigma\big(\boldsymbol{X}(t)\big)\sigma\big(\boldsymbol{X}(t)\big)^T\Big)_{i,j} = (c_0)_{i,j} + (c_1)_{i,j}^T\boldsymbol{X}(t), \qquad where\ (c_0, c_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n} \tag{3.8}$$

An interest rate component is added and assumed to satisfy:

$$r\big(\boldsymbol{X}(t)\big) = r_0 + r_1^T\boldsymbol{X}(t), \qquad where\ (r_0, r_1) \in \mathbb{R} \times \mathbb{R}^n \tag{3.9}$$

Thus a discounted characteristic function for (3.6), under the risk neutral measure, $\mathbb{Q}$, is given as (3.10) (see (Duffie et al, 2000)):

$$\phi(u, \boldsymbol{X}(t), t, T) = E_{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)ds + iu^T\boldsymbol{X}(t)\right)\Big|\mathcal{F}(t)\right] = \exp\big(A(u,\tau) + B^T(u,\tau)\boldsymbol{X}(t)\big).$$

As usual the time lag is given as $\tau = T - t$ and the coefficients $A(u,\tau)$ & $B^T(u,\tau)$ can be determined by solving the following ordinary differential equations (complex-valued):

$$\frac{d}{d\tau}B(u,\tau) = -r_1 + a_1^T B(u,\tau) + \frac{1}{2}B^T(u,\tau)c_1 B(u,\tau). \tag{3.11}$$

$$\frac{d}{d\tau}A(u,\tau) = -r_0 + B^T(u,\tau)a_0 + \frac{1}{2}B^T(u,\tau)c_0 B(u,\tau). \tag{3.12}$$

As an example consider the following system of SDE's:

$$dS(t) = r(t)S(t) + \sqrt{V(t)}S(t)dW_S(t). \tag{3.13}$$

$$dV(t) = \kappa\big(\Theta - V(t)\big)dt + \gamma\sqrt{V(t)}dW_V(t). \tag{3.14}$$

$$dr(t) = \lambda\big(\Theta_r(t) - r(t)\big)dt + \eta r^p(t) dW_r(t). \tag{3.15}$$

Clearly $\boldsymbol{X}(t) = [S(t), V(t), r(t)]^T$ and the left hand side of (3.8) can be formulated:

$$\sigma\big(\boldsymbol{X}(t)\big)\sigma\big(\boldsymbol{X}(t)\big)^T = \begin{pmatrix} V(t) & \rho_{S,V}\gamma V(t) & \rho_{S,r}\eta r^p(t)\sqrt{V(t)} \\ * & \gamma^2 V(t) & \rho_{r,V}\gamma\eta r^p(t)\sqrt{V(t)} \\ * & * & \eta^2 r^{2p}(t) \end{pmatrix}$$

Unfortunately this matrix cannot be written in the form of the right hand side of (3.8) unless the correlations between the interest rate and asset prices and the interest rate and volatility are set to zero. Additionally for the characteristic function to be constructed as above the logarithm of the asset price must be taken, i.e. $\boldsymbol{X}(t) = \big[\log\big(S(t)\big), V(t), r(t)\big]^T$.

The characteristic functions for the Double Heston, H2-CIR and the 4/2 Model were all given in Chapter 2, all of them for the log price process. The Double Heston Model is the only model whose characteristic function can be formulated using the above theory while the other two model's characteristic functions had to be determined by other means.

A priori it should then be expected that the Double Heston will not perform as well as the other two models as a result. The justification for this lies in the fact that certain market features might not be modelled using affine systems of SDE's. Whether this is true or not will be tested in Chapter 4.

3.1.2. **Advanced derivative pricing techniques part 2: Numerical integration techniques**

Equations (3.4) & (3.5) are used to recover the distribution and density function of the log price process at a given time $t > 0$ numerically. The issue here concerns how to calculate this quantity numerically, i.e. what method must the practitioner implement to calculate the integral quickly and effectively? The typical answer is to use the discrete Fourier transform as can be seen from (3.5).

However an investigation into the order of the number of operations reveals that this naïve approach demands arithmetical operations of the order N^2 as will now be explained (see (Kienitz and Wetterau, 2012), below minor corrections were made to their approach). Consider a grid consisting of N points and define the following vectors $\bar{F}$ & $\bar{f} \in \mathbb{C}^N$:

$$\bar{F} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{pmatrix}, \qquad \bar{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}$$

Additionally, consider the matrix M:

$$M := \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_N^1 & \omega_N^2 & \cdots & \omega_N^{N-1} \\ 1 & \omega_N^2 & \omega_N^4 & \cdots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \cdots & \omega_N^{(N-1)(N-1)} \end{pmatrix}$$

Here the notation is standard in Complex Analysis: $\omega_N = e^{-\frac{2\pi i}{N}}$. The quantities above relate to each other as follows:

$$\bar{F} = M\bar{f} \Rightarrow F_k = \sum_{n=1}^{N} f_n e^{-\frac{2\pi i}{N}(n-1)(k-1)} = \sum_{n=1}^{N} f_n \omega_N^{(n-1)(k-1)}. \tag{3.16}$$

Hence (3.16) serves as the starting point in applying the discretised Fourier Transform. Ultimately the goal is to recover some probability mass function (due to discretisation this cannot be considered a

true density function) as follows:

$$p_X(x) = \frac{1}{\pi}\mathcal{R}\left(\int_0^{\infty} e^{-itx}\varphi(t)dt\right). \tag{3.17}$$

Clearly there is the state space domain (x) and the time domain (t) to consider. Every value of the state space is integrated in (3.17) over the entire time domain. On a high level inspection of (3.17) if the space and time domains were discretised each over N points then the order of operations performed is roughly of the order N^2. To make this exact consider splitting the time interval $[0, T]$ into N subintervals with lengths of Δ_t and split the space domain into steps of size Δ_x. Next, set $T = N\Delta_t$ and denote each time point of valuation $t_n = (n-1)\Delta_t$ and each point of valuation in the space domain by $x = -b + \Delta_x(u-1), u = 1, \dots, N$. The number b can be arbitrary although here it is simply defined by $b = N\Delta_x/2$. For technical reasons, which we omit here, the values Δ_x & Δ_t are constrained and related by the following equation: $\Delta_x\Delta_t = \frac{2\pi}{N}$.

Keeping these considerations in mind implies the following (variants are possible) approximation to (3.17):

$$p_X(x) \approx \frac{1}{\pi}\mathcal{R}\left(\Delta_t\left[\sum_{n=2}^{N} e^{-i\frac{2\pi}{N}(n-1)(u-1)} e^{i(n-1)\Delta_t b}\varphi(t_n) - \frac{e^{ixt_1}\varphi(\mathrm{t}_1) + e^{ixt_N}\varphi(t_N)}{2}\right]\right). \tag{3.18}$$

Clearly (3.18) has N^2 multiplications and additions which implies the order of the arithmetical operations is N^2. (Note that is when calculating over all possible values of x.)

Originally a method of reducing the order of operations to $N \log_2 N$ was introduced by (Cooley and Tukey, 1967) and is known as the Fast Fourier Transform (FFT). The approach is relatively simple. Suppose that X is an array containing N complex numbers. For some $k = 1, \dots, N$ an array Y can is defined by

$$Y(k) = \sum_{l=1}^{N} X(l)\omega_N^{(l-1)(k-1)}, \qquad \omega_N = \exp\left(-\frac{2\pi i}{N}\right). \tag{3.19}$$

In order to reduce the number of operations of (3.19) we express $N = N_1N_2$,$l = l_1N_2 + l_2$ and $k = k_1 + k_2N_1$. Hence (3.19) can be reformulated as (3.20):

$$Y(k_1 + k_2N_1) = \sum_{l_2=1}^{N_2}\left[\left(\sum_{l_1=1}^{N_1} X((l_1-1)N_2 + l_2)\omega_{N_1}^{(l_1-1)(k_1-1)}\right)\omega_N^{(l_2-1)(k_1-1)}\right]\omega_{N_2}^{(l_2-1)(k_2-1)}.$$

The above is implemented in MATLAB® as the *fft* function. The alternative to applying the FFT is to use some numerical integration techniques. Luckily MATLAB® has an *integral* function, which uses global adaptive quadrature rules to calculate integral numerically. Details of those underlying techniques can be found in the article by (Shampine, 2008) and the references therein. Importantly, the *integral* function is "vectorised", meaning that computational times can significantly be reduced by passing a vector of values into the integral function.

The use of the *integral* function is used by the Author in preference to the *fft* function in the supporting material for its added robustness in recovering the density function and due to the considerations given in Appendix D. Additionally the Author's approach in recovering the density function, given in section 3.2 and Appendix D, works extremely well in both speed and accuracy and is also comparable to the FFT method.

3.1.3. **Risk neutral dynamics: General results on stochastic volatility models**

An overview of risk neutral dynamics can be found in financial mathematics textbooks, see (Björk, 2009) or (Campolieti and Makarov, 2014). Here familiarity with Girsanov's theorem, the Cameron-Martin formula and the general mechanics of the Radon-Nikodym derivative are assumed.

In this section some advance issues are considered, firstly from a general framework (that can be applied to a broad range of models) and secondly some model specific issues are considered for the models presented in table 3.1. Before doing so, we need to consider why discounting with a risk-free rate is not always appropriate.

We need not always consider risk neutral dynamics from the viewpoint of a risk-free rate r(t). Traditionally it has been assumed that an appropriate discount factor $D(t)$ is defined with reference to the bank account (asset) $B(t)$, e.g. $D(t) = \frac{1}{B(t)} = \exp\left(\int_0^t r(u)du\right) => dB(t) = r(t)B(t)dt, B(0) = 1$. Thus if V_T denotes an option payoff at then at any prior time $t < T$ the usual valuation formula reads:

$$V(t) = B(t)E^{\mathbb{Q}}\left[\frac{V_T}{B_T}\middle|\mathcal{F}_t\right] = E^{\mathbb{Q}}\left[\exp\left(-\int_0^t r(u)du\right)V_T|\mathcal{F}_t\right]. \tag{3.21}$$

For some exotic options, discounting by the risk-free rate is not necessary. Consider an exchange option (with two assets S_1 & S_2) with payoff $V_T = \left(S_2(T) - S_1(T)\right)^+ = \max(S_2(T) - S_1(T), 0)$. Under the BS Model the valuation (Campolieti and Makarov, 2014) given in terms of time to maturity ($\tau = T - t$) is:

$$V(\tau) = S_2 e^{-q_2\tau}\mathcal{N}\left(d_+\left(\frac{S_2}{S_1}, \tau\right)\right) - S_1 e^{-q_1\tau}\mathcal{N}\left(d_-\left(\frac{S_2}{S_1}, \tau\right)\right)$$

$$d_\pm(x, \tau) \coloneqq \frac{\ln(x) + \left(q_1 - q_2 \pm \frac{1}{2}v^2\right)\tau}{v\sqrt{\tau}}, x, \tau > 0$$

$$v \coloneqq \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}\,.$$

Here the respective dividend rates are given by $q_i's\,, i = 1{,}2$. Note that in the valuation formula there is no r. The same formula can be obtained whether one discounts with asset $B(t)$ or either of $S_2(t)$ or $S_1(t)$. In this specific example, the former approach leads to a tedious algebraic exercise while the latter two approaches simplifies the problem significantly. Mathematically, if $g(t)$ denotes a dividend paying asset price process, then (3.21) takes an equivalent form:

$$V(t) = g(t)E^{(g)}\left[\exp\left(-\int_t^T q_g(u)du\right)\left(\frac{V_T}{g(T)}\right)|\mathcal{F}_t\right]. \tag{3.22}$$

In fact it can be shown that discounting by the asset process $g(t)$ leads to case where all discounted asset prices relative to $g(t)$ are $\mathbb{P}^{(g)}$ martingales if we define this probability measure via the Radon-Nikodym derivative process:

$$\left(\frac{\mathrm{d}\mathbb{P}^{(g)}}{\mathrm{d}\mathbb{P}^{(B)}}\right)_t \coloneqq \exp\left(-\frac{1}{2}\int_0^t \left\|\sigma^{(g)}(s)\right\|^2 ds + \int_0^t \sigma^{(g)}(s)\cdot dW_B(s)\right).$$

The details of the above can be found in (Campolieti and Makarov, 2014).

Although the analysis in (3.21) & (3.22) allows asset dynamics to be discounted under any appropriate numeraire asset, where it might be appropriate for some exotic derivatives, it is often convenient to work with the risk-free interest rate (either constant or stochastic) and is traditional of much of the work done on derivative pricing. Hence consider the following general multifactor factor

stochastic volatility model:

$$\left.\begin{aligned} \frac{dS(t)}{S(t)} &= \mu\big(\boldsymbol{Y}(t)\big)dt + f\big(\boldsymbol{Y}(t)\big)dW^{(0)}(t) \\ d\boldsymbol{Y}(t) &= \alpha\big(\boldsymbol{Y}(t)\big)dt + \boldsymbol{\beta}\big(\boldsymbol{Y}(t)\big)d\boldsymbol{W}(t) \end{aligned}\right\} \tag{3.23}$$

Here $\alpha : \mathbb{R}^d \to \mathbb{R}^d, \boldsymbol{\beta}: \mathbb{R}^d \to \mathbb{R}_+^d \times R_+^d$ is a diagonal matrix with entries $\beta_i\big(\boldsymbol{Y}(t)\big)$ and $\boldsymbol{W}$ is a d-dimensional vector of correlated Brownian motions, all correlated with W^0. Here $Y \in \mathbb{R}^d$ is some d-dimensional diffusion process. The function f is assumed to be a smooth, positive and an increasing function of Y. The correlations between the Brownian motions underlying each volatility process is

$$d\langle W^{(i)}, W^{(j)}\rangle_t = \rho_{ij}dt, \qquad i,j = 1,\dots,d.$$

and the correlations between each volatility component and the asset price process is

$$d\langle W^{(0)}, W^{(j)}\rangle_t = \rho_i dt, \qquad i = 1,\dots,d.$$

It is often convenient to split the W^i Brownian motion into two orthogonal components:

$$W^{(i)} = \rho_i W^{(0)} + \sqrt{1-\rho_i^2}W^{(i)\perp}.$$

where the orthogonal set of Brownian motions is defined by $\boldsymbol{W}^{\perp} = (W^{(1)\perp}, \dots, W^{(d)\perp})$.

The problem with trying to construct an appropriate risk neutral measure to change the drift term in (3.23) to the risk-free rate (and constructing the risk neutral Brownian motion terms) is that any appeal to no-arbitrage arguments fails since it is not sufficient anymore to hedge any contingent claim with cash and the underlying asset. The problem lies with the fact that the $W^{(0)}$ can be hedged using cash and the underlying but the orthogonal component $W^{(i)\perp}$ requires another component. Essentially this implies that an appropriate partial differential equation with a risk neutral parameter r cannot be constructed in this way.

It is however possible to get around this problem and the arguments in (Fouque et al, 2011) show that hedging the contingent claim with the underling, cash and *another derivative with a different expiration date* effectively allows the construction of appropriate risk-neutral dynamics. Another approach by the same authors mentioned is given below which addresses this issue by incorporating the market price of volatility risk, with the implicit assumption that the market ultimately selects the appropriate risk-neutral measure to price derivatives (or that this is implicit in traded prices observed).

Recall that in the classical BS theory the market price of risk, the Sharpe Ratio, is used in the construction of the Brownian motion term under the risk neutral dynamics $\mathbb{P}^*$:

$$W^{(0)*}(t) = W^{(0)}(t) + \int_0^t \frac{\big(\mu\big(Y(s)\big) - r\big)}{f\big(Y(s)\big)}ds. \tag{3.24}$$

Due to the addition of stochastic volatility, each Brownian motion driving a particular volatility factor must be changed by incorporating the market price of volatility risk, denoted by $\gamma^{(i)}$, as follows:

$$W^{(i)\perp *}(t) = W^{(i)\perp}(t) + \int_0^t \gamma^{(i)}(s)ds. \tag{3.25}$$

Here $i = 1,\dots,d$. In order to make the notation clear the risk neutral dynamics, now dependent on the market price of volatility risk process $\boldsymbol{\gamma} = (\gamma^1, \dots, \gamma^d) \in \mathbb{R}^d$, is denoted by $\mathbb{P}^{*(\gamma)}$. Under this measure the discounted stock price is assured to be a $\mathbb{P}^{*(\gamma)}$- martingale. Consequently (3.23) can be reformulated under this risk neutral measure as (3.26):

$$\frac{dS(t)}{S(t)} = rdt + f\big(\boldsymbol{Y}(t)\big)dW^{(0)*}(t)$$

$$d\boldsymbol{Y}(t) = \Big(\alpha\big(\boldsymbol{Y}(t)\big) - \boldsymbol{\beta}\big(\boldsymbol{Y}(t)\big)\underline{\boldsymbol{\Lambda}}(t)\Big)\,dt + \boldsymbol{\beta}\big(\boldsymbol{Y}(t)\big)d\boldsymbol{W}^{*}(t)$$

$$W^{(i)*}(t) = \rho_i W^{(0)*}(t) + \sqrt{1-\rho_i^2}\,W^{(i)\perp *}(t).$$

Here $\boldsymbol{W}^{*}(t) = \big(W^{(1)*}(t), \dots, W^{(d)*}\big)$ and $\underline{\boldsymbol{\Lambda}}(t) = \Big(\Lambda^{(1)}(t), \dots, \Lambda^{(d)}(t)\Big)$ with each component defined by

$$\Lambda^{(i)}(t) = \rho_i \frac{\big(\mu\big(\boldsymbol{Y}(t)\big) - r\big)}{f\big(\boldsymbol{Y}(t)\big)} + \gamma^{i}(t)\sqrt{1-\rho_i^2}.$$

Using the results from theorem 2.3 in Chapter 2, the pricing of any European option $P(t,x,\boldsymbol{y})$ with payoff $h\big(S(t)\big)$ is

$$P(t,x,\boldsymbol{y}) = E^{*(\gamma)}\big[e^{-r(T-t)}h(S_T)\big|S(t)=S, \boldsymbol{Y}(t)=\boldsymbol{y}\big]. \tag{3.27}$$

and consequently the corresponding partial differential equation for (3.27) can formulated as

$$\frac{\partial P}{\partial t} + \frac{1}{2}f(\boldsymbol{y})^2x^2\frac{\partial^2 P}{\partial x^2} + \sum_{i=1}^{d}\rho_i\beta_i(\boldsymbol{y})xf(\boldsymbol{y})\frac{\partial^2 P}{\partial x \partial y_i}$$
$$+\frac{1}{2}\sum_{i,j=1}^{d}\rho_{i,j}\beta_i(\boldsymbol{y})\beta_j(\boldsymbol{y})\frac{\partial^2 P}{\partial y_i \partial y_j} + r\Big(x\frac{\partial P}{\partial x} - P\Big)$$
$$+\sum_{i=1}^{d}\Big(\alpha_i(\boldsymbol{y}) - \beta_i(\boldsymbol{y})\Lambda^{(i)}(t,x,\boldsymbol{y})\Big)\frac{\partial P}{\partial y_i} = 0.$$

The $W^{(0)*}(t)$ term appearing in (3.26) is typically decomposed as

$$W^{(0)*}(t) = \sqrt{1-c_0^2}\,\widehat{W}^{*} + \sum_{j=1}^{d}c_j W^{(j)*}. \tag{3.28}$$

Here $\widehat{W}^{*}$ is a standard Brownian motion independent of the $W^{(j)^{*}}$ terms and each of the constants $c_j's$ can be obtained as the solutions of the following equations:

$$\sum_{j=1}^{d}\rho_{ij}c_j = \rho_i\,, \qquad i = 1, \dots, d, \qquad c_0^2 = \sum_{i=1}^{d}\rho_i c_{i.}$$

Interestingly under this general approach the price of a European Call options is given in terms of the price of the BS price for European Call options:

$$C(t) = E^{*(\gamma)}\Big[C_{BS}\Big(t, S(t)\times\xi(t); K, T; \sqrt{(1-c_0^2)\bar{\sigma}^2}\Big)\Big|\mathcal{F}_t\Big]. \tag{3.29}$$

Here the BS volatility is replaced by the $\sqrt{(1-c_0^2)\bar{\sigma}^2}$ where

$$\bar{\sigma}^2 = \frac{1}{T-t}\int_t^T \sigma^2(s)ds\,, \qquad \sigma(t) = f\big(\boldsymbol{Y}(t)\big),$$

and finally the function $\xi(t)$ is defined as

$$\xi(t) = \exp\left(\sum_{j=1}^{d}c_j\int_t^T\sigma(s)dW_s^{(j)*} - \frac{1}{2}c_0^2\int_t^T\sigma^2(s)ds\right).$$

The above discussion is very general and to a degree model-independent, i.e. the choices of coefficients α & β in (3.26).

The formula (3.10) gives the correct risk neutral dynamics in terms of the characteristic function for a general system of SDE's that satisfies (3.6) – (3.8). Hence for all the models showed in table 3.1 the correct risk neutral dynamics can be specified using the characteristic function. Additionally, the techniques described are general and thus other stochastic volatility models not considered here can

be used by the reader. The techniques of this section are particularly useful if one wishes to price options via numerically solving PDE. Such an approach will however not be used here.

3.2. **Monte Carlo techniques and considerations**

This section explains some general techniques to simulate any of the three models considered in table 3.1. and the techniques can be extended to many other models. The only assumption is that all parameter values are known. To "simulate" a model means being able to generate a sample path or paths of values of the process across time. The behaviour of the corresponding process can then be examined and these sample paths can be used to price derivatives.

The following example will demonstrate some important points that will underlie the discussions of this section. The discussion closely follows that given in (Baldeaux and Platen, 2013).

Generally, the classical BS equation is governed by SDE as given in (1.1), which is rewritten here as:

$$dS(t) = S(t)\big((r+\sigma^2)dt + \sigma dW(t)\big). \tag{3.30}$$

under the real world probability measure, $\mathbb{P}$. The SDE in (3.30) has a strong solution, i.e.:

$$S(t) = S_0 \exp\left(\left(r + \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right). \tag{3.31}$$

Consequently the value of a Call option under the real world probability measure is given by

$$Call_{BS}\big(T,K,S(0)\big) = S(0)E^{\mathbb{P}}\left[\frac{\max(S(T)-K,0)}{S(T)}\right]. \tag{3.32}$$

The theoretical basis for (3.32) was given by (3.22) in section 3.1.3 under risk neutral dynamics where we noted that it is not necessary (always) to discount by the risk-free rate. Consequently, suppose we want to calculate (3.32) using Monte Carlo methods. If we simulate say N paths of (3.31) at time T, denoting each simulated value by $S^i(T), i = 1, \dots, N$, then we would obtain for each simulated value a corresponding sample Call option price:

$$C^i\big(T,K,S(0)\big) = S(0)\frac{\max\big(S^i(T)-K,0\big)}{S^i(T)}. \tag{3.33}$$

Consequently the value in (3.32) can be approximated by the estimator:

$$\widehat{Call}_{BS}\big(T,K,S(0)\big) = \frac{1}{N}\sum_{i=1}^{N} C^i\big(T,K,S(0)\big). \tag{3.34}$$

The core issue here is what method should be used to simulate the values $S^i(T)$? Since a strong solution is given in (3.31) an obvious method would be to simulate values from the normal distribution, i.e. $\mathcal{N}(0,T)$ since the Brownian motion $W(T) \sim N(0,T)$ and then obtain the values $S^i(T)$ using the formula in (3.31). This method of simulation, or what we will term a scheme, is known as an *exact simulation* scheme.

To understand this important concept, let us introduce a scheme which is not exact, such as the following discretization scheme:

$$\tilde{S}(t+\Delta t) = \tilde{S}(t)\big(r\Delta t + \sigma^2\Delta t + \sigma\Delta W(t)\big). \tag{3.35}$$

where the time steps are typically given as $\Delta t = \frac{T}{n}$, where n is the number of time steps over the interval $[0,T]$ and $\Delta W(t) = W(t+\Delta t) - W(t)$. The values $\tilde{S}^i(T)$ are obtained from recursively substituting in time points and calculating the next value from (3.35) until the terminal time T has been reached. Since the true distribution of $S(T)$ is approximated by $\tilde{S}(T)$ it follows that generally

$$E^{\mathbb{P}}\left[\frac{\max\left(\tilde{S}(T)-K,0\right)}{\tilde{S}(T)}\right] \neq E^{\mathbb{P}}\left[\frac{\max(S(T)-K,0)}{S(T)}\right]. \tag{3.36}$$

This leads to two notions, which quantify to what extent $\tilde{S}(T)$ differs from $S(T)$. The first refers to the concept of *strong order of convergence*, given by the number α, and defined through

$$\left(E\left(\left|\tilde{S}(T)-S(T)\right|^2\right)\right)^{\frac{1}{2}} \leq c\cdot(\Delta t)^{\alpha}. \tag{3.37}$$

The second notion is that of *weak order of convergence*, defined through the number β and some smooth function f as:

$$\left|E\left(f\left(\tilde{S}(T)\right)\right)-E\left(f\left(S(T)\right)\right)\right| \leq c\cdot(\Delta t)^{\beta}. \tag{3.38}$$

Here c is simply some constant number. The first concept is applied to measuring pathwise convergence while the second is more typically used to assess the quality of approximating functionals of diffusion processes through Monte Carlo simulations. These notions are related to the MSE error measure as explained below

Consider again the exact Call option price in (3.32) and the approximation (3.34). The MSE, or mean square error, measures the following quantity:

$$\begin{aligned} MSE &= E\left[\left(\widehat{Call}_{BS}-Call_{BS}\right)^2\right] \\ &= \left(E\left[\widehat{Call}_{BS}\right]-Call_{BS}\right)^2 + E\left[\left(\widehat{Call}_{BS}-E\left(\widehat{Call}_{BS}\right)\right)^2\right]. \end{aligned} \tag{3.39}$$

The first quantity in (3.39) is known as the *bias* and the second quantity is the *variance of the estimator*. Clearly the MSE is minimised when the bias and the variance of the estimator is small. Using (3.38) we see that the *bias depends on the weak order of convergence for the scheme* and hence the MSE can be rewritten as:

$$MSE = c^2\cdot\frac{T^{2\beta}}{n^{2\beta}}+\frac{Var\left(C^i\left(T,K,S(0)\right)\right)}{N}. \tag{3.40}$$

The result in (3.40) is extremely important as it shows us that the *bias is dependent on the size of the time step* taken to simulate values and the variance is related to the number of simulations N. Hence an *exact simulation* scheme is one in which the bias is zero. This is evident from the above discussions as simulating the asset price at time T can be done without discretising the time grid and then sequentially building a sample path. This follows if the exact transition density is known or if a strong solution to the corresponding SDE's exists.

Another important concept that is related to the MSE is the idea of the Monte Carlo Convergence rate. To understand this we assume that the computational complexity of a simulation scheme is given as $C = n\cdot N$. Looking at (3.40) the MSE could be minimized by balancing the denominators, e.g. $n^{2\beta} \asymp N$, (the symbol $\asymp$ means "equivalent"), which implies that $C \asymp n^{2\beta+1}$ and that consequently MSE $\asymp C^{-\frac{2\beta}{2\beta+1}}$. The latter quantity is known as the Monte Carlo Convergence rate. An unbiased scheme will typically exhibit a Monte Carlo Convergence rate of C^{-1}.

The two general schemes that will be given here each have various biases and a detailed analysis is omitted in the interest of saving space. After that, two more advance methods are presented. Before starting, let us to consider the Heston Model as a canonical example of how to apply the schemes, from which adjustments for other models can easily be made. In particular, we will use the log-price of the asset, $X(t) = \log\left(S(t)\right)$:

$$dX(t) = \left(r-\frac{V(t)}{2}\right)dt+\sqrt{V(t)}dW_1(t) \tag{3.41}$$

$$dV(t) = \kappa\big(\theta - V(t)\big) + \eta dW_2(t)$$

The correlations between the Brownian motions is ρ.

3.2.1. **Euler Scheme**

The Euler scheme is the most basic scheme and is used frequently in practice. It is what the Author would term the "fall-back" scheme, simply because one can always fall back to this scheme if other methods prove difficult to implement. The Euler scheme, as presented here, achieves a strong order of convergence of $\alpha = \frac{1}{2}$ and a weak order of convergence of $\beta = 1$. The Monte Carlo Convergence rate is $C^{-2/3}$ which is due to the bias introduced through discretising.

The method is simple and applied to (3.41) gives:

$$X(t + \Delta t) = X(t) + \left(r - \frac{V(t)}{2}\right)\Delta t + \sqrt{V(t)\Delta t}W_1(t). \tag{3.42}$$

$$V(t + \Delta t) = V(t) + \kappa(\theta - V(t)\Delta t + \eta\sqrt{V(t)\Delta t}W_2(t). \tag{3.43}$$

here the W_1 & W_2 are standard, correlated, Normal random variables. If Z denotes a standard Normal random variable, independent of W_2 then $W_1 = \rho W_2 + \sqrt{1-\rho^2}Z$. Hence for the scheme to work, one only needs the starting values of $V(t)$ & $X(t)$ and generate at each time step two Normal random variables.

A big problem with the Euler scheme is that, given $V(t) > 0$, the following, obtained from basic statistics, relationship holds:

$$P(V(t + \Delta t) < 0) = \Phi\left(\frac{-1(1 - \kappa\Delta t)V(t) - \kappa\theta\Delta t}{\eta\sqrt{V(t)\Delta t}}\right).$$

A solution to having a negative variance is to simply cap it, i.e. the value of $V(t)$ is either positive or zero. This solution has been given by (Lord et al, 2010). Consequently, if the variance is zero then the next value has no random component and is given by $\kappa\theta\Delta t$. We will work with this as our standard approach and build in the negative variance cap as a precaution when implementing the method in MATLAB®. Note that this is only a practical consideration which arises as a consequence of the Euler discretisation scheme.

3.2.2. **Milstein scheme**

A number of variations of this scheme as appeared, which is originally due to (Mil'shtejn, 1975) and although it will not be used for our purposes, we can consider briefly consider it. For the Heston Model the scheme simply adds in an additional factor to simulate the variance:

$$X(t + \Delta t) = X(t) + \left(r - \frac{V(t)}{2}\right)\Delta t + \sqrt{V(t)\Delta t}W_1(t). \tag{3.44}$$

$$V(t + \Delta t) = V(t) + \kappa(\theta - V(t)\Delta t + \eta\sqrt{V(t)\Delta t}\, W(t)^2 + \frac{\eta^2}{4}\Delta t(W_2(t) - 1)^2. \tag{3.45}$$

The last summand in (3.45) destroys the martingale property of the discounted asset price. On the other hand, an advantage of adding this factor is that the strong order of convergence for the Milstein Scheme (i.e. $\alpha = 1$) is better than for the Euler Scheme. A slightly more generalised setting for this scheme is given by (Kietnitz and Wetterau, 2012), but the Author formulates it differently for notational consistency:

Suppose that an SDE is given by

$$dX(t) = f\big(X(t)\big)dt + g\big(X(t)\big)dW(t). \tag{3.46}$$

The corresponding Milstein Scheme for (3.46) is then:

$$X(t+\Delta t) = X(t) + f\big(X(t)\big)\Delta t + \sqrt{\Delta t}Wg\big(X(t)\big) + \frac{1}{2}g\big(X(t)\big)'g\big(X(t)\big)\big(\sqrt{\Delta t}(W)^2 - \Delta t\big). \tag{3.47}$$

In (3.47) $g\big(X(t)\big)'$ denotes the first derivative with respect to $X(t)$, which is typically given through by a simple finite difference scheme (see section 3.5).

3.2.3. **Multilevel methods**

Conceptually, the Multilevel method introduced here will reach a Monte Carlo Convergence rate of C^{-1} by taking the basic Euler scheme and applying a mix of time steps and a number of simulations, of variable size, to derive a suitable trade-off between the bias and variance. The idea is to generate paths of variable time sizes such that paths generated with large time steps are simulated a large number of times but paths with a smaller time steps are simulated fewer times.

On the one hand the variance is reduced in the first instance and on the other hand the bias is reduced as the time step is small. If this mix is done properly then it can be shown that the Monte Carlo Convergence rate is as given above.

The basis for this method is introduced in the two articles by (Giles, 2008a) and (Giles, 2008b). It is a method we will use in conjunction with the Quasi-Monte Carlo method below. The discussion follows that of (Giles and Waterhouse, 2009) and (Giles, 2008a) in presenting this method. The aim of the Multilevel method is to reduce the order of the number of computational operations. Suppose that the $MSE < \epsilon$, $\epsilon > 0$, for some specified error ϵ, then in order to guarantee this condition the number of computational operations is of the order $O(\epsilon^{-3})$. The Multilevel approach reduces this to $O(\epsilon^{-2}(\log\epsilon)^2)$, see (Giles, 2008a). Note that an unbiased scheme will typically achieve $O(\epsilon^{-2})$ in computational operations.

Now, the method assumes that Monte Carlo paths are simulated with time steps of size $\Delta t_l = \frac{T}{M^l}$ where M^l indicates exponentiation of M with l and $l = 0,1,\dots,L$. The SDE we will be working should preferably be of the form:

$$dS(t) = a(S(t),t) + b(S(t),t)dW(t), \qquad 0 < t < T. \tag{3.48}$$

Note that for this exposition the variable $S(t)$ can equally denote the variance process $V(t)$ which is also assumed to have the same form as (3.48). Most of the models in table 3.1 have this form and the assumptions on a & b are for our purposes are generally satisfied. It is sufficient to consider the method for the single SDE in (3.48) as the adaptation is trivial when considering an added variance or interest rate process.

For a given particular path of (3.48), denote the payoff function $f\big(S(T)\big)$ by P and denote the approximations to $S(t)$ and P by $\hat{S}_{l,M^l}$ and $\hat{P}_l$ by using a discretised scheme with time step Δt_l. The following telescoping sum can then be given:

$$E\big[\hat{P}_L\big] = E\big[\hat{P}_0\big] + \sum_{l=1}^{L} E\big[\hat{P}_l - \hat{P}_{l-1}\big]. \tag{3.49}$$

The goal of this method is to sequentially estimate the $E\big[\hat{P}_l - \hat{P}_{l-1}\big]$ quantity. The starting point for this is to let $\hat{Y}_0$ be an estimator for $E\big[\hat{P}_0\big]$ using N samples and generally let $\hat{Y}_l$ be an estimator for

$E\left[\hat{P}_l - \hat{P}_{l-1}\right]$ using N_l paths. Clearly then:

$$\hat{Y}_l = \mathrm{M}_\mathrm{l}^{-1} \sum_{i=1}^{M} \left(\hat{P}^i{}_l - \widehat{P}^i{}_{l-1}\right). \tag{3.50}$$

The trick, as described by as (Giles, 2008a), is that $\hat{P}_l - \hat{P}_{l-1}$ is calculated as two discrete approximations using different time steps but the same Brownian path. This is achieved by first estimating $\hat{P}^i{}_l$ using the discrete path build from the Brownian motion term, i.e. $dW(t)$. In other words at each point in time a value for the Brownian motion term is simulated at time t. Each of these values is used to construct the next value of the underlying $S(t + \Delta t)$ in the discrete path until the final time step is reached.

However we collect the Brownian terms as they are simulated in a vector. This vector will then be used to construct a sample path for $S(T)$, however since the time steps differ we have to sum the elements of this vector in groups of M. Each group that is summed will then give the Brownian term needed at time t to construct the next value of the discrete path, i.e. $S(t + \Delta t)$ until the entire path is constructed and $\hat{P}_{l-1}$.

Figure 3.1 below illustrates this idea. A Geometric Brownian motion has been constructed using 4096 time steps. Here 4096 normal random variables were simulated and hence used in the Euler scheme to construct the path in blue. Each of these normal random variables were summed in groups of 4 and leaving 1024 variables. These variables, together with 1024 time steps, were used to construct the path in red. In other words the same Brownian motion is used to simulate paths at two different time steps. This is the idea of the multilevel method.

A further green path has been added by again summing in groups of 4 to show the effect of this "coarser" structure with less time steps in comparison to the "fine" structure of the blue path. This clearly shows the bias when discretising.

This procedure is typically repeated N_l times to estimate (3.50) and consequently (3.49), i.e. the combined estimator is:

$$E\left[\hat{P}_L\right] = \sum_{l=0}^{L} \hat{Y}_l. \tag{3.51}$$

The key insight here is that each of $\hat{Y}_l$'s are independent of each other and consequently the variance of the estimator is $\sum_{l=0}^{L} N_l^{-1} Var\left(\hat{P}^i{}_l - \widehat{P}^i{}_{l-1}\right)$ which we write as $\sum_{l=0}^{L} N_l^{-1} V_l$. It can then be shown that by specifying that the MSE should be less than some $\epsilon > 0$, and if $N_l \propto \sqrt{V_l \Delta_{t_l}}$ with $N_l = O\left(\epsilon^{-2} L \Delta_{t_l}\right)$ and $L = \frac{\log(\epsilon^{-1})}{\log(M)}$ then the Monte Carlo Convergence rate is equal to that of an unbiased (exact) scheme, up to a factor of $(\log(\epsilon))^2$. Lastly, the optimal number M is given in (Giles, 2008a) to be 7, although in the article the he preferred to work with $M = 4$, which is what we will also work with.

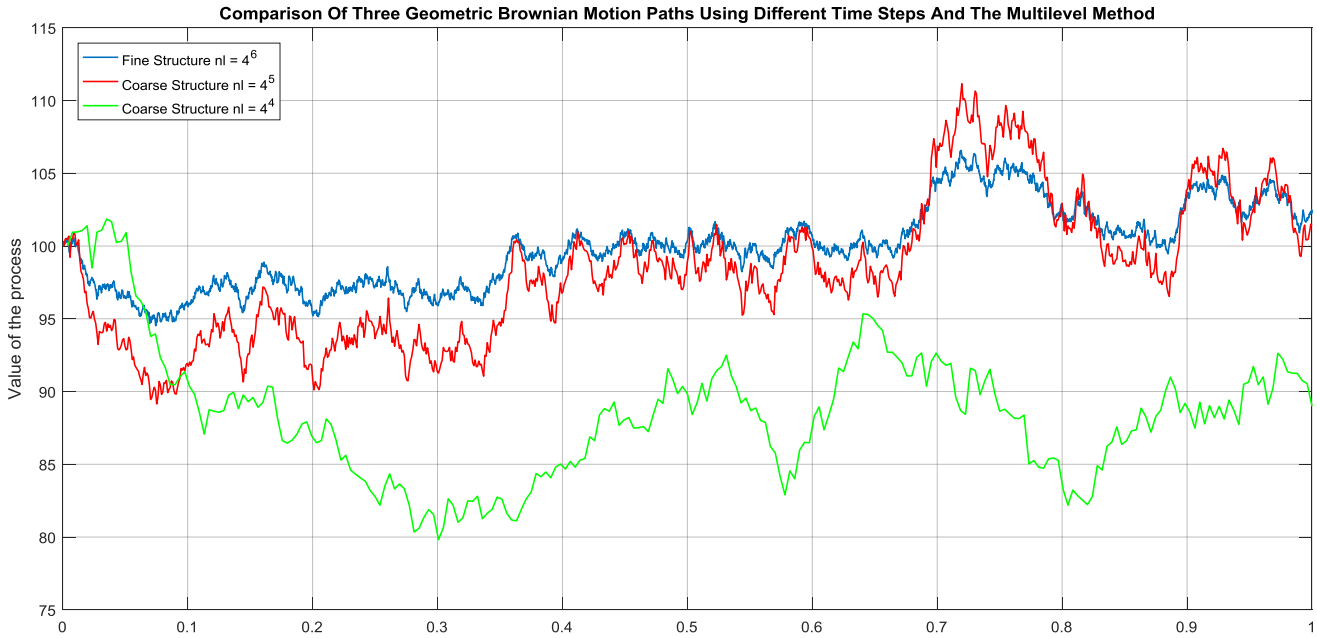

Figure 3.1: An illustration of the multilevel method using different time steps.

3.2.4. Exact HW and CIR simulation schemes and the Brownian Motion Reconstruction scheme.

Using exact simulation schemes are often computationally expensive for general Monte Carlo purposes, but ultimately such schemes enable the elimination of bias. An example of an exact scheme is given in (Baldeaux and Platen, 2013) for the 3/2 Model, however running this scheme is computationally expensive as the Bessel function takes complex valued arguments, which may not always give robust results.

Conversely, by using a Euler discretisation scheme, one has to discretize twice or even three times for some of the models considered here. Typically the variance process is discretized and then used to calculate the asset process, which itself is subject to a Euler discretisation scheme. The resulting option prices do not always converge to the correct option price. If one desires accuracy then simulating by using smaller time steps can work, but will lead to increase computational time.

A balance between these extremes can be struck. The effect on the overall bias of the scheme can be partially mitigated by simulating any interest rate or variance processes exactly and then only applying a Euler scheme to generate the asset process. This process however requires that we re-construct the Brownian motion increments when considering non-zero correlations. This is done by simulating the variance or interest rate process and then recovering the $dW(t)$ which are then used to simulate the asset process.

To the best of the Author's knowledge this method of reconstruction is not used in literature, nonetheless it is computationally efficient, as demonstrated in Chapter 4. This method of reconstruction is then an alternative method that can be used whenever a stochastic volatility model (with or without stochastic interest rates) is driven by a volatility or interest rate process that are either Hull-White or CIR processes. An example with supporting code of this method is given in

section 3.2.5.1 below.

**Note: This Brownian Motion Reconstruction scheme is not claimed to be original work. However, the absence of this technique in the significant literature works regarding derivative pricing and Monte Carlo simulations suggest that using this technique for comparison purposes, especially coupled with the results given below, is unique.

For the CIR process, recall that in Chapter 2 we examined the SQB process in detail and alluded to the exact simulation scheme of this model as given by (Makarov and Glew, 2009). Additionally, recall that the SQB process is related to the CIR process, which means that we need to sample the SQB only. The exact simulation scheme is relatively fast and the algorithm for simulating the CIR process can be formulated below.

Generally the form of the CIR process that we have largely considered is of the form:

$$dV(t) = \kappa\big(\Theta_V - V(t)\big)dt + \nu\sqrt{V(t)}dW(t).$$

which can be rewritten in a very trivial manner as:

$$dV(t) = \left(\alpha - \tilde{\beta}V(t)\right)dt + \nu\sqrt{V(t)}dW(t). \tag{3.52}$$

and this form is consistent with the notation in Chapter 2 under the CIR process. Hence, this form is going to be used in numerical work and a simple substitution is only required.

Note that if the volatility process is assumed to satisfy the Feller condition, as given under the Heston Model in Chapter 2, the volatility process cannot become zero. Additionally, the Feller condition makes the simulation process via the SQB process extremely easy. The reader is referred to Chapter 2, subsection 2.2.4.2 for details of the SQB process in order to understand the following algorithm:

Algorithm 3.1 Exact Simulation of the CIR Process Using the SQB Process

---

1.) Simulate a sample path using the SQB process, imputing the starting value of the variance of the CIR process,V_0, μ and a time vector, $e.g. \boldsymbol{T} = [t_0, t_1, t_2, \dots t_n], where\ t_0 = 0, t_1 < t_2 < \cdots < t_n$.

1.1) Compute the μ parameter, using the parameters in (3.52): $= \frac{\left(\frac{2\alpha}{\beta}\right)}{\nu^2} - 1$.

1.2) Compute an appropriate time change: $\boldsymbol{\tau} = (\exp(\beta \boldsymbol{T}) - 1)/\beta$, where the exponentiation and other operators are applied component wise to the vector $\boldsymbol{T}$.

1.3) Compute the initial value to start the SQB process: $X_0 = V_0 \times \frac{4}{\nu^2}$

1.4) Simulate a path using the SQB process:

1.4.1) Generate $Y_j = Poisson\left(\frac{X(\tau_j - 1)}{2(\tau_j - \tau_{j-1}}\right)$.

1.4.2) Generate $X(\tau_j) = Gamma\left(Y_j + \mu + 1{,}2(\tau_j - \tau_{j-1})\right)$

1.5) Repeat the process for all time values working forwards in time sequentially and returning a sample path $\{X(\tau_j)\}_{0 \le j \le n}$

1.6) Apply the transformation to recover the CIR process: $V(t_j) = \frac{\nu^2}{4} \times X(\tau_j) \times \exp(-\beta t_j)$.

In the supporting material (see Appendix A) further algorithms are given to simulate the volatility process if the Feller Condition is not fulfilled. In the interest of saving space these are not given here but the details can be found in (Makarov and Glew, 2009). In any case, the above algorithm does

behave well (reflects the volatility away from zero) if the Feller Condition is not fulfilled.

An example with supporting code is given below in section 3.2.5.1. on how to simulate asset prices using the exact CIR simulation with Brownian motion reconstruction technique.

The Hull-White model's simulation is very easy and without restrictions. The form that we will be working with is:

$$dr(t) = \lambda\big(\theta(t) - r(t)\big)dt + \sigma dW(t). \tag{3.53}$$

For our purposes the parameter θ is kept constant, although the example below, which was promised in Chapter 2, is given with a time dependent parameter $\theta(t)$. Recall from (2.24) that, with $t - s = \tau$:

$$r(t)|r(s) \sim N\left(\exp(-\lambda\tau)\, r(s) + \lambda \int_s^t \exp(-\lambda\tau)\, \theta(u)du\ , \frac{\sigma^2}{2\lambda}(1 - \exp(-2\lambda\tau))\right). \tag{3.54}$$

Thus one need only simulate from the Normal distribution at each time period and build a sample path. Comparing this to Algorithm 3.1 where at each time point we have to sample from the Poisson and the Gamma distribution, which takes more time. For the majority of our work the parameter θ will be constant, as the default assumption.

3.2.5. **Simulation by using density recovery techniques vs Monte Carlo simulation**

The transition density function of a stochastic volatility model can be given by inverting the characteristic function using (3.5). This numerical inversion technique gives the exact density function at some future time, given a starting value for the asset price process, and hence for pricing any non-path dependent derivatives, such as European options, the exact density function can be used.

As an example, suppose that the density function, representing the spread of values of an asset price, has been obtained numerically at time $t = 1$ given some starting value at time $t = 0$, say $S(0) = 1000$. Furthermore suppose that a European Call option has expiry at time $t = 1$ with strike $K = 1100$.

To price this option one can work out the expectation from the point $x = 1100$ onwards on the support of the density function and then simply discounting the value. The expectation is calculated by multiplying each point of the support with the value of the density function and then summing across all the values after subtracting 1100 from each point on the support (from point $x = 1100$ onwards) and adjusting for the number of points on the support. The procedure can be found in the support material (see Appendix A).

Below, in subsection 3.2.5.1 a few illustrations of the Monte Carlo simulation technique and the exact density recovery technique are given. After that the Author presents a new formula that allows for rapid density recovery.

3.2.5.1. **Example: Double Heston Model and transition density surfaces**

In literature, (Gauthier and Possamaï, 2010) compare various simulation schemes for the Double Heston Model. The schemes that are compared are the Euler, Zhu, Exact Zhu, Quadratic Exponential and variations of the Alfonsi scheme. Details regarding each of the schemes and the authors introducing them can be found in the referenced article. The final results of Gauthier & Possamaï (2010) suggest a 3rd Order Alfonsi scheme is appropriate for the Double Heston Model.

As mentioned earlier, we will not pursue other alternative simulation schemes. An explicit scheme is given here for the Double Heston Model in MATLAB® using exact simulation of volatility paths and Brownian increment recovery. The MATLAB® code for the Double Heston Model under this scheme is given here:

```
function y = simulateDH(X0,v1,v2,r,a1,a2,b1,b2,sigma1,sigma2,rho1,rho2,Time,NrSim)
s = length(Time);
%Simulate the Volatility Processes exactly
V1 = simulateCIR_mex(a1,b1,sigma1,v1,Time,NrSim);
V2 = simulateCIR_mex(a2,b2,sigma2,v2,Time,NrSim);

%Initialise some variables
dW1 = zeros(NrSim,s);
dW2 = zeros(NrSim,s);
X = zeros(NrSim,s);
Z = random('Normal',0,1,NrSim,s);
Z1 = random('Normal',0,1,NrSim,s);
%Reconstruct Brownian Motion Increments, remember that time is already incorporated in the
increment!:
for ii = 1:NrSim
for jj = 1:s

    if jj == 1
         dW1(ii,jj)=((V1(ii,jj)-v1) - (a1-b1*v1)*Time(jj))/(sigma1*sqrt(v1));
    else
         dW1(ii,jj)=((V1(ii,jj)-V1(ii,jj-1)) - (a1-b1*V1(ii,jj-1))*(Time(jj)-Time(jj-
1)))/(sigma1*sqrt(V1(ii,jj-1)));
    end
end
for jj = 1:s
    if jj == 1
         dW2(ii,jj)=((V2(ii,jj)-v2) - (a2-b2*v2)*Time(jj))/(sigma2*sqrt(v2));
    else
         dW2(ii,jj)=((V2(ii,jj)-V2(ii,jj-1)) - (a2-b2*V2(ii,jj-1))*(Time(jj)-Time(jj-
1)))/(sigma2*sqrt(V2(ii,jj-1)));
    end
end
%Simulate Log-Prices
for jj = 1:s
    if jj == 1
    X(ii,jj)=X0+(r-0.5*(v1+v2))*Time(1)+sqrt(v1)*(rho1*dW1(ii,jj)+...
        sqrt(1-rho1^2)*Z(ii,jj)*sqrt(Time(1)))+sqrt(v2)*(rho2*dW2(ii,jj)+...
        sqrt(1-rho2^2)*Z1(ii,jj)*sqrt(Time(1)));
    else
    X(ii,jj)=X(ii,jj-1)+(r-0.5*(V1(ii,jj-1)+V2(ii,jj-1)))*(Time(1))+sqrt(V1(ii,jj-
1))*(rho1*dW1(ii,jj)+sqrt(1-rho1^2)*Z(ii,jj)*sqrt(Time(1)))+sqrt(V2(ii,jj-
1))*(rho2*dW2(ii,jj)+sqrt(1-rho2^2)*Z1(ii,jj)*sqrt(Time(1)));
    end

end
end
%Obtain Matrix of Time series, each row vector is a simulated path:
```

```
X1 = zeros(NrSim,s+1);
X1(:,1) = X0;
X1(:,2:end)=X;
y = exp(X1)';
end
```

The function *simulateCIR* can be found in the supporting material. The above code takes longer to run than using a simple Euler Scheme, however it converges faster to the correct distribution of asset prices at some terminal time due to the elimination of the bias in simulating the volatility processes.

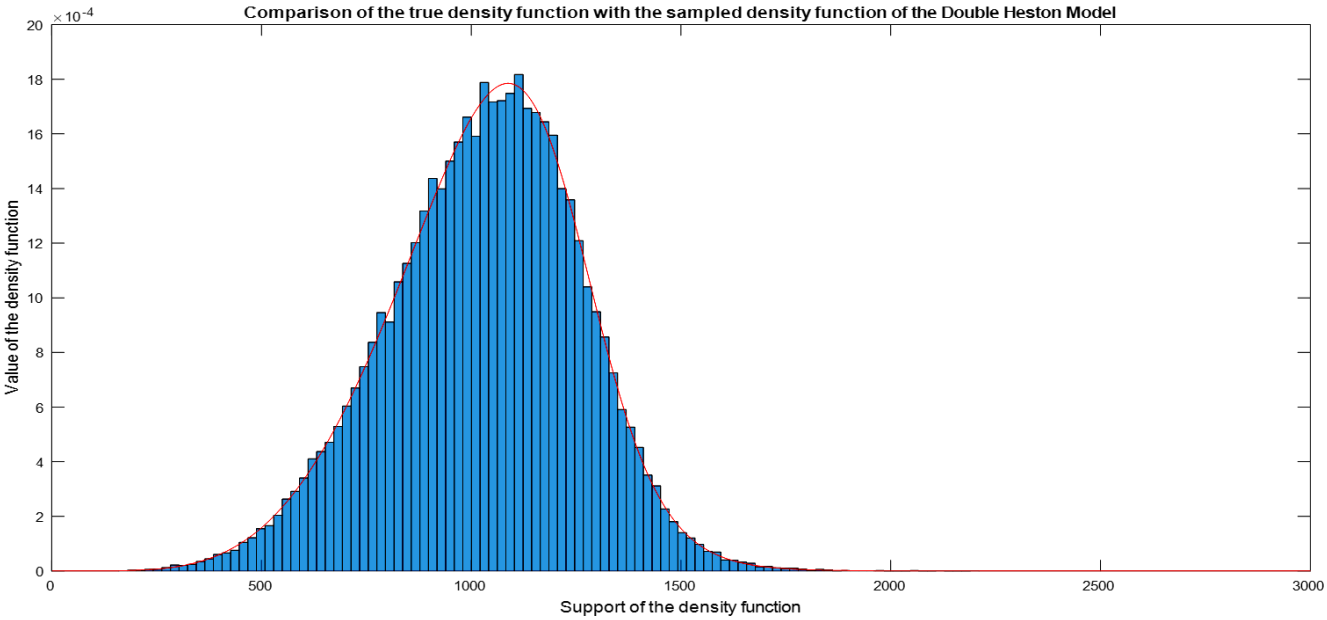

Figure 3.2: Comparison of the true density function with the sampled density function of the Double Heston Model.

The above histogram was generated from 50000 samples and 365 time steps in one year. Ultimately there is still a bias present from simulating the asset paths and this is evident from the figure as the resulting distribution is slightly flatter than expected. Note that the asset prices in the above figure are due to the following set of dummy parameters that were used to generate the histogram and the exact density function:

Parameter	$\log(S_0)$	$v_1(0)$	$v_2(0)$	a_1	a_2	b_1	b_2	σ_1	σ_2
Value	6.9077	225%%	220%%	0.04	0.03	1	0.8	0.35	0.3
Parameter	ρ_1	ρ_2	r						
Value	-0.7	-0.7	0.03						

Table 3.2: Double Heston Model parameters used in the Monte Carlo simulation example.

One can always take the characteristic function and produce a "surface" of probability densities across time that shows the evolution of how the density function of the price distribution changes as time changes using the above parameters. Effectively the transition densities are plotted at each future period where the conditioning is upon the current time $t = 0$ and the asset price $S(0) = 1000$. This is

done in such a way as to produce an "evolving surface" across time.

Note that the graph starts at $t = 0.2$:

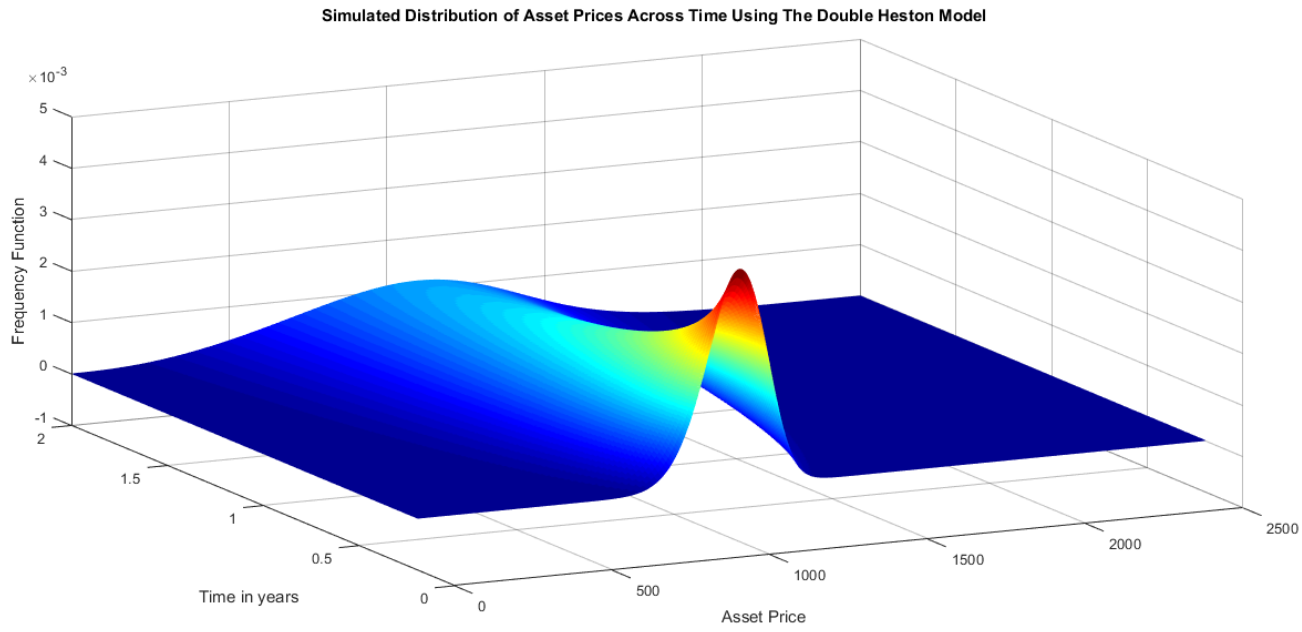

Figure 3.3: Simulated transition density surface of asset prices across time using the Double Heston Model.

Lastly, for clarity, when one obtains the log-price density function then one can obtain the density function of the asset price by dividing the log-price density function by the desired support of the asset price. Mathematically, this result is well-known and given in the undergraduate textbook by (Rice, 2007), specifically proposition B in Chapter 2:

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ whenever $x \notin I$. Then Y has the density function:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|.$$

for y such that $y = g(x)$ for some x. Here g^{-1} is the inverse function of g: $g^{-1}(y) = x$ if $y = g(x)$.

The function g is then simply $g(x) = e^x$ and clearly $\frac{d}{dy}g^{-1}(y) = \frac{1}{y}$. Since the function f_X is evaluated at the points of the log-price density's support, one only needs to divide f_X at the corresponding points with the support of the asset price's density function, i.e. the y values. In other words the density function f_X is divided by y at the corresponding points, e.g. if $f_X(4) = 3$ then $f_Y(e^4) = \frac{3}{e^4} = \frac{f_X(4)}{e^4} \equiv \frac{f_X(x)}{e^x} = \frac{f_X(x)}{y}$.

3.2.5.2. **Numerical techniques: A fast method to recovering the density function and sampling from the cumulative distribution function.**

In Chapter 4 a comparison will be made between pricing exotic equity options using Monte Carlo techniques and sampling from density functions using characteristic functions as mentioned in

subsection 3.2.5 above. Here a new technique has been discovered by the Author and is presented here. The aim of using this technique is to develop a numerical method that computes the transition density function faster than using the FFT inversion technique. A brief mathematical justification is given as to why it is expected that technique will work.

The motivation for deriving a new sampling technique stems from the fact that every time equation (3.4) or (3.5) is used then to recover the density function around 1000-10000 points are needed which results in 1000-10000 integrals that have to be evaluated numerically. The resulting density function can be used to sample values by using inversion techniques or a root finding algorithm. Essentially this technique calculates the (transition) density function quickly and is a comparable, although not revolutionary, alternative to the FFT.

Consider again equation (3.4), with a slight notational change:

$$Cumulative\ distribution:\ \ H(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty} \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)}{iu}\,du. \qquad (3.55)$$

As a starting point, by specifying some number on $H(x)$, say ω, one can use a root finding algorithm and solve for x which is exact, up to some specified error tolerance. However this process is likely to be computationally expensive and might be prone to numerical instabilities. Thus the Author's solution is built around (3.55) that is both accurate and fast and is stated in Theorem 3.1, which is the main result in this thesis.

The solution presented below consist of a theorem, a corollary and two numerical routines. Before giving the theorem and its proof, it should be mentioned that this result is investigated further in Appendix D that allows us to compute the density function in (3.5) around 2 times faster than the standard formula. The Author is not aware of any literature that contains or has used this result. Furthermore the algorithms and issues investigated in Appendix D do not appear in literature, to the best of the Author's knowledge.

Essentially the result removes all the complex-valued operations in (3.5) which results in greater stability and faster calculations. The resulting function is real valued and favours simpler quadrature rules to calculating integrals. Hence in preparation for the proof a few facts about characteristic functions are developed from first principles and other mathematical facts are also given.

Properties of characteristic functions and other techniques:

It is understood here that the function denoted as $imag()$ is a function that returns the imaginary part of a complex number, i.e. if $z = a + i \times b;\ a \in \mathbb{R}, b \in \mathbb{R}$ then $imag(z) = b$. Additionally recall that for any $z \in \mathbb{C}$ we have that $\cos(-z) = \cos(z)$ and $\sin(-z) = -\sin(z)$. Now, let $\phi_X(u)$ be the characteristic function of a random variable X with probability density function $H'(x) = h_X(x)$. Then for all $u \in \mathbb{R}$ we have

$$\begin{aligned} \phi_X(u) = E\left[e^{iuX}\right] &= E[\cos(uX) + i \times \sin(uX)] \\ &= E[\cos(uX)] + iE[\sin(uX)]. \end{aligned}$$

$$\therefore E[\cos(uX)] = \int_{-\infty}^{\infty} \cos(ux)\, f_X(x)dx \Rightarrow real\left(\phi_X(u)\right)\ is\ an\ even\ function\ in\ u$$

$$\therefore E[\sin(uX)] = \int_{-\infty}^{\infty} \sin(ux) f_X(x)dx \Rightarrow imag\left(\phi_X(u)\right)\ is\ an\ odd\ function\ in\ u$$

Clearly then any characteristic function can be written in the form $\phi_X(u) = f(u) + i \times g(u)$, where $f(u) = real\left(\phi_X(u)\right)$ and $g(u) = imag\left(\phi_X(u)\right)$.

Lastly, from calculus we have the following limit technique that will be used:
For any $a \neq 0$, we have that

$$\lim_{z\to 0}\frac{sin(z/a)}{z} = \lim_{z\to 0}\frac{sin(z/a)}{z/a}\times\frac{1}{a} = \left(\lim_{z\to 0}\frac{sin\left(\frac{z}{a}\right)}{\frac{z}{a}}\right)\times\frac{1}{a} = 1\times\frac{1}{a} = \frac{1}{a}.$$

The main result and its proof can now be given.

Theorem 3.1:

The cumulative distribution function for an asset price process can be expressed, by using the decomposition $\phi(u) = f(u) + i \times g(u)$*, as*

$$H(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)}{iu}du = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{f(u)\sin(ux)}{u} - \frac{g(u)\cos(ux)}{u}du.$$

Proof:

Firstly we start with formula (3.4):

$$H(x) = \omega = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)}{iu}du.$$

The following decomposition is then applied:

$$\phi(u) = f(u) + i \times g(u)$$

Thus

$$\begin{aligned}\omega &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{iux}\big(f(-u) + i\times g(-u)\big) - e^{-iux}\big(f(u) + i\times g(u)\big)}{iu}du \\ &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{e^{iux}\big(f(u) - i\times g(u)\big) - e^{-iux}\big(f(u) + i\times g(u)\big)}{iu}du \\ &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{f(u)\big(e^{iux} - e^{-iux}\big) - i\times g(u)(e^{iux} + e^{-iux})}{iu}du.\end{aligned}$$

Apply the identities:

$$sin(\theta) = \frac{(e^{i\theta} - e^{-i\theta})}{2i}, cos(\theta) = \frac{(e^{i\theta} + e^{-i\theta})}{2}$$

$$\omega = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{f(u)\,sin(ux)}{u}\,du - \frac{1}{\pi}\int_0^\infty \frac{g(u)\,cos(ux)}{u}du.$$

Define $I = \left(\omega - \frac{1}{2}\right)\pi$*, thus*

$$I = \int_0^\infty \left(\frac{f(u)\,sin(ux)}{u} - \frac{g(u)\,cos(ux)}{u}\right)du. \qquad \blacksquare$$

Some remarks:

The integral is well defined for all values of u. Clearly when using some numerical routine (e.g. basic trapezoidal rule, Simpson's rule, etc.) to calculate the integral one might worry about division by zero if $u = 0$ is substituted. For the first term the limit as $u \to 0$ is simply $x \times f(0) = x$. The second term

is trickier to consider but works out well. For any constant $a \neq 0$, we have

$$\lim_{z\to 0}\frac{g(z/a)}{z/a}=\lim_{z\to 0}\frac{\int_{-\infty}^{\infty}\sin\left(\frac{zx}{a}\right)f_X(x)dx}{z/a}=\lim_{z\to 0}\int_{-\infty}^{\infty}\frac{x\times\sin\left(\frac{zx}{a}\right)}{\frac{zx}{a}}f_X(x)dx=\int_{-\infty}^{\infty}xf_X(x)dx=E(X).$$

Expressing this in terms of characteristic functions:

$$E[X^n]=i^{-n}\left[\frac{d^n}{du^n}\phi_X(u)\right]_{u=0} => E[X]=-i\times\phi'(u)]_{u=0}=-i\phi'(0)=-i\times\big(f'(0)+i\times g'(0)\big)$$
$$\therefore E[X]=-i\times f'(0)+g'(0).$$

Note that

$$f(u)=\int_{-\infty}^{\infty}\cos(ux)\,f_X(x)dx => \ f'(0)=\int_{-\infty}^{\infty}-x\times\sin(0\times x)\,f_X(x)dx=0.$$

It follows that

$$E[X]=g'(0).$$

Since there are no other singularities or any other problems for other values of u it must be that the integral is well-defined. Additionally, the value of $g'(0)$, for some small h can be calculated as

$$g'(0)\approx\frac{g(0+h)-g(0)}{h}=\frac{g(h)}{h}.$$

An important corollary to theorem 3.1. now follows.

Corollary 3.1:

A formula for the density function can be given:

$$H'(x)=h(x)=\frac{1}{\pi}\int_0^{\infty}e^{-iux}\phi(u)du=\frac{1}{\pi}\int_0^{\infty}f(u)\,cos(ux)+g(u)\,sin(ux)\,du.$$

Proof:

The result follows from Theorem 3.1. by direct differentiation with respect to x. ■

An alternative proof of Corollary 3.1. provides insight into why this formula is expected to be computationally faster than formula (3.5). Start off with the formula (3.5):

$$\frac{1}{\pi}\int_0^{\infty}e^{-iux}\phi(u)du=\frac{1}{\pi}\int_0^{\infty}e^{-iux}\big(f(u)+i\times g(u)\big)du$$
$$=\frac{1}{\pi}\int_0^{\infty}\big(\cos(ux)-i\times sin(ux)\big)\big(f(u)+i\times g(u)\big)du$$
$$=\frac{1}{\pi}\int_0^{\infty}(f(u)\cos(ux)+g(u)\sin(ux))du+\frac{i}{\pi}\int_0^{\infty}(g(u)\cos(ux)-f(u)\sin(ux)\,du.$$

The density function is *real-valued* and is defined as the real part of the above equation (see for instance (3.17)). Consequently

$$h(x)=\frac{1}{\pi}\int_0^{\infty}f(u)\cos(ux)+g(u)\sin(ux)\,du.$$

Under this approach, in contrast to the FFT approach, all the complex number calculations are taken out from the start. Under the FFT approach, all the calculations are done and then the real part is taken (i.e. see (3.17)) which recovers the density function. The argument here is that by avoiding all complex-valued calculations the computational time taken to reach the same result by otherwise including such calculations should be less. This however requires expensive matrix multiplication and hence this quadrature method may be slower than the FFT method.

Computational efficiency: FFT vs new approach:

Below two numerical routines are given that allow the formulas in Theorem 3.1. and Corollary 3.1 to be calculated at the fastest possible speed while still retaining reasonable accuracy. The accuracy and speed of the calculations for the first routine are further investigated in Appendix D.

The conclusions in Appendix D are that the FFT is the faster method (0.067234567seconds to compute expanded $2^{16} = 32768$ points on the density function), but lacks accuracy from the third decimal point onwards. The Author's method is 3 times slower but accurate to the 10^{th} decimal place (0.175979252 seconds to compute 32768 points). These results only account for the core calculations and when the entire function is run then the differences in computational time disappear.

The numerical routines used in the investigation are given below are very simple and basic, which is the main reason the code runs particularly fast.

Numerical Routine 1: Calculating the density function quickly.

Simpson's rule is used to calculate the density function with a relatively *large step size* (see Appendix E) of $h = 0.25$:

$$\frac{1}{\pi}\int_0^\infty f(u)\cos(ux) + g(u)\sin(ux)\,du = \frac{1}{\pi}\int_0^\infty J(u;x)du$$

$$\approx \frac{h}{3\pi}\big(J(u_1;x) + 4J(u_2;x) + 2J(u_3;x) + 4J(u_4;x) + 2J(u_5;x) + \cdots 2J(u_{n-1};x) + J(u_n;x)\big).$$

Here the interval over which the variable u extends is truncated, since the characteristic function is essentially zero for all $u > N$, where N is some positive number, and u is subdivided into $n-1$ subintervals, appropriately indexed so that the i^{th} interval is given by $[u_i, u_{i+1}]$. Essentially to use this numerical approximation each of the function values at each u_i are multiplied alternately by 4's and 2's, except for the first and last function values.

An excerpt of the MATLAB® code is given here for convenience and the reader can pursue the details by examining the supporting material. The script calculates the density function using the parameters as given earlier for the transition density surface of the Double Heston Model.

```
load 'DHBImpliedVolParameters.mat'
%Computing the Characteristic Function at Time 1
x = logespace(log(5),log(2500),2496);
DBHChar = @(u)doublehestonChar(X0,v1,v2,r,a1,a2,b1,b2,sigma1,sigma2,rho1,rho2,tau(1),u);
tic
T = @(u) real(DBHChar(u));
W = @(u) imag(DBHChar(u));
u = 0:0.25:(40-0.25);
R = simpsonprep(T(u))';%Multiplies the function by alternating 4's and 2's
R2 = simpsonprep(W(u))';
R3 = u'*x;%Matrix Multiplication incorporating all values on the support
Q1 = 1/(pi).*(R*cos(R3)+R2*sin(R3))./exp(x).*0.25/3;
toc
```

```
Elapsed time to compute Q1 is 0.008730 seconds.
```

The total time to run the entire script is 0.010066 seconds (as determined by MATLAB® 's tic-toc function). Note that the density function is recovered with 2946 points and the error involved in the approximation is nearly zero.

The next numerical routine samples from the density function. It essentially extends the above numerical routine to give a full sampling algorithm that is extremely robust. Computationally this scheme is more costly to run, but can be optimised using some of the ideas in Appendix C.

Numerical Routine 2: Sampling from the cumulative density function via characteristic functions quickly.

Suppose that we are given a value from a uniform distribution, to sample from the cumulative distribution function, which is denoted by ω. Furthermore define

$$Q(x) = \omega - \left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{f(u)\sin(ux)}{u} - \frac{g(u)\cos(ux)}{u}du\right).$$

Hence

$$Q'(x) = -\frac{1}{\pi}\int_0^\infty f(u)\cos(ux) + g(u)\sin(ux)\,du.$$

Now apply Newton's Method to solve for x iteratively, given a starting value $x_1 = S(0)$ until $x_{n+1} - x_n < \epsilon, where\ \epsilon\ is\ some\ small\ number$ or until the maximum number of iterations has been exceeded:

$$x_{n+1} = x_n - \frac{Q(x_n)}{Q'(x_n)}.$$

An example of using this method is given by the following function, as implemented in MATLAB®

```
function x = FastCharSampling(S0,characteristic,omega)
u = 0:0.25:40-0.25;
y = log(S0);
T = @(u) real(characteristic(u));
W = @(u) imag(characteristic(u));
TT = simpsonprep(T(u))';
WW = simpsonprep(W(u))';
x = y;
for ii = 1:1000
    R = cos(u.*x);
    R1 = sin(u.*x);
    intexpression = @(u) T(u).*sin(x.*u)./u-W(u).*cos(x.*u)./u;
    Q = omega-(1/2+1/pi*integral(intexpression,0,50));%Ensures robust calculation of a
point on the cumulative distribution function
    Q1 = -1/pi*((TT*R'+WW*R1').*0.25/3);
    x1 = x-(Q/(Q1));
    if abs(x1-x) < 0.001 %epsilon = 0.001
       break
    end
    if ii == 1000
       display('Maximum number of iterations reached');
       break
```

```
    end
    x = x1;
end
x = x1;
end
```

This numerical routine is sufficiently robust and it converges to the correct density function for around 1000 simulations. The above sampling scheme is unbiased and simulates the true density function exactly. The script below gives the implementation in MATLAB®. While the procedure is fast, it takes around 35 seconds to produce 1000 simulations.

To increase speed and accuracy the Author proposes using Quasi-Monte Carlo techniques. Essentially the exact distribution is recovered for 1000 simulations in around 3.5 seconds.

Further improvements can be made by running vectorised code in C++ (see Appendix C) and calling the function from within MATLAB® which produces 1000 simulations in 0.057760 seconds. The code to achieve this as well as compare the accuracy of the simulation is given below for reference, (with some lines of code omitted to save space):

```
function DHBSampDemo(nrSim)

load 'DHBImpliedVolParameters.mat'
DBHChar = @(u)doublehestonChar(X0,v1,v2,r,a1,a2,b1,b2,sigma1,sigma2,rho1,rho2,1,u);

if nrSim > 1000
p = sobolset(1,'Skip',1e2,'leap',50);
p = scramble(p,'MatousekAffineOwen');
else
p = sobolset(1);
p = scramble(p,'MatousekAffineOwen');
end
U1 = net(p,nrSim)';
omega = U1;

tic
y = AppOptimalCharSampling_mex(1000,DBHChar,omega);
toc

histogram(y,50,'Normalization','pdf')

x = logespace(log(100),log(2500),2400);
R = @(p) exp(-1i.*p.*x).*DBHChar(p);
Q = 1/(pi).*real(integral(R,0,inf,'ArrayValued',true));
hold on
plot(x,Q,'r')
end
```

The function by *DHBSampDemo(1000) generates:*

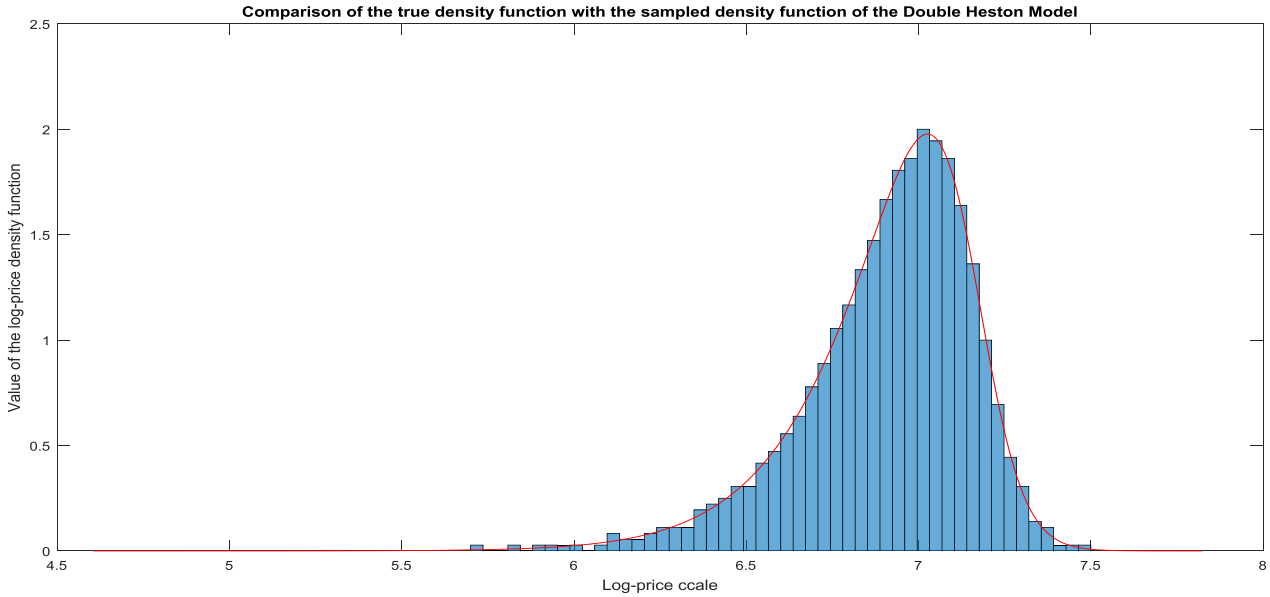

Figure 3.4: Comparison of the true density function with the sampled density function using an optimised sampling scheme.

Note that due to the code being optimised as it is written, one might expect that for instance with the Double Heston Model that if 1000 simulations take 0.057760 on average then 10000 simulations would take 0.57760 seconds. Due to the optimisation in the above scheme, the average time for 10000 simulations is around 0.3 seconds and for 100000 simulations the average time is 2.695671 seconds.

However as mentioned, using Quasi-Monte Carlo simulation techniques reduces the number of simulations as the convergence to the true distribution is faster for a smaller sample size. For the Double Heston Model considered here there is virtually no difference in the histograms produced by using 10000 or 100000 simulations.

Sampling from the transition density function can then be done quickly and certain non-path dependent European type options can be calculated. The above algorithms are effective for pricing non-path dependent derivatives, whilst most exotic equity derivatives are indeed path dependent. This calls for an alternative solution, given in the next subsection.

3.2.6. **An exact sampling scheme to calculate the price of any exotic derivative: A theoretical Monte Carlo technique to derivative pricing using characteristic functions.**

This section presents the Author's proposal to calculating any exotic derivative's price as an alternative to using Monte Carlo techniques. Drawing upon the material in the previous sections allows us to approach the problem in a slightly new manner. The approach below generates sample paths using the underlying model's characteristic function and the alternative formula for calculating the density function in the previous section.

The material in Appendix D is particularly relevant to the discussion in this section. It is recommended the reader works through Appendix D before proceeding. The Monte Carlo algorithm is given below how to improve this computationally. The technique is basic and is based on the recommendations in (Campolieti and Makarov, 2014), who have assumed that the transition density function is known. In our case we compute the transition density function numerically using

numerical routine 1.

The algorithm can now be specified. Note that we specify the log-density function's support in advance as a set of values, which can determined by using the Authors *logespace* function to ensure even spread (see the function in the supporting material). The Author's recommendation is to set the support as between 0 and 3 times the starting value of the asset for 1 year time to maturity, 0 to 10 times for 1-5 years of maturity and finally 0 to 20 times for up times to maturity up to 10 years.

There are also recommendations in literature, in particular one is given in section 3.3, as to determine the appropriate support for the log-price density function. Unfortunately such techniques are computationally expensive and for simplicity it works out better to specify the support in advance.

Algorithm 3.2 Exact Sample Path Generation Scheme Using Characteristic Functions

---

1.) Input a vector of times, $\boldsymbol{T} = \{t_j : j = 1,2, \dots, M\}$ with $t_0 \leq t_1 < t_2 < \cdots < t_M = T$ and obtain a vector of time differences $\mathbf{diffT} = \{t_k - t_{k-1}; k = 1,2, \dots M\}$.
2.) Generate n independent paths, $S_0^j, S_1^j, \dots, S_M^j$, using the transition PDF, i.e. each S_k^j is sampled from the density function $p(t_k - t_{k-1}; S_{k-1}^j, S_k^j)$, which is to be numerically recovered, with $j = 1,2, \dots n$ and fix the starting value at $S_0 = S_0^j \; \forall \, j$. The sample paths form a stochastic tree with $n * M$ nodes with the $(k,j)^{th}$ node corresponding to an exercise time t_k and asset price S_k^j.
3.) Density recovery is done using the following scheme:
3.1.) Define a maximal range for the support of the log-price density and the characteristic function. For the characteristic function this is set between 0 and 400, with intervals of 0.25. The number of elements on the support of the log-price density is S.
3.2.) Perform a matrix multiplication using the support of the density function and the support of the characteristic function:

$$u = [800 \text{ X } 1] \; matrix \; times \; x = [1 \text{ X } S] \; matrix$$

3.3) Let $RCos$ be the resulting matrix working out the trigonometric cos value of each element and similarly let $RSin$ be the resulting matrix by working out the $sine$ value of each element in the matrix formed in 3.2.
3.4) Take the real part of the characteristic function as TT and the imaginary part as WW, multiply all elements, except the first and last, in a sequence of alternating 2's and 4's.
3.5) Recover the log-price density function by using the formula:

$$(\text{TT} * \text{R4Cos} + \text{WW} * \text{R4Sin}) .* (0.25/(3\pi))$$

Here $*$ indicates matrix multiplication and $.*$ indicates element-wise multiplication.

Important Note: The recommendations given here is based on the numerical work the Author carried out and may not necessarily apply in all situations. For instance, in step 3.1. in Algorithm 3.2, the support of the characteristic function is set between 0 and 400. Naturally if the inputted times are spread far apart then if may be desirable to set the characteristic function's support to range between smaller values, e.g. 0 to 50. This is simply due to theory behind Fourier Analysis which for our purposes implies the following rule:

If the time to maturity is very small, then the support of the characteristic function must be large. Conversly, for large times to maturity, the support of the characteristic function can be small.

This rule follows from the fact that a density function with a thin and large single peak has a

characteristic function that dies down much more slowly as $u \to \infty$. Similarly, the intervals of step 0.25 in Algorithm 3.2 are determined both numerically and heuristically (see Appendix D for a justification) to provide reasonable accuracy. Nothing stops these intervals from being *smaller*, except additional computation time, should this be a requirement.

An example of implementing Algorithm 3.2 in the MATLAB® environment for the Double Heston Model:

```
function y = DHBMonteCarloProto(tau,nrSim)
Parameters = load('DHBImpliedVolParameters.mat');
X0 = Parameters.X0;
v1 = Parameters.v1;
v2 = Parameters.v2;
r = Parameters.r;
a1 = Parameters.a1;
a2 = Parameters.a2;
b1 = Parameters.b1;
b2 = Parameters.b2;
sigma1 =Parameters.sigma1;
sigma2 = Parameters.sigma2;
rho1 = Parameters.rho1;
rho2 = Parameters.rho2;
%Assume that appropriate parameters are loaded.
omega = gather(rand(length(tau),nrSim,'gpuArray'));
y = zeros(length(tau),nrSim,'gpuArray');
y(1,:) = gpuArray(X0);
u = 0:0.25:400-0.25;
w = logespace(log(1),log(3000),2999);
R3 = gpuArray(u'*w);%Matrix Multiplication incorporating all values on the support
R4Cos = cos(R3);
R4Sin = sin(R3);
[A,B1,B2] = doublehestonChar(r,a1,a2,b1,b2,sigma1,sigma2,rho1,rho2,tau(1),u);
V1 = zeros(length(tau)+1,nrSim);V2 = zeros(length(tau)+1,nrSim);V1(1,:) = v1;V2(1,:)=v2;
V1(2:end,:) = simulateCIR_mex(a1,b1,sigma1,v1,tau,nrSim)';
V2(2:end,:) = simulateCIR_mex(a2,b2,sigma2,v2,tau,nrSim)';
for ii = 1:nrSim
    X0 = log(1000);
    for jj = 2:length(tau)
        y(jj,ii) = AppOptimalCharSampling1(X0,V1(jj-1,ii),V2(jj-
1,ii),A,B1,B2,omega(jj,ii),R4Cos,R4Sin,w,u);
        X0 = y(jj,ii);
    end

end
y = gather(y);
end

function x = AppOptimalCharSampling1(X0,v1,v2,A,B1,B2,omega,R4Cos,R4Sin,w,u)
%Non-Parallel Version
Q = doublehestonCharComplete(X0,v1,v2,A,B1,B2,u);
T = real(Q);
W = imag(Q);
TT = simpsonprepGPU(T)';
```

```
WW = simpsonprepGPU(W)';
Q1 = gather(bsxfun(@times,mtimes(TT,R4Cos)+mtimes(WW,R4Sin),(0.25/(3*pi))));
Q2 = cumtrapz(w,Q1);
kk = roundnearest(omega,Q2);
x = w(kk);
end


function y = roundnearest(omega,cumdist)
diff = cumdist - omega;
diff(diff<0)=0;
diff(diff>0)=1;
cumulative = cumsum(gpuArray(diff));
k = gather(find(cumulative==1));
y =k;
end

function [A,B1,B2] = doublehestonChar(r,a1,a2,b1,b2,sigma1,sigma2,rho1,rho2,tau,u)

d1 = sqrt((rho1.*sigma1.*u.*1i-b1).^2+sigma1.^2.*(u.*1i+u.^2));
d2 = sqrt((rho2.*sigma2.*u*1i-b2).^2+sigma2.^2.*(u.*1i+u.^2));
g1 = (b1-rho1.*sigma1.*u.*1i-d1)./(b1-rho1.*sigma1.*u.*1i+d1);
g2 = (b2-rho2.*sigma2.*u.*1i-d2)./(b2-rho2.*sigma2.*u.*1i+d2);
B1 = (b1-rho1.*sigma1.*u.*1i-d1)./(sigma1.^2).*(1-exp(-d1.*tau))./(1-g1.*exp(-d1.*tau));
B2 = (b2-rho2.*sigma2.*u.*1i-d2)./(sigma2.^2).*(1-exp(-d2.*tau))./(1-g2.*exp(-d2.*tau));
A = r.*u.*1i.*tau+(a1./sigma1.^2).*((b1-rho1.*sigma1.*u.*1i-d1).*tau-2.*log((1-g1.*exp(-
d1.*tau))./(1-g1)))...
    +(a2/sigma2.^2).*((b2-rho2.*sigma2.*u.*1i-d2).*tau-2.*log((1-g2.*exp(-d2.*tau))./(1-
g2)));


end

function y = doublehestonCharComplete(Xt,v1,v2,A,B1,B2,u)
         y = exp(Xt.*u.*1i+A+B1.*v1+B2.*v2);
end
```

The code, as above, takes on average 1.75 seconds to produce 1 simulation (365 points on a sample path). The above code works, however it should truly only be considered as a theoretical alternative. Generating sample paths using the characteristic function takes considerable computational time, even when resorting to advance computational techniques. There is also no guarantee of the robustness of the algorithm.

It appears numerically that the finer the support of the density function and of the characteristic function the better the path wise and terminal distribution convergence is for the above scheme. However by choosing finer grids we are left with serious computational penalties which makes this method unattractive from a practical viewpoint.

Nonetheless, further investigation into using the characteristic function to generate exact sampling schemes is left to future research endeavours.

3.2.7. **Summary of the Monte Carlo techniques used to price exotic derivatives**

In this final section a short summary is given to bring together the material in the previous sections. Firstly, when European options need to be priced one can always use the transition density function, compute it quickly using the Author's quick method, and then work out the value of the option. The result is exact and requires no Monte Carlo techniques.

Secondly, when working out the price of path dependent options, where generally no pricing formula exists, one has to simulate values. The first method is to apply a Euler discretisation scheme to the entire SDE. This method will run fast, however around 200000+ simulations are needed to provide somewhat accurate results. Decreasing the time step reduces the bias but unfortunately increases computational time.

The second method is to apply an exact sampling scheme to the volatility or stochastic interest rate path, use Brownian reconstruction to recover the Brownian motion increments and then conditional on this volatility path, generate the asset path. In the next chapter a comparison is done on whether the Euler or the Brownian Motion Reconstruction technique is preferred for general path-dependent European option pricing.

The last method presented here uses only the characteristic function to create sample paths. This technique is computationally slow and can be unreliable, which is why it is given only as a theoretical alternative. We now move on to considering some analytical methods to pricing derivatives.

3.3. **General techniques to price Call options using stochastic volatility models**

The list of models in table 3.1.may make it seems as if we have a daunting task by specifying Call and (via Put-Call parity or otherwise) Put options for each of the 3 models. In fact there is effectively only "one model" that we need to consider. All of the models in table 3.1 have characteristic functions that are well-defined and thus any general pricing scheme, using characteristic functions, can be used for these models.

Hence the first part the following discussions is devoted to the pricing Call options under these types of model highlighting some of the pricing methods and material found in literature. Thereafter various exotic options are considered. For each exotic option it is instructive to see the BS approach and hence the usual BS formulas are given, where applicable.

3.3.1. **Call option strategies for models admitting characteristic functions**

Equations (3.1) to (3.3) have given us a workable method to calculating Call option prices, however there are other methods in literature that could also be used. A great overview of these methods can be found in (Kienitz and Wetterau, 2012) who also investigated some of these methods and compared them. Thus in the interest of saving space only three methods are considered here, partly based on the conclusions of (Kienitz and Wetterau, 2012). The methods are presented in the chronological order as they appeared in literature.

Before proceeding, since the Monte Carlo section gave a workable method to pricing Call options (section 3.2.5) the question remains why do we need to consider the literature regarding the calculation of Call options? The answer depends on the purpose of the calculation. For instance if we

are interested in computing the Greeks then the transition density function has to be numerically differentiated in order to do this. In fact, as given in section 3.5, there are semi-analytical and more powerful methods of calculating the Greeks based directly off of the method used to price Call options.

Secondly, for the calibration of stochastic volatility models it is desirable (though not necessary) to have some formula (semi-analytical or analytical) available to compute the objective function (see Appendix F) used in the calibration procedure. This certainly speeds up the calibration procedure instead of sampling option prices as the formulas given can be evaluated at a much faster speed.

3.3.1.1. **The Lewis method**

A brief outline of the Lewis Method (Lewis, 2001) method is as follows. Firstly to use the pricing formula (3.57) below, defined for non-path dependent payoffs, one has to compute the Fourier Transform of the payoff function given in terms of the log-price $X(t)$. For a call option this is given as:

$$C(T,K) = \max(S(T)-K,0) => \max\left(e^{X(T)}-K\right) = w\left(X(T)\right)$$
$$\therefore\ \mathcal{F}\left(w(x)\right)(z) = \int_{-\infty}^{\infty} e^{izx}\max(e^{x}-K,0)dx = -\frac{K^{iz+1}}{z^2-iz}. \tag{3.56}$$

Generally the Fourier Transform of the payoff is denoted $\widehat{w}(x) = \mathcal{F}\left(w(x)\right)$ and is assumed to satisfy some regularity (the derivative exists and the function is single-valued) strip in the complex plane, $a < \mathrm{Im}(z) < b$, which for (3.56) is by inspection $\mathrm{Im}(z) > 1$. The general option pricing formula is

$$V_{call}(T) = \frac{e^{-rT}}{2\pi}\int_{iv-\infty}^{iv+\infty} e^{-izY}\phi_T(-z)\widehat{w}(z)dz. \tag{3.57}$$

Here $Y = \ln(S_0) + (r-q)T$. We assume that $w(z)$ is Fourier integrable in some strip in the complex plane and the transform is given by $\widehat{w}(z)$ with $z \in \mathcal{S}_w$, where $\mathcal{S}_w$ is some strip in the complex plane. The log-price X_T is assumed to have the characteristic function ϕ_T that is assumed to be regular in the strip $\mathcal{S}_X = \{z = u + iv : v \in (a,b); a < -1\ \&\ b > 0\}$. Consequently the number v in (3.57) is $v = \mathrm{Im}(z)$ where $z \in\ \mathcal{S}_V = S_w \cap\ S_X^*$ and the $*$ notation simply indicates the conjugate or reflected strip about the real axis: $\mathcal{S}_X^* = \{z = u - iv : v \in (a,b); a < -1\ \&\ b > 0\}$.

Thus in order to apply (3.57) we need to carefully consider where the strip $\mathcal{S}_V$ is defined. If it is empty then formula (3.57) cannot be applied. The formula (3.57) can be specified for call options as:

$$V_{call}(S,K,T) = -\frac{Ke^{-rT}}{2\pi}\int_{iv_1-\infty}^{iv_1+\infty} e^{-izk}\phi_T(-z)\frac{dz}{z^2-iz}, \qquad v_1 \in (1,\beta), \tag{3.58}$$

with $k = \log\left(\frac{S}{K}\right) + (r-q)T$. Here the assumption still applies that $\mathcal{S}_X^* \notin\ \emptyset$ and $\mathcal{S}_X^* = \{z : \alpha < \mathrm{Im}(z) < \beta; \alpha < 0, \beta > 1\}$. (Lewis, 2001), however, uses Residue Calculus, to derive a more "useful" alternative to (3.58):

$$V_{call}(S,K,T) = Se^{-qT} - \frac{1}{\pi}\sqrt{SK}e^{\frac{(r+q)T}{2}}\int_0^{\infty}\mathcal{R}\left[e^{iuk}\phi_T\left(u-\frac{i}{2}\right)\right]\frac{du}{u^2+\frac{1}{4}}. \tag{3.59}$$

For clarity, note that $\mathcal{R}()$ denotes the real part of a complex number. We are not going to use the Lewis method and hence we proceed to the Attari method.

3.3.1.2. **The Attari method**

The Attari Method is a peculiar method that has not received much attention in literature yet it is

based on simple complex algebra (no pun intended…). It reduces the number of integrations in (3.1)-(3.3) and is due to (Attari, 2004). The formula is shown in the article to be fast and very accurate. The final result is stated for Call and Put options below:

$$V_{call}(S(t),K,T) = S(t) - e^{-r(T-t)}\left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty I_A(x)dx\right). \tag{3.60}$$

$$V_{Put}(S(t),K,T) = S(t) - e^{-r(T-t)}\left(\frac{1}{2} - \frac{1}{\pi}\int_0^\infty I_A(x)dx\right). \tag{3.61}$$

Here

$$I_A(x) = \frac{\left(\left[\mathcal{R}\left(\phi(x)\right) + x^{-1}\mathrm{Im}\left(\phi(x)\right)\right]\cos(\beta x) + \left[\mathrm{Im}\left(\phi(x)\right) - x^{-1}\mathcal{R}\left(\phi(x)\right)\right]\sin(\beta x)\right)}{1+x^2},$$

and $\beta = \log\left(\frac{e^{-r(T-t)}K}{S(t)}\right)$.

Unlike the Lewis method, we cannot apply the Attari method to the valuation of other payoff functions. However we can more easily calculate the Greeks as the above equations are easily differentiable. Since this pricing method is restrictive our discussion of the Attari method ends here.

3.3.1.3. **The COS method**

The COS method was introduced by (Fang and Oosterlee, 2008) and when examined against other Call option pricing methods (Kienitz and Wetterau, 2012) found this method outperformed many other pricing methods in terms of pricing various exotic derivatives, computational times and robustness. Hence the COS method will serve a priori as our preferred pricing choice when pricing European options. The techniques are based on Fourier Analysis. Various ideas from Fourier Analysis have already been introduced, nonetheless, let us recall some of the other basic ideas (see (Folland, 2009) for more detail).

Let $f(x)$ be a function on $[0,\pi]$ (for generality we assume f is complex valued and various hypothesis on f are handled in (Folland, 2009) and assume a Cosine expansion of f is

$$f(\theta) = \frac{1}{2}A_0 + \sum_{k=1}^{\infty} A_k\cos(k\theta)\,;\;\; A_k = \frac{2}{\pi}\int_0^\pi f(\theta)\cos(k\theta)\,d\theta. \tag{3.62}$$

For a general interval $[a,b]$ one simply needs to apply a change of variables in (3.62), i.e. from θ to x:

$$\theta = \frac{x-a}{b-a}\pi.$$

Hence (3.62) becomes

$$f(x) = \frac{1}{2}A_0 + \sum_{k=1}^{\infty} A_k\cos\left(k\pi\frac{x-a}{b-a}\right). \tag{3.63}$$

and

$$A_k = \frac{2}{b-a}\int_a^b f(x)\cos\left(k\pi\frac{x-a}{b-a}\right)dx. \tag{3.64}$$

Now following the arguments in (Fang and Oosterlee, 2008) we start with approximating the density function of the log-price using Cosine expansions, assuming the characteristic function of the log-price exists:

$$F_k = \frac{2}{b-a}\mathcal{R}\left(\phi\left(\frac{k\pi}{b-a}\right)\cdot\exp\left(-i\frac{ka\pi}{b-a}\right)\right). \tag{3.65}$$

Thus a truncated series expansion for the density function (with (truncated) support on $[a,b]$) can be

found by applying (3.65) to (3.63):

$$f(x) = \frac{1}{2}F_0 + \sum_{k=1}^{N-1} F_k \cos\left(k\pi \frac{x-a}{b-a}\right). \tag{3.66}$$

Even for small numbers N, the approximation in (3.66) is relatively accurate and in Appendix C a quick practical example is given by considering f as the density function of the Student's t-distribution. In the interest of space the end result for a general option price is given together with the pricing formula for a call option:

$$V(x,T) \approx e^{-rT} \sum\nolimits_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a}; x\right) e^{-ik\pi\frac{a}{b-a}}\right) V_k. \tag{3.67}$$

and

$$V_k = \frac{2}{b-a}\int_a^b v(y,T)\cos\left(k\pi\frac{y-a}{b-a}\right)dy. \tag{3.68}$$

Here $v(y,T)$ is the payoff function of the log-price and x represents the value of the process now ($T=0$) and y the value of the process at time T. The $*$ in (3.67) indicates that the $k=0$ term is weighted by a factor of ½. For a Call and Put options (3.68) can be formulated analytically by specifying the payoff in terms of the log-price as was done in (3.66):

$$V_k^{Call} = \frac{2}{b-a}K\big(\chi_k(0,b) - \psi_k(0,b)\big). \tag{3.69}$$

$$V_k^{Put} = \frac{2}{b-a}K\big(-\chi_k(0,b) + \psi_k(0,b)\big). \tag{3.70}$$

$$\chi_k(c,d) = \frac{1}{1+\left(\frac{k\pi}{b-a}\right)^2}\left[\cos\left(k\pi\frac{d-a}{b-a}\right)e^d - \cos\left(k\pi\frac{c-a}{b-a}\right)e^c + \frac{k\pi}{b-a}\sin\left(k\pi\frac{d-a}{b-a}\right)e^d - \frac{k\pi}{b-a}\sin\left(k\pi\frac{c-a}{b-a}\right)e^c\right]. \tag{3.72}$$

$$\psi_k(c,d) = \begin{cases}\left[\sin\left(k\pi\frac{d-a}{b-a}\right) - \sin\left(k\pi\frac{c-a}{b-a}\right)\right]\frac{b-a}{k\pi} & k \neq 0 \\ d-c & k=0\end{cases} \tag{3.73}$$

Lastly, the appropriate truncation range for the density function can be given by the following formula:

$$[a,b] \coloneqq \left[c_1 - L\sqrt{|c_2| + \sqrt{|c_4|}}, c_1 + L\sqrt{|c_2| + \sqrt{|c_4|}}\right]. \tag{3.74}$$

Here the $c_n's$ denote the n^{th} cumulant of $\log\left(\frac{S(T)}{K}\right)$ which can be derived from the characteristic function (often calculated numerically using finite difference schemes) via $k(u) = \log\big(\phi(iu)\big)$ and noting that $c_n = \frac{\partial^n k}{\partial u^n}\big|_{u=0}$. Here $L = 10$, as per (Fang and Oosterlee, 2008)'s recommendation.

3.3.1.4. **Adapting the COS method to stochastic volatility models with stochastic interest rates**

The apt reader would have noticed that formulas (3.1) to (3.3) and the formulas under the previous subsections implicitly assume some constant interest rate. This is problematic for the H2-CIR models, as by construction the interest rates are stochastic. Luckily this is not a problem and the solution is based on the previous discussions and the article by (Grzelak, Oosterlee and Van Weeren, 2011).

What is extremely important here is to realise that there is a difference between a discounted contingent claim ***at*** some future time T and discounting cashflows *from* time T. For instance, using a

stochastic interest rate model, cash flows can be discounted using zero coupon bonds derived from such a model. Discounted contingent claims arising from stochastic interest rates is shown below. Both of these ideas are important for the work that follows.

For instance, recall that (Campolieti and Makarov, 2014) gave the following pricing function for a zero-coupon bond price as stated in Chapter 2:

$$Z\big(t,T,r(t)\big) = E_Q\left[e^{-\int_t^T r(u)du}\ |\mathcal{F}_t\right]. \tag{3.75}$$

Under the CIR Model an explicit formula can be given for this. From (3.10), which is repeated for reference as (3.76) below, we simply set $u = 0$ and then (3.75) is calculated:

$$\phi(u,\boldsymbol{X}(t),t,T) = E_{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)ds + iu^T\boldsymbol{X}(t)\right)\middle|\mathcal{F}(t)\right] = \exp\big(A(u,\tau) + B^T(u,\tau)\boldsymbol{X}(t)\big). \tag{3.76}$$

In other words one simply uses the characteristic function of a stochastic volatility model with stochastic interest rates and set $u = 0$ to obtain $Z(t,T)$. However, one cannot simply discount by zero coupon bonds when pricing derivatives using semi-analytical formulas, in other words replacing the discount factor e^{-rT} by $Z(0,T)$, see (3.75). This is due in part because the interest rate process drives the drift of the underlying asset price through a non-zero correlation and that the interest rate process is stochastic. To understand this idea the discounted and undiscounted density functions from the H2-CIR Model are plotted below (being derived from numerically inverting (3.76)). The parameters to produce the bottom graph can be found in the supporting material. What is important to notice is that the discounted density function does not integrate to one. Indeed, if one calculates $Z(t,T)$ from (3.76) and divides each point $f(x)$ by this value then the density function integrates exactly to one, giving essentially the undiscounted density function:

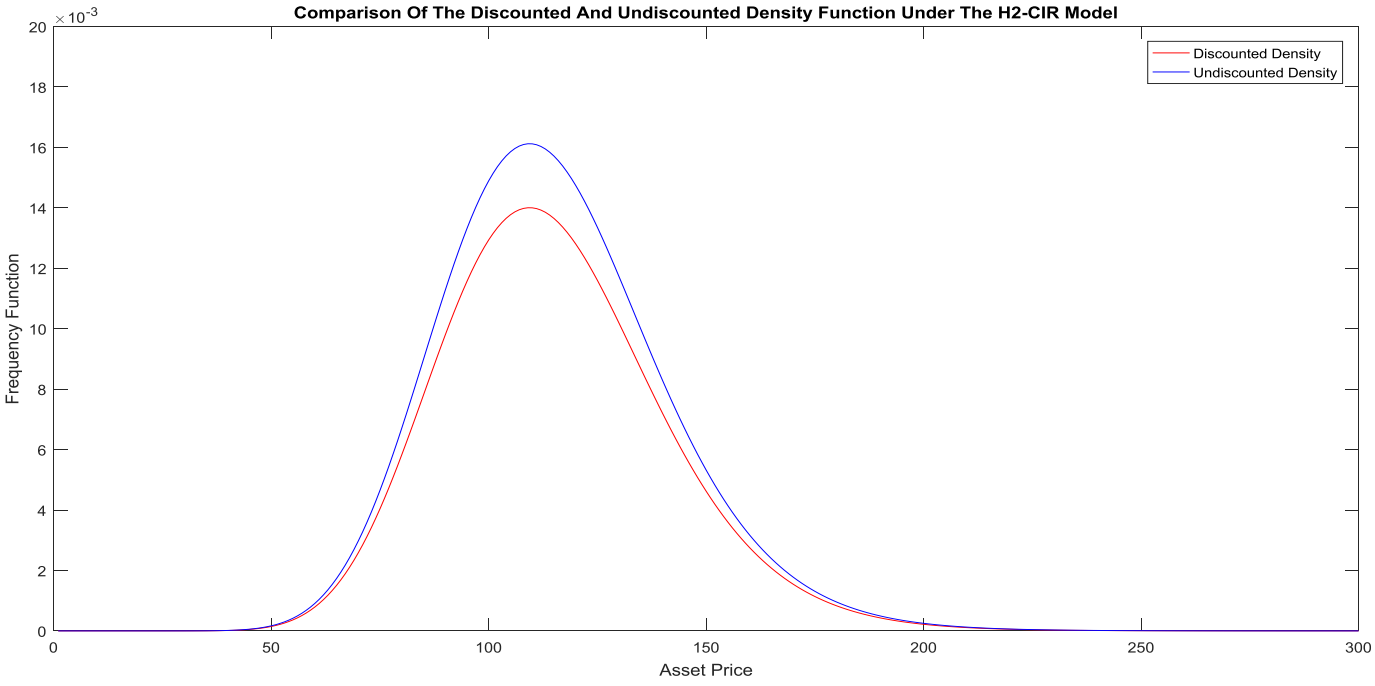

Figure 3.5: Comparison of the discounted and undiscounted density functions under the H2-CIR Model

The question is then how does one use the discounted density function? Firstly to understand how this works consider the question of sampling from the undiscounted density function. Suppose that a uniform random number is generated as 0.6064 and corresponds to sampling an asset value from the undiscounted density of 120 (this is exact regarding the undiscounted density above). Now the exact value of the zero-coupon bond is $Z(0,1) = 0.8686$. If one takes the uniform random number 0.6064 and deflate it by $Z(0,1)$ then one obtains 0.5267. Taking this value as a new uniform variate and

sampling from the *discounted* density function one obtains an asset value of 120. Thus one can sample faithfully from the discounted density function provided the uniform random numbers are adjusted by the zero-coupon bond price.

As given by (Grzelak, Oosterlee and Van Weeren, 2011) a more general formula for any European contingent claim can be developed through discounting by stochastic interest rates:

$$V(t,T;x,y) = E^{\mathbb{Q}}\left[e^{-\int_t^T r(s)ds}\, V(T,S_T)|\mathcal{F}(t)\right] = \int_{\mathbb{R}} V(T,y)\widehat{f_Y}(y|x)dy$$

$$= \int_{\mathbb{R}} V(T,y) \int_{\mathbb{R}} e^z f_{Y,Z}(y,z|x)dz\, dy\,;\;\; z = -\int_t^T r(s)ds.$$

Note that discounted characteristic function in (3.76) can be written as:

$$\phi(u,\boldsymbol{X}(t),t,T) = \int\int_{\mathbb{R}} e^{z+iuy} f_{Y,Z}(y,z|x)dzdy = \int_{\mathbb{R}} e^{iuy}\, \widehat{f_Y}(y|x)dy. \qquad (3.77)$$

as usual the function $\hat{f}$ is the (transition) density function, which we recover numerically from the characteristic function and $V(T,y)$ is the option payoff specified in terms of the log-price. Hence any European claim can be priced by simply inverting (3.77) as done in (3.5). Specifically this means that when using the COS method, one simply does not apply the discounting factor as it is already incorporated into the density function. Loosely speaking, the discount factor can be thought of as having been moved into the expected value, as is shown in (3.21). Here the discount factor "deflates" the density function instead of the payoff function. Using the COS method one can think of the discount factor in (3.76) being grouped with the characteristic function, to give a "discounted" characteristic function. Of course this is only an informal explanation.

The reader should recall that the characteristic function for the H2-HW and H2-CIR models involves solving a set of partial differential equations numerically. This is done by using the well-known Runge-Katta method, by using the MATLAB® built-in solver *ode45*.

All the models in table (3.1) have had their Call option prices stated in the various subsections above. Additionally, some hints were given as to a general option pricing strategy, although this is covered in section 3.4 in more depth. Once Call option prices are specified the calculation of the Greeks can be done as is in section 3.5.

3.4. **Techniques from literature and other suggestions to pricing exotic equity options**

In this section the discussions under Section 3.2.for pricing various exotic equity derivatives are extended. The term "exotic" is taken here to mean any option that is not a standard Call or Put option with European or American style exercise features. Additionally, the exotic options here do not involved foreign exchange rates, pricing on more than one underlying or any non-standard features. The focus is mainly on path-dependent options, although a strategy for non-path dependent options with American or Bermudan exercise styles is also given.

Numerous pricing strategies for and under specific models are given. Firstly, a definition is given for each type of exotic equity derivative given here (only some of the more popular exotic derivatives are considered). Additionally, the equivalent BS pricing formula is given for reference, if such a formula exists and can be found in the textbook by (Campolieti and Makarov, 2014).

It is impossible to cover all the strategies and there may well be some strategies in literature that the Author missed. Not all methods presented below are complete. For instance, the COS method may be

extended to Bermudan and Barrier options as found in literature but not necessarily to Asian options as there (to the Author's knowledge) has not been work done on it.

Also, the COS method for Bermudan options may work for some of the models in table 3.1 but not others like the H2-CIR Model. This is largely due to the derivation involved with the underlying methods. As a result, there are many subsections below, but where possible models are grouped according to a common strategy. Lastly, in some cases only a single algorithm is given by the Author that provides an implementation strategy as the scope of the pricing literature may not have solved the particular pricing problem meaningfully.

3.4.1. **General linear combinations of European options**

There are a number of European options that can be viewed as linear combinations of call and put options. Some common types include Butterfly Spreads, Straddles, Strips, Straps and Strangles. All of them are linear combinations of Call or Put options with the same or different strikes but with the same expiry dates and without dividends. Hence a general formula can be constructed for them.

The general payoff formula for a linear combination of n European Options, with underlying $S(t) \coloneqq S$, is:

$$\Lambda(S) = \sum_{i=1}^{n} (A_i S + B)\mathbb{I}_{\{a_i \leq S \leq b_i\}}. \tag{3.78}$$

A more useful equivalent alternative of (3.78) is given here:

$$\Lambda(S) = \sum_{i=1}^{n} \{A_i S\left(\mathbb{I}_{\{S \geq a_i\}} - \mathbb{I}_{\{S \geq b_i\}}\right) + B_i\left(\mathbb{I}_{\{S \geq a_i\}} - \mathbb{I}_{\{S \geq b_i\}}\right)\}. \tag{3.79}$$

Here the assumptions include that A_i & B_i are any real constants and that the $a_i's$ & $b_i's$ form non-overlapping intervals: $0 \leq a_1 < b1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n \leq \infty$. For a butterfly spread with strikes $K_1 < K_2 < K_3$, $K_2 = (K_1 + K_3)/2$ the payoff as given by (3.79) is

$$\Lambda(S) = (S - K_1)\mathbb{I}_{\{K_1 \leq S < K_2\}} + (K_3 - S)\mathbb{I}_{\{K_2 \leq S < K_3\}}. \tag{3.80}$$

where the parameters in (3.80) correspond to $A_1 = 1, B_1 = -K_1, a_1 = K_1, b_1 = K_2, A_2 = -1, B_2 = K_3, a_2 = K_2, b_2 = K_3$.

Pricing the resulting option that has payoff $\Lambda(S)$ under the BS formula is derived here and consequently gives us a strategy for pricing options by using (3.1) & (3.2) for the Fourier-Pricing Models. This is derived in Appendix C, while the BS derivation is given here under risk-neutral measure $\mathbb{Q}, E_{\mathbb{Q}} \coloneqq \tilde{E}; E_{\mathbb{Q}}(\cdot\,|F_t) \coloneqq \tilde{E}_t$:

$$\begin{aligned}
V(t,T,S) &= e^{-r(T-t)}\tilde{E}_t\left(\Lambda\big(S(T)\big)\right) \\
&= e^{-r(T-t)}\sum_{i=1}^{n}\Big\{A_i\big(\tilde{E}_t\big[S(T)\mathbb{I}_{\{S(T)\geq a_i\}}\big] - \tilde{E}_t\big[S(T)\mathbb{I}_{\{S(T)\geq b\}}\big]\big) \\
&\quad + B_i\left(\widetilde{\mathbb{P}}_t(S(T) \geq a_i) - \widetilde{\mathbb{P}}_t(S(T) \geq b_i)\right)\Big\} \\
&= \sum_{i=1}^{n}\left\{A_i S\left[\mathcal{N}\left(d_+\left(\frac{S}{a_i}, T-t\right)\right) - \mathcal{N}\left(d_+\left(\frac{S}{b_i}, T-t\right)\right)\right]\right. \\
&\quad \left. + e^{-r(T-t)}B_i\left[\mathcal{N}\left(d_+\left(\frac{S}{a_i}, T-t\right)\right) - \mathcal{N}\left(d_+\left(\frac{S}{b_i}, T-t\right)\right)\right]\right\}.
\end{aligned} \tag{3.81}$$

The $d_{\pm}(x,y)$ function is as defined in (1.4). Formula (3.81) is implemented in MATLAB® as *gbmderivativep.m* by the Author and the user is required to specify a matrix with the appropriate constants for the given payoff (see Appendix A for more details).

Pricing these types of derivatives is done by simply applying the Call option formulas in section 3.2 and using the Put-Call parity relationship to derive the value of Put options (or otherwise if a formula exists for the Put options).

3.4.2. **Bermudan and American options**

Bermudan options are options that can be exercised at a series of fixed dates which is denoted by the vector $\boldsymbol{T}$:

$$\boldsymbol{T} = \{t_j : j = 1,2,\dots,M\}.$$

In other words the set of exercise dates satisfy $t_0 \le t_1 < t_2 < \cdots < t_M = T$. Theoretically one would exercise the option when the option value, defined in (3.82) below, is the greatest during the life of the option at say time t_j, meaning that if $V(t)$ denotes the option value then $\forall\, t_j \in \boldsymbol{T} \Rightarrow V(t_{j-1}) < V(t_j)\ \&\ V(t_j) > V(t_{j+1})$. If no such t_j can be found then $t_j = \infty$, i.e. the option expires worthless.

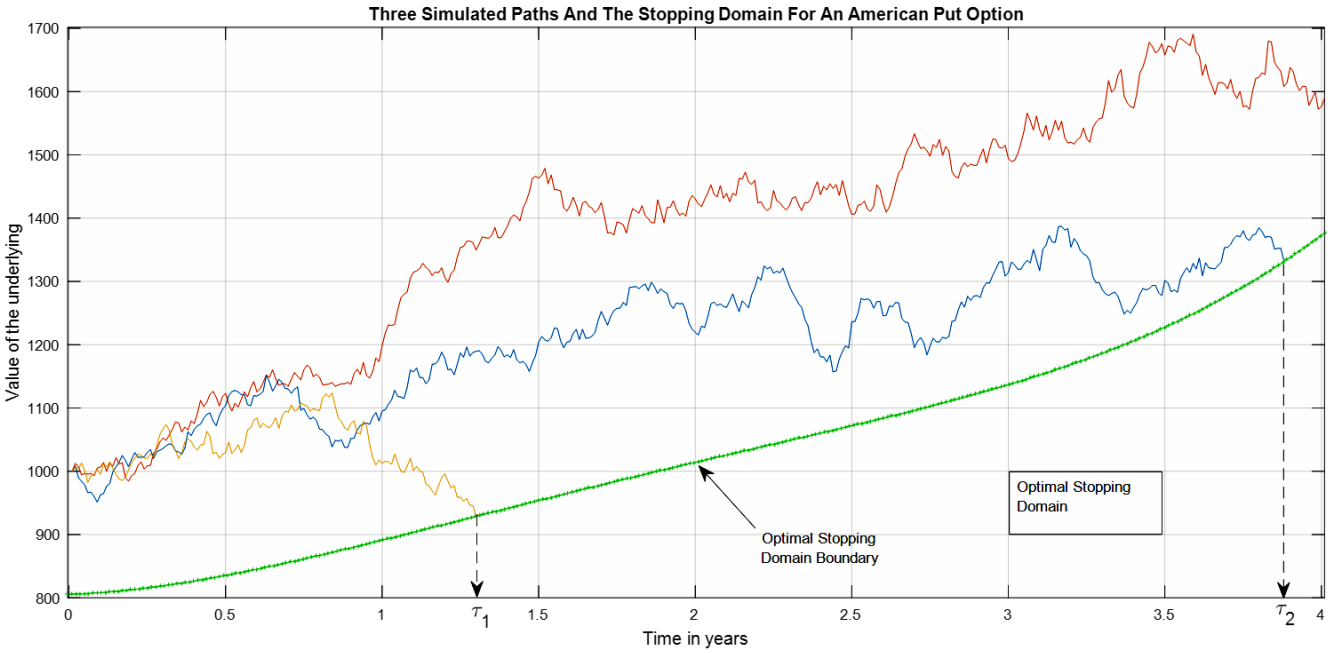

Figure 3.6: Three simulated paths and the stopping domain for an American Put option.

As (Campolieti and Makarov, 2014) state: "The price of such a contract is defined as the maximum arbitrage-free price of the contract under all possible exercise policies". Statistically speaking an exercise policy forms a stopping time $\mathcal{T}$ for the underlying process $S(t)$ and we denote the set of all stopping times, including ∞, by $\mathcal{S}_{\boldsymbol{T}}$. Hence at some time t, assuming the option hasn't been exercised already, the value of the option at time t is given by

$$V(t,S) = \sup_{\mathcal{T}\in\mathcal{S}_{\boldsymbol{T}}\cap[t,T]} E_{\mathbb{Q}}\left[e^{-r(T-t)}\Lambda\big(S(\mathcal{T})\big)\middle|S(t) = S\right]. \tag{3.82}$$

In the option pricing literature there is a distinction between the intrinsic value of the option and the continuation value at time t. The former is simply the payoff of the option if it were to be immediately exercised, denoted by $\Lambda\left(S(t_j)\right)$, while the latter has a more technical definition:

$$V^{cont}(t_j, S) = E_{\mathbb{Q}}\left[e^{-r(t_{j+1}-t_j)}V\left(t_{j+1}, S(t_{j+1})\right)\middle|S(t_j) = S\right]. \quad (3.83)$$

This simply represents the time t_j value of the option that has not been exercised yet. Rationally, if $\Lambda\left(S(t_j)\right) > V^{cont}(t_j, S)$ then an investor would definitely exercise the option as the cash value that he will obtain in the market by exercising the option is greater than the option value (i.e. the value he would obtain by selling it in the market). Conversely, if $\Lambda\left(S(t_j)\right) < V^{cont}(t_j, S)$ then the investor would not exercise his option as it is better to sell it at price $V^{cont}(t_j, S)$ than exercise it and receive less money. In fact $\Lambda\left(S(t_j)\right) = V^{cont}(t_j, S)$ corresponds to the optimal exercise strategy and forms the boundary of the stopping domain. An example is illustrated in figure 3.6 above.

Although American options are not strictly classified as exotic options, they are nonetheless extremely important as a large part of derivative contracts traded on exchanges are America in nature. Essentially American options generalise Bermudan options in that they can be exercised at any moment until expiry. Above, in Figure 3.6, is an example of three stock price simulations and the optimal stopping domain for an American Put option. Two of the paths lead to a situation where the option should be exercised while the last option expires worthless.

The determination of the optimal stopping domain is included in some of the discussion below. Although it might be possible to use the above stopping domain to value the option, see (3.82), it is generally difficult to construct the stopping domain for an arbitrary payoff (due to the intricate geometry involved). In any case, for the COS method that follows, the optimal stopping boundary will be calculated explicitly.

The only problem we now have is to develop some pricing strategy for our models. Firstly a general algorithm is given which forms the backbone of various strategies. Thereafter a few algorithms are suggested from literature and the order of model treatment corresponds to that presented in section 3.2.

The following algorithm is a general algorithm taken from (Campolieti and Makarov, 2014) and is based on the following expected value:

$$E_{\mathbb{Q}}\left[V_{t_{k+1}}\left(S(t_k+1)\right)\middle|S(t_k)\right] = \int_0^{\infty} V_{t_k}\left(S(t_k+1)\right)p\left(t_k, t_{k+1}; S(t_k); S(t_{k+1})\right)dS(t_{k+1}). \quad (3.84)$$

here $p(t, T; x, y)$ denotes the transition pdf or the density function of a process Y at time T given the value of the process now $X = x$ at time t. This is usually recovered using equation (3.5) and can be done for all the Fourier-Pricing models, although the method is not very computationally efficient.

Algorithm 3.3. The Stochastic Mesh Method for the estimation of Bermudan option prices by Campolieti & Makarov (Altered and Adapted by the Author).

1.) Input a vector of times, $\boldsymbol{T} = \{t_j : j = 1,2, \dots, M\}$ with $t_0 \leq t_1 < t_2 < \cdots < t_M = T$ and obtain a vector of time differences **diffT** $= \{t_k - t_{k-1}; k = 1,2, \dots M\}$.
2.) Generate n independent paths, $S_0^j, S_1^j, \dots, S_M^j$, using the transition PDF, i.e. each S_k^j is sampled from the density function $p(t_k - t_{k-1}; S_{k-1}^j, S_k^j)$, which is to be numerically recovered, with $j = 1,2, \dots n$ and fix the starting value at $S_0 = S_0^j \; \forall \, j$. The sample paths form a stochastic tree with $n * M$ nodes with the $(k, j)^{th}$ node corresponding to an exercise time t_k and asset price S_k^j.
3.) At the terminal nodes, (M, j), set

$$\hat{V}_M^j = \Lambda_M(S_M^j)\ for\ j = 1,2,\dots,n.$$

In other words calculate the payoff under the option pricing rule at time T for each of the simulated prices at that time.

4.) Work backwards in time and set

$$\hat{V}_k^j = \max\left\{\Lambda_k(S_k^i), \frac{1}{n}\sum_{j=1}^{n} W_k^{ij}\,\hat{V}_{k+1}^j\right\}$$

for $k = M-1, M-2, \dots, 1$ and $i = 1,2,\dots,n$. The weights W_k^{ij} are given by the equation

$$W_k^{ij} = \frac{p(t_{k+1}-t_k; S_{k-1}^j, S_k^j)}{p(t_{k+1}; S_0, S_{k+1}^j)}.$$

5. Finally, calculate the approximate (initial) value of the option by

$$\hat{V}_0 = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_1^i.$$

As can be seen from the algorithm, and in general for Bermudan and American options, the option value at time t_{k-1} is the maximum of the immediate payoff $\Lambda_{t_{k-1}}(x)$, given the asset price, and the continuation value. The latter has the general formula

$$V^{Cont}(t_{k-1}, x) = e^{-r(t_{k+1}-t_k)} \int_{\mathbb{R}} V(t_k, y) f(y|x)\,dy. \tag{3.85}$$

with the current log-price being given by x and the value of the log price at time t_k is y with the payoff function given in terms of the log-price. [Note that it does not matter whether we work with the log-price or the usual asset price, as in Algorithm 3.1, what is important is being consistent mathematically (e.g. specifying the correct payoff function in terms of the log-price or the asset price and calculating the correct density function) as either approach gives the same result.]

Clearly then, the option value at time zero can be found by working backwards, comparing the immediate payoff with the continuation value, until the option value can be computed at time zero. It should be self-evident that an American option can be approximated by computing the value of a Bermudan option with many expiration dates. Furthermore, it is possible to obtain accurate values in this manner by using Richardson Extrapolation* (see (Chang et al, 2007)).

Note*: Richardson Extrapolation is given in the Algorithms3.4 below. Hence the pricing formulae below apply to pricing Bermudan options while using this technique to derive the corresponding American option value from the price of the Bermudan option. We use the method suggested by (Fang and Oosterlee, 2009), altered slightly.

In the Black Scholes case no option formula (semi-analytical) can be given for Bermudan and American options in general (with the exception of an American call option, with no dividend payments and non-negative interest rates) and has to be solved using partial differential equations and inequalities or using some numerical algorithm.

In the interest of space the complete derivation of the Duhamel Solutions to the nonhomogeneous BS PDE in solving American Put and Call options is not given here. Thus a summary is given:

The value of an American Call/Put options can be written as

$$V^A(\tau, S) = V^E(\tau, S) + V^e(\tau, S). \tag{3.86}$$

where the value of the American option is decomposed, with time to maturity $\tau = T - t$, into the price of a the corresponding European option ($V^E(\tau, S)$) and an additional early exercise premium $V^e(\tau, S)$. Since we know how to calculate $V^E(\tau, S)$ under the BS Model the calculation in (3.86) requires only that $V^e(\tau, S)$ is computed numerically (Campolieti and Makarov, 2014):

$$V^e_{Call}(\tau, S; K) = \int_0^\tau \left[qSe^{-q\tau'}\mathcal{N}\left(d^*_+(\tau')\right) - rKe^{-r\tau'}\mathcal{N}\left(d^*_-(\tau')\right)\right]d\tau'. \tag{3.87}$$

$$V^e_{Put}(\tau, S; K) = \int_0^\tau \left[rKe^{-r\tau'}\mathcal{N}\left(-d^*_-(\tau')\right) - qSe^{-q\tau'}\mathcal{N}\left(-d^*_+(\tau')\right)\right]d\tau'. \tag{3.88}$$

Here the early exercise premium is evaluated with

$$d^*_\pm(\tau') = \frac{\ln\left(\frac{S}{S^*(\tau - \tau')}\right) + \left(r - q \pm \frac{1}{2}\sigma^2\right)\tau'}{\sigma\sqrt{\tau'}}. \tag{3.89}$$

and the Call or Put option is evaluated using

$$d_\pm(\tau) = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q \pm \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}. \tag{3.90}$$

Clearly the only requirement is to using these equations for obtaining the BS price is obtaining the early-exercise boundary. The procedure is that suggested by (Campolieti and Makarov, 2014) and involves solving what is known as "*nonlinear Volterra integral equations*."

A couple of suggestions from literature now follow. Note that most payoff functions in practice encompass either a Call or Put option payoff and will be the focus here, but where possible the solution for a more general (non-path dependent) payoff function will be given.

3.4.2.1 **Bermudan and American option strategies for models with constant interest rates**

In a follow up article (Fang and Oosterlee, 2009) show how to apply the COS method to early-exercise options and Barrier options. The computational complexity is $O((M-1)N * \log_2 N)$ with N being the number of terms of the series expansion in (3.67) and M the number of early exercise dates. This method applies to the 4/2 and the Double Heston Model.

The strategy is conceptually not too complicated, but becomes intricate when considering its implementation. Define the following state variables:

$$x = \log\left(\frac{S(t_{m-1})}{K}\right);\ \ y = \log\left(\frac{S(t_m)}{K}\right).$$

and denote by $v(x,t), c(x,t)$ and $g(x,t)$ the option value, the continuation value and the value of the payoff evaluated at time t. In terms of the notation used by (Fang and Oosterlee, 2009), we then have that:

$$c(x, t_m) = \mathrm{e}^{-r(t_{m+1}-t_m)}\int_{-\infty}^{\infty} v(y, t_{m+1}) f(y|x)dy. \tag{3.91}$$

$$v(x, t_m) = \max\left(g(x, t_m), c(x, t_m)\right). \tag{3.92}$$

The time coefficients are indexed with an m instead of k to avoid confusion with the Fourier-cosine coefficients. Now using the methodology under the COS method (3.91) can be rewritten as:

$$\hat{c}(x, t_{m-1}) = e^{-r(t_m - t_{m-1})}\sum_{k=0}^{*N-1}\mathcal{R}\left(\phi\left(\frac{k\pi}{b-a}; x\right)e^{-ik\pi\frac{a}{b-a}}\right)V_k(t_m). \tag{3.93}$$

where the notation $\hat{c}(x, t_{m-1})$ denotes the approximation to $c(x, t_{m-1})$. The option value at time 0 can finally be written as

$$\hat{v}(x,t_0) = e^{-r(t_1-t_0)} \sum_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a};x\right)e^{-ik\pi\frac{a}{b-a}}\right)V_k(t_1). \tag{3.94}$$

Comparing the formulas (3.93) and (3.94) it is clear that the problem comes into computing the V_k coefficients. Recall that these (Fourier-cosine) coefficients are given generally by formula (3.68), however in this context the formula is slightly altered:

$$V_k(t_m) = \frac{2}{b-a}\int_a^b v(y,t_m)\cos\left(k\pi\frac{y-a}{b-a}\right)dy. \tag{3.95}$$

For some of the models in table 3.1 and for some arbitrary payoff function our analysis would be cut short here. Luckily in the case of Bermudan Call and Put options and for Exponential Lêvy models such as the 4/2 and Double Heston models (3.94) and (3.95) can be further analysed. The formulas in (3.94) and (3.95) can be re-expressed, using a simple property of Exponential Lêvy models, as

$$\hat{c}(x,t_{m-1}) = e^{-r(t_1-t_0)} \sum_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a};0\right)e^{-ik\pi\frac{x-a}{b-a}}\right)V_k(t_m). \tag{3.96}$$

$$\hat{v}(x,t_0) = e^{-r(t_1-t_0)} \sum_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a};0\right)e^{-ik\pi\frac{x-a}{b-a}}\right)V_k(t_1). \tag{3.97}$$

Following the work of (Fang and Oosterlee, 2009), the goal here is to develop a backwards induction formula for $V_k(t_m)$ until the cosine coefficients $V_k(t_1)$ are computed, after which these are replaced in formula (3.97) to price the option. The formula in (3.95) can, for Call and Put options, be split into two parts, provided we have calculated the early exercise boundary point, $x^*(t_m)$, such that $c(x^*(t_m),t_m) = g(x^*(t_m),t_m)$. The calculation of the early exercise boundary point is given in Algorithm 3.2 below. Consequently the $V_k(t_m)$ cosine coefficients are, after being split into two intervals, $[a,x^*(t_m)]$ and $(x^*(t_m),b]$, defined as

$$V_k(t_m) = \begin{cases} C_k(a,x^*(t_m),t_m) + G_k(x^*(t_m),b), & Call\ option\ payoff \\ C_k(x^*(t_m),b,t_m) + G_k\big(a,x^*(t_m)\big), & Put\ option\ payoff \end{cases} \tag{3.98}$$

for $m = M-1, M-2, \dots, 1$ and at the terminal payoff time, t_M, (3.98) is

$$V_k(t_M) = \begin{cases} G_k(0,b), & Call\ option\ payoff \\ G_k(a,0), & Put\ option\ payoff \end{cases} \tag{3.99}$$

The functions C_k and G_k are

$$C_k(x_1,x_2,t_m) = \frac{2}{b-a}\int_{x_1}^{x_2} c(x,t_m)\cos\left(k\pi\frac{x-a}{b-a}\right)dx. \tag{3.100}$$

$$G_k(x_1,x_2) = \frac{2}{b-a}\int_{x_1}^{x_2} g(x,t_m)\cos\left(k\pi\frac{x-a}{b-a}\right)dx. \tag{3.101}$$

Recall from a previous discussion on the COS method that the coefficients in (3.101) can be calculated for Call and Put options and that the expressions are exactly as in (3.69)-(3.73) with the c & d's in (3.72)-(3.73) replaced by $x_1 = c, x_2 = d$. The inductive steps can now be given below.

At time t_M, the coefficients $V_j(t_M), j = 0,1,\dots,N-1; m = 1,2,\dots,M$ are known exactly as given by (3.95). Next, proceed to time t_{M-1}, use the formula in (3.96) to calculate the approximated continuation value $\hat{c}(x,t_{M-1})$ and insert this into (3.100). It can be shown that by inserting (3.96) into (3.100) we are allowed to swap integration and summation (the terms are positive and the integrals are finite) which leads to the approximated value

$$\hat{C}_k(x_1,x_2,t_{M-1}) = e^{-r(t_M-t_{M-1})}\mathcal{R}\left(\sum_{j=0}^{*N-1}\phi\left(\frac{j\pi}{b-a};0\right)V_j(t_M)\cdot\mathcal{M}_{k,j}(x_1,x_2)\right). \tag{3.102}$$

with, remembering $i = \sqrt{-1}$,

$$\mathcal{M}_{k,j}(x_1,x_2) = \frac{2}{b-a}\int_{x_1}^{x_2} e^{i*j\pi\frac{x-a}{b-a}}\cos\left(k\pi\frac{x-a}{b-a}\right)dx. \tag{3.103}$$

Thus working backwards in time leads to the general formula, for any time $t_m, m = M-2, M-$

3, … ,1:

$$\hat{C}_k(x_1, x_2, t_m) = e^{-r(t_{m+1}-t_m)}\mathcal{R}\left(\sum_{j=0}^{*N-1} \phi\left(\frac{j\pi}{b-a}; 0\right) V_j(t_{m+1}) \cdot \mathcal{M}_{k,j}(x_1, x_2)\right). \quad (3.104)$$

Since most of the analysis done here is in MATLAB®, which is specially designed to work with matrices, it is useful to rewrite (3.98)-(3.104) in matrix form:

$$\widehat{\boldsymbol{V}}(t_m) = \begin{cases} \widehat{\boldsymbol{C}}(a, x^*(t_m), t_m) + \boldsymbol{G}(x^*(t_m), b), & Call\ option\ payoff \\ \widehat{\boldsymbol{C}}(x^*(t_m), b, t_m) + \boldsymbol{G}\big(a, x^*(t_m)\big), & Put\ option\ payoff \end{cases} \quad (3.105)$$

$$\widehat{\boldsymbol{C}}(x_1, x_2, t_m) = \begin{cases} e^{-r(t_{m+1}-t_m)}\mathcal{R}(\mathcal{M}(x_1, x_2)\boldsymbol{\Lambda})\boldsymbol{V}(t_M), & m = M-1 \\ e^{-r(t_{m+1}-t_m)}\mathcal{R}(\mathcal{M}(x_1, x_2)\boldsymbol{\Lambda})\widehat{\boldsymbol{V}}(t_{m+1}), & m = 1,2, \dots, M-2 \end{cases} \quad (3.106)$$

Here the payoff vector in (3.106) is $\boldsymbol{V}(t_M) = \big(V_0(t_M), V_1(t_M), \dots, V_{N-1}(t_M)\big)^T$. The matrices in (3.106) are $\mathcal{M} = \{\mathcal{M}_{k,j}\}_{k,j=0}^{N-1}$ and the diagonal matrix $\boldsymbol{\Lambda} = \left\{\phi\left(\frac{j\pi}{b-a}; 0\right)\right\}_{j=0}^{N-1}$. On close inspection one can see that in order to compute (3.105) operations of $O(N^2)$ are required to be performed. This is computationally expensive and luckily (Fang and Oosterlee, 2009) show how to reduce this number, based on the FFT algorithm, to $O\big(Nlog_2(N)\big)$.

To begin, the exponential function in (3.103) can be replaced by the well-known identity $e^{i\theta} = \cos(\theta) + i * \sin(\theta)$ which essentially allows us to break up the corresponding matrix $\mathcal{M}$ into two parts:

$$\mathcal{M}_{k,j}(x_1, x_2) = -\frac{i}{\pi}\Big(\mathcal{M}_{i,j}^{c}(x_1, x_2) + \mathcal{M}_{i,j}^{s}(x_1, x_2)\Big). \quad (3.107)$$

Each of the matrices in (3.107) are

$$\mathcal{M}_{i,j}^{c}(x_1, x_2) = \begin{cases} \dfrac{(x_2 - x_1)\pi i}{(b-a)}, & k = j = 0 \\ \dfrac{\exp\left(i(j+k)\dfrac{(x_2-a)\pi}{(b-a)}\right) - \exp\left(i(j+k)\dfrac{(x_1-a)\pi}{(b-a)}\right)}{j+k}, & otherwise \end{cases} \quad (3.108)$$

and

$$\mathcal{M}_{i,j}^{s}(x_1, x_2) = \begin{cases} \dfrac{(x_2 - x_1)\pi i}{(b-a)}, & k = j = 0 \\ \dfrac{\exp\left(i(j-k)\dfrac{(x_2-a)\pi}{(b-a)}\right) - \exp\left(i(j-k)\dfrac{(x_1-a)\pi}{(b-a)}\right)}{j-k}, & otherwise \end{cases} \quad (3.109)$$

Hence by inserting (3.107) into (3.102) and (3.014) and writing the result in matrix form gives

$$\widehat{\boldsymbol{C}}(x_1, x_2, t_m) = \frac{e^{-r(t_{m+1}-t_m)}}{\pi}\mathrm{Im}\{(\mathcal{M}_c + \mathcal{M}_s)\boldsymbol{u}\}. \quad (3.110)$$

where the matrices and vector in (3.110) is

$$u = \{u_j\}_{j=0}^{N-1}, \qquad u_j = \phi\left(\frac{j\pi}{b-a}; 0\right) V_j(t_{m+1}), \qquad u_0 = \phi(0; 0)V_0(t_{m+1}) \quad (3.111)$$

$$\mathcal{M}_c = \{\mathcal{M}_{i,j}^{c}(x_1, x_2)\}_{k,j=0}^{N-1}, \qquad \mathcal{M}_s = \{\mathcal{M}_{i,j}^{s}(x_1, x_2)\}_{k,j=0}^{N-1}. \quad (3.112)$$

The matrices in (3.112) form a Hankel $(\mathcal{M}_c)$,i.e. ascending diagonals from left to right are constant (same number), and Toeplitz $(\mathcal{M}_s)$, i.e. descending diagonals from left to right are constant, matrices respectively:

$$\mathcal{M}_c = \begin{bmatrix} m_0 & m_1 & m_2 & \cdots & m_{N-1} \\ m_1 & m_2 & \cdots & \cdots & m_N \\ m_2 & & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{N-1} & \cdots & \cdots & \cdots & m_{2N-2} \end{bmatrix}_{N\times N} \tag{3.113}$$

$$\mathcal{M}_s = \begin{bmatrix} m_0 & m_1 & \cdots & m_{N-2} & m_{N-1} \\ m_{-1} & m_0 & m_1 & \cdots & m_{N-2} \\ \vdots & & \ddots & & \vdots \\ m_{2-N} & \cdots & m_{-1} & m_0 & m_1 \\ m_{1-N} & m_{2-N} & \cdots & m_{-1} & m_0 \end{bmatrix}_{N\times N} \tag{3.114}$$

The entries in (3.113) and (3.114) are

$$m_j = \begin{cases} \dfrac{(x_2 - x_1)\pi i}{(b-a)}, & j = 0 \\ \dfrac{\exp\left(ij\,\dfrac{(x_2-a)\pi}{(b-a)}\right) - \exp\left(ij\,\dfrac{(x_1-a)\pi}{(b-a)}\right)}{j}, & j \neq 0 \end{cases} \tag{3.115}$$

Before continuing, recall from that if $\boldsymbol{a}$ & $\boldsymbol{b}$ are vectors then the convolution, denoted $\boldsymbol{a} \circledast \boldsymbol{b}$, is defined as

$$\boldsymbol{a} \circledast \boldsymbol{b} = \mathcal{D}^{-1}\{\mathcal{D}(\boldsymbol{a}) \cdot \mathcal{D}(\boldsymbol{b})\}.$$

where $\mathcal{D}$ denotes the discrete Fourier transform and $\mathcal{D}^{-1}$ the inverse discrete Fourier transform. Then in order to compute (3.110) defined the following vectors

$$\boldsymbol{m}_s = [m_0, m_{-1}, m_{-2}, \dots, m_{1-N}, 0, m_{N-1}, \dots, m_1]^T. \tag{3.116}$$
$$\boldsymbol{m}_c = [m_{2N-1}, m_{2N-2}, \dots, m_1, m_0]^T. \tag{3.117}$$
$$\boldsymbol{u}_s = [u_0, u_1, \dots, u_{N-1}, 0, \dots, 0]^T. \tag{3.118}$$
$$\boldsymbol{u}_c = [0, \dots 0, u_0, u_1, \dots, u_{N-1}]^T. \tag{3.119}$$

Then $\mathcal{M}_s\boldsymbol{u} = \boldsymbol{m}_s \circledast \boldsymbol{u}_s$ and $\mathcal{M}_c\boldsymbol{u} = reverse\{\boldsymbol{m}_c \circledast \boldsymbol{u}_c\}$, where the *reverse* operator simply takes the elements of a vector and reverses their order so that the first element is last, the last element is first, the second element is second last, etc.

Lastly, the algorithm to calculate the corresponding American and Bermudan option can finally be formulated below. This algorithm is an amalgamation of the shorter algorithms given in (Fang & Oosterlee, 2009), additionally it corporates Richardson Extrapolation and an alternative to using Newton's method to calculating the early exercise boundary. The alternative to Newton's method is Halley's method which has a cubic rate of convergence instead of a quadratic rate and is the Author's own recommendation. Details of the method is not important and the formula is stated only.

Algorithm 3.4. Efficient calculation of Bermudan and American Call and Put option price using the COS method with a modified Richardson Extrapolation technique.

1.) For all the values of $k = 0,1, \dots, N-1$, compute the coefficients $V_k(t_M)$, as given by (3.99), using equation (3.101). For Call options (3.101) takes the form $G_k(0, b)$ and for Put options it takes the form $G_k(a, 0)$.
2.) Work backwards from $m = M - 1\ to\ 1$ to calculate the approximated value $\hat{V}_k(t_m)$:
2.1) Calculate the early exercise boundary point x_m^* by using (3.106): Let $f(x) = \hat{c}(x, t_{m+1}) - g(x, t_{m+1})$, remembering that $g(x, t_{m+1})$ is simply the payoff function, and apply Halley's Method, for some initial guess x_0:

$$x_{n+1} = x_n - \frac{2f(x_n)f'(x_n)}{2[f'(x_n)]^2 - f(x_n)f''(x_n)}, \qquad x_0 \in [a, b]$$

Note that this process can be specified for a limited number of iterations and some specified error

tolerance ϵ. If the method gives an x_m^* value outside $[a, b]$ then it is set to the nearest boundary point.
2.2) For $j = 0,1, \dots, N-1$, calculate m_j's using the formula in (3.115).
2.3) Construct the vectors $\boldsymbol{m}_s$ & $\boldsymbol{m}_c$ in (3.116) and (3.117).
2.4) Calculate u_j's using formula (3.111).
2.5) Construct the vectors $\boldsymbol{u}_s$ & $\boldsymbol{u}_c$ in (3.108) and (3.109).
2.6) Calculate $\mathcal{M}_s\boldsymbol{u}$, which is the first N elements of $\mathcal{D}^{-1}\{\mathcal{D}(\boldsymbol{m}_s) \cdot \mathcal{D}(\boldsymbol{u}_s)\}$.
2.7) Calculate $\mathcal{M}_c\mathbf{u}$, which is the first N elements of $reverse\{\mathcal{D}^{-1}\{\mathcal{D}(\boldsymbol{m}_c) \cdot \mathcal{D}(\boldsymbol{u}_c)\}\}$.
2.7.1) Calculate $\mathcal{D}(\boldsymbol{u}_c) = \boldsymbol{sgn} \cdot \mathcal{D}(\boldsymbol{u}_s)$, where $\boldsymbol{sgn} = [1, -1, 1, -1, \dots]^T$. This uses the value $\mathcal{D}(\boldsymbol{u}_s)$ calculated in step (2.6) to save computational time.
2.8) Calculate the continuation value $\widehat{\boldsymbol{C}}(x_1, x_2, t_m) = \frac{e^{-r(t_{m+1}-t_m)}}{\pi} \mathrm{Im}\{(\mathcal{M}_c + \mathcal{M}_s)\boldsymbol{u}\}$
2.9) Calculate $V_k(t_m)$ using the value in step 2.8 and the formula (3.99).
3.) Recover the option value at time t_0 as $\hat{v}(x, t_0)$ via using formula (3.97).
4.) Richardson extrapolation: The option value calculated in step 3 has M exercise dates and is denoted by $\mathcal{v}(M)$. Let $Q = \lfloor \log_2 M \rfloor$, where Q denotes the floor function. The approximated American option value via 4 Bermudan options, 3 of which is to be calculated by steps 1-3 above, is then given by

$$\mathcal{v}_{AM}(Q) = \frac{1}{21}\left(64\mathcal{v}(2^{Q+3}) - 56\mathcal{v}(2^{Q+2}) + 14\mathcal{v}(2^{Q+1}) - \mathcal{v}(M)\right).$$

The exercise dates for each $\mathcal{v}$ should be spread evenly (or reasonably evenly) across the time interval. If $\mathcal{v}(M)$ has largely uneven exercise dates spread across time then proceed to step 5.
5.) If the exercise dates of $\mathcal{v}(M)$ are not equally spread then the American option is still calculated using the formula for $\mathcal{v}_{AM}(Q)$, replacing $\mathcal{v}(M)$ with $\mathcal{v}(2^Q)$ and choosing Q arbitrarily and spreading the exercise dates evenly across time.

Some remarks regarding Algorithm 3.2:

It does not matter with what form of the characteristic function we work with. Although all our models are specified in terms of the log-price characteristic function, it is a trivial manner to change the above to incorporate this instead of the logarithm of the price to the strike price.

In step 2.2 the m_j's can be calculated efficiently by noting that $m_{-j} = -\overline{m_j}$, where the bar denotes the complex conjugate of the number, and

$$m_{j+N} = \frac{\exp\left(iN\frac{(x_2-a)\pi}{b-a}\right) \cdot \exp\left(ij\frac{(x_2-a)\pi}{b-a}\right) - \exp\left(iN\frac{(x_1-a)\pi}{b-a}\right) \cdot \exp\left(ij\frac{(x_1-a)\pi}{b-a}\right)}{j+N}$$

Clearly then in step 2.3 the quantities $\exp\left(ij\frac{(x_2-a)\pi}{b-a}\right)$ & $\exp\left(ij\frac{(x_1-a)\pi}{b-a}\right)$ need only be calculated once. Secondly, the Algorithm 3.2 works for any exponential Lêvy Model and hence can be used for a wider set of models than considered here. Lastly, the Richardson Extrapolation has been adjusted from the original since the Author believes that the contribution from the last term is relatively small and that for small number of M, which is often the case in practice, the adjustment has minimal effect of the resulting price. Hence the practitioner can use the value of the calculated Bermudan Contract to estimate the corresponding American price. Otherwise the last term can be replaced by $\mathcal{v}(2^Q)$ and Q can be chosen independently of M. This allows for some flexibility and incorporates the discretion of the practitioner.

For non-Exponential Lêvy models, we cannot use the above algorithm or the analysis from (3.95) onwards. Additionally the above algorithm is restricted in that only non-path dependent payoff types that is allowed. Hence a more general algorithm for all three of our stochastic volatility models is

given below, based on the COS method and mirroring Algorithm 3.3.

Algorithm 3.5. Calculation of Bermudan and American option prices using the COS method for constant interest rate models with characteristic functions including a modified Richardson Extrapolation technique.

1.) For all the values of $k = 0,1, \dots, N-1$ compute the coefficients $V_k(t_M)$ using equation (3.95).
2.) Work backwards from $m = M-1 \; to \; 1$ to calculate the approximate value $\hat{V}_k(t_m)$:
2.1) Calculate the continuation value $\hat{c}(x, t_m)$ using (3.93) by inserting the coefficients $\hat{V}_k(t_{m+1})$.
2.2) Calculate the immediate payoff at time t_m and hence the option value at t_m:

$$v(x, t_m) = \max\big(g(x, t_m), c(x, t_m)\big)$$

2.3) Calculate $\hat{V}_k(t_m)$ from using the value in step 2.2 and equation (3.95).
3.) Use the coefficients $\hat{V}_k(t_1)$ and calculate the option value at time t_0 using (3.94).
4.) Richardson extrapolation: The option value calculated in step 3 has M exercise dates and is denoted by $\mathcal{v}(M)$. Let $Q = \lfloor \log_2 M \rfloor$, where Q denotes the floor function. The approximated American option value via 4 Bermudan options, 3 of which is to be calculated by steps 1-3 above, is then given by

$$\mathcal{v}_{AM}(Q) = \frac{1}{21}\big(64\mathcal{v}(2^{Q+3}) - 56\mathcal{v}(2^{Q+2}) + 14\mathcal{v}(2^{Q+1}) - \mathcal{v}(M)\big).$$

The exercise dates for each $\mathcal{v}$ should be spread evenly (or reasonably evenly) across the time interval. If $\mathcal{v}(M)$ has largely uneven exercise dates spread across time then proceed to step 5.
5.) If the exercise dates of $\mathcal{v}(M)$ are not equally spread then the American option is still calculated using the formula for $\mathcal{v}_{AM}(Q)$, replacing $\mathcal{v}(M)$ with $\mathcal{v}(2^Q)$ and choosing Q arbitrarily and spreading the exercise dates evenly across time.

3.4.2.2 **Bermudan and American option strategy for models with stochastic interest rates**

The COS method wasn't invented to deal with the stochastic interest rate models, namely the H2-CIR Model, however the Author adopts the expanded COS method as given in (Grzelak, Oosterlee and Van Weeren, 2011) in a trivial manner to allow for stochastic interest rates. Consequently the material below is similar to that of the previous subsection and the same notational conventions apply.

Recall that under the stochastic interest rate approach the density function is recovered through the discounted characteristic function:

$$\phi(u) \coloneqq \phi(u, \boldsymbol{X}(t), t, T) = \int\int_{\mathbb{R}} e^{z+iuy} f_{Y,Z}(y, z|x)\,dz\,dy = \int_{\mathbb{R}} e^{iuy}\, \widehat{f_Y}(y|x)\,dy. \tag{3.120}$$

A more general formula for any European contingent claim is then given by

$$v(t, T; x, y) = \int_{\mathbb{R}} v(T, y)\widehat{f_Y}(y|x)\,dy. \tag{3.121}$$

For a Bermudan option the continuation value of each time period is effectively a Call option that is calculated during the time period $(t_{m+1} - t_m)$ and discounted to time t_m. This allows us to use (3.121) and (3.91) and define the continuation and option value at time t_m:

$$c(x, t_m) = \int_{-\infty}^{\infty} v(y, t_{m+1})\widehat{f_Y}(y|x)\,dy. \tag{3.122}$$

$$v(x, t_m) = \max\big(g(x, t_m), c(x, t_m)\big). \tag{3.123}$$

Using the methodology under the COS method, $\widehat{f_Y}(y|x)$ in (3.122) can be rewritten as:

$$\widehat{f_Y}(y|x) \approx \sum_{k=0}^{*N} \theta_k \cos\left(k\pi \frac{y-a}{b-a}\right). \tag{3.124}$$

here θ_k is given as

$$\theta_k = \frac{2}{b-a} Re\left(\phi\left(\frac{k\pi}{b-a}, x\right) \exp\left(-k\pi \frac{ia}{b-a}\right)\right). \tag{3.125}$$

Clearly Algorithm 3.5 can be adopted, with possibly one extra step, to price Bermudan and American options under this approach.

We know turn to Asian options.

3.4.3. **Asian options**

Asian options are an example of path-dependent exotic options. There are numerous ways of specifying the payoff of the option although for our purposes it is sufficient to address the payoff structure that has always led to more difficulties: Arithmetic Asian options. These are defined by the payoff functions with fixed strike:

$$\max\left(\frac{1}{n}\sum_{i=1}^{n} S_{t_i} - K, 0\right). \tag{3.126}$$

The option is assumed to be discretely monitored at some set of dates $0 = t_0 < t_1 < \cdots < t_n = T$ and the option can only be exercised at time T. Note that whether the process admits a fixed or floating strike is relatively trivial when using Monte Carlo simulations. For our purposes a fixed strike is assumed. Additionally, all Asian options are assumed to be "fresh" options, i.e. discrete monitoring starts at time 0 and not from some future time t: $0 < t < T$.

Now, under the BS Model no known closed formula for Arithmetic Asian options exists. Indeed, the literature regarding the models we which consider here is sparse on this topic and it is easier for our purposes to consider a Monte Carlo approach. The interested reader can however look at the article by (Forde and Jacquier, 2010) to see how to price Asian options (under a different payoff formulation than that considered here) under the Heston Model.

3.4.3.1 **An algorithm to calculate Asian options**

The Monte Carlo Strategy is rather simple. We simulate sample paths using the Euler scheme or the Brownian Motion Reconstruction scheme and consequently sample paths can be generated at each of the discrete monitoring times and hence the option value can be obtained at time 0. There are many improvements to the method that is not addressed here and the reader inspecting the supporting MATLAB® scripts can probably implement a variety of techniques to reduce the number of simulations, run the code more efficiently, implementing variance reduction techniques, etc.

The matter on whether interest rates are stochastic or deterministic should be trivial by now. The use of stochastic interest rates is presented in Algorithm 3.6 below and for all other models this is taken to be constant.

Thus the following algorithm is presented:

Algorithm 3.6 Monte Carlo strategy to pricing arithmetic Asian options using the characteristic function of the log-price.

1.) Sample any sufficient number of sample paths using the transition densities, where each simulated y value at time t_k forms the x value to simulate a value y at time t_k:
1.1.) For each discrete monitoring time $0 = t_0 < t_1 < \cdots < t_n = T$ compute the n transition densities numerically, i.e. compute $p(t_{k-1}, t_k, x, y)$ for some integer $k \leq n$ from the log-price characteristic functions.
2.) At the terminal date T, compute each of the option payoffs from the sample paths generated.
3.) Discount the values obtained in step 2.) with the appropriate discount factors, or in the case of stochastic interest rates, using the corresponding zero coupon bond values obtained from the characteristic function.
4.) Average the values obtained in step 3.) to obtain a sample option value.

3.4.4. **Barrier options**

Barrier options are path dependent options whose payoff depends on whether on or not the underlying option has hit a specified barrier B or not. Our work here will focus on down-and-out/up-and-out or *knocked out options*, which may be Call or Put options. Here the option payoff is similar to the a Call option payoff, provided the asset hasn't hit a the barrier B before time T. Consequently the value of a down-and-out Call option for instance must be less than the value of the corresponding Call option without the barrier. This discount is known as the *knock-out discount*. The general payoff formula may also include what is called a rebate amount, R, which compensates the buyer of the option if the barrier is hit. In the case of a Call option this payoff function, at time T, is given as:

$$h = \big(\max\big((S(T) - K), 0\big) - R\big)\mathbb{I}_{\{inf_{t \leq T} S(t) > B\}} + R. \tag{3.127}$$

In general our work will involve $R = 0$, although under the COS pricing method we may assume $R \neq 0$ and setting it to zero will not impact on the final result. Note, an investor receiving the rebate amount at any time prior to time T will simply receive $Re^{-r(T-t_B)}$, where t_B denotes the first time (monitored) when the underlying reaches the barrier.

In what follows the majority of the work is focused specifically on down-and-out call options, although any of the methods may be adjusted for the other types by modifying the methods given here.

3.4.4.1 **Barrier option strategies for models with constant interest rates**

The follow-up article by (Fang and Oosterlee, 2009), with which we priced Bermudan options above, also specifies the appropriate method to price Barrier options. The method they present is for up-and-out Call and Put options. Their method is presented here but the Author reformulates this from the perspective of a down-and-out Call or Put option. The difference between up-and-out and down-and-out options, for clarity, is simply whether $B > S_0$ (up-and-out) or $B < S_0$ (down-and out), i.e. where the barrier is set.

Although the options here are discretely monitored, the computational time taken to compute the option value is significantly less than the case for Bermudan options, due in part to the barrier B that is known in advance. Hence we can approximate the value of a continuously monitored barrier option

by increasing the number of times monitored.

In the notation of the COS methodology, we may write the payoff function for a down-and-out Call or Put option as:

$$v(x,T) = (\max(\alpha(S(T)-K),0)-R)\mathbb{I}_{\{inf_{t\leq T}S(t)>B\}} + R. \tag{3.128}$$

Here $\alpha = 1$ corresponds to a Call option and $\alpha = -1$ corresponds to a Put option. The set of observation dates is given by $t_1 < \cdots < t_M = T$, which corresponds to M monitoring dates. The following recursive formula describes the price of the option in (3.128):

$$\begin{cases} c(x,t_{m-1}) = e^{-r(t_m - t_{m-1})}\int_{\mathbb{R}} v(x,t_m)f(y|x)dy \\ v(x,t_{m-1}) = \begin{cases} e^{-r(T-t_{m-1})}R, & x \leq h \\ c(x,t_{m-1}), & x > h \end{cases} \end{cases} \tag{3.129}$$

Here $h = \log\left(\frac{B}{K}\right)$ and $m = M, M-1, \ldots, 2$. Remember that here $x = \log\left(\frac{S_{t_{m-1}}}{K}\right)$, and $y = \log\left(\frac{S_{t_m}}{K}\right)$. The same methodology as used for pricing Bermudan options allows us to formulate a backwards induction approach for calculating the $\hat{V}_k(t_m)$ coefficients, for $m = M-1, M-1, \ldots, 1$:

$$\hat{V}_k(t_m) = \hat{C}_k(x_1, x_2, t_m) + e^{-r(T-t_{m-1})}R\frac{2}{b-a}\psi_k(c,d). \tag{3.130}$$

where $\hat{C}_k(x_1,x_2,t_m)$ & $\psi_k(c,d)$ are given exactly as in (3.108) and (3.72). For $m = M$:
If $h < 0$

$$\hat{V}_k(t_M) = \begin{cases} \frac{2R\psi_k(c,d)}{b-a}, & for\ a\ Call\ option \\ G_k(x_1,x_2) + \frac{2R\psi_k(c,d)}{b-a}, & for\ a\ Put\ option \end{cases} \tag{3.131}$$

If $h \geq 0$

$$\hat{V}_k(t_M) = \begin{cases} G_k(0,x_2) + \frac{2R\psi_k(c,d)}{b-a}, & for\ a\ Call\ option \\ G_k(x_1,0), & for\ a\ Put\ option \end{cases} \tag{3.132}$$

The undetermined numbers in the above equation are given as follow:
For up-and-out options: $x_1 = a, x_2 = h, c = h$ & $d = b$.
For down-and-out options: $x_1 = h, x_2 = b, c = a$ & $d = h$.

Consequently, the algorithm from (Fang and Oosterlee, 2009) is reproduced here, with minor adjustments, as Algorithm 3.7:

Algorithm 3.7 General strategy for calculating the value of a barrier Call or Put option for up-and-out or down-and-out options under a stochastic volatility model admitting a general Exponential Lêvy characteristic function.

1.) Before the inductive step work out the following quantities:
1.1) The coefficient $\hat{V}_k(t_M)$ as given by formulas (3.131) & (3.132).
1.2) Construct the vectors $\boldsymbol{m_s}(x_1,x_2)$ & $\boldsymbol{m_c}(x_1,x_2)$ using the formulas in (3.115)-(3.117).
1.3) Set $d_1 = \mathcal{D}\{\boldsymbol{m_s}(x_1,x_2)\}, d_2 = \boldsymbol{sgn}\cdot\mathcal{D}\{\boldsymbol{m_c}(x_1,x_2)\}$.
1.4) Construct $\boldsymbol{G} = \frac{2}{b-a}R\{\psi_k(c,d)\}_{k=0}^{N-1}$
2.) For $m = M, \ldots, 2$, calculate the $\hat{V}(t_m)$ coefficient recursively:
2.1) Calculate $\boldsymbol{u}_{t_m}$ using the formula (3.111).
2.2) Construct $\boldsymbol{u}_s$ as done in (3.116).
2.3) Calculate $\boldsymbol{Msu}$ which is the first N elements of $\mathcal{D}^{-1}\{d_1\cdot\mathcal{D}(\boldsymbol{u}_s)\}$.

2.4) Calculate $\boldsymbol{Mcu}$ as the first N elements of $reverse\{\mathcal{D}^{-1}\{d_2 \cdot \mathcal{D}(\boldsymbol{u}_s)\}\}$.
2.5) Calculate $\widehat{\boldsymbol{C}}(t_{m-1}) = \frac{e^{-r(t_m - t_{m-1})}}{\pi}\ \mathrm{Im}\{\boldsymbol{Msu} + \boldsymbol{Mcu}\}$
2.6) Lastly calculate $\widehat{\boldsymbol{V}}(t_m) = \widehat{\boldsymbol{C}}(t_{m-1}) + e^{-r(T-t_{m-1})}\boldsymbol{G}$
3.) Finally, compute the option price as $\hat{v}(t_0, x)$ using the formula in (3.97).

Note that this algorithm does not work for the H2CIR Model and a more general algorithm is given below.

3.4.4.2 General Barrier option pricing strategy

There are no analytical shortcuts and thus the Author proposes the following basic (and generic) algorithm which incorporates stochastic or constant interest rates.

Algorithm 3.8 General strategy for calculating the value of a barrier Call or Put option for up-and-out or down-and-out options under a stochastic volatility model admitting a characteristic function with constant or stochastic interest rates.

1.) Sample any sufficient number of sample paths using the transition densities or Monte Carlo simulation, where each simulated y value at time t_k forms the x value to simulate a value y at time t_k:
1.1.) For each discrete monitoring time $0 = t_0 < t_1 < \cdots < t_n = T$ compute the n transition densities numerically, i.e. compute $p(t_{k-1}, t_k, x, y)$ for some integer $k \leq n$ from the log-price characteristic functions OR start the monitoring process at the last simulated Monte Carlo value of each sample path.
2.) At the terminal date T, compute each of the option payoffs from the sample paths generated:
2.1) If the option has hit the barrier, then calculate the discounted value at time zero of the rebate amount, if any.
2.2)If the option has not hit the barrier, and the payoff is nonzero, then calculate the present value of the cashlow.
2.3) Average all option values at time zero and determine the option value.
3.) Discount the values obtained in 2 with the appropriate discount factors, or in the case of stochastic interest rates, using the corresponding zero coupon bond values obtained from the characteristic function.
4.) Average the values obtained in step 3 to obtain a sample option value.

3.4.4.3 Remark on a general path-dependent exotic equity derivatives pricing strategy

Pricing any exotic derivative can be done through Monte Carlo simulation of the entire asset path, as needed. Generally this is the best way to approach unusual options whereby calculations are done on each path or all of them together depending on the derivative contract. The formulae given in the previous sections are meant to accelerate the implementation time and offer viable alternatives to the Monte Carlo method. Ultimately though, where no formulas exists, the Monte Carlo method can be used together with some of the suggestions in section 3.2, such as the Multilevel method, to ensure faster convergence.

This also concludes the section on pricing exotic equity derivatives.

3.5. **Calculating option sensitivity: The Greeks**

The purpose of building a model to price derivatives, apart from determining the price of the contract, is in practice to determine one's exposure to the underlying's movements The practitioner will typically build a model, calibrate it with today's option prices or implied volatility surface, and then from there determine the appropriate Greeks or even use the model to price other exotic options whose prices are not found in the market. The focus of this subsection is to look at the strategies used to calculate the Greeks.

For clarity, a Greek is defined as the change in the option value with respect to the change in a model parameter ϖ, under some payoff function h_{ϖ}:

$$Greek_{\varpi} = \frac{\partial}{\partial \varpi} E\big[h_{\varpi}\big(S(t)\big)\big]. \tag{3.133}$$

Clearly a model parameter to calculate the option price includes the underlying itself. The first derivative of the option with respect to the underlying is commonly called the "Delta" of the option. The second derivative with respect to the underlying is referred to as the "Gamma" of the option. The ordinary time derivative of the option price is called the "Theta" of the option. There are many others that can be defined, however for our purposes it sufficient to consider the Delta and Gamma of the option contract.

The first subsection will look at the classical finite difference schemes used. Finally, there are some shortcuts in literature that will be given for some of the models in table 3.1.

3.5.1 **A complex analysis alternative to finite difference schemes**

Using basic calculus, the derivative of a function $f(x)$ can be approximated by:

$$f'(x) = \frac{f(x+\epsilon) - f(x-\epsilon)}{2\epsilon}. \tag{3.134}$$

Here we assume that ϵ is some sufficiently small number, say $\epsilon = 10^{-16}$. Consequently the Greek in (3.133) can be approximated by

$$Greek_{\varpi} = \frac{E\big[h_{\varpi+\epsilon}\big(S(t)\big)\big] - E\big[h_{\varpi-\epsilon}\big(S(t)\big)\big]}{2\epsilon}. \tag{3.135}$$

By linearity of the expected value operator, the quantity in (3.135) can be computed using Monte Carlo simulations and thus an estimator for (3.135) is (Kietnitz and Wetterau, 2012):

$$\widehat{Greek}_{\varpi} = \frac{1}{N}\sum_{i=1}^{N} \frac{h\left(S^i(\varpi+\epsilon)\right) - h\left(S^i(\varpi-\epsilon)\right)}{2\epsilon}. \tag{3.136}$$

Unfortunately, as demonstrated below, choosing the value of ϵ is not easy, even for simple functions and consequently a more sophisticated method can be found as in (Spuire and Trapp, 1998). Their method is simple. Firstly use a Taylor expansion in the complex valued function:

$$f(x+i\epsilon) = f(x) + i\epsilon f'(x) - \frac{\epsilon^2}{2} f''(x) + \cdots. \tag{3.137}$$

Clearly taking the imaginary parts of (3.137) leads to the following formula:

$$f'(x) = \frac{Im\big(f(x+i\epsilon)\big)}{\epsilon} + O(\epsilon^2). \tag{3.138}$$

This formula is more stable than in (3.134) regarding values of ϵ, however the additional assumption

is that the expansion (3.137) holds. For interest sake, let us compare the schemes in (3.134) and (3.138) using $f(x) = \cos(\sqrt{x}) - \exp(1 - ln(x))$.

Suppose we are interested in finding the derivative value at $x = 2$. The exact value is $f'(2) = -\sin(\sqrt{2}) * \frac{1}{2\sqrt{2}} - \exp(1 - \ln(2)) * \frac{-1}{2} = 0.330342457796457$. Consider now the differences between implementing (3.134) and (3.138) for various values of ϵ to compute this derivative:

Epsilon ϵ	Finite Difference (3.134)	Finite Difference (3.138)
10^{-8}	0.330342453391808	0.330342457796457
10^{-9}	0.330342420085117	0.330342457796457
10^{-10}	0.330342420085117	0.330342457796457
10^{-11}	0.330335758746969	0.330342457796457
10^{-12}	0.330291349825984	0.330342457796457
10^{-13}	0.327515792264421	0.330342457796457
10^{-14}	0.344169137633799	0.330342457796457
10^{-15}	0.333066907387547	0.330342457796457
10^{-16}	0	0.330342457796457
10^{-17}	0	0.330342457796457
10^{-18}	0	0.330342457796457
10^{-19}	0	0.330342457796457
10^{-20}	0	0.330342457796457

Table 3.3: Comparison of numerical differentiation schemes

The table clearly shows that finite difference method in (3.134) yields somewhat stable answers when $\epsilon \in [10^{-8}, 10^{-12}]$ and otherwise the estimates are unreliable even when $\epsilon < 10^{-16}$. The method in (3.138) performs superbly and is stable for all values of ϵ.

One usage of the above method is to calculate the Greeks using Monte Carlo simulation techniques by applying the above formula as the sample paths are all real valued.

3.5.2 **Using the COS method to calculate the Greeks**

In the same article by (Fang and Oosterlee, 2008), that introduced the COS method, there are two formulas given which allow us to compute the Delta and Gamma of an option using the characteristic function. The formulas are stated below and their implementation can be found in the supporting material. The formulas given here:

$$Delta: \Delta \approx e^{-rT} \sum_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a}; x\right) e^{-ik\pi\frac{x-a}{b-a}} \frac{ik\pi}{b-a}\right) \frac{V_k}{S(0)}. \tag{3.139}$$

$$Gamma: \Gamma \approx e^{-rT} \sum_{k=0}^{*N-1} \mathcal{R}\left(\phi\left(\frac{k\pi}{b-a}; x\right) e^{-ik\pi\frac{x-a}{b-a}} \left(-\frac{ik\pi}{b-a} + \left(\frac{ik\pi}{b-a}\right)^2\right)\right) \frac{V_k}{S(0)}. \tag{3.140}$$

These formulas work for all the models in table 3.1. except the H2-CIR Model. For that specific

model it is recommended by the Author to use Monte Carlo simulations to calculate the Greeks.

It should be mentioned that the COS method allows rapid calculation of option prices and computing the Delta and Gamma of an option can be done quickly as well. Here the term “rapid calculation” means within 1 second. The Author has found, combined with the considerations given in Appendix D, that the COS method can calculate 1500 option prices s in under a second (using the Double Heston Model). The same calculations using a Monte Carlo based approach takes considerably longer.

3.6. **Conclusion**

In this chapter we primarily looked at the methods used to price exotic derivatives. Monte Carlo techniques provide a very flexible approach to pricing derivatives, although consideration has to be given to the overall time spent running the simulations and, where possible, use computational techniques to accelerate the simulations. Thereafter we considered the issue of using techniques from literature to price European and more general options, giving strategies and algorithms where needed. Finally, the issue of calculating option sensitivity has been considered briefly. This ends our formal literature review, although some additional references can be found in the Appendices where additional issues, such as model calibration, are investigated

In the next chapter we will compare pricing schemes and stochastic volatility models. Two research hypotheses are investigation and draw upon the materials in this and the previous chapter:

Chapter 4 – Main Results: Investigation and Conclusion into Two Research Questions

4.1. **Research questions**

In this chapter two research questions are investigated. The sections below outline in detail the methodology used to investigate these questions, followed by a discussion of the results.

Up to now we have assumed that all parameters values are known. Clearly this serves well for illustrative purposes, where parameters have been chosen to give workable examples, but in practice these values have to be determined from market data. This presents us with two problems.

First, in literature there is no definitive recommendation on what procedure should be used to calibrate stochastic volatility models. The Author has decided to use Differential Evolution (DE) and presents a short appendix (Appendix F) giving an example of how this procedure works. The justification for using (DE) is given in Appendix F as well as some of the implementation details underlying the results presented here.

Second, market data pertaining to Call or Put options are easy to obtain, but data on exotic options is much harder to obtain, unless one pays substantial fees. Now even with data available, the data has to be cleaned and adjusted to make it suitable for calibration purposes. In light of these problems, the Author has decided to present a Market Model that serves as the basis for generating option prices and to which the other models can be calibrated to.

Using the Market Model to generate simulated data has some distinct advantages over using market data. An exact analysis can be made on various exotic equity derivatives by pricing several different types of options. In this way, the prices obtained from the various models can be compared to an exact price and the exact transition density function across various times can be calculated. The significance of this is given later.

Ultimately if we decided to use option data for a specific index, say the S&P500, then any conclusion would likely only apply to that specific data set. So instead if our focus is wider than merely which model produces correct option prices for a specific index (or asset) then the argument is that using any suitable market data or market model will suffice. So we will perform a simulated study design.

4.1.1 **Market Model**

The Market Model proposed here is more complicated than any of the stochastic volatility models presented in the previous sections. The model is driven by two stochastic volatility factors and two stochastic interest rate factors. In this way the generated prices and sample paths will each represent more features than that which is captured by the stochastic volatility models presented. Additionally, this model does not appear in literature and has no closed form for the characteristic function (to the best of the Author's knowledge). The model is formulated using the following SDE, using the results in section 3.1.3:

$$\frac{dS(t)}{S(t)} = \frac{(r_1(t)+r_2(t))}{2} + \left(a\sqrt{\left(V_1(t)\right)^3} + \frac{b}{\sqrt{V_1(t)}}\right)dW_S^1(t) + c\left(V_2(t)\right)^{\frac{3}{4}}dW_S^2(t)$$

$$dV_1(t) = \left(d_1 - e_1 v_1(t)\right)dt + \sigma_1\sqrt{V_1(t)}dW_{V_1}^3(t)$$

$$dV_2(t) = \left(d_2 - e_2 v_2(t)\right)dt + \sigma_2\sqrt{V_2(t)}dW_{V_2}^4(t)$$

$$dr_1(t) = \lambda_1\left(\theta_1 - r_1(t)\right)dt + \eta_1\sqrt{r_1(t)}dW_{r_1}^5(t)$$

$$dr_2(t) = \lambda_2\left(\theta_2 - r_2(\,t)\right)dt + \eta_2 dW_{r_2}^6(t).$$

The correlation matrix between the Brownian motion processes is given by:

$$\begin{bmatrix} 1 & 0 & \rho_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \rho_2 & 0 & 0 \\ \rho_1 & 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This correlation structure is chosen to simplify the implementation of this model and reduce the number of model parameters.

An exact simulation scheme for this model is not available and hence the Author's method of Brownian Motion reconstruction will be used to simulate from this model. Whether a Euler or the Brownian reconstruction scheme is used ultimately has no impact on the conclusions reached for question two. The justification for this is given in the conclusion to research question one below.

The following parameter set will be used to generate sample paths and option prices. The time horizon will be fixed as 1 year, and a standard of 365 points will be generated per sample path. Here $S_0 = 1000$, the starting price of the underlying asset. Note that when referring to log() we mean the natural logarithm of a number (i.e. with respect to base e).

Parameter	$\log(S_0)$	a	b	c	d_1	d_2	e_1	e_2	σ_1
Value	6.9077	0.1	0.01	0.2	1.5	1	0.5	4	0.1
Parameter	σ_2	λ_1	λ_2	θ_1	θ_2	η_1	η_2	ρ_1	ρ_2
Value	0.15	0.4	0.2	0.03	0.04	0.1	0.05	-0.7	-0.9
Parameter	$V_1(0)$	$V_2(0)$	$r_1(0)$	$r_2(0)$					
Value	0.1	0.1	0.05	0.02					

Table 4.1: Market Model Parameters

It is instructive to see sample paths of this model and the simulated transition density function:

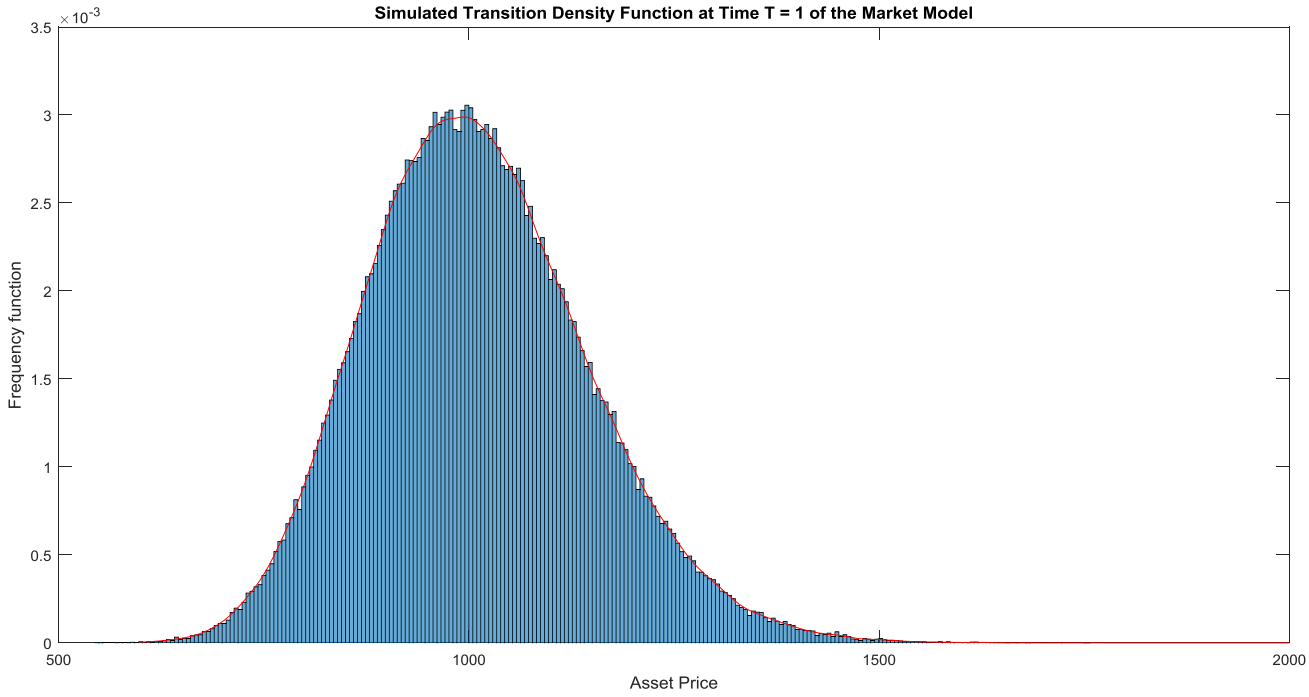

Figure 4.1: Simulated transition density function for the Market Model

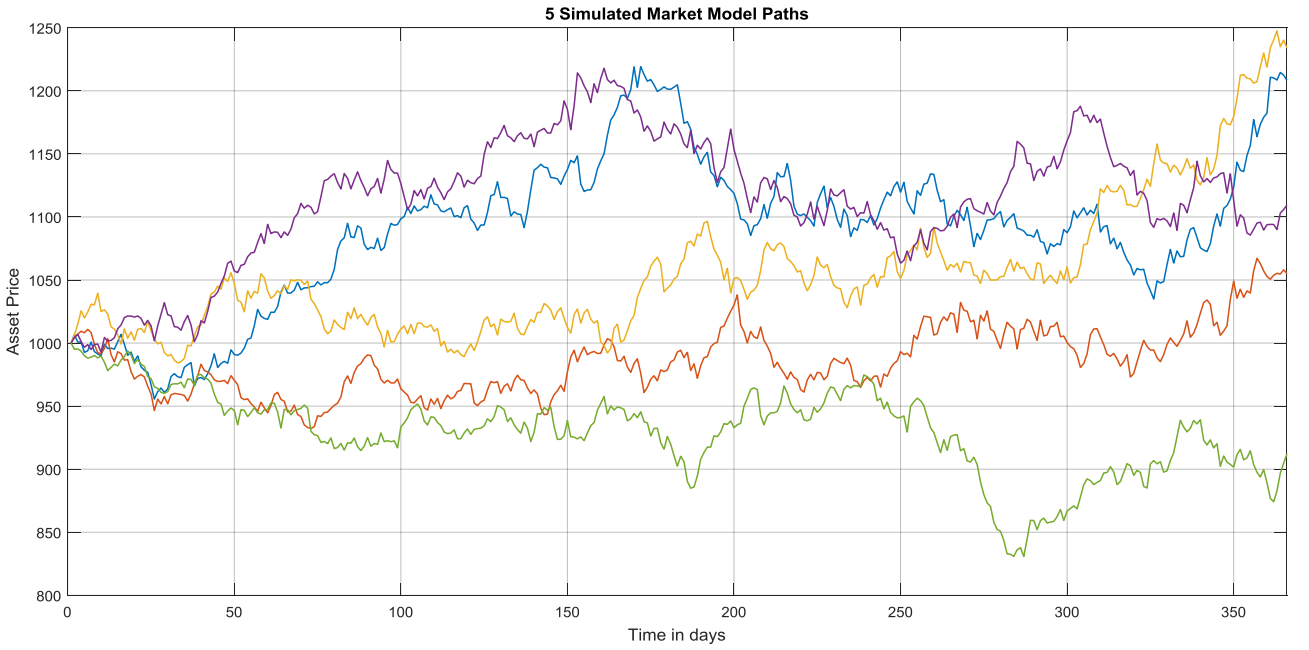

Figure 4.2: Five simulated Market Model paths

A table of sample Call options can be given here, for strike prices $K \in [800,1550]$, which will serve as our market data for calibration. Although the table contains 16 prices, for the actual calibration 151 option prices will be used. These values have been obtained by simulating 200000 sample paths with each sample path having 365 points. The time taken to simulate the 73 000 000 points is 3 minutes and 30 seconds using C++ code for acceleration and two workers via a parallel computing acceleration (see Appendix D).

Strike:	**800**	**850**	**900**	**950**	**1000**	**1050**	**1100**	**1150**
Price:	204.75	159.97	119.19	84.19	56.09	35.17	20.77	11.61
Strike:	**1200**	**1250**	**1300**	**1350**	**1400**	**1450**	**1500**	**1550**
Price:	6.15	3.08	1.48	0.69	0.31	0.134	0.057	0.023

Table 4.2: Table of sample Call option prices for the Market Model

This table has been obtained by simulating 200 000 values and working out the average of all option prices obtained from the simulation. The discount factor applied is simply the average of the zero-coupon bond prices implied by the two stochastic interest rate factors. For the CIR interest rate process formulas for the zero-coupon bond price has been given in Chapter 2. The formula for the HW Model (which in the case of constant parameters reduces to the well-known Vasicek Model) is given in Appendix B.

The Market Model has now been specified and a rough indication has been given on its qualitative behaviour. We now move onto the research questions that are our main focus.

4.1.2. **Research Question 1**

In the Monte Carlo, section 3.2 in the previous chapter, three methods were introduced for simulating sample paths to be used for pricing options: The Euler method, the Author's Brownian Motion reconstruction technique, and simulating paths sequentially using the characteristic function. Recall that the latter mentioned technique was shown to be unreliable for fine path simulation and hence will not be considered here.

The first research question can thus be formulated:

Which of the two Monte Carlo Techniques given works best to price exotic derivatives?

Where "best" is interpreted according to the following criteria:

First Criterion:

The number of runs on average that are needed to reach a specified accuracy in comparison to the true transition density function. The two methods will then be compared on average run times to achieve this accuracy. Clearly a faster average run time to reach a specified accuracy is preferable.

Second Criterion:

We need to investigate the rate at which the given Monte Carlo Techniques converge to the true option price and thus we are interested in the *fastest convergence rate* to the true option price. Measuring the convergence rate is simply comparing the errors with the number of simulations and how successive errors decrease. This is done by a comparison of option prices and transition densities, as outlined in the next section.

In summary, we will look at average run time, and the rate at which the given Monte Carlo Techniques converge to the true option price and draw conclusions based on these quantities to

answer research Question 1.

4.1.3. **Research Question 2: Comparison of select stochastic volatility models**

In Chapter 2 we investigated a select number of stochastic volatility models that are present in literature. In Chapter 3, table 3.1 listed the three of the models in Chapter 2 that we will be using. Recall that these are the Double Heston, H2-CIR and 4/2 models and that the stochastic volatility process of each of these models are driven by (a) CIR process(es). The CIR process was discussed in subsection 2.2.4.2, Chapter 2.

The second research question can thus be stated as follows:

Which of the select stochastic volatility models performs the best relative to the Market Model?

Criteria used to investigate the second research question:

First Criterion:

Select some error measure (such as MSE) and see which stochastic volatility model comes closest to the predicted market prices for a range of different options. Thus we want to see which of the select stochastic volatility models provides exotic option prices that are closest to the given market prices.

Second Criterion:

Which of the select stochastic volatility models exhibits the fastest time to compute the characteristic function for various times to maturity? For all the models the methods used to compute option prices or run simulations are practically the same. Computing the characteristic function quickly is a desirable trait of a model, especially when applying calibration techniques. The issue is thus one of model implementation.

Third Criterion:

The last criterion relates to that of parameter stability. Here the transition density function of the stochastic volatility models will be examined when the time of maturity is long, i.e. 5 years. In this way we will be able to distinguish the long run behaviour of models and compare them to the market model. Parameter stability is then assessed under this criterion. Thus we are concerned with which stochastic volatility model exhibits the most stable parameter values across time.

In summary, for the second research question we are concerned with pricing accuracy, run times to compute the characteristic function and parameter stability.

4.2. **Investigation into the research questions**

The methodology used to answer the two research questions is given below in two separate subsections, one for each research question. All the results can be closely replicated by examining the supporting material as detailed in Appendix A. At the end of each subsection an analysis and conclusion is given. The final conclusions and discussion are then given at the end of the chapter in section 4.3.

4.2.1. **Investigating Research Question 1**

In order to investigate Research Question 1 we need to specify some model in order to work with. The Double Heston Model will serve as our working model due to the fact that it is easy to implement. The comparison here is essentially between using a Euler and an exact simulation scheme for the CIR process driving the volatility of the model.

The parameters used to simulate paths from the model are:

Parameter	$\log(S_0)$	$v_1(0)$	$v_2(0)$	a_1	a_2	b_1	b_2	σ_1	σ_2
Value	6.9077	225%%	220%%	0.04	0.03	1	0.8	0.35	0.3
Parameter	ρ_1	ρ_2	r						
Value	-0.7	-0.7	0.03						

Table 4.3: Double Heston Model parameters used to investigate research Question 1

The comparison between the different schemes depends heavily on how they are implemented in the C++ and called from the MATLAB® environment. The Author has optimised all the schemes to run as fast as possible while utilising most of the computational speed-up techniques in Appendix D where possible. Ultimately, we have to recognise that *different* schemes are being compared here and that includes their implementation.

Starting the investigation, the following table give the run times for the models to simulate a certain number of paths, each having 365 points. The given two schemes are simulated in 10 blocks of simulations running 100, 1000, 10 000, 20 000 simulations per block respectively. These blocks are distributed across two MATLAB® workers in a parallel structure. The best average running times are reported here.

Path simulations	Euler Method (Seconds)	BM Reconstruction Method (Seconds)
100	0.006971	0.062950
1000	0.078413	0.529655
10000	0.824843	4.729090
100000	8.393877	46.971224
200000	17.514301	94.513494

Table 4.4: Running times under different Monte Carlo Schemes, using 365 time steps.

It should be noted that the total number of points simulated is 365 times the number of path simulations and 1000 simulations means that 365 000 points have been simulated.

Consider now the pricing of an American Call option with a strike price of 1100 and a one year time horizon. The current asset price is 1100. The table below gives sample option prices per number of simulations and the relative absolute error involved. Note that the final reference price used is 80.7931

for the Euler scheme and 80.6757 for the BM reconstruction scheme, each obtained after 2 million simulations.

Nr of path simulations	Euler Method Price	Euler Method Error	BM Reconstruction Price	BM Reconstruction Error
1000	80.3925	0.4006	81.1433	0.4676
10000	80.7649	0.0282	80.6865	0.0108
100000	80.6723	0.1208	80.6705	0.0052
500000	80.5834	0.2097	80.3252	0.3505
1000000	81.0415	0.2484	80.7955	0.1198
2000000	**80.7931**	N/A	**80.6757**	N/A

Table 4.5: Errors under each scheme relative to the reference price for 365 time steps

Some commentary: The random fluctuations involved in the errors are evident and no patterns in the errors are discernible if we repeat the exercise a number of times. The final reference prices obtained after 2 million simulations also fluctuate but less than the prices for 10000 simulations. The convergence rate to the true price is indeed slow.

Continuing, since our work involves time homogeneous stochastic volatility models, it follows that the kernel distribution of the asset paths at time $T = 1$ must be the same as the distribution of values obtained irrespective of the subintervals of time used to generate paths. In other words, the transition density function at time $T = 1$, which we obtain from the characteristic function, must be equal to the density function resulting from simulating many paths across various time intervals at the terminal time ($T = 1$).

In general this means that the spread of values at any time point in the interval, say $t^* \in (0,1]$ resulting from the simulation of asset paths must be equal to the transition density function with time to maturity as defined by $\tau = t^* - 0 = t^*$. Hence, correct convergence to the transition density function at any time interval resulting from the spread of simulated paths implies directly that the convergence to the true option price will be reached (after suitably many simulations).

Below the table gives the error involved between the kernel density fitted to the spread of values obtained after N simulations and the true density function. The error is defined as the mean square error here as many points are being compared. Both simulation schemes eventually convergence 100% to the correct transition density function at time $T = 1$. Note that kernel density is fitted with 3000 fine points in each case.

Nr of path simulations	Euler Method Error ($Value * 10^{-10}$)	BM Reconstruction Error ($Value * 10^{-10}$)
10000	3.4662	5.4926
20000	2.6239	2.0324
50000	0.58085	0.92612
100000	0.55384	0.30512
200000	0.33666	0.29031
500000	0.26979	0.092751

Table 4.6: Relative errors of the simulations schemes vs the true transition density function at 365 time steps a year.

Our analysis began by considering the simulation of 365 points per sample path. Decreasing the time step also needs to be investigated. Here the 3 part analysis is repeated using 52 and 12 time steps per year with commentary given at the end only. An additional table is also given with the Euler method having 365 sample points used to price Bermudan options with the monitoring dates corresponding to the time steps.

Essentially by decreasing the time step we are now pricing Bermudan options and not American options. Note below how the option premium decreases, as expected (early-exercise benefit diminishes).

Table of time taken to simulate a given number of sample paths, with 52 points per sample path.

Path simulations	Euler Method (Seconds)	BM Reconstruction Method (Seconds)
100	0.002136	0.052167
1000	0.063556	0.099960
10000	0.159921	0.597754
100000	1.279209	5.599902
200000	2.295506	10.867286

Table 4.7: Running times under different Monte Carlo Schemes, using 52 time steps.

Table of Bermudan option prices with the number of simulations, using 52 points per simulated path:

Path simulations	Euler Method (52 Paths)	Euler Method (365 Paths)	BM Reconstruction (52 Paths)
1000	77.1056	78.4155	78.1051
10000	76.5693	75.0282	76.5466
100000	76.6513	75.9154	75.2962
500000	76.6438	75.8959	75.3280
1000000	76.7700	75.5809	74.9250

Table 4.8: Option prices under each scheme with 52 time steps vs the 365 time step Euler Scheme

Table of errors of the fitted kernel density function vs the transition density function at time $T = 1$.

Path simulations	Euler Method Error ($Value * 10^{-10}$)	BM Reconstruction Error ($Value * 10^{-10}$)
10000	6.3653	3.5600
20000	4.7703	2.7437
50000	2.6712	0.95603
100000	1.4725	0.44052
200000	0.7948	0.22815
500000	0.6402	0.15060

Table 4.9: Relative errors of the simulations schemes vs the true transition density function at 52 time steps a year.

Table of time taken to simulate a given number of sample paths, with 12 points per sample path.

Path simulations	Euler Method (Seconds)	BM Reconstruction Method (Seconds)
100	0.001452	0.010394
1000	0.012296	0.056252
10000	0.066783	0.141748
100000	0.331540	1.163510
200000	0.628954	2.197024

Table 4.10: Running times under different Monte Carlo Schemes, using 12 time steps.

Table of Bermudan option prices with the number of simulation, using 12 points per simulated path:

Path simulations	Euler Method (12 Paths)	Euler Method (365 Paths)	BM Reconstruction (12 Paths)
1000	71.0216	69.8577	62.9658
10000	70.0505	66.2537	65.5288
100000	71.3304	67.1249	65.9176
500000	71.3038	67.1018	66.5687
1000000	70.9383	66.8527	66.1426

Table 4.11 Option prices under each scheme with 12 time steps vs the 365 time step Euler scheme.

Table of errors of the fitted kernel density function vs the transition density function at time $T = 1$.

Nr of path simulations	Euler Method Error ($Value * 10^{-10}$)	BM Reconstruction Error ($Value * 10^{-10}$)
10000	9.2519	8.2079
20000	10.961	4.1216
50000	8.1570	0.8937
100000	5.0516	0.67781
200000	6.4109	0.86671
500000	6.4923	0.45823

Table 4.12: Relative errors of the simulations schemes vs the true transition density function at 12 time steps a year.

4.2.2. **Analysis and conclusion on the results obtained for research question 1**

Analysis:

An analysis of the run times in the tables 4.4, 4.7 and 4.10 above show that the Euler scheme generates sample paths faster than the BM Reconstruction scheme. This is expected as simulating exact volatility paths are theoretically more complicated and involved than simulating such paths using finite time steps.

The difference between the run times suggest that on average, for 365 time steps per path, the Euler scheme is approximately 5.75 times faster than the BM Reconstruction scheme. For 52 time steps this reduces to being around 4-4.4 times faster than the BM Reconstruction scheme. Lastly, for 12 time

steps the discrepancy reduces further to 3.5 times, although for very small sample paths (100) the result can vary.

A very interesting observation occurs when one considers the accuracy in pricing options using the Euler scheme for variable time steps. In the tables 4.5, 4.8 and 4.11 with the given sample option prices, one sees that as the time step increases the accuracy of the Euler scheme comes into question, especially when the time steps are very short (12 steps). This shows that in order to obtain relatively accurate prices one must use smaller time steps and hence more computational time to maintain accuracy.

Recall, in Chapter 3 subsection 3.2, that the bias of a scheme is dependent on its (time) step size and that this implies that the expected option value and the true option value will not be the same. The expected values do however converge as the time step is taken closer to zero.

Now, the observation that the Euler scheme with 365 time steps and the BM reconstruction scheme with 12 steps show similar prices indicates that BM reconstruction scheme is more accurate. The exact simulation of the volatility processes leaves only the bias associated with constructing the asset price which is the main reason for its accuracy. This is also in line with the theoretical design of the scheme and the results are as would be expected.

Tables 4.6, 4.9 and 4.12 clearly show that for a large number of simulations the BM reconstruction scheme gives more accurate values, as measured by the MSE error criterion, than the Euler scheme. Again, this is expected as the bias under the BMR scheme should be less than that of the Euler scheme.

Conclusion and limitations of the investigation:

The conclusion from the analysis and the results in the tables suggest that when pricing options with many monitoring dates or options that are American in nature, the Euler scheme outperforms the BMR scheme. The Euler scheme's bias reduces when the time steps are decreased and shows faster run times when simulating a large number of paths with many time steps. Accuracy is also then maintained when pricing options and although the BMR scheme provides greater accuracy benefits it still takes longer to run.

In contrast to this, pricing options that are more European or that have fewer exercise dates will considerably favour the BMR scheme. If accuracy is desired, then pricing these options with a Euler scheme that simulates with a smaller time steps (e.g. 1/365) will actually lead to worse running times than under the BMR scheme (compare table 4.4 against table 4.10). The accuracy will either be on par or slightly better than the BMR scheme (see table 4.6 against 4.12). Thus it is recommended to price with the BMR scheme when pricing options with fewer exercise dates.

A first issue we need to consider now is whether there are any limitations to the conclusions reached here or other issues that were not addressed. In the course of the work the simulation was carried out using the Double Heston Model. It is within reason to believe that the conclusion given here will also apply to other stochastic volatility models that have additional processes such as volatility or interest rate CIR processes. This is mainly due to the fact that the CIR processes has been simulated exactly.

Second, implicit in our assumptions and calculations is the matter of having equal time steps in a year. The algorithms as found in the supporting material do however allow for uneven time steps and

the Author's recommendation here is that using the BMR scheme will allow for better treatment of uneven time steps as the volatility paths will show no bias. If uneven time steps are not too large then the Euler scheme may also be used, although there is no guarantee this is accurate.

Lastly, the Euler scheme poses no problems with division by zero. Under the BMR scheme when the volatility or interest rate simulated becomes arbitrarily close to zero then the division by the square root of the volatility or interest rate might lead to large Brownian Motion increments. This will lead to inaccurate results.

Usually this is rarely problematic, provided the starting volatility or interest rate values, e.g. $V(0)$, are not close to zero and the mean volatility, e.g. α in (2.28), is not close to zero. This is ultimately however a drawback of this scheme compared to the Euler scheme

Comparing the schemes as they are suggest that the BMR scheme will likely be better than the Euler scheme, without regard to run times. The final conclusion here is thus that the BMR scheme improves upon the Euler scheme in accuracy and, given certain option pricing conditions, gives faster run times as well.

4.2.3. **Investigating Research Question 2**

We will investigate the calibrated results for the various models to the market model's Call option prices. The exact details are given in Appendix F on using Differential Evolution to calibrate the models, as well as the finer details in implementing the method. The calibration exercise is determined to be a success if any evaluation of the objective function has an MSE less than 0.1 or until the maximum number objective function evaluations are reached (Appendix F). The results follow:

Double Heston Model Calibrated Parameters, with minimum objective value (MSE) of 0.0242 reached after 190 objective function evaluations:

Parameter	$v_1(0)$	$v_2(0)$	r	a_1	a_2	b_1	b_2	σ_1	σ_2
Value	0.0260	0.0130	0.0004	0.0070	0.0341	4.7720	2.4128	0.0306	0.1497
Parameter	ρ_1	ρ_2							
Value	-0.027	-0.530							

Table 4.13: Double Heston Model Calibrated Parameters

H2-CIR Model Calibrated Parameters, with minimum objective value (MSE) of 0.0907 reached after 980 objective function evaluations:

Parameter	$V(0)$	$r(0)$	κ	Θ_V	ν	λ	Θ_r	η	$\rho_{x,r}$*
Value	0.0179	0**	0.0013	0.0342	0.0007	0.0004	19.999	0.3207	0.0019
Parameter	$\rho_{x,V}$								
Value	0.9999								

Table 4.14: H2-CIR Model Calibrated Parameters

*The correlation between the asset price's Brownian Motion and the interest rate's Brownian Motion can be incorporated using Cholesky decomposition.

**The exact value is 0.0000091355, which for all practical purposes is near zero.

Note that when running simulations the inherent approximations in the model are not needed although essentially the full model is calibrated according to the approximated model.

4/2 Model Calibrated Parameters, with minimum objective value (MSE) of 0.0737 reached after 260 objective function evaluations:

Parameter	a	b	κ	θ	σ	V(0)*	r	ρ	
Value	0.0801	0.0551	6.2808	0.8716	0.4641	0.1	0.0033	-0.857	

Table 4.15: 4/2 Model Calibrated Parameters

*The parameter value has been fixed to 0.1 to avoid division by zero in the simulation algorithms.
Now that all the models are calibrated, we can price various options and compare them to the true prices. The choice of simulation scheme here does not impact on the results, even with slight variations in price. All the stochastic volatility models have been simulated using a Euler scheme, although the results are repeatable using the other scheme.

The exotic options that will be priced are Asian, Barrier, Cliquet, Lookback and Binary options. Note that the current asset value is by default $S(0) = 1000$. All options are priced in Dollar units $ for simplicity and reference. The Asian and Barrier options are priced for a range of strike prices. The specification used are:

Asian Option Specifications:

Type: Put Option

Exercise Time: 1-year, only at end of year.

Payoff: Classic arithmetic payoff structure, with the average taken over last 75 days before exercise.

Strikes: $K = 900,910,920, \dots ,1100$

Barrier Option Specifications:

Type: Up-and-out Put Option

Exercise Time: 1-year, only at end of year, option terminates if Barrier is hit during monitoring date.

Monitoring Dates: End of each day, 365 monitoring dates in year.

Barrier (Upper): $B = 1100$

Rebate if barrier is hit: $\$50$

Payoff: Put option payoff on the asset value at the end of the year less the rebate amount.

Strikes: $K = 900,910,920, \dots ,1050$

Cliquet:

Type: Call

Exercise Times: Expires in 3 monthly intervals with final expiry at the end of 1 year.

Strike: Floating, reset as current value of the underlying at the start of every 3 month interval.

Payoff: Standard Call option payoff per option, floating strike.

Lookback:

Type: Call

Exercise Time: End of the year.

Payoff: $\max(S(T) - S_{min}, 0)$

Strike: Floating, minimum of asset value during the year, monitored at the end of each day (365 monitoring dates).

Binary:

Type: Call

Exercise Time: End of the year.

Payoff: $100 if the price of the underlying at the end of the year is above 1100 and zero otherwise.

For each type of option a table is given below. Each table gives the market price, prices from each stochastic volatility model, obtained from running 250000 simulations using a Euler scheme. Lastly, the MSE error for all the prices relative to the Market Model are given at the end (where applicable).

Asian Put Options:

Strike:	Market Model	Double Heston	H2CIR	4/2
900	10.10	16.67	12.17	12.88
910	12.02	19.05	14.36	15.12
920	14.19	21.68	16.82	17.64
930	16.64	24.56	19.56	20.42
940	19.37	27.72	22.60	23.51
950	22.39	31.15	25.94	26.90
960	25.73	34.87	29.59	30.61
970	29.37	38.87	33.57	34.63
980	33.35	43.18	37.87	38.98
990	37.64	47.79	42.49	43.64
1000	42.27	52.71	47.44	48.63
1010	47.21	57.93	52.72	53.93
1020	52.48	63.45	58.30	59.54
1030	58.06	69.27	64.21	65.45
1040	63.95	75.38	70.41	71.65
1050	70.13	81.78	76.90	78.13
1060	76.60	88.45	83.67	84.89
1070	83.35	95.39	90.71	91.91

1080	90.36	102.59	98.00	99.17
1090	97.61	110.02	105.53	106.68
1100	105.10	117.69	113.29	114.40
MSE:	N/A	105.4796	30.1622	42.9481

Table 4.16: Asian Put option prices for all stochastic volatility models relative to the Market Model

Cliquet Call Option:

Model:	Market Model	Double Heston	H2CIR	4/2
Price:	107.25	110.43	109.57	109.84

Table 4.17: Cliquet Call option prices for all stochastic volatility models relative to the Market Model

Lookback Call Option:

Model:	Market Model	Double Heston	H2CIR	4/2
Price:	100.09	107.25	102.33	101.89

Table 4.18: Lookback Call option prices for all stochastic volatility models relative to the Market Model

Binary Call Option:

Model:	Market Model	Double Heston	H2CIR	4/2
Price:	23.11	23.33	23.01	22.74

Table 4.19: Binary Call option prices for all stochastic volatility models relative to the Market Model

Barrier Up-and-out Put Option:

Strike:	Market Model	Double Heston	H2CIR	4/2
900	8.69	12.59	8.51	9.70
910	10.74	14.91	10.74	11.98
920	13.00	17.43	13.21	14.48
930	15.49	20.16	15.91	17.22
940	18.20	23.11	18.85	20.19
950	21.14	26.27	22.03	23.40
960	24.31	29.66	25.46	26.85
970	27.71	33.27	29.12	30.52
980	31.34	37.10	33.01	34.43
990	35.18	41.14	37.13	38.55
1000	39.23	45.38	41.46	42.88
1010	43.48	49.83	46.00	47.40
1020	47.91	54.47	50.72	52.10
1030	52.51	59.28	55.62	56.96
1040	57.27	64.24	60.66	61.97
1050	62.15	69.35	65.83	67.11
MSE:	N/A	32.5078	4.0810	10.3000

Table 4.20: Barrier Up-and-out Put option prices for all stochastic volatility models relative to the Market Model

The following table gives the average time taken to compute the characteristic function of each model, using the calibrated parameters. The support of the characteristic function has 10000 points

spread evenly between 0 and 100.

Model:	Double Heston	H2-CIR	4/2
Time (Seconds):	0.009054	25.391072	136.525815

Table 4.21: Run times to compute 10000 points on the characteristic function

The apparent differences are due to mathematics involved in computing the characteristic function. The Double Heston model's implementation can be fully vectorised and the operations involved are all elementary. In contrast, calculating the H2-CIR or 4/2 model's characteristic function takes significantly more time to calculate. Additionally, the significantly longer run times are also due to implementing 100% robust versions for the H2-CIR and 4/2 models.

The last table gives the errors involved between the 1 and 5 year transition density function of the market model and the 1 and 5 year transition density functions obtained from each of the models:

Model:	Double Heston	H2-CIR	4/2
MSE at T =1:	$1.6879 * 10^{-9}$	$9.002 * 10^{-10}$	$1.074 * 10^{-9}$
MSE at T =5:	$2.0779 * 10^{-8}$	$1.18 * 10^{-8}$	N/A**

Table 4.22: Relative errors of the transition density function of all stochastic volatility models relative to the Market Model

**The 4/2 Model is not stable at time $T = 5$ when calculating the transition density function. From time $T = 4$ onwards, the characteristic function exhibits non-standard behaviour and thus the numerical inversion fails. For this reason the 4/2 Model is not included in Figure 4.4. below.

As a visual reference, the various transition densities are plotted on the same surface, for the two given times:

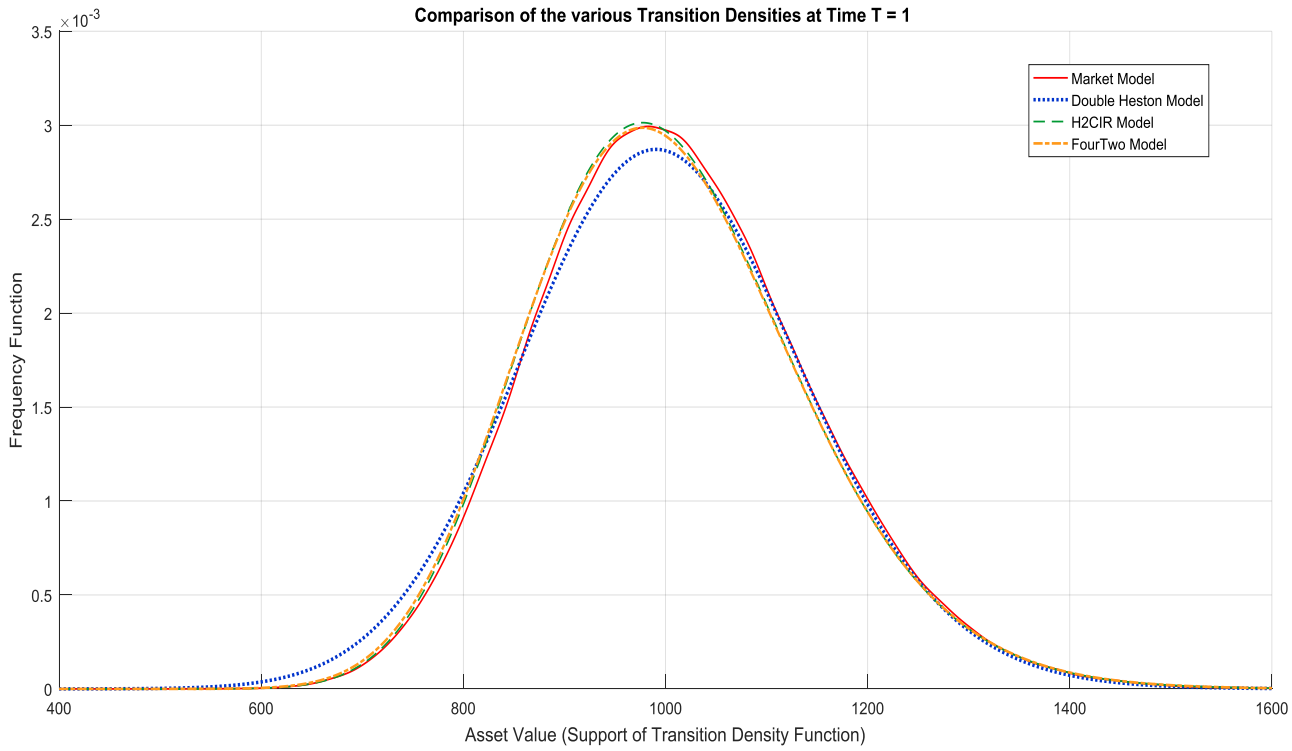

Figure 4.3: Transition density function of all stochastic volatility models relative to the Market Model at time $T = 1$

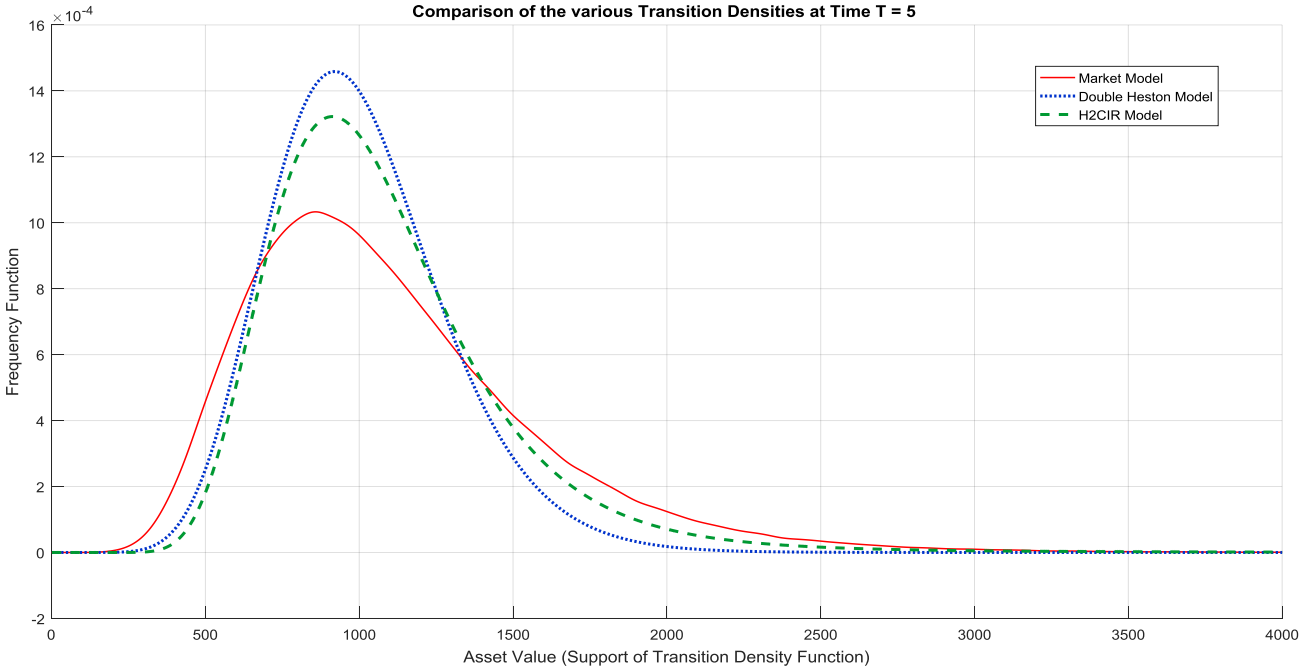

Figure 4.4: Transition density function of all stochastic volatility models relative to the Market Model at time $T = 5$

4.2.4. **Analysis and conclusion on the results obtained for research question 2**

Analysis:

Inspecting tables 4.16-4.20 and 4.22 shows very clearly that the Double Heston Model performed the worse out of the three stochastic volatility models with almost all the option prices being very far off from the market model's. This appears in spite of the fact that during the calibration of the model the Double Heston Model showed the lowest objective function evaluation, meaning its calibration to the sample prices was closer than for the other models.

In answering why this is the case we resort to exact transition density of the Double Heston Model at the time duration $T = 1$ in comparison to that of the market model. Recall that in section 4.2.1 the statement was made that convergence to the true transition density implies convergence to the true option price. This is correct provided the different discount rates implied by each of the different models correctly deflates their transition density functions to one discounted transition density function.

For the Market Model the true averaged discount rate implied by the zero coupon bond prices obtained from the stochastic interest rate models is $r = 0.033$. The rate under the Double Heston Model is $r = 0.0004$. The problem then comes when looking at the transition density function at time $T = 1$ in Figure 4.3, namely if the discount factor is applied to obtain a discounted transition density function then the Market Model would flatten down more than the Double Heston Model.

In essence the discounted transition density functions would be closer matched in the region of the support from 900 to 1100, which is where the majority of the option prices' strike prices were. This would give Call option prices that are closer matched in that region. For values greater than 1100 the transition densities for all the models die down quickly and the impact of the discount factor is

negligible. The problem lies in the transition density values seen for values of the support less than 900. Clearly the Double Heston mismatches this section. Having values on this transition density greater than the market model's implies that the Double Heston Model predicts lower asset values than the market at time $T = 1$, and generally for $T \in (0,1]$, and hence the Double Heston Model values Put options higher, as seen from the tables.

The cause of the Double Heston Model's poor performance has been determined. This however begs the question as to why the other two models exhibit transition densities and also discounted transition densities that are more closely matched to the Market Model in the tails of the density functions? The other two models have strictly been calibrated less robustly than the Double Heston Model (when looking at the objective function evaluation) and yet they outperform. This is also in spite of the fact that fewer model parameters are involved.

It is clear that the H2CIR Model is the winner in terms of having evaluated option prices that are the closest to the Market Model, with the 4/2 Model following (see MSE values at tables 4.16-4.20). Why both of these models performed well is not easily explainable. In fact, if we go to the extreme and introduce an even tighter MSE calibration criteria (i.e. 0.01) for the Double Heston Model, as done in Appendix F, then we find that the option prices still do not match near the accuracy of the H2-CIR and 4/2 models.

It would seem that then the answer lies in terms of each model's structure. For the Double Heston and 4/2 Model the $dS(t)/S(t)$ increment in the SDE follows a randomised normal distribution centred at $r * dt$, r being the risk free rate. The stochastic volatility components for the Double Heston Model effective means that we are adding two randomized normal distributions with a variance that is the sum of the individual volatility components. Mathematically for all the models we can construct this randomized normal distribution:

For the 4/2 Model:

$$\frac{dS(t)}{S(t)} \sim \mathcal{N}\left(r, dt * \left(a^2 V(t) + \frac{b^2}{V(t)} + 2ab\right)\right); a, b > 0.$$

For the Double Heston Model:

$$\frac{dS(t)}{S(t)} \sim \mathcal{N}\left(r, dt * \left(V_1(t) + V_2(t)\right)\right).$$

For the H2CIR Model:

$$\frac{dS(t)}{S(t)} \sim \mathcal{N}\left(r(t)dt, V(t)\right).$$

For the Market Model:

$$\frac{dS(t)}{S(t)} \sim \mathcal{N}\left(\frac{\left(r_1(t) + r_2(t)\right)}{2} dt, a^2 V_1^3(t) + \frac{b^2}{V_1(t)} + 2abV_1^2(t) + c^2 V_2^{\frac{3}{2}}(t)\right).$$

It is clear that the H2CIR and the Market Model have a stochastic mean induced by the stochastic interest rate factor. This allows the H2CIR Model to better match the Market Model. The stochastic variance of the Double Heston and 4/2 Model would on the other hand match the variance of the Market Model with it being expected that the 4/2 Model would outperform the Double Heston Model. The 4/2 Model captures more dynamic features of the Market Model's variance given its inverted CIR component.

Ultimately over the short horizon of 1 year the H2CIR outperforms the other models but this seemingly good outperformance shows over a longer period no real advantage. Over a period of 5 years the transition density functions do not match the Market Model's. Without any further calculations it is also clear that the discounted transition density functions will also not match the Market Model's as the Market Model's density function is flatter and more widespread

Conclusion and limitations of the investigation:

We conclude that if we have *a priori* beliefs or empirical evidence about the market not having any significant stochastic interest rate component then using either of the Double Heston or 4/2 Model would be suitable. For longer term considerations the 4/2 Model is not suitable as there is no guarantee of its parameter stability although for shorter term pricing considerations it is better suited than the Double Heston Model provided calculation time is not a significant factor.

If the market is believed or shown to exhibit a stochastic short rate process then the H2CIR Model is preferred. If the realised volatility in the market is shown to be sufficiently volatile then the 4/2 Model would also work well here, in spite of the assumption of a constant interest rate.

The calculations here suggest the H2CIR Model is sufficiently complex and computationally efficient to capture very complex market behaviour but ultimately, as with the other models, its parameters are not stable across time. This may suggest a time dependent H2CIR Model, which can be implemented analytically, although the implementation time on a computer might suffer because of this.

The study presented here does have the obvious limitation that no recommendation can be made as to a single purpose model ready for all real world market dynamics. Additionally the conclusions here might be different if a different market model were chosen. Lastly, time dependent parameters were not investigated and this extension to each of the stochastic volatility models might also result in different conclusions.

In conclusion, there is no single purpose stochastic volatility model that could sufficiently trump all others. Rather each model must be examined for its merits relative to other models and the purpose of using the model. Ultimately though, if a single model had to be chosen, then the H2CIR Model would be the recommended choice based on the calculations in subsection 4.2.3.

4.3. **Final Conclusion**

We have examined three stochastic volatility models with constant and stochastic interest rates that can be used to price exotic derivatives. The conclusions reached to the research question 1 and 2 suggest that the H2CIR Model should be used together with either a Euler scheme or the BMR scheme, depending on the characteristics of the option to be priced.

Finally, it must be recognised that it is unlikely that there will be a universal model to price exotic derivatives and that the markets are continually changing. The progression from the BS methodology to the sophisticated models today affirm this statement. Each model and pricing method must be suited for the purposes that it is to be used and in that regards it is hoped that the work done in this thesis can assist the practitioner to obtain greater clarity in assessing and vetting models for real world use.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. 1st ed. New York: Wiley, pp.257,505.

Ackerer, D., Filipović, D. and Pulido, S. (2016). *The Jacobi stochastic volatility model*. 1st ed. [Geneva]: Swiss Finance Institute.

Alam, M. and Uddin, G.S. (2009). Relationship between interest rate and stock price: Empirical evidence from developed and developing countries. *International Journal of Business and Management*, 4(3).

Albrecher, H., Mayer, P., Schoutens, W. and Tistaert, J. The Little Heston Trap. *Wilmott Magazine*, January issue, 83-92.

Attari, M. (n.d.). Option pricing using Fourier Transforms: A Numerically Efficient Simplification. *SSRN Electronic Journal*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=520042 [Accessed 12 May 2016].

Baldeaux, J. (2012). Exact simulation of the 3/2 model. *International Journal of Theoretical and Applied Finance (IJTAF), World Scientific Publishing Co. Pte. Ltd.*, vol. 15(05), pages 1-13.

Baldeaux, J. and Platen, E. (2013). *Functionals of Multidimensional Diffusions with Applications to Finance*. 1st ed. Cham: Springer International Publishing, pp.299-322.

Ball, C. and Roma, A. (1994). Stochastic volatility option pricing. *The Journal of Financial and Quantitative Analysis*, 29(4), pp.589-607.

Bardet, J.-M. et al., 2003. Semi-parametric estimation of the long-range dependence parameter: A Survey. *Theory and Applications of Long-Range Dependence*, (October 2015), pp.557–578.

Bayraktar, E.,Poor,H. and Sircar, K. (2004). Estimating the fractal dimension of the S&P500 index using wavelet analysis. *International Journal of Theoretical and Applied Finance*, 07(05), pp.615-643.

Barndorff-Nielsen, O.E. Normal inverse Gaussian distributions and the modeling of stock returns. *Research Report, Department of Theoretical Statistics, Aarhus University*, 300, 1995.

Bender, C., Lindner, A. and Schicks, M. (2011). Finite variation of fractional Lévy processes. *Journal of Theoretical Probability*, 25(2), pp.594-612.

Berberan-Santos, M. (2006). Expressing a probability density function in terms of another PDF: A Generalized Gram-Charlier Expansion. *Journal of Mathematical Chemistry*, 42(3), pp.585-594.

Björk, T. (2009). *Arbitrage theory in continuous time*. 3rd ed. Oxford: University press.

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

Brigo, D. and Mercurio, F. (2007). *Interest rate models - theory and practice*. 2nd ed. Berlin [u.a.]: Springer.

Bru, M. (1991). Wishart processes. *Journal of Theoretical Probability*, 4(4), pp.725-751.

Calvet, L. and Fisher, A. (2008). Multifractal volatility. 1st ed. Amsterdam: Academic Press, pp.7-50.

Campolieti, G. and Makarov, M. (2014). Financial Mathematics: A Comprehensive Treatment. 1st ed. Chapman and Hall/CRC.

Cariboni, J and Schoutens, W. (2009). Lévy processes in credit risk. 1st ed. Chichester: Wiley.

Carr, P., Geman, H., Madan, D. and Yor, M. (2003). Stochastic Volatility for Levy Processes. *Mathematical Finance*, 13(3), pp.345-382.

Carr, P., Geman, H., Madan, D. and Yor, M. (2002). The fine structure of asset returns: An Empirical Investigation. *The Journal of Business*, 75(2), pp.305-333.

Carr, P. and Wu, L. (2003). The finite moment log stable process and option pricing. *The Journal of Finance*, 58(2), pp.753-777.

Carr, P. and Wu, L. (2016). Analyzing volatility risk and risk premium in option contracts: A new theory. *Journal of Financial Economics*, 120(1), pp.1-20.

Cheung, Y. and Ng, L. (1992). Stock price dynamics and firm size: An empirical investigation. *The Journal of Finance*, 47(5), pp.19-85.

Chen, B. and Grzelak, L. (2012). Calibration and Monte Carlo pricing of the SABR–Hull–White model for long-maturity equity derivatives. *The Journal of Computational Finance*, 15(4), pp.79-113.

Christoffersen, P., Heston, S. and Jacobs, K. (2009). The Shape and Term Structure of the Index Option Smirk:

Why Multifactor Stochastic Volatility Models Work So Well. *Management Science*, 55(12), pp.14-32.
Chang, C., Chung, S. and Stapleton, R. (2007). Richardson extrapolation techniques for the pricing of American-style options. *Journal of Futures Markets*, 27(8), pp.791-817.
Cooley, J. and Tukey, J. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90), pp.297-297.
Cox, J. (1975). Notes on option pricing I: Constant elasticity of variance diffusion, *Technical report*, Stanford University, Graduate School of Business.
Reprinted in: *Journal of Portfolio Management*, Vol. 22, 1996
Derman, E. and Kani, I. (1994). Riding on a smile. *RISK* **7**: 9-32
Dieker, T., 2004. Simulation of fractional Brownian motion. *MSc theses, University of Twente, Amsterdam.* Available at: http://www.wstein.org/home/wstein/www/home/simuw/simuw08/refs/fractal/dieker-2004-thesis-simulation_of_fractional_Brownian_motion.pdf.
Duffie, D., Pan, J. and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-diffusions. *Econometrica*, 68(6), pp.43-76.
Dufresne, D. (2001). The integrated square-root process. Working paper, University of Montreal.
Dupire, B. (1994). Pricing with a smile. *RISK* **7**: 18-20
Fan, J. and Pan, J. (2005). Convergence Properties of a Self-adaptive Levenberg-Marquardt Algorithm Under Local Error Bound Condition. *Computational Optimization and Applications*, 34(1), pp.47-62.
Fang, F. and Oosterlee, C. (2009). A novel pricing method for European options based on Fourier-Cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2), pp.826-848.
Fang, F. and Oosterlee, C. (2009). Pricing early-exercise and discrete barrier options by Fourier-Cosine series expansions. *Numerische Mathematik*, 114(1), pp.27-62.
Figlewski, S. (2010). Estimating the implied risk neutral density for the U.S. market portfolio, in M. Watson, T. Bollerslev and J. Russell (eds), *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle.* Oxford: Oxford University Press.
Filipović, D. and Larsson, M. (2016). Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4), pp.931-972.
Folland, G. (2009). *Fourier analysis and its applications*. 1st ed. Providence: American Mathematical Society.
Forde, M. and Jacquier, A. (2010). Robust approximations for pricing Asian options and Volatility Swaps under stochastic volatility. *Applied Mathematical Finance*, 17(3), pp.241-259.
Fouque, J., Papanicolaou, G. & Sircar, K., (2000). Stochastic volatility correction to Black-Scholes. *Risk*, (August 1999), pp.1–9. Available at: http://143.248.27.21/mathnet/paper_file/stanford/papani/risk.pdf.
Fouque, J., Papanicolaou, G., Sircar, R. and Solna, K. (2003). Multiscale Stochastic Volatility Asymptotics. *Multiscale Modeling & Simulation*, 2(1), pp.22-42.
Fouque, J., Papanicolaou, G., Sircar, R. and Sølna, K. (2011). *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives*. 1st ed. Cambridge: Cambridge University Press, pp.51-147.
Gatheral, J. (2006). *The volatility surface*. 1st ed. Hoboken, N.J.: John Wiley & Sons, pp.1-8.
Gauthier, P. and Possamai, D. (2010). Efficient Simulation of the Double Heston Model. *SSRN Electronic Journal*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1434853 [Accessed 19 Aug. 2016].
Giles, M.B (2008a). Improved multilevel Monte Carlo convergence using the Milstein scheme. *Monte Carlo and Quasi-Monte Carlo Methods 2006 (A. Keller, S. Heinrich and H. Niederreiter, eds), Springer, pp. 343–358.*
Giles, M.B. (2008b). Multilevel Monte Carlo Path Simulation. *Operations Research*, 56(3), pp.607-617.
Giles, M.B and Waterhouse, B. (2009). Multilevel quasi-Monte Carlo path simulation. *Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics.* De Gruyter, pp. 165–181.
Goard, J., 2011. Pricing of volatility derivatives using 3/2-stochastic models. *S. I. Ao, l. gelman, d. WL. hukins, A. Hunter & a. m. korsunsky (Eds.), World Congress on Engineering 2011,* pp. 433-438.
Gourieroux, C., Jasiak, J. and Sufana, R. (2009). The Wishart Autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2), pp.167-181.
Grasselli, M. (2016). THE 4/2 STOCHASTIC VOLATILITY MODEL: A UNIFIED APPROACH FOR THE HESTON AND

THE 3/2 MODEL. *Mathematical Finance*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2523635 [Accessed 7 Jun. 2016].
Grzelak, L. and Oosterlee, C. (2011). On the Heston Model with Stochastic Interest Rates. *SIAM Journal on Financial Mathematics*, 2(1), pp.255-286.
Grzelak, L., Oosterlee, C. and Van Weeren, S. (2011). The affine Heston model with correlated Gaussian interest rates for pricing hybrid derivatives. *Quantitative Finance*, [online] 11(11), pp.1647-1663. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1434829 [Accessed 23 Aug. 2016].
Grzelak, L., Oosterlee, C. and Van Weeren, S. (2012). Extension of stochastic volatility equity models with the Hull–White interest rate process. *Quantitative Finance*, 12(1), pp.89-105.
Hagan, P., Kumar, D., Lesniewski, A. and Woodward, D. (2002). Managing Smile Risk. *WILMOTT Magazine*, (wilm003.qxd), pp.84-108.
Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), pp.327-343.
Heston, S. (1997). A simple new formula for options with stochastic volatility*. Manuscript: John M. Olin, School of Business, Washington University.*
Hull, J. and White, A. (1987). The Pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2), pp.281-300.
Hull, J. and White, A. (1990). Pricing interest-rate-derivatives securities. *Review of Financial Studies*, 3(4), pp.573-592.
Hunter, C. (2006). *Hybrid derivatives*. 1st ed. Hoboken, N.J.: Wiley.
Jiang, L., Chen, Q., Wang, L. and Zhang, J. (2003). A new well-posed algorithm to recover implied local volatility. *Quantitative Finance*, 3(6), pp.451-457.
Kallenberg, O. (2010). *Foundations of modern probability*. 1st ed. New York, NY: Springer.
Kendall, M.G., Stuart, A., and Ord, J.K. (1994). K*endall's advanced theory of statistics: Distribution theory vol. 1*. 6th ed. New York. Wiley, pp 78-121.
Kienitz, J. and Wetterau, D. (2012). *Financial modelling*. 1st ed. Hoboken, N.J.: Wiley, pp.17-183.
Kleinert, H. and Korbel, J. (2016). Option pricing beyond Black–Scholes based on double-fractional diffusion. *Physica A: Statistical Mechanics and its Applications*, 449, pp.200-214.
Kou, S. and Wang, H. (2004). Option Pricing Under a Double Exponential Jump Diffusion Model. *Management Science*, 50(9), pp.78-92.
Kyprianou, A. (2006). *Introductory lectures on fluctuations of Levy processes with applications*. 1st ed. Berlin [etc.]: Springer, pp.7-68.
Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), pp.164-168.
Lewis, A. (2001). A Simple Option Formula for General Jump-Diffusion and Other Exponential Levy Processes. *SSRN Electronic Journal*. [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=282110 [Accessed 24 May 2016].
Liu, J. and Lampinen, J. (2002). On setting the control parameter of the differential evolution method. *Proceedings of the 8th International Conference on Soft Computing (MENDEL). Brno, Czech Republic. pp. 11–18.*
Lord, R., Koekkoek, R. and Dijk, D. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2), pp.177-194.
Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4), pp.422-437.
Madan, D., Carr, P. and Chang, E. (1998). The Variance Gamma Process and Option Pricing. *Review of Finance*, 2(1), pp.79-105.
Makarov, R. and Glew, D. (2010). Exact simulation of Bessel diffusions. *Monte Carlo Methods and Applications*, [online] 16(3-4). Available at: https://arxiv.org/pdf/0910.4177.pdf [Accessed 5 Mar. 2016].
Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), pp.431-441.
Meintanis, S. and Taufer, E. (2012). Inference procedures for stable-Paretian stochastic volatility models.

Mathematical and Computer Modelling, 55(3-4), pp.1199-1212.
Melnikov, M. and Melnikov, Y.A. (2007). Construction of Green's function for the Black-Scholes equation, *Electronic Journal of Differential Equations,* Vol. 2007, No. 153, Pp. 1-14.
Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2), pp.125-144.
Mil'shtejn, G. (1975). Approximate Integration of Stochastic Differential Equations. *Theory of Probability & Its Applications*, 19(3), pp.557-562.
Nolan, J. (2013). Financial modeling with heavy-tailed stable distributions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1), pp.45-55.
Oehlert, G. (1992). A Note on the Delta Method. *The American Statistician*, 46(1), pp.27-29.
Overhaus, M., Bermudaz, A., Buehler, H., Ferraris, A., Jordinson, C. and Lamnouar, A. (2007). *Equity hybrid derivatives*. 1st ed. Hoboken, N.J: Wiley, pp.1-18, 23-74.
Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. Ph.D. Albert-Ludwigs-Universit¨at Freiburg.
Price, K., Lampinen, J. and Storn, R. (2005). *Differential Evolution*. 1st ed. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
Rebonato, R. (2004). *Volatility and correlation*. 1st ed. Chichester, West Sussex, England: J. Wiley, pp.1-85.
Renault, E and Touzi, N. (1996). Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance* **6**(3): 279-302.
Rice, J. (2007). *Mathematical statistics and data analysis*. 1st ed. S.l.: Cengage Learning.
Rodrigo, M. and Mamon, R. (2007). Recovery of Time-Dependent Parameters of a Black-Scholes-Type Equation: An Inverse Stieltjes Moment Approach. *Journal of Applied Mathematics*, 2007, pp.1-8.
Ross, S., Cox, J. and Ingersoll, J. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), pp.385-407.
Samorodnitsky, G. and Taqqu, M. (1994). Levy Measures of Infinitely Divisible Random Vectors and Slepian Inequalities. *The Annals of Probability*, 22(4), pp.30-56.
Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. 1st ed. Cambridge: Cambridge University Press.
Schroder, M. (1989). Computing the Constant Elasticity of Variance Option Pricing Formula. *The Journal of Finance*, [online] 44(1), pp.211-219. Available at: http://www.ase.ro/upcpr/profesori/167/schroder.pdf [Accessed 11 Feb. 2016].
Shampine, L. (2008). Vectorized adaptive quadrature in MATLAB. *Journal of Computational and Applied Mathematics*, 211(2), pp.131-140.
Sircar,R. and Papanicolaou, G. (1999). Stochastic volatility, smile and asymptotics. *Applied Mathematical Finance* **6**(2): 107-45.
Squire, W. and Trapp, G. (1998). Using Complex Variables to Estimate Derivatives of Real Functions. *SIAM Review*, 40(1), pp.110-112.
Stein, E. and Stein, J. (1991). Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, [online] **4**(4), pp.727-752. Available at: https://scholar.harvard.edu/files/stein/files/rfs-1991.pdf [Accessed 17 Mar. 2016].
Storn, R. and Price, K. (1995). *Differential evolution*. 1st ed. Berkeley, Calif: ICSI, pp.1-347.
Wallace, D. (1958). Asymptotic Approximations to Distributions. *The Annals of Mathematical Statistics*, 29(3), pp.635-654.
Wu, Q. (2010). Series Expansion of the SABR Joint Density. *Mathematical Finance*, 22(2), pp.310-345.
Zolotarev, V. (1986). *One-dimensional stable distributions*. 1st ed. Provindence, R.I.: American Mathematical Society.

Appendices

Appendix A: Programming code

Please note that all code used to produce this thesis remains the intellectual property of the University of the Free State unless otherwise specified that the code is from a third party, in which case due acknowledgement will be given. Very little code used to produce this thesis will be published, this is in the interest of saving space, but some segments of code can be found either in Appendix E or in some of the chapters in this thesis.

Note that almost all of the programming is done in MATLAB® R2016b Student Edition and in C++. A license (40374188) has been granted by Mathworks to perform academic work with the aim of completing course work and writing a thesis. Additional packages were purchased in addition to the basic package offered in order to facilitate research and reduce the programming load on the Author.

Additionally, all source code and data used to produce every figure can be found under the author's MATLAB® file exchange. All figures are saved and can simply be viewed in MATLAB® or as JPEG files. Note that some of the code my not run depending on the toolboxes the user has installed and/or the hardware the user has available.

The Author has tried, where possible, to create programming code from "scratch", however, since the number of MATLAB® lines of code is already over 200 pages, the Author has taken shortcuts to reduce the programming load. For instance, using the MATLAB® Statistics and Machine Learning Toolbox has allows the programming of random number generators which would take even more code and time to produce.

Details to obtaining the programming code and files for the data used:

If you are possession of a physical copy of this thesis then the accompanying CD should contain all of the relevant programming code with instructions to replicate the results in this thesis. All figures and code are saved on this CD and it is assumed the reader has at least MATLAB® (R2016a) installed.

Alternatively, if you are reading an electronic copy then I may be contacted on my public email jaundre.s1@gmail.com and from thereon I can provide instructions to obtaining the material from a cloud storage facility (since the ZIP file sizes are mostly too large to be sent via email). On special request a USB containing all the code can be couriered to an individual, *residing within* South Africa.

If any of the above fail, then a request can be sent to the University of The Free State, Department of Statistics, Mathematical Statistics and Actuarial Science in order to obtain the programming code as the code rightly falls under the University of The Free State's jurisdiction.

Appendix B: Supplementary Explanations and Proofs.

This appendix contains various supplementary explanations, proofs and summaries of material as referenced from the previous chapters. The material is split by chapter and each subsection varies in length and content. The entire purpose of including this material is to aid the reader in understanding some concepts and to fill in the gaps for readers that are unfamiliar with the body of literature regarding specific issues. The numbering of equations is by chapter with the prefix "B" attached, for instance the first equation below in under the heading "Chapter 1" is B1.1.1 where the last 1 indicates the equation number under the subsection B1.1.

Chapter 1: Supplementary Explanations

The material in this section relates to the content as found in Chapter 1. This material is quite short and the proof can best be understood after the reader has read the section on risk neutral dynamics in Chapter 3, subsection 3.1.3.

B.1.1. Determining the Minimum of the Graph of the Volatility Smile: Renault-Touzi Theorem

Consider the following stochastic volatility model:

$$\frac{dS(t)}{S(t)} = rdt + \sqrt{V(t)}dW_0^{\mathbb{Q}}(t). \qquad \text{(B. 1.1.1)}$$

Here the volatility process $V(t)$ is some Markov process driven by another Brownian motion $dW_1^{\mathbb{Q}}(t)$ that is uncorrelated to $dW_0^{\mathbb{Q}}(t)$. Here the assumption is that a suitable risk-neutral measure exists so that (B.1.1.1) can be used to price contingent claims.

Pricing of a European Call option under (B.1.1.1) can be done using some simplifying assumption by using the Hull-White Formula (See Fouque et al 2011). The gist of the argument is given by defining

$$\bar{V} = \frac{1}{T-t}\int_t^T V(s)ds. \qquad \text{(B. 1.1.2)}$$

the price of a European Call option can be given by

$$C(t) = E^{\mathbb{Q}(\gamma)}\left[C_{BS}\left(t, T, S(t), K, \sqrt{\bar{V}}\right)\middle|\mathcal{F}_t\right]. \qquad \text{(B. 1.1.3)}$$

Here C_{BS} is the Black-Scholes price for a European Call option. The market price of risk, as referenced by (3.25), under the risk neutral measure $\mathbb{Q}$ is given by γ and its inclusion in pricing the Call option in (B.1.1.3) is referenced by the notation $\mathbb{Q}(\gamma)$.

To show that the Stochastic Volatility in (B.1.1.1) implies a volatility smile, the statement by (Renault and Touzi, 1996) is formulated below and proved:

In an uncorrelated Stochastic Volatility model, provided $\bar{V}$ is an L^2 random variable, the implied volatility curve $I(K)$ for fixed t, x, T is a smile, that is , it is locally convex around the minimum $K_{min} = S(t)e^{r(T-t)}$, which is the forward price of the stock.

Proof:

Without loss of generality, assume for the moment that $\bar{V}$ can take only the two values:

$$\bar{V} = \begin{cases} V_1 \; with \; probability \; p \\ V_2 \; with \; probability \; 1-p \end{cases} \tag{B. 1.1.4}$$

under the probability measure $\mathbb{Q}(\gamma)$. Applying the expectation in (B. 1.1.3) then yields:

$$C_{BS}\big(K; I(p,K)\big) = p * C_{BS}(K; V_1) + (1-p) * C_{BS}(K, V_2). \tag{B. 1.1.5}$$

Here the implied volatility level, depending on p, K, with other parameters being fixed, is defined by $I(p,K)$. Now, by differentiating (B.1.1.5) with respect to K yields

$$\frac{\partial C_{BS}}{\partial V}\big(I(p,K)\big)\frac{\partial I}{\partial K} + \frac{\partial C_{BS}}{\partial K}\big(I(p,K)\big) = p * \frac{\partial C_{BS}}{\partial K}(V_1) + (1-p)\frac{\partial C_{BS}}{\partial K}(V_2). \tag{B. 1.1.6}$$

Note that

$$sign\left(\frac{\partial I}{\partial K}\right) = sign\big(g(p)\big).$$

where the function g is defined, for some fixed value for K, as:

$$g(p) = p * \frac{\partial C_{BS}}{\partial K}(V_1) + (1-p)\frac{\partial C_{BS}}{\partial K}(V_2) - \frac{\partial C_{BS}}{\partial K}\big(I(p,K)\big). \tag{B. 1.1.7}$$

It follows from simple algebra that if $p = 0{,}1$ then $I(p,K) = V_2, V_1$ which implies that $g(0) = g(1) = 0$. The next step is to differentiate (B.1.1.7) with respect to p:

$$g'(p) = \frac{\partial C_{BS}}{\partial K}(V_1) - \frac{\partial C_{BS}}{\partial K}(V_2) - \frac{\partial^2 C_{BS}}{\partial K \partial V}\big(I(p,K)\big)\frac{\partial I}{\partial p}. \tag{B. 1.1.8}$$

This expression can be simplified by replacing $\frac{\partial I}{\partial p}$ with an expression obtained from differentiating (B.1.1.5) with respect to p:

$$\frac{\partial C_{BS}}{\partial V}\big(I(p,K)\big)\frac{\partial I}{\partial p} = C_{BS}(V_1) - C_{BS}(V_2). \tag{B. 1.1.9}$$

Differentiating (B.1.1.8) with respect to p and $\frac{\partial I}{\partial p}$ with respect to p, by using (B.1.1.9), gives the following result:

$$g''(p) = \left(\frac{C_{BS}(V_1) - C_{BS}(V_2)}{\frac{\partial C_{BS}}{\partial V}}\right)^2 \left(\frac{\frac{\partial^2 C_{BS}}{\partial K \partial V}\frac{\partial^2 C_{BS}}{\partial V^2}}{\frac{\partial C_{BS}}{\partial V}} - \frac{\partial^2 C_{BS}}{\partial K \partial V}\right). \tag{B. 1.1.10}$$

Note that all partial derivatives are evaluated at I. Using the known BS equation in (1.4) and taking appropriate derivatives leads us to simplify (C.1.1.10):

$$g''(p) = \frac{\big(C_{BS}(V_1) - C_{BS}(V_2)\big)^2}{\frac{\partial C_{BS}}{\partial V}(I)}\frac{2L}{(T-t)KI^3}. \tag{B. 1.1.11}$$

Here $L = \log\left(\frac{Se^{r(T-t)}}{K}\right)$. Some comments on (B.1.1.11):

Since $\frac{\partial C_{BS}}{\partial V} > 0$ and thus C_{BS} is increasing in V. Trivially, since $I > 0$ it follows that $sign\big(g''(p)\big) = sign(L)$. Hence for any $K < K_{min} = Se^{r(T-t)}$, we have that $L > 0 \Rightarrow g''(p) > 0$ and thus g achieves a minimum in [0,1] which implies that $g < 0$ for $0 < p < 1$. Thus $\frac{\partial I}{\partial K} < 0$ since as mentioned we have that $sign\left(\frac{\partial I}{\partial K}\right) = sign\big(g(p)\big)$. Similarly if $K > K_{min}$, then $L < 0$ which implies that $g''(p) < 0$ and hence that g achieves a maximum in [0,1] and thus $g > 0$ for $0 < p < 1$. Thus $\frac{\partial I}{\partial K} > 0$ which finally implies that at $K = K_{min}$ we must have $\frac{\partial I}{\partial K} = 0$. ■

The above argument can be expanded to incorporate any number of values that $\bar{V}$ can take by using an inductive argument as can be found in (Sircar and Papanicolaou, 1999)). The proof above is based on the arguments of the mentioned authors. On a final note, as mentioned in Chapter 1, the minimum derived in the above proof for the volatility smile is not always seen in practice. This implies that,

where this is observed, the market dynamics cannot be described by uncorrelated stochastic volatility factors or that more than one stochastic volatility factor is needed.

Chapter 2 Supplementary Explanations And Proofs

The material in this section relates to a wide variety of content as found in Chapter 2.

B.2.1. Derivation of the Ordinary Differential Equations Underlying the Zero Coupon Bond Prices As Determined By the CIR Model

The derivation is quite simple and the general case will be presented here. The reader can find more information in (Campolieti and Makarov, 2014), p.624-625. The material is best read after working through the derivation of the $A(t,T)$ & $C(t,T)$ functions of the CIR Model in Chapter 2. The material here simply completes the proof by considering the derivation of the ODE's underlying the proof in Chapter 2.

Recall that we are interest in a short rate model that produces zero-coupon bond prices and that takes the form:

$$Z\big(t,T,r(t)\big) = e^{A(t,T)-C(t,T)r(t)}. \tag{B. 2.1.1}$$

Additionally, recall that a general affine term structure model for the short rate process can be expressed as

$$dr(t) = \tilde{a}\big(t,r(t)\big)dt + b\big(t,r(t)\big)d\widetilde{W}(t). \tag{B. 2.1.2}$$

here $\tilde{a}\big(t,r(t)\big) = a\big(t,r(t)\big) - \gamma\big(t,r(t)\big)b\big(t,r(t)\big)$ and the Brownian motion term is a standard Brownian motion under some risk-neutral measure $\mathbb{Q}$. Hence using Itô's formula on (B.2.1.1) one readily derives

$$\frac{\mathrm{d}Z\big(t,T,r(t)\big)}{Z\big(t,T,r(t)\big)} = \left[\frac{\partial A(t,T)}{\partial t} - \frac{\partial C(t,T)}{\partial t}r(t) - C(t,T)\tilde{a}\big(t,r(t)\big) + \frac{1}{2}C^2(t,T)b^2\big(t,r(t)\big)\right]dt - C(t,T)b\big(t,r(t)\big)d\widetilde{W}(t). \tag{B. 2.1.3}$$

Since the model in (B.2.1.2) has been specified in terms of risk-neutral dynamics we know that the drift term in (B.2.1.3) must be equal to the risk-free rate $r(t)$:

$$g\big(t,r(t)\big) = \frac{\partial A(t,T)}{\partial t} - \frac{\partial C(t,T)}{\partial t}r(t) - C(t,T)\tilde{a}\big(t,r(t)\big) + \frac{1}{2}C^2(t,T)b^2\big(t,r(t)\big) - r(t). \tag{B. 2.1.4}$$

Clearly then g must be identically zero for all values of t & $r(t)$. Note that if the drift and diffusion coefficients in (B.2.1.2) are affine in $r(t)$, then they can be written in the following forms:

$$\tilde{a}\big(t,r(t)\big) = a_0(t) + a_1(t)r(t);\;\; b^2\big(t,r(t)\big) = b_0(t) + b_1(t)r(t). \tag{B. 2.1.5}$$

It is understood that all the functions in (B.2.1.5) are non-random and only functions of time (or constant).

As is shown by (Campolieti and Makarov, 2014) the following pricing function

$$Z\big(t,T,r(t)\big) = E_{\mathbb{Q}}\left[e^{-\int_t^T r(u)du}\;|\mathcal{F}_t\right].$$

satisfies the following PDE via the Discounted Feynman-Kac Theorem:

$$\frac{\partial Z}{\partial t} + \frac{b^2(t,r)}{2}\frac{\partial^2 Z}{\partial r^2} + a(t,r) - \gamma(t,r)b(t,r)\frac{\partial Z}{\partial r} - rZ = 0. \tag{B. 2.1.6}$$

where the notation is shortened for readability, e.g. $r(t) = r$ and the terminal condition for this equation is $Z\big(T,T,r(t)\big) = 1$. Now by substituting (B. 2.1.1) into (B. 2.1.6) one obtains:

$$\frac{\partial A(t,T)}{\partial t} - \left(1 + \frac{\partial C(t,T)}{\partial t}\right) r(t) - C(t,T)\tilde{a}\big(t, r(t)\big) + \frac{1}{2} C^2(t,T) b^2\big(t, r(t)\big) = 0. \quad \text{(B. 2.1.7)}$$

This equation has terminal conditions $A(T,T) = 0$ & $C(T,T) = 0$. Substituting (B.2.1.7) into (B.2.1.5) yields

$$\frac{\partial A(t,T)}{\partial t} - a_0(t)C(t,T) + \frac{b_0(t)}{2} C^2(t,T)$$
$$-\left(\frac{\partial C(t,T)}{\partial t} + a_1(t)C(t,T) - \frac{b_1(t)}{2} C^2(t,T) + 1\right) r(t) = 0.$$

Since this equation is identically zero it implies that the following ODE's can be formed, which when solved analytically or numerically, yield the appropriate functions that define the zero-coupon bond prices in (B.2.1.1):

$$\frac{\partial C(t,T)}{\partial t} + a_1(t)C(t,T) - \frac{b_1(t)}{2} C^2(t,T) + 1 = 0$$
$$\frac{\partial A(t,T)}{\partial t} - a_0(t)C(t,T) + \frac{b_0(t)}{2} C^2(t,T) = 0.$$

Solving these equations numerically is done using the Runge-Katta method. For certain choices of the coefficients in (B. 2.1.5) these ODE's can be solved analytically, as was done in Chapter 2 under the CIR Model.

B.2.2. **Green's Function and an introduction into Fractional Calculus**

It is important to gain a high level understanding of the material below rather than get bogged down with the specific details. The Author has reworked the material significantly from the original articles and source material to better aid understanding for the reader who is likely to be unfamiliar with the concepts.

Green's Function

Assume that $L = L(x)$ is a linear differential operator. An example of this operator is $L = \frac{\partial}{\partial x} + a$ so that $Ly(x) = \frac{\partial y}{\partial x} + ay(x)$, a is some arbitrary constant.

Green's function, denoted by $G(x,s)$, is a solution to the following problem:

$$LG(x,s) = \delta(s - x). \quad \text{(B. 2.2.1)}$$

The Dirac delta function $\delta(x)$ can be defined rigorously under a measure theoretic framework (see (Campolieti and Makarov, 2014), Chapter 9), however it suffices to consider it here as the limit of a normal distribution with zero mean and variance tending to zero:

$$\delta(x) = \begin{cases} +\infty, if\ x = 0 \\ 0\ , otherwise \end{cases}$$

Additionally $\int_{-\infty}^{\infty} \delta(x)dx = 1$. Essentially this function is used within the context of the following integration problem concerning (B.2.2.1):

$$\int LG(x,s)f(s)ds = \int \delta(s - x)f(s)ds = f(x). \quad \text{(B. 2.2.2)}$$

Clearly when faced with the problem of finding a function $u(x)$ such that $Lu(x) = f(x)$, then (using (B.2.2.1)) the function $u(x)$ can be found by:

$$Lu(x) = \int LG(x,s)f(s)ds = L\int G(x,s)f(s)ds$$

$$\therefore u(x) = \int G(x,s)f(s)ds. \qquad \text{(B. 2.2.3)}$$

In other words an ordinary differential equation $Lu(x) = f(x)$ can be solved by knowing the Green Function associated with the operator L. This is the essence of the Green Function. The concept applies to solving partial differential equations as well. Naturally for various operators L the corresponding Green Function has been determined, for example consider the BS operator is given as

$$L = \frac{\partial}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2}{\partial x^2} + r\left(x\frac{\partial}{\partial x} - \cdot\right).$$

The corresponding Green Function (with source point S') is given by

$$G(S,t:S') = \frac{\exp\big(-r(T-t)\big)}{S'[2\pi\sigma^2(T-t)]^{\frac{1}{2}}} \exp\left(-\frac{\left[ln\left(\frac{S}{S'}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)\right]^2}{2\sigma^2(T-t)}\right).$$

For details regarding this and for a general method of constructing Green Function from various terminal conditions associated with BS operator, see (Melnikov and Melnikov, 2007).

Fractional Calculus

It is well known in calculus that $\frac{\mathrm{d}^m}{\mathrm{d}x^m}(x^n) = \frac{n!}{(n-m)!}x^{n-m}$ for some positive integers $n \geq m$. The question may be asked whether any non-integer values for n, m are allowable or even if the concept of a "fractional" derivatives makes sense? There is a formula, known as the Cauchy formula, that serves as a starting point:

$$\int_{x_0}^{x}\int_{x_0}^{x_1}\dots\int_{x_0}^{x_{n-1}} f(x_n)\mathrm{d}x_n \dots \mathrm{d}x_1 = \frac{1}{(n-1)!}\int_{x_0}^{x}(x-y)^{n-1}f(y)\mathrm{d}y. \qquad \text{(B. 2.2.4)}$$

The formula (B.2.2.4) applies only to integer values for n. Recall from statistics that the gamma function exhibits the following property: $\Gamma(n) = (n-1)!$ for integer values n. Clearly the gamma function can take on non-integer values and thus (B.2.2.4) can be generalized by defining:

$$_{x_0}I_x^v f(x) \coloneqq \frac{1}{\Gamma(v)}\int_{x_0}^{x}(x-y)^{v-1}f(y)\mathrm{d}y. \qquad \text{(B. 2.2.5)}$$

Note that $_{x_0}I_x^v$ is an operator that defines a fractional integral with respect to some function and value $v \in \mathbb{R}$. There are a few formulations of a fractional derivative which are based on the operator in (B.2.2.5). The first is known as the Riemann-Liouville (RL) fractional derivative:

$$_{x_0}D_x^v f(x) = \begin{cases} \dfrac{d^{\lceil v\rceil}}{dx^{\lceil v\rceil}}\left(_{x_0}I_x^{\lceil v\rceil - v}f(x)\right), & v > 0 \\ f(x), & v = 0 \\ _{x_0}I_x^{-v}f(x), & v < 0 \end{cases} \qquad \text{(B. 2.2.6)}$$

Here the notation $\lceil v\rceil$ denotes the lowest integer exceeding v, e.g. $\lceil 4.3\rceil = 5$.
Two concrete examples of (B.2.2.6) are given here (the 1st is original and the 2nd shows the explicit calculation steps not found in the article of (Kleinert and Borbel, 2016):
1.) Let $f(x) = x^2$, find the 1.4^{th} derivative of $f(x)$ (assuming $x_0 = 0$):

$$_0I_x^{2-1.4}f(x) = {}_0I_x^{0.6}f(x) = \frac{1}{\Gamma(0.6)}\int_0^x (x-y)^{0.6}y^2dy$$

$$\therefore \quad _0D_x^{1.4}f(x) = \frac{d^2}{dx^2}\left(_{x_0}I_x^{\lceil v\rceil - v}[f]\right)(x) = \frac{1}{\Gamma(0.6)}\frac{d^2}{dx^2}\int_0^x (x-y)^{-0.4}y^2dy$$

$$= 0.6715 * \frac{d^2}{dx^2}\int_0^x y^2 \sum_{k=0}^{\infty} \binom{-0.4}{k} x^k (-y)^{-0.4-k}\, dy, \qquad \binom{k}{m} = \frac{\Gamma(k+1)}{\Gamma(m+1)\Gamma(k-m+1)}.$$

This equation does not (to the Author's knowledge) have any closed form expression and the reader can apply a truncation of the series to obtain an approximate expression and evaluate the derivative at some point if desired

2.) Let $f(x) = C, C\ is\ some\ constant$. Find the v^{th} derivative, assuming $x_0 = 0$:

$${}_0I_x^{\lceil v\rceil - v} f(x) = {}_0I_x^{\lceil v\rceil - v} C = \frac{C}{\Gamma(\lceil v\rceil - v)}\int_0^x (x-y)^{\lceil v\rceil - v - 1} dy = \frac{C}{\Gamma(v')}\int_0^x (x-y)^{v'-1} dy$$

$$= \frac{C}{\Gamma(v'+1)} x^{v'}; v' = \lceil v\rceil - v.$$

$$\therefore\ {}_0D_x^v C = \frac{d^{\lceil v\rceil}}{dx^{\lceil v\rceil}} \frac{C}{\Gamma(v'+1)} x^{v'} = \frac{d^{\lceil v\rceil}}{dx^{\lceil v\rceil}} \frac{C}{\Gamma(\lceil v\rceil - v + 1)} x^{\lceil v\rceil - v} = \frac{C * \lceil v\rceil * x^{-v}}{\lceil v\rceil \Gamma(1-v)} = \frac{C x^{-v}}{\Gamma(1-v)}.$$

This derivative is zero $iff\ v \in \mathbb{N}$ since the gamma function explodes to infinity at these values and the derivative then tends to zero.

As the last example shows, an unwanted property (there are others as well) of the RL derivative is that a constant function has a non-zero (fractional) derivative. This prompts us to consider a different formulation for the fractional derivative. The new definition is known as the Caputo (C) fractional derivative, which is a bit easier to follow:

$$ {}_{x_0}^{*}D_x^v f(x) \coloneqq \frac{1}{\Gamma(\lceil v\rceil - v)}\int_{x_0}^x \frac{f^{\lceil v\rceil}(y)}{(x-y)^{v+1-\lceil v\rceil}} dy. \qquad \text{(B. 2.2.7)}$$

In fact by looking at the second example given above the apt reader would have noticed how the definition of this fractional derivative could have come about. In any case, the definition in (B.2.2.7) correctly allows any constant function to have fractional derivative of zero, however the price to pay is that the function f must be $\lceil v\rceil$ times differentiable. In any case, the connection between the two types of derivatives is

$$ {}_0^{*}D_x^v f(x) = {}_0D_x^v f(x) - \sum_{k=0}^{\lceil v\rceil} \frac{x^k}{k!} f^k(0).$$

The Caputo derivative can be involved in the formulation of fractional differential equations such as

$$ {}_0^{*}D_x^v f(x) = \lambda f(x). \qquad \text{(B. 2.2.8)}$$

Here λ is some constant parameter. Problems such as (B.2.2.8) are interesting, not just academically, but practically in derivative pricing as will be seen. The solution to (B.2.2.8) is given by the Mittag-Leffler function defined as:

$$E_{v,\xi}(x) = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(vn+\xi)}. \qquad \text{(B. 2.2.9)}$$

The solution to (B.2.2.8) is then given by $f(x) = E_{v,1}(\lambda x^v)$. In fact (B.2.2.8) is actually a generalization of the well-known formula involving the exponential function:

$$\frac{d^m}{dx^m}\exp(\lambda x) = \lambda^m \exp(\lambda x). \qquad \text{(B. 2.2.10)}$$

Recall that the exponential function (the "$f(x)$" in (B.2.2.8)) is defined by $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, thus the function in (B.2.2.9) essentially generalizes the exponential function. An interesting question which turns the previous problem in (B.2.2.8) around by asking is there a value for $x_0\ (= 0)$ in the definition of the Caputo derivative in (B.2.2.7) such that property in (B.2.2.10) holds for values $m = v$ (fractional values)? Mathematically find a value x^* such that:

$$ {}_{x^*}^{*}D_x^v \exp(\lambda x) = \lambda^v \exp(\lambda x). \qquad \text{(B. 2.2.11)} $$

Essentially solving this problem will lead to another type of fractional derivative that plays an important role in the analysis to come namely the Riesz-Feller (RF) derivative defined as ${}_{x^*}^{*}D_x^v$.
To solve problem (B.2.2.11), start off with definition (B.2.2.7):

$$ LHS = \frac{1}{\Gamma(\lceil v \rceil - v)} \int_{x^*}^{x} \frac{f^{\lceil v \rceil}(y)}{(x-y)^{v+1-\lceil v \rceil}} dy = \frac{1}{\Gamma(\lceil v \rceil - v)} \int_{x^*}^{x} \frac{\lambda^{\lceil v \rceil} \exp(\lambda y)}{(x-y)^{v+1-\lceil v \rceil}} dy $$

$$ = \frac{\lambda^{\lceil v \rceil}}{\Gamma(\lceil v \rceil - v)} \int_{x^*}^{x} \exp(\lambda y)\,(x-y)^{\lceil v \rceil - v - 1} dy $$

$$ = \frac{\lambda^{\lceil v \rceil}}{\Gamma(\lceil v \rceil - v)} \int_{\lambda x^*}^{\lambda x} \exp(u) \left(x - \frac{u}{\lambda}\right)^{\lceil v \rceil - v - 1} \frac{du}{\lambda} = \frac{\lambda^v}{\Gamma(\lceil v \rceil - v)} \int_{\lambda x^*}^{\lambda x} \exp(u)\,(\lambda x - u)^{\lceil v \rceil - v - 1} du. $$

Solving this integral requires us to recall that the gamma function can be split into the incomplete lower and upper gamma functions (the first was introduced under the CIR Model earlier):

$$ \Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt = \int_x^{\infty} t^{a-1} e^{-t} dt + \int_0^{x} t^{a-1} e^{-t} dt = \Gamma(a,x) + \gamma(a,x). $$

It then becomes a straightforward exercise to show that the following indefinite integral is related to the upper incomplete gamma function:

$$ \int \exp(u)\,(b-u)^v du = \exp(b)\,\Gamma(v+1, b-u). \qquad \text{(B. 2.2.12)} $$

Hence solving the definite integral above:

$$ LHS = \frac{\lambda^v}{\Gamma(\lceil v \rceil - v)} \left(\exp(\lambda x) * \Gamma(\lceil v \rceil - v, \lambda x - \lambda x) - \exp(\lambda x) * \Gamma(\lceil v \rceil - v, \lambda x - \lambda x^*) \right) $$

$$ = \frac{\lambda^v}{\Gamma(\lceil v \rceil - v)} \Big(\exp(\lambda x) \left(\Gamma(\lceil v \rceil - v) - \Gamma(\lceil v \rceil - v, \lambda x - \lambda x^*) \right) \Big). $$

The function $\Gamma(\lceil v \rceil - v, \lambda x - \lambda x^*)$ can be made zero if $x^* \to -\infty$ and consequently:

$$ LHS = \frac{\lambda^v}{\Gamma(\lceil v \rceil - v)} \Big(\exp(\lambda x) \left(\Gamma(\lceil v \rceil - v) \right) \Big) = \lambda^v \exp(\lambda x) = RHS. \qquad \blacksquare $$

Since the value of $x_0 = -\infty$ the corresponding derivative function that preserves (B.2.2.11) under fractional derivative taking is ${}_{-\infty}^{*}D_x^v$ which we define as the RF derivative. To improve clarity in the notation we set ${}_{-\infty}^{*}D_x^v := D_x^v$ and thus to summarise, the three fractional derivatives considered so far are denoted by:
RL: ${}_0D_x^v$
C: ${}_0^*D_x^v$
RF: D_x^v

The last piece needed to continue our discussions is the introduction of a new differential operator which is related to the characteristic function defining α-stable distributions in (2.122). This differential operator is defined indirectly via the Fourier transform and is known as a Lêvy pseudo-differential operator, denoted by ${}^{\theta}D_x^v$. The Fourier transform of this operator is given the notation $\overline{{}^{\theta}D_x^v}$. The definition is:

$$ \overline{\left[{}^{\theta}D_x^v f \right]}(p) = \phi(v, \theta, 1, 0)(p) \bar{f}(p). \qquad \text{(B. 2.2.13)} $$

here $f(p)$ is some arbitrary function and $\bar{f}(p)$ is the Fourier transform of $f(p)$.

B.2.3. **Detailed look at the Jacobi Model**

Before carrying out our analysis, recall that Gram-Charlier A expansions are used in statistics to

approximate the characteristic function of a probability density function (unknown) via using the characteristic function of another probability density function that is known. For instance, see (Wallace, 1958), the general method can be briefly summarised:

Let ϕ be the characteristic function of an unknown density function F and denote by κ_r the cumulants of this unknown probability distribution. Suppose that another known density function has a probability density function Ψ, characteristic function ψ and cumulants φ_r. As is known, characteristic functions can be expressed in terms of cumulants (Berberan-Santos, 2006):

$$\phi(t) = \exp\left[\sum_{r=1}^{\infty} \kappa_r \frac{(it)^r}{r!}\right]. \tag{B. 2.3.1}$$

$$\psi(t) = \exp\left[\sum_{r=1}^{\infty} \varphi_r \frac{(it)^r}{r!}\right]. \tag{B. 2.3.2}$$

Dividing expressions (B.2.3.1) and (B.2.3.2) and rearranging leads to

$$\phi(t) = \exp\left[\sum_{r=1}^{\infty} (\kappa_r - \varphi_r) \frac{(it)^r}{r!}\right] \psi(t). \tag{B. 2.3.3}$$

Using the properties of the Fourier transform and taking the Fourier transform of (B.2.3.3) with a change of variables "$x = -x$" leads to:

$$F(x) = \exp\left[\sum_{r=1}^{\infty} (\kappa_r - \varphi_r) \frac{(-D)^r}{r!}\right] \Psi(x). \tag{B. 2.3.4}$$

here $D = \frac{d^r}{dx^r}$. Typically $\Psi(x)$ is chosen to be the distribution function of a Normal random variable and the cumulants for the unknown density are set as $\mu = \kappa_1$ and $\sigma^2 = \kappa_2$ where μ, σ are the mean and standard deviation of the Normal distribution. In this case (B.2.3.4) can be reformulated:

$$F(x) = \exp\left[\sum_{r=3}^{\infty} \kappa_r \frac{(-D)^r}{r!}\right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{B. 2.3.5}$$

With this background knowledge in mind the discussion of the Jacobi Model is continued.

The true power for the Jacobi Model lies in its efficient computation of option prices. To gain an understanding of the dynamics a general pricing formula is derived here. Firstly, consider the distribution of X_T in (2.137). It is shown in the article mentioned that X_T has a density function $g_T(x)$ with support on $\mathbb{R}$ that satisfies the following condition if $C(T) > 0$:

$$\int_{-\infty}^{\infty} e^{\epsilon x^2} g_T(x) dx < \infty. \tag{B. 2.3.6}$$

where $\epsilon < 1/(2v_{max}T)$ and C(T) is defined by

$$C(t) = \int_0^t \Big(V(s) - \rho^2 Q\big(V(S)\big)\Big) ds. \tag{B. 2.3.7}$$

Furthermore if

$$E\left[C_T^{-\frac{1}{2}-k}\right] < \infty. \tag{B. 2.3.8}$$

holds for some $k \in \mathbb{N}_0$ then $g_T(x)$ and $e^{\epsilon x^2} g_T(x)$ are uniformly bounded and $g_T(x)$ is k-times continuously differentiable. The condition in (B.2.3.8) holds for any $k \in \mathbb{N}$ provided $v_{min} > 0$ and $\rho^2 < 1$.

Indeed, if (B.2.3.7) holds then the moment generating function of $g_T(x)$, i.e.

$$\widehat{g_T}(z) = \int_{\mathbb{R}} e^{zx} g_T(x) dx. \tag{B.2.3.9}$$

extends to an entire function in $z \in \mathbb{C}$ and as will be shown, a Gram-Charlier A approximation of the density $g_T(x)$ does converge in an L^2-sense (as defined below). Importantly, if (2.202) holds for $k = 0$, then for *any* Gaussian density $w(x)$ with mean μ_w and variance $\sigma_w > \frac{v_{max}T}{2}$ the following relationship holds

$$\int_{\mathbb{R}} \frac{\left(g_T(x)\right)^2}{w(x)} dx < \infty. \tag{B.2.3.10}$$

Keeping these considerations in mind will allow us to formally derive the price of a European option with a discounted payoff function $f(X_T)$. The only assumption needed is that (B.2.3.10) holds for $k = 0$. Firstly define the following weighted Lebesgue space:

$$L^2_w = \left\{ f(x) \middle| \|f\|^2_w = \int_{\mathbb{R}} f(x)^2 w(x) dx < \infty \right\}. \tag{B.2.3.11}$$

The Hilbert space in (B.2.3.11) is defined with the following "weighted" inner product:

$$\langle f, g \rangle_w = \int_{\mathbb{R}} f(x) g(x) w(x) dx. \tag{B.2.3.12}$$

An orthonormal basis (a review of the definition of a *basis* is given a few paragraphs below) for L^2_w can be constructed out of the so-called generalized Hermite polynomials, for $n \geq 0$:

$$H_n(x) = \frac{1}{\sqrt{n!}} \mathcal{H}_n\left(\frac{x - \mu_w}{\sigma_w}\right). \tag{B.2.3.13}$$

The function $\mathcal{H}_n(x)$ is simply given by:

$$\mathcal{H}_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}. \tag{B.2.3.14}$$

It is clear from inspection that the degree of $H_n(x) = n$ and that $\langle H_m, H_n \rangle = 1 \; iff \; m = n$ and zero otherwise. Note that $l(x) = \frac{g_T(x)}{w(x)} \in L^2_w$ and assume that the any contingent claim (on the log price X_t) with payoff function satisfies $f(x) \in L^2_w$. This latter condition is satisfied for many of the well-known exotic derivative pricing rules. Denote the price of the corresponding derivative contract by π_f and hence given equations (B.2.3.8)-(B.2.3.14) this price is well-defined and given by

$$\pi_f = \int_{\mathbb{R}} f(x) g_T(x) dx = \langle f, l \rangle_w = \sum_{n \geq 0} f_n l_n. \tag{B.2.3.15}$$

Each of coefficients in (B.2.3.15) are defined by

$$f_n = \langle f, H_n \rangle_w. \tag{B.2.3.16}$$

$$l_n = \langle l, H_n \rangle_w = \int_{\mathbb{R}} H_n(x) g_T(x) dx. \tag{B.2.3.17}$$

It should be clear that (B.2.3.15) needs to be truncated to some order, say $n = N$, and consequently the price is

$$\pi^N_f = \sum_{n=0}^{N} f_n l_n. \tag{B.2.3.18}$$

As would be expected, $\pi^N_f \to \pi_f$ as $N \to \infty$, although if $f(x)$ is some polynomial function then $\pi^N_f = \pi_f$ whenever $N \geq \deg f(x)$, this result being obvious from the orthogonality of the basis polynomials in (B.2.3.13).

As matter of curiosity, the Author believes non-standard (other than the exotic derivatives given in the article by (Ackerer et al, 2016) payoff functions can potentially be priced using (B.2.3.18) without

truncation by using the just mentioned property. For instance, if $f(x) = \sqrt{x} * \log(x)$ then the "degree" of f can be found using either of the standard formula $\deg(f) = \lim_{x\to\infty} \frac{xf'(x)}{f(x)}$ or $\deg(f) = \lim_{x\to\infty} \frac{\log(|f(x)|)}{\log(x)}$. The degree of $f(x)$ in this case is then 0.5 which *could* imply that $\pi_f^N \to \pi_f$ for $N \geq 1$, by considering the Hermite polynomials as forming a *Schauder basis* for (B.3.24). However, this is left as a topic for future research.

Continuing the discussion, the equations in (B.2.3.15), (B.2.3.17) leads us to define (B.2.3.18) via the Gram-Charlier A approximation of the log price density $g_T(x)$**:

$$\pi_f^N = \sum_{n=0}^{N} \langle f, l_n H_n \rangle_w = \int_{\mathbb{R}} f(x) g_T^N(x)\, dx. \tag{B.2.3.19}$$

$$g_T^N(x) = \sum_{n=0}^{N} l_n H_n(x) w(x). \tag{B.2.3.20}$$

Clearly the pricing approaches in (B.2.3.19) and (B.2.3.20) are designed to avoid numerical integration where possible, although in (B.2.3.21) it is not left out completely. Clearly the last ingredient needed to use (B.2.3.20) is a formula for the coefficients l_n in order to compute the density g_T^N or to use (B.2.3.19) directly.

Technical note**: The Gram-Charlier A series in (B.2.3.20) converges in a $L^2_{1/w}$ sense for the density function of the log price X_T, while interestingly for the Heston Model it does not converge. This is essentially because of the $Q(v)$ function in (2.137).

The mathematics is more involved to find l_n as the formulas (B.2.3.17) & (B.2.3.20) essentially involve a "tautology" where l_n is expressed in terms of $g_T^N(x)$ and vice-versa, which does not help. The solution to this problem is the following result:

$$l_n = [h_1(V_0, X_0), \dots, h_M(V_0, X_0)] e^{TG} \boldsymbol{e}_{\pi(0,n)}, \qquad 0 \leq n \leq N. \tag{B.2.3.21}$$

However, it is difficult to understand (B.2.3.21) unless it is derived and partly because the notation, i.e. the indexation, used in the derivation (as in the article) is tricky in itself and makes it difficult to just simply state the form of the $M \times M$-matrix G (defined more precisely below). To better understand the concepts behind the solution (B.2.3.21), a review of some facts from Linear Algebra and Calculus is given and the notation is fixed.

The notation ∇ in the following sections mean $\nabla = \left(\frac{\partial}{\partial v}, \frac{\partial}{\partial x}\right)$ which is a vector containing two partial derivatives, i.e. $\nabla f = \left(\frac{\partial f}{\partial v}, \frac{\partial f}{\partial x}\right)$ and the notation ∇^2 is the Hessian Matrix defined by:

$$\nabla^2 = \begin{bmatrix} \frac{\partial^2}{\partial v^2} & \frac{\partial^2}{\partial x \partial v} \\ \frac{\partial^2}{\partial v \partial x} & \frac{\partial^2}{\partial x^2} \end{bmatrix}.$$

Next, denote by P_k^n the vector space of polynomials in k-variables of degree less than n. For instance consider P_2^3, where some examples of this vector space includes $xy^2 - 4xy + 5, y^3 - x^3 + 7x, xy - 2x, 4, xy + x^2 - y^3$, etc. in the variables x, y (for more than two variables we index the variables, i.e. for $k = 5$ we have the variables $x_1, x_2, \dots x_5$). Note that each one of the examples given can be written as a linear combination of one of the monomials in the set

$$\{x^0 y^0 (= 1), y, y^2, y^3, x, xy, xy^2, x^2, x^2 y, x^3\}. \tag{B.2.3.22}$$

i.e. $xy^2 - 4xy + 5 = 5 * 1 + 0 * y + 0 * y^2 + 0 * y^3 + 0 * x + (-4) * xy + 1 * xy^2 + 0 * x^2 + 0 * x^2 y + 0 * x^3$. In fact we call this set of monomials a *basis* for the vector space (or more informally,

"collection") of polynomials, meaning any polynomial in P_2^3 can be written as a linear combination of these monomials (strictly speaking vectors) in (B.2.3.22). Important to realise is that the each of the vectors in the basis (B.2.3.22) should be linearly independent, meaning one must not be able to write a vector as a linear combination of the others in the basis. Indeed, in the above example no monomial can be written as a linear combination of the others.

The *dimension* of the vector space (of P_k^n and any other vector space in general) is simply the number of vectors needed in the basis such that the whole space can be generated (*spanned*) through linear combinations of the basis vectors. For our purposes the dimension of P_k^n is given by $\binom{n+k}{n}$ and indeed applying this formula to P_2^3 gives 10 which is exactly the number of monomials in the set given above. If there are less than 10 basis vectors to span P_2^3 then some polynomials will be missing from this vector space that otherwise should be included per definition and if there are more than 10 then these vectors can be removed as they can be written as linear combinations of the 10 basis vectors.

Lastly, consider again the basis in (B.2.3.21) and the example P_2^3. In Linear Algebra, there is a theorem that says any n-dimensional vector space is isomorphic to $\mathbb{R}^n$ (Theorem 8.2.3. Anton & Rorres (2011)) which means (for our purposes here) that we can associate a matrix of real numbers to the basis in (B.2.3.22), e.g. the first element of (B.2.3.22) can be described by the unit vector:

$$(1,0,0,0,0,0,0,0,0,0)$$

and subsequently the i^{th} element in (B.2.3.22) is associated (via a linear transformation map) with the i^{th} unit vector. This gives a 10×10-matrix with ones along the diagonal and zeros otherwise. While this is simplistic, the *idea* is important for the discussion below.

Taking these considerations in mind, the formula for l_n can now be derived. Denote by $\mathcal{G}$ the generator of the Jacobi Model in (2.137) and (2.138) which is given by:

$$\mathcal{G}f(v,x) = b(v)^T \nabla f(v,x) + \frac{1}{2} Tr\big(a(v)\nabla^2 f(v,x)\big). \qquad \text{(B. 2.3.23)}$$

where drift vector $b(v)$ and diffusion matrix $a(v)$ are

$$b(v) = \begin{bmatrix} \kappa(\theta - v) \\ r - q - v/2 \end{bmatrix}, \qquad a(v) = \begin{bmatrix} \sigma^2 Q(v) & \rho\sigma Q(v) \\ \rho\sigma Q(v) & v \end{bmatrix}. \qquad \text{(B. 2.3.24)}$$

Clearly $b(v) \in P_1^n$ and $a(v) \in P_2^n$. Thus the operator $\mathcal{G}$ essentially maps any polynomial of degree n to a polynomial of degree n or less. Since we have two state variables, the volatility $V(t)$ and the log price $X(t)$, the discussion will involve only the vector space P_2^N. The dimension of this vector space is denoted by $M = (N+2)(N+1)/2$, where the notation is changed from n to N to denote the degree of the polynomials. A basis for this vector space is

$$\{h_1(v,x), \dots, h_M(v,x)\}. \qquad \text{(B. 2.3.25)}$$

and (B.2.3.25) will be given explicitly below. Next, denote by G the matrix of the linear map $\mathcal{G}$ restricted to P_2^n with respect to the basis (B.2.3.25). The entries of G will be made explicit later on. In order to move on we need a powerful theorem from (Filipović and Larsson, 2016), Theorem 3.1, which is formulated as Theorem 2.6 here:

Theorem 2.6

For any polynomial $p(v,x) \in P_2^n$ and $0 \le t \le T$ we have

$$E[p(V_T, X_T)|\mathcal{F}_t] = [h_1(V_t, X_t), \dots, h_M(V_t, X_t)]e^{(T-t)G}\vec{p},$$

where $\vec{p} \in \mathbb{R}^M$ is the coordinate representation of the polynomial $p(v,x)$ with respect to the basis $h_1(v,x), \dots, h_M(v,x)$. In particular, all moments of (V_T, X_T) are finite.

Assume $N \ge 1$. The tricky part is now to define some index $\pi \in (1, \dots, M)$ that is an enumeration of the set of exponents

$$\mathcal{E} = \{(m,n): m,n \geq 0; m+n \leq N\}, \qquad \text{(B. 2.3.26)}$$

and use this as a means of indexing the vectors in the basis (B.2.3.26). This is best explained via an example. Suppose that we are concerned with the vector space P_2^3 and the basis is given by (B.2.3.25). The following table gives one example of how to apply (B.2.3.26) to (B.2.3.25):

Normal Indexation	Indexation via $\mathcal{E}$	Normal Indexation	Indexation via $\mathcal{E}$
$h_1(v,x)$	$h_{\pi(0,0)}$	$h_6(v,x)$	$h_{\pi(1,2)}$
$h_2(v,x)$	$h_{\pi(0,1)}$	$h_7(v,x)$	$h_{\pi(0,2)}$
$h_3(v,x)$	$h_{\pi(1,0)}$	$h_8(v,x)$	$h_{\pi(2,0)}$
$h_4(v,x)$	$h_{\pi(1,1)}$	$h_9(v,x)$	$h_{\pi(0,3)}$
$h_5(v,x)$	$h_{\pi(2,1)}$	$h_{10}(v,x)$	$h_{\pi(3,0)}$

Table B.1.1: Enumeration of the basis of a vector space under the Jacobi Model.

So for instance, the 4^{th} basis vector is associated with the index $\pi(1,1)$, in other words $4 = \pi(1,1)$. Also, it is impossible in this instance to have an index $\pi(3,1)$ as then $m+n > N$ in (B.2.3.26) and thus $(3,1) \notin \mathcal{E}$ and $\pi(3,1)$ cannot be associated with any number in $(1,\dots,M)$. Note that the number of ways to form the pair (m,n) in (B.2.3.26) is exactly equal to M and the arrangement as given in the above table is not unique. The particular arrangement has no effect on the end result and we could associate any number index $\pi(m,n)$ with any number in $(1,\dots,M)$ provided this is specified uniquely and no single index specifies two different numbers. Hence the basis (B.2.3.25) can be specified:

$$h_{\pi(m,n)} = v^m H_n(x), \qquad (m,n) \in \mathcal{E} \qquad \text{(B. 2.3.27)}$$

Using the property

$$H_n'(x) = \frac{\sqrt{n}}{\sigma_w} H_{n-1}(x), \qquad n \geq 1 \qquad \text{(B. 2.3.28)}$$

the $M \times M$-matrix G can be constructed. The exact construction details are omitted but what is remarkable is that each column, or more precisely the $\pi(m,n)^{th}$ column, has at most 7 nonzero elements (with all other elements zero!) and these are defined by:

$$G_{\pi(m-2,n),\pi(m,n)} = -\frac{\sigma^2 m(m-1)v_{max}v_{min}}{2\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}, \qquad m \geq 2$$

$$G_{\pi(m-1,n-1),\pi(m,n)} = -\frac{\sigma\rho m\sqrt{n}v_{max}v_{min}}{\sigma_w\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}, \qquad m,n \geq 1$$

$$G_{\pi(m-1,n),\pi(m,n)} = \kappa\theta m + \frac{\sigma^2 m(m-1)(v_{max}+v_{min})}{2\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}, \qquad m \geq 1$$

$$G_{\pi(m,n-1),\pi(m,n)} = \frac{(r-q)\sqrt{n}}{\sigma_w} + \frac{\sigma\rho m\sqrt{n}(v_{max}+v_{min})}{\sigma_w\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}, \qquad n \geq 1$$

$$G_{\pi(m+1,n-2),\pi(m,n)} = \frac{\sqrt{n(n-1)}}{2\sigma_w^2}, \qquad n \geq 2$$

$$G_{\pi(m,n),\pi(m,n)} = -\kappa m - \frac{\sigma^2 m(m-1)}{2\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}$$

$$G_{\pi(m+1,n-1),\pi(m,n)} = -\frac{\sqrt{n}}{2\sigma_w} - \frac{\sigma\rho m\sqrt{n}}{\sigma_w\left(\sqrt{v_{max}}-\sqrt{v_{min}}\right)^2}, \qquad n \geq 1.$$

For example suppose we are in the $\pi(2,1)^{\text{th}}$ column, i.e. the 5^{th} according to the table. Then clearly $m = 2$ and $n = 1$ and some of the above nonzero values exist. For instance $G_{\pi(m+1,n-1),\pi(m,n)}$

becomes $G_{\pi(3,0),\pi(2,1)} = G_{10,5}$ and hence the 10^{th} row and the 5^{th} has a non-zero entry. If the index is out of bounds, $G_{\pi(3,-1),\pi(2,1)}$ then no entry can be defined, although this is no problem as all entries are assumed zero and only if an index is within bounds then that particular entry is filled. In any case let us now repeat the final result for l_n as given above:

$$l_n = [h_1(V_0, X_0), \dots, h_M(V_0, X_0)]e^{TG}\boldsymbol{e}_{\pi(0,n)}, \qquad 0 \le n \le N. \qquad \text{(B. 2.3.29)}$$

Here e_i is the i^{th} standard basis vector in $\mathbb{R}^M$ as defined earlier.

The Jacobi Model is primarily, in the Author's view, a model that optimises the pricing of options and the main contribution is the peculiar pricing methodology given above, which works for various types of derivatives. Other models considered in this chapter are based on giving more realistic asset price dynamics. Thus our focus in the next chapter is based on models that lean towards explaining the implied volatility surface, rather than optimising the pricing approach.

A side note**: MATLAB® has a distinct advantage out of many programming packages in that it is designed to work efficiently with matrices and a very fast calculation of e^{TG} can be done even for large matrices. In practice the allowable size of the matrix M is restricted by the computer's memory.

Chapter 4: Supplementary Explanations and Proofs

This section is very short.

B.4.1. Hull-White model's zero coupon bond price formula

Recall the HW formula and the general zero coupon bond formula:

$$dr(t) = \lambda\big(\theta - r(t)\big)dt + \sigma dW(t).$$
$$Z(t, T, r) = \exp(A(t, T) - C(t, T)r(t)).$$

Formulas for $A(t, T)$ & $C(t, T)$ can be given (Mamon, 2004):

$$A(t, T) = \left(\theta - \frac{\sigma^2}{2\lambda^2}\right)\big(B(t, T) - (T - t)\big) - \frac{\sigma^2}{4\lambda}\big(B(t, T)\big)^2$$
$$B(t, T) = \frac{1 - e^{-\lambda(T-t)}}{\lambda}.$$

Appendix C: Computational Considerations: C++ Coding and Parallel Computing to reduce simulation times

In many mathematical finance applications it is necessary to run simulations of asset paths. As mentioned in section 3.2, using exact simulation schemes are computationally expensive, but they require less simulation as the Monte Carlo convergence rate is faster than for biased schemes. This means that to compute some quantity, e.g. such as option prices, less simulations are needed. Using non-exact schemes such as the Euler Scheme takes less time to run, but require more simulations to achieve accuracy.

The Author believes that a balance between computational speed, the number of simulations, and the accuracy of the quantities calculated can be achieved by the reconstruction scheme given in subsection 3.2.4. Now, having decided upon a simulation scheme, the question is how to accelerate the computational time using computational techniques. In what follows below the perspective is given from the computational environment in MATLAB®, although the techniques also apply to other high-level programming languages such as R, Python, Julia, etc.

The first technique that we will us is the idea of taking MATLAB® code and converting this into C++ code. It is well-known that C++ provides considerable speed advantages over high-level programming languages such as those mentioned in the previous paragraph. The down-side is that development time takes considerably longer in C++ to code the algorithms and ensure that they work accurately.

In MATLAB® it is possible to call C++ code through what is known as MEX files and functions. As an example, consider the simulation scheme for the Double Heston Model in subsection 3.2.5.1.

Running the code in the MATLAB® environment yields, using the parameters in subsection 3.2.5.1, a 1000 simulations with each simulated path over a year having 365 points (thus 365000 simulated points), an average computational time of 68.974398 seconds. Now, by using the same random numbers, the computational time using C++ code and Mex files is reduced to 0.687596 seconds, which is around a 100 times faster on average.

C++ is considerably faster for running *for* loops and doing any logical branching (*if* and *else* type statements). Since the simulations run sequentially, it is difficult to vectorise the above code and hence the MATLAB® environment does not give optimal run times for the code. This is partly the reason why converting to C++ code gives a considerable speed-up. (Note that these conclusions are valid for other high level programming languages as well, e.g. Python, R, etc.)

Below is a plot 10000 simulations, taking an average of 6.592456 seconds. The plot below generates a heatmap of values. The histogram of values at the terminal date, i.e. the kernel distribution of values at $T = 1$, resembles the exact density function at that time quite closely, as was shown in section 3.2.51. Note however that running there are still further improvements that can be made in order to reduce the computational time.

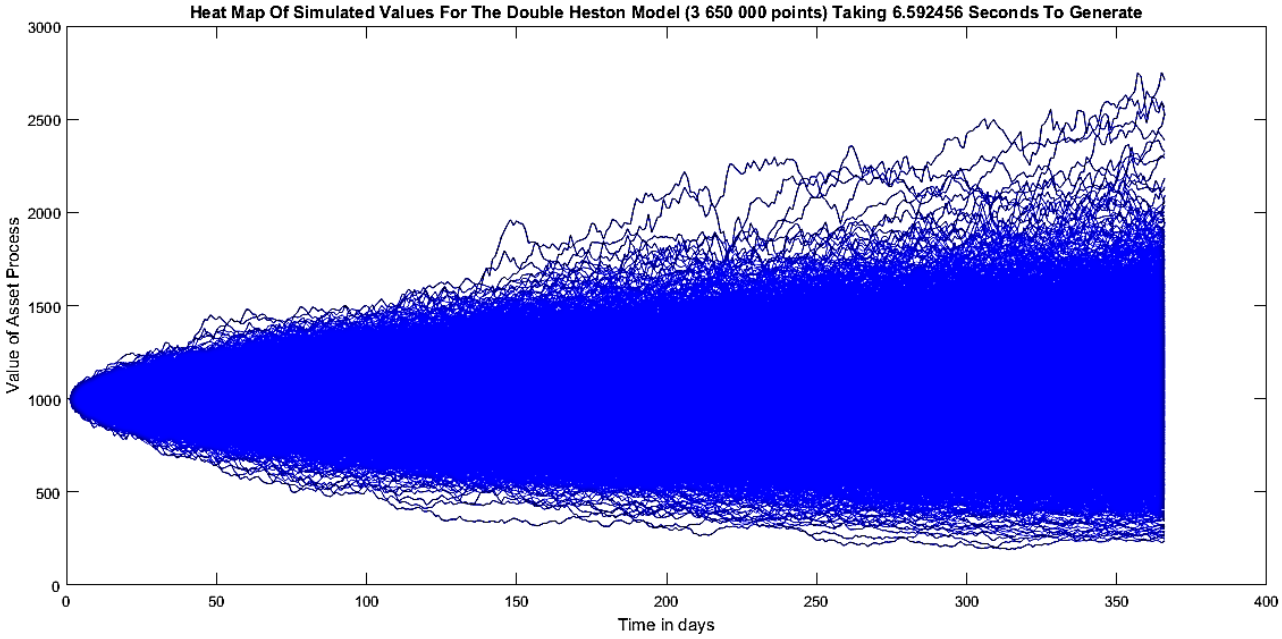

Figure C.1.1: Heatmap of the Double Heston Model produced in 6.6 seconds using C++ acceleration.

Continuing with the above example, the second computational technique is to use parallel computing in order to accelerate the code. The idea here is use multiple CPU cores to do the computations simultaneously by spreading independent and repetitive tasks across the cores. For our purposes we will use what is called 2 "workers" that are dedicated to parallel processing. Essentially this is similar to having 2 MATLAB® programs open, each running on a different core on the computer. The result can be seen from this code segment:

```
y = zeros(366,5000,2);
tic
parfor ii = 1:2
y(:,:,ii) = simulateDH_mex(1/365:1/365:1,5000);
end
toc
```

```
Elapsed time is 4.961427 seconds.
```

The *parfor* expression divides the simulation task into two groups of 5000 simulations each and takes on average 4.921982 seconds (while the time given in the above code segment is for a single run). On the face of it around 1.7 seconds are saved.

Now in order to understand how the parallel computing technique can further speed up the computational time we break up the 10000 simulations into ten blocks of 2500 simulations and running the *parfor* loop above (here ii runs from 1 to 4), again using only two workers. The result is that 10000 simulations are computed on average in 4.557868 seconds, which gives an additional 0.37 seconds saved.

This concludes this short section on computational accelerations, please see the supporting material and code for additional detail.

Appendix D: Investigation into the accuracy and speed of new density recovery technique

The purpose of this investigation is to determine how accurate and fast the new density recovery technique is. Recall that this is based on the formula:

$$F'(x) = f(x) = \frac{1}{\pi}\int_0^{\infty} e^{-iux}\phi(u)du = \frac{1}{\pi}\int_0^{\infty} f(u)\cos(ux) + g(u)\sin(ux)\,du.$$

The methodology employed here is simple. The true density function will be given by evaluating the integral expression using advanced quadrature techniques. This will be compared to the resulting density function recovered from using the FFT technique and second numerical routine given in section 3.3. in chapter 3.

The stochastic volatility model used here is the Double Heston Model, with parameters as given in Appendix F. For reference, the true log-price density function is plotted here:

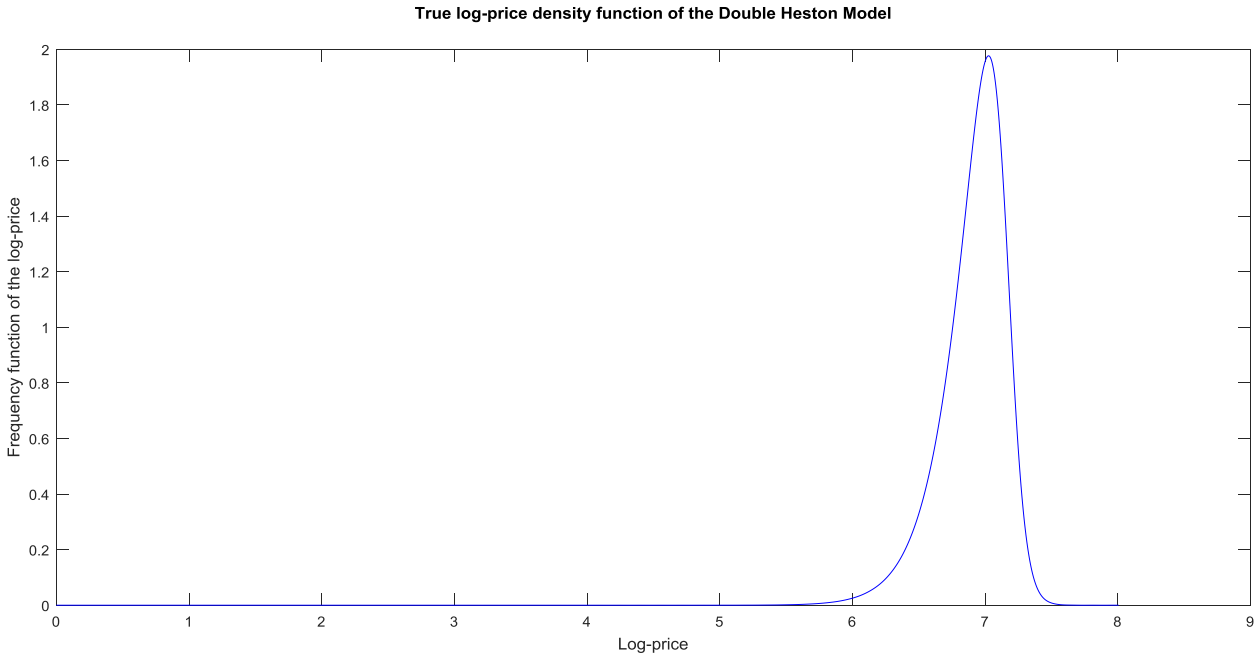

Figure D.1.1: True log-price density function of the Double Heston Model

The results below can be replicated by investigating the scripts in the supporting material. The following table summarises the results and time taken to compute the density function given a certain number of points on the support. The errors relative to the true density function are given as well. The errors are calculated as the mean-square error. The cumulative absolute error is given in brackets (defined as the sum of the absolute value of the differences).

For practicality, the separate tables below are treated as one unit and are collectively referred to as table D.1.1

Method:	Nr of Points	Error: MSE [ABS]	Time (seconds)
Advanced Quadrature	1024	n/a	0.338064035
FFT	1024	0.7222 [7.4886]	0.001856859
Author's	1024	0.0013 [0.0237]	0.006670719

Method:	Nr of Points	Error: MSE [ABS]	Time (seconds)
Advanced Quadrature	4096	n/a	0.804112012
FFT	4096	0.3695 [7.9214]	0.005523998
Author's	4096	0.0027 [0.0947]	0.026403401

Method:	Nr of Points	Error: MSE [ABS]	Time (seconds)
Advanced Quadrature	8192	n/a	1.356300255
FFT	8192	0.2830 [8.6610]	0.011686292
Author's	8192	0.0038 [0.1894]	0.047329383

Method:	Nr of Points	Error: MSE [ABS]	Time (seconds)
Advanced Quadrature	16384	n/a	2.395839830
FFT	16384	0.2556 [10.6017]	0.038759962
Author's	16384	0.0054 [0.3789]	0.095060612

Method:	Nr of Points	Error: MSE [ABS]	Time (seconds)
Advanced Quadrature	32768	n/a	4.851768240
FFT	32768	0.2912 [15.8236]	0.067234567
Author's	32768	0.0076 [0.7577]	0.175979252

Table D.1.1: Results on the various density function recovery techniques measuring only the core calculation engines.

It is known that the FFT implementation works optimally when the number of points is a power of 2, i.e. $32768 = 2^{15}$. The FFT algorithm can actually run slower when this isn't the case, although the resulting times are still faster than the other two techniques given.

Examining the numbers in the table suggest that the FFT technique is superior in terms of the speed obtained when performing the calculations. This is due to the fact that the slightly more expensive matrix multiplications performed by using the Author's method is avoided in the FFT implementation. It is also noteworthy that FFT method generates comparatively larger errors when compared to the true density function.

The observation is interesting. The trade-off between speed and accuracy is clearly visible here and thus a recommendation in favour of either method, as an alternative to the expensive calculations of calculating the true density, depends on the purpose of performing the calculations. Although another criterion can be applied.

The times obtained in the above table D.1.1 measures only the raw computational times taken to calculate the density function. In other words the time taken to run the single lines of code that perform the actual numerical inversion are measured in the above table (*core calculation engine*). The full computational time taken to prepare the data, define anonymous functions, increasing and shrinking the grids of the FFT, and performing other operations that are required to compute the density function has been left out. Including these times, yields:

Method:	Nr of Points	Time (seconds)
Advanced Quadrature	1024	0.337623504
FFT	1024	0.033199648
Author's	1024	0.008768765

Method:	Nr of Points	Time (seconds)
Advanced Quadrature	4096	0.804865176
FFT	4096	0.162561234
Author's	4096	0.031473448

Method:	Nr of Points	Time (seconds)
Advanced Quadrature	8192	**1.341053799**
FFT	8192	0.245679763
Author's	8192	0.057718163

Method:	Nr of Points	Time (seconds)
Advanced Quadrature	16384	2.483607217
FFT	16384	0.449365455
Author's	16384	0.098632224

Method:	Nr of Points	Time (seconds)
Advanced Quadrature	32768	**4.737152974**
FFT	32768	0.827839467
Author's	32768	0.209786406

Table D.1.2: Results on the various density function recovery techniques measuring the entire calculation engines.

Note: All of the times above have been computed using MATLAB® 's tic/toc functions and generally by performing a series of runs one can obtain an average computational time that should be taken as representative of the true computational times.

Of all the results in the tables given, only the Advanced Quadrature exhibits the anomaly where the computational time taken employed by using the core calculations is longer than the time taken to run the entire function. Perhaps this is not as unexpected seeing as the relative differences between the times reported in both sets of tables for the Advanced Quadrature method is the smallest. This is also expected as no expensive operations are performed in preparation to passing the arguments into the core calculation engine.

The interesting conclusion however is that it takes computationally longer to prepare the data to use the FFT method. This does not mean that the FFT method is computationally slower than the Author's as the given results could depend on the specific method the FFT is implemented. The correct conclusion is that given a choice between the Author's method and the FFT *as implemented* in the supporting material, the computationally less expensive option is to use the Author's method.

A last issue to investigate is why the quadrature technique works with the large time step given ($h = 0.25$). This is probably the reason why the code runs fast. This step size works for all the stochastic volatility models given here. Decreasing the step size, i.e. smaller h, increases the computational time taken but only yields a small improvement in accuracy (using MSE or ABS).

Firstly recall the function that is to be integrated by using Simpson's Rule:

$$\frac{1}{\pi}\int_0^{\infty} f(u)\cos(ux) + g(u)\sin(ux)\,du.$$

Now if we vary the support (x-values) and vary the values of u we can obtain an "integration

surface" by evaluating the given function $z = H(u,x) = f(u)\cos(ux) + g(u)\sin(ux)$ and making a three-dimensional plot. Clearly then, by fixing a value for x, the corresponding u-z plane represents a function that is to be integrated and yields as integrant the value for the log-price density associated with the fixed value x. A few plots are given here to obtain insight into this surface. This surface was generated from the Double Heston Model, with parameters as given in Appendix E.

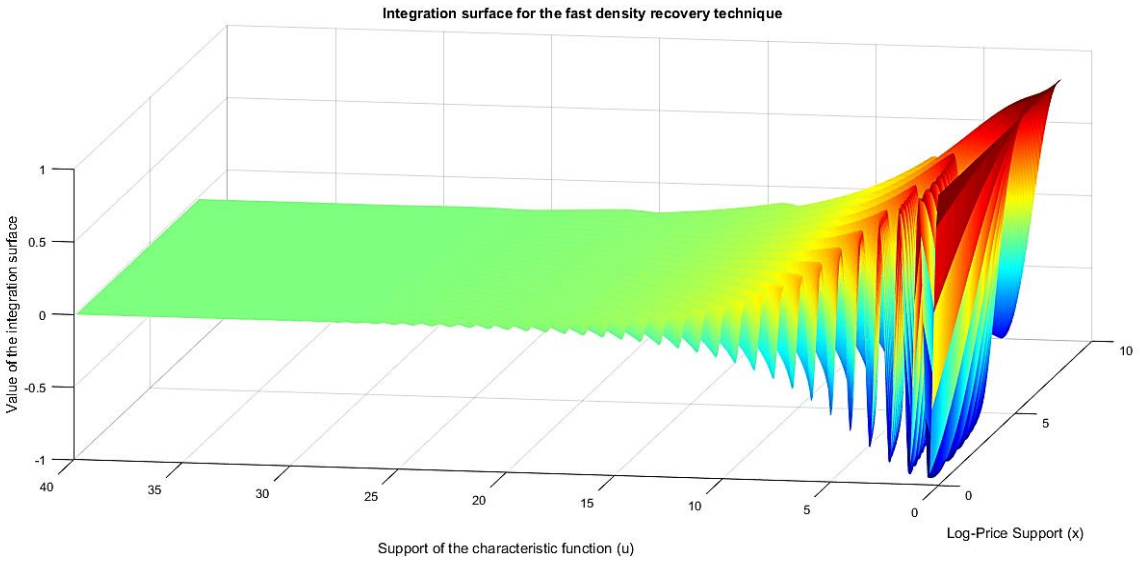

Figure D.1.2: Left-hand view of the integration surface for the fast density recovery technique

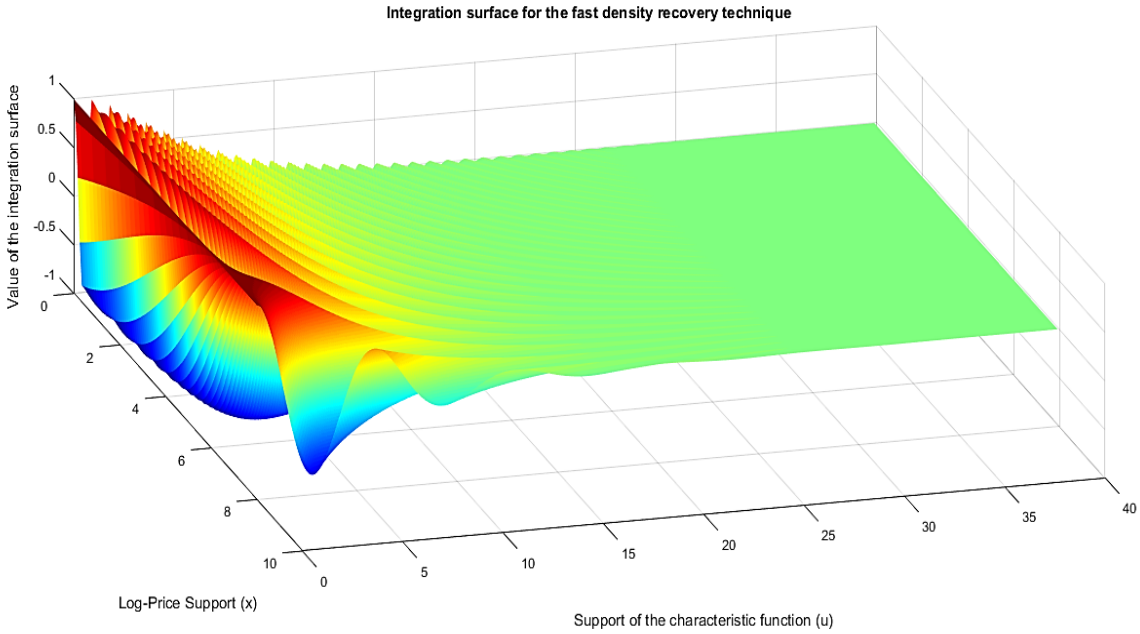

Figure D.1.3: Right-hand view of the integration surface for the fast density recovery technique

The answer to understanding why the integration works well can be seen from fixing certain x-values and taking u-z plane slices. Taking these planar slices will yield insight into why Simpson's rule works particularly well here.

The real value of characteristic function of the model can almost be obtained by taking small values, near zero, on the log-price support. This is given in a two-dimensional plot below. A far more

smoothed graph can be obtained by taking values near the upper bounds of the log-price support. Again such a graph is plotted in the figures below.

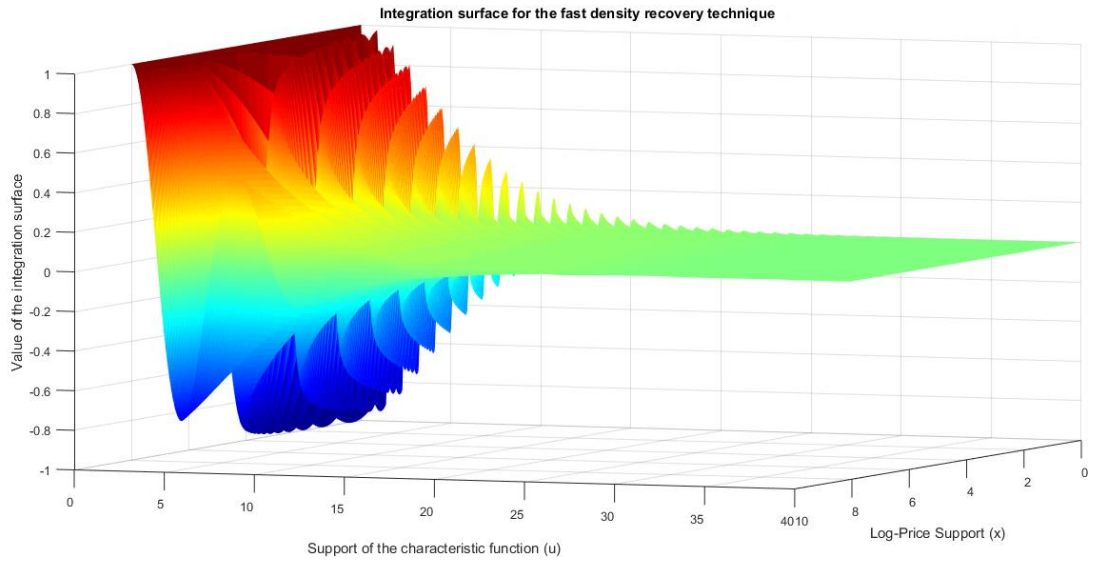

Figure D.1.4: Front view of the integration surface for the fast density recovery technique

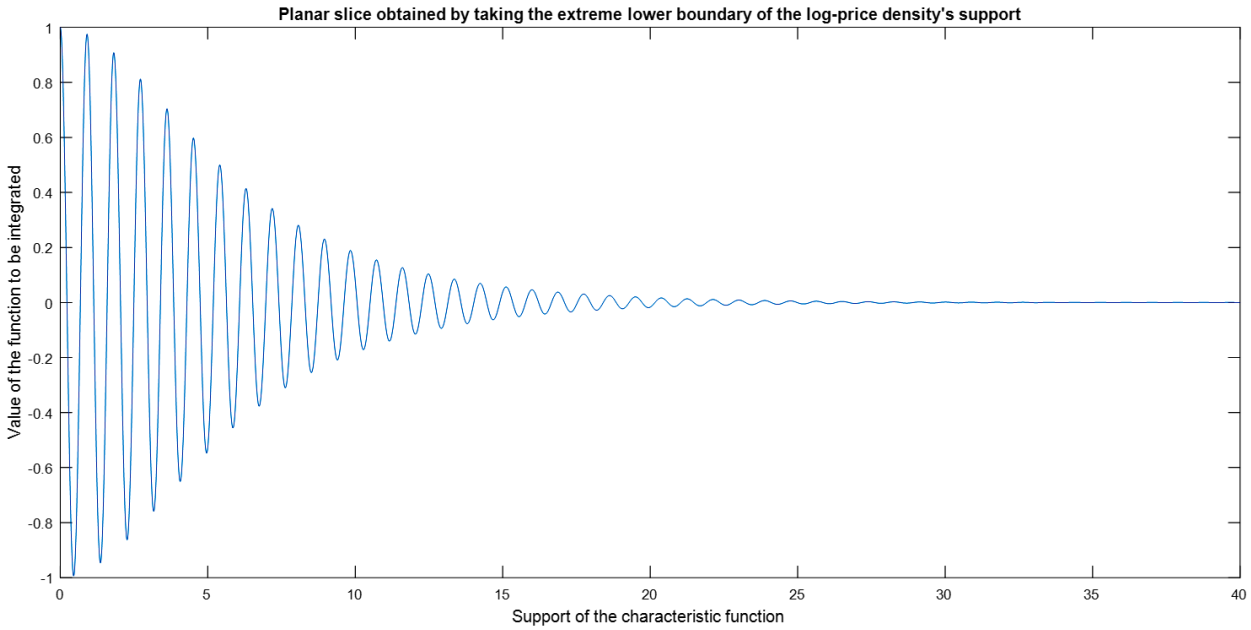

Figure D.1.5: Planar slice obtained by taking the extreme lower boundary of the log-price density's support

The (net) area of this 2-dimensional plot is approximately 0.000012811 as the positive values above the x-axis cancel out the negative parts underneath. Consequently quadrature rules such as the Simpson rule which approximate the function ($H(u,x)\ around\ x = 0$) has positive and negative errors that cancel out in the approximation of the integral. This allows us to use larger step sizes while still maintaining sufficient accuracy.

For the plot other 2-dimensional plot (below) we can see that the function is far smoother. Here the smooth function is well-approximated by the quadratic approximations implicit in the Simpson's method. Any errors in the mis-fitting incurred usually cancel out as parts of the fitted polynomial is

usually above and below the actual curve. For sufficiently smooth functions the accuracy is extremely high.

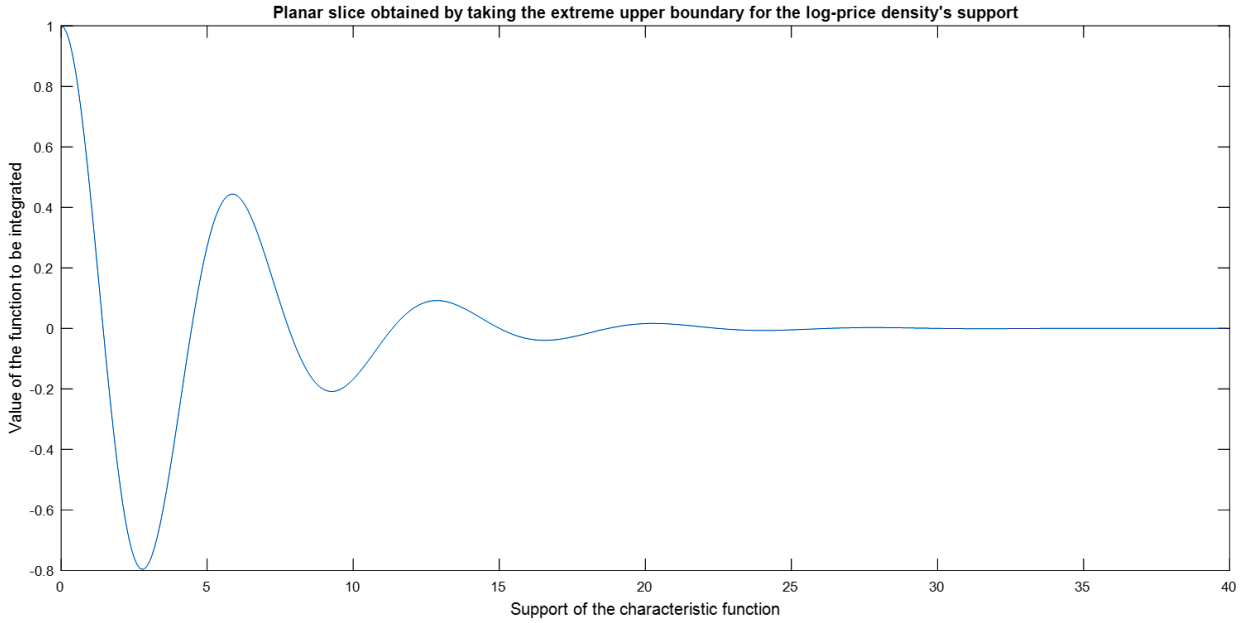

Figure D.1.6: Planar slice obtained by taking the extreme upper boundary of the log-price density's support

Lastly, for intermediate planar slices on this integration surface that yield non-zero areas the functions are actually smoother than that given in the above figure. Having even smoother functions ensures that Simpson's rule functions optimally and reduces the integration error. For reference, the planar slice corresponding to a point around peak of the log-price density is given here, note how smooth the function is:

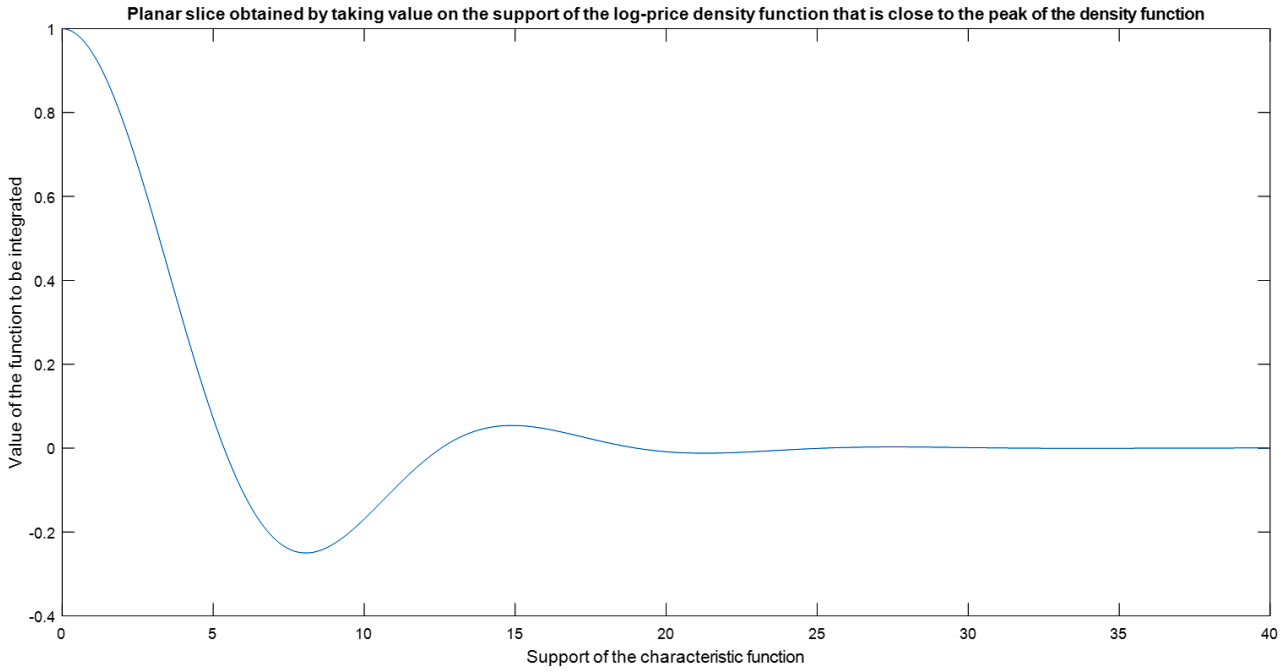

Figure D.1.7: Planar slice obtained by taking values on the support of the log-price density function that are close to the peak.

The step size of $h = 0.25$ is entirely arbitrary, but in numerical experiments the Author has found this value to work quite well. Using larger step sizes work well, except at very small values on the support

where the approximation yields a negative density function (which is obviously not correct). Using $h = 0.25$ does ensure that the resulting density integrates to one and this can easily be verified by numerically integrating the resulting density.

Admittedly, using the step size of $h = 0.25$ does yield very small negative values of the density function (having values only starting in the 6^{th} decimal place onwards). These values are usually negligible and can be found on the extreme ends of the log-price support. Now, using the resulting cumulative density function to sample values yields virtually no problems when computing option values (i.e. the accuracy is very decent). This can be checked in the supporting material with the instructions that the Author gives.

Thus, in summary, we find that using the fast density recovery technique does yield the correct density function, has lower errors that the FFT transform, runs faster than FFT function when considering all the code involved and can yield small negative values for the log-price density function at the extrema of its support. The conclusion obtained from this small investigation is that this new technique does have advantages and disadvantages relative to the traditional FFT, nonetheless the results indicate that it is viable alternative to the FFT technique.

Appendix E: Implied volatility surfaces for stochastic volatility models: Illustrative graphs and informal comparisons

Traditionally the approach, and the one advocated in this thesis, is that the practitioner pricing some exotic option should specify the underlying dynamics of the asset price and the unobserved volatility process. The implied volatility surface is then recovered by using formula (1.4) or some other suitable BS pricing equation (depending on the option payoff) by inputting the output prices of the stochastic volatility model into the BS pricing equation and plotting the various values of σ against both K & $(T-t)$. This approach has been questioned by (Carr and Wu, 2016). They suggest that the implied volatility surface should be modelled directly and proceed to suggest a simple model to do so.

However, the Author remains unconvinced of this approach. In fact concerns of this practice have already been addressed. Echoing the words of (Figlewski, 2010):

"The smile is stable enough over short time intervals that traders use the BS Model anyway, by inputting different volatilities for different options according to their moneyness. This jury-rigged procedure, known as 'practitioner Black-Scholes,' is an understandable strategy for traders, who need some way to impose pricing consistency across a broad range of related financial instruments, and don't care particularly about theoretical consistency with academic models. This has led to extensive analysis of the shape and dynamic behaviour of volatility smiles, even though it is odd to begin with a model that is visibly inconsistent with the empirical data and hope to improve it by modelling the behaviour of the inconsistency."

The question that we must investigate is how each of the 3 models selected produce the implied volatility surface. For each of the 3 models a set of parameters is specified, the pricing method, a table of call option prices across various strikes and maturities and finally a graph of the implied volatility surface. The figures here are given to show the dynamics of each model and not so much as to how they compare. A formal set of comparisons between all the models is given in Chapter 4 on different grounds.

All parameters and graphs below are original and in each subsection the scripts to reproduce all the results are referred to for the reader that wishes to experiment with the parameters and seeing the resulting effects. Note that the scripts can be found in the supporting material to this thesis. The material here can thus be seen as supplementary to the discussions given in this thesis.

Double Heston model's implied volatility surface

The following parameters will be used (note that $S(0) = 1000$) for the Double Heston Model:

Parameter	$\log(S_0)$	$v_1(0)$	$v_2(0)$	a_1	a_2	b_1	b_2	σ_1	σ_2
Value	6.9077	225%%	220%%	0.04	0.03	1	0.8	0.35	0.3
Parameter	ρ_1	ρ_2	r						
Value	-0.7	-0.7	0.03						

Table E.1.1: Parameters used to demonstrate implied volatility surface of the Double Heston Model

Main pricing method: COS method with constant interest rates. Note that the time to maturity is

specified as a vector of times starting at time 1 and increasing in intervals of $1/2^{th}$ of a year for 10 years.

The characteristic function and corresponding density function at time 1 is given below for reference, although for the other stochastic volatility models these will be omitted in the interest of saving space. They can however be found in the supporting material.

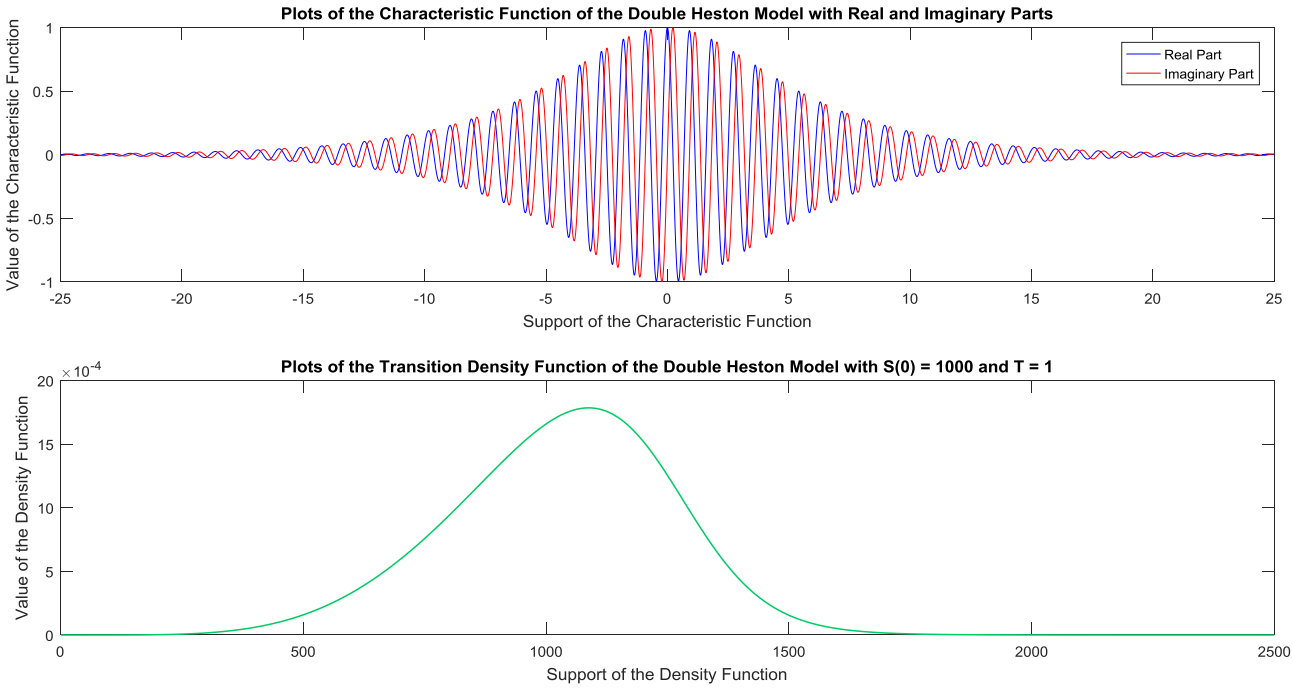

Figure E.1.1: Plots of the characteristic and transition density functions of the Double Heston Model

A table of Call options with differing maturities (note that not all results are given):

Strike/Time	1	2	3	4	5	6	7	8	9	10
K = 500	516.42	535.41	554.99	573.91	591.83	608.58	623.67	636.34	645.84	651.69
K = 600	421.76	448.01	473.64	497.69	520.06	540.73	559.33	575.19	587.59	596.11
K = 700	330.36	365.31	397.36	426.55	453.24	477.65	499.56	518.38	533.48	544.47
K = 800	245.06	288.92	327.14	361.11	391.74	419.53	444.42	465.90	483.42	496.62
K = 900	168.58	220.48	263.84	301.81	335.77	366.43	393.86	417.65	437.27	452.42
K = 1000	104.85	161.35	208.11	248.93	285.40	318.32	347.80	373.49	394.87	411.66
K = 1100	57.51	112.54	160.29	202.56	240.59	275.06	306.07	333.22	356.01	374.15
K = 1200	27.13	74.47	120.41	162.58	201.16	236.47	268.45	296.64	320.49	339.69
K =1300	10.72	46.63	88.18	128.72	166.85	202.30	234.72	263.53	288.10	308.08
K = 1400	3.56	27.65	62.95	100.55	137.32	172.24	204.61	233.65	258.62	279.12
K = 1500	1.25	15.59	43.86	77.52	112.17	145.99	177.86	206.76	231.84	252.61

Table E.1.2: Call option prices of differing maturities for the Double Heston Model.

The implied volatility surface for the given parameters can be constructed:

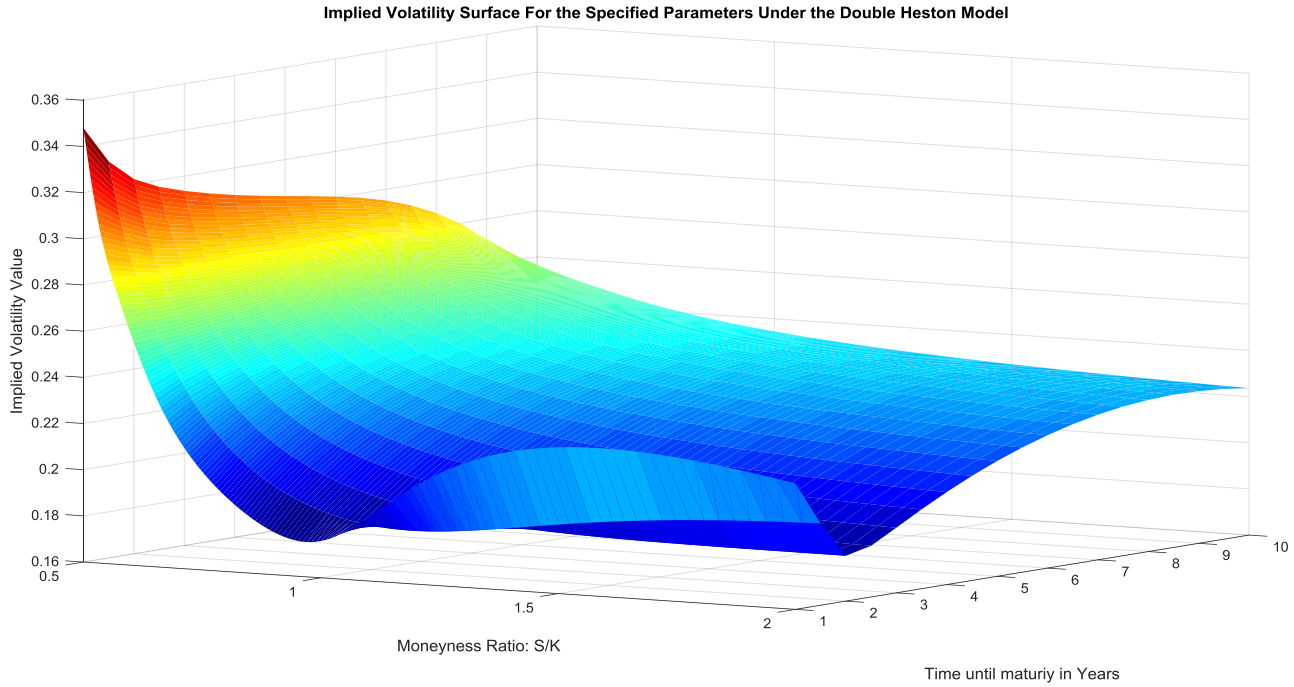

Figure E.1.2: Implied volatility surface for the specified parameters under the Double Heston Model.

Note that by changing the correlation coefficients from being strongly negative to being strongly positive the implied volatility surface can produce a pronounced forward volatility skew and the volatility for in-the-money Call options reduces. The strong positive correlations induce the characteristics of a commodity market on the equity model. Hence the upside potential for the stock has increased resulting in out-of-the-money Call options' prices to increase.

All of the above parameters were carefully chosen, as they will be for most of the other models. It is important to realise that there is a fine "inter-play" between all the parameters and in general conditions such as the Feller condition need to be obeyed to avoid nonsensical results. This is important when calibrating options as negative option prices can arise if the underlying parameters give characteristic functions that do not decay sufficiently fast to zero.

H2-CIR model's implied volatility surface

The following parameters will be used (note that $S(0) = 1000$) for the H2-CIR Model:

Parameter	$\log(S_0)$	$V(0)$	$r(0)$	κ	Θ_V	ν	λ	Θ_r	η
Value	6.9077	0.1	0.02	1	0.04	0.35	0.03	0.02	0.05
Parameter	$\rho_{x,V}$	$\rho_{x,r}$							
Value	-0.8	0.6							

Table E.1.3: Parameters used to demonstrate implied volatility surface of the H2-CIR Model.

Main pricing method: COS method with Stochastic interest rates. Note that the time to maturity is specified as a vector of times starting at time 1 and increasing in intervals of $1/2^{th}$ of a year for 10 years. The characteristic function and corresponding density function at time 1 is given below. Note

that one can obtain zero-coupon bond prices from the characteristic function and derive a single constant risk-free interest rate to use in the BS formula to recover the implied volatility surface.

Merely deriving a set of spot yield rates and averaging them is not a serious option to determine the BS risk-free rate. The Author suggests working out the Internal Rate of Return (IRR) from the zero-coupon bond prices which should provide a risk-free rate that is constant across time. Applying different risk-free rates to the BS equation implies using different BS Model to evaluate the implied volatility surface. This is a contradiction in the definition of the implied volatility surface.

Table of Call options with differing maturities (note that not all results are given):

Strike/Time	1	2	3	4	5	6	7	8	9	10
K = 500	531.15	545.65	558.36	570.16	581.34	592.08	602.44	612.47	622.12	631.34
K = 600	440.72	460.59	476.93	491.82	505.82	519.16	531.96	544.26	556.06	567.30
K = 700	353.99	379.76	399.82	417.85	434.66	450.58	465.76	480.26	494.11	507.27
K = 800	273.08	304.33	327.96	349.00	368.49	386.83	404.21	420.74	436.48	451.40
K = 900	200.11	235.71	262.32	285.99	307.83	328.27	347.58	365.89	383.27	399.74
K = 1000	137.03	175.07	203.78	229.42	253.07	275.19	296.04	315.79	334.54	352.29
K = 1100	85.95	123.48	153.04	179.74	204.50	227.72	249.66	270.46	290.23	308.99
K = 1200	48.34	81.88	110.56	137.20	162.24	185.91	208.40	229.84	250.27	269.72
K =1300	23.80	50.59	76.49	101.80	126.20	149.64	172.15	193.77	214.50	234.32
K = 1400	9.99	28.90	50.50	73.28	96.17	118.70	140.70	162.06	182.73	202.61
K = 1500	3.57	15.17	31.72	51.11	71.74	92.76	113.75	134.46	154.72	174.39

Table E.1.4: Parameters used to demonstrate implied volatility surface of the H2-CIR Model.

The implied volatility surface for the given parameters can be constructed as seen on the next page.

Some remarks on the H2-CIR Model versus the H2-HW Model can also be made here. The table below gives the zero-coupon prices and the constant interest rate derived from each of these models. The parameters used for the H2-HW Model are exactly those given in the table for the H2-CIR Model. In this way a comparison can be made between the effect of the different interest rate processes.

The table excludes half-yearly zero-coupon bond prices, but the constant (annual) interest rate calculated includes half-yearly zero-coupon bond prices. Note: Zero-coupon bond prices quoted to two decimals.

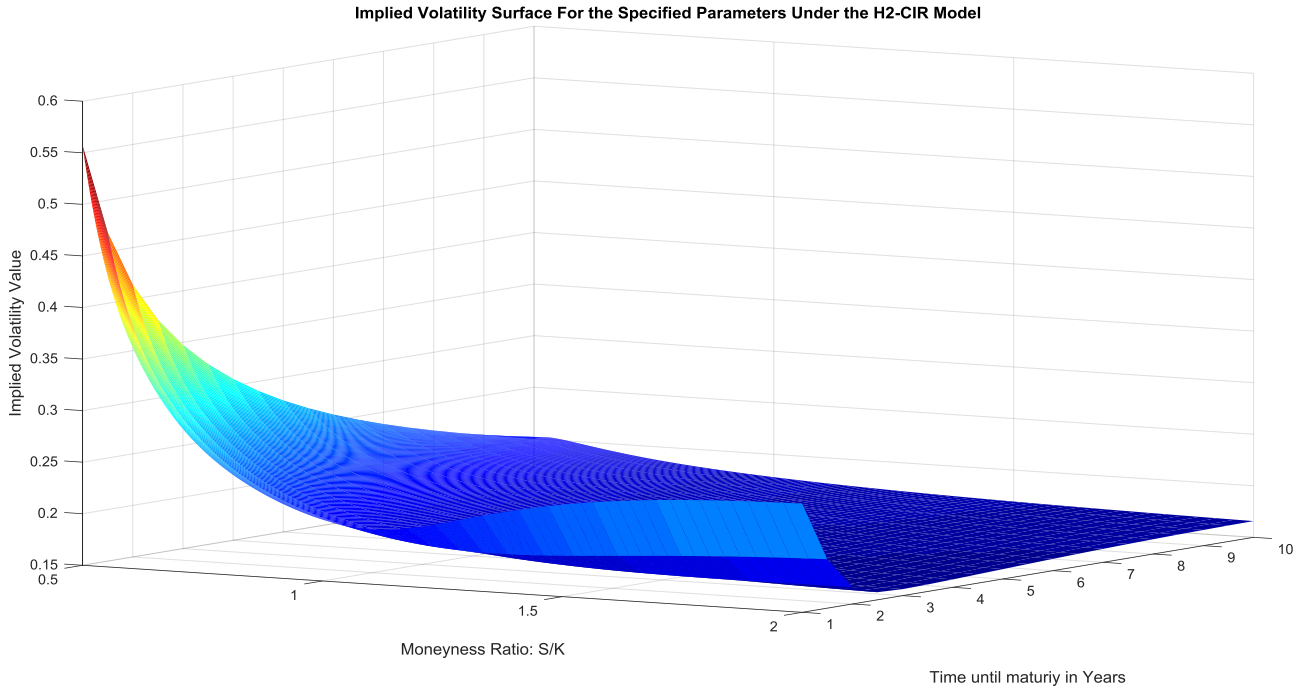

Figure E.1.3: Implied volatility surface for the H2-CIR Model

H2-HW Model, Implied Constant Interest Rate = 2.0495%										
Time:	1	2	3	4	5	6	7	8	9	10
Value:	0.98	0.96	0.94	0.92	0.90	0.88	0.87	0.85	0.83	0.82
H2-CIR Model, Implied Constant Interest Rate = 2.7764%										
Time:	1	2	3	4	5	6	7	8	9	10
Value:	0.94	0.92	0.90	0.89	0.87	0.85	0.83	0.82	0.80	0.79

Table E.1.5: Differences in zero coupon bond prices between the H2-HW and H2-CIR Model with the same parameters

The differences in the above table are attributed to the fact that the Hull-White short rate process can go negative while the CIR short rate process is bounded below by zero. Essentially this means that on average, for the *same model parameters*, the CIR process produces higher interest rate paths and hence zero-coupon bond prices that are lower than those given by the Hull-White process.

Additionally, since the CIR process gives higher interest rate paths on average, the rate at which the underlying grows is higher in the beginning which results in higher Call option prices. As the time horizon increases this higher growth in the underlying is countered with a higher discount rate, as reflected in the discounted density function, which essentially reduces Call option prices.

The effect of these dynamics is that when pricing Call options under these two models, using the same parameters for the driving interest rate process, the Call options should have values that are close in value. However, the question can be asked what the effect would be on the resulting implied volatility surfaces between the two models?

The implied volatility surfaces of the two models should be closer in value as the duration increases. The reason for this that the lower H2-CIR Call prices at longer durations are passed into a BS Model which calculates the implied volatility at a higher interest rate. The result is that the implied volatility, which should be lowered by lower Call option prices, is increased by a higher risk-free rate. The

opposite arguments with higher Call prices and lower risk-free rates apply to the H2-HW Model. Hence the implied volatility is calculated by two offsetting factors and is hence more stable which results in the two implied volatility surfaces not differing significantly.

4/2 Model's implied volatility surface

There are a few technical notes as to calculating the characteristic function of this model quickly and efficiently. These are the Author's own recommendations.

The confluent hypergeometric function in (2.184) can be calculated quickly, see Abramowitz and Stegun (1972), page 505, by using the formula:

$$ {}_1F_1(a,b,z) = \frac{\Gamma(b)\Gamma(a)}{\Gamma(b-a)} \int_0^1 \exp(z*u) * u^{(a-1)} * (1-u)^{(b-a-1)}\, du, $$

This formula does however require that $real(b) - real(a) > 0$ and that $real(a) > 0$ for which there is no guarantee (it is difficult to establish whether these conditions hold analytically). Under these assumptions the function may be computed as given. Unfortunately the given function is not always robust and at times it might be desirable to move away from the restrictive assumptions imposed. Hence the computationally slower, but more robust function, *hypergeom*, a built in function in MATLAB®, will be used to calculate the characteristic function.

Technical note*: Calling the function *hypergeom* using for-loops is computationally intensive in MATLAB®. Using the Symbolic Toolbox allows us to define this function using the extended computer algebra system only once and hence reduces computational times significantly. This is done by defining a *symfun* and then running the for-loop on this function.

Lastly, the Gamma function in (2.184) may take complex arguments. Unfortunately MATLAB®does not implement this function and hence the Author's *complexgamma* function is built around the approximation by Abramowitz and Stegun (1972), page 257, for some complex argument z:

$$ x = (z-0.5) * log(z) - z + 0.5 * log(2\pi) + \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5} - \frac{1}{1680z^7} + \frac{1}{1188z^9} - \frac{691}{360360z^{11}} + \frac{1}{156z^{13}} $$

$$ \therefore \Gamma(z) = \exp(x). $$

Now that all the technical details were attended to, the following parameters will be used (note that $S(0) = 1000$) for the 4/2 Model's implementation:

Parameter	$\log(S_0)$	a	b	κ	θ	σ	V(0)	r	ρ
Value	6.9077	0.1	0.05	1.8	0.3	0.4	400%	0.03	-0.9

Table E.1.6: Parameters used to demonstrate implied volatility surface of the 4/2 Model

Main pricing method: COS method with constant interest rates. Note that the time to maturity is specified as a vector of times starting at time 1 and increasing in intervals of 1/2th of a year for 10 years. The characteristic function and corresponding density function at time 1 is given below:

Table of Call options with differing maturities (note that not all results are given):

Strike/Time	1	2	3	4	5	6	7	8	9	10
K = 500	523.45	532.38	544.55	557.88	571.29	584.40	597.06	609.13	620.51	631.06
K = 600	426.35	438.31	453.70	470.28	486.86	503.02	518.58	533.40	547.37	560.40
K = 700	329.57	345.75	365.45	385.98	406.13	425.51	444.01	461.54	478.05	493.44
K = 800	235.57	258.70	283.61	308.22	331.76	354.09	375.21	395.14	413.85	431.30
K = 900	152.43	183.06	212.27	239.95	265.98	290.47	313.55	335.28	355.67	374.73
K = 1000	89.25	123.14	153.94	182.81	209.93	235.50	259.64	282.44	303.91	324.03
K = 1100	48.23	79.56	108.87	136.91	163.66	189.18	213.49	236.62	258.53	279.19
K = 1200	24.76	49.95	75.58	101.23	126.43	150.93	174.63	197.43	219.22	239.91
K =1300	12.38	30.80	51.82	74.20	97.04	119.85	142.33	164.26	185.47	205.79
K = 1400	6.11	18.81	35.27	54.08	74.18	94.87	115.73	136.42	156.69	176.32
K = 1500	3.00	11.46	23.92	39.31	56.58	74.98	93.98	113.19	132.27	150.96

Table E.1.7: Call option prices for the 4/2 Model

The implied volatility surface for the given parameters can be constructed:

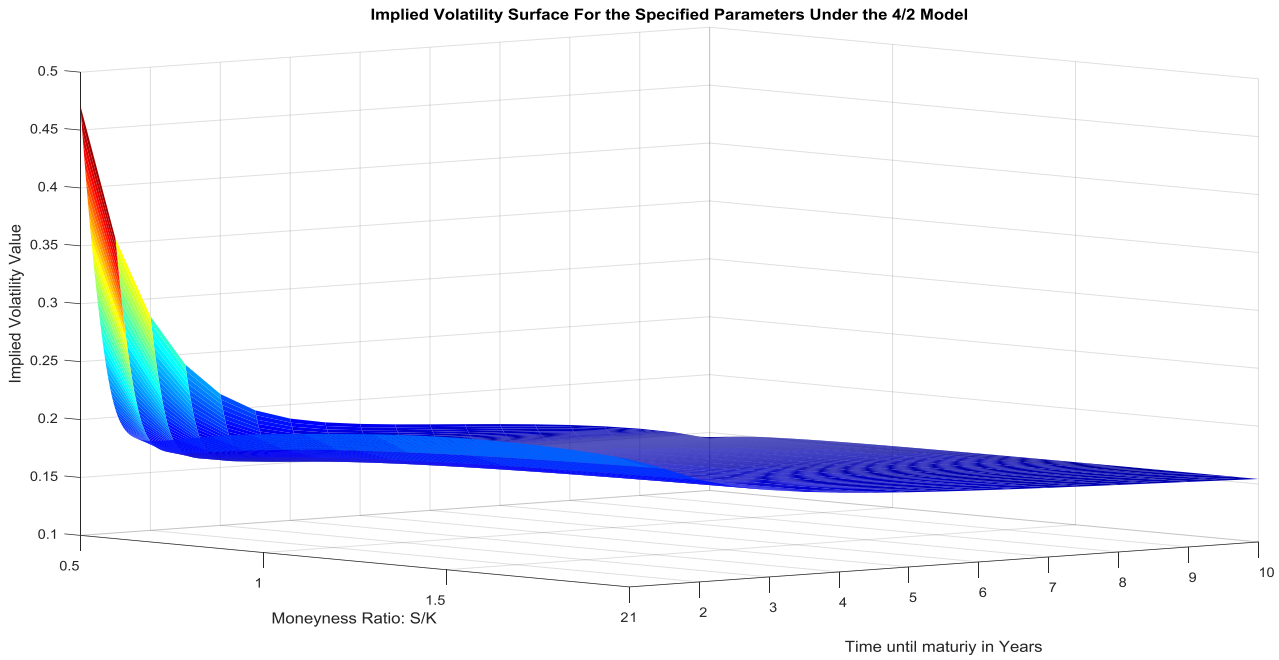

Figure E.1.4: Implied volatility surface for the 4/2 Model

Appendix F: Calibration of stochastic volatility models using Differential Evolution

"To calibrate does not mean to understand. A perfect fit is not a theory"

- Unknown

(See Fouque et al (2011), page 85, for reference to the origin of this quote.)

Calibration is the computational procedure to ascribing parameter values to some asset price (and hence stochastic volatility) model by fitting the model to financial data, which may be current or past data. The entire idea with calibration is that the model parameters should be consistent with financial data in order to improve risk management, make forecasts, pricing OTC derivatives, etc. In literature this is done by fitting the model to today's option data, instead of using time series data. Once a calibration procedure is done, one has to typically re-calibrate a model, perhaps every day depending on the time horizon and use of the model, to the then current options data.

This now brings in the question of parameter stability and the use of time dependent parameters. A discussion or investigation into these specific issues is not done in this thesis and thus an underlying assumption behind the work done here is that any appropriately obtained parameter value is valid up to a time duration of 5 years. This is perhaps justified by considering the calibration of parameters to deeply out- and in-the-money option prices. The reason for doing this is simple: Those prices are unlikely to move in a short period of time and thus recalibration of model parameters is unnecessary.

In this chapter 2 calibration techniques are given and only one of these techniques is used for the calibration of all the models in table 3.1. A set of options data is given at the end, to which the each of the 3 stochastic volatility models are calibrated and the parameter values given. This set of data and the parameter values are then used in the following chapters to investigate certain research hypothesis, outlined in the next chapter.

There are a number of other calibration techniques not considered here, however the techniques given below are powerful and are more than sufficient for our purposes here. Additionally, using one calibration technique for most of the models eliminates some bias when assessing and comparing the models against each other.

Background knowledge to calibration

Let ϵ be some suitably small positive number. In the calibration procedures considered here one tries to minimize the distance between model, V^{model}, and market prices, V^{market}, using some function such that the minimized distance is below ϵ or according to some other criteria. The set of parameters that produces the model prices is denoted by $\boldsymbol{x} = [x_1 \ldots., x_n]$. The functions that are used are called *objective functions* and typical examples include:

1.) Average absolute error:

$$f(\boldsymbol{x}) = \frac{1}{M}\sum_{j=1}^{M}\left|V_j^{model}(\boldsymbol{x}) - V_j^{market}\right|.$$

2.) Root-mean-square error:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{M}} \sqrt{\sum_{j=1}^{M} \left(V_j^{model}(\boldsymbol{x}) - V_j^{market}\right)^2}.$$

For our purposes the function f will be specified as a *nonlinear least squares minimization objective function*:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \coloneqq \frac{1}{2} \sum_{j=1}^{M} r_j(\boldsymbol{x})^2. \tag{F.1.1}$$

Here $r_j(\boldsymbol{x}) = V_j^{model}(\boldsymbol{x}) - V_j^{market}$ and if we denote the vector of these values by $R(\boldsymbol{x}) = (r_1, \dots, r_M)^T$, collectively forming a vector of residuals, then (F.1.1) can be written succinctly as

$$f(\boldsymbol{x}) = \frac{1}{2} R(\boldsymbol{x})^T R(\boldsymbol{x}). \tag{F.1.2}$$

Hence our goal will be to find a set of parameters $\boldsymbol{x}$ such that we have minimised (F.1.2) or such that when successive evaluations of the objective function have a difference within ϵ. There are many techniques to accomplish this task, but the methods must be distinguished between those that are locally convergent or globally convergent. In the former case the algorithm may be caught giving a value of $\boldsymbol{x}$ that corresponds to a local minimum instead of the global minimum. The first of the techniques below, the famous Levenberg-Marquardt calibration technique, generally provides solutions that are generally locally convergent, while the second of these is designed to find the global minimum.

Levenberg-Marquardt technique

This technique was originally introduced by Levenberg (1944) and an improvement was to the basic procedure by Marquardt (1963). The main problem with using this algorithm is that it is based on evaluating derivatives of the vector $R(\boldsymbol{x})$, which in the case of calibrating stochastic volatility models is an enormous undertaking. Typically this algorithm is much better suited to curve fitting problems, such as fitting a density function or fitting time dependent functions in specifying parameters. Nonetheless, it is an important technique in literature, and although it won't be used here, it summarised below:

The derivatives of the function $f(\boldsymbol{x})$ can be written using what is called a Jacobian Matrix:

$$J(\boldsymbol{x}) = \frac{\partial r_j}{\partial x_i}, 1 \leq j \leq m, 1 \leq i \leq n. \tag{F.1.3}$$

Generally the residuals, r_j, are nonlinear and thus we define

$$\nabla f(\boldsymbol{x}) = \sum_{j=1}^{m} r_j(\boldsymbol{x}) \nabla r_j(\boldsymbol{x}) = J(\boldsymbol{x})^T r(\boldsymbol{x}). \tag{F.1.4}$$

$$\nabla^2 f(\boldsymbol{x}) = J(\boldsymbol{x})^T J(\boldsymbol{x}) + \sum_{j=1}^{m} r_j(\boldsymbol{x}) \nabla^2 r_j(\boldsymbol{x}). \tag{F.1.5}$$

One would choose to approximate the residuals in (F.1.5) by a linear function, assuming that the residuals are relatively small, then (F.1.5) can be reduced to a simpler formula:

$$\nabla^2 f(\boldsymbol{x}) = J(\boldsymbol{x})^T J(\boldsymbol{x}). \tag{F.1.6}$$

In order to obtain a new parameter vector the following formula is typically used, assuming that there are Q iterative steps, i.e. $q = 1, \dots, Q \in \mathbb{N}$ and that $\lambda \in \mathbb{R}$:

$$\boldsymbol{x}_{q+1} = \boldsymbol{x}_q - \lambda \nabla f(\boldsymbol{x}_q). \tag{F.1.7}$$

Note that (F.1.7) suffers from convergence problems and hence can be improved by using the

following formula that uses both the information of the gradient of the surface and well as its curvature (or roughly its second derivatives):

$$\boldsymbol{x}_{q+1} = \boldsymbol{x}_q - \left(\nabla^2 f(\boldsymbol{x}_q)\right)^{-1} \nabla f(\boldsymbol{x}_q). \quad \text{(F. 1.8)}$$

Here the approximation in (F.1.6) is used to compute (F.1.8). Although (F.1.8) is an improvement on (F.1.7) there are still better improvements that can be made and this is where the insights of Levenberg (1944) come into play. He suggests that (F.1.8) be replaced by

$$\boldsymbol{x}_{q+1} = \boldsymbol{x}_q - (H + \lambda I)^{-1} \nabla f(\boldsymbol{x}_q). \quad \text{(F. 1.9)}$$

Here H is the Hessian matrix associated with $f(\boldsymbol{x})$, evaluated at $\boldsymbol{x}_q$ and I is the identity matrix. Consequently when implementing the above, one would compute (F.1.1) using the vector $\boldsymbol{x}_{q+1}$ and if the error goes down one can reduce the value of λ by some factor, say 5. The opposite is done if the error actually increases.

A problem with (F.1.9) is that if λ becomes large then the contribution of the Hessian matrix is dwarfed and we are essentially back to (F.1.7). Hence Marquardt (1963) suggested the following and final improvement:

$$\boldsymbol{x}_{q+1} = \boldsymbol{x}_q - \left(H + \lambda diag(H)\right)^{-1} \nabla f(\boldsymbol{x}_q). \quad \text{(F. 1.10)}$$

The $diag()$ notation is used to denote the diagonal along the Hessian matrix that replaces the identity matrix, i.e. the main diagonal of the Hessian matrix replaces the main diagonal of the identity matrix.

There are various updating rules to determine the value of λ as well as other advanced considerations such as gain ratios ρ_q that are not presented here. However (F.1.10) represents the very basic formulation of algorithm that would be used to determine the optimal parameter set. For details on further expansions and advance techniques related to this technique the reader can consult Fan & Pan (2006).

Reaching the Global Minimum: Using Differential Evolution to calibrate models

The method of Differential Evolution, or DE, was developed by originally by Storn and Price (1995) and below their DE1 scheme is presented here. Although the Author generally prefers to implement algorithms from scratch (or doing this in conjunction with some MATLAB® built in functions), the latest version of the free DeMat software for MATLAB® by Price et al (2005) is used for all DE problems. Details can be found in the appendix.

Now, DE is a very easy method to understand and is based on some principles from biological evolution. The algorithm is globally convergent and does not require the evaluation of derivatives. In order to explain the procedure an example will be used to guide us through the process which is an easier alternative than giving the full theory.

Example: How Differential Evolution Works

Firstly define the objective function that needs to be minimised as

$$f(\boldsymbol{x}) = \sum_{i=1}^{4} x_i^2. \quad \text{(F. 1.11)}$$

subject to $x_i \in (-10,10)$. The first step in the DE framework is to initialise a population of vectors, randomly, and label them. Hence for our purposes assume that there are 6 population vectors with corresponding values in (F.1.11):

$$\boldsymbol{x}_1 = [9, -4.5, 2,4] => f(\boldsymbol{x}_1) = 121.25$$

$$\boldsymbol{x}_2 = [3.8, -4, -3{,}1] => f(\boldsymbol{x}_2) = 40.44$$
$$\boldsymbol{x}_3 = [-7{,}0.2, 6.7, -3] => f(\boldsymbol{x}_3) = 102.93$$
$$\boldsymbol{x}_4 = [5, -9, 3.1{,}0] => f(\boldsymbol{x}_4) = 115.61$$
$$\boldsymbol{x}_5 = [2, -9\ 4, -8] => f(\boldsymbol{x}_5) = 165$$
$$\boldsymbol{x}_6 = [1, -1, 9, -9] => f(\boldsymbol{x}_6) = 164$$

Once the population is initialised, the next step is to select one of the above vectors at random, calling it the *target vector*, and three other vectors, where these three vectors will be used to derive a new vector, through a process known as *mutation*. Suppose that $\boldsymbol{x}_5$ is selected as a target vector and vectors $\boldsymbol{x}_1, \boldsymbol{x}_2$ & $\boldsymbol{x}_6$ are selected for mutation. A new mutated vector, $\boldsymbol{v}_1$, is constructed by:

$$\boldsymbol{v}_1 = \boldsymbol{x}_1 + F * (\boldsymbol{x}_2 - \boldsymbol{x}_6) \qquad \text{(F. 1.12)}$$

Here F is simply a scaling factor, typically taking values in the interval: $F \in [0{,}2]$. For argument's sake, let's choose $F = 0.1$ and hence $\boldsymbol{v}_1 = [9.28, -4.8, 0.8, 5]$.

The next step is to perform a *crossover* where for each component in the target vector $\boldsymbol{x}_5$ that specific component is either replaced by a corresponding component from $\boldsymbol{v}_1$ with some probability cp or kept as is with probability $1 - cp$. The crossover probability is kept fixed and uniform random numbers are drawn. If the uniform random number is below cp then the corresponding component is switched out. This is illustrated in the following steps:

1.) Set $cp = 0.6$.

2.) Generate a vector uniform random numbers: $\boldsymbol{u} = [0.67, 0.59, 0.91, 0.38]$.

3.) The 2nd and 4th components in the vector $\boldsymbol{u}$ are below 0.6 and hence the 2nd and 4th components in the vector $\boldsymbol{x}_5$ are replaced by the 1st and 4th components in $\boldsymbol{v}_1$.

4.) The resulting vector, called a *child* of $\boldsymbol{x}_5$ and $\boldsymbol{v}_1$, is denoted by $\boldsymbol{x}_5^* = [2, \mathbf{-4}, 4, \mathbf{5}]$. The entries of $\boldsymbol{x}_5$ that have been replaced are marked in bold red.

The last step of the DE process is known as *selection* and simply involves evaluating the objective function with the child vector $\boldsymbol{x}_5^*$ and comparing the two. In this case $f(\boldsymbol{x}_5^*) = 61 < f(\boldsymbol{x}_5) = 165$ and hence the child vector is selected and *replaces* the parent vector in the population:

$$\boldsymbol{x}_1 = [9, -4.5, 2{,}4] => f(\boldsymbol{x}_1) = 121.25$$
$$\boldsymbol{x}_2 = [3.8, -4, -3{,}1] => f(\boldsymbol{x}_2) = 40.44$$
$$\boldsymbol{x}_3 = [-7{,}0.2, 6.7, -3] => f(\boldsymbol{x}_3) = 102.93$$
$$\boldsymbol{x}_4 = [5, -9, 3.1{,}0] => f(\boldsymbol{x}_4) = 115.61$$
$$\boldsymbol{x}_5^* = [2, -4, 4, 5] => f(\boldsymbol{x}_5) = 61$$
$$\boldsymbol{x}_6 = [1, -1, 9, -9] => f(\boldsymbol{x}_6) = 164$$

This process is repeated M times, and every M iterations forms a *population-evolution*. The whole process is repeated for k population evolutions. At the end of each population-evolution a comparison is made to the objective function and if this is below ϵ for at least one i in the k^{th} population then the best ranking member, the i^{th} member, is chosen and the parameters are the values in side the vector $\boldsymbol{x}_i^k$. Per the suggestion of Storn & Price (1995), k is usually set to be a multiple of the number of population vectors, Np, i.e. $k \in (3 * Np, 10 * Np);\ k \in \mathbb{N}$. The vector with the lowest value when evaluated using the objective function is then taken as the solution to the minimisation problem in (F.1.1).

To handle problems that introduce some additional constraint on the parameters, for example the Feller condition when working with CIR processes, Lampinen's "rule" (Liu and Lampinen (2002)) can be used which simply modifies the selection step slightly to check whether the constraints are satisfied.

Basically if both the parent and child vectors in the selection step are admissible (satisfy the constraints) then the one with the smaller objective function is taken as the new population member.

If only the child vector satisfies the constraints then it is chosen. Lastly if both vectors are inadmissible then the vector with the smaller objective function evaluation is taken.

Note that a parent vector can be inadmissible, particularly when initialising the population, although one can initialise all vectors to be admissible from the beginning.

We are now in a position to calibrate the stochastic volatility models as described in Chapter 4.

Calibration of Stochastic Volatility Models in Chapter 4

This section supplements the investigation in Chapter 4, section 4.2. Additional details are given here as well as a post analysis on the Double Heston Model.
Now in order to calibrate the stochastic volatility models it is firstly necessary to define a vector of upper and lower parameter constraints. There are no guidelines as to choosing these, although through experimentation the Author has the following sets to be adequate:

Double Heston Model:
Vector of lower bounds for the parameters:

$$[0, 0, 0, 0, 0, 0, 0, 0, 0, -1, -1]$$

Vector of upper bounds for the parameters:

$$[0.1, 0.1, 0.1, 0.5, 0.5, 5, 5, 0.5, 0.5, 0, 0]$$

H2-CIR Model:
Vector of lower bounds for the parameters:

$$[0, 0, 0, 0, 0, 0, 0, 0, , -1,0]$$

Vector of upper bounds for the parameters:

$$[5, 5, 10, 10, 10, 1, 20, 10, 1, 1]$$

4/2 Model:
Vector of lower bounds for the parameters:

$$[0, 0, 0, 0, 0, 0.1, 0, -1,]$$

Vector of upper bounds for the parameters:

$$[1, 0.1, 10, 1, 1, 0.1, 1, 1]$$

For all the models the following global parameters were set, which are used in the calculation engine for each of the stochastic volatility models (this coincides naming conventions in the scripts given in the supporting material):

Value to reach of the objective function before the simulation terminates: 0.1

I_NP: Number of population members
I_NP = 10;

I_itermax: Maximum number of iterations (generations)
I_itermax = 110;

F_weight: DE-stepsize, factor by which parameter vector differences are scaled, i.e. the scaling factor $F \in [0,2]$
F_weight = 0.85;

F_CR: Crossover probabililty constant ex [0, 1]
F_CR = 1;

There parameters form part of a classical differential evolution strategy as described above. The approximate time taken to calibrate the models are as follows:

Double Heston Model: 3 min.

H2-CIR Model: 20 min

4/2 Model: 50-80 min

Post Analysis On The Double Heston Model:

As mentioned in section 4.2, the question is whether we can improve the performance of the Double Heston Model by introducing stricter criteria for the calibration process. Here the calibration is a success if the objective function evaluates to a values less than 0.01 (for the MSE). In order to achieve the calibration parameters are adjusted and other more advanced differential evolution techniques have been applied (see supporting material):

F_weight: DE-stepsize, factor by which parameter vector differences are scaled, i.e. the scaling factor $F \in [0,2]$
F_weight = 0.3;

F_CR: Crossover probabililty constant ex [0, 1]
F_CR = 0.3;

Double Heston Model Calibrated Parameters, with minimum objective value (MSE) of 0.0086 reached after 840 objective function evaluations:

Parameter	$v_1(0)$	$v_2(0)$	r	a_1	a_2	b_1	b_2	σ_1	σ_2
Value	0.0375	0.0034	0.0001	0.0013	0.0630	4.6190	4.1827	0.0632	0.3614
Parameter	ρ_1	ρ_2							
Value	-0.516	-0.341							

Table F.1.1: Tight calibration parameters for the Double Heston Model

Repeating the analysis with the new parameters:

Asian Put Options:

Strike:	Market Model	Double Heston
900	10.10	17.04
910	12.02	19.42
920	14.19	22.51
930	16.64	24.94
940	19.37	28.10
950	22.39	31.53
960	25.73	35.25
970	29.37	39.26

980	33.35	43.58
990	37.64	48.20
1000	42.27	53.13
1010	47.21	58.36
1020	52.48	63.89
1030	58.06	69.72
1040	63.95	75.83
1050	70.13	82.23
1060	76.60	88.91
1070	83.35	95.85
1080	90.36	103.06
1090	97.61	110.50
1100	105.10	118.18
MSE:	N/A	114.6506

Table F.1.2: Asian Put option prices of the Double Heston Model vs the Market Model

Cliquet Call Option:

Model:	Market Model	Double Heston
Price:	107.25	108.59

Table F.1.3: Cliquet Call option price of the Double Heston Model vs the Market Model

Lookback Call Option:

Model:	Market Model	Double Heston
Price:	100.09	107.60

Table F.1.4: Lookback Call option price of the Double Heston Model vs the Market Model

Binary Call Option:

Model:	Market Model	Double Heston
Price:	23.11	23.15

Table F.1.5: Binary Call option price of the Double Heston Model vs the Market Model

Barrier Up-and-out Put Option:

Strike:	Market Model	Double Heston
900	8.69	11.88
910	10.74	14.17
920	13.00	16.66
930	15.49	19.37
940	18.20	22.30
950	21.14	25.45
960	24.31	28.82
970	27.71	32.43
980	31.34	36.25

990	35.18	40.29
1000	39.23	44.54
1010	43.48	49.00
1020	47.91	53.65
1030	52.51	58.48
1040	57.27	63.48
1050	62.15	68.63
MSE:	N/A	24.1516

Table F.1.6: Barrier Up-and-out prices of the Double Heston Model vs the Market Model

Analysis:

It is clear that there is an improvement with respect to the Barrier option prices, the binary option price and the cliquet option. For the Asian options and the lookback option there is not any real improvement. Ultimately these findings do not contradict the conclusions in section 4.2. that the Double Heston Model was the worse performing model.

Lastly, this also shows that even though adhering to stricter calibration criteria the improvement in the resulting prices is not significant. This further affirms the analysis given in section 4.2.