European option pricing under the rough Heston model using the COS method

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

Master of Science
in
Applied Mathematics,
Financial Engineering

Kamuran Emre Erkan

August, 2020

MSc THESIS APPLIED MATHEMATICS

European option pricing under the rough Heston model using the COS method

Kamuran Emre Erkan

Delft University of Technology

Daily Supervisor
S. Liu, PhD

Responsible Professor
Prof.dr.ir. C.W. Oosterlee

Other members of the thesis committee
Dr. R.J. Fokkink
Dr. ir. L.A. Grzelak

August, 2020

Delft, The Netherlands

Acknowledgements

This thesis has been submitted for the degree of Master of Science in Applied Mathematics at Delft University of Technology. The supervisor of this thesis is Prof.dr.ir. C. W. Oosterlee, professor at the Numerical Analysis Group of Delft Institute of Applied Mathematics.

I would like to express my gratitude to my supervisor Kees Oosterlee for his excellent guidance and support throughout the writing of this thesis. I would also like to thank my daily supervisor Shuaiqiang Liu, especially for his help in the calibration part of this thesis. Finally, I am grateful to my friends and family, especially my partner Deniz Hazer with whom I have shared a home throughout the COVID-19 pandemic, for their support and encouragement.

Abstract

This thesis is about pricing European options using a Fourier-based numerical method called the COS method under the rough Heston model. Besides examining the efficiency and accuracy of the COS method for pricing options under the rough Heston model, it is also investigated if the rough Heston model produces the advantages of the so-called rough volatility models. To do so, the characteristic function of the rough Heston model is derived, and the COS method for the rough Heston model and also a Monte Carlo simulation scheme is introduced. Throughout the thesis, the theoretical background of the rough Heston model, the numerical techniques and some numerical experiments on European option prices and implied volatility behaviors are presented. Also, a calibration of the rough Heston model is performed using Artificial Neural Networks. As a result of this thesis, pricing of European options using COS method is succeeded. Moreover, it is shown that the rough Heston model produces the rough volatility behaviors as expected.

Keywords: option pricing, European options, rough Heston model, characteristic function of the rough Heston model, rough volatility, COS method, Artificial Neural Networks, Monte Carlo simulation, fractional Brownian motion.

List of Figures

List of Tables

Contents

1 Introduction

Financial markets like stock exchange markets are major driving forces for the current economic systems worldwide. An option is a financial instrument that is exchangeable on the exchange markets. It is a type of contract between some parties and its value is determined according to the conditions of this contract and the behavior of an underlying financial asset, e.g., stocks. Thus, the value of an option is based on another financial asset. This implies that an option is a type of financial derivative.

Determining the value of an option has been a significant problem for the quantitative finance. One solution for this problem is modelling the dynamics of an option's elements like stock, volatility, interest rate and etc. These models are called the option pricing models and computing the value of an option is called option pricing. In this thesis, we are going to concentrate on one of these models that is called the rough Heston model.

The rough Heston model is an example of the so-called rough volatility models. Rough volatility is a financial concept that aims to increase the accuracy of option pricing models. The rough volatility models are a relatively new area of the option pricing models. Thus, we work on an easier type of option, European option, pricing in this thesis since we believe it is a better starting point.

The COS method is a numerical method that is used to compute the option prices according to the related option pricing model. Our main work in this thesis is deriving European option prices under the rough Heston model using the COS method. Hence, our research question is: "Is the COS Method an efficient and accurate way of pricing European options under the rough Heston model dynamics?"

In addition to pricing options, looking for a parameter set that fits into the dynamics of the rough Heston model according to the real market prices of the options is also an important stage of the option pricing process. This stage is called calibration and we use Artificial Neural Networks for this stage. Thus, we have also side objectives throughout this thesis, which are different than answering the research question we determined. The objectives of this thesis can be listed as:

- Exploring the rough volatility concept and its advantages,
- Deriving the rough Heston model dynamics from stylized facts of the electronic trading markets where European options are traded,
- Applying and evaluating the performance of the COS Method for pricing European options under the rough Heston dynamics using MATLAB,
- Performing calibration to the real market option prices with Artificial Neural Networks using Python,

- Examining the results of the numerical applications to observe if the advantages of the rough volatility is ensured.

Therefore, the aim of the thesis can be described as finding an alternative and more accurate solution to the option pricing problem of the quantitative finance. The contributions to this aim that are newly introduced in this thesis are an application of the COS method for pricing European options under the rough Heston model, an application of a newly introduced (*Section 5.4*) Monte Carlo simulation that is used as a reference for European option pricing and the calibration of the European option prices for the rough Heston model parameters.

In *Section 2*, a literature review of the background information needed for the content in the thesis is given. First, a brief description of the options and option pricing is given. Then, some knowledge about volatility and implied volatility that is useful to understand the advantages of the rough volatility concept is presented. Finally, the rough volatility concept which is the base of the rough Heston model is explained by introducing the fractional Brownian motion and the rough fractional stochastic volatility model.

In *Section 3*, after introducing the Hawkes Processes, the derivation of the rough Heston model from the microstructural market interactions (and the following facts of the electronic trading market) is presented.

In *Section 4*, to use for the COS method, the characteristic function of the rough Heston model is derived. In *Section 5*, the numerical techniques that we used in numerical application; the COS method, the fractional Adams method, Volterra integral equation and Monte Carlo simulation is presented. And in *Section 6*, the numerical experiments and results are shown.

In *Section 7*, the calibration stage is presented with some numerical results.

Finally, in *Section 8*, the thesis is concluded with a summary of the results and some future research recommendations.

2 Literature Review

In this chapter, we present the prerequisite knowledge to understand how the characteristic function of the rough Heston model is derived and how it is used to price options. To do that, in *Section 2.1*, the basics of the option pricing and the volatility modelling are shown. Then, in *Section 2.2* rough volatility model basics are presented with fractional Brownian Motion, also the rough fractional stochastic volatility model is briefly explained.

2.1 Option Pricing and Volatility

An option is a right to trade an underlying asset in future with a pre-determined price [21, 39]. Options are one of the most sophisticated financial instruments in the derivatives market. Thus, deciding the value of these instruments is a significant problem of quantitative finance. The quantitative process to find the value of the option is called option pricing.

Before going into the details of the option pricing, more details on how options work should be given. Since we will concentrate on the European type options in this thesis, only the details of the European options will be given. There are mainly two types of options which are the call options and the put options. Call options give the buyer the right to buy the underlying asset whereas put options give the right to sell at a specified pre-determined price [21].

One important notification in this aspect is the fact that the buyer of the option does not have an obligation to exercise the option. This means that, if exercising the option does not give profit to the buyer, then the option will not be exercised and will be valueless. European options can only be exercised at a specified date which we call the maturity time [21, 39]. Hence, if we denote the maturity time as T, price of the underlying asset as S, the pre-determined price (strike price) as K and the value of the option as V, then,

$$V_{call,T} = max(S_T - K, 0),$$

$$V_{put,T} = max(K - S_t, 0),$$

are the fundamental value equations for the European options. These are called the payoff functions [21] and they are shown in *Figure 1.*

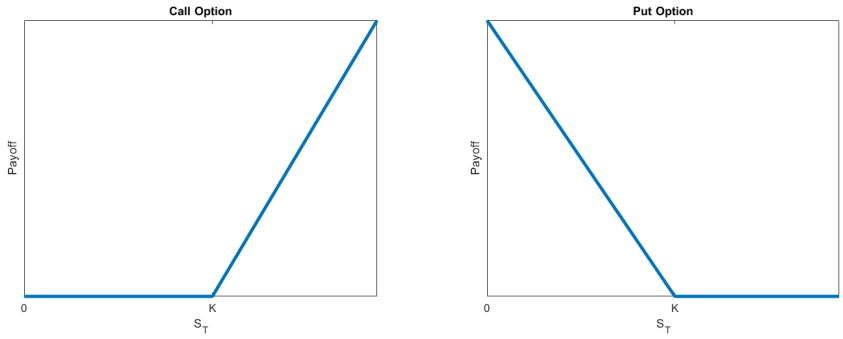


Figure (1) Payoff Functions

Following the payoff functions, some models are introduced in order to price the options before the maturity time. The most well-known and popular one of these functions is the Black-Scholes model offered by Fischer Black and Myron Scholes in [6]. Even though, this model is known to have non-negligible problems regarding the real-world market data, it is still the fundamental model used in practice. In this model, the interest rate r and the volatility v is assumed to be constant or a homogeneous function of time, and the stock price process is assumed to follow a geometric Brownian motion (Brownian motion is explained in detail in *Section 2.2.1*), so the dynamics of the stock price process, S_t, are

$$dS_t = \mu_t S_t dt + v_t S_t dW_t,$$

where μ is called the drift and W_t is the Brownian motion [21].

One important advantage of the dynamics of the Black-Scholes model is the fact that it allows an analytical solution for the option prices. First, projection of the Feynman-Kac theorem on option pricing should be briefly introduced. This theorem presents a relationship between the value of the option and the expectation of the payoff function, which is:

$$V_t = e^{-r(T-t)}\mathbb{E}[V_T|\mathcal{F}_t],$$

under the risk-free measure [21]. This relationship allows us to get an analytical solution for the option pricing since we know the probability distribution of the stock price process. Thus, from [21]

$$V_{call,t} = e^{-r(T-t)}\mathbb{E}[(S_T - K)\mathbf{1}_{S_T>K}|\mathcal{F}_t],$$

which results in,

$$V_{call,t} = S_T\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

where Φ is the cumulative distribution function of a standard normal and

$$d_1 = \frac{log(S_t/K) + (r + \frac{1}{2}v^2)(T-t)}{v\sqrt{T-t}},$$

$$d_2 = d_1 - v\sqrt{T-t}.$$

Similarly for the put options we have,

$$V_{put,t} = Ke^{-r(T-t)}\Phi(-d_2) - S_T\Phi(-d_1).$$

The analytical solution of the Black-Scholes model is also useful to calculate the implied volatility of the market or model price data. In general the volatility of the future market is unknown, thus using the analytical solution of the Black-Scholes model, the volatility that is implied by the market price and other parameters can be found. This is called implied volatility [39]. Hence, if we call the Black-Scholes analytical formula as $V_{BS}(v_{market}, S_t, r, t, K, T)$,

$$V_{BS}(v_{market}, S_t, r, t, K, T) = V_t.$$

From this relationship, if we know the option price data, V_t, then by applying an inverse function, we can get the volatility value which is the implied volatility [39]. Implied volatility is an important tool for option pricing, since the real market implied volatility behavior tends to have an overall shape [18, 19]. The most well-known indicator of this behavior is the smile or skew shape of the implied volatility - strike price graph as shown in *Figure 2*. We are going to name this behavior as the implied volatility smile behavior throughout the thesis.

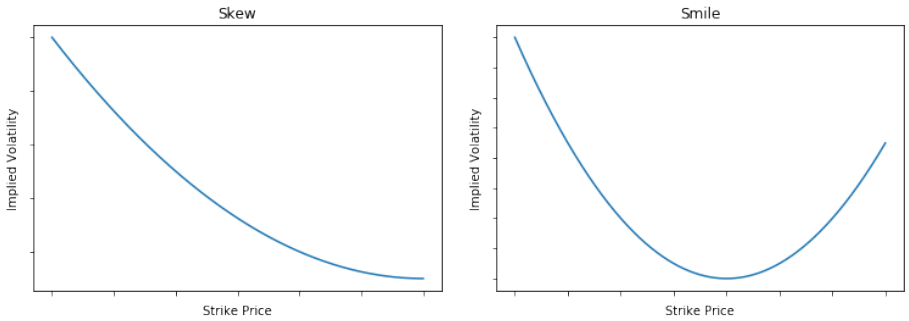


Figure (2) Volatility Skew and Smile.

The Black-Scholes model has been a revolutionary mathematical model for asset prices in the financial markets with derivative instruments. However, in this classical model the volatility process is taken as constant or a deterministic function of time. This results in an inconsistency between the model implied volatilities and market implied volatilities. Because of this unrealistic scheme of the volatility, new models with more sophisticated volatility dynamics called stochastic volatility models like the Heston model [25] or the SABR model [22]

are introduced to solve this problem. One important aspect of these stochastic volatility models is that the randomness of the volatility is generated with Brownian motion. Even though these stochastic volatility models give more realistic results, still these are not successful too on being consistent with the empirical results obtained from European options.

Especially in equity markets, looking at the behavior of the empirical data it is a generally accepted fact that the volatility surface of such markets has a certain shape. This means that the level and orientation of the volatility surface shape may change according to the situation, but the overall shape stays the same [19]. The most traditional stochastic volatility models like the SABR model and Heston model are designed to have a time-homogeneous volatility process with parameters independent of time and price of the underlying asset. However, this attempt to fit the real-world behavior of the market does not work well. In [18] and [19], it is shown in detail that these models do not fit well to the volatility surface.

One small experiment is done to give an example for this. At-the-money option prices are computed for different maturity times under the Heston model. At-the-money (ATM) options are options with strike price equal to the underlying asset price at the moment of the trade meaning $S_t = K$. Then, the implied volatility for these option prices is computed. In *Section 2.2.2*, a relationship between the ATM volatility skew (the derivative of the implied volatility with respect to $log(\frac{K}{S_t})$) and maturity time shown by Fukasawa in [17] is explained in detail. We will show the relationship for the Heston model and compare it with the Fukasawa's as shown in *Figure 5*.

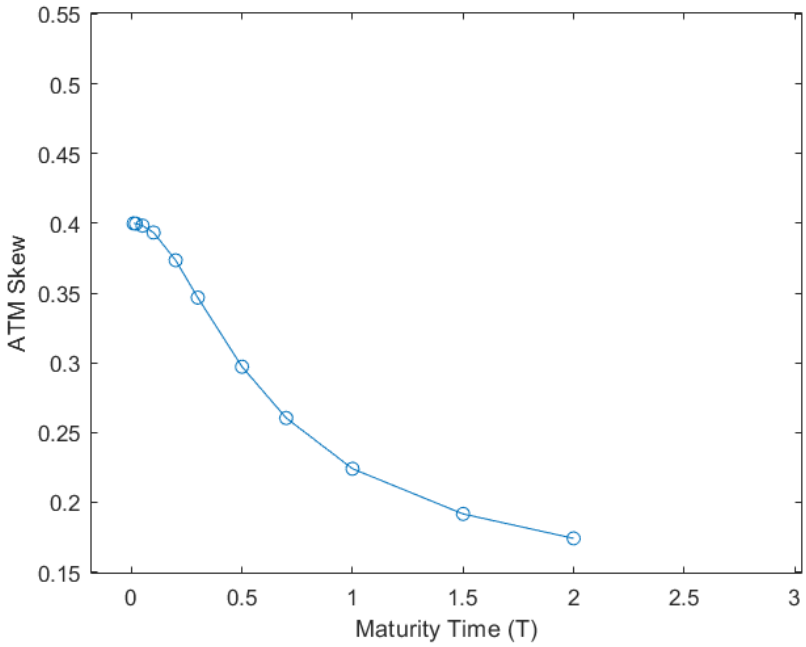


Figure (3) Heston model behavior

As we can see in *Figure 5*, the ATM skew tends to go to infinity for the real market data at maturity time close to zero. However, for the experiment with the Heston model shown in *Figure 3*, it is easily observed that the ATM skew tends to converge to around 0.4 for this example. This result implies that the Heston model does not reflect the ATM skew behavior of the real market data. The details of this experiment and the relationship are given in *Section 2.2.2 and 6.2*.

2.2 Rough Volatility

This section is to show the foundations of the rough volatility models and why they are significant to understand actual financial markets.

2.2.1 Fractional Brownian Motion

A new emerging area of research with fractional volatility models was one of the essential steps that led to rough volatility models. Fractional volatility models are basically generating the randomness of the volatility by a fractional Brownian motion instead of the classical Brownian motion.

Before going into fractional Brownian motion, classical Brownian motion should

be presented briefly. Brownian motion is a real valued continuous stochastic process with four main characteristics that define the process [39]. Hence, Brownian motion,$\{W(t) : t \geq 0\}$,

- starts at $t = 0$ with $W(0) = 0$.
- has a continuous mapping $t \to W(t)$.
- has increments which follow a normal distribution. Moreover, for $0 \leq s < t$, $W(t) - W(s) \sim \mathcal{N}(0, t - s)$.
- has increments that are independent of each other.

As a result of these characteristics, Brownian motion also shows the martingale property which makes it a critical component in the option pricing models [39]. To briefly explain, say there is a random sequence X_n as $X_1, ..., X_N$. The σ-algebra of these variables is represented by filtration $\mathcal{F}_n$. Thus, $\mathcal{F}_n$ is the natural filtration for X_n, meaning that X_n is adapted to F_n and F_n-measurable. Under these conditions a stochastic process Y_n is a martingale if $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] = Y_{n-1}$.

Fractional Brownian motion is a generalized version of Brownian motion where the increments are not independent, unlike classical Brownian motion. It may be defined by means of the Mandelbrot - van Ness representation [34],

$$W_t^H = \frac{1}{\Gamma(H + 1/2)} \Bigg[\int_{-\infty}^{0} \Big((t-s)^{H-1/2} - (-s)^{H-1/2} \Big) dW_s + \int_0^t (t-s)^{H-1/2} dW_s \Bigg]. \tag{1}$$

$H \in (0, 1)$ is called the Hurst parameter where if $H = \frac{1}{2}$, the process is classical Brownian motion. $\Gamma(.)$ is the gamma function. The covariance function of the fractional Brownian motion(fBM) can be shown to be

$$\mathbb{E}[W_t^H W_s^H] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

Thus, it is trivial that if $H < \frac{1}{2}$ the increments of the fBM have a negative correlation and if $H > \frac{1}{2}$ the increments of the fBM have a positive correlation. One other increment relationship can be shown for any $t \in \mathbb{R}, dt \geq 0, q > 0$, i.e.,

$$\mathbb{E}[|W_{t+dt}^H - W_t^H|^q] = K_q dt^{qH},$$

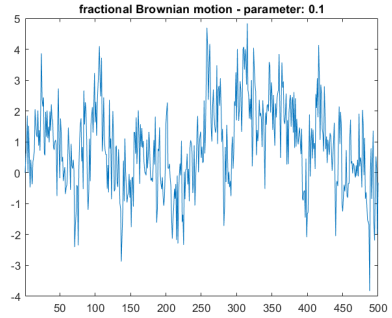
where K_q is the moment of order q for the absolute value of a standard normal variable [19]. Thus, it can be seen that the Hölder property holds for any $h < H$ for the paths of the fBM with Hurst parameter H.

Hölder Property *If a function f on a Euclidean space has the Hölder property which implies it is α-Hölder continuous where C, α are nonnegative constants, then*

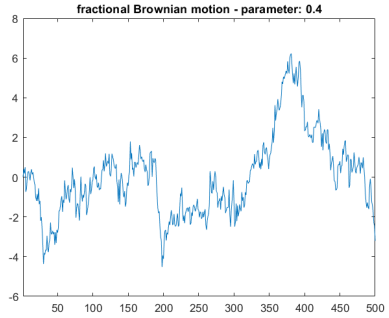
$$|f(x) - f(y)| \leq C||x - y||^{\alpha}$$

for all x, y in the domain of f. Having the Hölder property means that for any $\alpha > 0$ the function is uniformly continuous and for $\alpha = 1$ the function satisfies the well known Lipschitz condition.

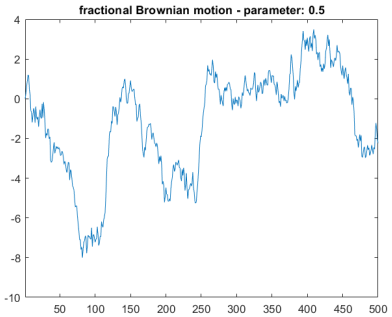
To be able to show these aspects visually, a simulation of fractional Brownian motion with different Hurst parameters is made. The movements of the fBMs are presented in *Figures 4 a,b,c,d,e.* Especially, the decreasing roughness with increasing Hurst parameter of the fBms can be clearly observed visually.



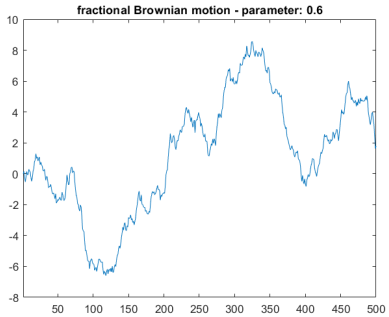
(a) Hurst parameter $H = 0.1$



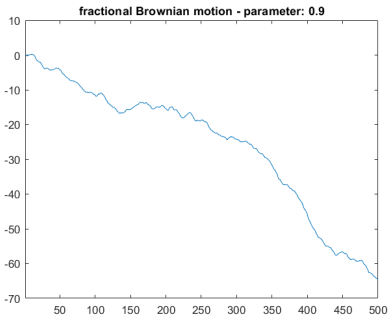
(b) Hurst parameter $H = 0.4$



(c) Hurst parameter $H = 0.5$



(d) Hurst parameter $H = 0.6$



(e) Hurst parameter $H = 0.9$

Figure (4) Fractional Brownian motion with different Hurst parameters

One of the first proposals on fractional volatility was made in [8] by Comte and Renault. By picking the Hurst parameter bigger than $\frac{1}{2}$, they wanted to include long memory effects, since it was seen as a fact that volatility process has a long memory property (more detailed information can be found in the analysis in [11]). However, in [19], it is claimed that, looking at the autocorrelation function's asymptotic behaviour, long memory can not be seen as a fact. The details of the fractional stochastic volatility (FSV) model offered by Comte and

Renault will not be discussed in this thesis since it is not directly related to the topic.

2.2.2 Rough Fractional Stochastic Volatility Model

A milestone for the rough volatility model was the paper [19] written in 2014 by Gatheral, Jaisson and Rosenbaum claiming that the volatility is rough and presenting the rough fractional stochastic volatility (RFSV) model. As a result of the availability of high frequency data, they were able to get empirical results showing that there is an approximately constant smoothness parameter which implies the volatility in practice is rough. Volatility being rough means that it follows a path like fractional Brownian motion with a Hurst parameter smaller than 0.5 instead of the classical one with $H = 0.5$. As we can see in *Figures 4 a,b,c,d,e*, the fBM with small Hurst parameter has rougher movements whereas the fBm with big Hurst parameter has smoother movements.

Before going into the details of [19], another pointer should be given. In [17], Fukasawa found an interesting relationship for small maturity time. The ATM volatility skew of a fractional stochastic volatility model with Hurst parameter H has a form $\psi(\tau) \sim \tau^{H-1/2}$ where ψ is the volatility skew and τ is time to maturity. This shows that the explosion of the ATM volatility skew as the maturity time gets closer to zero is not directly implying the presence of jumps. It is showed clearly in *Figure 5* that in a case without jumps but with rough fractional stochastic volatility, the explosion of the ATM volatility skew going to infinity is observed when Fukasawa's relationship applies. If $H = 0.5$, the ATM volatility skew would be a straight line, while if $H > 0.5$, it would tend to go to zero assuming Fukasawa's relationship applies. *Figure 5* is an oversimplification of the concept but rather is a good indicator of it. This means that rough stochastic volatility is a second possibility in the case of the explosion rather than the presence of jumps. A more solid version of the figure can be seen in *Section 6.2*.

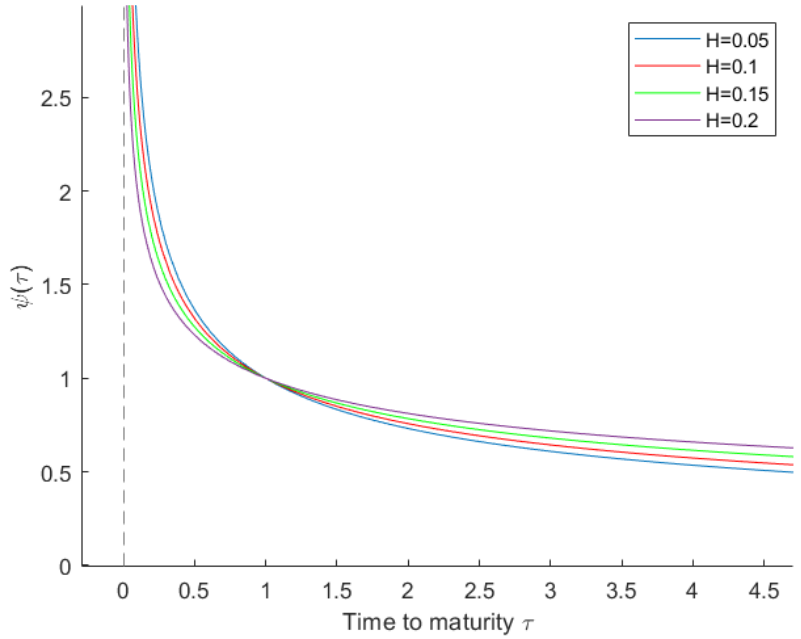


Figure (5) Fukasawa's relationship assuming $\psi(\tau) = \tau^{H-1/2}$ with different values of H.

Gatheral et al. first showed the roughness of the volatility in their paper [19] with the help of an estimation of smoothness. Assuming the availability of N discrete observations of the volatility with a mesh size of h on time grid $[0, T]$: $v_0, v_h, ..., v_{kh}, ...$, where $k \in \{0, N\}$. $N = \lfloor T/h \rfloor$, then for any $q \geq 0$ we define m [19]:

$$m(q, h) = \frac{1}{N} \sum_{k=1}^{N} |log(v_{kh}) - log(v_{(k-1)h})|^q.$$

For some $s_q > 0$ and $b_q > 0$, it is assumed that as $h \to 0$:

$$N^{qs_q} m(q, h) \to b_q.$$

Using [37], it is known that within this assumption, the volatility process is within the Besov generalized smoothness space $\mathbb{B}^{s_q}_{q,\infty}$ while it does not belong to the any Besov generalized smoothness space of order greater than s_q. Since any function in $\mathbb{B}^{s}_{q,\infty}$ has the Hölder property with parameter $h < s$ [19], the Hölder property with any parameter $s < s_q$ applies to the volatility process. In our case it implies that s_q is an estimation for the smoothness parameter, hence for an fBM with Hurst parameter H, for any $q \geq 0$, $s_q = H$ in probability because of the Hölder regularity explained in *Section 2.2.1.* Thus, to find the empirical values for the Hurst parameter which also implies the smoothness,

this estimator is used by Gatheral et al in [19].

Besov Space *Besov spaces are a generalization of the Sobolev function spaces, which are used as a measure for the regularity (smoothness) of a function [30]. Thus, Besov spaces can be used as a smoothness indicator in our case too. There are different definitions for the Besov spaces, we are going to follow the definitions of classical Besov space and Besov space of generalized smoothness which are given seperately in [7].*

Regarding the classical Besov space, say a function $f \in \mathbb{R}^N$ where $\mathbb{R}^N$ is the N-dimensional real Euclidean space; A is a positive integer and $h \in \mathbb{R}^N$, and the iterated difference operator is defined as

$$\Delta_h^A f(x) = \sum_{i=0}^{A} (-1)^{A-i} \binom{A}{i} f(x + hi).$$

Following these and $p > 0$, $q \leq \infty$ and $s \in (0, A)$ with $s > \zeta_p$ where $\zeta_p = N(\frac{1}{p} - 1)$, the Besov space $\mathcal{B}^s_{p,q}$ is the function space which consists of functions f which are measurable in $\mathbb{R}^N$ (L^P:=space of measurable functions in $\mathbb{R}^N$) such as

$$[f]_{\mathcal{B}^s_{p,q}} := \left(\int_{|h|\leq 1} ||\Delta_h^A f||^q_{L^p} \frac{dh}{|h|^{N+sq}}\right)^{\frac{1}{q}} < \infty,$$

with the case $q = \infty$,

$$[f]_{\mathcal{B}^s_{p,\infty}} := \sup_{|h|\leq 1} \frac{||\Delta_h^A f||_{L^p}}{|h|^s} < \infty.$$

For the Besov space of generalized smoothness, which is used in [37], and the smoothness estimation we presented, first we need to define admissible functions. Say there is a real function Φ on the interval (0,1], if it is positive and monotone on this interval and,

$$\Phi(2^{-j}) \sim \Phi(2^{-2j}) \quad \text{for every } j \in \mathbb{N},$$

then, Φ is admissible. Thus with the same conditions for the classical Besov space and Φ being an admissible function, the Besov space of generalized smoothness $\mathbb{B}^s_{p,q}$ is the function space which consists of functions f which are measurable in $\mathbb{R}^N$ such that,

$$[f]_{\mathbb{B}^s_{p,q}} := \left(\int_0^1 \sup_{|h|\leq t} ||\Delta_h^A f||^q_{L^p} \frac{\Phi(t)^q}{t^{1+sq}} dt\right)^{\frac{1}{q}} < \infty,$$

with the case $q = \infty$,

$$[f]_{\mathbb{B}^s_{p,\infty}} := \sup_{0<t\leq 1} t^{-s}\Phi(t) \sup_{|h|\leq t} ||\Delta_h^A f||_{L^p} < \infty.$$

As a result of an empirical estimation applied on DAX and Bund future contracts and S&P and NASDAQ indices in [19] where $m(q, \Delta)$ from [19] is used as the empirical counterpart of $\mathbb{E}[|log(v_\Delta) - log(v_0)|^q]$, they reached the result that the Hurst parameter can be taken constant as $H \approx 0.1$. According to this result, the volatility process is obviously a rough process. Also in [5] and [33], similar empirical findings are made to estimate the roughness of the volatility which gave supporting results to [19].

After showing that the volatility is rough, Gatheral et al. introduced a model called Rough Fractional Stochastic Volatility (RFSV) model where the log-volatility is an Ornstein-Uhlenbeck process. At first, they suggested

$$log(v_{t+h}) - log(v_t) = \nu(W^H_{t+h} - W^H_t),$$

where W^H is a fBM and ν is a positive constant, since it should be a volatility model with constant smoothness and with a distribution similar to a Gaussian process. However, this is problematic because it is an unsteady process according to [19]. Choosing a fractional Ornstein-Uhlenbeck process solved this problem and the model was given as:

$$v_t = exp(X_t),$$

$$dX_t = \nu dW^H_t - \alpha(X_t - x),$$

where $x \in \mathbb{R}$ and α is positive constant. For any $\varepsilon \geq 0$, the volatility process has the Hölder property of order $H - \varepsilon$, and since we know that H is accepted as approximately 0.1, this process is rough.

The main difference between the FSV and RFSV models is obviously the different Hurst parameters which leads to different smoothness behaviour. The FSV model has a Hurst parameter $H > \frac{1}{2}$, which creates a problem looking at Fukasawa's analysis [14] that is mentioned earlier. The observed term structure has a behavior of approximate skew of $1/\sqrt{\tau}$. However, in Fukasawa's analysis with $H > \frac{1}{2}$, it can be seen that the volatility skew function should increase with time to expiration. As a result, there is a contradiction with the volatility skew of FSV and the observed term structure. However, in [4], Bayer, Friz and Gatheral developed a pricing model of a Rough Bergomi model which is a specific case of RFSV model and it is an example of a better fit of the volatility surface than with the FSV model.

Besides these models in [19], it is shown that the RFSV autocovariance function is also better fitting with the empirical observations whereas the FSV autocovariance function does not fit. Also, RSFV outperforms the well-known predictors like the AR model in the forecasting applications, which also supports the claim of the volatility being rough.

3 The Rough Heston Model

In this section, a brief introduction to the Hawkes processes in *Section 3.1* is given since these processes are fundamental in the derivation of the rough Heston model. Then, in *Section 3.2*, the microstructural foundations and the derivation of the rough Heston model is presented.

3.1 The Hawkes Processes

Before continuing with the rough volatility processes, we need to introduce one important aspect in the construction of the rough stochastic volatility processes, the Hawkes processes that are introduced and developed by A.G. Hawkes with D. Oakes in [23] and [24]. However, many types of point processes started to be called Hawkes processes, which were different from the process derived in [23] and [24]. Hence, we will define Hawkes processes as self-exciting temporal point processes.

First, we need to define a point process. If a sequence of non-negative random variables $(t_i)_{\forall i \in N^*}$ is such that $t_i < t_{i+1}$, $(t_i)_{\forall i \in N^*}$ is said to be a simple point process. Then, there is a counting process $N(t)$ associated with the point process:

$$N(t) = \sum_{i \in N^*} \mathbf{1}_{t_i < t}.$$

In general, the terminology of the point process and counting process is interchangeable. The point processes may be characterised by a distribution function of the next arriving point conditional on the past.

However, it is hard to work with these conditional probability distributions for the point processes. Hence, besides the conditional probability distributions, the intensity of the point process conditional to the past is often used. The intensity process λ of a point process adapted to the filtration $\mathcal{F}_t$ is defined as, using $h = \Delta t$ [29],

$$\lambda(t) = \lim_{h \to 0} \mathbb{E}\left[\frac{N(t+h) - N(t)}{h} \middle| \mathcal{F}_t \right].$$

The intensity function can be used to characterise the behavior of the point process as self-exciting and self-regulating. If for every introduction of a new point in the process, the conditional intensity function increases, then it is a self-exciting process. If it decreases, then the point process is a self-regulating process.

Following the information given above, the Hawkes Process is defined as [29]:

$$\mathbb{P}(N(t+h) - N(t) = m | \mathcal{F}_t) = \begin{cases} \lambda(t)h + o(h), & m = 1 \\ o(h), & m > 1 \\ 1 - \lambda(t)h + o(h), & m = 0 \end{cases}$$

Moreover, the univariate self-exciting Hawkes process is a point process N with a conditional intensity function λ given by, [29]:

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} \phi(t-s) dN_s,$$

where $\mu(t)$ is a deterministic base intensity and $\phi(t)$ is a function for the positive influence from the past events of the point process to ensure that it is a self-exciting process, so it is called the excitation function.

There are two frequently used examples for the excitation function [29]. The first one is called exponential decay which is in the form $\phi(t) = ae^{-bt}$ where a, b are positive constants. The processes with this excitation function are also called Hawkes Processes with exponentially decaying intensity. The other example is called the power law function which is in the form $\phi(t) = \frac{k}{(c+t)^p}$ with parameters c,k and p are positive scalars.

One other important aspect for us to show about Hawkes processes is the relation, by definition, between the expectation of the intensity and the expectation of the point process, i.e.

$$\mathbb{E}[dN_s] = \mathbb{E}[\lambda(s)]ds.$$

A simple proof [29]: Let's define

$$g(t) = \mathbb{E}[\lambda(t)].$$

It is trivial to show that

$$\lambda(t) = \lim_{h \to 0} \mathbb{E}\left[\frac{N(t+h) - N(t)}{h} \middle| \mathcal{F}_t \right] = \mathbb{E}\left[\frac{dN(t)}{dt} \middle| \mathcal{F}_t \right].$$

Thus,

$$g(s) = \mathbb{E}[\lambda(s)] = \frac{\mathbb{E}[\mathbb{E}[dN_s|\mathcal{F}_s]]}{ds} = \frac{\mathbb{E}[dN_s]}{ds},$$

resulting in,

$$\mathbb{E}[dN_s] = g(s)ds = \mathbb{E}[\lambda(s)]ds.$$

3.2 Microstructural Foundations and the Rough Heston Model

As mentioned in *Section 2.2*, it is expected to see a rough volatility behaviour in real market. Additional to that, the so-called leverage effect is also seen as a stylized fact in finance. The leverage effect is a phenomena observed in the financial markets which is the negative correlation between the underlying asset price returns and volatility increments [13]. El Euch, Fukasawa and Rosenbaum were able to generate these effects using microstructural interactions between

market agents in [13]. This is also an important paper in the sense that these microstructural foundations lead to the rough Heston model.

Before getting into the microstructural foundations, we introduce a generalization for the rough Heston model presented in [14]. To do that, first the dynamics in the classical Heston model without drift are shown with asset price S and volatility v [21]:

$$dS_t = S_t\sqrt{v_t}dW_t,$$
$$dv_t = \gamma(\theta - v_t)dt + \gamma\nu\sqrt{v_t}dB_t,$$

where γ,θ, ν and v_0 are positive parameters and W and B are two classical Brownian motions with $\langle dW_t, dB_t\rangle = \rho dt$. As a result of the microstructural foundations that are explained in the derivation of the rough Heston model, there is not a presence of drift. Thus, we concentrate on the no drift case of the Heston model.

The Heston model is one of the most popular models because of its success in representing the stylized facts of the low frequency data like fat tails, leverage effect or time-varying volatility. It also gives the availability to control the implied volatility smile with the ν parameter and the skew with the ρ, as a result it gives reasonable fit for the implied volatility smile and skew [14]. Besides, there is a formula for the characteristic function of the Heston model which gives great advantage for the pricing purposes. However, it does not reflect the rough behavior of the volatility, thus a version of the Heston model with rough volatility is derived.

The rough Heston model can be derived fundamentally by implementing a kernel $(t-s)^{\alpha-1}$ to get the fractional Brownian motion into the volatility equation (it will be explained in the following sections in more detail). As follows,

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

$$v_t = v_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma\nu\sqrt{v_s}dB_s, \quad (2)$$

with $\alpha \in (1/2, 1)$ determining the roughness of the volatility process, where $\alpha = H + 1/2$. H is the Hurst parameter.

To show the typical micro behaviours of the market agents, a tick by tick price model in the form of ultra high frequency dynamics is shown in [2]:

$$P_t = N_t^+ - N_t^-, \quad (3)$$

where N_t^+ represents the number of upward jumps, N_t^- represents the number of downward jumps in $[0, t]$ and P_t is the price. The tick by tick price model introduced in [13] is inspired by the Hawkes processes presented in [23]. A

bi-dimensional Hawkes process is used to be the base of the tick by tick price model:

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \varphi_1(t-s) & \varphi_3(t-s) \\ \varphi_2(t-s) & \varphi_4(t-s) \end{pmatrix} \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix}. \tag{4}$$

This is a bivariate point process of the upward and downward jumps where the λ_t values are representing the jump intensity. Vector μ contains positive constants and non-negative φs are excitation functions described in *Section 3.1* that forms the kernel matrix ϕ. A very basic description of the fundamental property of this process is the fact that the jump probabilities are associated with the location of the past events. Thus, this model facilitates to generate the bid-ask bounce effect, which is an effect commonly observed in financial markets stating that the probability of a bid order to take place increases after an ask order and vice versa. It can be done by choosing appropriate shapes for φ functions that would result in a high probability of upward jump after a downward jump and vice versa.

Developing the model, four essential facts of high frequency electronic trading markets are encoded in [13]. These are:

1. As shown in [16], market movements are mostly generated by algorithms reacting to other orders rather than real economic principles, which implies that the markets are endogeneous.

2. There is no statistical arbitrage on the high frequency markets.

3. Asymmetric liquidity is observed, see [38]. To be more precise, the action taking mechanism of buying and selling is not the same. Because of the change of inventory size, after a buying order, a market maker tends to raise the price by less than the amount that market maker tends to lower after a selling order.

4. Metaorders are the most commonly seen transactions in the market and are not carried out at once but split over time, see [1].

Encoding the Fact 1 into the tick by tick price model leads us to the so-called nearly Unstable Hawkes processes (the stability condition is close to saturation because of the high endogeneity) as explained in *Section 3.2.2*. Hence, we can say that the spectral radius of the kernel matrix integral is close to 1, but smaller [16]. Fact 2 and Fact 3 are also some specific modifications for the kernel matrix. Fact 4 results in heavy tailed distributions for the φ functions [27].

To summarize the effects of the encoding facts, Fact 1, Fact 2 and Fact 3 results in the classical Heston model of [25] by generating the leverage effect as will be shown in *Sections 3.2.2, 3.2.3 and 3.2.4*, whereas Fact 4 constructs the rough volatility which will lead to the rough Heston model.

3.2.1 Hawkes Framework

Recalling the tick by tick price model that is described earlier in equations (3) and (4) as the basis of the Hawkes framework, we need to clarify the intensity, $\lambda_t^+ dt$ is the probability of having an upward jump at time t on $[t, t+dt]$. Vector μ corresponds to exogeneous reasons and behaves as Poissonian, whereas the φ terms were explained earlier.

Finally, we define

$$\phi = \begin{pmatrix} \varphi_1 & \varphi_3 \\ \varphi_2 & \varphi_4 \end{pmatrix} : \mathbb{R}_+ \to \mathcal{M}^2(\mathbb{R}_+^*)$$

as the kernel matrix.

3.2.2 Fact 2 and Fact 3

To encode the facts mentioned at the beginning of the *Section 3.2*, the parameters of the process need to be modified as in [13]. First, Fact 2 and Fact 3 will be discussed.

Fact 2 states that there is no statistical arbitrage in the high frequency markets. This translates to the fact that the expected number of upward jumps and expected number of downward jumps should be the same essentially, i.e.,

$$\mathbb{E}[N_t^+] = \mathbb{E}[N_t^-].$$

In the Hawkes framework, it is known, by definition, that,

$$\mathbb{E}[N_t^+] = \int_0^t \mathbb{E}[\lambda_s^+] ds, \ \mathbb{E}[N_t^-] = \int_0^t \mathbb{E}[\lambda_s^-] ds,$$

and thus,

$$\begin{pmatrix} \mathbb{E}[\lambda_t^+] \\ \mathbb{E}[\lambda_t^-] \end{pmatrix} = \begin{pmatrix} \mu^+ \\ \mu^- \end{pmatrix} + \int_0^t \begin{pmatrix} \varphi_1(t-s) & \varphi_3(t-s) \\ \varphi_2(t-s) & \varphi_4(t-s) \end{pmatrix} \begin{pmatrix} \mathbb{E}[\lambda_s^+] ds \\ \mathbb{E}[\lambda_s^-] ds \end{pmatrix}$$

Looking at these, equating the expectation of the intensity of the upward jump and the intensity of the downward jump is a way to ensure there is no statistical arbitrage. Thus, taking

$$\mu^+ = \mu^-, \quad \varphi_1 + \varphi_3 = \varphi_2 + \varphi_4,$$

is encoding Fact 2 into the model.

Fact 3 basically says that the ask side is more liquid. Hence, it is more likely to see a downward jump after a downward jump than an upward jump after an upward jump. It translates to $\varphi_1(x) < \varphi_4(x)$, which implies $\varphi_3(x) > \varphi_2(x)$. To implement this fact,

$$\varphi_3 = \beta \varphi_2,$$

is set, where $\beta > 1$.

These modifications project in the Hawkes framework as

$$\phi = \begin{pmatrix} \varphi_1 & \beta\varphi_2 \\ \varphi_2 & \varphi_1 + (\beta - 1)\varphi_2 \end{pmatrix},$$

and

$$\begin{pmatrix} \lambda_t^+ \\ \lambda_t^- \end{pmatrix} = \mu \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \int_0^t \phi(t-s) \begin{pmatrix} dN_s^+ \\ dN_s^- \end{pmatrix},$$

with μ as a positive scalar constant.

3.2.3 Fact 1

To encode Fact 1, the population representation of the Hawkes processes presented in [24] and [13] is used. In this representation, the counting process $N(t)$ is considered as a population process. The population increases with migrants incoming with a Poisson process with intensity μ. Also, each of these migrants have children with the inhomogeneous Poisson process with intensity φ. The same applies to the children having children with the inhomogeneous Poisson process with intensity φ and so on. This representation is also used and explained in more detail in *Section 4.4*.

Using this representation, N being the number of transactions, migrants being the exogenous orders and children being the endogeneous orders, for the one-dimensional Hawkes process the degree of endogeneity is found to be the L^1-norm of the function that defines the influence of the past events of the point process, $||\phi||_1$.

This is explained in [26] using the population approach. The endogeneous parameter $||\phi||_1$ represents the average number of the children of one person in the one-dimensonal Hawkes process, average number of the grandchildren is $||\phi||_1^2$, average number of the children of the grandchildren is $||\phi||_1^3$ and so on. Thus, using the geometric series, the average number of total descendants of this person is as follows,

$$\sum_{k \geq 1} ||\phi||_1^k = \frac{||\phi||_1}{1 - ||\phi||_1}.$$

The degree of endogeneity is the proportion of the endogeneous orders to the total number of orders and the number of descendants represents the number of endogeneous orders. Hence, the proportion of number of descendants to total number of people (which is the number of descendants plus one in this case because there is only one migrant) represents the degree of endogeneity. As a result, it is trivial to see that this proportion is equal to $||\phi||_1$.

Since Fact 1 represents the high endogeneity of the market, $||\phi||_1$ should be close to 1 but smaller. This results in the case called nearly unstable Hawkes process, which is also in parallel with the empirical results from [16].

The projection of this result into the bi-dimensional Hawkes process gives a spectral radius of the kernel matrix integral close to 1, but smaller, to ensure Fact 1:

$$\mathcal{S}(\int_0^\infty \phi(s)ds) = ||\varphi_1||_1 + \beta||\varphi_2||_1,$$

where $\mathcal{S}$ is the spectral radius operator. Inspired by [27], to get to the spectral radius, an asymptotic framework is used in [13]. It is done basically by indexing the probability space by $T > 0$, where the bi-dimensional Hawkes process is defined on time interval $[0, T]$. Thus, the tick by tick price process $N^T = (N^{T,+}, N^{T,-})$ may be redefined with intensity

$$\begin{pmatrix} \lambda_t^{T,+} \\ \lambda_t^{T,-} \end{pmatrix} = \mu_T \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \int_0^t \phi^T(t-s) \begin{pmatrix} dN_s^{T,+} \\ dN_s^{T,-} \end{pmatrix}.$$

With this modified process in the asymptotic framework, a first assumption is made in [13] to provide the work done in the previous sections.

Assumption 3.1 *We assume $\mu_T > 0$ and*

$$\phi^T = a_T \phi, \quad \phi = \begin{pmatrix} \varphi_1 & \beta\varphi_2 \\ \varphi_2 & \varphi_1 + (\beta - 1)\varphi_2 \end{pmatrix},$$

where $\beta \geq 1$ and φs are positive measurable functions with

$$\mathcal{S}(\int_0^\infty \phi(s)ds) = ||\varphi_1||_1 + \beta||\varphi_2||_1 = 1,$$

and a_T is a positive increasing sequence converging to 1.

With *Assumption 3.1*, the Facts 1,2 and 3 are ensured with endogeneity a_T, which is close to 1 but smaller. We find

$$\mathcal{S}(\int_0^\infty \phi^T(s)ds) = a_T \mathcal{S}(\int_0^\infty \phi(s)ds) = a_T.$$

3.2.4 The Heston Model

The convergence result of the asymptotic model found by encoding the Facts 1,2 and 3 leads to a Heston model similar to the one in [26], where $\beta = 1$ is taken which is contradictory to the liquidity asymmetry, since liquidity asymmetry implies $\beta > 1$ as in *Section 3.2.2*. Thus, directly taken from [13], two new assumptions are made.

Assumption 3.2 *There exist positive parameters μ, λ and n, such that*

$$T(1-a_T) \underset{T\to\infty}{\to} \lambda, \quad \mu_T = \mu,$$

and

$$\mathcal{S}(\int_0^\infty x\phi(x)dx) = n < \infty.$$

We need to define

$$\psi^T = \sum_{k\geq 1} (\phi^T)^{*k}$$

where $(\phi^T)^{*k} = \int_0^t \phi^T(s)(\phi^T)^{*(k-1)}(t-s)ds$ for $k > 1$, $(\phi^T)^{*1} = \phi^T$.

Assumption 3.3 *The function ψ^T is uniformly bounded and ϕ is differentiable such that each component ϕ_{ij} satisfies $||\phi'_{ij}||_\infty < \infty$ and $||\phi'_{ij}||_1 < \infty$.*

The technical details for these assumptions and how it leads to the following theorem can be found in [13] and [26]. Following [26], if we take $\beta = 1$, the price process is rescaled as follows after *Assumptions 3.1, 3.2 and 3.3*:

$$\frac{1}{T}P^T_{tT} = \frac{N^{T,+}_{tT} - N^{T,-}_{tT}}{T}.$$

The rescaled process under the time interval $[0,1]$ converges in law as $T \to \infty$ to a Heston model:

$$P_t = \frac{1}{1-(||\varphi_1||_1 - ||\varphi_2||_1)} \int_0^t \sqrt{X_s} dW_s,$$

$$dX_t = \frac{\lambda}{m}(\frac{2\mu}{\lambda} - X_t)dt + \frac{1}{m}\sqrt{X_t}dB_t,$$

with W and B being uncorrelated Brownian motions. To project this result into the case which we are interested in where $\beta > 1$, El Euch et al. presents the following theorem directly taken from [13]:

Theorem 3.1 *Under Assumptions 3.1, 3.2 and 3.3, as T tends to infinity, the rescaled microscopic price*

$$\frac{1}{T}P^T_{tT} = \frac{N^{T,+}_{tT} - N^{T,-}_{tT}}{T}, \quad t \in [0,1],$$

converges in law for the Skorokhod topology to the following Heston model:

$$P_t = \frac{1}{1-(||\varphi_1||_1 - ||\varphi_2||_1)} \sqrt{\frac{2}{1+\beta}} \int_0^t \sqrt{X_s} dW_s,$$

with

$$dX_t = \frac{\lambda}{m}((\beta+1)\frac{\mu}{\lambda} - X_t)dt + \frac{1}{m}\sqrt{\frac{1+\beta^2}{1+\beta}}\sqrt{X_t}dB_t,$$

where (W,B) is a correlated bi-dimensional Brownian motion with

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

Thus, as a result encoding Facts 1,2 and 3 results in a Heston model in the long run with the leverage effect, because $\beta > 1$ implies a negative correlation. The proof of *Theorem 3.1* can be found in *Appendix C*.

3.2.5 Fact 4 and Rough Heston Model

Fact 4 states that there are commonly seen metaorders in the high frequency trading markets which are executed as numerous split in time rather than once. This translates into the heavy-tailed behaviour of the kernel matrix as explained in [27]. To briefly explain, metaorders that are split in time cause unexpected orders to take place at unexpected times, this implies that an order with small probability at a time has a greater probability than expected. Thus, it implies that the probability distribution has a heavy-tailed behaviour.

Encoding Fact 4 does not change the fact that the *Assumptions 3.1* applies. However, even though *Assumption 3.1* does not change, *Assumption 3.2* needs to be modified, as shown in [13], to ensure Fact 4 into the process.

Assumption 3.4 *For $C > 0$ and $\alpha \in (\frac{1}{2}, 1)$, it follows that*

$$\alpha x^\alpha \int_x^\infty \lambda_1(s)ds \underset{x\to\infty}{\to} C.$$

Moreover, for some $\lambda^* > 0$ and $\mu > 0$,

$$T^\alpha(1-a_T) \underset{T\to\infty}{\to} \lambda^* > 0,\ T^{1-\alpha}\mu_T \underset{T\to\infty}{\to} \mu.$$

It should be noted that the empirical results from [36] show that α is always close to $\frac{1}{2}$ in practice.

As a result of encoding the Facts 1,2,3 and 4 from the previous sections and setting $\lambda = \alpha\lambda^*/(C\Gamma(1-\alpha))$, the following theorem is presented in [13]:

Theorem 3.2 *Under Assumptions 3.1 and 3.4, the rescaled price,*

$$\sqrt{\frac{1-a_T}{\mu T^\alpha}}P^T_{tT},\ t \in [0,1],$$

converges in the sense of finite dimensional laws, as $T \to \infty$, to a rough Heston model:

$$P_t = \frac{1}{1-(||\varphi_1||_1 - ||\varphi_2||_1)}\sqrt{\frac{2}{1+\beta}}\int_0^t \sqrt{Y_s}dW_s,$$

$$Y_t = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda((1+\beta) - Y_s) ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \lambda \sqrt{\frac{1+\beta^2}{\lambda^* \mu (1+\beta)}} \sqrt{Y_s} dB_s,$$

with Y_t being the unique solution and (W, B) is a correlated bi-dimensional Brownian motion with

$$d\langle W, B \rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}} dt.$$

Also, the process Y_t has Hölder regularity $\alpha - 1/2 - \varepsilon$ for any $\varepsilon > 0$.

The main modification different from *Theorem 3.1* is the addition of the kernel $(t-s)^{\alpha-1}$ in the integrals of the stochastic volatility process which gives us the Heston model with rough volatility. Thus, it can be said that Fact 4 includes rough volatility into the process, because the kernel occurring is similar to the one presented in the Mandelbrot-van Ness representation of the fractional Brownian Motion shown in *Equation (1)*.

Now, it is trivial to see that the Hurst exponent H for the rough Heston model presented in [13] equals $\alpha - 1/2$. From the empirical results of [3], it is known that $\alpha \in (1/2, 1)$ is close to $1/2$, which implies that the Hurst parameter is close to zero meaning the volatility process is rough. Hence, the rough Heston model is derived with microstructural foundations of the high frequency electronic trading markets.

As a final note, the convergence and the proof of *Theorem 3.2* can be found in [13]. It is not presented in this thesis, because to present the proof of *Theorem 3.2*, we would need a framework that is not presented in this thesis. However, the proof of *Theorem 3.2* is similar to the proof of *Theorem 4.1* that is presented in *Appendix A* which is also a derivation of the rough Heston model.

4 Characteristic Function of the Rough Heston Model

To perform option pricing under the rough Heston model, pricing methods using the characteristic function of the model seem a reasonable possibility since the characteristic function of the rough Heston model is similar to the one of the classical Heston model. However, it is not trivial to derive the characteristic function of the rough Heston model, since using the same approach as used for the classical Heston model does not work. The reason behind this is the fact that the formula for the classical Heston model is found using its Markovian property, but the roughening of the Heston model gives rise to the loss of the Markovian property.

Let's show the derivation of the characteristic function of the Heston model. To do that, first, we need to present the pricing partial differential equation (PDE) of the Heston model that is derived using the martingale property of the discounted option price and the dynamics of the Heston model. Hence the pricing PDE is given by [21],

$$\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2}+\rho\gamma\nu Sv\frac{\partial^2 V}{\partial S\partial v}+\frac{1}{2}v\gamma^2\nu^2\frac{\partial^2 V}{\partial v^2}+rS\frac{\partial V}{\partial S}+\gamma(\theta-v)\frac{\partial V}{\partial v}-rV=0,$$

where r is the interest rate and V is the option price.

Then, the fact that the Heston model is an Markovian affine process, as shown in [21], is critical in the derivation of the Heston model characteristic function. Since from [12], we know that for the affine processes, the solutions of the characteristic function,L_P, satisfies the pricing PDE,

$$\frac{\partial L_P}{\partial t}+\frac{1}{2}v\frac{\partial^2 L_P}{\partial P^2}+\rho\gamma\nu v\frac{\partial^2 L_P}{\partial P\partial v}+\frac{1}{2}v\gamma^2\nu^2\frac{\partial^2 L_P}{\partial v^2}+(r-\frac{1}{2}v)\frac{\partial L_P}{\partial P}+\gamma(\theta-v)\frac{\partial L_P}{\partial v}-rL_P=0.$$

where $P(t)=logS(t)$.

Now, under the affinity, the form of the characteristic function is as follows [21],

$$L_X(u,t,T)=exp(\bar{A}(u,\tau)+\bar{B}(u,\tau)P(t)+\bar{C}(u,\tau)v(t))$$

where $\tau=T-t$. Also, there are some initial conditions like $\bar{A}(u,0)=0$,$\bar{B}(u,0)=iu$ and $\bar{C}(u,0)=0$ as in [21]. Putting this form inside of the pricing PDE, we get the following ordinary differential equations (ODEs),

- $\frac{d\bar{B}}{d\tau}=0$,
- $\frac{d\bar{C}}{d\tau}=\bar{B}(\bar{B}-1)/2-(\gamma-\gamma\nu\rho\bar{B})\bar{C}+\gamma^2\nu^2\bar{C}^2/2$,
- $\frac{d\bar{A}}{d\tau}=\gamma\theta\bar{C}+r(\bar{B}-1)$.

The solution of these ODEs are also solvable and gives the solution as follows [21],

- $\bar{B}(u,\tau) = iu$,
- $\bar{C}(u,\tau) = \frac{1-e^{-D_1\tau}}{\gamma^2\nu^2(1-ge^{-D_1\tau})}(\gamma - \gamma\nu\rho iu - D_1)$,
- $r(iu-1)\tau + \frac{\theta\tau}{\gamma\nu^2}(\gamma - \gamma\nu\rho iu - D_1) - \frac{2\theta}{\gamma\nu^2}log(\frac{1-ge^{-D_1\tau}}{1-g})$,

where $D_1 = \sqrt{(\gamma - \gamma\nu\rho iu)^2 + (u^2 + iu)\gamma^2\nu^2}$ and $g = \frac{\gamma-\gamma\nu\rho iu-D_1}{\gamma-\gamma\nu\rho iu+D_1}$. Hence, we have a solution for the characteristic function of the Heston model.

To derive the characteristic function of the rough Heston model, the microstructural properties are used in [14]. Thus, we need to give more detail about the microstructural properties and the asymptotic behaviour of the rough Heston model. However before going into the derivation, the Mittag-Leffler density function should be introduced.

The Mittag-Leffler distribution is introduced in *Section 4.1*. Then, in *Section 4.2* a new and proper baseline intensity is derived. After that a modified theory for the Rough Heston model is introduced in *Section 4.3*. In *Section 4.4* and *Section 4.5* the characteristic functions of the multidimensional Hawkes process and the Rough Heston model are derived, respectively.

4.1 Mittag-Leffler Distribution

All the information for this section is based on the appendix of [14]. For $z \in \mathbb{C}$ and $(\alpha,\beta) \in (\mathbb{R}_+^*)^2$, $E_{\alpha,\beta}$, the Mittag-Leffler function is defined:

$$E_{\alpha,\beta}(z) = \sum_{n\geq 0} \frac{z^n}{\Gamma(\alpha n + \beta)}.$$

For $(\alpha,\gamma) \in (0,1)\times\mathbb{R}_+$, the Mittag-Leffler density function is given by

$$f^{\alpha,\gamma}(t) = \gamma t^{\alpha-1} E_{\alpha,\alpha}(-\gamma t^\alpha), \quad t > 0,$$

and the cumulative density function is given by

$$F^{\alpha,\gamma} = \int_0^t f^{\alpha,\gamma}(s)ds, \quad t \geq 0.$$

Some properties that are important to us can also be shown as:

$$f^{\alpha,\gamma}(t) \underset{t\to 0^+}{\sim} \frac{\gamma}{\Gamma(\alpha)} t^{\alpha-1}, \qquad f^{\alpha,\gamma}(t) \underset{t\to\infty}{\sim} \frac{\alpha}{\gamma\Gamma(1-\alpha)} t^{-(\alpha+1)},$$

and

$$F^{\alpha,\gamma}(t) = 1 - E_{\alpha,1}(-\gamma t^\alpha), \qquad F^{\alpha,\gamma}(t) \underset{t\to 0^+}{\sim} \frac{\gamma}{\Gamma(\alpha+1)} t^\alpha, \qquad 1 - F^{\alpha,\gamma}(t) \underset{t\to\infty}{\sim} \frac{1}{\gamma\Gamma(1-\alpha)} t^{-\alpha}.$$

And for $\alpha \in (1/2, 1)$, $f^{\alpha,\gamma}$ is square integrable with a Laplace transform where $z \geq 0$:

$$\hat{f}^{\alpha,\gamma}(z) = \frac{\gamma}{\gamma + z^{\alpha}},$$

where the notation $\hat{\cdot}$ represents the Laplace transform.

4.2 A Proper Choice for the Baseline Intensity

Before starting to derive the characteristic function, a proper baseline intensity $\hat{\mu}_T$, indexed by $T > 0$, should be formulated which will be useful later on. We are working with the bidimensional Hawkes process, $N^T = (N^{T,+}, N^{T,-})$, again indexed by same $T > 0$, with intensity:

$$\lambda_t^T = \begin{pmatrix} \lambda_t^{T,+} \\ \lambda_t^{T,-} \end{pmatrix} = \hat{\mu}_T(t) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \int_0^t \phi^T(t-s) dN_s^T.$$

To get a rough Heston model, the assumptions made in *Section 3* still apply. Thus, we can choose Mittag-Leffler functions with heavy-tailed distributions for the excitation functions in kernel matrix ϕ^T. Mittag-Leffler functions are chosen in [14] directly as the excitation functions since it is sufficient to find functions which are suitable to excitation function properties and it makes the computations relatively easier.

Definition 4.1(from [14]) *There exist* $\beta \geq 0, 1/2 < \alpha < 1$ *and* $\gamma > 0$ *such that*

$$a_T = 1 - \gamma T^{-\alpha}, \qquad \phi^T = \varphi^T \chi,$$

where

$$\chi = \frac{1}{\beta + 1} \begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix}, \qquad \varphi^T = a_T \varphi, \quad \varphi = f^{\alpha,1},$$

where $f^{\alpha,1}$ is the Mittag-Leffler density function.

Examining the behavior of $P(X > x)$ at the limit, where X is a random variable with the Mittag-Leffler distribution, it can be seen that it is a heavy-tailed distribution. Looking at the given properties of the Mittag-Leffler density in *Definition 4.1*, it is known that:

$$\alpha x^{\alpha} \int_x^{\infty} \varphi(t) dt \underset{x \to \infty}{\to} \frac{\alpha}{\Gamma(1-\alpha)}.$$

Thus, it is trivial that

$$lim_{x \to \infty} e^{\lambda x} P(X > x) = \infty,$$

for every $\lambda > 0$ which implies that we have a heavy-tailed distribution. With *Definition 4.1*, since the two rows of the χ are equal, it is now known that

$\lambda_t^{T,+} = \lambda_t^{T,-}$.

Now, following [14], we need to show that the intensity process should be a non-constant Poisson intensity $\hat{\mu}_T(t)$. Hence, we need to look for the asymptotic behavior of the renormalized intensity to provide the projection in our rough Heston process, because with the given conditions the renormalized intensity behaves like the volatility at the limit.

Since, by definition, $\mathbb{E}[N_t^T] = \int_0^t \mathbb{E}[\lambda_s]ds$, the martingale property of the point process N_t^T can be defined as:

$$(M_t^{T,+}, M_t^{T,-}) = M_t^T = N_t^T - \int_0^t \lambda_s^T ds,$$

following $\mathbb{E}[M_t^T] = \mathbb{E}[\mathbb{E}[M_t^T|\mathcal{F}_0]] = \mathbb{E}[M_0^T] = 0$. Thus, we have,

$$\lambda_t^{T,+} = \hat{\mu}_T(t) + \int_0^t \varphi^T(t-s)\lambda_s^{T,+}ds + \frac{1}{1+\beta}\int_0^t \varphi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}).$$

Let's define

$$\Psi^T = \sum_{k\geq 1}(\varphi^T)^{*k},$$

where $*$ behaves like introduced in *Assumption 3.2*. Now a lemma derived from Wiener-Hopf equations, like in [36], is introduced before going further.

Lemma 4.1 (from [14]) *Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^d$ and $\phi : \mathbb{R}_+ \to \mathcal{M}^d(\mathbb{R})$ be a matrix valued with integrable components such that $\mathcal{S}(\int_0^\infty \phi(s)ds) < 1$. Then there exists a unique locally bounded function e from $\mathbb{R}$ to $\mathbb{R}^d$ solution of*

$$e(t) = g(t) + \int_0^t \phi(t-s)e(s)ds, \quad t \geq 0$$

given by

$$e(t) = g(t) + \int_0^t \Psi(t-s)g(s)ds, \quad t \geq 0$$

where $\Psi = \sum_{k\geq 1}\phi^{*k}$.

Then, using *Lemma 4.1*, we have

$$\lambda_t^{T,+} = \hat{\mu}_T(t) + \int_0^t \Psi^T(t-s)\hat{\mu}_T(s)ds + \frac{1}{1+\beta}\int_0^t \Psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}).$$

We know that the intensity $\hat{\mu}_T(t)$ should be of order μ_t from *Assumption 3.4*, meaning that $\hat{\mu}_T \to \mu_T$, hence $T^{1-\alpha}\hat{\mu}_T \to \mu$ where $\mu_T = \mu T^{\alpha-1}$ and μ is a

positive constant. And from [19], it is known that the normalization for the intensity should be

$$C_t^T = \frac{1-a_T}{\mu_T}\lambda_{tT}^{T,+},$$

where C_t^T is the renormalized intensity. Without this normalization, the limit behavior of the intensity would degenerate. Defining,

$$B_t^T = \int_0^{tT} \frac{dM_s^{T,+} + \beta dM_s^{T,-}}{\sqrt{T(\lambda_s^{T,+} + \beta^2\lambda_s^{T,-})}}, \qquad \nu = \sqrt{\frac{1+\beta^2}{\gamma\mu(1+\beta)^2}},$$

we get

$$C_t^T = \frac{1-a_T}{\mu_T}\hat{\mu}_T(tT) + \int_0^t T(1-a_T)\Psi^T(T(t-s))\frac{\hat{\mu}_T(Ts)}{\mu_T}ds + \nu\int_0^t T(1-a_T)\Psi^T(T(t-s))\sqrt{C_s^T}dB_s^T.$$

In [14], it was shown that,

$$(1-a_T)T\Psi^T(T.) = a_T f^{\alpha,\gamma},$$

is obtained using the Laplace transform of the Mittag-Leffler density function. Thus, as a result we have

$$C_t^T = \frac{1-a_T}{\mu_T}\hat{\mu}_T(tT) + \int_0^t a_T f^{\alpha,\gamma}(t-s)\frac{\hat{\mu}_T(Ts)}{\mu_T}ds + \nu\int_0^t a_T f^{\alpha,\gamma}(t-s)\sqrt{C_s^T}dB_s^T.$$

Since B^T would converge to a Brownian motion in the limit (can be seen by computing its quadratic variation), it is clear that taking $\hat{\mu} = \mu_T$ would mean that C^T in the limit would have an initial value equal to zero since $a_T \to 1$. However, it is also possible to have a nontrivial initial value with a non-constant intensity (which should be avoided). As a result of a convenient equality from the proof of *Theorem 4.1* and following computations that is presented in *Appendix A*, a proper baseline intensity $\hat{\mu}_T$ that would not give an initial value which is zero and non-trivial is derived in [14]:

$$\hat{\mu}_T(t) = \mu_T + \xi\mu_T\Big(\frac{1}{1-a_T}\Big(1 - \int_0^t \varphi^T(s)ds\Big) - \int_0^t \varphi^T(s)ds\Big),$$

where $\xi > 0$. It can be rewritten as:

$$\hat{\mu}_T(t) = \mu_T + \xi\mu_T\Big(\frac{T^\alpha}{\gamma}\int_t^\infty \varphi(s)ds + \gamma T^{-\alpha}\int_0^t \varphi(s)ds\Big), \tag{5}$$

which implies that the baseline intensity is a positive function assuring our model to be well-defined.

4.3 A New Theorem for the Rough Heston Model

Since there is a newly defined baseline intensity $\hat{\mu}_T(t)$ for the rough Heston model and more specific kernel matrix defined with Mittag-Leffler functions, *Theorem 3.2* that we got from [13] should be modified. To derive a new theorem, we need to look at the limiting behavior of the Hawkes process presented in the previous sections. First, some definitions in [14] for $t \in [0,1]$ are needed:

$$X_t^T = \frac{1-a_T}{T^\alpha \mu} N_{tT}^T, \quad \Lambda_t^T = \frac{1-a_T}{T^\alpha \mu} \int_0^{tT} \lambda_s^T ds, \quad Z_t^T = \sqrt{\frac{T^\alpha \mu}{1-a_T}} (X_t^T - \Lambda_t^T)$$

This leads to:

Theorem 4.1 [14] *Within the given kernel matrix and baseline intensity,* $(\Lambda^T, X^T, Z^T)_{t\in[0,1]}$ *converges in law under Skorokhod topology to* (Λ, X, Z) *as* T *goes to infinity:*

$$\Lambda_t = X_t = \int_0^t Y_s ds \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad Z_t = \int_0^t \sqrt{Y_s} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix},$$

where Y *is the unique solution of:*

$$Y_t = \xi + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma(1-Y_s) ds + \gamma \sqrt{\frac{1+\beta^2}{\gamma\mu(1+\beta)^2}} \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \sqrt{Y_s} dB_s,$$

with

$$B = \frac{B^1 + \beta B^2}{\sqrt{1+\beta^2}}.$$

(B^1, B^2) *is a bidimensional Brownian motion.*

The proof of *Theorem 4.1* can be found in [14] and in *Appendix A*. This theorem, showing that our Hawkes process has a differentiable limit with a derivative behaving like a Cox-Ingersoll-Ross process (one of the option pricing processes that has a behavior similar to Y_t) and with a nonzero initial value, states that the Hawkes process designed is suitable for the limiting volatility we look for. In [14], a new microscopic price process is defined to get a new modified rough Heston model:

$$P_{new}^T = \sqrt{\frac{\theta}{2}} \sqrt{\frac{1-a_T}{\mu T^\alpha}} (N_{.T}^{T,+} - N_{.T}^{T,-}) - \frac{\theta}{2} \frac{1-a_T}{\mu T^\alpha} N_{.T}^{T,+} = \sqrt{\frac{\theta}{2}} (Z^{T,+} - Z^{T,-}) - \frac{\theta}{2} X^{T,+},$$

for $\theta > 0$. For this microscopic price process a corollary using *Theorem 4.1* is derived:

Corollary 4.1 from [14], *As* $T \to \infty$ *the sequence of processes* $(P_{new,t}^T)_{t\in[0,1]}$ *converges in law for the Skorokhod topology to*

$$P_{new,t} = \int_0^t \sqrt{v_s} dW_s - \frac{1}{2} \int_0^t v_s ds,$$

where v is the unique solution of the rough stochastic differential equation

$$v_t = \theta\xi + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds + \gamma\sqrt{\frac{\theta(1+\beta^2)}{\gamma\mu(1+\beta)^2}}\frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{v_s}dB_s,$$

with (W,B) a correlated bidimensional Brownian motion whose bracket satisfies

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt.$$

Now, this corollary reflects the parameters of our rough Heston model and will be used to find the characteristic function of the rough Heston model.

Proof of Corollary 4.1: [14] Using *Theorem 4.1*, it is trivial to see that the microscopic price process converges in law under the Skorokhod topology, $P^T_{new} \to P_{new}$ as $T \to \infty$:

$$P_{new,t} = \sqrt{\frac{\theta}{2}}\int_0^t \sqrt{Y_s}(dB^1_s - dB^2_s) - \frac{\theta}{2}\int_0^t Y_s ds.$$

Defining $v_t = \theta Y_t$ and $dW_t = \frac{1}{\sqrt{2}}(dB^1_t - dB^2_t)$, we have

$$P_{new,t} = \int_0^t \sqrt{v_s}dW_s - \frac{1}{2}\int_0^t v_s ds,$$

with

$$v_t = \theta\xi + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds + \gamma\sqrt{\frac{\theta(1+\beta^2)}{\gamma\mu(1+\beta)^2}}\frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\sqrt{v_s}dB_s,$$

and with trivial

$$d\langle W, B\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}dt,$$

since $dW = \frac{dB^1 - dB^2}{\sqrt{2}}$ and $dB = \frac{dB^1 + \beta dB^2}{\sqrt{1+\beta^2}}$.

4.4 Characteristic Function of the Multidimensional Hawkes Process

Hawkes and Oakes presented a cluster-based representation of the multidimensional Hawkes process in [24]. This representation and its description with a population approach is useful to understand how to derive the characteristic function of the multidimensional Hawkes process.

The cluster-based representation is shown now. Say $N = (N^1, ..., N^d)$ is a d-dimensional Hawkes process with intensity:

$$\lambda_t = \begin{pmatrix} \lambda_t^1 \\ \vdots \\ \lambda_t^d \end{pmatrix} = \mu(t) + \int_0^t \phi(t-s) dN_s,$$

with $\phi : \mathbb{R}_+ \to \mathcal{M}^d(\mathbb{R}_+)$ is the kernel matrix with the properties given in the previous sections and $\mu : \mathbb{R}_+ \to \mathbb{R}_+^d$. A population approach from [14] is explanatory for the process. Say there are d types of individuals, and an individual is a migrant or the child of an other individual. Thus, from time $t = 0$:

- Migrants of type $k \in \{1, ..., d\}$ arrive with a non-homogeneous Poisson process with rate $\mu_k(t)$.
- From every migrant of type $k \in \{1, ..., d\}$, children of type $j \in \{1, ..., d\}$ are born with a non-homogeneous Poisson process with rate $\phi_{j,k}(t)$.
- Again from every child of type $k \in \{1, ..., d\}$, children of type $j \in \{1, ..., d\}$ are born with a non-homogeneous Poisson process with rate $\phi_{j,k}(t)$.

Define N_t^k with $k \in \{1, ..., d\}$ as the number of people with type k up to time t. To derive the characteristic function $L(a, t) = \mathbb{E}[exp(iaN_t)]$ with $t \geq 0$ and $a \in \mathbb{R}^d$, we first define a new population setting inspired by [14]:

- Migrants of type $j \in \{1, ..., d\}$ arrive with a non-homogeneous Poisson process with rate $\phi_{j,k}(t)$.
- From every migrant of type $j \in \{1, ..., d\}$, children of type $l \in \{1, ..., d\}$ are born with a non-homogeneous Poisson process with rate $\phi_{l,j}(t)$.
- Again from every child of type $j \in \{1, ..., d\}$, children of type $l \in \{1, ..., d\}$ are born with a non-homogeneous Poisson process with rate $\phi_{l,j}(t)$.

Let's call the corresponding point process of this population setting for a given $k \in \{1, ..., d\}$ $\tilde{N}_t^{k,j}$, which is the number of people with type j up to time t. It is easy to see that this is again a multidimensional Hawkes process with $(\phi_{j,k})_{1 \leq j \leq d}$ instead of μ_k since it has the same pattern where the only difference is the Poisson rate of the migrants. Thus, its characteristic function is given by:

$$L_k(a, t) = \mathbb{E}\left[exp\left(ia(\tilde{N}_t^{k,j})_{1 \leq j \leq d} \right) \right], \quad t \geq 0, \quad a \in \mathbb{R}^d.$$

Following [14], we return to the first population setting, define $N_t^{0,k}$ as the number of migrants of type k arrived up to time t. Also define $T_1^k < ... < T_{N_t^{0,k}}^k$ as the arrival times of these migrants. Thus, the number of descendants with

different types from a migrant of type k is represented by the second population setting. Hence:

$$N_t^k = N_t^{0,k} + \sum_{1\leq j\leq d}\sum_{1\leq l\leq N_t^{0,j}} \tilde{N}^{j,k,(l)}_{t-T_l^j},$$

where $\tilde{N}^{j,k,(l)}_{t-T_l^j}$ are independent increments of $(\tilde{N}^{j,k})$ and also independent of $N^0 = (N^{0,k})_{1\leq k\leq d}$. Hence, we have:

$$\mathbb{E}[exp(iaN_t)|N^0] = exp(ia.N_t^0)\prod_{1\leq j\leq d}\prod_{1\leq l\leq N_t^{0,j}}\mathbb{E}\left[exp\left(ia(\tilde{N}^{j,k,(l)}_{t-T_l^j})_{1\leq k\leq d}\right)|N^0\right],$$

which implies

$$\mathbb{E}[exp(iaN_t)|N^0] = exp(ia.N_t^0)\prod_{1\leq j\leq d}\prod_{1\leq l\leq N_t^{0,j}} L_j(a,t-T_l^j).$$

Note that $(T_1^k,...,T^k_{N_t^{0,k}})$ has the same law with the order statistics built from iid variables $(X_1,...,X_{N_t^{0,k}})$ with density [14],

$$\frac{\mu_k(s)1_{s\leq t}}{\int_0^t \mu_k(s)ds}.$$

With that we get:

$$\mathbb{E}[exp(iaN_t)|N^0] = exp(ia.N_t^0)\prod_{1\leq j\leq d}\left(\int_0^t L_j(a,t-s)\frac{\mu_k(s)}{\int_0^t\mu_k(s)ds}ds\right)^{N_t^{0,j}}.$$

Moreover,

$$L(a,t) = \prod_{1\leq j\leq d} exp\left(\left(\int_0^t e^{ia_j}L_j(a,t-s)\frac{\mu_k(s)}{\int_0^t\mu_k(s)ds}ds - 1\right)\int_0^t\mu_j(s)ds\right).$$

It is trivial that

$$L(a,t) = exp\left(\sum_{1\leq j\leq d}\int_0^t (e^{ia_j}L_j(a,t-s)-1)\mu_j(s)ds\right). \tag{6}$$

Hence, from this relationship we can write the same as for the second population setting:

$$L_k(a,t) = exp\left(\sum_{1\leq j\leq d}\int_0^t (e^{ia_j}L_j(a,t-s)-1)\phi_{j,k}(s)ds\right). \tag{7}$$

Now, define

$$D(a,t) = \left(e^{ia_j}L_j(a,t)\right)_{1\leq j\leq d}.$$

As a result, the characteristic function of the multidimensional Hawkes process is found.

Result [14] *Using Equation (6) it is found that*

$$L(a,t) = exp\left(\int_0^t (D(a,t-s) - \mathbf{1})\mu(s)ds \right),$$

where, using Equation (7), $\mathcal{D} : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{C}^d$ is the solution of

$$D(a,t) = exp\left(ia + \int_0^t \phi^*(s)(D(a,t-s) - \mathbf{1})ds \right)$$

where $\phi^*(s)$ is the transpose of $\phi(s)$.

4.5 The Characteristic Function of the Rough Heston Model

A generalization of the rough Heston model is given in [14] as

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

$$v_t = v_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma\nu\sqrt{v_s}dB_s.$$

Now, it is easy to say that the rough Heston model that is found in *Corollary 4.1* fits in this generalization. Taking $dP_t = d(logS_t)$ and applying Itô formula will give us the equation in *Corollary 4.1.*

The characteristic function of the rough Heston model is derived following [14]. First, we need to assign:

$$V_0 = \xi\theta, \quad \rho = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}, \quad \nu = \sqrt{\frac{\theta(1+\beta^2)}{\gamma\mu(1+\beta)^2}}.$$

Reminding ourselves from *Definition 4.1* that $\gamma = (1 - a_T)T^\alpha$, in Section 4.3 we also showed that

$$P^T_{new} = \sqrt{\frac{\gamma\theta}{2\mu}}T^{-\alpha}(N^{T,+}_{.T} - N^{T,-}_{.T}) - \frac{\gamma\theta}{2\mu}T^{-2\alpha}N^{T,+}_{.T},$$

converges in law to P_{new} in the *Corollary 4.1.* Define the characteristic function of the bidimensional Hawkes process N^T as $L^T((a,b),u)$ at time u. Also define the characteristic function of P_{new} as $L_p(a,u)$. Thus, if we have

$$a^+_T = a\sqrt{\frac{\gamma\theta}{2\mu}}T^{-\alpha} - a\frac{\gamma\theta}{2\mu}T^{-2\alpha}, \quad a^-_T = -a\sqrt{\frac{\gamma\theta}{2\mu}}T^{-\alpha},$$

$L^T((a_T^+, a_T^-), tT)$ converges in law to $L_p(a,t)$. In the result of the previous section the characteristic function for the multidimensional Hawkes process is found. Hence [14]:

$$L^T((a_T^+, a_T^-), tT) = exp\bigg(\int_0^{tT} \hat{\mu}_T(s)\Big((D^{T,+}((a_T^+, a_T^-), tT-s)-1)+(D^{T,-}((a_T^+, a_T^-), tT-s)-1)\Big)ds \bigg),$$

where

$$D^T((a_T^+, a_T^-), t) = exp\bigg(i(a_T^+, a_T^-) + \int_0^t (D^T((a_T^+, a_T^-), t-s) - (1,1))\phi^T(s)ds \bigg),$$

with $D^T = (D^{T,+}, D^{T,-}) \in \mathcal{M}^{1\times 2}(\mathbb{C})$. Then, in [14], a new variable Y^T is defined:

$$Y^T(a,.) = (Y^{T,+}(a,.), Y^{T,-}(a,.)) = D^T((a_T^+, a_T^-), .T) : [0,1] \to \mathcal{M}^{1\times 2}(\mathbb{D}).$$

Hence,

$$Y^T(a,t) = exp\bigg(i(a_T^+, a_T^-) + T\int_0^t (Y^T(a, t-s) - (1,1))\phi^T(Ts)ds \bigg), \quad (8)$$

and,

$$L^T(a_T^+, a_T^-, tT) = exp\bigg(\int_0^t (T^\alpha(Y^{T,+}(a,t-s)-1)+T^\alpha(Y^{T,-}(a,t-s)-1))(T^{1-\alpha}\hat{\mu}(sT))ds \bigg).$$

Since we derived a proper equation for the baseline intensity in *Section 4.2* and we know from *Section 4.2* that φ has a heavy-tailed distribution with

$$\alpha x^\alpha \int_x^\infty \varphi(t)dt \underset{x\to\infty}{\to} \frac{\alpha}{\Gamma(1-\alpha)},$$

by implementing *Equation (5)* it is trivial to see that

$$\begin{aligned} T^{1-\alpha}\hat{\mu}(sT) &= T^{1-\alpha}\mu_T + \xi T^{1-\alpha}\mu_T\bigg(\frac{T^\alpha}{\gamma}\int_{sT}^\infty \varphi(u)du + \gamma T^{-\alpha}\int_0^{sT}\varphi(u)du \bigg) \\ &= \mu\bigg(1 + \frac{\xi}{\gamma}s^{-\alpha}(sT)^\alpha \int_{sT}^\infty \varphi(u)du \bigg) + \mu\xi\gamma T^{-\alpha}\int_0^{sT}\varphi(u)du \\ &\underset{T\to\infty}{\to} \mu\bigg(1 + \frac{\xi}{\gamma\Gamma(1-\alpha)}s^{-\alpha} \bigg). \end{aligned}$$

Also $T^\alpha(Y^T(a,t)-(1,1))$ converges to the solutions of a Volterra-type equation which leads to the characteristic function of the rough Heston model. The solution is based on a Taylor expansion of *Equation (8)*, which can be seen in [14]

and in *Appendix B* resulting in the following theorem. Before introducing the theorem, two definitions are needed to be shown.

Fractional integral of order r, for $r \in (0, 1]$, we write [14],

$$In^r f(t) = \frac{1}{\Gamma(r)} \int_0^t (t-s)^{r-1} f(s) ds.$$

Fractional derivative of order r, for $r \in (0, 1]$, we write [14],

$$De^r f(t) = \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s) ds.$$

Theorem 4.2 from [14] *Consider the rough Heston model with a correlation between the two Brownian motions ρ satisfying $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$. For all $t \geq 0$ and fixed $a \in \mathbb{R}$, we have*

$$L_p(a,t) = exp\big(\theta\gamma In^1 h(a,t) + v_0 In^{1-\alpha} h(a,t)\big),$$

where h(a,.) is solution of the fractional Riccati equation given by

$$De^\alpha h(a,t) = \frac{1}{2}(-a^2 - ia) + \gamma(ia\rho\nu - 1)h(a,s) + \frac{(\gamma\nu)^2}{2} h^2(a,s), \quad In^{1-\alpha} h(a,0) = 0,$$

which admits a unique continuous solution.

A curious thing about this theorem is the fact that in the derivation of the characteristic function for the classical Heston Model, the classical Riccati equation is solved. The difference for the rough Heston Model is the fact that it uses a fractional Riccati equation instead of the classical one. This means that if we take $a = 1$, it will project as the characteristic function of the classical Heston model.

As a final note, the proof of *Theorem 4.2* can be found in *Appendix B*.

5 Numerical Techniques

In this section, several methods that we used for the numerical application of the option pricing under the rough Heston model will be shown. In *Section 5.1*, we explain the COS method which we use to derive the option prices using the characteristic function. In *Section 5.2*, the fractional Adams Method to numerically compute the characteristic function of the rough Heston model will be shown. In *Section 5.3* more practical details in order to compute the characteristic function will be shown using the Volterra integral equation. Then, a modified Monte-Carlo scheme to use as a reference for option pricing is shown in *Section 5.4*.

To summarize, the mix of the first three methods are used in the pricing process of the options. The Monte-Carlo simulation derives a confidence interval to be used as a reference check for the option prices derived using the other methods. The Monte-Carlo scheme that is introduced in this thesis is a partly new approach that is open to improvements.

5.1 The COS Method

For pricing of the European options, the Feynman-Kac Theorem's version which is related to the option pricing has a significant place. It is the basis of the Fourier methods and Monte Carlo methods for the pricing of European options. Since the COS Method that is going to be described is a Fourier method, we need to present the risk-neutral valuation formula obtained via the Feynman-Kac Theorem. As before, $S(t)$ is the asset price at time t and thus $P(t) := logS(t)$ with $P(t_0) = x$, $P(T) = y$, where T is the maturity time and $V(t, x)$ being the value of a European option:

$$V(t_0, x) = e^{-r\tau}\mathbb{E}[V(T, y)|\mathcal{F}(t_0)] = e^{-r\tau}\int_{\mathbb{R}} V(T, y)f_P(y)dy,$$

where r is the interest rate and $\tau = T - t_0$. $f_P(y) := f_P(t_0, x; T, y)$ is the transition probability density.

The problem with this risk-neutral valuation formula is the fact that the transition probability density is not known in general. However, the characteristic function is often known. Moreover, the characteristic function $L_P(u) := L_P(u, t)$ and the probability density function form a Fourier pair:

$$L_P(u) = \int_{\mathbb{R}} e^{iyu} f_P(y)dy,$$

$$f_P(y) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iyu} L_P(u)du.$$

As an efficient Fourier method, the COS Method introduced in [15] is using this relationship within the risk-neutral valuation formula for the pricing of vanilla

options (options with a basic structure like European options). The result of the derivation of the COS method can be described as:

$$V(t_0,x) = e^{-r\tau}\sum_{k=0}^{N-1}{}'Re\left\{L_P(\frac{k\pi}{b-a})exp(-ik\pi\frac{a}{b-a})\right\}.H_k, \tag{9}$$

where N, a, b are decided by the user of the method, H_k is called the payoff coefficient which will be explained and $\sum'$ is the sum operator with the first term weighted by one-half. To get this result, first the risk-neutral valuation formula's integration range is truncated to $[a,b] \in \mathbb{R}$:

$$V(t_0,x) \approx V_I(t_0,x) = e^{-r\tau}\int_a^b V(T,y)f_P(y)dy.$$

From Fourier-cosine series, we know that:

$$f_P(y) = \sum_{k=0}^{\infty}{}'\bar{A}_k cos(k\pi\frac{y-a}{b-a}),$$

$$\bar{A}_k = \frac{2}{b-a}\int_a^b f_P(y)cos(k\pi\frac{y-a}{b-a})dy.$$

Then, we define the truncated characteristic function, as

$$\hat{L}_P(u) = \int_a^b e^{iuy}f_P(y)dy \approx L_P(u).$$

Now, it is also trivial to see that

$$Re\{L_P(u)e^{ia}\} = \int_{\mathbb{R}} cos(uy+a)f_P(y)dy,$$

which also applies for the truncated characteristic function. Thus,

$$\bar{A}_k = \frac{2}{b-a}Re\left\{\hat{L}_P(\frac{k\pi}{b-a})exp(-ik\pi\frac{a}{b-a})\right\}.$$

Replacing the truncated characteristic function with the actual one, gives us

$$\bar{F}_k = \frac{2}{b-a}Re\left\{L_P(\frac{k\pi}{b-a})exp(-ik\pi\frac{a}{b-a})\right\} \approx \bar{A}_k.$$

Hence, by replacing $\bar{A}_k$ with $\bar{F}_k$, and truncating the series summation, we have:

$$\hat{f}_P(y) = \sum_{k=0}^{N-1}{}'\bar{F}_k cos(k\pi\frac{y-a}{b-a}) \approx f_P(y).$$

Thus, we can define

$$V_{II}(t_0,x) = e^{-r\tau}\int_a^b V(T,y)\sum_{k=0}^{N-1}{}'\bar{F}_k cos(k\pi\frac{y-a}{b-a})dy,$$

and

$$H_k = \frac{2}{b-a}\int_a^b V(T,y)cos(k\pi\frac{y-a}{b-a})dy,$$

resulting in

$$V_{II}(t_0,x) = \frac{b-a}{2}e^{-r\tau}\sum_{k=0}^{N-1}{}'\bar{F}_k.H_k.$$

As a result we have *Equation (9)* [15]:

$$V(t_0,x) \approx V_{II}(t_0,x) = e^{-r\tau}\sum_{k=0}^{N-1}{}'Re\left\{L_P(\frac{k\pi}{b-a})exp(-ik\pi\frac{a}{b-a})\right\}.H_k.$$

One important aspect here is that the payoff coefficient H_k can be derived analytically for vanilla options. Now, we are going to show an example of this analytical solution for the European options as in [15].

Payoff Coefficient We start with adjusting the log-price setting by defining:

$$y(T) = log\frac{S(T)}{K},$$

leading to

$$V(T,y) = [\alpha.K(e^y-1)]^+, \text{ where } \alpha = \begin{cases}1 & \text{for a call,}\\ -1 & \text{for a put,}\end{cases}$$

where $[.]^+ = max(.,0)$.

Now let's define two functions on an integration range $[c,d] \subset [a,b]$:

$$\chi_k(c,d) = \int_c^d e^y cos(k\pi\frac{y-a}{b-a})dy,$$

$$\varphi_k(c,d) = \int_c^d cos(k\pi\frac{y-a}{b-a})dy.$$

From basic calculus these two cosine series coefficients are known in closed form,

$$\chi_k(c,d) = \frac{1}{1+(\frac{k\pi}{b-a})^2}\left[cos(k\pi\frac{d-a}{b-a})e^d - cos(k\pi\frac{c-a}{b-a})e^c + \frac{k\pi}{b-a}sin(k\pi\frac{d-a}{b-a})e^d - \frac{k\pi}{b-a}sin(k\pi\frac{c-a}{b-a})e^c\right],$$

and

$$\varphi_k(c,d) = \begin{cases} \left[sin(k\pi \frac{d-a}{b-a}) - sin(k\pi \frac{c-a}{b-a}) \right] \frac{b-a}{k\pi} & k \neq 0 \\ d-c & k = 0 \end{cases}.$$

Thus, for $a < 0 < b$:

$$\begin{aligned} H_{k,call} &= \frac{2}{b-a} \int_0^b K(e^y - 1) cos(k\pi \frac{y-a}{b-a}) dy \\ &= \frac{2}{b-a} K(\chi_k(0,b) - \varphi_k(0,b)), \end{aligned}$$

and

$$\begin{aligned} H_{k,put} &= \frac{2}{b-a} \int_a^0 K(1 - e^y) cos(k\pi \frac{y-a}{b-a}) dy \\ &= \frac{2}{b-a} K(-\chi_k(a,0) + \varphi_k(a,0)). \end{aligned}$$

Also, it is trivial to see that for $a < b < 0$, $H_{k,call} = 0$ and for $0 < a < b$, $c = a, d = b$. The reverse of this relationship applies for the put options.

Integration Range Deciding the integration range $[a, b]$ is also an issue that is affecting the size of the numerical error made using the COS method. We are not going to show how to decide this range in this thesis in detail since it can be found in [15]. However there is a simple version to decide the integration range which is also introduced in [15]:

$$[a,b] = [-L\sqrt{T}, L\sqrt{T}],$$

where $L \in [6, 12]$.

5.2 The Fractional Adams Method

A first step to use the characteristic function of the rough Heston model in the COS method is finding $h(a,t)$ from *Theorem 4.2* which is the solution of the fractional Riccati equation:

$$De^{\alpha} h(a,t) = F(a, h(a,t)), \quad In^{1-\alpha} h(a,0) = 0,$$

with

$$F(a,x) = \frac{1}{2}(-a^2 - ia) + \gamma(ia\rho\nu - 1)x + \frac{(\gamma\nu)^2}{2} x^2.$$

To find the solution for this equation numerically we will use the fractional Adams method presented in [9,10] by Diethelm, Ford and Freed. Following [14], from the fractional integral given in *Section 4.5*, we have the following Volterra equation:

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

Let's define a regular time grid, $(t_k)_{k\in\mathbb{N}}$, where $t_k = k\Delta$, and $f(a,t) := F(a,h(a,t))$. Then, we have

$$h(a,t_{k+1}) \approx \frac{1}{\Gamma(\alpha)}\int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1}\hat{f}(a,s)ds,$$

with

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

Thus, using the trapezoidal rule and well-known techniques from quadrature theory for the fractional integral, we find the following approximation,

$$\hat{h}(a,t_{k+1}) = \frac{1}{\Gamma(\alpha)}\int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1}\left[\frac{s_{j+1}-s}{s_{j+1}-s_j}\hat{f}(a,s_j)+\frac{s-s_j}{s_{j+1}-s_j}\hat{f}(a,s_{j+1})\right]ds,$$

where, with a manipulation of notation,

$$s_k = k\Delta, \quad s\in[s_j,s_{j+1})$$

and as a straightforward solution of the integral, after messy calculations, we have [14]:

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

with

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

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

and

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

which are derived as the straightforward solutions of the integral.

The problem with this scheme is the fact that it is an implicit scheme. Thus, as a substitution for the variable causing the implicit scheme, we will employ a predictor resulting from a Riemann sum [14]:

$$h^P(a,t_{k+1}) = \frac{1}{\Gamma(\alpha)}\int_0^{t_{k+1}} (t_{k+1}-s)^{\alpha-1}\tilde{f}(a,s)ds,$$

where

$$\tilde{f}(a,t) = \hat{f}(a,t_j), \quad t\in[t_j,t_{j+1}), \quad 0\le j\le k.$$

Following, similarly,

$$h^P(a,t_{k+1}) = \sum_{0\le j\le k} b_{j,k+1} F(a,\hat{h}(a,t_j)),$$

with

$$b_{j,k+1} = \frac{\Delta^\alpha}{\Gamma(\alpha+1)}((k-j+1)^\alpha - (k-j)^\alpha), \quad 0 \le j \le k.$$

Hence, our final scheme turns out to be [14]:

$$\hat{h}(a,t_{k+1}) = \sum_{0\le j\le k} a_{j,k+1} F(a,\hat{h}(a,t_j)) + a_{k+1,k+1} F(a,h^P(a,t_{k+1})).$$

5.3 Volterra Integral Equation

After applying the fractional Adams method, in order to find the characteristic function numerically, inspired by [20], we use the Volterra integral equation. In [20], it is showed that the fractional differential equation,

$$De^\alpha h(a,t) = F(a,h(a,t)), \quad In^{1-\alpha} h(a,0) = 0,$$

is applicable if and only if it satisfies the Volterra integral equation:

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

Hence, inspired by this relationship, it is trivial to see that the fractional derivative process and fractional integral process have counter effects on the function. To clarify, $In^\alpha De^\alpha h(a,t) = h(a,t)$.

The rough Heston characteristic function from *Theorem 4.2*, has both $In^1 h(a,t)$ and $In^{1-\alpha} h(a,t)$ terms inside that need to be solved numerically. Since, we have a mesh of $h(a,t)$ obtained from the fractional Adams method, the numerical solution of the first term $In^1 h(a,t)$, which is the classic integral, is trivial. For the second term, we can use the relationship between the fractional derivative and fractional integral,

$$In^{1-\alpha} h(a,t) = In^1 De^\alpha h(a,t) = In^1 F(a,h(a,t)).$$

Since, thanks to fractional Adams method, we have a discretization for $h(a,t)$, we can easily get a mesh for $F(a,h(a,t))$ too. Thus, taking the classical integral of $F(a,h(a,t))$ is also trivial. As a result of these, the value of the rough Heston characteristic function can be computed numerically.

Finally, since the characteristic function of the rough Heston model can be computed, it is also easy to apply the COS method that is explained in *Section 5.1.*

5.4 Monte Carlo Simulation

In order to set a reference confidence interval for the European option prices computed by the COS method, we apply a Monte-Carlo simulation for the rough Heston model. To briefly explain Monte-Carlo simulation for option pricing, it is a process that uses a discretization scheme for the option model dynamics. A population of stock price results at maturity time, S_T, can be found with the simulation. From this stock price values, we can find the payoff values of the options, e.g. for call options, it is $[S_T - K]^+$. From Feynman-Kac theorem, we know that the option value is an expectation of the payoff values, so, by taking the average of the payoff values we got from the simulations, we can have an estimation of the option value. However, this result is not constant because of the randomness in the dynamics. Thus, we prefer to have a confidence interval for the option values.

For the stock price process of the generalized rough Heston model shown in *Equation (2)*, a basic discretization scheme is used:

$$S_{i+1} \overset{d}{=} S_i + S_i\sqrt{v_i}\sqrt{\Delta t}z_{1,i}$$

where the maturity time T is meshed into N equidistant points implying $T = N\Delta t$, $t_i = i\Delta t$, $S_{t_i} := S_i$ and $v_{t_i} := v_i$. For the dB_s term, we use $\sqrt{\Delta t}z_{1,i}$ where $z_{1,i}$ is a randon number generated from a standard Gaussian distribution.

We are going to present two different discretizations for the volatility process. The first one is a more sophisticated approach which has a poor performance in practice. We aim to suggest a starting point for the future research by presenting it. The second discretization is a similar but simpler way which shows a better performance.

We should remark that the first discretization uses two slightly different approaches for the different parts of the volatility process. Let's concentrate on the first term of the generalization of the rough Heston volatility process in *Equation (2)*:

$$Term_1 \approx \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds.$$

For this term, the critical thing to realize is the fact that it has the same structure as the fractional integral defined in *Section 4*. Thus, we use a discretization scheme for the fractional integral inspired by the one shown in [35]. Again, the same technique based on the trapezoidal rule from the quadrature theory used in the fractional Adams method is used for this approach too. Let's define:

$$k_i := \gamma(\theta - v_i),$$

then

$$Term_1 = \frac{(\Delta t)^\alpha}{\Gamma(\alpha+2)}\sum_{i=0}^{N} c_{1,i}k_i$$

where coefficients $c_{1,i}$ are defined as follows,

$$c_{1,i} = \begin{cases} (1+\alpha)N^{\alpha} - N^{1+\alpha} + (N-1)^{1+\alpha}, & i = 0, \\ (N-i+1)^{1+\alpha} - 2(N-i)^{1+\alpha} + (N-i-1)^{1+\alpha}, & 0 < i < N, \\ 1, & i = N. \end{cases}$$

Now, let's concentrate on the second term:

$$Term_2 \approx \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma\nu\sqrt{v_s} dB_s.$$

For this term, the same approach we used for the first term can not be used. The reason behind this is the fact that the second term is a stochastic integral in which the technique from the quadrature theory would not apply directly. Thus, a different approach based on the product rule is used. From product rule, we have the following relationship where $l(t)$ is a real function,

$$d(l(t)B(t)) = B(t)dl(t) + l(t)dB(t),$$

$$\int_0^t d(l(s)B(s)) = \int_0^t B(s)dl(s) + \int_0^t l(s)dB(s),$$

$$l(t)B(t) = \int_0^t B(s)dl(s) + \int_0^t l(s)dB(s).$$

Using this relationship, we can write that,

$$\begin{aligned} \int_0^t (t-s)^{\alpha-1} dB_s &= (t-t)^{\alpha-1} B_t + \int_0^t B_s(\alpha-1)(t-s)^{\alpha-2} ds \\ &= \int_0^t B_s(\alpha-1)(t-s)^{\alpha-2} ds. \end{aligned}$$

Then,

$$dB_s = \frac{B_s(\alpha-1)}{(t-s)} ds,$$

Using this, it can be written that,

$$\frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \gamma\nu\sqrt{v_s} dB_s = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-2} \gamma\nu\sqrt{v_s} B_s(\alpha-1) ds.$$

Thus,

$$Term_2 \approx \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-2} \gamma\nu\sqrt{v_s} B_s(\alpha-1) ds,$$

which is now similar to the first term. Hence, we can use the same technique with the first term now. Let's define:

$$l_i := \gamma\nu\sqrt{v_i} B_i(\alpha-1),$$

with $B_i = B_{i-1} + z_{2,i-1}\sqrt{\Delta t}$, $z_{2,i} = \rho z_{1,i} + \sqrt{1-\rho^2}\epsilon$ and ϵ is a random number generated by standard Gaussian distribution. Now, due to the different integration, there are some small changes in the constants of the discretization. After the proper integration, it can be seen that,

$$Term_2 = \frac{(\Delta t)^{\alpha-1}}{\Gamma(\alpha)\alpha(\alpha-1)} \sum_{i=0}^{N} c_{2,i} l_i,$$

where coefficients $c_{2,i}$ are defined as follows,

$$c_{2,i} = \begin{cases} \alpha N^{\alpha-1} - N^\alpha + (N-1)^\alpha, & i = 0, \\ (N-i+1)^\alpha - 2(N-i)^\alpha + (N-i-1)^\alpha, & 0 < i < N, \\ 1, & i = N. \end{cases}$$

Hence, as a result the first discretization scheme for the volatility process can be shown to be,

$$V_{i+1} = V_i + \frac{(\Delta t)^\alpha}{\Gamma(\alpha+2)} c_{1,i} k_i + \frac{(\Delta t)^{\alpha-1}}{\Gamma(\alpha)\alpha(\alpha-1)} c_{2,i} l_i.$$

This discretization is inspired from the fractional Adams method. However, we neglect the implicit part of the method due to the huge computational cost in practice. We believe that this might be a good starting point for the future research.

One other problem is observed in practice for this discretization. The problem is that the option prices that we got from this discretization are unstable. It means that the length of the confidence interval we got for every other simulation fluctuates a lot. The mean of these discretizations also fluctuate. Thus, we introduce a similar but simpler discretization using the left rectangle rule instead of the trapezoidal rule. This approach leads to solutions with better stability.

The only difference between these two discretizations is the coefficients due to the integration differences. The integration for the left rectangle rule discretization is the same with the implicit part of the fractional Adams method. Thus, for the first term of the volatility, we have,

$$Term_1 = \frac{(\Delta t)^\alpha}{\Gamma(\alpha+1)} \sum_{i=0}^{N} d_{1,i} k_i,$$

where

$$d_{1,i} = (N-i)^\alpha - (N-i-1)^\alpha, \qquad 0 \le i < N.$$

Similarly, for the second term, we have,

$$Term_2 = \frac{(\Delta t)^{\alpha-1}}{\Gamma(\alpha)(\alpha-1)} \sum_{i=0}^{N} d_{2,i} l_i,$$

where

$$d_{2,i} = (N-i)^{\alpha-1} - (N-i-1)^{\alpha-1}, \quad 0 \leq i < N-1,$$

and

$$d_{2,N-1} = -1.$$

Finally, the second discretization scheme for the volatility process is as follows,

$$V_{i+1} = V_i + \frac{(\Delta t)^{\alpha}}{\Gamma(\alpha+1)} d_{1,i} k_i + \frac{(\Delta t)^{\alpha-1}}{\Gamma(\alpha)(\alpha-1)} d_{2,i} l_i.$$

The second discretization is used for the numerical experiments throughout the thesis.

5.4.1 Convergence of the Monte Carlo Simulation

To observe the convergence behavior of the Monte Carlo simulation with the new scheme that we modified for the rough Heston model, an experiment is done. This experiment is under the rough Heston model given in *Equation (2)* with $H = 0.1$, and with parameter set:

$$r = 0,\ \gamma = 0.1,\ \theta = 0.0398,\ \nu = 0.3647,\ \rho = -0.5711,\ v_0 = 0.0175, S_0 = 100.$$

In the first part of the experiment, the behavior of the European put option prices with strike price, $K = 100$, and maturity time, $T = 1$, for different numbers of Monte Carlo paths is observed. The European put option prices are computed as a 95% confidence interval. Hence, the behavior of the upper and lower limits of the confidence intervals with their mean is observed in *Figure 6*.

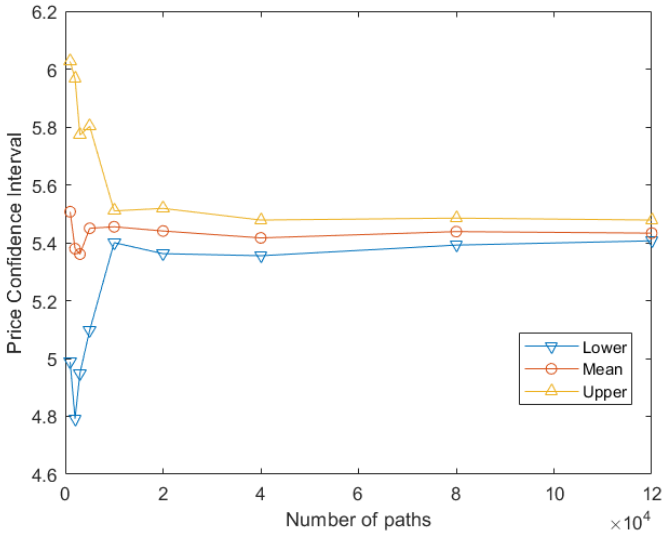
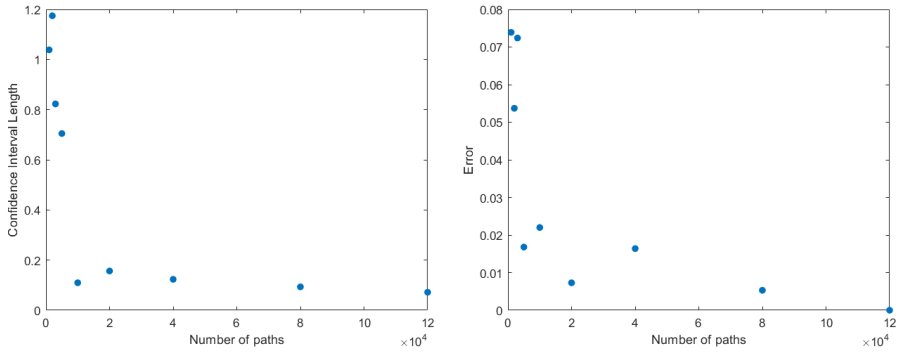


Figure (6) 95% Confidence interval Monte Carlo simulation convergence

It is clear that the length of the confidence interval tends to get smaller as the number of paths increases. At around 10000 paths, we start to get a reasonable confidence interval which gets better around 120000 paths which however, gives a huge computational time. In *Figure 6*, we also see that the mean of the CI tends to converge around 10000 number of paths.

In *Figure 7*, we present two more graphs to support our results where one is observing the length of the confidence interval, while the other is concentrating on the error of the mean. To calculate the error, we appointed a reference point by making a simulation with 120000 paths. This simulation gave a 95% confidence interval of $[5.4065 - 5.4596]$ with a mean of 5.4331, which we accepted as a reference point.

Looking at *Figure 7.a*, we see the decrease of the confidence interval length with increasing number of Monte Carlo paths. In *Figure 7.b*, we see that for the mean of the confidence interval, even for a small number of paths we get reasonable errors smaller than 0.03 and increasing the number of paths has a negligible effect on the amount of error. Since we are using Monte Carlo simulation as a reference in our experiments, we can conclude that if the option price of our interest is close to the mean of the confidence interval even if we have a wide confidence interval, then the present option pricing technique is working well.



(a) Confidence interval length (b) Option price error

Figure (7) Supporting graphs for MC simulation convergence

In the second part of the experiment, we perform the same experiment but change the number of time steps instead of the number of Monte Carlo paths. Hence, under same conditions with the first part of the experiment, we observe the 95% confidence interval with an increasing number of time steps in *Figure 8*.

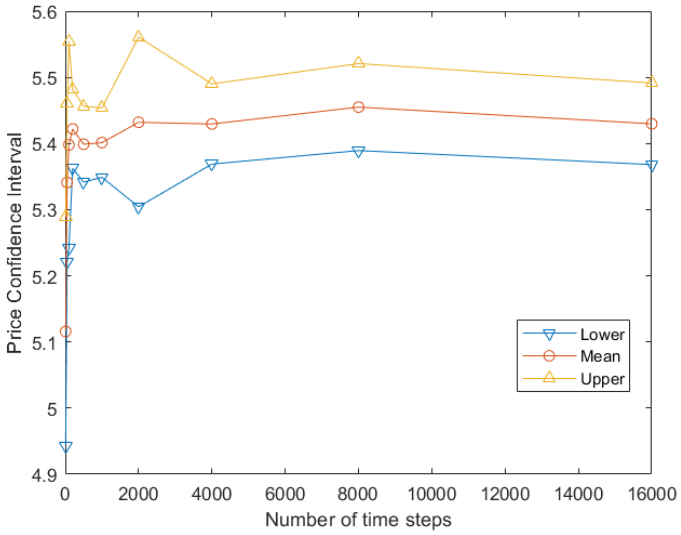


Figure (8) 95% Confidence interval Monte Carlo simulation convergence

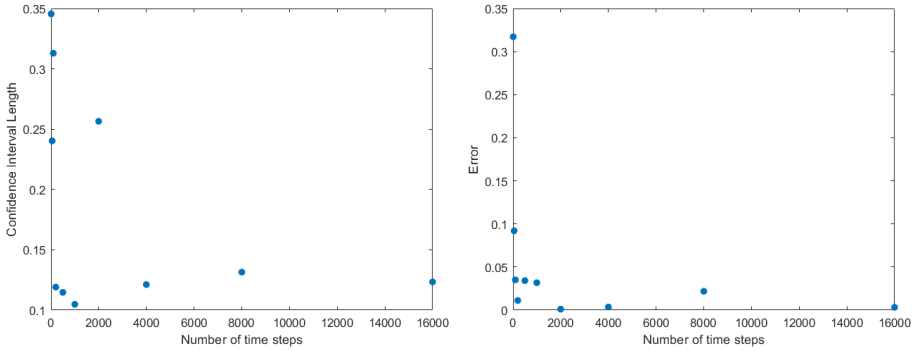
From *Figure 8*, we can conclude that after around 4000 time steps, increasing the number of time steps does not have a strong effect on the length of the confidence interval. Thus, we can say that the number of paths is more significant in order to narrow down the confidence interval. Again, around 2000 time steps, the mean of the confidence interval tends to converge at a reasonable amount.

Similar to the first part of the experiment, we have again supporting graphs in *Figure 9*. The reference point that is used to calculate the error is also the same as in the first part of the experiment.

We see in *Figure 9.a* that there is no convergence or not a regular pattern for the length of the confidence interval. Thus, it supports our conclusion that the number of time steps has less significance on narrowing down the confidence interval. In *Figure 9.b*, we see that even for a smaller number of time steps, the error for the mean price of the confidence interval is a sufficiently small error. Thus, having a large number of time steps does not compensate for the computational cost of it.

As a result of this experiment, we can say that our simulation is accurate enough to use as a reference for other experiments, especially in the sense that the width of the confidence interval is not significant in the case of having close results to the mean of the confidence interval. Throughout the numerical experiment, it is observed that the length of the confidence interval is not always narrow. Hence, as the main method to price the European options, this Monte Carlo simulation

performs poorly in the way that it has a large computation time in order to achieve narrower confidence intervals. Thus, the convergence properties of the simulation need to be improved.



(a) Confidence interval length (b) Option price error

Figure (9) Supporting graphs for MC simulation convergence

6 Numerical Experiments

In this section, some numerical experiments will be applied to test the feasibility of the numerical application. In *Section 6.1*, the procedure and the results of an European option pricing experiment is presented. In *Section 6.2*, some other numerical experiments that are supporting to the advantages of the rough Heston model are shown.

6.1 European Option Pricing Experiment

The generalization of the rough Heston model is given in *Equation (2)*. Let's first give a parameter set for this version of the model:

$$r = 0.3,\ \gamma = 0.1,\ \theta = 0.3156,\ \nu = 0.331,\ \rho = -0.681,\ v_0 = 0.0392, S_0 = 100,$$

where r is the constant interest rate. The first experiment using these parameters aims to test the COS method and characteristic function that is explained in the previous sections. We used different values for the α and compared the results we got from the COS Method and Monte Carlo simulation. The tests are done in MATLAB version R2019b.

By choosing α as 0.6, 0.99 and 1, we aim to see the differences between the 'rougher' rough Heston model, 'smoother' rough Heston model and classic Heston model. Also, we want to observe the transition from the rough Heston model to the classical Heston model for same parameters except α. We also show results using a well-established classical Heston model characteristic function (from https://github.com/LechGrzelak/QuantFinanceBook/tree/master/MatlabCodes) for the COS method in order to test the numerical solution of the characteristic function. As a final note, Monte Carlo simulation is done only for $\alpha = 0.6$ and $\alpha = 0.99$ and the results are given as a 95% confidence interval.

In *Table 1*, the resulting prices of the experiment of the European call options with maturity time, $T = 1$. K represents the strike price of the option.

European Call Option with T=1						
K	$\alpha = 0.6$		$\alpha = 0.99$		$\alpha = 1$	CHeston
	MC	COS	MC	COS	COS	COS
100	6.5393-6.6688	6.6198	6.6753-6.7919	6.7416	6.7417	6.7402
80	15.8708-16.1700	16.1349	16.0081-16.3038	16.2115	16.2116	16.2107
120	2.0842-2.2267	2.0529	2.1013-2.2852	2.1383	2.1384	2.1372

Table (1) European call option prices with maturity time 1

In *Table 2*, similarly we present the same results for the European put options with maturity time, $T = 1$. Besides these, in *Table 3* we present European call option prices with strike price $K = 100$ for different maturity times.

European Put Option with T=1						
K	$\alpha = 0.6$		$\alpha = 0.99$		$\alpha = 1$	CHeston
	MC	COS	MC	COS	COS	COS
100	6.5355-6.7076	6.6198	6.7248-6.7959	6.7416	6.7417	6.7402
80	1.2079-1.2879	1.3186	1.3219-1.4072	1.3951	1.3952	1.3944
120	16.7991-17.1215	16.8692	16.7992-17.0909	16.9546	16.9547	16.9535

Table (2) European put option prices with maturity time 1

European Call Option with K=100						
T	$\alpha = 0.6$		$\alpha = 0.99$		$\alpha = 1$	CHeston
	MC	COS	MC	COS	COS	COS
0.5	5.1538-5.3105	5.2394	5.1589-5.2670	5.1971	5.1943	5.1934
2	7.2003-7.3971	7.2846	7.7376-7.9999	7.8051	7.8156	7.8134
5	5.0186-5.2053	5.0830	5.9945-6.1218	6.0210	6.0453	6.0433

Table (3) European call option prices with strike price 100

As a result of these tests, we can say that the COS method, with integration range [-6,6] and number of cosine terms equal to 160, gives good results for the at-the-money European option prices under the rough Heston model, if we take the Monte Carlo simulations as our reference. There is also a consistency between the COS method and Monte Carlo results for different maturity times too. However, the consistency starts to disappear as we go to different strike prices.

There is a consistency between the relationship of roughness too, meaning that as α gets closer to 1, the option price gets closer to the option price under the classical Heston model. In all cases, $\alpha = 0.99$ gives similar results to $\alpha = 1$ as expected and also $\alpha = 1$ gives reasonably close results to the one with the classical Heston characteristic function which implies that our approach to solve characteristic function numerically is also acceptable. This also implies that the reason for the loss of consistency between the COS method and Monte Carlo results is due to the poor performance of Monte-Carlo simulations at strike prices 80 and 120.

6.1.1 Convergence of the COS Method

After the validation of the COS method, we applied a convergence experiment for the COS method too. This experiment is also consisted of two parts like the experiment for the convergence of Monte Carlo simulation in *Section 5.4.1*. The parameter set for the experiment is:

$r = 0\ \gamma = 1.5768,\ \theta = 0.0398,\ \nu = 0.3647,\ \rho = -0.5711,\ v_0 = 0.0175, S_0 = 100.$

The price results for this experiment is for put options with $H = 0.1$, $T = 1$ and $K = 100$.

In the first part of the experiment, we observe the convergence of the COS method with increasing number of the cosine terms shown in *Section 5.1*. As a result of this, we obtain a behavior of option price against the number of cosine terms shown in *Figure 10*.

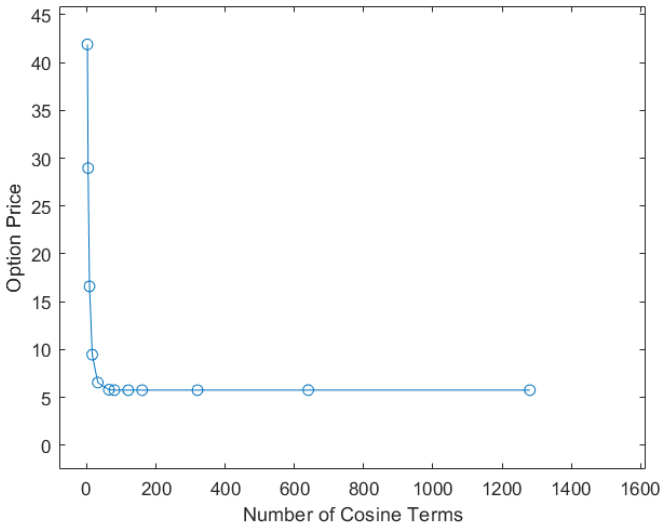


Figure (10) Convergence of the COS method with increasing number of cosine terms

Looking at *Figure 10*, we can say that the price of the option tends to converge before 200 cosine terms. To get in more details with the convergence, taking the option price result with 2^{12} cosine terms as reference, we have the following *Table 4*:

N	64	80	120	160	320	640
Abs. Error	0.0433	0.0018	0.0046	9.0e-04	0	0

Table (4) Absolute Error for the COS method with different number of cosine terms

Table 4 implies that around 160 cosine terms, we have a sufficient convergence for the option prices found with COS method. For the second part of the experiment we concentrate on the effect of the number of time steps to calculate the characteristic function instead of number of cosine terms. Again, under same conditions, we obtain *Figure 11*. Looking at it, we see that around 2000 time steps, it tends to converge.

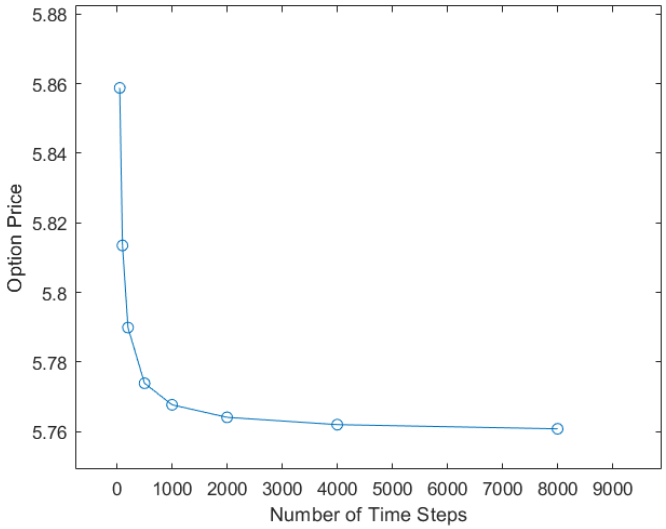


Figure (11) Convergence of the COS method with increasing number of time steps

One important note about the performance of the COS method is the low computational time. By choosing the number of cosine terms as 160 and the number of time steps as 2000 (good enough numbers for the convergence), the price of an European option is computed in between 1 and 1.1 seconds. This computation is done with MATLAB, using a Laptop PC with 2.20 GHz processor and 16 GB RAM. Thus, it can be said that the COS method is efficient.

6.2 Other Numerical Experiments

It is explained in *Section 2.2.2* and in [19] that the empirical studies about the real-life market volatility show that the volatility should be rough. One way to show this is found in the stylized facts about the implied volatility surface. For the real market data, the general shape of the implied volatility surface is observed to be similar for even different parameters. Thus, to show that the rough Heston European option pricing model fits to the real market behavior we will present two experiments. One is looking for the implied volatility smile of the rough Heston model and the other is looking for the behavior of the at-the-money volatility skew behavior from Fukasawa's [17] shown in *Section 2.2.2*.

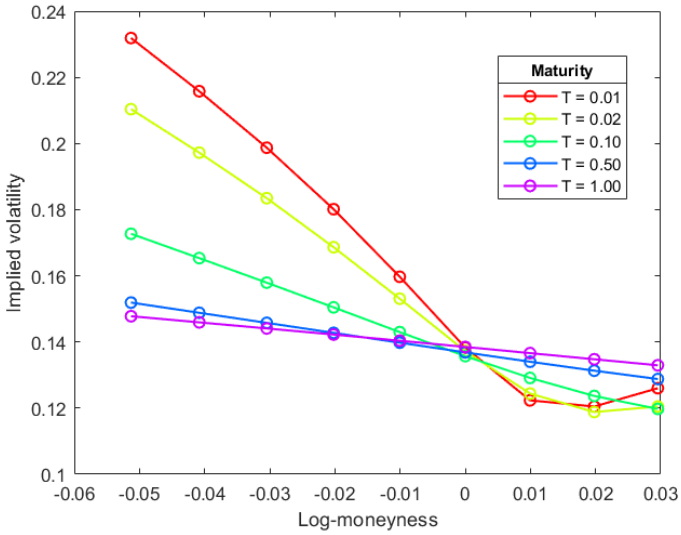
First, a parameter set for the rough Heston model is defined for these experiments:

$$r = 0,\ \gamma = 2,\ \theta = 0.0225,\ \nu = 0.2,\ \rho = -0.6,\ v_0 = 0.0225, S_0 = 100.$$

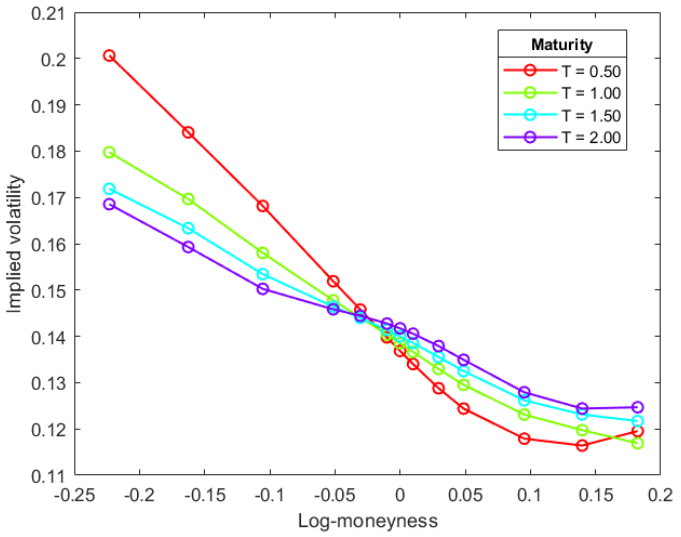
Then, let's define log-moneyness, k,

$$k := log(\frac{K}{S_0}).$$

Now, the smile behavior of the Black-Scholes implied volatility, σ_{BS}, of the rough Heston European option prices under the rough Heston model with $\alpha = 0.6$ for different maturity times is shown in *Figures 12 a,b*:



(a) Small Maturity Times



(b) Larger Maturity Times

Figure (12) Rough Heston model implied volatility smiles

For the second experiment, to get to the Fukasawa's relationship, we first need to define the implied volatility at-the-money skew, which is the absolute value of the partial derivative of the implied volatility to the log-moneyness

when the log-moneyness is equal to zero. Hence, the ATM skew is given by:

$$\left| \frac{\partial}{\partial k} \sigma_{BS} \right|_{k=0}.$$

Thus, again using the European option prices under the rough Heston model with $\alpha = 0.6$ and under the classical Heston model, we find the behavior shown in the graph in *Figure 13*.

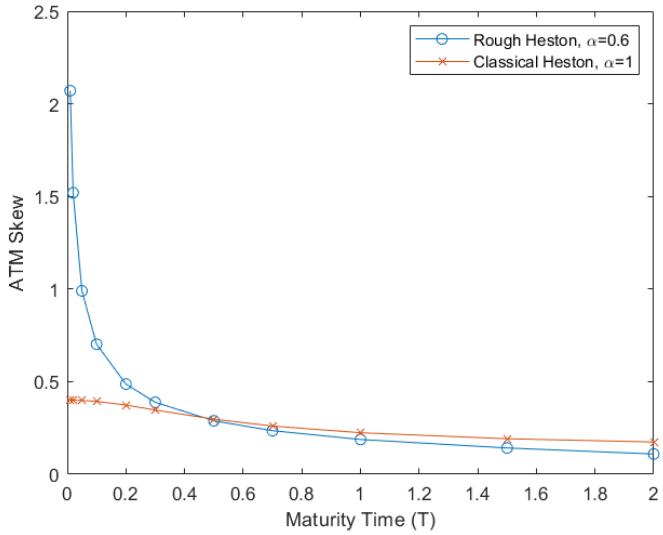


Figure (13) At-the-money skew behavior with changing maturity

As a result, observing *Figure 12 and 13*, European option prices achieved by the COS method give a consistent result with the rough volatility behavior described in Gatheral et. al.'s [19], implying that the behavior of the rough Heston model is also consistent with the real-life market data.

7 Calibration

In this section, the calibration process applied for the rough Heston model European option pricing is explained. In *Section 7.1*, a brief introduction to the calibration process in general is presented. In *Section 7.2*, the calibration method that we used, ANNs, are shown. In *Section 7.3*, the optimization methods that we used for the calibration is explained. Finally, in *Section 7.4*, the numerical results of the calibration process is presented.

7.1 A Brief Introduction to Calibration Procedure

One essential problem with the financial option pricing models is the fact that the model parameters that should be fitted to the real market option prices are unknown. The calibration procedure is the stage of the option pricing process that aims to solve this problem. It is a procedure which aims to estimate the model parameters with the available market option prices. Sometimes instead of the option prices, the related implied volatilities are used as the market data for the calibration procedure [31]. We use both as an output, but we are going to concentrate on the option prices in this section. The only change is the usage of implied volatility instead of the option prices as the output.

The market option prices can be seen as functions of specific log-moneyness, k, and time to maturity, $\tau = T - t$, and is denoted as $V^*(\tau, k)$, whereas the model option price is denoted as $V(\tau, k, \Theta)$ with Θ being the parameter vector. For the rough Heston model, we have the model parameters, $\Theta = [r, v_0, \rho, \theta, \gamma, \nu, \alpha]$.

The calibration procedure can be seen as an optimization problem where the distance between the market option price, $V^*(\tau, k)$, and the model option prices, $V(\tau, k, \Theta)$ is minimized. Thus, we need an error measure first. In [31], the difference between these two prices is used as an error measure given by,

$$e_i := ||V^*(\tau, k) - V(\tau, k, \Theta)||, \quad i = 1, .., N,$$

where N is the number of available market data that is used for the calibration procedure and $||.||$ is a measure of distance. Thus, following [31], the total difference is measured using a target function given by,

$$J(\Theta) := \sum_{i=1}^{N} w_i e_i + \lambda ||\Theta||,$$

where w_i is the weight, λ is the regularization parameter. One example for the target function is the mean squared error which is the case with $w_i = 1/N$, the measure of distance is $||.||_2$ and $\lambda = 0$. For the calibration procedure for option pricing models, in general the weights of the at the money options are higher since they are more liquid in the market [31].

As a result, the calibration procedure is a minimization problem as follows,

$$arg\min_{\Theta \in \mathbb{R}^n} J(\Theta)$$

with solutions giving the optimal parameter values that are minimizing the total difference between the market data and model option prices. This is an over-determined problem since $N > n$, so this problem is generally solved iteratively until a stopping criterion is met [31].

7.2 Artificial Neural Networks

Machine learning is becoming a popular approach as a computational tool. Thus, we used a machine learning based approach for solving the optimization problem of the calibration procedure. Following [31] and [32], Artificial Neural Networks (ANN) which are used to determine the optimal parameter values for the rough Heston model are presented in this section.

Let's first give a brief summary of the calibration procedure using ANNs step by step [31]:

- Generate a data set for the input parameters for the financial model,
- Using the option pricing model compute the corresponding option prices which are the output,
- With this data set train the ANNs, this is called the training phase,
- Evaluate the performance of the ANN, this is called the prediction phase,
- Using the market data as the input and doing a backwards pass with the ANN, estimate model parameters, this is called the calibration phase.

In a standard ANN structure, there are three main components which are the layers, neurons and the architecture of the network. The architecture determines how the layers are located, and the layers consist of multiple neurons. Throughout the training process, the weights and biases assigned to a neuron are modified which made them 'learnable'. Thus, a neuron is the fundamental unit of the ANN [32].

The layers are ordered one by one, and the adjacent layers are connected to pass a signal. Thus, the output signal of a previous layer is the input signal for the next one. There are an input layer, some hidden layers (might be one) and an output layer, in which the signal may circuit in a cyclic or recurrent way [32].

The simplest version of the ANN is the multi-layer perceptron (MLP). A MLP has at least three layers which are the input layer, the hidden layer and the output layer. In our case, there are the rough Heston model parameters in the input parameter, and the output signal of the output layer is the option price

or the corresponding implied volatility. The parameter set of a MLP is defined as,

$$\Theta' = (W_1, B_1, W_2, B_2, ..., W_N, B_N),$$

where W_i is the weight matrix and B_i is the bias vector of the i-th layer and N is the number of layers [32]. Thus, if we call the input of the ANN as x, and the output of the ANN as y, we have the following function,

$$y(x) = F(x|\Theta').$$

Now, we need to show what happens inside of a neuron. Let's call the output signal of the i-th neuron in the n-th layer as z_i^n. Thus, we have the following equation,

$$z_i^n = \varphi^n\left(\sum_j w_{ji}^n z_j^{n-1} + b_i^n\right),$$

where w_{ji}^n and b_i^n are the corresponding weight and bias in the weight matrix and bias vector, and $\varphi(.)$ is called the activation function [32]. When $n = 0$, we have the input layer and $n = N$, we have the output layer, otherwise we are at the hidden layers. The aim of having an activation function is adding non-linearity to the ANN and some examples of the activation function function are as follows [32],

- Sigmoid, $\varphi(x) = \frac{1}{1+e^{-x}}$,
- ReLu, $\varphi(x) = max(x, 0)$,
- Leaky ReLu, $\varphi(x) = max(x, ax), \;\; 0 < a < 1$.

An illustration of the neuron explained above is shown in *Figure 14* [32].

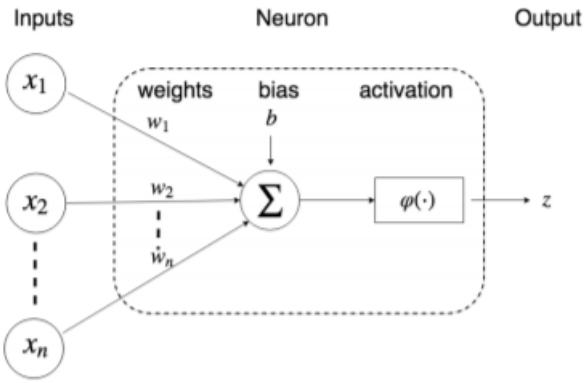


Figure (14) Neuron illustration

An example of the MLP is the one with single hidden layer. The formula for such MLP is given by,

$$\begin{cases} y = \varphi^2\Big(\sum_j w_j^2 z_j^1 + b^2\Big), \\ z_i^1 = \varphi^1\Big(\sum_j w_{ji}^1 x_j + b_i^1\Big), \end{cases}$$

and the illustration for this MLP can be seen in *Figure 15* [32].

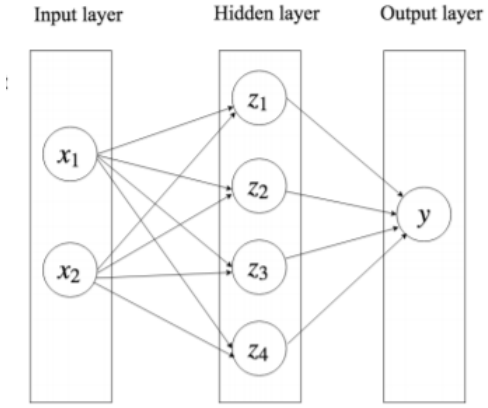


Figure (15) An example of MLP

There are two main phases for the calibration using ANNs. The first phase is called the forward pass in which the solution is learned, the second phase is called the backward pass in which the calibration takes place.

7.2.1 The Forward Pass

The training and the prediction phases are done in the forward pass. In this pass, the numerical method (the COS method in our case) that is used to price the options is learned by the ANN. The learning is done with the optimization of the MLP architecture, which is deriving optimal weights and biases, that gives a mapping from the input to the output according to the numerical method that is used [31].

To be more specific, we use the COS method to derive European option prices from the rough Heston parameters. ANN also learns to derive the option prices from the same rough Heston parameters using the learning methodology which is called the rough-Heston-ANN. Besides, instead of option prices, we may use the Black-Scholes implied volatilities as an output. We use Brent's method to derive the implied volatilities from the option prices, thus the ANN also learns

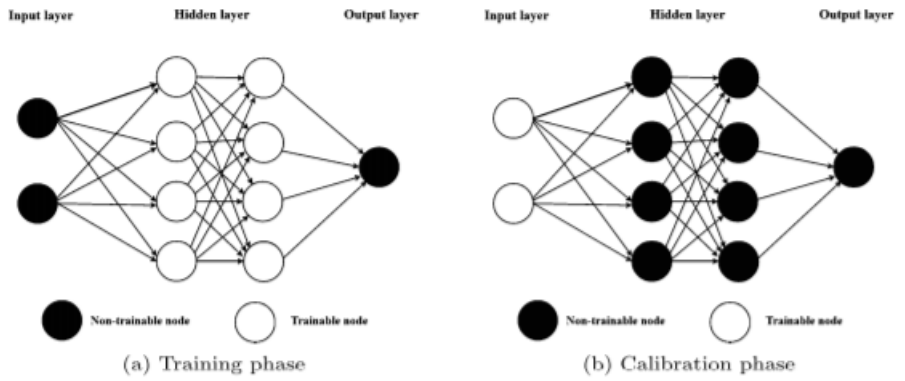


Figure (16) Trainable layers for the training and calibration phase [31]

the Brent algorithm which is called IV-ANN. When, we use the implied volatilities as the output in the calibration, we merge rough-Heston-ANN with the IV-ANN to use one ANN instead of two seperate ANNs. This is called rough-Heston-IV-ANN.

7.2.2 The Backward Pass

Until this section, we concentrated on the optimization of the ANN parameters like the weights and the bias inside the neurons. Since we are doing the calibration phase in this pass, the parameters that are optimized are the parameters of the rough Heston model. The corresponding objective function is the error measure between the market option prices and the model option prices (or implied volatilities instead of option prices) [31]. Hence, again, we need an optimization scheme for this pass.

During the training phase of the forward pass, the hidden layers are optimized to train the ANN to give an optimal output for the corresponding input. In the prediction phase, the hidden layers of the ANN are frozen, since this phase is to evaluate the performance of the ANN [31].

During the calibration phase, the hidden layers are again frozen, but the input layer becomes the learnable layer as can be seen in *Figure 16*. The hidden layers are already trained in the forward pass, thus to get the optimal rough Heston parameters, the input layer should be trained. As a result, from the output (option prices or implied volatilities), we may get an optimal set of rough Heston parameters in this phase. It should also be noted that the training phase is computationally expensive, but done only once and off-line, whereas, the calibration phase is computationally cheap and done on-line [31].

7.3 Optimization

Both in the training phase and the calibration phase, optimization plays a significant role for the ANNs. However, the training phase and the calibration phase have different objectives and requirements, thus we are using different optimization methods for both phases. For the training phase, Stochastic Gradient Descent is used as an optimization method, whereas, for the calibration phase the global optimizer Differential Evolution is used.

7.3.1 Stochastic Gradient Descent

In the training phase of the ANN, we have a high-dimensional, non-convex optimization problem due to the large numbers of weights in the neurons that need to be optimized. Thus, Stochastic Gradient Descent is a popular choice for the training phase of the ANN. We do not need a global optimum for this phase, since it is enough to get a sufficiently low cost function value. Thus, a high-quality local minima is not problematic.

In the Stochastic Gradient Descent algorithm, the optimization is done with [31, 32],

$$W^{(i+1)} \leftarrow W^{(i)} - \eta(i)\frac{\partial L}{\partial W},$$

$$b^{(i+1)} \leftarrow b^{(i)} - \eta(i)\frac{\partial L}{\partial b},$$

$$\text{for i=}0, 1, ..., N_T,$$

where L is the loss function (distance measure between the target value and the output), η is called the learning rate and N_T is the number of training iterations. Using this algorithm, the optimal values for the weights W and biases b are computed. Also, we need to note that there are different versions of the Stochastic Gradient Descent, e.g., RMSprop and Adam [31].

7.3.2 Differential Evolution

Differential Evolution is an optimization process which gives a global optimum. This is important for the calibration phase, since a specific set of parameters is the objective of the calibration. There is no requirement for the specific initial values, and the algorithm is population based without any derivatives [31].

There are four steps in the algorithm of Differential Evolution that we used [31]:

1. Initialization: Determine a lower and upper bound for the rough Heston parameters. Then, generate a population with N_P parameter vectors with random positions between the lower and upper bounds,

$$(\Theta_1, \Theta_2, ..., \Theta_{N_P}).$$

2. Mutation: After initialization, a mutant candidate Θ_i' is derived for every parameter vector as follows,

$$\Theta_i' = \Theta_a + F.(\Theta_b - \Theta_c),$$

where Θ_i is the i-th candidate, Θ_b and Θ_c are randomly picked from the population and Θ_a is the best candidate from the previous population. $F \in [0, \infty]$ is called the differential weight and it is used to determine the step size of the evolution. Thus, large F provides a larger search radius, but it may give rise to slow convergence. The mutant candidates are called the donor candidates and they create a new intermediate population.

3. Crossover: In this stage, a random pick of the mutated candidates that are allowed in the next stage is chosen. For every $i \in \{0, 1, ..., N_P\}$ a random probability, p_i is used. Also, a crossover probability, Cr, is determined. Then,

$$\Theta_i'' = \begin{cases} \Theta_i', & p_i \leq Cr, \\ \Theta_i, & \text{otherwise.} \end{cases}$$

If a donor candidate has a probability larger than crossover probability, then it will be eliminated.

4. Selection: The left donor candidate, Θ_i'', after the crossover stage, is compared with the initial Θ_i according to their performances for the objective function. Hence, if the donor candidate has an improved performance, then $\Theta_i \leftarrow \Theta_i''$ for the next generation.

The stages 2-4 are repeated creating new generations, until a pre-decided stopping criterion is satisfied.

7.4 Numerical Results

Before showing the numerical results for the calibration procedure, the settings used for the ANN method is presented.

7.4.1 Calibration Settings

ANN	Parameters	Value Range	Generating Method
ANN Input	$m = K/S_0$	[0.6, 1.4]	LHS
	Time to maturity, τ	[0.05, 3.0](year)	LHS
	r	[0.0, 0.10]	LHS
	ρ	[-0.90, 0.0]	LHS
	γ	(0.01, 3.0]	LHS
	ν	(0.01, 1.0]	LHS
	θ	(0.15, 0.50]	LHS
	ξ	(0.5, 1.0]	LHS
	α	(0.5, 1.0)	LHS
ANN Output	European put price, V	$(0, 0.6)$	COS

Table (5) Sampling range for the rough-Heston parameters; LHS means Latin Hypercube Sampling, COS stands for the COS method.

Parameters	Setting
Hidden layers	4
Neurons(each layer)	200
Activation	ReLu
Optimizer	Adam

Table (6) The configuration of the ANN.

Parameter	Setting
Population size	60
Mutation Strategy	Best candidate
Mutation	(0.5, 1.0)
Crossover recombination	0.7
Convergence tolerance	0.01

Table (7) The setting of DE

The objective function used for rough Heston calibration is:

$$J(\Theta) = \sum (V_{ANN} - V_{Market})^2$$

From some initial experiments done by only calibrating 4 parameters (ξ, γ and r are fixed), some important conclusions that may be useful for future research is presented:

- There are some cases with multiple solutions. That is because movements of two different parameters may have similar effects on the option price. Thus, there may not be a global minimum over some error threshold.
- θ is the most sensitive parameter, so it tends to dominate the calibration accuracy.
- α is more sensitive than γ and ρ.
- The effect of ρ to the option price is limited. Thus, a change in ρ has small changes on the option price.

7.4.2 Calibration Test

To evaluate the performance of the ANN calibration for the rough Heston model, we performed a test of ANN calibration with the rough-Heston-ANN and compared its performance with a calibration process directly using the Differential Evolution algorithm (which we call MATLAB-DE). This test is done using the European option prices, not implied volatilities.

Since the aim of this test is evaluating the performance of the ANN calibration, instead of real market option price data, we used a set of option prices derived from the COS method for a pre-determined parameter set for every other case and assigned them as the market prices. The parameter set is $[\gamma, \theta, \nu, \rho, \xi, \alpha]$.

The market option prices are derived from these parameters for different $m = K/S_0$, τ and interest rate values. Thus, the result of the calibration procedure is estimations of these six parameters. After performing 729 calibration cases, the following results are achieved. The average calibration time for this test is 0.6055 seconds.

Parameter	Average Deviation from the Market Value
γ	0.2091
θ	0.0199
ν	0.1416
ρ	0.1038
ξ	0.0636
α	0.0721

Table (8) Average deviation of the calibrated parameters from the pre-determined market parameter

Now, we are going to examine only one specific case of these calibration cases. The parameter set is:

$$\gamma = 0.2,\ \theta = 0.15,\ \nu = 0.3,\ \rho = -0.75,\ \xi = 0.6,\ \alpha = 0.7.$$

The calibration process took 0.4562 seconds for this case and gave the following results:

Parameter	Market value	Calibration Result	Absolute Difference
γ	0.2	0.5672	0.3672
θ	0.15	0.1123	0.0377
ν	0.3	0.1298	0.1702
ρ	-0.75	-0.5037	0.2463
ξ	0.6	0.7982	0.1982
α	0.7	0.7786	0.0786

Table (9) Calibrated parameters of one specific case.

The results in *Table 8* and *Table 9* agrees with the conclusions we got from the initial experiments. It can be seen that the calibrated parameters are not perfect fit for the pre-determined market parameters. However, the important thing is the performance of the calibrated parameters in computing the option prices. Thus, we need to compare the option prices resulting from the pre-determined market parameters with the option prices resulting from the calibrated parameters. Hence, the option prices for different τ, $m = K/S_0$ and interest rate values are computed for both of the parameter sets. For every other 25 different cases, the absolute difference between the market and model option prices is smaller than $5e-4$. This results in *Figure 17*.

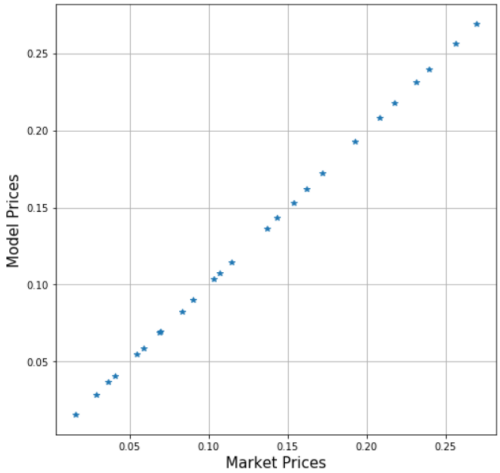


Figure (17) Option price comparison for pre-determined market parameters and calibrated parameters. They are called market prices and model prices, respectively.

The same calibration process is also done using directly the Differential Evolution algorithm (MATLAB-DE) instead of the rough-Heston-ANN to evaluate

the rough-Heston-ANN's performance. The Differential Evolution algorithm is applied in the MATLAB-DE as it is explained in *Section 7.3.2*. Only for the mutation strategy, instead of the using the best initial parameter, a random parameter is used, since this strategy gives better convergence. The time of the calibration process using the MATLAB-DE algorithm varies from 6 to 10 hours for the rough Heston parameters. This specific case took around 8.5 hours and gave the following results:

Parameter	Market value	Calibration Result	Absolute Difference
γ	0.2	0.2611	0.0611
θ	0.15	0.1504	0.0004
ν	0.3	0.6890	0.3890
ρ	-0.75	-0.2729	0.4771
ξ	0.6	0.6009	0.0009
α	0.7	0.5991	0.1009

Table (10) Calibration with Differential Evolution algorithm.

Looking at *Table 10*, it can be seen that even though for parameters γ, θ and ξ, the calibration is a good fit, for general, there is not a significant difference in the accuracy comparing to the rough-Heston-ANN. Similar to what we do with ANN calibration, the option prices for different cases are computed. Again the difference of market and model option prices for every other case is smaller than $5e-4$, and *Figure 18* is got.

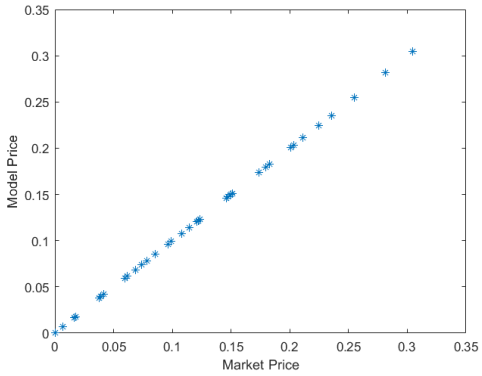


Figure (18) Option price comparison for pre-determined market parameters and calibrated parameters with Differential Evolution.

According to these results, there is not a significant change between the calibration accuracy of the rough-Heston-ANN and the MATLAB-DE algorithm. However, there is a huge difference between their computational times, which implies that using ANN method instead of using directly the Differential Evolution algorithm for the rough Heston parameter calibration is more advantageous.

7.4.3 Implied Volatility Behavior

In *Section 6.2*, the implied volatility behavior of the rough Heston model is examined. It is observed that the implied volatility behavior of the rough Heston model is consistent with the market implied volatility behavior. In this section, this result is going to be tested with calibrated parameters from the real market data. For the calibration, DAX option prices of July 22, 2020 (taken from eurexchange.com/exchange-en/markets/idx/dax/DAX-Options-139884) are used.

First, the implied volatility smile behaviors of the market and the rough Heston model are compared for different maturity times. The model values are computed using the calibrated parameters from the rough-Heston-ANN algorithm. The results are presented in *Figure 19*.

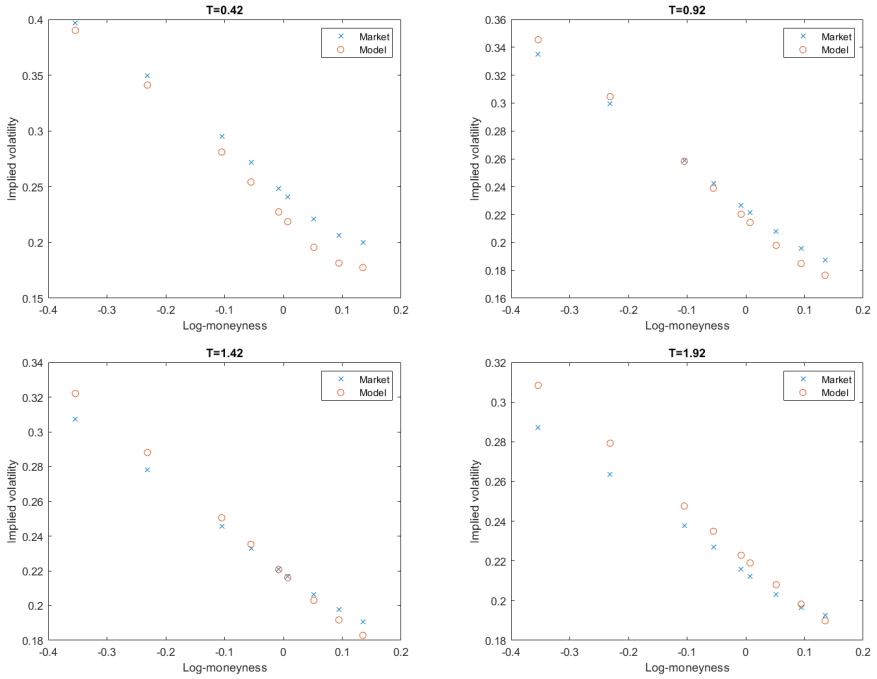


Figure (19) Market vs. Model implied volatility smiles using the rough Heston-ANN calibration

In *Figure 19*, we can see that the calibrated parameters of the rough Heston model is able to generate a reasonable fit for the smile behavior of the market implied volatility.

For *Figure 19*, the parameters are calibrated with the rough-Heston-ANN algorithm. We also produce the same results for the parameter calibration with the rough-Heston-IV-ANN in *Figure 20*.

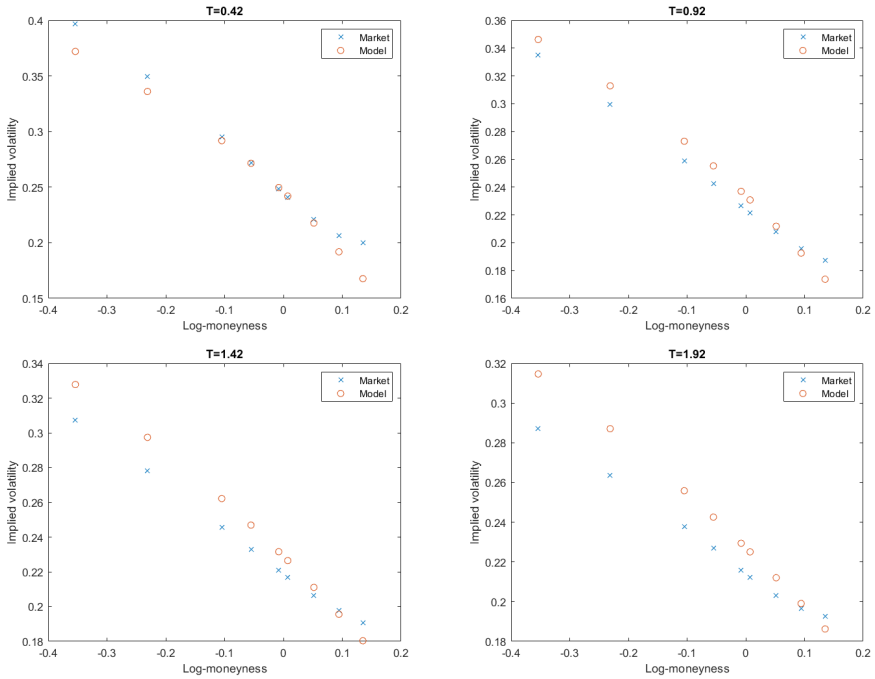


Figure (20) Market vs. Model implied volatility smiles with calibration based on implied volatility data.

Looking at *Figure 19* and *Figure 20*, we see that the implied volatility values of the rough-Heston-ANN calibration is a better fit numerically. Also, especially for $T = 0.42$ and $T = 0.92$, the shape of the smile for the rough-Heston-ANN case is better at reflecting the behavior of the market values. Thus, it can be said that the rough-Heston-ANN calibration outperforms rough-Heston-IV-ANN calibration for the DAX data we use. Thus, we continue to use the rough-Heston-ANN calibration for the next results.

The second critical market implied volatility behavior is the ATM skew behavior that is explained in *Section 6.2*. The comparison of the market ATM skew and the rough Heston model ATM skew is presented in *Figure 21.* From this comparison, it can be seen that both the market data and the rough Heston model under calibrated parameters show the expected ATM skew behavior. The rough Heston model also generates a reasonable fit for the DAX market ATM implied volatility skew.

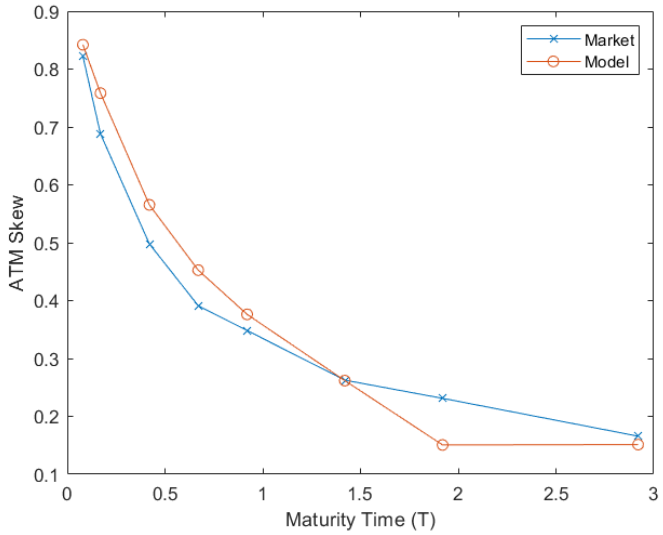


Figure (21) At-the-money skew of Market vs. Model.

Under the same calibration, by fixing other calibrated parameters and changing only α, we aim to show the effect of different α values on the ATM skew behavior. For the DAX option prices, the calibrated alpha is relatively close to 1. In *Figure 22*, we present the behavior of the case with α close to 0.5 (the 'rougher' case) and the case with $\alpha = 1$ (the classical Heston model case).

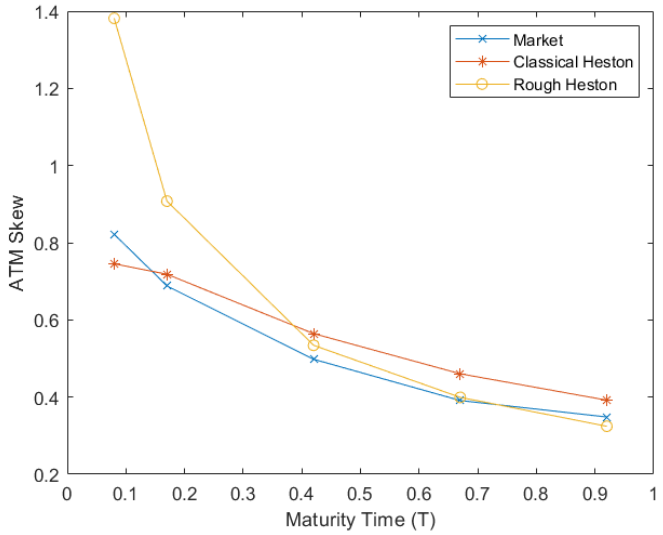


Figure (22) At-the-money skew behavior for different α values.

From *Figure 22*, we see that for larger maturity times the effect of α is limited. However, the critical advantage of the 'rougher' case can be seen for small maturity times. Even though, the ATM skew for the classical Heston is numerically more close to the market ATM skew (this is due to the fact that the calibrated α is closer to 1), both the market ATM skew and the rough Heston ATM skew tends to go higher, while the classical Heston ATM skew tends to converge around 0.8. As explained in [19], we see the tendency to go to infinity as maturity time goes to zero for the average market ATM skew. Hence, the rough Heston model is more advantageous compared to the classical Heston at reflecting this property of the average market ATM implied volatility skew behavior.

8 Conclusion

Proposing an alternative method to price European options under the rough volatility model called the rough Heston model is the main purpose in this thesis. The COS method introduced in [15] is a Fourier-based pricing method that uses essentially the characteristic function of the option pricing model. The rough Heston model is an option pricing model based on the properties of the micro-interactions of the electronic trading market as shown in [13] and its characteristic function is derived in [14]. Hence, the COS method is a good fit for the option pricing under the rough Heston model and an efficient and accurate European option pricing process is constructed throughout this thesis.

In *Section 6.1*, an experiment evaluating the performance of the COS method European option pricing process that is proposed in this thesis is done. The numerical techniques presented in *Section 5.1, 5.2 and 5.3*, is used for this experiment. The performance of the COS method under the rough Heston model is tested for different roughness levels of the volatility, for both put and call options, for different maturity times and for different strike prices. As a result of this experiment, it is showed that the pricing process gives coherent and accurate results with reasonable time costs. Also, an experiment to evaluate the convergence properties is performed. It is observed that within this reasonable time costs, the COS method pricing results tend to converge. Hence, as a result of these experiments, it can be said that the COS method is a promising alternative for European option pricing under the rough Heston model.

Besides the COS method, a Monte-Carlo scheme is also proposed in *Section 5.4*. The aim of this Monte-Carlo scheme is to derive a confidence interval for European option prices under the rough Heston model that is going to be used as a reference to evaluate the accuracy of the other pricing methods, e.g. the COS method. It is showed that it has a good enough accuracy and convergence property to use as a reference, especially for ATM options. However, a loss of accuracy is also observed for different strike price values and it is computationally expensive to get a narrow confidence interval. Thus, it is open to improvements.

The rough volatility models are complex and harder to work on, but they have some better qualities over other option pricing models as explained in *Section 2*. First, the real market data for option prices implies that the volatility behavior is rough [19]. Also, as shown in *Section 2.2*, the implied volatility behavior of the rough volatility models has a great resemblance to the market implied volatility behavior. Hence, it can be said that the rough volatility models tend to reflect the real market behavior. In *Section 6.2*, another experiment is performed in order to examine if the rough Heston model shows the rough volatility properties under the COS method pricing proposed in this thesis. As a result of this experiment, the desirable market behavior of the rough volatility models resembling the market behavior is achieved.

Finally in *Section 7*, it is showed that using Artifical Neural Networks for the calibration of the rough Heston model parameters is a preferable method. It is a computationally cheap method after the learning process and it has also a good accuracy for the calibrated parameters. Performing the calibration for DAX market option prices, it is also showed that the rough Heston model option prices has a consistent implied volatility behavior with the market option prices.

8.1 Further Research

In this section some further research is recommended for the option pricing under the rough Heston model.

1. This thesis concentrates only on the European option pricing under the rough Heston model using the COS method. However, pricing of the other type of options, e.g. American options, Asian options, exotic options, etc. for the COS method may be performed in the future. The pricing process for these type of options are not trivial.

2. The pricing PDE of the rough Heston model can be derived. Introducing option pricing methods based on the PDE of the model for the rough Heston model have a potential to increase the accuracy and efficiency of the option pricing under the rough Heston model.

3. The mathematical background of the Monte-Carlo scheme proposed in this thesis should be examined more. As a result of that, the convergence behavior could be understood better or a modified version of the Monte-Carlo scheme can be proposed. This new scheme may be a more robust and stable version with better convergence properties.

4. Besides of the Monte Carlo scheme proposed in this thesis, a whole new and better Monte-Carlo scheme can be proposed. Throughout the research for the thesis, it is seen that there is a lack of research on the option pricing models for the rough Heston model. Our intention was to use an already proposed Monte Carlo scheme for this thesis, but it could not be found.

References

[1] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[2] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[3] E. Bacry, T. Jaisson, and J.-F. Muzy. Estimation of slowly decreasing hawkes kernels: Application to high frequency order book modelling. *Quantitative Finance*, 16(8):1179–1201, 2016.

[4] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[5] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Decoupling the short and long term behavior of stochastic volatility. *Available at SSRN 2846756,*, 2016.

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

[7] J. Brasseur. On restricitions of besov functions. 2018.

[8] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

[9] K. Diethelm, N. J. Ford, and A. D. Freed. Detailed error analysis for a fractional adams method. *Numerical Algorithms*, 36(1):31–52, 2004.

[10] K. Diethelm, N. J. Ford, A. D. Freed, and Y. Luchko. Algorithms for the fractional calculus: A selection of numerical methods. *Computer Methods in Applied Mechanics and Engineering*, 194(6-8):743–773, 2005.

[11] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

[12] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[13] O. E. Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2019.

[14] O. E. Euch and M. Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[15] F. Fang and C. W. Oosterlee. A novel pricing method for european option based on fourier-cosine series expansions. *SIAM Journal of Scientific Computing*, 31(2):826–848, 2008.

[16] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the hawkes self-excited point process model: Application to high-frequency financial data. 2014.

[17] M. Fukasawa. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[18] J. Gatheral. *The volatility surface: A practitioner's guide*. John Wiley & Sons, 2006.

[19] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.

[20] S. Gerhold, C. Gerstenecker, and A. Pinter. Moment explosions in the rough heston model. *Decisions in Economics and Finance*, 42:575–608, 2019.

[21] L. Grzelak and C. Oosterlee. *Mathematical modelling and computation in finance: With exercices and Python and MATLAB computer codes*. World Scientific, 2020.

[22] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *Wilmott Magazine*, 1:84–108, 2002.

[23] A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443, 1971.

[24] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[25] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

[26] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

[27] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy-tailed hawkes processes. *The Annals of Applied Probability*, 2016.

[28] A. Jakubowski, J. Mémin, and G. Pages. Convergence en loi des suites d'intégrales stochastiques sur l'espace d1 de skorokhod. *Probability Theory and Related Fields*, 81(1):111–137, 1989.

[29] P. Laub, T. Taimre, and P. K. Pollett. Hawkes processes. 2015.

[30] G. Leoni. *A first course in Sobolev spaces: Second edition*. American Mathematical Society, 2017.

[31] S. Liu, A. Borovykh, L. Grzelak, and K. Oosterlee. A neural network-based framework for financial model calibration. *Journal of Mathematics in Industry*, 9(1):1–28, 2019.

[32] S. Liu, C. W. Oosterlee, and S. M. Bohte. Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):1–22, 2019.

[33] G. Livieri, S. Mouti, A. Pallavicini, and M. Rosenbaum. Rough volatility: evidence from option prices. *IISE Transactions*, pages 1–21, 2018.

[34] B. Mandelbrot and J. W. V. Ness. Fractional brownian motions, fractional noises and application. *SIAM Review*, 10(4):422–437, 1968.

[35] T. Marinov, N. Ramirez, and F. Santamaria. Fractional integration toolbox. *Fractional Calculus and Applied Analysis*, 16(3):670–681, 2013.

[36] B. Noble. *Methods based on the Wiener-Hopf technique for the solution of partial differential equations*. Pergamon Press, Belfast, Northern Ireland, 1958.

[37] M. Rosenbaum. A new microstructure noise index. *Quantitative Finance*, 11(6):883–899, 2011.

[38] R. Tayal and S. Thomas. Measuring and explaining the asymmetry of liquidity. *Available at SSRN 2239492*, 2012.

[39] P. Wilmott. *Paul Wilmott introduces quantitative finance*. John Wiley & Sons, 2007.

Appendices

Appendix A Proof of Theorem 4.1

For the proof of Theorem 4.1, we follow the proof that is presented in [14]. First, the C-tightness (C for space of continuous functions) of (X^T, Λ^T, Z^T) that are defined in *Section 4.3* is shown. Then, we look at the convergence of X^T and Z^T.

A.1 C-Tightness of (X^T, Λ^T, Z^T)

We know that,

$$\lambda_t^{T,+} = \lambda_t^{T,-} = \hat{\mu}_T(t) + \int_0^t \Psi^T(t-s)\hat{\mu}_T(s)ds + \frac{1}{1+\beta}\int_0^t \Psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}),$$

with M_t^T being the martingale defined in *Section 4.2*. Using the fact that M^T is a martingale and $\int_0^{\cdot}(f * g) = (\int_0^{\cdot} f) * g$, we see that

$$\mathbb{E}[N_T^{T,+}] = \mathbb{E}[N_T^{T,-}] = \mathbb{E}[\int_0^T \lambda_s^{T,+} ds] = \int_0^T \hat{\mu}_T(s)ds + \int_0^t \Psi^T(t-s)(\int_0^s \hat{\mu}_T(u)du)ds.$$

We can also show that

$$1 + \int_0^\infty \Psi^T(s)ds = 1 + \sum_{k\geq 1}\int_0^\infty (\varphi^T)^{*k} = \sum_{k\geq 0}(a_T)^k = \frac{T^\alpha}{\gamma},$$

and we know that $\hat{\mu}$ is a positive function, thus,

$$\mathbb{E}[N_T^{T,+}] \leq \int_0^T \hat{\mu}_T(s)ds(1 + \int_0^\infty \Psi^T(s)ds) \leq \frac{T^{\alpha+1}}{\gamma}\int_0^1 \hat{\mu}_T(Ts)ds.$$

From the definition of the baseline intensity, it is trivial to see that

$$\int_0^1 \hat{\mu}_T(Ts)ds \leq cT^{\alpha-1},$$

with c being a positive constant. Thus, $\mathbb{E}[N_T^{T,+}] \leq cT^{2\alpha}$ and,

$$\mathbb{E}[X_1^T] = \mathbb{E}[\Lambda_1^T] \leq c,$$

for each component. Also, we should note that every component is increasing, thus have the same behavior in the context of tightness. The size of the maximum jump for both X^T and Λ^T is $\frac{1-a_T}{T^\alpha \mu}$, which goes to zero. As a result of these, it can be stated that X^T and Λ^T have C-tightness property.

For Z^T, we know that,

$$\langle Z^T, Z^T \rangle = diag(\Lambda^T),$$

which implies that Z^T is C-tight.

It is also important to show that, since,

$$X_t^T - \Lambda_t^T = \frac{1 - a_T}{T^\alpha \mu} M_{tT}^T,$$

by definition. From Doob's inequality we can show that,

$$\mathbb{E}[sup_{t\in[0,1]} |X_t^T - \Lambda_t^T|^2] \leq cT^{-4\alpha}\mathbb{E}[M_T^T]^2.$$

Thus, using $[M^T, M^T] = N^T$, it is easy to see that,

$$\mathbb{E}[sup_{t\in[0,1]} |X_t^T - \Lambda_t^T|^2] \leq cT^{-4\alpha}\mathbb{E}[N_T^T] \leq cT^{-2\alpha}.$$

As a result, we see that $X_t^T - \Lambda_t^T$ has a uniform convergence to zero in probability. One other important note is that, as explained in [14], (X, Z) being a possible limit of (X^T, Z^T), Z is a continuous martingale with $[Z, Z] = diag(X)$.

We conclude this section with the following proposition.

Proposition A.1: *The sequence (X^T, Λ^T, Z^T) is C-tight with*

$$sup_{t\in[0,1]} ||\Lambda_t^T - X_t^T|| \underset{T\to\infty}{\to} 0,$$

in probability. Also Z is a continuous martingale with $[Z, Z] = diag(X)$.

A.2 Convergence of X^T and Z^T

Before going into the convergence, we first show that, since,

$$\Lambda^{T,+} = \Lambda^{T,-},$$

and,

$$sup_{t\in[0,1]} ||\Lambda_t^T - X_t^T|| \underset{T\to\infty}{\to} 0,$$

we have,

$$sup_{t\in[0,1]} ||X_t^{T,+} - X_t^{T,-}|| \underset{T\to\infty}{\to} 0.$$

This implies that the convergence X of $X^{T,+}$ is also the convergence of $X^{T,-}$. Now, we first look into a convenient equality that can be checked using *Lemma 4.1*. It is also important to note that $\hat{\mu}_T$ is considered unknown for now. The equality is,

$$\hat{\mu}_T(t) + \int_0^t \Psi^T(t-s)\hat{\mu}_T(s)ds = \mu_T + \xi\mu_T \frac{1}{1-a_T} + \mu_T(1-\xi)\int_0^t \Psi^t(t-s)ds.$$

Using convolution with φ^T for both sides of the equality and using the fact that $\Psi^T * \varphi^T = \Psi^T - \varphi^T$, the left hand side of the equality becomes as follows,

$$\begin{aligned}
&\int_0^t \hat{\mu}_T(s)\varphi^T(t-s)ds + \int_0^t \int_0^s \Psi^T(s-u)\hat{\mu}_T(u)du\varphi^T(t-s)ds \\
&= \int_0^t \hat{\mu}_T(s)\varphi^T(t-s)ds + \int_0^t (\Psi^T(t-u) - \varphi^T(t-u))\hat{\mu}_T(u)du \\
&= \int_0^t \Psi^T(t-s)\hat{\mu}_T(s)ds.
\end{aligned}$$

The right hand side becomes,

$$\begin{aligned}
&\int_0^t \varphi^T(t-s)(\mu_T + \xi\mu_T \frac{1}{1-a_T})ds + \mu_T(1-\xi)\int_0^t \varphi^T(t-s)\int_0^s \Psi^T(s-u)duds \\
&= (\mu_T + \xi\mu_T \frac{1}{1-a_T})\int_0^t \varphi^T(t-s)ds + \mu_T(1-\xi)\int_0^t (\Psi^T(t-u) - \varphi^T(t-u))du.
\end{aligned}$$

As a result, the equality becomes,

$$\int_0^t \Psi^T(t-s)\hat{\mu}_T(s)ds = \mu_T\xi(1+\frac{1}{1-a_T})\int_0^t \varphi^T(t-s)ds + \mu_T(1-\xi)\int_0^t \Psi^T(t-s)ds.$$

Thus, the only possible solution for the baseline intensity $\hat{\mu}_T$ is,

$$\hat{\mu}_T(t) = \mu_T + \xi\mu_T\Big(\frac{1}{1-a_T}\Big(1 - \int_0^t \varphi^T(s)ds\Big) - \int_0^t \varphi^T(s)ds\Big).$$

Now, we can apply the baseline intensity to $\lambda_t^{T,+}$ from *Section 4.2* as follows,

$$\begin{aligned}
\lambda_t^{T,+} = \mu_T + \mu_T \int_0^t \Psi^T(t-s)ds + \xi\mu_T\Big(\frac{1}{1-a_T} - \int_0^t \Psi^T(t-s)ds\Big) + \\
\frac{1}{\beta+1}\int_0^t \Psi^T(t-s)(dM_s^{T,+} + \beta dM_s^{T,-}).
\end{aligned}$$

Then, using Fubini's theorem with the fact that $\int_0^{\cdot}(f * g) = (\int_0^{\cdot} f) * g$, we have,

$$\begin{aligned}
\int_0^t \lambda_s^{T,+}ds = \mu_T t + \mu_T \int_0^t \Psi^T(t-s)sds + \xi\mu_T\Big(\frac{t}{1-a_T} - \int_0^t \Psi^T(t-s)sds\Big) + \\
\frac{1}{\beta+1}\int_0^t \Psi^T(t-s)(M_s^{T,+} + \beta M_s^{T,-})ds.
\end{aligned}$$

Moreover, for $t \in [0,1]$, we have the following decomposition,

$$\Lambda_t^{T,+} = T_1 + T_2 + T_3,$$

where,

$$T_1 = (1 - a_t)t,$$

$$T_2 = \int_0^t T(1 - a_T)\Psi^T(T(t - s))sds + \xi\Big(t - T(1 - a_T)\int_0^t \Psi^T(T(t - s))sds\Big),$$

and,

$$T_3 = \frac{1}{\sqrt{\gamma\mu(1 + \beta)^2}}\int_0^t T(1 - a_T)\Psi^T(T(t - s))(Z_s^{T,+} + \beta Z_s^{T,-})ds.$$

We know that,

$$T(1 - a_T)\Psi(T.) = a_T f^{\alpha,\gamma},$$

then we can see that,

$$T_2 \underset{T\to\infty}{\to} \int_0^t f^{\alpha,\gamma}(t - s)sds + \xi(t - \int_0^t f^{\alpha,\gamma}(t - s)sds),$$

and,

$$T_3 \underset{T\to\infty}{\to} \frac{1}{\sqrt{\gamma\mu(1 + \beta)^2}}\int_0^t f^{\alpha,\gamma}(t - s)(Z_s^+ + \beta Z_s^-)ds.$$

Therefore, using *Proposition A.1*, we have,

$$X_t = \int_0^t f^{\alpha,\gamma}(t-s)sds+\xi(t-\int_0^t f^{\alpha,\gamma}(t-s)sds)+\frac{1}{\sqrt{\gamma\mu(1 + \beta)^2}}\int_0^t f^{\alpha,\gamma}(t-s)(Z_s^+ + \beta Z_s^-)ds.$$

Similarly with theorem 3.2 in [27], we know that,

$$X_t = \int_0^t Y_s ds,$$

with Y satisfying,

$$Y_t = F^{\alpha,\gamma}(t) + \xi(1 - F^{\alpha,\gamma}(t)) + \frac{1}{\sqrt{\gamma\mu(1 + \beta)^2}}\int_0^t f^{\alpha,\gamma}(t - s)(dZ_s^+ + \beta dZ_s^-).$$

From *Proposition A.1*, we have,

$$[Z, Z] = \int_0^t Y_s ds \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and we know the existence of the bidimensional Brownian motion (B^1, B^2) with,

$$Z_t^+ = \int_0^t \sqrt{Y_s}dB_s^1, \quad Z_t^- = \int_0^t \sqrt{Y_s}dB_s^2.$$

Hence, we define the Brownian motion B as follows,

$$B = \frac{B^1 + \beta B^2}{\sqrt{1 + \beta^2}}.$$

Finally, Y_t satisfies,

$$Y_t = F^{\alpha,\gamma}(t) + \xi(1 - F^{\alpha,\gamma}(t)) + \sqrt{\frac{1 + \beta^2}{\gamma\mu(1 + \beta)^2}}\int_0^t f^{\alpha,\gamma}(t - s)\sqrt{Y_s}dB_s.$$

A.3 Final part of the proof

In [14], there is the following proposition with its proof in [13].

Proposition A.2: *Let γ, ν, θ and V_0 be positive constants, $\alpha \in [1/2, 1]$ and B be a Brownian motion. The process V is a solution of the following fractional stochastic differential equation:*

$$v_t = v_0(1 - F^{\alpha,\gamma}(t)) + \theta F^{\alpha,\gamma}(t) + \nu \int_0^t f^{\alpha,\gamma}(t-s)\sqrt{v_s}dB_s,$$

if and only if it is a solution of

$$v_t = v_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\gamma(\theta - v_s)ds + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}\gamma\nu\sqrt{v_s}dB_s.$$

Furthermore, both equations admit a unique strong solution.

Thus, combining this proposition with the convergence of X^T, Z^T and *Proposition A.1*, we prove *Theorem 4.1*.

Appendix B Proof of Theorem 4.2

The proof of *Theorem 4.2* follows the proof in [14]. The proof is done for $t \in [0,1]$, but it can be extended for any $t \geq 0$. In this proof, first the process $Y^T(a,t) - (1,1)$ is controlled, then the convergence of this process is showed, and finally the proof is finished with some last steps.

We are not going to show the controlling process in this thesis, because the flow of the proof is not relevant for us. But, the resulting proposition of this step will be presented.

Proposition B.1: *For any* $t \in [0,1]$,

$$T^{\alpha}||Y^T(a,t) - (1,1)|| \leq c(a),$$

where $c(a)$ *is a positive constant independent from* t *and* T.

B.1 Convergence of $T^{\alpha}(Y^T(a,t) - (1,1))$

Before going into the convergence, let's define $\kappa = \frac{\gamma\theta}{2\mu}$. It is known that the complex logarithm is analytic on the set $\mathbb{C}/\mathbb{R}^-$, then there exists $c > 0$ such that for any $x \in \mathbb{C}$ with $|x| < 1/2$,

$$|log(1+x) - x + \frac{1}{2}x^2| \leq c|x|^3.$$

Therefore, we have the following,

$$log(Y^T(a,t)) = Y^T(a,t) - (1,1) - \frac{1}{2}(Y^T(a,t) - (1,1))^2 - \varepsilon^T(a,t),$$

with $|\varepsilon^T(a,t)| \leq c(a)T^{-3\alpha}$. Moreover, we may have $|Y^{T,+}(a,t) - 1| \leq \frac{1}{2}$ and $|Y^{T,-}(a,t) - 1| \leq \frac{1}{2}$ for large enough T from *Proposition B.1.* Also, we can write that,

$$\Big|\Big|i(a_T^+, a_T^-) + T\int_0^t (Y^T(a,t-s) - (1,1))\phi^T(Ts)ds\Big|\Big| \leq c(a)T^{-\alpha} \underset{T\to\infty}{\to} 0.$$

Now,

$$log\Big(exp\Big(i(a_T^+, a_T^-) + T\int_0^t (Y^T(a,t-s) - (1,1))\phi^T(Ts)ds\Big)\Big),$$

is equal to

$$i(a_T^+, a_T^-) + T\int_0^t (Y^T(a,t-s) - (1,1))\phi^T(Ts)ds,$$

since for large enough T, the imaginary part of this expression has a norm smaller than π. Hence, following from the *Equation (8)*,

$$Y^T(a,t) - (1,1) = \frac{1}{2}(Y^T(a,t) - (1,1))^2 + \varepsilon^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) \\ -ia\kappa T^{-2\alpha}(1,0) + T\int_0^t (Y^T(a,t-s) - (1,1))\phi^T(Ts)ds.$$

It is a known fact that,

$$\sum_{k\geq 1}(T\phi^T(T.))^{*k} = a_T\frac{T^\alpha}{\gamma}f^{\alpha,\gamma}\chi,$$

thus using *Lemma 4.1*, we can get,

$$Y^T(a,t) - (1,1) = \frac{1}{2}(Y^T(a,t) - (1,1))^2 + \varepsilon^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) - ia\kappa T^{-2\alpha}(1,0) \\ + \frac{a_T T^\alpha}{2\gamma}\int_0^t (Y^T(a,t-s) - (1,1))^2 f^{\alpha,\gamma}(s)\chi ds + \frac{a_T T^\alpha}{\gamma}\int_0^t \varepsilon^T(a,t-s)f^{\alpha,\gamma}(s)\chi ds \\ + ia\sqrt{\kappa}\frac{a_T}{\gamma}(1,-1)F^{\alpha,\gamma}(t)\chi - ia\kappa T^{-\alpha}\frac{a_T}{\gamma}(1,0)F^{\alpha,\gamma}(t)\chi.$$

Define,

$$\varepsilon_1^T(a,t) = \frac{1}{2}(Y^T(a,t)-(1,1))^2 + \varepsilon^T(a,t) - ia\kappa T^{-2\alpha}(1,0) + \frac{a_T T^\alpha}{\gamma}\int_0^t \varepsilon^T(a,t-s)f^{\alpha,\gamma}(s)\chi ds,$$

and

$$\varepsilon_2^T(a,t) = \frac{-1}{2}\int_0^t (Y^T(a,t-s) - (1,1))^2 f^{\alpha,\gamma}(s)\chi ds + ia\frac{\kappa}{\beta+1}T^{-2\alpha}F^{\alpha,\gamma}(t)(1,\beta).$$

Thus, we can write,

$$Y^T(a,t) - (1,1) = \varepsilon_1^T(a,t) + \varepsilon_2^T(a,t) + ia\sqrt{\kappa}T^{-\alpha}(1,-1) - ia\frac{\kappa}{(\beta+1)\gamma}T^{-\alpha}F^{\alpha,\gamma}(t)(1,\beta) + \\ \frac{T^\alpha}{2\gamma}\int_0^t (Y^T(a,t-s) - (1,1))^2 f^{\alpha,\gamma}(s)\chi ds.$$

Then, we set,

$$\theta^T(a,t) = (\theta^{T,+}(a,t), \theta^{T,-}(a,t)) = T^\alpha(Y^T(a,t) - (1,1)),$$

and,

$$r^T(a,t) = T^\alpha(\varepsilon_1^T(a,t) + \varepsilon_2^T(a,t)).$$

From *Proposition B.1*, it can be seen that $T^{2\alpha}\varepsilon_1^T(a,t)$ and $T^{2\alpha}\varepsilon_1^T(a,t)$ are uniformly bounded in t and T. Thus, $T^\alpha r^T$ is also uniformly bounded in t and T.

We have the following,

$$\theta^T(a,t) = r^T(a,t) + ia\sqrt{\kappa}(1,-1) - ia\frac{\kappa}{(\beta+1)\gamma}F^{\alpha,\gamma}(t)(1,\beta) + \frac{1}{2\gamma}\int_0^t (\theta^T(a,t-s))^2 f^{\alpha,\gamma}(s)\chi ds.$$

Using Lemma A.3 from [14], it can be shown that $\theta^T(a,t)$ is a Cauchy sequence. With this knowledge, $\theta^T(a,t)$ converges uniformly in t to (c,d) which are solutions of the following,

$$c(a,t) = ia\sqrt{\kappa} - ia\frac{\kappa}{(\beta+1)\gamma}F^{\alpha,\gamma}(t) + \frac{1}{2\gamma(\beta+1)}\int_0^t (c^2(a,t-s) + d^2(a,t-s))f^{\alpha,\gamma}(s)ds,$$

and,

$$d(a,t) = -ia\sqrt{\kappa} - ia\frac{\beta\kappa}{(\beta+1)\gamma}F^{\alpha,\gamma}(t) + \frac{\beta}{2\gamma(\beta+1)}\int_0^t (c^2(a,t-s) + d^2(a,t-s))f^{\alpha,\gamma}(s)ds.$$

B.2 Final part of the proof

In Section 4.5, we encountered the following argument in the process of deriving the characteristic function for the rough Heston model,

$$L^T(a_T^+, a_T^-, tT) = exp\left(\int_0^t (T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1))(T^{1-\alpha}\hat{\mu}(sT))ds\right).$$

From the convergence in *Appendix B.1*, we can see that,

$$T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1),$$

converges uniformly in t to $c(a,t) + d(a,t)$. Also, from *Section 4.5*, we know that $T^{1-\alpha}\hat{\mu}(tT)$ converges as $T \to \infty$ to,

$$\mu\left(1 + \frac{\xi}{\gamma\Gamma(1-\alpha)}t^{-\alpha}\right).$$

Moreover, from *Proposition B.1*, we see that,

$$(T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1))(T^{1-\alpha}\hat{\mu}(sT)),$$

is controlled for $t \in [0,1]$ with,

$$|T^\alpha(Y^{T,+}(a,t-s)-1) + T^\alpha(Y^{T,-}(a,t-s)-1)|(T^{1-\alpha}\hat{\mu}(sT)) \leq c(a)(1+s^{-\alpha}).$$

Now, applying the dominated convergence theorem, $L^T(a_T^+, a_T^-, tT)$ converges to $L_p(a,t)$,

$$L_p(a,t) = exp\Big(\int_0^t g(a,s)(1 + \frac{\xi}{\gamma\Gamma(1-\alpha)}(t-s)^{-\alpha})ds\Big),$$

where $g(a,t) = \mu(c(a,t)+d(a,t))$. By a simple manipulation, we easily see that,

$$L_p(a,t) = exp\Big(\int_0^t g(a,s)ds + \frac{v_0}{\theta\gamma} In^{1-\alpha} g(a,t) \Big).$$

Now, we need to show that g is a solution for an integral equation. From $g(a,t) = \mu(c(a,t) + d(a,t))$, it is trivial to see that,

$$g(a,t) = \mu(\beta+1)(c(a,t) - ia\sqrt{\kappa}).$$

Hence,

$$\begin{aligned} g(a,t) = -ia\frac{\mu\kappa}{\gamma} F^{\alpha,\gamma}(t) + \frac{1+\beta^2}{2\mu\gamma(1+\beta)^2} \int_0^t (g(a,s))^2 f^{\alpha,\gamma}(t-s)ds \\ -a^2\frac{\mu\kappa}{\gamma} F^{\alpha,\gamma}(t) + ia\frac{\sqrt{\kappa}(1-\beta)}{\gamma(1+\beta)} \int_0^t g(a,s) f^{\alpha,\gamma}(t-s)ds. \end{aligned}$$

Using the definition of ρ and ν from *Section 4.5* and implementing the definition of κ, $g(a,t)$ becomes,

$$\frac{\theta}{2}(-a^2 - ia)F^{\alpha,\gamma}(t) + ia\rho\nu \int_0^t g(a,s) f^{\alpha,\gamma}(t-s)ds + \frac{\nu^2}{2\theta} \int_0^t (g(a,s))^2 f^{\alpha,\gamma}(t-s)ds,$$

which can also be written as,

$$g(a,t) = \int_0^t \Big(\frac{\theta}{2}(-a^2 - ia) + ia\rho\nu g(a,s) + \frac{\nu^2}{2\theta}(g(a,s))^2 \Big) f^{\alpha,\gamma}(t-s)ds.$$

Now, if we set $h = \frac{g}{\theta\gamma}$, we have,

$$L_p(a,t) = exp\Big(\int_0^t h(a,s)(\theta\gamma + v_0 \frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)})ds \Big),$$

where,

$$h(a,t) = \int_0^t \Big(\frac{1}{2}(-a^2 - ia) + ia\gamma\rho\nu h(a,s) + \frac{(\gamma\nu)^2}{2}(h(a,s))^2 \Big) \frac{1}{\gamma} f^{\alpha,\gamma}(t-s)ds.$$

Using Lemma A.2 from [14], we can rewrite this as,

$$De^{\alpha} h(a,t) = \frac{1}{2}(-a^2 - ia) + \gamma(ia\rho\nu - 1)h(a,s) + \frac{(\gamma\nu)^2}{2} h^2(a,s), \quad In^{1-\alpha} h(a,0) = 0.$$

Thus, *Theorem 4.2* is proved.

Appendix C Proof of Theorem 3.1

For the whole proof of *Theorem 3.1*, we would need to present the whole framework of the [13] that is not in this thesis. Thus by following [13], we are going to give a more brief version of the proof. In the proof, first, P^T is rewritten in a convenient way, then some convergence results from this rewriting is shown and finally the proof is finished with some last steps.

C.1 Rewriting of P^T

First, it is trivial to see that, we can rewrite the rescaled price P^T_{tT}/T as follows,

$$\frac{P^T_{tT}}{T} = \frac{N^{T,+}_{tT} - N^{T,-}_{tT}}{T} = \int_0^{tT} \frac{dM^{T,+}_s - dM^{T,-}_s}{\sqrt{T(\lambda^{T,+}_s - \lambda^{T,-}_s)}} \sqrt{\frac{\lambda^{T,+}_s - \lambda^{T,-}_s}{T}} + \int_0^{tT} \frac{\lambda^{T,+}_s - \lambda^{T,-}_s}{T} ds.$$

We can also show that,

$$\lambda^{T,+}_t - \lambda^{T,-}_t = \int_0^t a_T(\varphi_1(t-s) - \varphi_2(t-s))(dN^{T,+}_s - dN^{T,-}_s)$$

$$= \int_0^t a_T(\lambda_2(t-s))(dM^{T,+}_s - dM^{T,-}_s) + \int_0^t a_T(\lambda_2(t-s))(\lambda^{T,+}_s - \lambda^{T,-}_s) ds,$$

where $\lambda_2 = \varphi_1 - \varphi_2$. Given that, $\psi^T_2 = \sum_{k\geq 1} a^k_T \lambda^{*k}_2$, and using *Lemma 4.1*, we get,

$$\lambda^{T,+}_t - \lambda^{T,-}_t = \int_0^t \psi^T_2(t-s)(dM^{T,+}_s - dM^{T,-}_s).$$

Then, from Fubini theorem, we have,

$$\int_0^x (\lambda^{T,+}_s - \lambda^{T,-}_s) ds = \int_0^x (\int_0^{x-s} \psi^T_2(u) du)(dM^{T,+}_s - dM^{T,-}_s).$$

Thus, the rescaled price P^T_{tT}/T can be written as,

$$\int_0^t \sqrt{C^{T,+}_s + C^{T,-}_s} dW^T_s - \int_0^t (\int_{T(t-s)}^{\infty} \psi^T_2(u) du)(d\overline{M}^{T,+}_s - d\overline{M}^{T,-}_s) + \int_0^{\infty} \psi^T_2(u) du (\overline{M}^{T,+}_t - \overline{M}^{T,-}_t),$$

where $\overline{M^T_t} = \frac{M^T_{tT}}{T}$ and,

$$W^T_t = \int_0^{tT} \frac{dM^{T,+}_s - dM^{T,-}_s}{\sqrt{T(\lambda^{T,+}_s - \lambda^{T,-}_s)}}.$$

Therefore,

$$\frac{1}{T} P^T_{tT} = \frac{1}{1 - a_T(||\varphi_1||_1 - ||\varphi_2||_1)} \int_0^t \sqrt{C^{T,+}_s + C^{T,-}_s} dW^T_s - R^T_t,$$

where,

$$R_t^T = \int_0^t \int_{T(t-s)}^{\infty} \psi_2^T(u) du (d\overline{M}_s^{T,+} - d\overline{M}_s^{T,-}).$$

One important note is that R_t^T converges to zero as T goes to infinity. This is not going to be shown in this proof, but it can be seen in [13].

C.2 Convergence of $[W^T, B^T]$

We know from [13] that using the quadratic variation $[W^T, W^T]_t$ and $[B^T, B^T]_t$ converges in probability to t,

$$[W^T, W^T]_t \underset{T\to\infty}{\to} t, \quad [B^T, B^T]_t \underset{T\to\infty}{\to} t.$$

Now, using the fact that $[M^T, M^T] = diag(N^T)$, it can be written that,

$$\begin{aligned}[W^T, B^T]_t &= \int_0^{tT} \frac{dN_s^{T,+} - \beta dN_s^{T,-}}{T\sqrt{\lambda_s^{T,+} + \lambda_s^{T,-}}\sqrt{\lambda_s^{T,+} + \beta^2 \lambda_s^{T,-}}} \\ &= \int_0^t \frac{C_s^{T,+} - \beta C_s^{T,-}}{\sqrt{C_s^{T,+} + C_s^{T,-}}\sqrt{C_s^{T,+} + \beta^2 C_s^{T,-}}} ds + \varepsilon_t^T,\end{aligned}$$

with

$$\varepsilon_t^T = \int_0^{tT} \frac{dM_s^{T,+} - \beta dM_s^{T,-}}{T\sqrt{\lambda_s^{T,+} + \lambda_s^{T,-}}\sqrt{\lambda_s^{T,+} + \beta^2 \lambda_s^{T,-}}}.$$

Then, using $\langle M^T, M^T \rangle = diag(\int_0^t \lambda^T)$ and $\lambda^T > \mu \mathbf{1}$, we can see that,

$$\mathbb{E}[(\varepsilon_t^T)^2] = \mathbb{E}[\int_0^{tT} \frac{1}{T^2(\lambda_s^{T,+} + \lambda_s^{T,-})}] \leq \frac{1}{2\mu T} \underset{T\to\infty}{\to} 0.$$

It is known from [13] that $(C^{T,+}, C^{T,-})$ converges in law for the Skorokhod topology to $(\frac{1}{\beta+1}X, \frac{1}{\beta+1}X)$ with the dynamics of X is as defined in *Theorem 3.1.* Thus, the term,

$$\frac{C_t^{T,+} - \beta C_t^{T,-}}{\sqrt{C_t^{T,+} + C_t^{T,-}}\sqrt{C_t^{T,+} + \beta^2 C_t^{T,-}}},$$

converges as $T \to \infty$ to,

$$\frac{1-\beta}{\sqrt{2(1+\beta^2)}}.$$

Then, finally,

$$[W^T, B^T]_t \underset{T\to\infty}{\to} \frac{1-\beta}{\sqrt{2(1+\beta^2)}} t.$$

C.3 End of the proof

Theorem 2.6 from [28] implies that,

$$\int_0^t \sqrt{C_s^{T,+} + C_s^{T,-}} dW_s^T,$$

converges in law for the Skorokhod topology to,

$$\int_0^t \sqrt{\frac{2X_s}{1+\beta}} dW_s.$$

Thus, this ends the proof.