

UNIVERSITY OF LAUSANNE

DEPARTMENT OF FINANCE

Rough Volatility

Author:
Elyes Mahjoubi

Supervisor:
Dimitris Karyampas

Submitted in partial fulfillment of the requirements for the MSc degree in
Computational Finance of HEC Lausanne

September 2021

Abstract

Equity Markets have long been studied and it dates since the beginning of the 20th century by Louis Bachelier. However, Theories and models about the pricing of options only began to emerge after 1970, especially with the Black Scholes model developed by Fischer Black, Myron Scholes and Robert C. Merton. Since then, the research in the pricing options' field didn't stop to evolve in order to price options accurately and efficiently.

Two of the main drawbacks of the Black Scholes model were that the implied volatility was assumed to be constant across strike prices and maturities and in addition, the model assumed that equity log returns are normally distributed. These assumptions were proven wrong as empirical studies demonstrated the existence of a skew in the implied volatility across strikes, that itself differs across maturities, and that log equity returns' distributions are generally not normal but rather leptokurtic and negatively skewed.

In order to fix these flaws, in 1993, Steven L. Heston asserted in his paper that the volatility depends on a mean reverting stochastic process and thus allowed to price options using non constant volatility.

A major drawback of this model is that it fails to correctly fit the volatility skew of very short term maturities by greatly underestimating them. Moreover, the model is complex, then not understood like the Black Scholes model by traders, and accounts for many variables, which makes its optimisation hard as the calibration of the Heston model wasn't designed to be mathematically convex and therefore needs the use of a good initial guess along of a local optimizer or the use of a global optimizer¹ to find good parameters.

In 1996, David S. Bates introduced the stochastic volatility jump model, which helped to fix the underestimation of the stochastic volatility model for the very short term maturity. Even if original stochastic volatility jump model didn't take into account the non-continuity of the volatility process, but instead only the one of the equity process, the results between the stochastic Double Jump volatility model are very similar to the one developed by David S. Bates.

Even if the jump models are backed by empirical results, a major flaw exhibited by them is that they increase the number of parameters, which makes them harder than the original stochastic models to calibrate and to understand for traders.

A new generation of stochastic volatility models appeared by the beginning of 2010, triggered and latter called by Jim Gatheral, Thibault Jaisson and Mathieu Rosenbaum as

¹It may not find a global minimum meanwhile

"rough volatility models". These new models are backed by empirical studies demonstrating that the volatility process is not a semi-Martingale mean reverting process. Moreover, the usefulness of these timely developed models reside in their tractability. Through the Thesis we will cover these new developed options' pricing models and compare them to some of the already existing ones.

Keywords - Stochastic modelling , Volatility , Fractional Brownian motion

JEL Classification - Option pricing , Mathematical Methods , Financial Forecasting and Simulation

Acknowledgments

First of all, I would like to thank my supervisor Dimitris Karyampas for providing his invaluable insight , support, as well as guidance throughout the research and writing process of this thesis. Furthermore, I would also like to offer my gratitude to Süleyman Ceran for the support and assistance in the process of obtaining the data. Finally, I would like to express appreciation to my family and close ones for their patience, understanding and encouragements.

Declaration of authorship

I, Elyes Mahjoubi , declare that this thesis titled, ‘Rough Volatility’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Contents

1	Introduction	1
2	Preliminaries	2
2.1	Keywords and definitions	2
2.2	Data introduction	3
3	The Black and Scholes Model	5
3.1	Introduction to the Black and Scholes Model	5
3.1.1	The Black and Scholes environment	5
3.1.2	Definitions	6
3.2	Derivation of the Black and Scholes Model	7
3.2.1	The Greeks	8
3.3	The Merton Jump Model	9
4	Local Volatility Model	12
4.1	The Fokker Plank Equation	12
4.2	Computation of the Local Volatility	13
5	Stochastic Volatility Models	15
5.1	The Heston Model	15
5.1.1	SDEs under the \mathbb{P} measure	15
5.1.2	SDEs under the \mathbb{Q} measure	15
5.1.3	The semi-closed form of the Heston model	16
5.2	The Heston Jump Model	20
5.3	The SVJJ Model	23
5.4	The SVI Model	25
6	Bergomi's Model	37
6.1	One factor model	37
6.2	N factors model	39
7	The Fractional Brownian Motions	41
7.1	Definitions	41
7.2	Volterra Process	44
7.3	Simulation of Fractional Brownian Motion	44
7.3.1	Fraction Brownian motion using Cholesky decomposition : . .	45

7.3.2	FBm using spectral decomposition :	45
7.3.3	FBm using efficient spectral decomposition :	46
8	The rough Bergomi model	51
8.1	The unconventional Bergomi model	51
8.2	Variance forecast	53
8.3	FSV or RFSV under the \mathbb{P} measure	55
8.4	Rough Bergomi model under the \mathbb{P} measure	55
8.5	Rough Bergomi under the \mathbb{Q} measure	56
8.6	Volatility skew	58
8.7	The smoothness of volatility log increments	59
8.8	Volatility skew of the rough Bergomi model	62
9	Calibration	64
9.1	Cumulative Distribution functions approximation for SV models : . .	64
9.2	Calibration of the SV models	65
9.3	Optimizations	68
9.4	SVI calibration	73
10	Interpolations	75
10.1	Multivariate normal distribution	75
10.2	Properties of multivariate normal distribution	76
10.3	Covariance matrix estimation	77
10.4	Gaussian Process Regression	78
11	Appendix	81
11.1	Black and Scholes	81
11.1.1	Call option price derivation	81
11.1.2	Greeks' derivation	82
11.2	Heston Model	83
11.2.1	Gil Pelaez Proof(based on the paper [1])	83

List of Figures

2.1	Empirical implied volatility	4
3.1	Simulation of the evolution of the S&P500,an ATM option and the Delta hedge expiring the 29 October 2021 with an initial date,the 21 June 2021 using $r = 3\%$	9
3.2	Monte Carlo Simulation of stock prices using a jump process	11
4.1	Local volatility surface of the S&P500 the 16th April 2021	14
5.1	Monte Carlo Simulation of stock and variance paths using a Stochastic Volatility Model	17
5.2	Visualization of the characteristic function of a stock price S_{dt} following a Log Normal Distribution with $r = 0$, $q = 0$, $V_{dt} = 0.09$, $dt = \frac{1}{252}$ and $S_0 = 4360.30$	19
5.3	Heston probability density functions for d_1 and d_2	20
5.4	Stochastic volatility surface of a simple Heston model	20
5.5	Heatmap of the AARE errors across strikes and maturities for the Heston model	20
5.6	Heatmap of the AARE errors across strikes and maturities for the Heston model with jumps	23
5.7	Evolution of the SP500 and the VIX index from 1st January 2015 to 18th June 2021	24
5.8	Stochastic volatility model skews,the one factor Heston is created using a constrained argument such $\rho = -1$	26
5.9	Stochastic volatility "inspired" of the S&P500,the 16th April 2021 and its Durrleman's condition g)	32
5.10	Comparison between SSVI of the S&P500 and the market's implied volatility surface ,the 16th April 2021.	35
5.11	Heston Parameters across maturities ,the 20th April 2021.	36
6.1	Monte Carlo simulations of 100 forward curves under Bergomi's model with parameters $\kappa = 1$, $T = 2$, $dt = \frac{1}{252}$, $\xi_0^T = 0.10$ and $w = 0.01$	39
7.1	Koch curve is a self-similar and expresses regularity.	42
7.2	Fractional Brownian motions with different Hurst parameters	46
7.3	Covariance matrix with daily lags	48
7.4	Generated fractional Gaussian noises	50

8.1	Forecast and realized volatility for $\Delta = \frac{1}{252}$ given different parameters ν	54
8.2	A log-moment comparison between RFSV and FSV	56
8.3	Monte Carlo simulation of the rough Bergomi model under the \mathbb{P} measure	57
8.4	Term structure of S&P500 at-the-money-forward volatility skew and its alternative fit	60
8.5	Log-increment volatility distribution for different lags	60
8.6	Linear fit between $\log m(q, \Delta)$ and qH	61
8.7	Log Moments of the Log volatility increments	61
8.8	Rough Bergomi's volatility skew	63
9.1	ϑ weights	67
9.2	IVMSE scatter for the calibration of a simple Heston surface	68
9.3	IVMSE scatter for the calibration of a surface SVI	68
9.4	SVJ surface generated by Gaussian process interpolation	72
9.5	IVMSE errors of the SSVI surface	73
11.1	Heaviside Function	84

List of Tables

8.1	Comparison of norm-errors between the realized and the forecasted variance for a defined parameter ν	54
8.2	Result of the rough Bergomi model calibration using a differential evolution algorithm	62
9.1	Difference in the execution time between using the Gil Peleaz integral and the fast Fourier algorithm	65
9.2	Difference in the accuracy of the Gil Peleaz integral and the fast Fourier algorithm	65
9.3	SSVI Calibration Results	68
9.4	Simple Heston Calibration Results	68
9.5	SVJ Heston calibration results	72

Chapter 1

Introduction

The stochastic volatility models enhanced in the past are showing important signs of weakness: Hard to understand, not very tractable and, in addition, containing a lot of variables to calibrate, which makes the stochastic volatility models a nightmare to grasp and use for an average trader.

The financial industry today is in total need for more simple and tractable models that allow for non-perfect but good enough what I would call "dynamic data reflectors". The rough volatility models are perfectly fitting with the role, as they usually contain few variables with a practical sense.

Through our thesis we will first remind the properties of the old developed option pricing models. In addition, for each one of them, we will show what are their main flaws and advantages to finally analyze and show why rough volatility models are one of the best actual models to use in order to, in one hand, perfectly calibrate option prices to original market data, and in other, offering an extremely realistic option dynamic insight.

So as to deliver the full insights behind the "rough volatility" option pricing topic for the thesis's reader, we will use a model called the rough Bergomi to modelize option prices and their dynamics. Finally, the main objective of this thesis is to compare the newly developed model to the past stochastic volatility and non-stochastic models so the reader could understand the whole differences between the properties of these models.

Chapter 2

Preliminaries

2.1 Keywords and definitions

During the whole thesis we used specific appellations that will sometimes appear not mentioned for brevity and simplicity reasons. The following table defines all the main reiterated terms used in the thesis :

Table: General Item Data	
S_t	Price of the underlying at time t
C	Value of a call option
$\sigma_{IV}(\tau, s, k)$	Implied volatility function
τ	Time to maturity
W_t	A Brownian motion
K	Strike price
s and k	Respectively the log price ,equivalent to $\log(S_t)$ and the log strike ,equivalent to $\log(K)$
T	Maturity date
\mathbb{P}	Physical measure
\mathbb{Q}	Risk neutral measure
Φ	Standard normal cumulative distribution function
ϕ	Characteristic function
W_t^H	Fractional Brownian motion of Hurst parameter H
\mathcal{F}_t	Filtration of the 2-dimensional Brownian Motions $W_t^{\mathbb{P},1}$ and $W_t^{\mathbb{P},2}$
$P(x_t, V_t, \tau)$	Non-actualized call price
V_t	Spot variance
\mathbb{R}	Real part
\mathbb{I}	Imaginary part
w	Total implied variance in the SSVI chapter or a weight during the Calibration chapter

Table: General Item Data	
Ω	All the possible outcomes of the 2-dimensional Brownian motions
r and q	Respectively the risk-free rate and the dividend yield
ω	Log-normal volatility of the variance, expressed as ν on the Rough Bergomi model chapter
$F_t(S_t, \tau)$	Forward price with a time to maturity τ
H_i	Log-normal jump intensity
$w^{imp}(\tau)$	Total implied variance for a time-to-maturity τ

2.2 Data introduction

The options data were taken the 16th April 2021 at the opening of the stock market. We choose to study the call options of the S&P500, as it is considered as one of the major benchmark of the largest US stock market (NYSE). We selected a wide range of strikes from deep in-the-money to deep out-of-the money and all our data were extracted from an Eikon terminal available at the CEDIF (Centre de documentation et d'initiation financière). On the second part of the thesis, clean realized volatility data of the S&P500 were extracted from the website of the Oxford-Man Institute's "realised library". The website compute the annualized realized volatility by sampling the data each 5 minutes, making the estimate less biased.

Day	Maturities	Strikes	N	S_0	r	q
April 16, 2021	15	14	210	4180.50	0.03%	1.45%

As a risk free interest rate r we use the corresponding 3 month treasury bill rate and S_0 is the current S&P500 value at the time of obtaining the data. A Gaussian interpolation of the volatility surface is giving us the following surface :

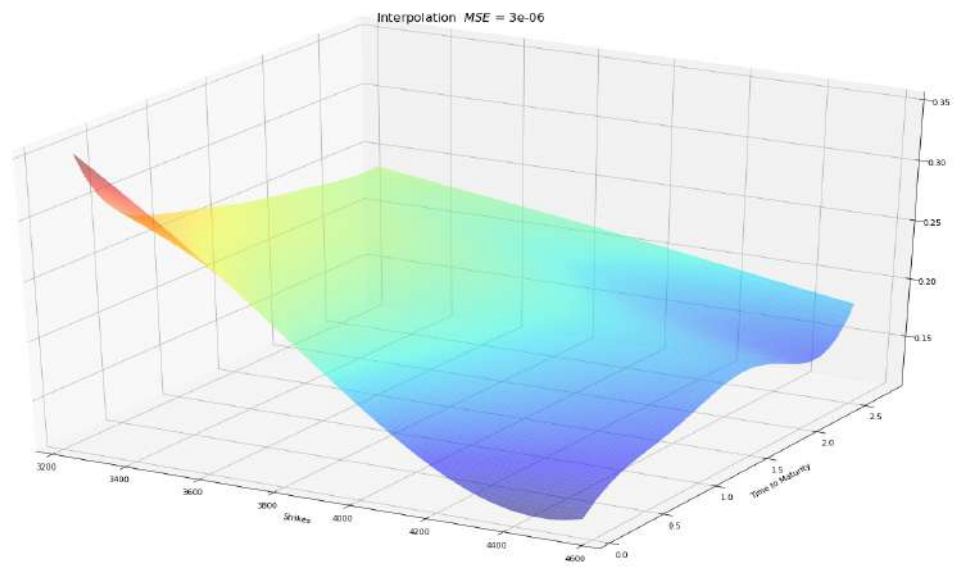


Figure 2.1: Empirical implied volatility

Chapter 3

The Black and Scholes Model

3.1 Introduction to the Black and Scholes Model

3.1.1 The Black and Scholes environment

The most basic one is considering an economy with 2 assets, which are the available underlying and a zero coupon bond. The originality of the Black and Scholes model [2] is that it allows for an payoff option to be replicated using a basket of the 2 available assets.:

Assumption 1. *Before telling more about the hedging techniques, we need to cite the most important BS assumptions, which are listed such:*

- *The option is European and can only be exercised at expiration.*
- *No dividends are paid out during the life of the option.*
- *Markets are efficient (i.e., market movements cannot be predicted and there is no possible arbitrage).*
- *There are no transaction fees in buying the option.*
- *The risk-free rate and volatility of the underlying are known and constant.*
- *The returns on the underlying asset are log-normally distributed.*

Therefore, we can write the option value of a call under the BS theory as :

$$P(S_t, t, T) = \Delta S_t + \Sigma B_t$$

With S_t the Underlying price, B the price of the zero coupon bond, Σ the shorted quantity of bonds sold and Δ the quantity of the underlying bought. We use this relation to derive the price of an option with $Ke^{-r(T-t)}$ symbolizing the price of a zero coupon bond, $\Delta = N(d_1)$ the probability for the underlying to exceed the strikes price K at maturity with a basis of e^x such $x = \log \frac{K}{S_t}$ and $\Sigma = -N(d_2)$, the same probability but with only a base of 1. Although the Black and Scholes assumptions don't seem very realistic, they allowed for far more advanced models to be developed along with hedging techniques.

3.1.2 Definitions

Definition 1 (Random Variable). Let $(\Omega, \mathcal{F}_t, \mu)$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset B of \mathbb{R} , the subset of Ω given by :

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$$

Which is in the σ -algebra \mathcal{F}_t .

Definition 2 (Stochastic Process). A stochastic process X is a collection of random variables

$$(X_t, t \geq 0) = (X_t(\omega), t \geq 0, \omega \in \Omega)$$

defined on the space Ω . By definition, a stochastic process X can be viewed as a function of two variables, the mapping ω and the time t .

Definition 3 (Standard Brownian Motion). Let $(\Omega, \mathcal{F}_t, \mu)$ be a probability space. For each $\omega \in \Omega$, suppose there is a continuous function $W(t)_{t \geq 0}$ which satisfies $W(0) = 0$. Then $W(t)_{t \geq 0}$, is a Brownian motion if for all $0 = t_0 < t_1 < \dots < t_m$, the increments :

$$W(t_1) - W(t_0), W(t_2) - W(t_1), \dots, W(t_m) - W(t_{m-1})$$

are independent and each of these increments is normally distributed with

$$\begin{aligned} E[W(t_{i+1}) - W(t_i)] &= 0 \\ \text{Var}[W(t_{i+1}) - W(t_i)] &= t_{i+1} - t_i \end{aligned}$$

Proposition 1. The path of the Brownian motion has the following properties:

- for almost every $\omega \in \Omega$, the path $W(t)(\omega)$ is continuous.
- for almost every $\omega \in \Omega$, the path $W(t)(\omega)$ is not differentiable.

Assumption 2. The assumptions on the integrand process $H(t, \omega)$ of an Ito integral $\int_0^T H(u, \omega) dW(u)$ are:

- H is adapted to Brownian motion on $[0, T]$.
- The integral $\int_0^T \mathbb{E} H^2(u, \omega) du$ is finite.
- $\mathbb{E}(H^2(t, \omega)) < \infty$ for each $0 \leq t \leq T$.
- let \mathcal{B} denote the smallest σ -field that contains all of the open subsets of $[0, T]$.

$H(t, \omega)$, as a map from $[0, T] \times \Omega$, is jointly $\mathcal{B} \times \mathcal{F}$ -measurable.

Definition 4 (Martingale). The stochastic process $Y(t, \omega)$ ($t \geq 0$ and $\omega \in \Omega$), which is adapted to the filtration \mathcal{F}_t , is called a continuous-time martingale if $\mathbb{E}|Y_t| < \infty$ for all $t > 0$ and

$$\mathbb{E}(Y_t | \mathcal{F}_s) = Y_s, \text{ for all } 0 \leq s < t$$

Remark 1. If $H(t, \omega)$ satisfies the assumption 2, $\int_0^T H(s, \omega) dW(s)$ is a martingale.

Remark 2. What a Martingale truly means is that the future change is expected to be 0, having all information at the current time.

Proposition 2 (Isometry Property of a Simple Process).

$$\mathbb{E} \left(\int_0^T H(s, \omega) dW(s) \right)^2 = \int_0^T \mathbb{E} H^2(s, \omega) ds$$

Where $H(t, \omega)$ is the simple process as defined above.

Definition 5 (Ito process). Let $W(t)$ be a Brownian motion and $\mathcal{F}(t)$ be an associated filtration, given $t \geq 0$. An Ito process is a stochastic process of the form :

$$X(t) = X(0) + \int_0^t H(u) dW(u) + \int_0^t M(u) du$$

where $X(0)$ is non-random and $H(u), M(u)$ are adapted to a stochastic process.

Proposition 3 (Adapted stochastic Process). An adapted stochastic process is a process incapable of foreseeing the future. An example can be given by trying to forecast a Martingale process (in the case $M(u) = 0$) at time 0 for time t :

$$\mathbb{E}(X_t) = X_0$$

3.2 Derivation of the Black and Scholes Model

For a probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$, for \mathcal{F} the possible σ -algebra sets available at time t $\forall t \in \mathbb{R}_+$ and \mathbb{P} a Lebesgue probability measure such as $d\mathbb{P} \exists \forall t \in \mathbb{R}_+$.

The stochastic differential equations is represented by :

$$dS_t = \mu S_t dt + \sigma S_t dW_t^{\mathbb{P}, 1}$$

With μ being the drift term, σ the standard deviation of the process and W_t a standard Brownian motion. The process is expressed such as :

$$\log\left(\frac{S_{t+dt}}{S_t}\right) \sim N\left((\mu - \frac{1}{2}\sigma^2)dt, \sigma^2 dt\right) \forall T \geq t$$

This result can be found using a Taylor Approximation along with the Ito's lemma such :

$$\begin{aligned} df(t, S_t) &= \frac{\partial f(t, S_t)}{\partial t} dt + \frac{\partial f(t, S_t)}{\partial S_t} dS_t + \frac{1}{2} \frac{\partial^2 f(t, S_t)}{\partial S_t^2} d\langle S_t, S_t \rangle \\ &\iff df(t, S_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t^{\mathbb{P}, 1} \end{aligned}$$

With $f(t, S_t) = \log(S_t)$

Using the Randon-Nikodym and the principle of Risk Neutrality under a \mathbb{Q} measure, we obtain :

$$\text{Log}\left(\frac{S_{t+dt}}{S_t}\right) \sim N\left(\left(r - \frac{1}{2}\sigma^2\right)dt, \sigma^2 dt\right)$$

With r being the risk free rate.

From this we can easily compute the Option Price by taking the expectation of the difference between the Spot price S_T and the strike K at a specific time. Ending Up with the an option price $C(S_t, K, t, T)^1 \forall T \geq t$:

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}(e^{-r(T-t)}(S_T - K)^+) &= S_t \Phi\left(\frac{\ln \frac{S_t}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad - K e^{-r(T-t)} \Phi\left(\frac{\ln \frac{S_t}{K} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}\right) \\ &= S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2) \\ &= C(S_t, K, t, T) \end{aligned}$$

With Φ being the CDF of a standard normal distribution. The main drawback of the Black Sholes is that the volatility is taken as constant while it is empirically definitely not the case as shown in figure 5.7.

3.2.1 The Greeks

The Greeks are very used tools in finance. Generally traders are using them so as to determine how to hedge and adjust their positions in risky assets to limit losses. The most used Greeks considered are listed below :

Name	Symbol	Derivative
Delta	Δ	$\frac{\partial V}{\partial S_t}$
Gamma	Γ	$\frac{\partial^2 V}{\partial S_t^2}$
Rho	ρ	$\frac{\partial V}{\partial r}$
Theta	Θ	$\frac{\partial V}{\partial \tau}$
Vega	ϑ	$\frac{\partial V}{\partial \sigma}$

With respectively $\Delta, \Gamma, \rho, \Theta, \vartheta$ do represent the sensibility of the option's value to the change of : the price of the underlying, the change in the price of the underlying, the interest rate, the volatility of the underlying and the time to maturity. For a call option, These coefficients are found on the appendix. An example of the evolution of the Δ in function of the asset price movement S_t is shown on the figure 9.1.

¹Derivations are available on the appendix

3.3 The Merton Jump Model

Jump stochastic models were initially created by Robert C. Merton [3]. One of the main goals of adding jumps to the log-normal process defined in the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ is made in order to reproduce the discontinuous reality of stock movements by therefore adding discrete brutal jumps to the asset price. It is empirically shown that the asset prices are non-continuous and thus, we can justify the use of a Poisson process in the original SDEs.

The modified SDE takes then the form of :

$$dS_t = \mu S_{t-} dt + \sigma S_{t-} dW_t^{\mathbb{P},1} + S_{t-} dJ_t$$

$$dJ_t = \prod_{i=1}^{dN_t} H_i - 1$$

Where $dN(t)$ is a Poisson process distributed with a probability of λdt , S_{t-} is the price before the jump, μ and σ are respectively the drift and the standard deviation of the process and V_i is the log-normal intensity of the jump. After an ap-

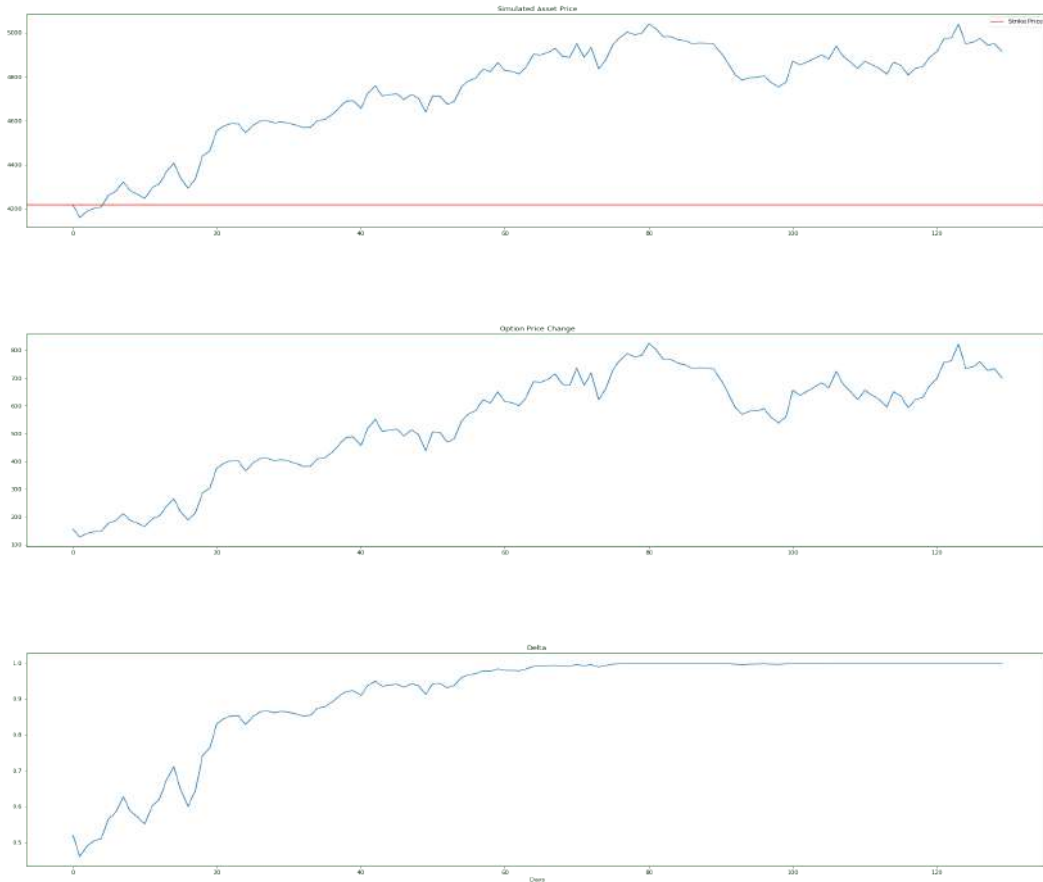


Figure 3.1: Simulation of the evolution of the S&P500, an ATM option and the Delta hedge expiring the 29 October 2021 with an initial date, the 21 June 2021 using $r = 3\%$

propriate application of a Taylor serie using f , a smooth and a twice differentiable function, $\log(S_t)$, in our case. We obtain :

$$\begin{aligned} f(T, S_T) - f(0, S_0) &= \int_0^T \left(\frac{\partial f}{\partial u} + \mu S_{u-} \frac{\partial f}{\partial S_{u-}} - \frac{1}{2} \sigma^2 S_{u-}^2 \frac{\partial^2 f}{\partial S_{u-}^2} \right) du \\ &\quad + \sigma \int_0^T S_{u-} \frac{\partial f}{\partial u} dW_u^{\mathbb{P},1} + \sum_{i=1}^T (f(T, S_i) - f(0, S_{i-})) \mathbf{1}_{\{N_i - N_{i-1} = 1\}} \end{aligned}$$

Finally we find the log price of the asset is obtained by integrating the continuous part and adding the log normally distributed jump's intensity such $\log(H_i) \sim N(m, \nu)$:

$$\begin{aligned} \log \left(\frac{S_T}{S_0} \right) &= \left(\mu - \frac{1}{2} \sigma^2 \right) T + \sigma W_T^{\mathbb{P},1} + \sum_{i=0}^T \mathbf{1}_{\{N_i - N_{i-1} = 1\}} \log \left(\frac{S_{i-}(H_i)}{S_{i-}} \right) \\ &= \left(\mu - \frac{1}{2} \sigma^2 \right) T + \sigma W_T^{\mathbb{P},1} + \sum_{i=0}^T \mathbf{1}_{\{N_i - N_{i-1} = 1\}} \log(H_i) \end{aligned}$$

After an appropriate application of risk neutrality to the log process we finally obtain:

$$S_T = S_0 e^{\left(\left(r - \frac{\sigma^2}{2} \right) T + \sigma W_T^{\mathbb{P},1} \right)} \prod_{i=1}^{N_T} H_i$$

Yielding to a deformation of the Black and Scholes Formula such :

$$C = \sum_{n=0}^{\infty} \frac{e^{-\lambda T} (\lambda T)^n}{n!} BS(S_0, \sigma_n, r_n, T, K) \quad (3.1)$$

With :

$$\sigma_n = \sqrt{\sigma^2 + n\nu/T} \quad (3.2)$$

$$r_n = r - \lambda(e^{m+\frac{1}{2}\nu} - 1) + \frac{1}{T} \left(\frac{n\nu}{2} + nm \right) \quad (3.3)$$

Knowing that (3.2) and (3.3) are the transformed drift and standard deviation of the process, we can easily derive the non-actualized option price C of an option using (3.1). The Merton-jump model is a good try to fix the flaw of the constant volatility and non-leptokurtic distribution present in the Black and Scholes model while it maintains its analytical tractability.

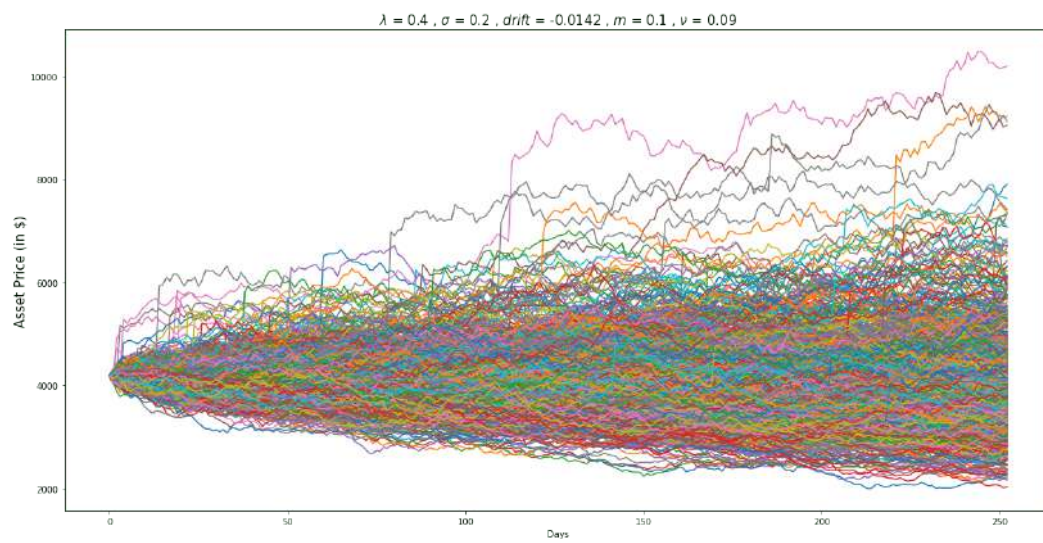


Figure 3.2: Monte Carlo Simulation of stock prices using a jump process

Chapter 4

Local Volatility Model

Bruno Dupire in his research paper, suggested that volatility could be computed directly from given strikes and maturities. Dupire ingeniously used the Fokker-Planck equation, which describes the dynamic of a known distribution across time.

4.1 The Fokker Plank Equation

Defining a probability space such as $(\Omega, \mathcal{F}_t, \mathbb{P})$, $\forall t \in \mathbb{R}_+$ and as $p \ni \forall t \in \mathbb{R}$. In addition, The dynamic of a variable S_t such as $S_t \in \mathbb{R}$ in the probability space under the physical measure is written as:

$$dS_t = (r - q)S_t dt + \sigma S_t dW^{\mathbb{P}}$$

However, as the dividend yield rate q is introduced. In order for us to have a risk neutral environment would require the use of the forward F_t such we have :

$$dF_t(S_t, \tau) = \sigma F_t(S_t, \tau) dW^{\mathbb{Q}}$$

with $F_t(S_t, \tau) = S_t e^{(r-q)\tau}$, for τ the time to maturity. Because we want to work in a risk-neutral \mathbb{Q} space we will use the forward rate and we define the dynamic of our probability density function such $p : \mathbb{R} \rightarrow \mathbb{R}_+$:

$$\frac{\partial p(S_t, t)}{\partial t} = \frac{1}{2} \frac{\partial D(S_t, t) p(S_t, t)}{\partial S_t^2} \quad (4.1)$$

A proof of 4.1 is given on the article [4]. Here we define $D(S_t, t) = \sigma^2 S_t^2$, respectively the drift term and the standard deviation of the process. Here the first term comes from a 2 step partial integration differentiation. Plugging this PDE into the expectation of an option price. A call in our case can be expressed such :

$$\begin{aligned} C(S_t, K, T, t) &= e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+] \\ &= e^{-r(T-t)} \int_K^\infty (S - K) p dS \end{aligned}$$

$$= e^{-r(T-t)} \int_K^\infty S p dS - K e^{-r(T-t)} \int_K^\infty p dS$$

Setting $\tau = T - t$, the precedent formula devolves directly to the equation :

$$\frac{\partial C}{\partial \tau} = \frac{1}{2} \sigma^2 K^2 \frac{\partial^2 C}{\partial K^2} \quad (4.2)$$

A proof of 4.2 is given on the article [5].

4.2 Computation of the Local Volatility

We have now a direct relation between the volatility and the strikes along with the maturities. By arranging the last equation we find :

$$\sigma^2 = \sigma(K, \tau)^2 = \frac{\frac{\partial C}{\partial \tau}}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}} \quad (4.3)$$

In addition, for implementation purpose we can derive the following variance using the following formulas as we know that C depends on S_t, K, t and T :

$$\begin{aligned} \frac{\partial C}{\partial \tau} &= \frac{C(S_t, K, \tau + \Delta\tau) - C(S_t, K, \tau - \Delta\tau)}{2\Delta\tau} \\ \frac{\partial^2 C}{\partial K^2} &= \frac{C(S_t, K + 2\Delta K, \tau) - 2C(S_t, K, \tau) + C(S_t, K - 2\Delta K, \tau)}{(\Delta K)^2} \end{aligned}$$

An exemplified interpolated surface coming from these proceed is showed in the figure 4.1. As we can compare with 2.1, the main advantage of the local volatility models are that we can fit the data extremely well without pain. However, its main drawback locates in the fact that its dynamic is wrong and that of the deterministic nature of the local volatility doesn't allow us to assess the possible future evolution of the volatility skew given some market anticipations.

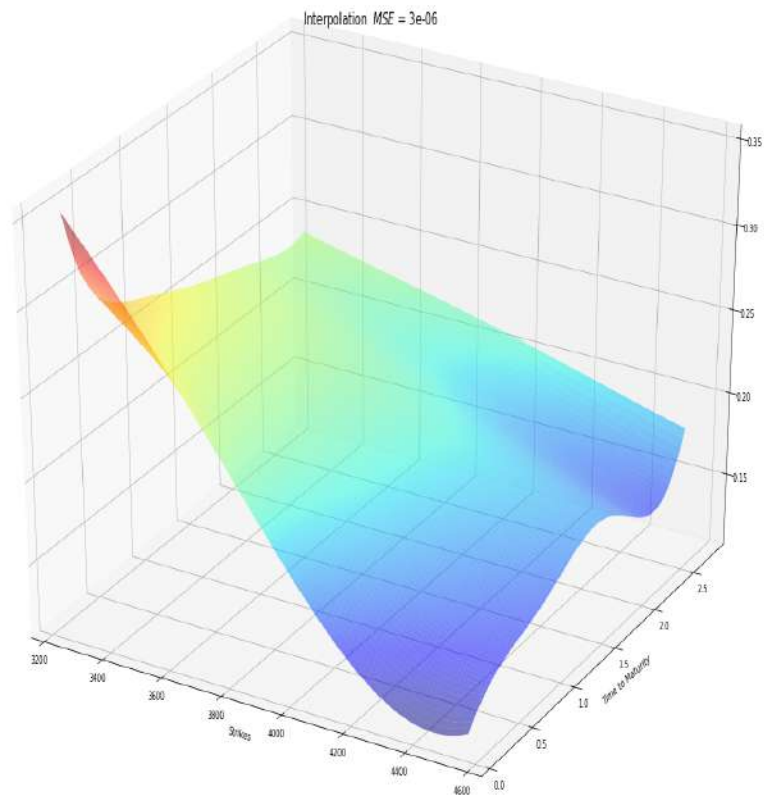


Figure 4.1: Local volatility surface of the S&P500 the 16th April 2021

Chapter 5

Stochastic Volatility Models

5.1 The Heston Model

5.1.1 SDEs under the \mathbb{P} measure

One of the main drawbacks of basic options' pricing models lately reviewed is that volatility is granted to be constant across time. Empirically it has been shown that the volatility tends to be mean revertible. A better approach to price options has its foundations on a Black and Scholes standard SDE along with an Ornstein-Uhlenbeck process describing the dynamic of the variance in a probabilistic space $(\Omega, \mathcal{F}_t, \mathbb{P})$ represented by :

$$\begin{aligned} dS_t &= (r - q)S_t dt + \sqrt{V_t}S_t dW_t^{\mathbb{P},1} \\ dV_t &= \kappa(\theta - V_t) dt + \xi\sqrt{V_t}dW_t^{\mathbb{P},2} \end{aligned}$$

With ξ being the volatility of the variance, κ being the mean reversion parameter and finally θ being the long term variance. The other particularity of this models is that it can be plurifactorial. Indeed, the volatility process and the stock prices tend to be correlated but not mono-factorial, meaning that $dW_t^{\mathbb{P},2} = \pm dW_t^{\mathbb{P},1}$, but instead :

$$\mathbb{E}(dW_t^{\mathbb{P},1}dW_t^{\mathbb{P},2}) = \rho dt$$

5.1.2 SDEs under the \mathbb{Q} measure

In order to keep our κ or θ fixed. We set the market premium of the volatility to 0. Our risk netral SDEs become :

$$\begin{aligned} dF_t &= \sqrt{V_t}S_t dW_t^{\mathbb{Q},1} \\ dV_t &= \kappa(\theta - V_t) dt + \xi\sqrt{V_t}dW_t^{\mathbb{Q},2} \\ \mathbb{E}(dW_t^{\mathbb{Q},1}dW_t^{\mathbb{Q},2}) &= \rho dt \end{aligned}$$

5.1.3 The semi-closed form of the Heston model

In 1993, Steve L. Heston [6] developed a semi closed form formula to price European options based on a stochastic volatility model for $[S_t, V_t] \in [\mathbb{R}_+, \mathbb{R}_+]$. This model allows to treat the volatility path as a diffusion. In order to construct the semi-closed formula we would need to use the Taylor serie and Ito's lemma as usual, such we obtain the PDE :

$$0 = \frac{\partial C}{\partial t} + \frac{1}{2} V_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} + \rho \xi V_t S_t \frac{\partial^2 C}{\partial S_t \partial V_t} + \frac{1}{2} \xi^2 V_t \frac{\partial^2 C}{\partial V_t^2} + (r - q) S_t \frac{\partial C}{\partial S_t} - rC - \kappa(V_t - \theta) \frac{\partial C}{\partial V_t}$$

Using a change of variable such as :

$$x_t = \ln \frac{F_t(S_t, \tau)}{K}$$

In the same time, we define the present value of the option price as $e^{-r\tau} P(x_t, V_t, \tau) = C(S_t, K, V_t, T, t)$. Thereafter, by applying the chain rule we obtain the following equation :

$$0 = -\frac{\partial P}{\partial \tau} + \frac{1}{2} V_t \frac{\partial^2 P}{\partial x_t^2} - \frac{1}{2} V_t \frac{\partial P}{\partial x_t} + \frac{1}{2} \rho^2 V_t \frac{\partial^2 P}{\partial V_t^2} + \rho \xi V_t \frac{\partial^2 P}{\partial x_t \partial V_t} - \kappa(V_t - \theta) \frac{\partial P}{\partial V_t} \quad (5.1)$$

Before that, we need to rewrite the formula of an option price as :

$$P(x_t, V_t, \tau) = K \{ e^{x_t} D_1(x_t, V_t, \tau) - D_0(x_t, V_t, \tau) \} \quad (5.2)$$

Substituting of $P(x_t, V_t, \tau)$ in the found PDE for $j \in [0, 1]$, we find out :

$$0 = -\frac{\partial D_j(x_t, V_t, \tau)}{\partial \tau} + \frac{1}{2} V_t \frac{\partial^2 D_j(x_t, V_t, \tau)}{\partial x_t^2} - \left(\frac{1}{2} - j\right) V_t \frac{\partial D_j(x_t, V_t, \tau)}{\partial x_t} + \frac{1}{2} V_t \xi^2 \frac{\partial^2 D_j(x_t, V_t, \tau)}{\partial V_t^2} + \rho \xi V_t \frac{\partial^2 D_j(x_t, V_t, \tau)}{\partial V_t \partial x_t} + (a - b_j V_t) \frac{\partial D_j(x_t, V_t, \tau)}{\partial V_t} \quad (5.3)$$

Where :

$$a = \kappa\theta, b_j = \kappa - j\rho\xi \quad (5.4)$$

Remark 3. In order to pass from 5.1 to 5.3, we set $K = 0, S = 1$ to express the PDE in terms of D_1 . To express it in terms of D_0 , we first set $K = -1, S = 0$.

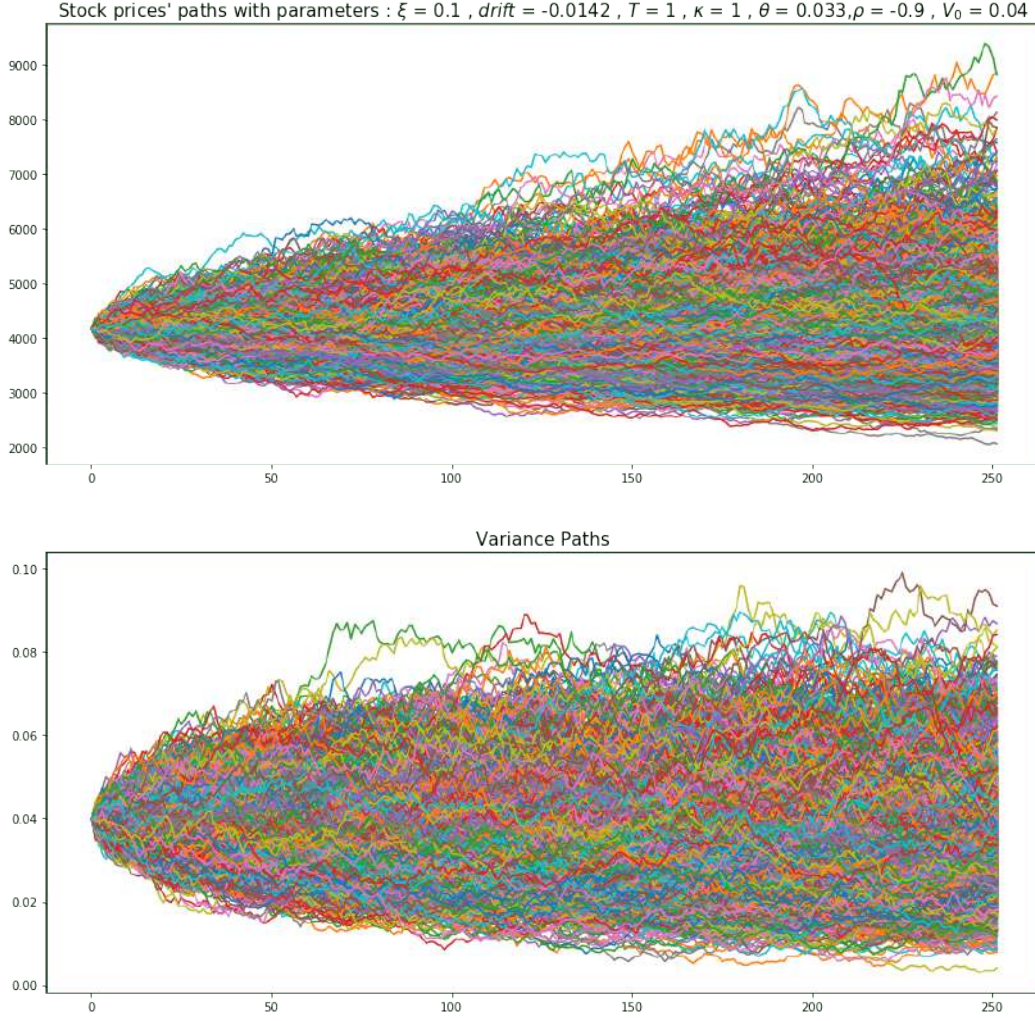


Figure 5.1: Monte Carlo Simulation of stock and variance paths using a Stochastic Volatility Model

Convolution We define $D_1(x_t, V_t, \tau)$ as the probability of exercise of the option under the base e^{x_t} and $D_0(x_t, V_t, \tau)$ the natural probability of exercise with $D_j \in \mathbb{R}_+ \forall t \in \mathbb{R}_+$. In addition, many distributions don't have a defined probability distributions function inversely to their characteristic functions. Moreover, The convolution property of characteristic functions allows us to write a characteristic function of the sum of two continuous random i.i.d variables $A + B$ such :

$$(\phi_A * \phi_B)(x) = \int_{-\infty}^{\infty} \phi_A(x - y) \phi_B(y) dy$$

With ϕ_A and ϕ_B being two characteristic functions for $\phi_i: \mathbb{R} \rightarrow \mathbb{C}$. For A and B having two Lebesgue measures in \mathbb{R} . The important properties of the characteristic functions are listed as :

- $\phi(0) = 1$
- ϕ is uniformly continuous

- $\phi(t)$ is Hermitarian, meaning $\phi(t) = \overline{\phi(-t)}$
- $|\phi(t)|$ is bounded such as $|\phi(t)| < 1$
- The real part of ϕ is even and the imaginary part is odd

Therefore, it is natural for us to use the inverse Fourier transform as we working with characteristic so as to retrieve the original probability distribution function and the wanted cumulative distribution. We can transform our density function written $d_j(x_t, V_t, \tau) = \partial_{x_t} D_j(x_t, V_t, \tau)$ into signal through the Fourier transform :

$$\begin{aligned}\tilde{d}_j(u, V_t, \tau) &= \mathbb{E}^Q(e^{iux_t}) \\ \tilde{d}_j(u, V_t, \tau) &= \int_{-\infty}^{\infty} dx_t e^{iux_t} d_j(x_t, V_t, \tau)\end{aligned}$$

Applying the inverse Fourier transform gives us the original Probability density function :

$$d_j(x_t, V_t, \tau) = \int_{-\infty}^{\infty} \frac{du}{2\pi} e^{iux_t} \tilde{d}_j(u, V_t, \tau)$$

Integrating both side over x_t gives use the Cumulative Probability function :

$$\int_{x_t}^{\infty} d_j(x_t, V_t, \tau) dx_t = \int_{x_t}^{\infty} \int_{-\infty}^{\infty} \frac{du}{2\pi} e^{-iux_t} \tilde{d}_j(u, V_t, \tau) dx_t$$

Our result is Lebesgue integrable and then satisfy the condition for which $g = \frac{1}{2\pi} e^{-iux_t} \tilde{d}_j(u, V_t, \tau)$ such that $g: \mathbb{R}_1 \times \mathbb{R}_2 \rightarrow \mathbb{R}_+$. The condition could be summarized by the inequation $\int_{\mathbb{R}_1 \times \mathbb{R}_2} \left| e^{-iux_t} \frac{\tilde{d}_j(u, V_t, \tau)}{2\pi} \right| d(x_t, u) < \infty$. The presented figure 5.2 gives the intuition why the measure is Lebesgue integrable, which makes the application of the Fubini's theorem possible. Therefore, by applying Fubini's theorem, the calculation is straightforward and we obtain :

$$D_j(x_t, V_t, \tau) = \int_{-\infty}^{\infty} \frac{du}{2\pi iu} e^{-iux_t} \tilde{d}_j(u, V_t, \tau)$$

In order to continue the derivation of the semi-closed formula of European options, we need for $D_j(x_t, V_t, \tau)$ to be tractable in order to compute its derivatives. For that reason we assume that $\tilde{D}_j(x_t, V_t, \tau)$ has affine dynamic structure defined in \mathbb{R}_+ under the form :

$$\tilde{D}_j(x_t, V_t, \tau, u) = \frac{e^{C_{SV}(u, \tau)\theta + L(u, \tau)V_t}}{iu} \quad (5.5)$$

For brevity we rewrite the equation 5.3 such :

$$V_t \left\{ \alpha \tilde{D}_j - \beta \frac{\partial \tilde{D}_j}{\partial V_t} + \gamma \frac{\partial^2 \tilde{D}_j}{\partial V_t^2} \right\} + a \frac{\partial \tilde{D}_j}{\partial V_t} - \frac{\partial \tilde{D}_j}{\partial \tau} = 0 \quad (5.6)$$

By defining :

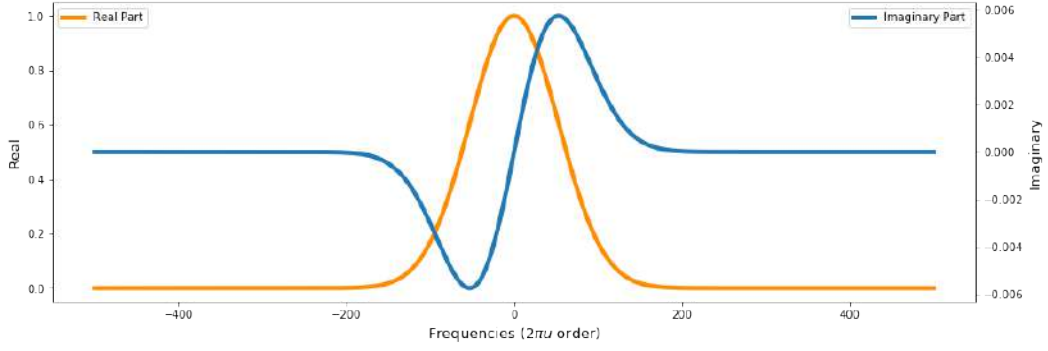


Figure 5.2: Visualization of the characteristic function of a stock price S_{dt} following a Log Normal Distribution with $r = 0$, $q = 0$, $V_{dt} = 0.09$, $dt = \frac{1}{252}$ and $S_0 = 4360.30$.

$$\begin{aligned}\alpha &= -\frac{u^2}{2} - \frac{iu}{2} + iju \\ \beta &= \kappa - \rho\xi j - \rho\xi iu \\ \gamma &= \frac{\xi^2}{2}\end{aligned}$$

And by deriving 5.6:

$$\begin{aligned}\frac{\partial \tilde{D}_j}{\partial \tau} &= \left\{ \theta \frac{\partial C_{SV}}{\partial \tau} + V_t \frac{\partial L}{\partial \tau} \right\} \tilde{D}_j \\ \frac{\partial \tilde{D}_j}{\partial V_t} &= L \tilde{D}_j \\ \frac{\partial^2 \tilde{D}_j}{\partial V_t^2} &= L^2 \tilde{D}_j\end{aligned}\tag{5.7}$$

and finally plugging what we obtain from 5.7 in 5.6 leads to :

$$\frac{\partial L}{\partial \tau} = \alpha - \beta L + \gamma L^2\tag{5.8}$$

$$\frac{\partial C_{SV}}{\partial \tau} = \kappa L\tag{5.9}$$

We then obtain an ODE in 5.8, which we can solve in order to find $L(u, \tau)$ in a first time thanks to the Riccati Differential Equation's solution. Nevertheless, for the equation 5.9 we only need to use simply integrate and thereafter. The two variables are then equal to :

$$\begin{aligned}L(u, \tau) &= r_- \frac{1 - e^{-d\tau}}{1 - g e^{-d\tau}} \\ C_{SV}(u, \tau) &= \kappa \left\{ r_- \tau - \frac{2}{\xi^2} \log \left(\frac{1 - g e^{-d\tau}}{1 - g} \right) \right\}\end{aligned}$$

with :

$$g := \frac{r_-}{r_+}$$

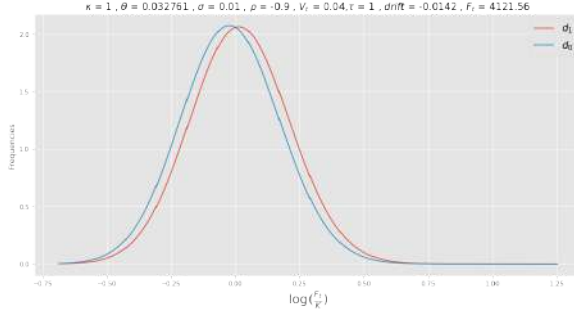


Figure 5.3: Heston probability density functions for d_1 and d_2

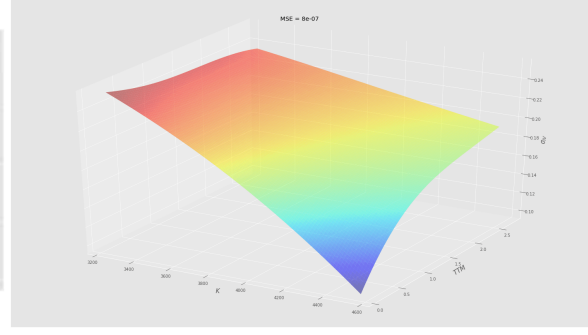


Figure 5.4: Stochastic volatility surface of a simple Heston model

and

$$r_{\pm} = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\gamma} =: \frac{\beta \pm d}{\xi^2}$$

After that, we finally end up reversing back the characteristic function 5.9 defined in \mathbb{C} to the original PDE for $D_j \in \mathbb{R}_+$ using the 5.10 Gil-Pelaez integral, which is not a proper Lebesgue integral but rather directly coming from Dirichlet's one [1].

$$D_j(x, V_t, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} du \operatorname{Re} \left\{ \frac{\exp \{ C_{SV}(u, \tau) \theta + L(u, \tau) V_t + i u x \}}{i u} \right\} \quad (5.10)$$

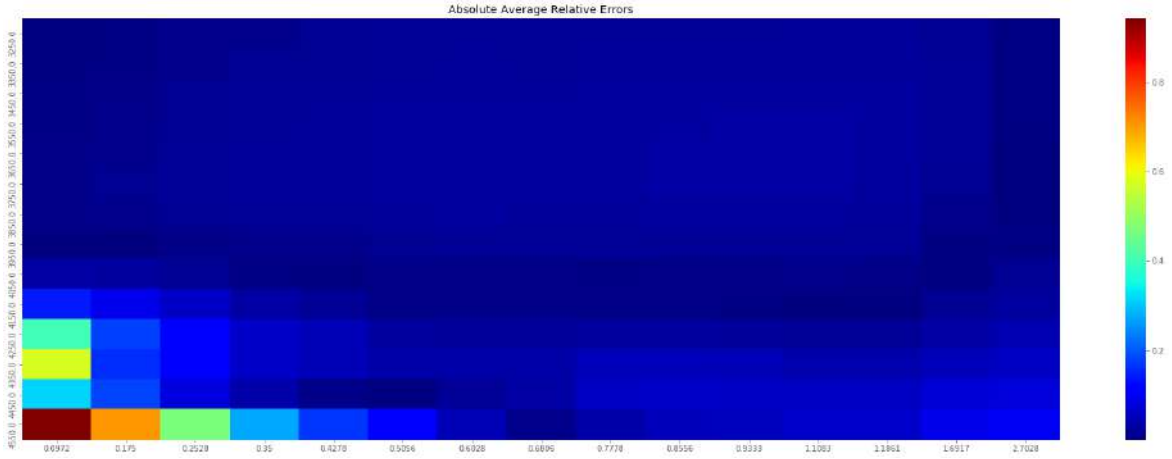


Figure 5.5: Heatmap of the AARE errors across strikes and maturities for the Heston model

5.2 The Heston Jump Model

One of the main flaws of the Heston Stochastic models is that it dramatically underweights the volatility surface for short maturities. David S. Bates on his paper [7]

suggested to add jumps to the natural Standard differential equation of the price. One explanation for that can be explained by the happening of crisis or major events (Black Monday in 1987 ,Lehman Brothers collapse in 2008 , Coronavirus pandemic in 2020).In order to do so ,we set a new standard differential equation in the same probability space as defined before under \mathbb{P} physical measure such :

$$\begin{aligned} dS_t &= (r - q)S_t dt + \sqrt{V_t}S_t dW_t^{\mathbb{P},1} + S_t dJ_t \\ dV_t &= \kappa(\theta - V_t) dt + \xi\sqrt{V_t}dW_t^{\mathbb{P},2} \\ dJ_t &= \prod_{i=1}^{dN_t} H_i - 1, \mathbb{E}(dW_t^{\mathbb{P},1}dW_t^{\mathbb{P},2}) = \rho dt \\ \log(H_i) &\sim N(m, \nu), N_t \sim \text{Poisson}(\lambda t) \end{aligned}$$

By rewriting our Taylor serie in order to include the discrete jumps we can observe a PDE written as :

$$\begin{aligned} dC(t, S_t, V_t) &= \frac{\partial C}{\partial t}dt + \frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}d\langle S_{-t}, S_{-t} \rangle + \frac{\partial^2 C}{\partial S_{-t}\partial V_t}d\langle S_{-t}, V_t \rangle + \frac{1}{2}\frac{\partial^2 C}{\partial V_t^2}\langle V_t, V_t \rangle \\ &+ \frac{\partial C}{\partial S_{-t}}dS_{-t} + \frac{\partial C}{\partial V_t}dV_t + \sum_{i=0}^{dN_t} [C(t, S_t, V_t) - C(t, S_{-t}, V_t)] \end{aligned}$$

However we are still in the physical risk world. Therefore we apply the Girsanov theorem, A proof and explanations of the theorem are available on the paper [8], by using our usual forward rate $F(S_t, \tau)$ along with the fact that we fix the risk premium of the volatility to 0. After applying the Girsanov theorem and changed the risk measure to \mathbb{Q} , our risk neutral SDE becomes :

$$\begin{aligned} dF_t &= -\lambda t(e^{m+\frac{\nu}{2}} - 1)\sqrt{V_t}F_t dW_t^{\mathbb{Q},1} + F_t dJ_t \\ dV_t &= \kappa(\theta - V_t) dt + \xi\sqrt{V_t}dW_t^{\mathbb{Q},2} \\ dJ_t &= \prod_{i=1}^{dN_t} H_i - 1, E(dW_t^{\mathbb{Q},1}dW_t^{\mathbb{Q},2}) = \rho dt \\ \log(H_i) &\sim N(m, \nu), N_t \sim \text{Poisson}(\lambda t) \end{aligned}$$

We then transform our old PDE to a PIDE ,which refers to the mixture of a partial differential equation with integrals. We end up with this PIDE as as we are using the Black and Scholes assumption that $\mathbb{E}^{\mathbb{Q}}(dC(t, S_t, V_t)) = rCdt$. The integral appears because the log-jump $\log(H_i)$ is normally distributed. Therefore, after having expressed the call price in term of 5.3, we suddenly obtain :

$$\begin{aligned} 0 &= -\frac{\partial P}{\partial \tau} + \frac{1}{2}V_t\frac{\partial^2 P}{\partial x_t^2} - \frac{1}{2}V_t\frac{\partial P}{\partial x_t} + \frac{1}{2}\rho^2V_t\frac{\partial^2 P}{\partial V_t^2} + \rho\xi V_t\frac{\partial^2 P}{\partial x_t\partial V_t} - \kappa(V_t - \theta)\frac{\partial P}{\partial V_t} \quad (5.11) \\ &+ e^{-r\tau}\lambda \int_0^\infty (P(x_{t-} + \log(H_i), V_t, \tau) - P(x_{t-}, V_t, \tau))p(\log(H_i))d\log(H_i) \end{aligned}$$

Thereafter, we define $p(x)$, the probability distribution of the log jumps as we evaluate options using log-prices such as shown in 5.1:

$$p(x) = \frac{e^{-\frac{(x-m)^2}{2\nu}}}{2\sqrt{\pi\nu}}$$

A major problem here is that we don't know how to analytically compute the equation directly not because of the jump but rather of the involved function that is controlling the movements of $C(t, S_t, V_t)$. An extremely elegant way of doing so would be to pass through the characteristic function as we know by convolution that the sum of k i.i.d random variables $X_1 + X_1 + \dots + X_{k+1}$ is distributed such $(f_1 * f_2 * \dots * f_{k+1})(x)$ with f_i the respective i density function. As a result, we derive straightforwardly the density function of our process :

$$(f_1 * f_2 * \dots * f_{k+1})(x) = \int_{-\infty}^{\infty} f_1(x-t)f_2(t)dt \int_{-\infty}^{\infty} f_1(x-z)f_3(z)dz \dots \int_{-\infty}^{\infty} f_k(x-g)f_{k+1}(g)dg$$

In our case we already know what is the distribution of our increments. Hence, a direct intuition that comes to our mind is to derive the convoluted distribution function such as:

$$(f_1 * f_2 * \dots * f_{k+1})(x) = \left(\int_{-\infty}^{\infty} p(x)dx \right)^k$$

Moreover, a simple derivation of a jump Poisson process is exemplified in 5.12. However, In our case, we are dealing with a compounded Poisson Process as we are mixing simple Poisson process with another log-normal distribution altering the intensity of the jumps. A simple Poisson process has a characteristic function given by :

$$\begin{aligned} \tilde{p}_\lambda(u) &= \sum_{k=0}^{\infty} e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!} e^{iku} \\ &= e^{-\lambda\tau} e^{\lambda\tau e^{iu}} \\ &= e^{\lambda\tau(e^{iu}-1)} \end{aligned} \tag{5.12}$$

We can clearly notice that the major change to evaluate the price of an option under the SVJ Heston is only altered by As a result by reproducing the steps 5.1 in addition with the discrete term expressed in 5.11. We find out that the characteristic function of the Heston Model becomes 5.13:

$$\phi_T^{SVJ}(u) = e^{C_J(u,\tau)\theta + L(u,\tau)V_t} \hat{\Xi}_\lambda(u, \tau) \tag{5.13}$$

We could derive 5.13 because the characteristic function of the Poisson process derived as written in 5.12. However, as our equation includes a log-normally distributed term. We define $\hat{\Xi}_\lambda(u, \tau)$, with p the probability density of the log-jump process for $t \in \mathbb{R}_+$ and $p: \mathbb{R}_+ \rightarrow \mathbb{R}_+$:

$$\hat{\Xi}_\lambda(u, \tau) = \mathbb{E}^{\mathbb{Q}}(E(e^{iu \sum_{i=1}^{N_t} \log(H_i)} | N_t = k)) \iff \sum_{k=1}^{\infty} E(e^{iu \log(H_i)} | N_t = k)^k P(N_t = k)$$

$$= \sum_{k=1}^{\infty} \frac{e^{-\lambda\tau}}{k!} \left(\lambda\tau \int_{-\infty}^{\infty} e^{iu \log(H_i)} p(\log(H_i)) d\log(H_i) \right)^k$$

Set $j = \log(H_i)$

$$\begin{aligned} &= e^{-\lambda\tau} e^{\lambda\tau \int_{-\infty}^{\infty} e^{iuj} p(j) dj} \\ &= e^{-\lambda\tau} e^{\lambda\tau \int_{-\infty}^{\infty} e^{iuj} \frac{e^{-\frac{(j-m)^2}{2\nu}}}{2\sqrt{\pi\nu}} dj} \\ &= e^{\lambda\tau(e^{i um - \frac{v u^2}{2}} - 1)} \end{aligned}$$

In addition, as we are in a risk neutral world determined by the Lebesgue dynamic measure \mathbb{Q} , the term $C(u, \tau)$ transforms into :

$$C_{SVJ}(u, \tau) = -iu\lambda\tau \left(\frac{e^{m - \frac{v}{2}} - 1}{\theta} \right) + \lambda\tau(e^{i um - \frac{v u^2}{2}} - 1) + \kappa \left\{ r_- \tau - \frac{2}{\xi^2} \log \left(\frac{1 - g e^{-d\tau}}{1 - g} \right) \right\}$$

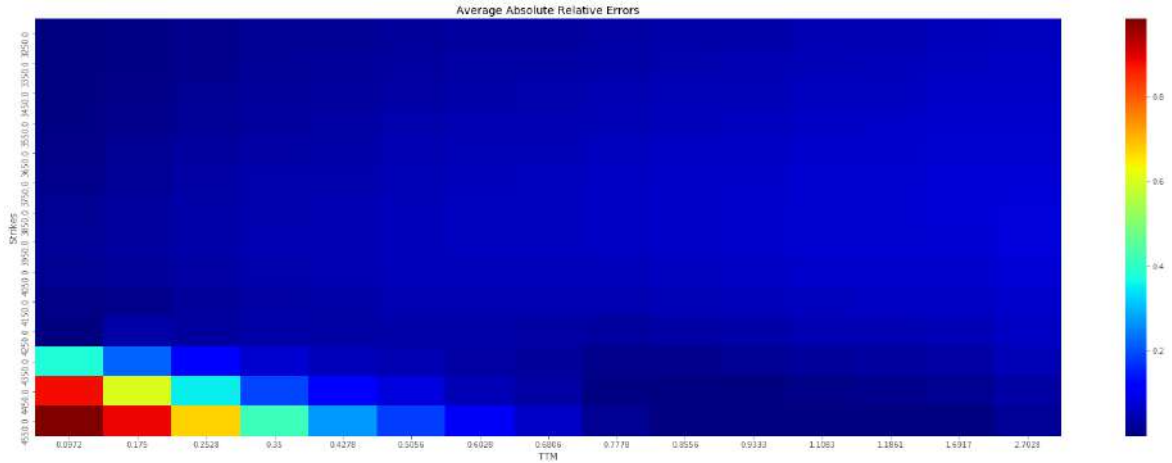


Figure 5.6: Heatmap of the AARE errors across strikes and maturities for the Heston model with jumps

5.3 The SVJJ Model

The Heston stochastic model with jumps allows us to replicate accurately the volatility skew available in the market for very short term maturities. A major drawback of Bate's model is that the correlation between the volatility and the log stock price is assumed to be 0. In other words, the SVJ model assumes that an impact on the price won't have an impact on the volatility. However, empirically it is untrue as seemingly indicated on the figure 5.7. In order to implement the SVJJ model we need to reset our main SDEs such they transform to :

$$dS_t = (r - q)S_t dt + \sigma S_t dW_t^{\mathbb{P},1} + S_t dJ_t^1$$



Figure 5.7: Evolution of the SP500 and the VIX index from 1st January 2015 to 18th June 2021

$$\begin{aligned}
 dV_t &= \kappa (\theta - V_{t-}) dt + \xi \sqrt{V_{t-}} dW_t^{\mathbb{P},2} + dJ_t^2 \\
 dJ_t^1 &= \prod_{i=1}^{dN_t} H_i - 1, \quad \mathbb{E}(dW_t^{\mathbb{P},1} dW_t^{\mathbb{P},2}) = \rho dt, \quad dJ_t^2 = d\left(\sum_{i=1}^{N_t} B_i\right) \\
 \log(H_i) &\sim N(m, \nu), \quad N_t \sim \text{Poisson}(\lambda t), \quad B_i \sim \exp\left(\frac{1}{\gamma_v}\right)
 \end{aligned}$$

By reusing the exact same scheme as with the Heston-Jump stochastic model, we are able to retrieve the original PDE of the option's price such :

$$\begin{aligned}
 rCdt &= \frac{\partial C}{\partial t} dt + \frac{1}{2} \frac{\partial^2 C}{\partial S_{-t}^2} d\langle S_{-t}, S_{-t} \rangle + \frac{\partial^2 C}{\partial S_{-t} \partial V_{-t}} d\langle S_{-t}, V_{-t} \rangle + \frac{1}{2} \frac{\partial^2 C}{\partial V_{-t}^2} d\langle V_{-t}, V_{-t} \rangle \\
 &+ \frac{\partial C}{\partial S_{-t}} dS_{-t} + \frac{\partial C}{\partial V_{-t}} dV_{-t} + \lambda dt \int_0^\infty (C(t, S_{-t} H_i, V_{-t} + \gamma) - C(t, S_{-t}, V_{-t})) p(H_i) d(H_i)
 \end{aligned}$$

As usual we translate our SDEs into a risk neutral world by applying the Girsanov theorem such it became :

$$\begin{aligned}
 dF_t &= \lambda t e^{m - \frac{v}{2}} + \sigma S_{t-} dW_t^{\mathbb{Q},1} + S_{t-} dJ_t^1 \\
 dV_t &= \kappa' (\theta' - V_{t-}) dt + \xi \sqrt{V_{t-}} dW_t^{\mathbb{Q},2} + dJ_t^2 \\
 dJ_t^1 &= \prod_{i=1}^{dN_t} H_i - 1, \quad \mathbb{E}(dW_t^{\mathbb{Q},1} dW_t^{\mathbb{Q},2}) = \rho dt, \quad dJ_t^2 = \sum_{i=1}^{dN_t} B_i \\
 \log(H_i) &\sim N(m, \nu), \quad N_t \sim \text{Poisson}(\lambda t), \quad B_i \sim \exp\left(\frac{1}{\gamma_v}\right)
 \end{aligned}$$

Where κ' and θ' are risk neutral adjusted parameters. Thereafter, we apply our change of variable $x_t = \log \frac{F_t(S_t, \tau)}{K}$, rewrite the option price C such as showed in 5.3 and take the expectation of the obtained PDE. Leading us to the following PIDE :

$$\begin{aligned}
 0 &= -\frac{\partial P}{\partial \tau} + \frac{1}{2} V_t \frac{\partial^2 P}{\partial x_t^2} - \frac{1}{2} V_t \frac{\partial P}{\partial x_t} + \frac{1}{2} \rho^2 V_t \frac{\partial^2 P}{\partial V_t^2} + \rho \xi V_t \frac{\partial^2 P}{\partial x_t \partial V_t} - \kappa' (V_t - \theta') \frac{\partial P}{\partial V_t} \\
 &+ e^{-r\tau} \lambda \int_0^\infty (P(x_{t-} + \log(H_i), V_{t-} + \gamma_v, \tau) - P(x_{t-}, V_{t-}, \tau)) p(\log(H_i)) d(\log(H_i))
 \end{aligned}$$

A good approach developed by Matytsin [9] is to link the jump of the stock price to the jump of the volatility. Therefore, a jump in the stock price simultaneously provokes a jump in the volatility. The approach is allowed when :

$$C_{SVJJ}(u, \tau) = C_{SV}(u, \tau) + \lambda \tau \left[e^{i u m - \frac{u^2 v}{2}} I(u, \tau) - 1 \right] - i u \lambda \tau \left(\frac{e^{m - \frac{v}{2}}}{\theta} \right)$$

$$L_{SVJJ}(u, \tau) = L(u, \tau)$$

Where $I(u, \tau) = \frac{1}{T} \int_0^\tau e^{\gamma_v L(u, t)} dt$. The interpretation is mainly that a jump in the stock price will provoke a follow up jump in the variance such $V_t = V_{t-} + \gamma_v$. However, multiple studies, among them in [10], showed that the SVJJ model doesn't improve substantially the volatility skew obtained under the SVJ model. In contrary, it adds more time to calibrate and difficulties to minimize mean square errors so as to fit market prices.

5.4 The SVI Model

The SVI implied volatility model is a parametric model for stochastic implied volatility. It aims to provide a simple function for which we calibrate the implied volatility following the wanted tenor. The first SVI model had been disclosed by Jim Gatheral in [11]. Its main advantage resides in its easy application for traders because they only have to remember the parameters to plug into the model, no characteristic function and no Dirichlet integrals guaranteed. The SVI model as its name indicates, exists for a probability space $(\Omega, \mathcal{F}_t, P)$, for \mathcal{F} the possible σ -algebra sets available at time t , $\forall t \in \mathbb{R}_+$ without a possibility of arbitrage. The most 'Raw' model is set as :

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b \left(\rho(x - m) + \sqrt{(x - m)^2 + \sigma^2} \right) \quad (5.14)$$

Where :

- w_{imp} is the Total implied variance
- a is the vertical shift of the variance
- b is the slope of the variance across Log moneyness
- ρ is the correlation of the variance across Log moneyness and represents the rotation of the smile around the log moneyness
- m is the distortion level of the smile starting from the log moneyness
- σ is the level of the curvature of the smile

Properly speaking, the SVI is more of a volatility skew fitter than a market price replicator. One of the originality of this parametrization comes from its dependency to τ as the total implied variance is expressed as $w_{\text{IV}} = \sigma(K, \tau)_{\text{IV}}^2 \tau$. In order for the SVI to be free of arbitrage opportunities, meaning to avoid free trading gains and

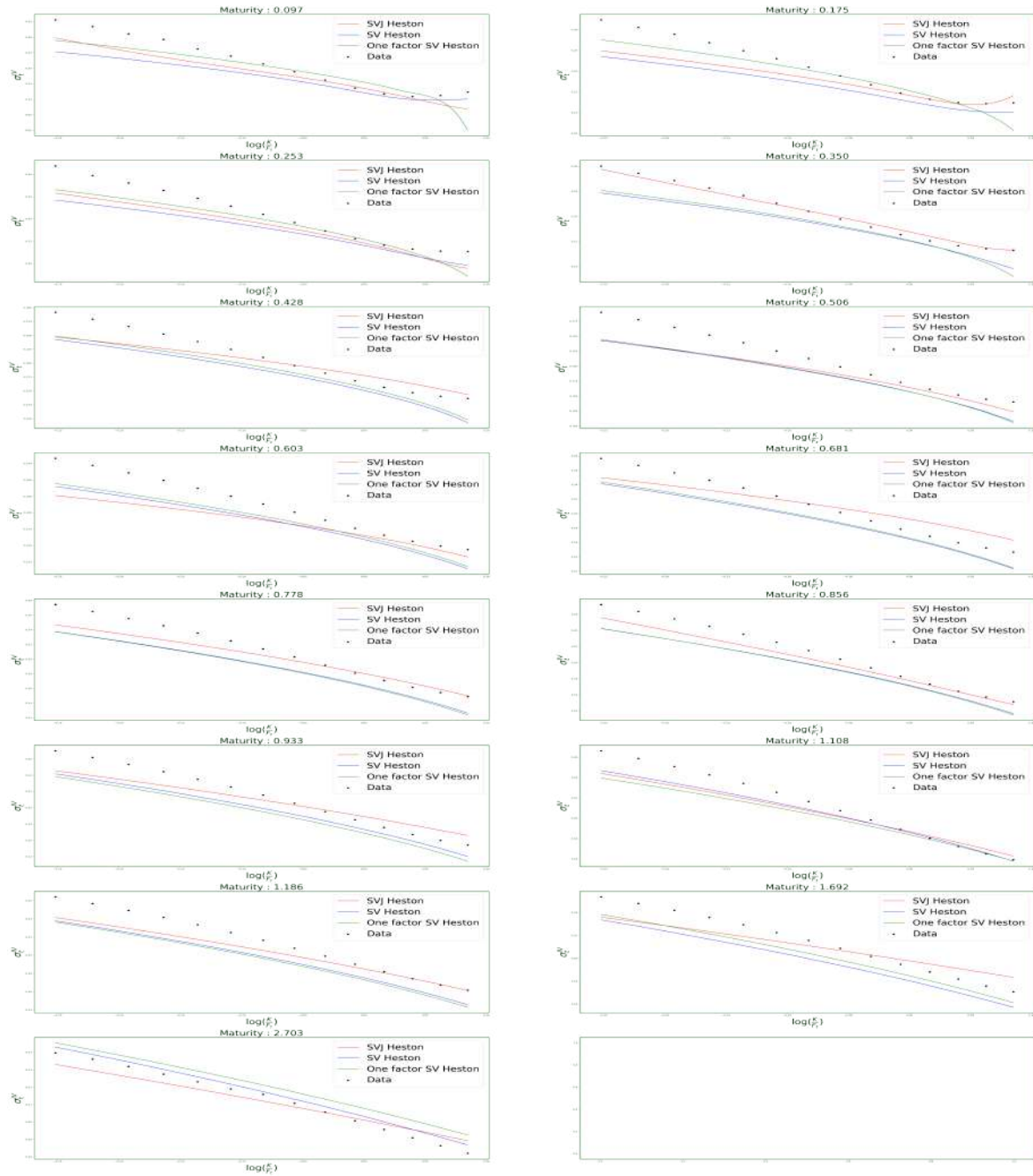


Figure 5.8: Stochastic volatility model skews, the one factor Heston is created using a constrained argument such $\rho = -1$

negative option prices requires in a first time to operate a variable change on the option price defined as :

$$\begin{aligned}
 C_{BS}(S_0, K, \sigma_{IV}(K, \tau), \tau) &= C(S_0, F_0 e^x, w, \tau) \\
 &= C(S_0, K, w, \tau) \\
 &= F_0 \left(N(d_1) - e^x N(d_2) \right)
 \end{aligned}$$

Where $x = \log \frac{K}{F_0}$. As we introduce the total implied variance into the Black Scholes equation we need to rewrite the terms d_1 and d_2 in a way of Black did in 1976 on his paper [12]. Meaning we need to pass through the :

$$\begin{aligned} d_1 &= \frac{\ln \frac{S_0}{K} + \int_0^\tau (r - q) dt + \frac{w}{2}}{\sqrt{w}} = -\frac{x}{\sqrt{w}} + \frac{\sqrt{w}}{2} \\ d_2 &= d_1 - \sqrt{w} \\ F_0 &= S_0 e^{\mu_\tau} \\ \mu_\tau &= (r - q)\tau \end{aligned}$$

By applying the chain rule we shall finally derive easily the first 2 required derivatives in a way:

$$\begin{aligned} \frac{\partial C_{BS}}{\partial \tau} &= \frac{\partial C}{\partial \tau} + \frac{\partial C}{\partial K} \frac{\partial K}{\partial \tau} \\ &= \frac{\partial C}{\partial \tau} + \frac{\partial C}{\partial K} K \mu_\tau \\ &= \frac{\partial C}{\partial \tau} + \frac{\partial C}{\partial x} \mu_\tau \end{aligned} \quad (5.15)$$

$$\text{By substitution as } \frac{\partial C_{BS}}{\partial x} = \frac{\partial C}{\partial K} \frac{\partial K}{\partial x} = \frac{\partial C}{\partial K} K \quad (5.16)$$

The third derivative is straightforward to compute such :

$$\frac{\partial^2 C_{BS}}{\partial K^2} K^2 = \frac{\partial^2 C}{\partial x^2} - \frac{\partial C}{\partial x} \quad (5.17)$$

Therefore, substituting 5.15, 5.16, 5.17 on the original Dupire's formula gives the subsequent formula :

$$\frac{\partial C_{BS}}{\partial \tau} = \frac{\sigma_w^2}{2} \left[\frac{\partial^2 C}{\partial x^2} - \frac{\partial C}{\partial x} \right] + \mu_\tau C$$

The derivation is still not finished. We need to express the local volatility in term of the total variance w . In order to do that some derivations are mandatory to express $\frac{\partial^2 C_{BS}}{\partial x^2}$ in terms of total implied volatility. The first derivation is done such :

$$\begin{aligned} \frac{\partial C_{BS}}{\partial w} &= F_0 \left[N'(d_1) \frac{\partial d_1}{\partial w} - e^x N'(d_2) \frac{\partial d_2}{\partial w} \right] \\ &= F_0 \left[N'(d_2) e^x \left(\frac{\partial d_2}{\partial w} + \frac{1}{2} w^{-\frac{1}{2}} \right) - e^x N'(d_2) \frac{\partial d_2}{\partial w} \right] \\ &= \frac{1}{2} F_0 e^x \left[N'(d_2) w^{-\frac{1}{2}} \right] \end{aligned}$$

The second derivation is obtained through :

$$\frac{\partial^2 C_{BS}}{\partial w^2} = \frac{1}{2} F_0 e^x \left[-N'(d_2) d_2 \frac{\partial d_2}{\partial w} w^{-\frac{1}{2}} - \frac{1}{2} N'(d_2) w^{-\frac{3}{2}} \right]$$

$$\begin{aligned}
&= \frac{1}{2} F_0 e^x N'(d_2) w^{-\frac{1}{2}} \left[-d_2 \frac{\partial d_2}{\partial w} - \frac{1}{2} w^{-1} \right] \\
&= \frac{\partial C}{\partial w} \left[\left(x w^{-\frac{1}{2}} + \frac{1}{2} w^{\frac{1}{2}} \right) \left(\frac{1}{2} x w^{-\frac{3}{2}} - \frac{1}{4} w^{-\frac{1}{2}} \right) - \frac{1}{2} w^{-1} \right] \\
&= \frac{\partial C}{\partial w} \left[-\frac{1}{8} - \frac{1}{2w} + \frac{x^2}{2w^2} \right]
\end{aligned}$$

the third derivative is :

$$\begin{aligned}
\frac{\partial^2 C_{BS}}{\partial w \partial x} &= \frac{1}{2} F_0 w^{-\frac{1}{2}} \frac{\partial}{\partial x} [e^x N'(d_2)] \\
&= \frac{1}{2} F_0 w^{-\frac{1}{2}} [e^x N'(d_2) - e^x N'(d_2) d_2 \partial_x d_2] \\
&= \frac{\partial C}{\partial w} [1 - d_2 \partial_x d_2] \\
&= \frac{\partial C}{\partial w} \left(\frac{1}{2} - \frac{x}{w} \right)
\end{aligned}$$

And finally, we obtain our most needed Graal :

$$\begin{aligned}
\frac{\partial^2 C_{BS}}{\partial x^2} &= -F_0 [e^x N(d_2) + e^x N'(d_2) \partial_x d_2] \\
&= -F_0 e^x N(d_2) + F_0 e^x N'(d_2) w^{-\frac{1}{2}} \\
&= \frac{\partial C}{\partial x} + 2 \frac{\partial C}{\partial w} \\
\text{With } \frac{\partial C}{\partial x} &= F_0 e^x N(d_2)
\end{aligned}$$

Re-applying the Chain rule transforms the precedent result into:

$$\begin{aligned}
\frac{\partial^2 C_{BS}}{\partial x^2} &= \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial x \partial w} \frac{\partial w}{\partial x} + \frac{\partial C}{\partial w} \frac{\partial^2 w}{\partial x^2} + \left[\frac{\partial^2 C}{\partial w \partial y} + \frac{\partial^2 C}{\partial w^2} \frac{\partial w}{\partial x} \right] \frac{\partial w}{\partial x} \\
&= \frac{\partial^2 C}{\partial x^2} + 2 \frac{\partial^2 C}{\partial x \partial w} \frac{\partial w}{\partial x} + \frac{\partial C}{\partial w} \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 C}{\partial w^2} \left(\frac{\partial w}{\partial x} \right)^2
\end{aligned}$$

Using the fact that $w = w(x)$ we can derive the last two most needed derivatives :

$$\begin{aligned}
\frac{\partial C_{BS}}{\partial x} &= \frac{\partial C}{\partial x} + \frac{\partial C}{\partial w} \frac{\partial w}{\partial x} \\
\frac{\partial C_{BS}}{\partial \tau} &= \frac{\partial C}{\partial \tau} + \frac{\partial C}{\partial w} \frac{\partial w}{\partial \tau} \\
&= \mu_\tau C + \frac{\partial C}{\partial w} \frac{\partial w}{\partial \tau}
\end{aligned}$$

Afterward, we can derive straightforwardly a deterministic formula to price options such as proposed by Dupire's formula 5.4 such :

$$\frac{\partial C_{BS}}{\partial w} \frac{\partial w}{\partial \tau} = \frac{\sigma_{lv}}{2 \left[\frac{\partial^2 C}{\partial x^2} + 2 \frac{\partial^2 C}{\partial x \partial w} \frac{\partial w}{\partial y} + \frac{\partial C}{\partial w} \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 C}{\partial w^2} \left(\frac{\partial w}{\partial y} \right) - \frac{\partial C}{\partial x} + \frac{\partial C}{\partial w} \frac{\partial w}{\partial x} \right]}$$

By factoring and removing the factor $\frac{\partial C}{\partial w}$ on both side we reformulate 5.4 such :

$$\frac{\partial w}{\partial \tau} = \sigma_{lv}^2 \left[1 - \frac{x}{w} \frac{\partial w}{\partial x} + \frac{1}{2} \frac{\partial^2 w}{\partial x^2} + \frac{1}{4} \left(-\frac{1}{4} - \frac{1}{w} + \frac{y^2}{w} \right) \left(\frac{\partial w}{\partial x} \right)^2 \right] \quad (5.18)$$

Finally, we rewrite 5.18 and obtain something a similar formula of the local variance such :

$$\sigma_{lv}^2 = \frac{\frac{\partial w}{\partial \tau}}{\left[1 - \frac{x}{w} \frac{\partial w}{\partial x} + \frac{1}{2} \frac{\partial^2 w}{\partial x^2} + \frac{1}{4} \left(-\frac{1}{4} - \frac{1}{w} + \frac{x^2}{w} \right) \left(\frac{\partial w}{\partial x} \right)^2 \right]}$$

In order for our local variance to be positive and therefore eliminate risks of having a calendar spread and a butterfly arbitrage, which consists into increasing personal wealth instantaneously by simply buying an option. A simple way of removing these kinds of arbitrages is to implement some constraints defined as :

$$w \geq 0 \quad \forall x \in \mathbb{R}, \forall \tau \in [0, \infty[$$

$$\frac{\partial w}{\partial \tau} \geq 0 \quad \forall x \in \mathbb{R}, \forall \tau \in [0, \infty[\quad (5.19)$$

$$g(x) = 1 - \frac{x}{w} \frac{\partial w}{\partial x} + \frac{1}{2} \frac{\partial^2 w}{\partial x^2} + \frac{1}{4} \left(-\frac{1}{4} - \frac{1}{w} + \frac{x^2}{w} \right) \left(\frac{\partial w}{\partial x} \right)^2 \geq 0 \quad (5.20)$$

$$\forall x \in \mathbb{R}, \forall \tau \in [0, \infty[, g : \mathbb{R} \rightarrow \mathbb{R}$$

$$\lim_{x \rightarrow \infty} d_2 = -\infty \quad \forall x \in \mathbb{R}, \forall \tau \in [0, \infty[\quad (5.21)$$

One might ask why are we talking about local volatility in the stochastic volatility chapter. It is because local volatility originally inspired the principle of stochastic volatility. Concerning the constraints, 5.19 is easily verified and its direct meaning is that across time the Total implied variance increases. However, 5.20 means that a negative value would involve that the probability density function at the strike price $\frac{\partial^2 C}{\partial K^2}$ can be negative using Fokker-Plank's equation, which cannot be true. This condition is sometimes called the Durrleman's condition.

In the meantime, SVI has an interesting condition 5.19 that doesn't allow traders to take advantages of calendar spreads unlike in other stochastic volatility models.

The proof of 5.21 demonstrates the behavior of d_2 towards very large in-moneyness, using the arithmetic-geometric mean lemma such :

$$\begin{aligned}
\lim_{x \rightarrow \infty} d_2 &= -\frac{x}{\sqrt{w}} - \frac{\sqrt{w}}{2} \\
&= -\frac{1}{2} \left(\frac{2x}{\sqrt{w}} + \sqrt{w} \right) \\
&\leq -\sqrt{\frac{2x\sqrt{w}}{\sqrt{w}}} \\
&\leq -\sqrt{2x} \\
&\leq -\infty
\end{aligned}$$

However, even if the SVI allows to have good fits very easily and rapidly. A main problem persists : the SVI is not intuitive at all for traders. The parameters only wear mathematical understandings because the whole process is centered around total implied variance w , which traders are not familiar with. In order to fix the problem, a reparametrization, called SVI-JW, was developed such the main parameters become :

$$\begin{aligned}
v &= \frac{a + b \left\{ -\rho m + \sqrt{m^2 + \sigma^2} \right\}}{\tau} \\
\psi &= \frac{1}{\sqrt{w}} \frac{b}{2} \left(-\frac{m}{\sqrt{m^2 + \sigma^2}} + \rho \right) \\
p &= \frac{1}{\sqrt{w}} b(1 - \rho) \\
c &= \frac{1}{\sqrt{w}} b(1 + \rho) \\
\tilde{v} &= \frac{1}{\tau} \left(a + b\sigma\sqrt{1 - \rho^2} \right)
\end{aligned}$$

The SVI-JW parameters are defined as :

- v is the at-the-money variance. It is defined as :

$$- v = \frac{w}{\tau} \Big|_{x=0}$$

- ψ is the at-the-money skew. its derivation is straightforward :

$$\begin{aligned}
- \frac{\sqrt{\frac{w_{\text{imp}}^{\text{SVI}}}{\tau}}}{\partial x} \Big|_{x=0} &= \frac{1}{2\sqrt{w_{\text{imp}}^{\text{SVI}}\tau}} \Big|_{x=0} \frac{w}{\partial x} \Big|_{x=0} \\
&= \frac{1}{2\sqrt{v\tau}} b \left(\rho + \frac{2x-2m}{2\sqrt{(x-m)^2 + \sigma^2}} \right) \Big|_{x=0} \\
&= \frac{1}{\sqrt{w\tau}} \frac{b}{2} \left(\rho - \frac{m}{\sqrt{m^2 + \sigma^2}} \right) \\
&= \psi_{\tau}
\end{aligned}$$

- p is the slope for deep out-of-the-money puts (left wing). Its derivation involves the use of limits such :

$$\begin{aligned}
- \lim_{x \rightarrow -\infty} \sqrt{\frac{w}{x\tau}} &= \lim_{x \rightarrow -\infty} \sqrt{\frac{\frac{1}{x} \left(a + b(\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}) \right)}{\tau}} \\
&= \lim_{x \rightarrow -\infty} \sqrt{\frac{1}{\tau} \left(\frac{a}{x} + b(\rho(1 + \frac{m}{x}) + \sqrt{(1 + \frac{m}{x})^2 + \frac{\sigma^2}{x^2}}) \right)} \\
\text{We can notice for } x < 0, (1 + \frac{m}{x}) &= (-1 - \frac{m}{|x|}), \text{ then:} \\
&= \sqrt{\frac{1}{\tau} \left(b(1 - \rho) \right)} = \sqrt{\frac{p\sqrt{w}}{\tau}} \\
p &= \frac{b}{\sqrt{w}} (1 - \rho)
\end{aligned}$$

- c is the slope for deep out-of-the-money calls (right wing). Its derivation is very similar to p :

$$\begin{aligned}
- \lim_{x \rightarrow \infty} \sqrt{\frac{w}{x\tau}} &= \lim_{x \rightarrow \infty} \sqrt{\frac{\frac{1}{x} \left(a + b(\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}) \right)}{\tau}} \\
&= \lim_{x \rightarrow \infty} \sqrt{\frac{1}{\tau} \left(\frac{a}{x} + b(\rho(1 - \frac{m}{x}) + \sqrt{(1 - \frac{m}{x})^2 + \frac{\sigma^2}{x^2}}) \right)} \\
\text{We can notice for } x > 0, (1 - \frac{m}{x}) &= (1 - \frac{m}{|x|}), \text{ then:} \\
&= \sqrt{\frac{1}{\tau} \left(b(1 + \rho) \right)} = \sqrt{\frac{c\sqrt{w}}{\tau}} \\
c &= \frac{b}{\sqrt{w}} (1 + \rho)
\end{aligned}$$

- \tilde{v} is the minimum implied variance. We can easily find a minimum by equalizing the derivative ψ_τ as our function is strictly convex, therefore :

$$\begin{aligned}
- 0 &= \frac{(x^* - m)}{\sqrt{(x^* - m)^2 + \sigma^2}} + \rho \\
(x^* - m)^2 &= \frac{\sigma^2 \rho^2}{1 - \rho^2} \\
x^* &= -\frac{\sigma \rho}{\sqrt{1 - \rho^2}} + m
\end{aligned}$$

Replacing x by x^* into 5.14, we find the minimum total implied variance such :

$$\tilde{v} = \frac{1}{\tau} \left(a + b\sigma\sqrt{1 - \rho^2} \right)$$

Finally our parametrization becomes extremely easy to understand for traders. Moreover the fit it provides are good. However, to be 100% sure that our SVI is free of arbitrage Gatheral and Jaquier re-picked the minimum implied volatility and the deep out-of-the-money call's slope such :

$$\begin{aligned}
c' &= p + 2\psi \\
\tilde{v}' &= \tilde{v} \frac{4pc'}{(p + c')^2}
\end{aligned} \tag{5.22}$$

On the figure 5.9, we can see that the left tail on the Durrleman's condition, $g(k)$, of the very short maturity appears to stabilize very quickly using the arbitrage free SVI in contrary to the unconstrained SVI, which offers best fits but without certainty that arbitrage opportunities won't appear. Later in time, Gatheral and Jaquier developed

the SSVI, called the Surface SVI. The specificity of the former is that we express our SVI in terms of the at-the-money variance time instead of the standard calendar time. In other words, we are trying to give more weights to the ATM total implied variance for the fit. Despite sounding more involved, SSVI allows us to pass from 5 to 3 parameters to estimate, such we start from :

$$w = \Delta + \frac{\omega}{2} \left(1 + \zeta \rho (x - \mu) + \sqrt{(\zeta (x - \mu) + \rho)^2 + (1 - \rho^2)} \right)$$

To end up with :

$$w = \frac{\theta_\tau}{2} \left(1 + \rho \phi(\theta_\tau) x + \sqrt{(\phi(\theta_\tau) x + \rho)^2 + (1 - \rho^2)} \right)$$

$$\text{with } \theta_\tau = \sigma_{ATMBS}^2 \tau$$

$$\text{and } \phi(\theta_\tau) = \frac{\eta}{\theta^\gamma} \text{ With } \phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \forall x \in \mathbb{R}, \forall t \in \mathbb{R}_+, \eta \geq 0, \gamma \geq 0, |\rho| \leq 1$$

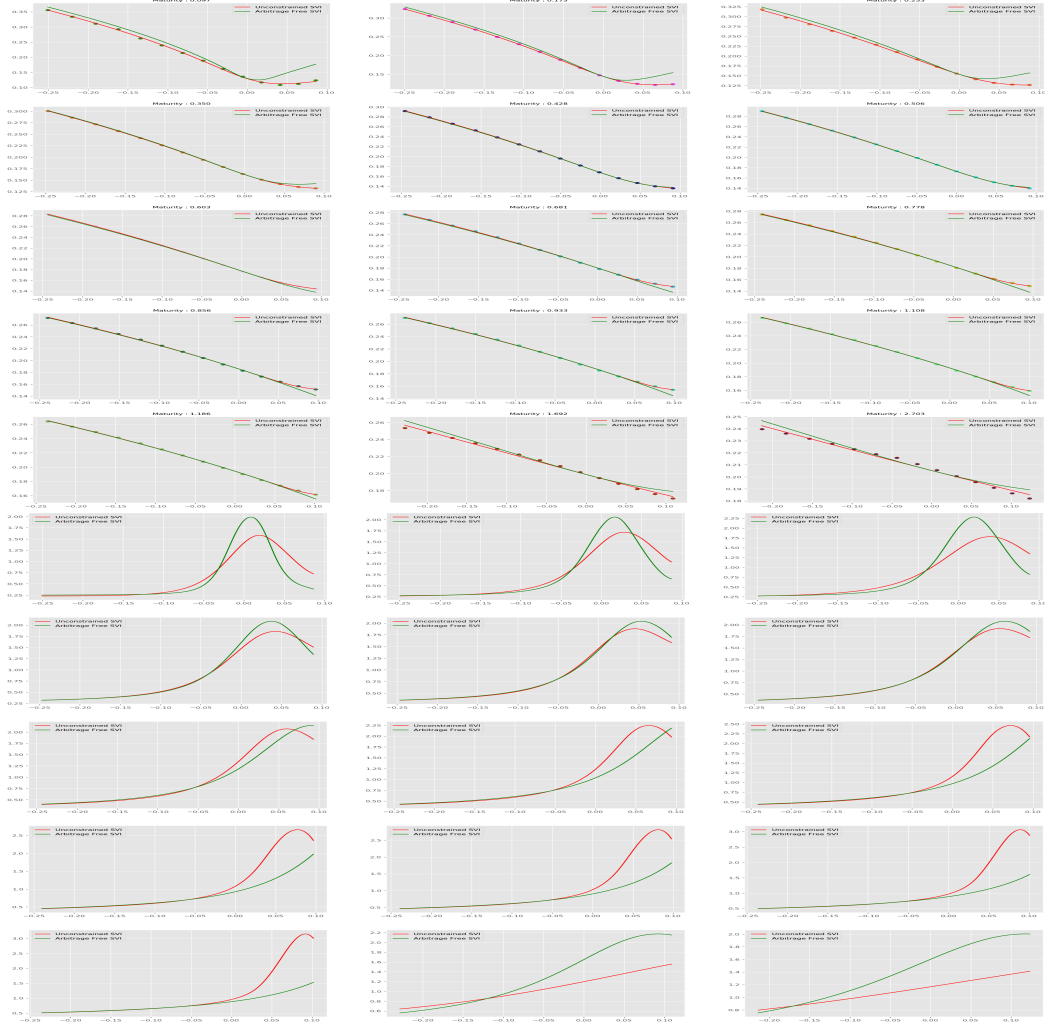


Figure 5.9: Stochastic volatility "inspired" of the S&P500, the 16th April 2021 and its Durrleman's condition g)

Therefore, by using analogous calculation previously made to find SVI-JW, the SSVI-JW parameters are exemplified as :

$$\begin{aligned} v &= \frac{\theta_\tau}{\tau} \\ \psi &= \frac{1}{2}\rho\sqrt{\theta_\tau}\phi(\theta_\tau), \\ p &= \frac{1}{2}\sqrt{\theta_\tau}\phi(\theta_\tau)(1-\rho) \\ c &= \frac{1}{2}\sqrt{\theta_\tau}\phi(\theta_\tau)(1+\rho) \\ \tilde{v} &= \frac{\theta_\tau}{\tau}(1-\rho^2) \end{aligned}$$

We can now by extreme easiness, which is the strong point of SVI models, express a volatility surface without calendar spread, neither butterfly arbitrage using the two reformulated conditions:

Firstly, The SSVI surface is free of calendar-spread arbitrage if and only if:

$$\begin{aligned} \frac{\partial \theta_\tau}{\partial \tau} &\geq 0 \quad \forall \tau \geq 0 \\ 0 &\leq \frac{\partial \theta_\tau \phi(\theta_\tau)}{\partial \theta_\tau} \leq \frac{1}{\rho^2} \left(1 + \sqrt{1 - \rho^2}\right) \phi(\theta_\tau) \quad \forall \theta_\tau > 0 \end{aligned} \quad (5.23)$$

We reformulated the original condition 5.4 using the chain rule such :

$$\frac{\partial w}{\partial \tau} = \frac{\partial w}{\partial \theta_\tau} \frac{\partial \theta_\tau}{\partial \tau}$$

Proof of 5.23.

$$\begin{aligned} \frac{\partial w}{\partial \theta_\tau} &= \frac{1}{2} \left(1 + \rho \phi(\theta_\tau)x + \sqrt{(\phi(\theta_\tau)x)^2 + 2\rho\phi(\theta_\tau)x + 1} + \frac{\partial \phi(\theta_\tau)}{\partial \theta_\tau} \rho x \theta_\tau \right. \\ &\quad \left. + x \frac{\partial \phi(\theta_\tau)}{\partial \theta_\tau} \frac{\theta_\tau(\phi(\theta_\tau)x + \rho)}{\sqrt{(\phi(\theta_\tau)x)^2 + 2\rho\phi(\theta_\tau)x + 1}} \right) \end{aligned}$$

we can rewrite $\theta_\tau \frac{\partial \phi(\theta_\tau)}{\partial \theta_\tau}$ as being equal to : $\frac{\partial \theta_\tau \phi(\theta_\tau)}{\partial \theta_\tau} - \phi(\theta_\tau)$ and $k = x\phi(\theta_\tau)$

$$\begin{aligned} &= \frac{1}{2} \left(1 + k\rho + \sqrt{k^2 + 2k\rho + 1} + \left(\frac{\partial \theta_\tau \phi(\theta_\tau)}{\partial \theta_\tau} \frac{x}{\phi(\theta_\tau)} - x \right) \left(\frac{x + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right) \\ &= \frac{1}{2} \left(1 + k\rho + \frac{k^2 + 2\rho k + 1 - k^2 - k\rho}{\sqrt{k^2 + 2k\rho + 1}} - k\rho + k \frac{\partial \theta_\tau \phi(\theta_\tau)}{\partial \theta_\tau} \frac{1}{\phi(\theta_\tau)} \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right) \\ &= \frac{1}{2} \left(1 + \frac{k\rho + 1}{\sqrt{k^2 + 2k\rho + 1}} + \frac{\partial \theta_\tau \phi(\theta_\tau)}{\partial \theta_\tau} \frac{k}{\phi(\theta_\tau)} \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right) \end{aligned} \quad (5.24)$$

We made appear our derivative 5.24 linear such:

$$\frac{\partial w}{\partial \theta_\tau} = \frac{1}{2} \left(\psi_0(k, \rho) + \gamma(\theta) \psi_1(k, \rho) \right)$$

With :

$$\begin{aligned} \psi_0(k, \rho) &= 1 + \frac{k\rho + 1}{\sqrt{k^2 + 2k\rho + 1}} \\ \psi_1(k, \rho) &= k \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \end{aligned}$$

For any $|\rho| < 1$, $\psi_0(k, \rho)$ is strictly positive :

$$1 + \frac{1 + k\rho}{\sqrt{k^2 + 2k\rho + 1}} = \frac{\sqrt{k^2 + 2k\rho + 1} + 1 + k\rho}{\sqrt{k^2 + 2k\rho + 1}}$$

The square root denominator cannot be negative, only the numerator could. In order to show that the numerator is strictly positive, we need to use a valid identity such :

$$\sqrt{k^2 + 2k\rho + 1} > \sqrt{k^2 + 2k\rho + \rho^2} = |k + \rho|$$

Therefore :

$$\begin{aligned} \sqrt{k^2 + 2k\rho + 1} + 1 + k\rho &> |k + \rho| + 1 + k\rho \\ \text{Knowing that } |k + \rho| + 1 &> |k| \quad \forall |\rho| < 1 \\ \iff |k + \rho| + 1 + k\rho &> |k| + k\rho \\ &> 0 \end{aligned}$$

However, $\psi_1(k, \rho)$ is strictly positively defined on the intervals \mathcal{M}_ρ such :

$$\mathcal{M}_\rho = \begin{cases} (-\infty, 0) \cup (-2\rho, \infty), & \text{if } \rho < 0 \\ (-\infty, -2\rho) \cup (0, \infty), & \text{if } \rho > 0 \\ \mathbb{R} \setminus \{0\}, & \text{if } \rho = 0 \end{cases}$$

The definition intervals are straightforward to find as we are in presence of a linear and a quadratic equation. Then it follows that :

$$\frac{\partial w}{\partial \theta_\tau} \geq 0 \text{ if and only if } \begin{cases} \gamma(\theta) \geq -\frac{\psi_0(x, \rho)}{\psi_1(x, \rho)}, & \text{for } x \in \mathcal{M}_\rho, \\ \gamma(\theta) \leq -\frac{\psi_0(x, \rho)}{\psi_1(x, \rho)}, & \text{for } x \in \mathbb{R} \setminus (\mathcal{M}_\rho \cup \{0, -2\rho\}), \end{cases}$$

The upper bound of the condition 5.23, is obtained by finding the supremum of $\gamma(\theta)$ on the domain $\mathcal{M}_\rho \cup \{0, -2\rho\}$ such we have :

$$\inf_{x \in \mathbb{R} \setminus (\mathcal{D}_\rho \cup \{0, -2\rho\})} -\frac{\psi_0(x, \rho)}{\psi_1(x, \rho)} = \frac{1 + \sqrt{1 - \rho^2}}{\rho^2}$$

□

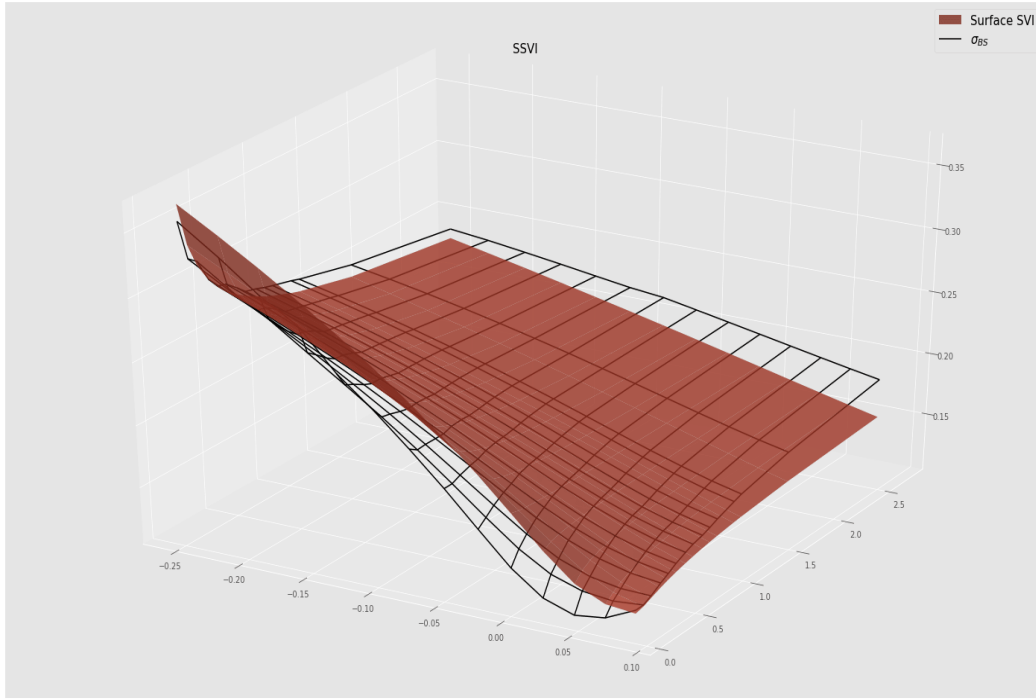


Figure 5.10: Comparison between SSVI of the S&P500 and the market's implied volatility surface ,the 16th April 2021.

The supremum for positive $\psi_1(x, \rho)$ is equal to 0, as the function keeps tending towards 0, without reaching it for $|x| \rightarrow \infty$. Finally, the last 2 conditions are meant to avoid the SSVI to have butterfly arbitrage and are listed such :

$$\begin{aligned} \theta\phi(\theta_\tau)(1 + |\rho|) &< 4 \\ \theta\phi(\theta_\tau)^2(1 + |\rho|) &\leq 4 \end{aligned} \quad (5.25)$$

The two conditions are obtained using the Durrleman's condition 5.20. Proofs are voluntarily omitted and available at [13]. Meanwhile the SSVI and SVI with free of static arbitrages are originally inspired from the Heston stochastic volatility model, which fails to accommodate the extreme tails of returns' and treats volatility as a diffusion process showed on the following figure 7.1.

Finally, stochastic volatility models have strength in the sense that their parameters have a clear economic meaning. For instance, across maturity we can see the anticipation of the market presented on the figure 5.11. However, as cited previously stochastic volatility models fail to reproduce the real volatility distribution and are hard to fit to market prices because of the high number of parameters. Moreover, the very short maturity term structure is hard to reproduce without using jump models.

However, Even we add jumps to the model, it makes the calibration harder to fit market's price as a direct consequence of adding more dimensions to the already involved optimization. Had the stochastic volatility 'inspired' brought an original model, where the calibration is relatively easy and where the prices are not uphold to static

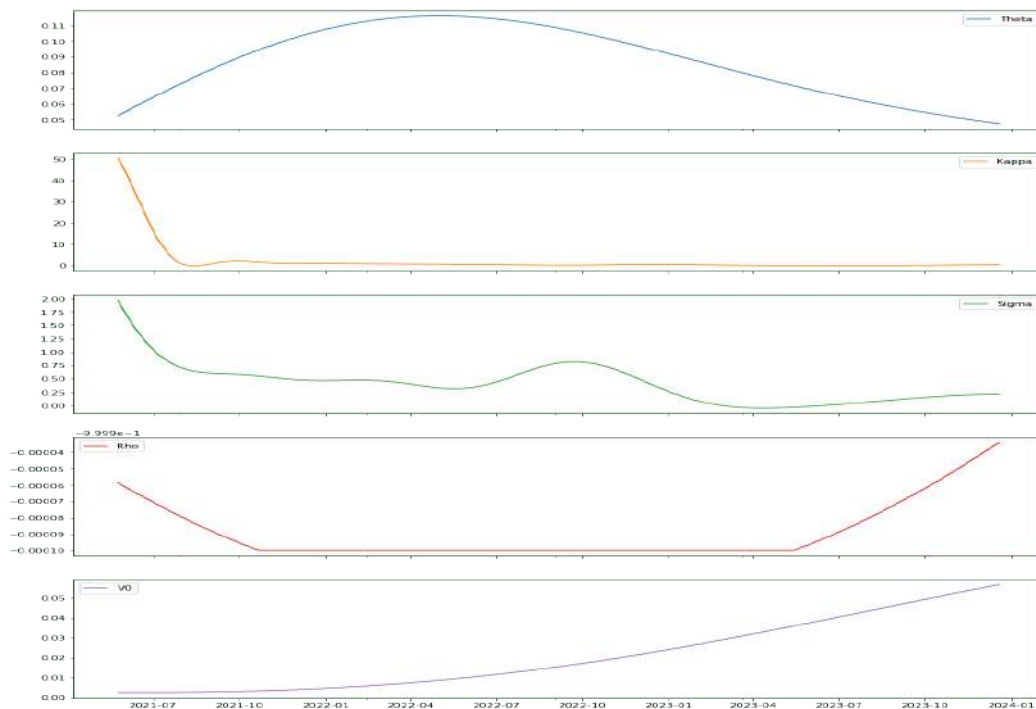


Figure 5.11: Heston Parameters across maturities ,the 20th April 2021.

arbitrage opportunities, it may produce a relevant volatility surface but doesn't give a more realistic dynamic insight about the volatility. Nonetheless, it is not surprising that the SVI dynamic is not very relevant because Gatheral, on his paper [13] showed that the SVI is a limit case of a standard Heston model.

Chapter 6

Bergomi's Model

The Bergomi model presents an interesting way of pricing and modelling options using forward variance developed by L.Bergomi on [14]. An advantage of such model is that forward variance can be traded on secondary markets even if the lack of standardized market would force us to carry the position on the realized variance unlike on the Heston model because of the presence of very liquid instruments to hedge the volatility, such as futures and options on the VIX index. In order to set up the Bergomi model, we need to issue an assumption where the forward variance only depends on the time to maturity of our option. Before that we need to clarify that the fair strike of a variance swap at time t , for a maturity T , is computed by using an average of forward variance for a time duration of $T - t$ such :

$$\sigma_T^2(T) = \frac{1}{T-t} \int_t^T \xi_s^T ds$$

We can therefore deduce the direct formula of the forward variance ξ_t^T as :

$$\xi_t^T = \frac{1}{dT} d\left((T-t)\sigma_t^2(T)\right)$$

Definition 6 (Instantaneous Volatility). *The instantaneous variance is a forward variance contract that expires instantaneously for an increment $\delta T \rightarrow 0$ between the receipt and the terminal repayment. Through the chapter we will note it $\xi_t^{T=t}$.*

on the two following sections, we will discuss about the technical insights this model is offering us :

6.1 One factor model

The SDE of a one factor Bergomi model developed by L.Bergomi and explained in his book [14] is given such :

$$d\xi_t^T = \omega(\tau)\xi_t^T dW'_t$$

- ξ_t^T is the instantaneous forward variance.

- $\omega(\tau)$ is a weighted average of the future forward volatility at time t for an horizon of T .
- W'_t is expressed here as standard Brownian path with length t .
- $\log(\xi_t^T) \sim N\left(\log(\xi_0^T) - \frac{1}{2} \int_0^t \omega^2(T-s)ds, \int_0^t \omega^2(T-s)ds\right)$. The proof is trivial and similar to the one of the Black and Scholes.

The probability space is always $(\Omega, \mathcal{F}_t, \mathbb{P}) \forall t \in \mathbb{R}_+$, as exemplified on chapter 1. In order to have the possibility of hedging our position through variance swap strikes we need to have a mean reversible process from 0 to our maturity t , explaining the dynamic of our variance evolution such our dW'_t follows a mean reversion path across time modeled by the equations :

$$\begin{aligned} \omega(u) &= \omega e^{-\kappa u} \\ \int_0^t \omega(T-s) dW'_s &= e^{-\kappa T} \int_0^t \omega e^{\kappa s} dW'_s \end{aligned}$$

Doesn't it seem like an Ornstein-Uhlenbeck (OU) process of the form :

$$dX_t = -\kappa \omega X_t dt + \omega dW_t^{\mathbb{P},2}, \quad X_0 = 0$$

Thus, we rewrite our initial equation of the forward volatility from Bergomi's model using this time, a standard Brownian motion W_t :

$$d\xi_t^T = e^{-\kappa(T-t)} \xi_t^T dX_t \quad (6.1)$$

Such that $X_t = \omega(\tau) dW'_t$. After that, we obtain the hedgeable forward variance solution as we can control the volatility of the volatility skew of our model such :

$$\xi_t^T = \xi_0^T \exp(e^{-\kappa(T-t)} X_t - \frac{e^{-2\kappa(T-t)}}{2} E[X_t^2]) \quad (6.2)$$

Proof of 6.1. First of all, we need to use a variation of parameters such that we can use a Taylor serie of degree 1 given by :

$$\begin{aligned} df(X_t, t) &= \frac{f(X_t, t)}{\partial X_t} dX_t + \frac{f(X_t, t)}{\partial t} dt \\ d(X_t e^{\kappa t}) &= -e^{\kappa t} \kappa \omega X_t dt + e^{\kappa t} \kappa \omega X_t dt + \omega e^{\kappa t} dW_t^{\mathbb{P},1} \end{aligned}$$

We set $f(X_t, t) = X_t e^{\kappa t}$ and by integration from t to 0 obtain :

$$\begin{aligned} X_t e^{\kappa t} &= \int_0^t \omega e^{\kappa s} dW_s^{\mathbb{P},2} \\ X_t &= \int_0^t \omega e^{-\kappa(t-s)} dW_s^{\mathbb{P},2} \end{aligned}$$

Through Ito's isometry we can derive the term X_t^2 such :

$$\begin{aligned}
E[X_t^2] &= \int_0^t \omega^2 e^{-2\kappa(t-s)} ds \\
&= \omega^2 \left(\frac{1 - e^{-2\kappa t}}{2\kappa} \right)
\end{aligned}$$

□

it is now straightforward to derive the relation 6.4. We define ω as the log-normal volatility of $\xi_t^{T=t}$, the instantaneous log-normal volatility at time t to make it simple. As shown by 7.4, each forward curve depends solely on its standard Brownian motion. For this reason we are forced to use Monte Carlo simulations in order to price our options.

6.2 N factors model

A major drawback of a single factor model is that our log-normal volatility of the volatility will only depend on a single Brownian movement such as it is done in the Heston model, which is inaccurate. L. Bergomi showed [14] showed that we can add more factors to our models and make it dependent on N correlated Brownian motions such given by :

$$d\xi_t^T = \xi_t^T \sum_i \omega_i \lambda_{it}^T(\xi_t^T) dW_t^{\mathbb{P},i} \quad (6.3)$$

- $\lambda_{it}^T(\xi_t^T) = \frac{e^{-k_i(T-t)}}{\sqrt{\sum_{ij} Cov_{i,j}}}$ being equal to a normalization term.

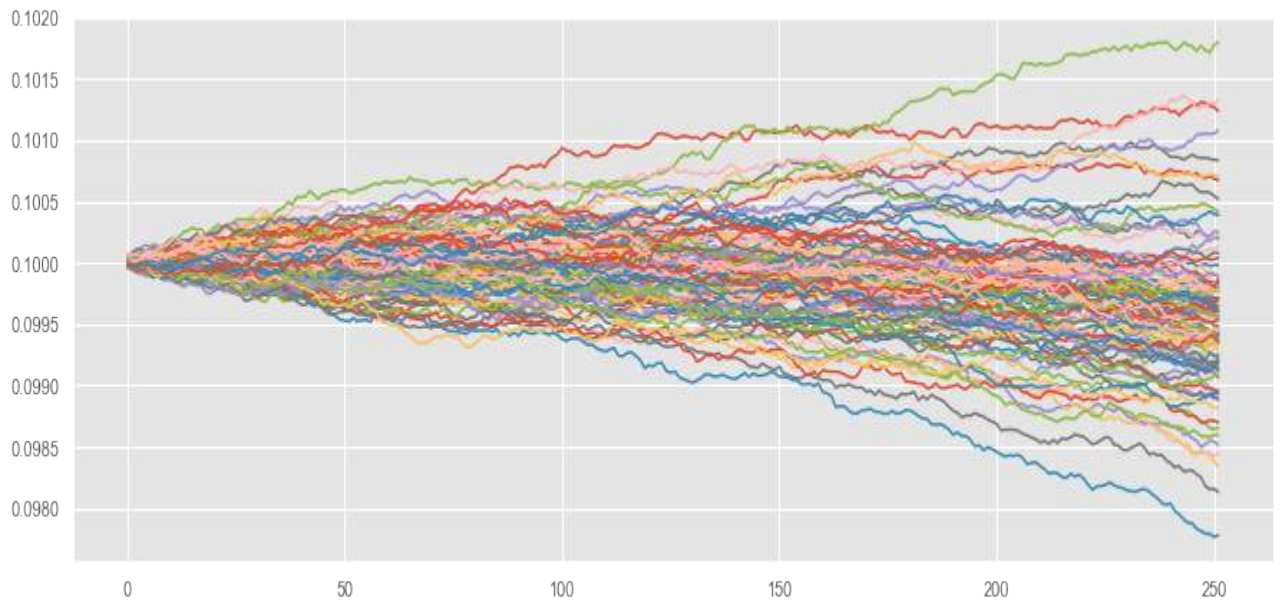


Figure 6.1: Monte Carlo simulations of 100 forward curves under Bergomi's model with parameters $\kappa = 1$, $T = 2$, $dt = \frac{1}{252}$, $\xi_0^T = 0.10$ and $w = 0.01$.

- ω_i being the log-normal variance of volatility factor noted in terms of the volatility of the volatility using the equivalence if a mono-factorial such : $\omega_i = 2v_i$, v_i being the log-normal volatility of the volatility for the i th Brownian path.
- $W_t^{\mathbb{P},i}$ is the i th correlated Brownian motion

In the case of an N factor model in 6.4, with correlation of 1, the forward rate can be approximated by :

$$\xi_t^T \approx \xi_0^T \left(1 + \omega \int_0^t \sum_i \left(\lambda_{is}^T \right) \left(\xi_0^T \right) dW_s^{\mathbb{P},i} \right) \quad (6.4)$$

Proof of 6.4. We begin by taking our SDE equation 6.3 and integrate such :

$$\xi_t^T = \xi_0^T + \omega \int_0^t \xi_s^T \sum_i \lambda_{is}^T(\xi_s^T) dW_s^{\mathbb{P},i}$$

We then plug the RHS into the original 6.3 SDE ,leading to:

$$d\xi_t^T = \omega \left(\xi_0^T + \omega \int_0^t \xi_s^T \sum_i \lambda_{is}^T(\xi_s^T) dW_s^{\mathbb{P},i} \right) \sum_i \lambda_{it}^T \left(\xi_0^T + \omega \int_0^t \xi_s^T \sum_k \lambda_{ks}^T(\xi_s^T) dW_s^{\mathbb{P},i} \right) dW_t^{\mathbb{P},i}$$

We expand at first order in ω :

$$d\xi_t^T \approx \xi_0^T \sum_i \lambda_{it}^T \left(\xi_0^T \right) dW_s^{\mathbb{P},i}$$

Which we reintegrate to retrieve 6.4 :

$$\begin{aligned} \xi_t^T &\approx \xi_0^T + \int_0^t \xi_0^T \sum_i \lambda_{is}^T \left(\xi_0^T \right) dW_s^{\mathbb{P},i} \\ &\approx \xi_0^T \left(1 + \int_0^t \sum_i \lambda_{is}^T \left(\xi_0^T \right) dW_s^{\mathbb{P},i} \right) \end{aligned}$$

We finally have a tractable formula that can be computed using a Riemann Approximation for small enough increments. As a result of that, we can translate the Bergomi's mathematical problem model into those two SDEs :

$$\begin{aligned} dF_t &= \sqrt{\xi_t^T} S_t dW_t^{\mathbb{Q},1} \\ d\xi_t^T &= \xi_t^T \sum_{i=2} \lambda_{it}^T \left(\xi_t^T \right) dW_t^{\mathbb{Q},i} \quad \text{With } \omega, \lambda_{it}^T > 0, \forall (T-t) \in \mathbb{R}_+ \end{aligned}$$

With W_t a Standard Brownian Motion.

Chapter 7

The Fractional Brownian Motions

Throughout this chapter, we will study processes defined in $(\Omega, \mathcal{F}_t, \mathbb{P})$, for $t \in \mathbb{R}_+$, a probability space where \mathcal{F}_t is the natural filtration of a standard Brownian motion.

7.1 Definitions

Definition 7 (Non independence). *Fractional Brownian motion is neither Markovian nor Martingale, its covariance matrix is time dependant and is written such :*

$$\text{Cov} \left(W_t^H, W_s^H \right) = \frac{1}{2} \left(t^{2H} + s^{2H} - |t - s|^{2H} \right), t, s \in \mathbb{R} \quad (7.1)$$

With $H \in (0, 1)$, being the Hurst parameter, a parameter that determines the level of time memory in the fractional Brownian Motion W_t^H .

Definition 8 (Gaussian process). *A fractional Brownian motion is a Gaussian process without independant increments, for finite increments at determined $(s, t) \in (M_s \cup J_t)$ noted as :*

$$|W_s - W_t| \sim N \left(0, \text{Cov} \left(W_s^H, W_t^H \right) \right) \quad (7.2)$$

With $W_{t_1}, W_{t_2}, \dots, W_{t_n}$ multivariate Gaussian random variables with $\min(t_1, t_2, \dots, t_n) \geq 0$

Definition 9 (Self Similarity Principle). *Fractional Brownian motions exhibit self-similarity such that in compact Borel space B self-similarity exists if and only if it exists a finite set S , where $S = t_1, t_2, \dots, t_n$ indexing a set of non-surjective homeomorphisms functions, which has the property of being continuously reversible such $\{f_g : g \in S\}$.*

Proposition 4. *Let's define a self-similar fractional Brownian motion W_t^H such that for any $h > 0$ and any $t \in \mathbb{R}$:*

$$\left\{ W^H(t_0 + \tau) - W^H(t_0) \right\} \triangleq \left\{ h^{-H} \left[W^H(t_0 + h\tau) - W^H(t_0) \right] \right\}$$

Definition 10 (Stationarity).

$$W_H(t_1) - W_H(t_2) \sim W_H(t_1 - t_2)$$

With $\mathbb{E}(W_H(t_1 - t_2)) = \mathbb{E}(W_H(t_n - t_{n-1})) = 0$ and $\mathbb{E}((W_H(t_1 - t_2))^2) = \text{Cov}(W_{t_n}^H, W_{t_{n-1}}^H)$ for an evenly spaced set S .

Definition 11 (Regularity). W_t^H is almost nowhere differentiable. However, all its paths are locally Hölder continuous. It means that for any $H > 0$, $t < T$, $s > 0$ and $\epsilon \rightarrow 0$, there exists a constant c , for which:

$$\left| W_t^H - W_s^H \right| \leq c |t - s|^{H-\epsilon}$$

Remark 4. The Koch curve 7.1, is regular, as if we scale the figure by $\frac{1}{N}$, its mass is scaled by $(\frac{1}{N})^d$, leaving us with the expression, for $N \in \mathbb{R}_+$:

$$M = cm^{1.262} \leq km^{1.262}$$

For c , a constant with $c \leq k$, $m = \frac{1}{N} > 0$, the scaling factor, and $d = 1.262$, the Minkowski–Bouligand dimension of the Koch fractal.

Remark 5 (Hölder Continuity). A function $h : R \rightarrow R$ is Hölder continuous on $\mathbb{I} \subset \mathbb{R}$ with Hölder exponent $\eta \in (0, 1)$ if and only if:

$$\limsup_{r \rightarrow 0} \frac{\sup_{\substack{s, t \in I \\ |s - t| < r}} |f(t) - f(s)|}{r^\eta} < \infty$$

An intuition can be given by looking to the rough figure 7.1. As the images' differences become very small. Then, the function is smooth on the interval $|t - s| < r, \forall r \rightarrow 0_+$.

Definition 12 (Differential Fractional Equation). Mandelbrot and Van Ness [15] discovered that we can use add memory to Brownian motion using interdependence with time. The following representation of a fractional Brownian motion will be used in the remaining of our paper:

$$W_H(t) = C_H \left(\int_{-\infty}^0 [(t - s)^{H-1/2} - (-s)^{H-1/2}] dW(s) + \int_0^t (t - s)^{H-1/2} dW(s) \right) \quad \forall t \in \mathbb{R}_+ \quad (7.3)$$

With :

$$\mathbb{E} \left[(W_H(t))^2 \right] = \mathbb{E} \left[(W_H(t) - W_H(0))^2 \right] = t^{2H} \quad (7.4)$$

We can guess from where does the " $T^{\frac{1}{2}}$ law" of the Standard Brownian motion comes.

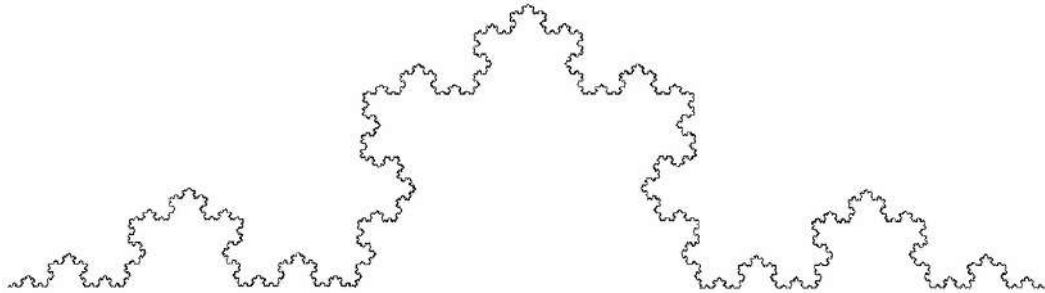


Figure 7.1: Koch curve is a self-similar and expresses regularity.

Proof of 7.4.

$$\begin{aligned}\mathbb{E}((W_{t+}^H - W_{t-}^H)^2) &= \mathbb{E}\left((C_H\left(\int_{-\infty}^0 \left\{(dt-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right\} dW_s\right) \right. \\ &\quad \left. + \int_0^{dt} \left\{(dt-s)^{H-\frac{1}{2}}\right\} dW_s\right)^2\end{aligned}$$

Through Ito's isometry we have :

$$\begin{aligned}\Leftrightarrow \mathbb{E}((W_{t+}^H - W_{t-}^H)^2) &= C_H^2 \left(\int_{-\infty}^0 \left\{(dt-s)^{H-\frac{1}{2}/2} - (-s)^{H-\frac{1}{2}}\right\}^2 ds \right. \\ &\quad \left. + \int_0^{dt} \left\{(dt-s)^{2H-1}\right\} ds \right) \\ &= C_H^2 \left(\int_{-\infty}^0 \left\{ dt^{H-\frac{1}{2}} \left[\left(1 - \frac{s}{dt}\right)^{H-\frac{1}{2}} - dt^{H-\frac{1}{2}} \left(-\frac{s}{dt}\right)^{H-\frac{1}{2}} \right] \right\}^2 ds + \int_0^{dt} \left\{ dt^{2H-1} \left(1 - \frac{s}{dt}\right)^{2H-1} \right\} ds \right) \\ &= dt^{2H-1} C_H^2 \left(\int_{-\infty}^0 \left\{ \left(1 - \frac{s}{dt}\right)^{H-\frac{1}{2}/2} - \left(-\frac{s}{dt}\right)^{H-\frac{1}{2}} \right\}^2 ds + \int_0^{dt} \left\{ \left(1 - \frac{s}{dt}\right)^{2H-1} \right\} ds \right)\end{aligned}$$

We proceed with a variable change such : $u = \frac{s}{dt}$

$$= dt^{2H} C_H^2 \left(\int_{-\infty}^0 \left\{ (1-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \right\}^2 du + \frac{1}{2H} \right)$$

and finally get :

$$\mathbb{E}((W_{t+}^H - W_{t-}^H)^2) = dt^{2H} \tag{7.5}$$

That is not what we were originally looking, as we want the variance $\mathbb{E}((W_{t+}^H - W_{t-}^H)^2)$ to be exclusively dependant on dt^{2H} . For this reason, We rewrite :

$$\begin{aligned}\mathbb{E}((W_{t+}^H - W_{t-}^H)^2) &= \mathbb{E}\left[C_H^2 \left(\int_{-\infty}^0 \left\{(dt-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right\} dW_s \right. \right. \\ &\quad \left. \left. + \int_0^{t+} \left\{(dt-s)^{H-\frac{1}{2}}\right\} dW_s \right)^2\right]\end{aligned}$$

Where C_H is equal to :

$$\begin{aligned}C_H &= \left(\int_{-\infty}^0 \left\{(dt-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right\} dW_s \right. \\ &\quad \left. + \int_0^{t+} \left\{(dt-s)^{H-\frac{1}{2}}\right\} dW_s \right)^{-\frac{1}{2}}\end{aligned}$$

As a result the covariance matrix is proven to be equal to :

$$\begin{aligned}\mathbb{E} \left[W_{t+}^H W_{t-}^H \right] &= \frac{1}{2} \left(\mathbb{E} \left([W_{t+}^H]^2 \right) + \mathbb{E} \left([W_{t-}^H]^2 \right) - \mathbb{E} \left[|W_{t+}^H - W_{t-}^H|^2 \right] \right) \\ &= \frac{1}{2} \left(|t+|^{2H} + |t-|^{2H} - |dt|^{2H} \right)\end{aligned}$$

Remark 6. The original covariance matrix used by Paul Levis took the form of an $H - \frac{1}{2}$ fractional Integration of a Standard Brownian Motion ,represented such :

$$W_{(}^H dt) = I^{-(H-\frac{1}{2})} f(dt) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^{dt} (dt - s)^{H-\frac{1}{2}} f(s) ds$$

However,the repeated Cauchy integral displayed here,uses a local minimum for the integration such $a = 0$,putting too much weight on the origin and then becoming ill-suited for many applications.That's why we prefer to use the Weyl integral 7.3 .

7.2 Volterra Process

The definitions of the Rough processes are very attractive.However,so as to be computationally useful,it requires some transformations.A Volterra Serie is written as a convolution of a kernel $K(t - s)$ with dW_s .It translates into :

$$V(t) = \int_0^t K(t - s) dW_s$$

With $K(t - s)$ a smooth function for which, $K : \mathbb{R} \rightarrow \mathbb{R}_+$.In this context,references found that for $H < \frac{1}{2}$, $K(t - s)$.Moreover,McCrickerd and Pakkanen [16] showed that the Volterra process can be efficiently expressed as :

$$\begin{aligned} K(t - s) &= \sqrt{2H}(t - s)^{H-\frac{1}{2}} \\ \mathbb{E}(V(t)V(s)) &= 2Hs^{2H} \int_0^1 (1 - u)^{H-1/2} \left(\frac{t}{s} - u\right)^{H-1/2} du \\ &= s^{2H} G\left(\frac{t}{s}\right) \end{aligned}$$

for all $t > s > 0$ and $H \in (0, 1)$.With ${}_2F_1$, an hyper-geometric function. This derivation leads us to a simpler and efficient formula to derive fractional Brownian using Gaussian quadrature ${}_2F_1$ function written as :

$$\begin{aligned} \mathbb{E}(V(t_+)V(t_-)) &= \left(\int_0^{dt} (t - s)^{2H-1} \cdot {}_2F_1^2\left(\frac{1}{2} - H, H - \frac{1}{2}, H + \frac{1}{2}; 1 - \frac{t}{s}\right) ds \right. \\ &= \left[\sum_{k=0}^M w_k f_k \right]^2 \end{aligned}$$

$n = j$ and $j \in \mathbb{N}_+$, $M_i \in \mathbb{N}_+$,with M representing the number of points in the $2M_i - 1$ dimension or less polynomial in the function f_k .

7.3 Simulation of Fractional Brownian Motion

The first techniques coming to our mind so as to simulate correlated random paths are inevitably either the Cholesky decomposition or the Eigenvector decomposition.Let's firstly begin by explaining the former.

7.3.1 Fraction Brownian motion using Cholesky decomposition :

We need to create the positive covariance matrix 7.1 such : $\Gamma(\mathcal{S}, \frac{T}{dt} \cdot \frac{T}{dt})$ for equally spaced and finite set \mathcal{S} , as defined in 9. The matrix is written as :

$$\Gamma = \begin{pmatrix} t_1^{2H} & \dots & \frac{1}{2} (t_1^{2H} + t_n^{2H} - |t_1 - t_n|^{2H}) \\ & \ddots & \vdots \\ * & & t_n^{2H} \end{pmatrix} \quad (7.6)$$

As we said, Γ is an Hermitian matrix, therefore we can rewrite 7.6 using Cholesky Triangulation method :

$$\Gamma = A^T \cdot A$$

With A^T , the higher triangular and A the lower one for $A \in \mathbb{R}^{n \times n}$, where $n = \frac{T}{dt}$.

$$A = \begin{pmatrix} a_{1,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{pmatrix} \quad (7.7)$$

The last step to obtain our so desired FBM is to multiply our Lower triangular A 7.7 by i.i.d random vectors of the dimension $n \cdot L$:

$$I = \begin{pmatrix} \epsilon_{1,1} & \dots & \epsilon_{L,n} \\ \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \dots & \epsilon_{n,L} \end{pmatrix} \quad \forall \epsilon_{i,j} \in \{\epsilon_{1,1}, \dots, \epsilon_{n,L}\} \sim N(0, 1)$$

Remark 7. If Γ is positive definite. Then the lower triangular matrix A won't contain 0 in its filled part.

Finally, our L random i.i.d correlated random paths are expressed in the matrix of dimension $n \cdot L$ composed of L fractional Brownian vectors :

$$W_H^L(T, dt) = \begin{pmatrix} W_{H1}^1 & \dots & W_{H1}^L \\ \vdots & \ddots & \vdots \\ W_{Hn}^1 & \dots & W_{Hn}^L \end{pmatrix} \quad \forall T \in \mathbb{R}_+ \quad \text{and} \quad \forall H \in (0, 1)$$

7.3.2 FBM using spectral decomposition :

Let's express the $n \times n$ dimensional covariance matrix using spectral decomposition as we already know that it has linear dependency in its columns following its positiveness :

$$\Gamma = Q \Theta Q^{-1}$$

With Q , being the eigenvector of Gamma and Θ , the diagonal eigenvalues' matrix for $Q \in \mathbb{R}^{n \times n}$ and $\Theta \in \mathbb{R}_+^{n \times n}$. In order to find the volatility, we just have to add a power $\frac{1}{2}$ to our diagonal eigenvalues' matrix such :



Figure 7.2: Fractional Brownian motions with different Hurst parameters

$$\Gamma^{\frac{1}{2}} = Q\Theta^{\frac{1}{2}}$$

Resulting that we can generate a fractional Brownian motion by multiplying $\Gamma^{\frac{1}{2}}$ to a standard normally distributed matrix Z of dimensions $n \times L$, for L different discrete random paths, through a dot product to finally end up with :

$$W_T^{H,L} = Q\Theta^{\frac{1}{2}}Z$$

7.3.3 FBm using efficient spectral decomposition :

We can efficiently compute a fractional Brownian motion using the Fast Fourier algorithm. In order to do so we need to express our covariance matrix in a circulant

way such it satisfies the following shift-ward to the bottom property illustrated as :

$$C = \begin{pmatrix} c_0 & c_1 & \dots & c_{n-1} \\ c_{n-1} & c_0 & \dots & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \dots & c_0 \end{pmatrix} \quad (7.8)$$

Of course our matrix is not the same as what we need it to be in order to apply the Fast Fourier algorithm. A way to circumvent that is to begin to use a topalix matrix covariance matrix. With $\gamma(t, s)$, representing the covariance of the Gaussian noise that can be de-incremented to form our fractional Brownian motion, satisfies :

$$\gamma(k) = \frac{1}{2}(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) \quad \forall k \in \mathbb{R}$$

Proof. Let's begin to exemplify the covariance of our fractional Brownian's increments across time:

$$\begin{aligned} \mathbb{E}(\delta W_{t,s}^H \delta W_{u,m}^H) &= \frac{1}{2} \left(|t|^{2H} + |u|^{2H} - |t-u|^{2H} - |t|^{2H} - |m|^{2H} + |t-m|^{2H} - |s|^{2H} - |u|^{2H} \right. \\ &\quad \left. + |s-u|^{2H} + |s|^{2H} + |m|^{2H} - |s-m|^{2H} \right) \end{aligned}$$

Which simplifies to :

$$= \frac{1}{2} (-|t-u|^{2H} + |t-m|^{2H} + |s-u|^{2H} - |s-m|^{2H})$$

Let's introduce $s = t - 1$, $u = t + k$, $m = t + k + 1$

$$\begin{aligned} &= \frac{1}{2} (-|k|^{2H} + |k-1|^{2H} + |1+k|^{2H} - |k|^{2H}) \\ &= \frac{1}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) \end{aligned}$$

If we want to express the covariance matrix in terms of s and t in order to be able to visualize the matrix we could express it as :

$$\gamma(t, s) = \frac{1}{2} (|t-s-1|^{2H} - 2|t-s|^{2H} + |t-s+1|^{2H}) \quad \forall t, s \in \mathbb{R}$$

In our case as we are working in a defined amount of days per year. We want to standardize the time to be realistic, the expression of the covariance matrix then transforms to :

$$\gamma(k) = \frac{1}{2ADays^{2H}} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) \quad \forall ADays \in \mathbb{N}_+$$

However, we still have a Toeplitz matrix that is not in the form of the required 7.8 that looks like:

$$\Gamma = \begin{pmatrix} \rho(0) & \rho(1) & \dots & \rho(n-1) \\ \rho(1) & \rho(0) & \dots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \dots & \rho(0) \end{pmatrix} \quad (7.9)$$

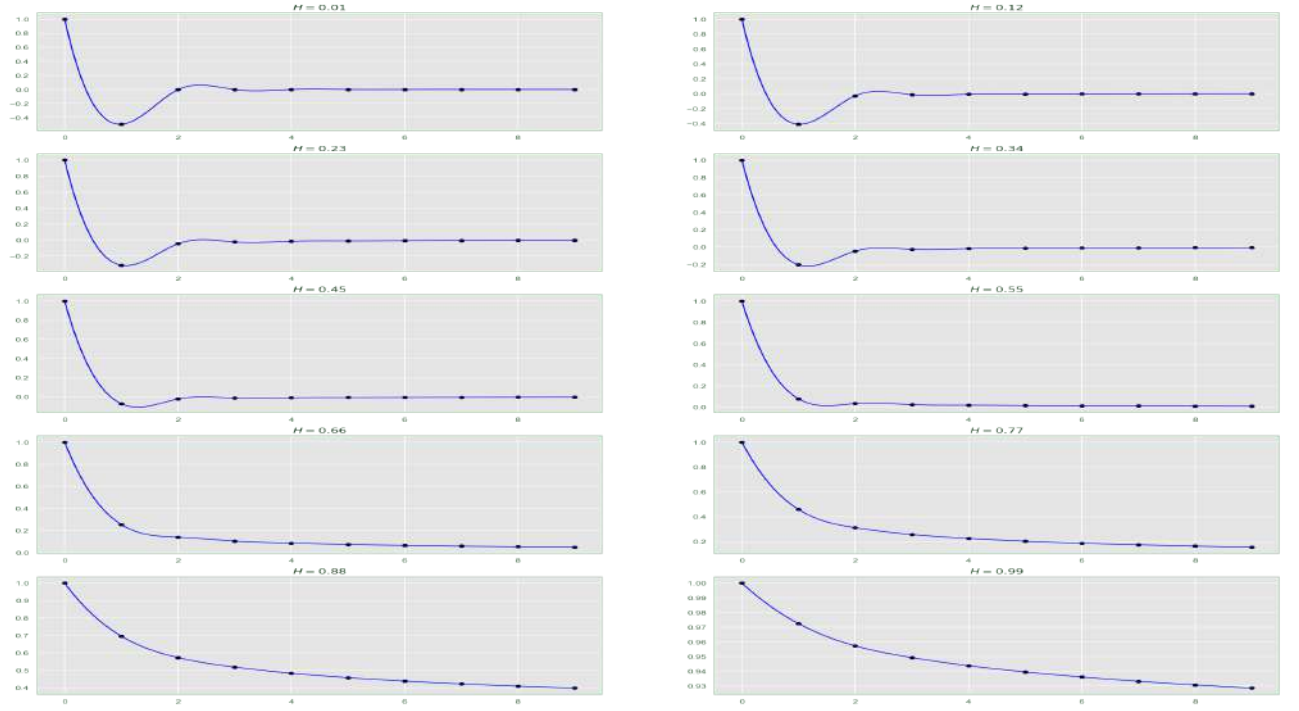


Figure 7.3: Covariance matrix with daily lags

Definition 13. A Toeplitz matrix is a matrix in which each of the descending diagonal from left to right is constant.

Nonetheless, it is not lost as a Circulant matrix is a particular case of a Toeplitz matrix and a fast transformation can easily fix the problem. First of all, we need to embed the covariance matrix 7.9 by a Circulant matrix M such:

$$M = \left(\begin{array}{cccc|cccc} \rho(0) & \rho(1) & \dots & \rho(n-1) & \dots & \rho(1) & & \\ \rho(1) & \rho(0) & \ddots & \rho(n-2) & \ddots & \rho(2) & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & & \\ \rho(n-1) & \rho(n-2) & \ddots & \rho(0) & \dots & \rho(n-1) & & \\ \hline \rho(n-2) & \rho(n-3) & \dots & \dots & \ddots & \rho(n-2) & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & & \\ \rho(1) & \rho(2) & \dots & \rho(n-2) & \dots & \rho(0) & & \end{array} \right)$$

Afterwards, our matrix becomes Circulant and we can now use the fast Fourier transform algorithm that allows us to reduce the complexity of our operation from $O(n^3)$ for the Cholesky decomposition and the spectral decomposition to $O(n \log(n))$, which is very nice when we work with huge amount of Monte Carlo simulations along with lengthy time frame.

The last steps to generate our Gaussian noises would be to multiply our orthonormal inverse $2(n-1) \times 2(n-1)$ DFT matrix, using a dot product, to the square root of our diagonal eigenvalues' matrix. Thereafter, we multiply the obtained matrix by a

$2(n-1) \times L$ matrix that contains our L standard normal samples. Lastly, we take the real part of the matrix and multiply it by $\sqrt{2}$ such as showed on the following equation :

$$\text{Re} \left(\sqrt{2(2n-2)} \times \text{ifft} \left(\sqrt{\text{fft}(M_{1,:})} \times Z \right) \right)$$

For usefulness the steps are listed such :

Algorithm 1 Generate Fractional Gaussian Noise

procedure GAUSSIAN NOISE GENERATOR(*self*)

Construct the $((2n-2) \times (2n-2))$ Circulant covariance matrix C .

Apply an FFT to the first column of C such $\sqrt{\text{fft}(C_{1,:})} = \Theta$.

Condition 1:

if All $\text{Diag}(\Theta) > 0$ **then return** continue :

end if

Construct a $((2n-2) \times L)$ Standard Normal matrix $Z \sim \mathcal{N}(0, I_{2n-2})$.

Do $\Theta \times Z$ using matrix multiplication, calling the matrix result F .

Do $\sqrt{(2n-2)} \text{ifft}(F)$ and multiply by $\sqrt{2}$, calling the result Q .

Take the real part of Q such $\text{Re}(Q)$.

end procedure

All of the steps were taken from the paper of Dietrich, CR and Newsam, Garry Neil [17]. However, they forgot on the paper to multiply the orthonormal inverse discrete Fourier transform matrix by $\sqrt{2}$, which we corrected.

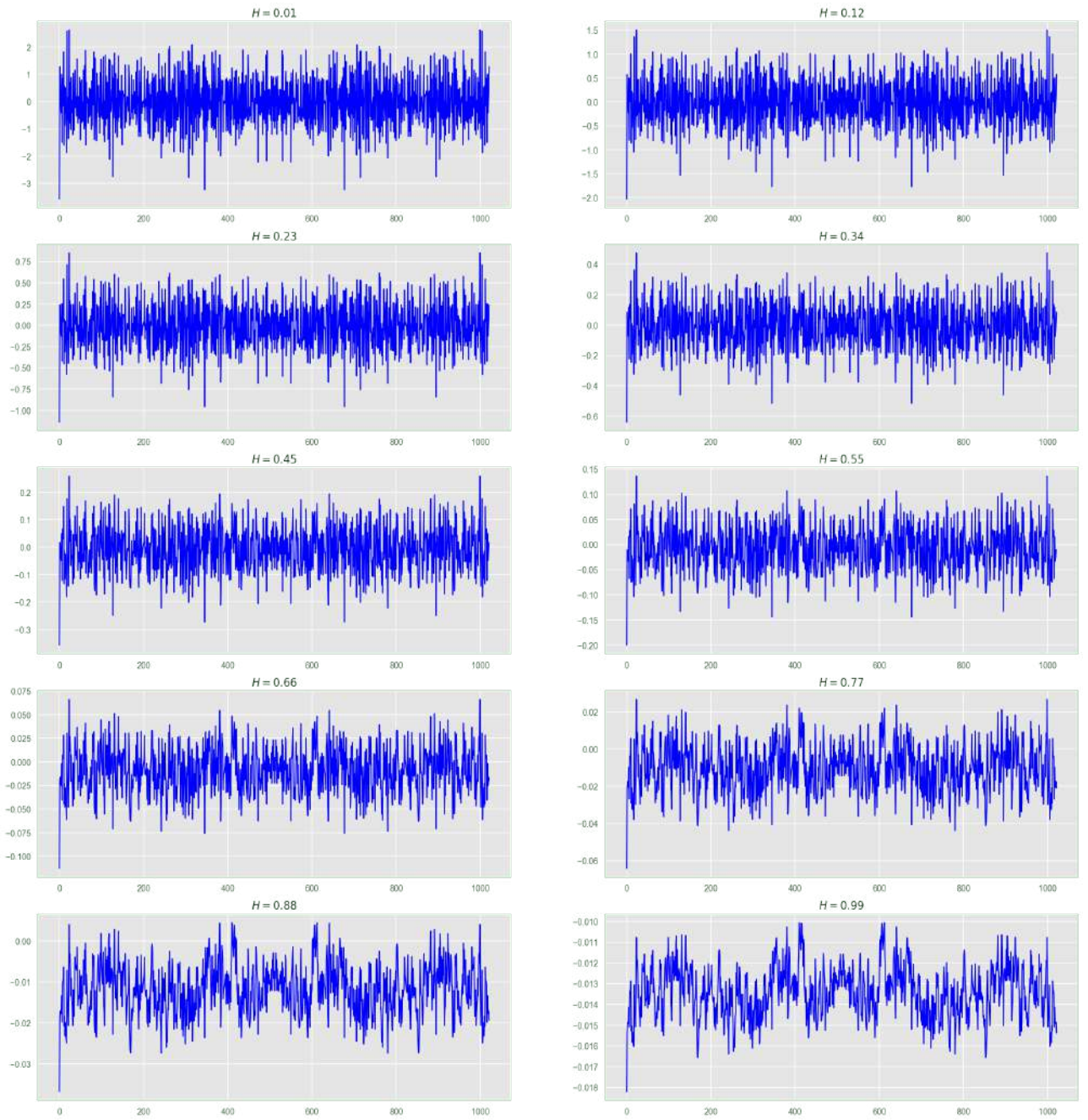


Figure 7.4: Generated fractional Gaussian noises

Chapter 8

The rough Bergomi model

8.1 The unconventional Bergomi model

Estimates of realized variance distributions in past studies were found Gaussian such showed on 8.5. Secondly, for reasonable timescales of practical interest (from one day to a few years), the time series of realized variance was found to be consistent with the simplest RFSV model [18], that we will cover later, such :

$$\log \sqrt{v_{t+\Delta}} - \log \sqrt{v_t} = \nu \left(W_{t+\Delta}^H - W_t^H \right)$$

- With $v_{t+\Delta}$, the variance of the asset.

Remark 8. As said before, on 6.1 if the long term reversal slope on the OU process $\kappa \rightarrow 0$ we could conclude that :

$$\mathbb{E} \left[\sup_{t \in [0, T]} \left| X_t - X_0 - \omega W_t^H \right| \right] \rightarrow 0$$

- With X_t a variable following an OU process with a mean reversal slope κ that tends to 0.

Gatheral, Bayer and Friz [19] found that this relationship held for all 21 equity indexes in the Oxford-Man database along with bond futures, crude oil futures, and gold futures. Now, considering the original expression of the Fractional Brownian motion 7.3 derived by Mandelbrot. under the physical measure \mathbb{P} :

$$\begin{aligned} \log v_u - \log v_t &= 2\nu \left(W_u^H - W_t^H \right) \\ &= 2\nu C_H \left(\int_{-\infty}^u |u-s|^{H-\frac{1}{2}} dW_s^{\mathbb{P},2} - \int_{-\infty}^t |t-s|^{H-\frac{1}{2}} dW_s^{\mathbb{P},2} \right) \\ &= 2\nu C_H \left(\int_t^u |u-s|^{H-\frac{1}{2}} dW_s^{\mathbb{P},2} + \int_{-\infty}^t \left[|u-s|^{H-\frac{1}{2}} - |t-s|^{H-\frac{1}{2}} \right] dW_s^{\mathbb{P},2} \right) \\ &=: 2\nu C_H \left[M_t(u) + Z_t(u) \right] \end{aligned}$$

In this expression, the left integral $M_t(u)$ is independent of the filtration \mathcal{F}_t because its time series is non-independent across the future time-frame. However, The right integral $Z_t(u)$ is \mathcal{F}_t measurable because of the sole dependence to the past data contrary to $M_t(u)$.

Remark 9. the expression C_H is equivalent to :

$$C_H = \sqrt{\frac{2H\Gamma(3/2 - H)}{\Gamma(H + 1/2)\Gamma(2 - 2H)}}$$

The Mandelbrot-Van Ness representation of fractional Brownian motion W_t^H in terms of Wiener integrals is :

$$W_t^H = C_H \left\{ \int_{-\infty}^t \frac{dW_s^{\mathbb{P},2}}{(t-s)^\gamma} - \int_{-\infty}^0 \frac{dW_s^{\mathbb{P},2}}{(-s)^\gamma} \right\}$$

where $\gamma = \frac{1}{2} - H$

We then define the semi-stationary fractional Brownian motion derived such :

$$\tilde{W}_t^{\mathbb{P}}(u) := \sqrt{2H} \int_t^u \frac{dW_s^{\mathbb{P},2}}{(u-s)^\gamma}$$

Remark 10. We can say that $\sqrt{2H}M_t(u)$ is a fractional Brownian motion according to its "weak" definition developed by Lévy in 1953 [20] such :

$$W_t^H = \sqrt{2H} \int_t^u \frac{dW_s^{\mathbb{P},2}}{(u-s)^\gamma}$$

It is therefore straightforward to see that $\sqrt{2H}M_t(u)$ is a Gaussian Noise with 0 mean and variance $(t-u)^{2H}$. Setting the parameter $\eta = \frac{2\nu C_H}{\sqrt{2H}}$, we can rewrite the formula of the conditional expected variance such as :

$$\mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] = v_t \exp \left\{ \eta Z_t(u) + \frac{1}{2} \eta^2 (u-t)^{2H} \right\}$$

As a consequence, because $v_u | \mathcal{F}_t$ is log normal, we have :

$$v_u = v_t \exp \left\{ \eta \tilde{W}_t^{\mathbb{P}}(u) + \eta Z_t(u) \right\}$$

Expressing that in terms of Wick exponential integral gives :

$$v_u = \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{P}}(u) \right)$$

Definition 14. For a continuous martingale or semi martingale Z , a classical attribution of Taylor series using log functions such $f(x) = \log x$. If Z is a stochastic process depending on a Brownian Motion, therefore we will have $\mathcal{E}(Z)_t = \exp(\text{drift}_t \mathbb{1}_{\setminus \mathcal{M}} + Z_t - Z_0 - \frac{1}{2}[Z, Z]_{0,t})$, where \mathcal{M} stands for Martingale. In general, on the field of option pricing the best-known expression is using Ito's isometry for its great property such for a random variable Ψ , we have:

$$\mathcal{E}(\Psi) = \exp \left(\Psi - \frac{1}{2} \mathbb{E} [|\Psi|^2] \right)$$

8.2 Variance forecast

The key formula on which our prediction method to forecast log-volatility, is explained by Gatheral, Jaisson and Rosenbaum [18]

$$\mathbb{E} \left[W_{t+\Delta}^H \mid \mathcal{F}_t \right] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^t \frac{W_s^H}{(t-s+\Delta)(t-s)^{H+1/2}} ds$$

where W^H is a fBM with $H < 1/2$ and \mathcal{F}_t our the usual filtration generated. By construction, over any reasonable and equally spaced time scale we may approximate the fractional Ornstein-Uhlenbeck volatility process in the RFSV model as

$$\log \sigma_t^2 \approx 2\nu W_t^H + C \quad (8.1)$$

For some constants ν and C . The prediction formula for log-variance given by the authors are then :

$$\mathbb{E} [\log \sigma_{t+\Delta}^2 \mid \mathcal{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^t \frac{\log \sigma_s^2}{(t-s+\Delta)(t-s)^{H+1/2}} ds \quad (8.2)$$

The Riemann sum approximation of the 8.2 formula ,as the volatility is assumed easily observable in the market(because our data are cleaned), is used to forecast the log-volatility for the periods of 1,5 and 20 days ahead ($\Delta = 1, 5, 20$). Recall the relation 8.1, starting from that point we can, knowing that our original expression of a fractional Brownian motion, formulate the variance forecast such :

$$\text{Var} \left[W_{t+\Delta}^H \mid \mathcal{F}_t \right] = C_H \Delta^{2H}$$

With C_H , as expressed before. Therefore, we can easily forecast the variance $\sigma_t^2 \mid \mathcal{F}_t$ for a determined lags such :

$$\mathbb{E} [\sigma_{t+\Delta}^2 \mid \mathcal{F}_t] = \exp \left(\mathbb{E} [\log \sigma_{t+\Delta}^2 \mid \mathcal{F}_t] + 2C_H \nu^2 \Delta^{2H} \right)$$

As a result of that our discretized Riemann Sum is straightforward to derive :

$$\begin{aligned} \int_0^t \frac{\log \sigma_s^2}{(t-s+\Delta)(t-s)^{H+1/2}} ds &= \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \frac{\log \sigma_s^2}{(t-s+\Delta)(t-s)^{H+1/2}} ds \\ &\simeq \sum_{i=1}^N \frac{\log \sigma_{t_{i-1}}^2}{(t-t_{i-1}+\Delta)(t-t_{i-1})^{H+1/2}} (t_i - t_{i-1}) \\ &= \delta \sum_{i=1}^N \frac{\log \sigma_{t_{N-i}}^2}{(t_i + \Delta) t_i^{H+1/2}} \end{aligned}$$

We can see that if we choose $\Delta = \frac{1}{252}$, therefore our approximation transforms to :

$$\int_0^t \frac{\log \sigma_s^2}{(t-s+\Delta)(t-s)^{H+1/2}} ds = \delta \sum_{i=1}^N \frac{\log \sigma_{t_{N-i}}^2}{(t_i + \Delta) t_i^{H+1/2}}$$

We compared different fit of volatility forecasts given some values ν . We found that for the S&P500, the best fit is given for values of ν close to 0.444.

ν	0	0.11	0.22	0.33	0.44	0.55	0.66	0.77	0.88	1
$ \sigma_{realized}^2 - \sigma_{forecasted}^2 $	8.15	8.09	7.93	7.74	7.67	8.05	9.54	13.57	23.13	44.76

Table 8.1: Comparison of norm-errors between the realized and the forecasted variance for a defined parameter ν

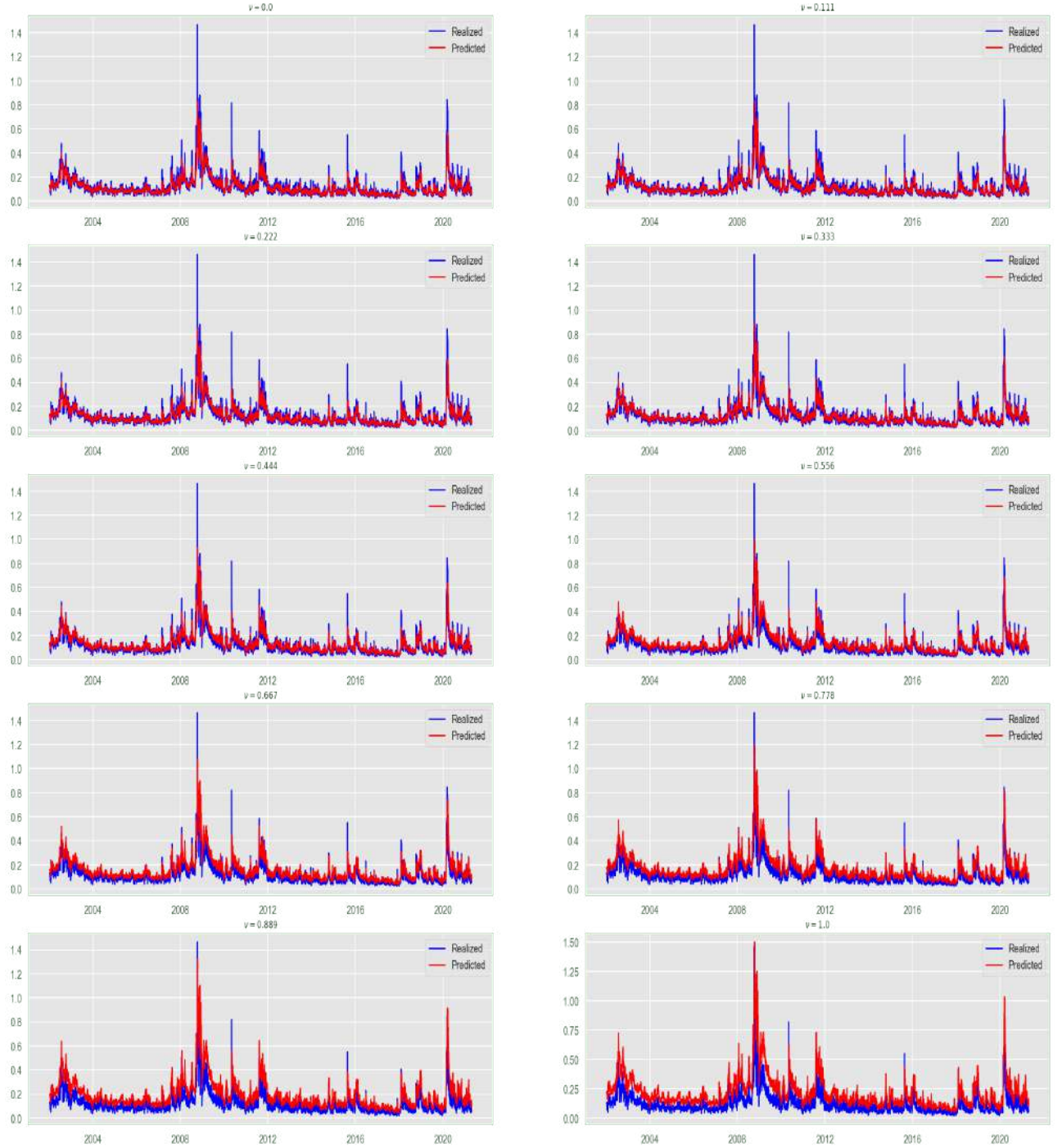


Figure 8.1: Forecast and realized volatility for $\Delta = \frac{1}{252}$ given different parameters ν

8.3 FSV or RFSV under the \mathbb{P} measure

The first expression of the rough variance was expressed by Comte and Renault [21]. Their main idea was to replace the standard Brownian motion expressed on the SDEs by a mean reversible one. This mean reversible Brownian motion will later be called fractional Brownian Motion. Let's call a probability space $(\Omega, \mathcal{F}_t, \mathbb{P}) \forall t \in \mathbb{R}_+$. In Comte and Renault fractional stochastic volatility (FSV) model :

$$\begin{aligned} \frac{dS_t}{S_t} &= (r - q)dt + \sigma_t dW_t^{\mathbb{P},1} \\ d \log \sigma_t &= \kappa(\log \sigma_t - \theta)dt + \nu dW_t^{H,\mathbb{P}} \end{aligned}$$

These SDEs are easily solved as we recognize the second equation as an OU process. The two solutions are given by :

$$\begin{aligned} S_T &= S_0 e^{(r-q-\frac{1}{2}\sigma_T^2)T + \sigma_T W_T^{\mathbb{P},1}} \\ \sigma_T &= e^{\theta + e^{-\kappa T}(\log \sigma_0 - \theta) + \nu \int_0^T e^{-\kappa(T-s)} dW_s^{H,\mathbb{P}}} \end{aligned}$$

The FSV has the specificity to require that the fraction Brownian Motion W_t^H exhibits long term memory such as its Hurst parameter $H > 0.5$ in order for the variance σ_t^2 to be mean reversible. A question that we would ask is why the authors used Fractional Brownian motions with long term memory. The answer resides from the fact that in the past, it has been widely-accepted stylized fact that the volatility time series exhibits long memory, never heard the myth 'Periods of high volatility follow up'.

Later on, Gatheral, Jaisson and Rosenbaum [18] showed that the long memory models such FSV models don't hold with empirical data because of the non linearity of their log moments across their first log moments compared to the original data. Moreover, they showed that the RFSV hold with empirical data such as shown on the figure 8.2. An RFSV is just an FSV that exhibits short memory on its fractional Brownian motion and with a mean reversion speed of $\kappa = 0$. It writes down as :

$$\sigma_T = \sigma_0 e^{\nu W_T^H} \quad \text{For } W_0^H = 0$$

8.4 Rough Bergomi model under the \mathbb{P} measure

The rough Bergomi model is defined as :

$$\begin{aligned} \frac{dS_u}{S_u} &= (r - q)du + \sqrt{v_u} dW_u^{\mathbb{P},1} \\ \tilde{W}_u^H &= \sqrt{2H} \int_0^u (u - s)^{H-\frac{1}{2}} dW_s^{\mathbb{P},2} \\ v_u &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{P}}(u) \right) \\ \mathbb{E} \left(dW_u^{\mathbb{P},1} dW_u^{\mathbb{P},2} \right) &= \rho du \end{aligned}$$

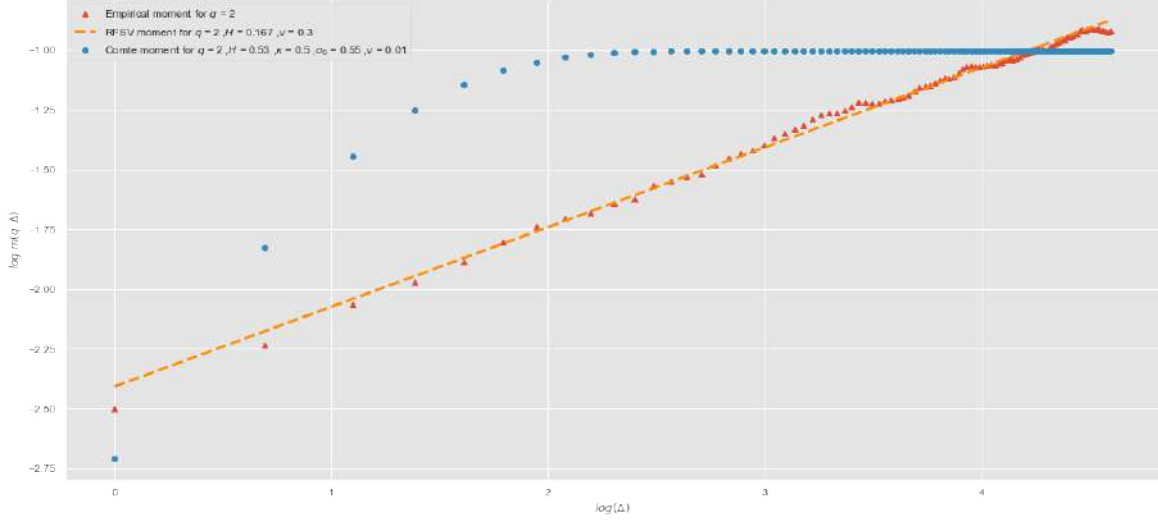


Figure 8.2: A log-moment comparison between RFSV and FSV

Knowing that the log volatility is usually normal(backed with 8.5) we could easily find a way to solve the SDE of both the asset price and the volatility such that our final result is given by :

$$\begin{aligned} S_u &= S_0 e^{\sum_{i=1}^{u \times \frac{1}{\Delta u}} (r - q - \frac{v_i \times \Delta u}{2}) du + \sqrt{v_i \times \Delta u} dZ_{i \times \Delta u}} \\ v_u &= \mathbb{E}(v_u | \mathcal{F}_{t \geq 0}) e^{\eta \tilde{W}_t^{\mathbb{P}}(u) - \frac{\eta^2}{2} u^{2H}} \end{aligned} \quad (8.3)$$

We could derive 8.3 because we know that $\tilde{W}_t^{\mathbb{P}} \sim N(0, u^{2H})$ and therefore using Ito's Isometry :

$$\begin{aligned} Var(\tilde{W}_t^{\mathbb{P}}(u)) &= \mathbb{E}\left((\tilde{W}_t^{\mathbb{P}}(u))^2\right) - E\left(\tilde{W}_t^{\mathbb{P}}(u)\right)^2 \\ &= E\left((\tilde{W}_t^{\mathbb{P}}(u))^2\right) - 0 \\ &\equiv E\left((\tilde{W}_t^{\mathbb{P}}(u))^2\right) = Var(\tilde{W}_t^{\mathbb{P}}(u)) = u^{2H} \end{aligned}$$

A huge drawback of the rough Bergomi's model is the non existence of tractable closed formula because of the main reason that the fractional Brownian motion is neither a Markovian ,nor Martingale or semi-Martingale process :

$$\mathbb{E}(v_u | \mathcal{F}_{t \geq 0}) \neq \mathbb{E}(v_u | v_t)$$

8.5 Rough Bergomi under the \mathbb{Q} measure

Under the physical probability measure \mathbb{P} , the SDEs are given such :

$$\begin{aligned} dS_u &= \mu S_u du + \sqrt{v_u} S_u dW_u^{\mathbb{P},1} \\ v_u &= v_t \exp\left(\eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u)\right) \end{aligned}$$

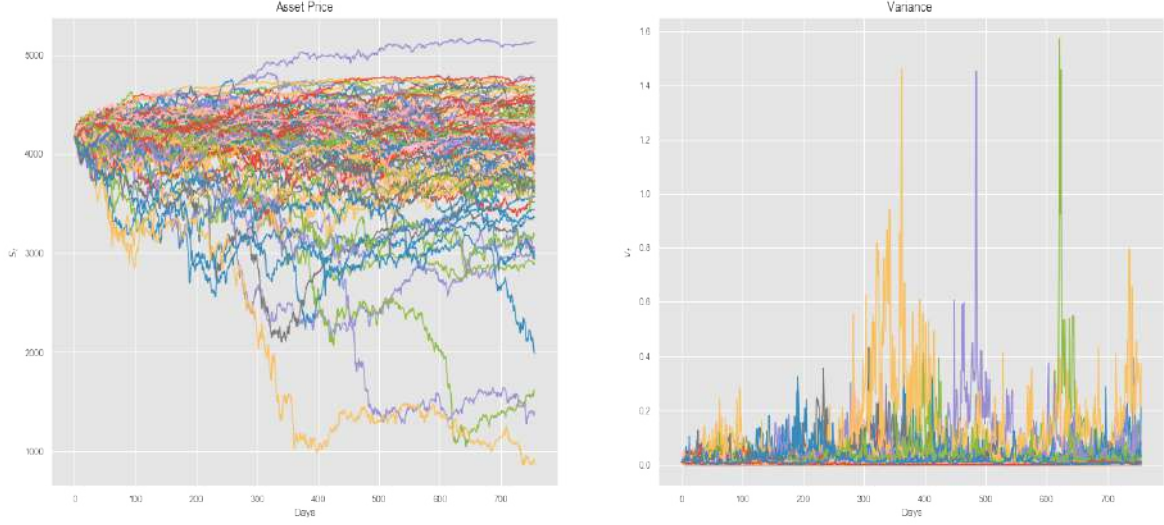


Figure 8.3: Monte Carlo simulation of the rough Bergomi model under the \mathbb{P} measure

with $\mathbb{E}(W_u^{\mathbb{Q},1}, W_u^{\mathbb{Q},2}) = \rho du$. As mentioned for the \mathbb{P} measure, the pricing of options is made under an equivalent martingale measure $\mathbb{Q} \sim \mathbb{P}$ on $[t, T]$:

$$W_u^{\mathbb{Q},1} = \frac{\mu - r}{\sqrt{v_u}} dt + dW_u^{\mathbb{P},1}$$

Obtained by a Girsanov change of measure on $[t, T]$ from \mathbb{P} to \mathbb{Q} . In addition, the Brownian motion $W_u^{\mathbb{P},1}$, which is used to construct the Volterra-type process $\tilde{W}_t^{\mathbb{P}}(u)$, the correlation factor of $W_u^{\mathbb{P},1}$ and $W_u^{\mathbb{P},2}$ is equal to ρ , which has been shown to be negative empirically for the equity. The standard Brownian motion used in the Volterra-type process is therefore equal to:

$$dW_u^{\mathbb{P},2} = \rho dW_u^{\mathbb{P},1} + \sqrt{1 - \rho^2} dZ_u^{\mathbb{P}}$$

where $W_u^{\mathbb{P},1}$ is independent of $Z_u^{\mathbb{P}}$. Thereafter, we first apply an initial change of measure for $Z_u^{\mathbb{P}}$ in order for us to be neutral against the market price of the volatility risk. The risk neutral Brownian motion $Z_u^{\mathbb{Q}}$ would then transform to the form:

$$dZ_u^{\mathbb{Q}} = dZ_u^{\mathbb{P}} + \gamma_u du$$

where γ is a suitable process on $[t, T]$ seen as the market price of volatility risk. Combining the last steps $dW_u^{\mathbb{Q},2}$ can be expressed as:

$$\begin{aligned} dW_u^{\mathbb{Q},2} &= \rho dW_u^{\mathbb{Q},1} + \sqrt{1 - \rho^2} dZ_u^{\mathbb{Q}} \\ &= dW_u^{\mathbb{P},2} + \left(\rho \frac{\mu - r}{\sqrt{v_u}} + \sqrt{1 - \rho^2} \gamma_u \right) du \end{aligned}$$

The change of measure from \mathbb{P} to \mathbb{Q} can therefore be noted as:

$$dW_u^{\mathbb{Q},2} = dW_t^{\mathbb{P},2} - \lambda_u du$$

Under the physical measure \mathbb{P} , the variance process v simulates the realized variance from market data. The variance process can be expressed via the pricing measure \mathbb{Q} so that, assuming

$$\begin{aligned} v_u &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{P}}(t) \right) \exp \left(\eta \sqrt{2H} \int_t^u (u-s)^{H-\frac{1}{2}} \lambda_s ds \right) \\ &= \xi_t^u \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{Q}}(u) \right) \end{aligned}$$

with the risk neutral forward variance curve expressed as :

$$\begin{aligned} \xi_t^u &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \exp \left(\eta \sqrt{2H} \int_t^u (u-s)^{H-\frac{1}{2}} \lambda_s ds \right) \\ &= \mathbb{E}^{\mathbb{Q}} [v_u | \mathcal{F}_t] \end{aligned}$$

However, practically it is not possible to evaluate λ without altering the log-normal property of the variance. As a result, for the same reason we are using for the stochastic volatility jump models concerning the volatility risk premium, we should just set the parameter $\lambda_s = 0$, and move forward. The Rough Bergomi model is non-Markovian in the instantaneous variance v_u and the the risk-neutral SDEs in \mathbb{Q} are given by :

$$\begin{aligned} dF_u &= \sqrt{v_u} S_u dW_u^{\mathbb{Q},1} \\ v_u &= \xi_t^u \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{Q}}(u) \right) \end{aligned}$$

8.6 Volatility skew

A huge deficiency of the stochastic volatility models is that the dynamic of the ATM volatility skew across time given by the simple formula :

$$\psi(\tau) := \left| \frac{\partial}{\partial x} \sigma_{IV}(\tau, x) \right|_{x=0} \quad (8.4)$$

With x being the log moneyness of the option and τ the time to maturity. Volatility skews across maturities are very sensitive to the choice of volatility dynamics in a stochastic volatility model and for the simplest models such showed on the figure 8.4, the volatility skew stays flat even for very short maturities.

Adding jumps into stochastic volatility models can improve the volatility skew across maturities but requires the understanding of involved models along with extremely heavy calibrations. Therefore, we can see that in general the stochastic volatility models are inconsistent to form an accurate at-the-money term structure skew across maturities and especially for very short term maturities.

However, empirical studies find that the ATM term structure skew can be well approximated by the power law $\frac{1}{\tau^\alpha} \quad \forall \alpha \in (0.3, 0.5)$, as showed by Fukasawa on [22], and $A \in (0, 1)$ generally such $\psi(\tau) \sim A\tau^{-\alpha}$. With α such :

$$\alpha = H - \frac{1}{2}$$

In order to compute 8.4 we must adjust our formula to find the implied volatility in the market given a strike K and a time to maturity τ . In order to do so, we use the equivalence of 8.4 using the chain rule along with the change of variable $x = \log \frac{K}{S_t}$:

$$\begin{aligned} \frac{\partial \sigma_{IV}^{mkt}(\tau, S_t, K)}{\partial K} &= \frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \frac{\partial x}{\partial K} \\ &= \frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \frac{1}{K} \end{aligned}$$

Yielding to the formula in terms of strikes K as :

$$\frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \Big|_{x=0} = K \frac{\partial \sigma_{IV}^{mkt}(\tau, S_t, K)}{\partial K} \Big|_{K=S_t}$$

This expression is very nice but is hardly implementable. A solution we propose for this is to replace the strikes by forwards such as we obtain :

$$\frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \Big|_{x=0} = F_t \frac{\partial \sigma_{IV}^{mkt}(\tau, S_t, K)}{\partial K} \Big|_{K=F_t}$$

In order to have a good approximation we might have $F_t = S_t e^{(r-q)\tau}$ with $(r-q)\tau$ sufficiently small ($(r-q)\tau \rightarrow 10^{-3}$ would allow a good approximation of the derivative) such :

$$\frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \Big|_{x=0} = \frac{F_t^{+'} \sigma_{IV}^{mkt}(F_t^{+'}, \tau) - F_t^{-'} \sigma_{IV}^{mkt}(F_t^{-'}, \tau)}{2(r-q)\tau}$$

For sufficiently small $(r-q)\tau$ we can exemplify the skew using the approximation that $F_t^{+'} \approx F_t^{-'}$ because we know that the implied volatility is very sensible to drift:

$$\frac{1}{F_t} \frac{\partial \sigma_{IV}(\tau, x)}{\partial x} \Big|_{x=0} \approx \frac{\sigma_{IV}^{mkt}(\tau, S_t, F_t) - \sigma_{IV}^{mkt}(\tau, S_t, F_{-\tau})}{2(r-q)\tau}$$

8.7 The smoothness of volatility log increments

It is not new that the volatility log increments are known to be normally distributed for a huge amount of traded assets. In order to assess that, we need to study the log increments of volatility $\delta \log \sigma$ at a time $t \in \mathbb{R}_+$ for a given time lag Δ for which $\min\{t_i\} = \Delta \forall i \in \{1, 2, \dots, N\}$ with $N\Delta = T = \max\{t_i\}$:

$$\delta \log \sigma = \left(\log \sigma_{t+\Delta} - \log \sigma_t \right)$$

However, a big question is still unanswered. How to find the Hurst H parameter so as we can simulate our Monte Carlo paths. The question can simply be answered using

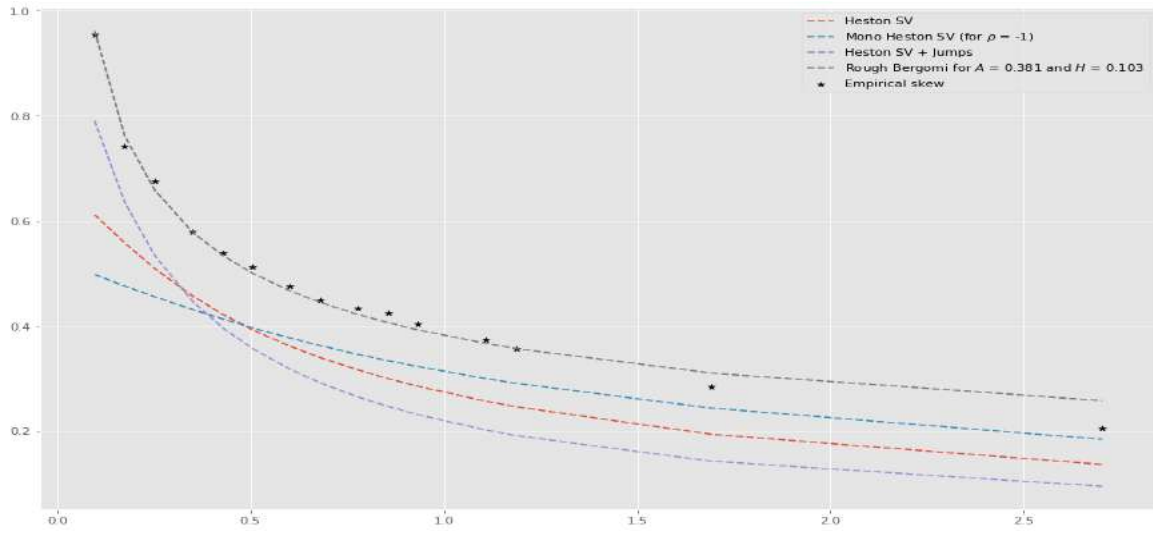


Figure 8.4: Term structure of S&P500 at-the-money-forward volatility skew and its alternative fit

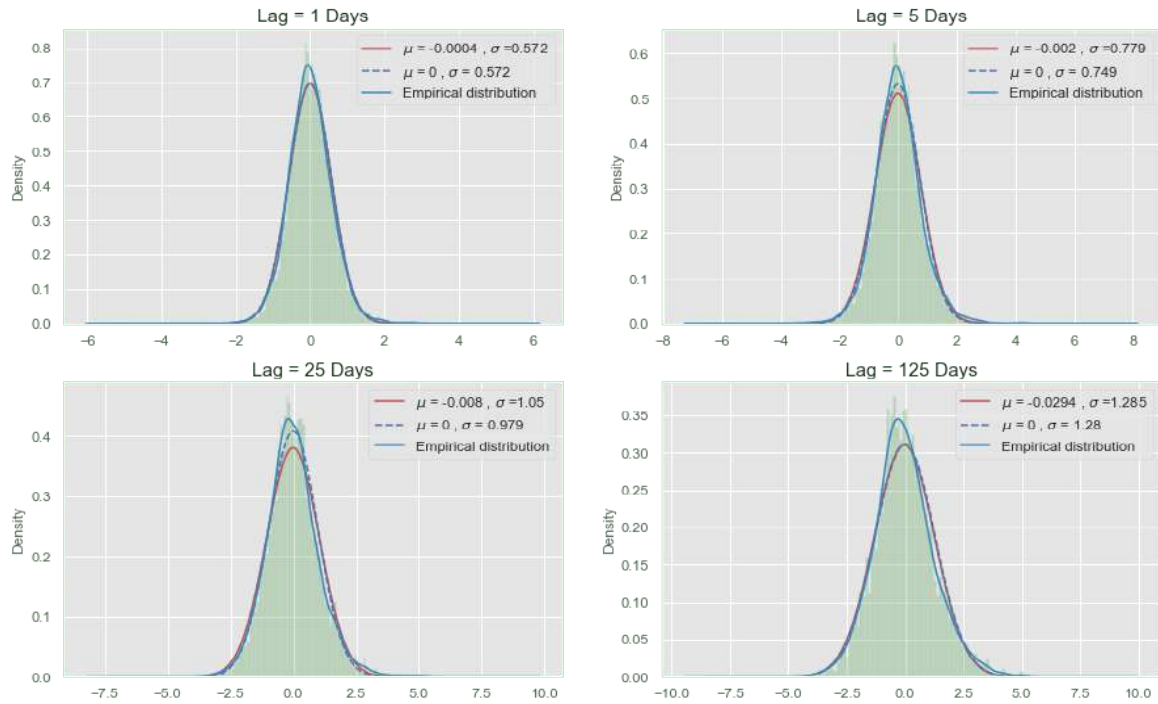


Figure 8.5: Log-increment volatility distribution for different lags

the moments. First we fix our q th moment and our time lag Δ sample a moments such :

$$m(q, \Delta) = \frac{|\log \sigma_{t+\Delta} - \log \sigma_t|^q}{N}$$

By looking to the generated moment graph we can approximate each of the q th moments by a power law such that $\log m(q, \Delta) \sim \Delta^{\zeta_q}$ showed by the figure 8.7. We finally can try to estimate the fractal scaling relationship such as :

$$\zeta_q = qH$$

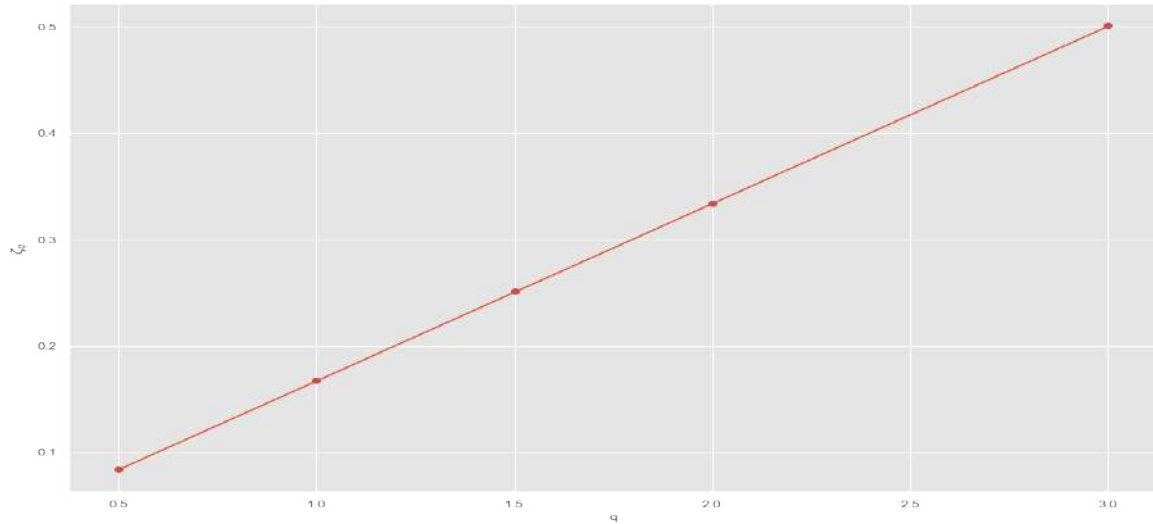


Figure 8.6: Linear fit between $\log m(q, \Delta)$ and qH

For $H \in (0, \frac{1}{2})$. A noticeable drawback in this estimation is the over dependency on the empirical data that can be very noisy. Moreover, Gatheral empirically found that our estimator $H = 0.166$ is biased high because the log-volatility increments are equally weighted. All of that even if the mean squared error of our linear estimator is equal to 1.95×10^{-08} for 5 evaluations.

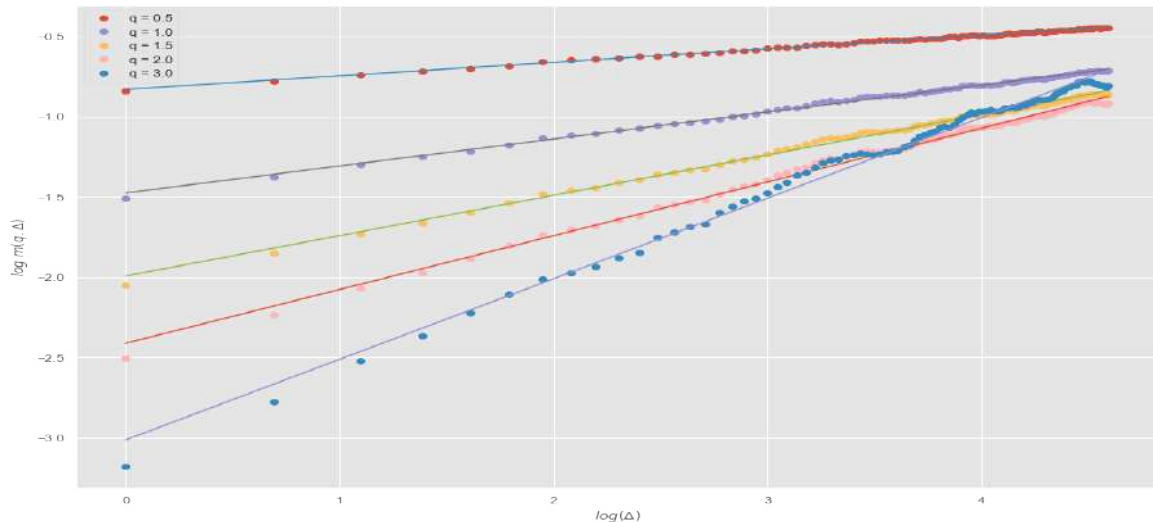


Figure 8.7: Log Moments of the Log volatility increments

Finally, Gatheral, Jaisson and Rosenbaum [18] showed that even the at-the-money skewness fitting method doesn't deliver satisfactory results in order to fit option prices. For this reason, we might try to estimate all our parameters using Monte Carlo simulations at our last resort.

8.8 Volatility skew of the rough Bergomi model

We observe that the volatility skew of the rough Bergomi model are fitting the data extremely well. A major drawback for now is the non existence of a rapid solving method, forcing us to use global optimizers or very good initial guesses, and in addition, there is no reason for the parameters to be stable. The differential algorithm, although gives very nice results, is extremely time consuming and can take up to several hours to produce results. The following initial guesses were used in order to calibrate the model using the Nelder-Mead algorithm, which is a local optimizer :

$$\begin{array}{cccc} \tilde{H} & \tilde{\eta} & \tilde{\rho} & \tilde{\xi}_0^T \\ \hline 0.102 & 2.173 & -0.871 & 0.0209 \end{array}$$

For the concern of the thesis, we used the Differential Algorithm, it took 3 hours to calibrate our 30000 Monte Carlo data points. For the Nelder-Mead algorithm during the optimisation we noticed that the parameters were smoothly and slowly evolving in the same direction optimization after optimization. For that reason we actualized the parameter after each optimisation from the next maturity. On our thesis, we calibrated the 4th parameter ξ_0^T but it shouldn't be mandatory as the volatility swap strike ξ^{mkt} are observable on the market for the S&P500 index.

τ	H	η	ρ	ξ_0^T	$IVMSE$
0.097	0.129	2.418	-0.825	0.031	0.249×10^{-5}
0.175	0.230	2.906	-0.722	0.037	0.139×10^{-5}
0.253	0.115	2.359	-0.871	0.047	0.349×10^{-5}
0.350	0.160	2.504	-0.810	0.055	0.150×10^{-5}
0.428	0.210	2.631	-0.754	0.059	0.040×10^{-5}
0.506	0.115	2.257	-0.870	0.062	0.068×10^{-5}
0.603	0.098	2.136	-0.919	0.063	0.025×10^{-5}
0.681	0.075	1.970	-0.996	0.061	0.016×10^{-5}
0.778	0.202	2.393	-0.795	0.076	0.007×10^{-5}
0.856	0.082	1.990	-0.999	0.068	0.021×10^{-5}
0.933	0.190	2.178	-0.826	0.077	0.025×10^{-5}
1.108	0.118	2.085	-0.911	0.083	0.011×10^{-5}
1.186	0.167	2.113	-0.837	0.086	0.009×10^{-5}
1.692	0.175	2.003	-0.833	0.096	0.016×10^{-5}
2.703	0.217	1.795	-0.808	0.112	0.030×10^{-5}

Table 8.2: Result of the rough Bergomi model calibration using a differential evolution algorithm

Although showing good results, the calibration time using a global optimizer is enormous. Which makes it hard to implement the rough Bergomi model in real time for financial institutions. However, we obtain a sort of compromise between time and fitting performance if we decide to use a local optimizer (it is still slow but a lot less than with a global optimizer) but we noticed in this case that the skew begins to deteriorate for very far maturities.

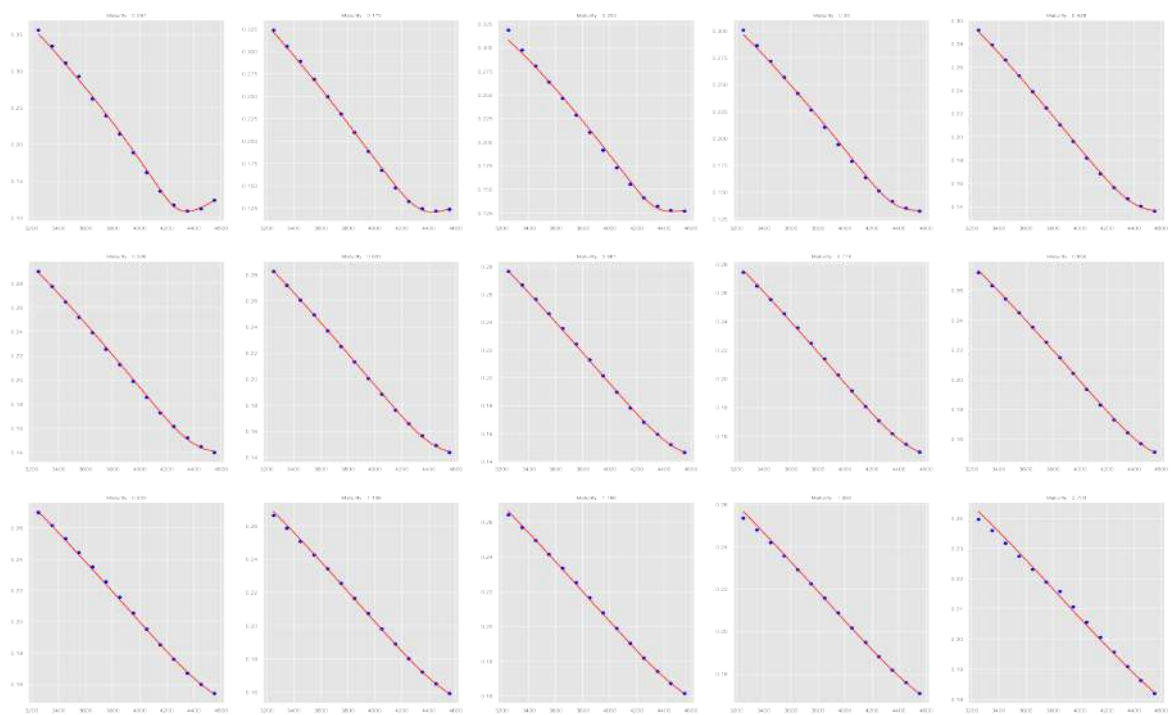


Figure 8.8: Rough Bergomi’s volatility skew

Chapter 9

Calibration

9.1 Cumulative Distribution functions approximation for SV models :

A huge drawback of the Gil-Peleaz approximation of the cumulative distribution written on 5.10 is the computational time required .For this reason, we preferred to use Lewis(2000) [23] formula:

$$C(S, K, T) = S_t - \sqrt{S_t K} \frac{1}{\pi} \int_0^\infty \frac{du}{u^2 + \frac{1}{4}} \operatorname{Re} \left[e^{-iuk} \phi_T(u - i/2) \right]$$

We then rewrite the initial call value $C_T(k)$ related to the risk-neutral density $d(s, \tau)$ by :

$$C(\tau, s, k) \equiv \int_k^\infty e^{-rT} \left(e^s - e^k \right) d(s, \tau) ds$$

We can see that $C(\tau, s, k)$ tends to S_0 as k tends to $-\infty$, and hence the call pricing function is not \mathbb{L}^1 defined and in other word $\lim_{k \rightarrow -\infty} C(\tau, s, k) = S_0 \neq 0$ and therefore doesn't converge . To obtain a measurable function Carr and Madan (1999) proposed to use a dumping factor [24] $e^{-\alpha k}$ such the modified call option price $c(\tau, s, k)$ is equal to :

$$c(\tau, s, k) \equiv e^{-\alpha k} C(\tau, s, k) \quad \alpha > 0$$

for $\alpha > 0$. For a range of positive values of α , we expect that $c(\tau, s, k)$ is square integrable in k over the entire real axis. We should now consider the Fourier transform of $c(\tau, s, k)$ defined such

$$\psi(\tau, u) = \int_{-\infty}^\infty e^{iuk} c(\tau, s, k) dk$$

We then use the inverse Fourier transform such we could retrieve our initial function $C(\tau, s, k)$:

$$C(\tau, s, k) = \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^\infty e^{-iuk} \psi(\tau, u) du$$

$$= \frac{e^{-\alpha k}}{\pi} \int_0^\infty \mathbb{R} \left(e^{-iuk} \psi(\tau, u) \right) du$$

And finally we end up with :

$$\begin{aligned} \psi(\tau, u) &= \int_{-\infty}^\infty e^{iuk} \int_k^\infty e^{\alpha k} e^{-rT} \left(e^s - e^k \right) d(s, \tau) ds dk \\ &= \int_{-\infty}^\infty e^{-rT} d(s, \tau) \int_{-\infty}^s \left(e^{s+\alpha k} - e^{(1+\alpha)k} \right) e^{iuk} dk ds \\ &= \int_{-\infty}^\infty e^{-rT} d(s, \tau) \left(\frac{e^{(\alpha+1+iu)s}}{\alpha + iv} - \frac{e^{(\alpha+1+iu)s}}{\alpha + 1 + iu} \right) ds \\ &= \frac{e^{-rT} \phi(u - (\alpha + 1), \tau) i}{\alpha^2 + \alpha - u^2 + i(2\alpha + 1)u} \end{aligned}$$

We can retrieve the original function using the fast Fourier transform algorithm for an N power of 2 so as to gain fast computing power (FFT divides and conquer matrix dimensions). In our case we took a number equal to 2^{15}

$$C(\tau, s, k) \approx \frac{\exp(-\alpha k)}{\pi} \mathbb{R} \left[\sum_{j=1}^N (e^{-iuk} \psi(\tau, u) \delta u) \right]$$

Time of execution(in ms)	Gil Peleaz Integral	Fast Fourier Algorithm
SV Heston for a single strike	10.7	24
SV Heston for 14 strikes	150	24.2
SVJ Heston for a single strike	12.5	25.4
SVJ Heston for 14 strikes	170	24.6

Table 9.1: Difference in the execution time between using the Gil Peleaz integral and the fast Fourier algorithm

The prices given by both integrals are generally very close in terms of values. We tested that using parameters $v_0 = 0.20$, $\tau = \frac{1}{12}$, $\kappa = 1$, $\theta = 0.181$, $\rho = -1$, $r_f = 0$, $q = 0$, $\lambda = 0.1$, $m = 0.1$, $\nu = 0.1$.

Call Price in \$	Gil Peleaz Integral	Fast Fourier Algorithm
SVJ Heston for an ATM option	97.20187	97.20187

Table 9.2: Difference in the accuracy of the Gil Peleaz integral and the fast Fourier algorithm

9.2 Calibration of the SV models

The SV models calibrations are extremely picky. The non convexity form of the problems makes the optimization hard to converge and requires the use of stochastic optimizers such as the differential evolution or other genetic algorithms. In practice

it is not possible to match the exact observed prices. Hence, calibration problem is formulated as a minimization problem such our aim is to minimize the pricing errors between the model prices and the market prices for a set of traded options. A common approach to measure this error is to use the squared difference between the market's and model's implied volatility, this approach leads to the nonlinear least square problem, for a vector parameter Θ , presented as :

$$\operatorname{argmin}_{\Theta} IVMSE = \sum_{\tau} \sum_k \left(\sigma_{IV}^2(\tau, s, k, \Theta) - \sigma_{IV}^{mkt2} \right)^2$$

In order to make the cost function $IVMSE$ convex enough would require convergent optimisation, which consists into changing the cost function into a convergent one by, for instance using a dumping factor and a function G such:

$$IVMSE' = e^{-\alpha G(\tau, k)} \left[IVMSE + G(\tau, k) \right]$$

The lack of liquidity for some options create a huge bid and ask spread that affects the option middle price in such a way it slows the optimizer so as to fit these abnormal prices, which will basically consume more time to calibrate. A slight change to improve the convergence speed of the optimizer would be to add weights in function of the bid-ask spread so our optimization, using a vector parameter $\Theta = (v_0, \kappa, \theta, \sigma, \rho, \lambda, m, \nu)$ for SV models or $\Theta = (\sigma, \eta, \rho, \gamma)$ for SSVI surfaces, would become :

$$IVMSE = \frac{1}{\sum_i w_{i,\tau}} \sum_{\tau} \sum_k \left(w_{i,\tau} \left(\sigma_{IV}^2(\tau, s, k, \Theta) - \sigma_{IV}^{mkt2}(\tau, s, k) \right) \right)^2$$

With weights that could be determined for instance as:

$$\begin{aligned} \text{Weight A : } w_{i,\tau} &= \frac{1}{|\text{spread}_i|} \\ \text{Weight B : } w_{i,\tau} &= \left(\frac{1}{\text{spread}_i} \right)^2 \\ \text{Weight C : } w_{i,\tau} &= \left(\frac{1}{\sqrt{|\text{spread}_i|}} \right) \\ \text{Weight D}^1 : w_{i,\tau} &= \vartheta_i^2 \\ \text{Weight E : } w_{i,\tau} &= \frac{1}{N} \end{aligned}$$

For weight D, the reason vega is used in quoting a spread is two fold:

- First, vega gives the change in price with respect to a change in volatility. So when we obtain a bid ask volatility, we can multiply by vega to get the bid ask in dollars.
- The second reason is that spreads are made wider for options whose prices have higher implied volatility in order to account for the risk,

For weights A, B, C the wider is the spread, the lighter is the weight of the point in the eye of the optimizer. Moreover, it has been shown on the paper [25], that the vega weighting underperforms every other type of weighting in the calibration. In our case, the best results were found using the weight A.

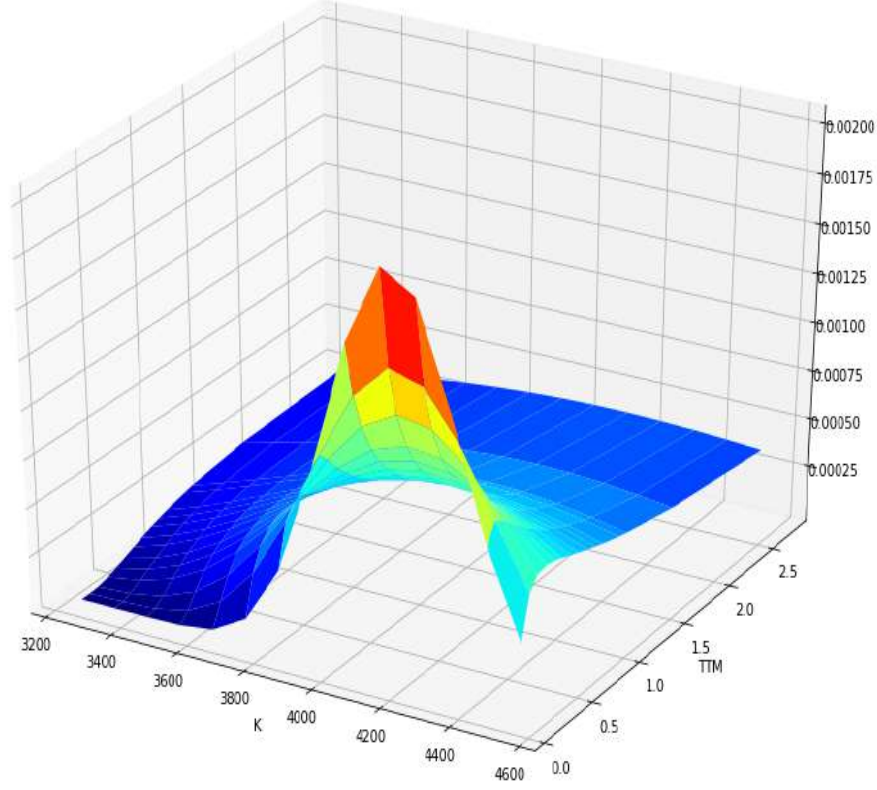


Figure 9.1: ϑ weights

Considered Errors Metrics : To evaluate the performance of all optimization methods, we measure the following errors for $i \in \{1, \dots, N\}$, :

$$\text{Average absolute relative error AARE}(\Theta) = \frac{1}{N} \sum_{i=1}^N \frac{|C_i^{\Theta} - C_i^*|}{C_i^*}$$

$$\text{Root-mean-square error: RMSE}(\Theta) = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i^{\Theta} - C_i^*)^2}$$

$$\text{Mean-square error: MSE}(\Theta) = \frac{1}{N} \sum_{i=1}^N (C_i^{\Theta} - C_i^*)^2$$

As we can see on the 4 dimension IV MSE error surfaces 9.3, the mispricing errors coming from stochastic volatility models are not convex at all and are extremely sensible to change of parameters. As a result, multiple combinations seem to give great fits and optimizers can fall very easily in local optimums.

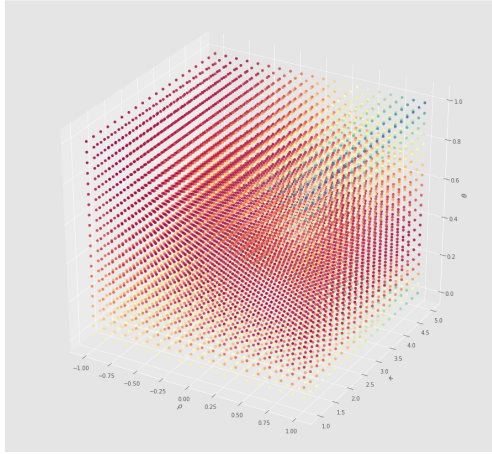


Figure 9.2: IVMSE scatter for the calibration of a simple Heston surface

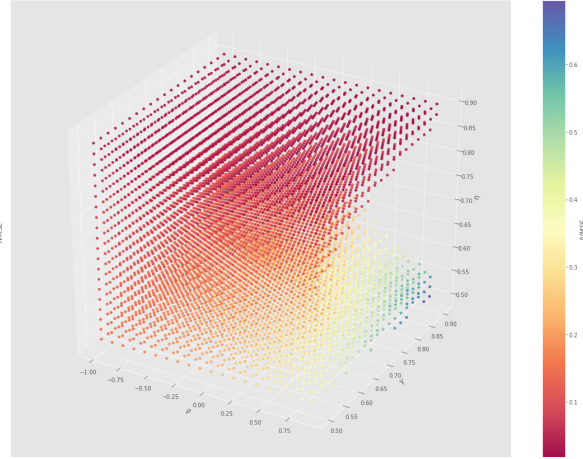


Figure 9.3: IVMSE scatter for the calibration of a surface SVI

	Summary
MSE	458.604
$AARE(in\%)$	6.516
$RMSE$	21.415
$IVMSE$	2.130×10^{-4}
N	210
σ	0.428
γ	1.406
η	0.172
ρ	-0.712

Table 9.3: SSVI Calibration Results

	Summary
MSE	153.992
$AARE(in\%)$	4.648
$RMSE$	12.409
$IVMSE$	3.460×10^{-4}
N	210
ρ	-0.914
σ	0.479
θ	0.061
κ	1.888
v_0	0.023

Table 9.4: Simple Heston Calibration Results

9.3 Optimizations

As showed on figure 9.3, the errors cost function is not linear along with parameters choices. A gradient descend algorithm working as: $\Theta_{k+1} = \Theta_k - \eta \nabla f(\Theta)$, such suggests the paper [26], won't certainly converge towards a global solution unless a very good guess is given. Instead we can use a global optimizer, which uses Metaheuristic methods to find good solutions candidate.

Definition 15 (Metaheuristic methods). *Metaheuristics methods are algorithms used to solve hard problems, generally involving high dimensional optimizations. These optimization algorithms do not insure to find good global optimum. However, Metaheuristic algorithms make very few assumptions about optimization problems and, couldn't they provide global optimum, would provide good enough approximations generally.*

Differential Evolution Algorithm In order to fit all our surface, we tried to use a stochastic optimizer developed by Storn and Price [27]. The selected algorithm works using the steps as follow:

- Define the objective cost function, in our case $IVMSE(\Theta)$ for $IVMSE : \mathbb{R}_{\mathbf{1}_{\rho}+}^n \rightarrow \mathbb{R}_+$
- Fix a high enough population number so as to get closer to the optimum. We set up the parameter to $Popsize = 50$.
- Choose a degree of mutation M such as $M \in (0, 2)$. This parameter choice is critical for the success and speed of the optimization. A too low degree of mutation would likely get the algorithm stuck at local optimums more often. On the other side if the degree of mutation is too high, the algorithm will converge more slowly because the large search radius becomes wider.
- The algorithm now mutates each candidate solution by mixing with other 2 candidates randomly in order to create a trial candidate. We used a well known strategy called "best1bin", which determines the newly candidate such:

$$b' = b_0 + \text{mutation} * (\text{population}[\text{random0}] - \text{population}[\text{random1}]) \quad (9.1)$$

- We also must choose a recombination number such $C \in (0, 1)$. The recombination dictates the probability of changing the population using in the strategy described in the step 9.1. The lower is the parameter the most stochastic is the research. Usually for problems we want fixed and converging solution. Therefore, this parameter is crucial and we set it up as $C = 0.70$.

During the last decade huge numbers of scientific and industrial problems were solved by the differential evolution because of its stochastic nature and ability to not be trapped very easily on local optimum using relevant parameters (parameters are very crucial in the process). However, constraints such as the feller condition need to be implemented in a loop 2 each time we select new population candidates. This process is very time consuming because of the need to rewrite efficiently the whole Differential Algorithm. A simpler but "unhealthy" way of doing so would be to penalize errors for which parameters don't obey fixed constraints. The penalty was set as a multiplier equal to 10^{50} in our case.

Nelder-Mead Algorithm In order to polish the result we also used a Nelder-Mead optimizer using the initial guess of the differential evolution algorithm. The Nelder-Mead algorithm employs a gradient free heuristic process that doesn't guarantee the finding of a global optimum. In contrary to the differential equation algorithm, Nelder-Mead's algorithm is not stochastic and can be stuck easily on local optimum, that's why we are using it only to polish our results. The solution's search method of the algorithm is explained through the following steps and available on the original paper of J. Nelder and R. Mead [28] :

Algorithm 2 Differential Evolution Algorithm

```

procedure DE HESTON OPTIMIZER (self)           ▷ We don't need an initial guess
   $t = 0; j = 0; x = 0; tol = 1; CR = 0.7; M = 1; popsize = 50$ 
  Set dimension  $N = 5$                                ▷  $N = 8$  for SVJ Heston
  Set bounds  $bnds$  for the different parameters       ▷ for instance  $|\rho| \leq 1$ 
  Set constraints  $cons$  for the different parameters  ▷ for instance  $2\kappa\theta - \sigma^2 > 0$ 
  while  $tol \geq e^{-9}$  do
    if  $x = 0$  then
       $pen[j] = 1$ 
      for  $i$  in range(0,  $popsize$ ) do
         $X[i] = random(size = N, bounds = bnds)$ 
      end for
       $X_j = X[randint]$ 
       $b_0 = X_j$ 
       $X_t = b_0$ 
      if  $cons$  are violated then
         $pen[j] = e^{50}$ 
      end if
       $f[X_t] = pen[j] \times IVMSE(X_t)$ 
    else:
       $pen[j] = 1$ 
      for  $i$  in range(0,  $popsize$ ) do
         $X[i] = random(size = N, bounds = bnds)$ 
      end for
       $b' = b_0 + M[X[randint_1] - X[randint_2]]$ 
       $C = randuniform(0, 1)$ 
      if  $C \leq CR$  then
         $X_j = b'$ 
        if  $cons$  are violated then
           $pen[j] = e^{50}$ 
        end if
        if  $pen[j] \times IVMSE(X_j) < f[X_t]$  then
           $X_t = X_j; b_0 = X_t$ 
           $f[X_t] = pen[j] \times IVMSE(X_t)$ 
        else:
           $X_t = X_{t-1}$ 
        end if
      else:
         $X_j = b_0; X_t = X_{t-1}$ 
      end if
       $tol = f[X_t] - f[X_{t-1}]$ 
    end if
     $t = t + 1; j = j + 1$ 
  end while
  The best set of parameters is  $X_t$ 
end procedure

```

- Define the N dimension problem such and set an initial parameter x_{init} and we generate n points to rank them such as $IVMSE(x_0) < IVMSE(x_1) \dots < IVMSE(x_n)$. The dimension for this problem is $n + 1$.
- Using these $n + 1$ points we set a simplex in $n + 1$ dimensions.
- We compute the centroid c of all the $IVMSE(x_i) \quad \forall i \in \{0, \dots, n - 1\}$ such $c = \frac{\sum_{i=0}^{n-1} x_i}{n}$ and call the point. This will serve us during all the next steps.
- We firstly implement the reflection method, which consists of choosing a point from a basis of the worst one (x_n in our case) such :

$$x_r = x_n + \alpha(c - x_n)$$

- if the point's image is superior to the worst one we can try to have a "momentum" chance by using the expansion method such :

$$x_r = x_r + \gamma(x_r - c)$$

- Lastly, we must try the contraction method, in which we try to find a point from the centroid towards the interior of the simplex such :

$$x_r = c + \rho(x_n - c)$$

- We replace the worst point x_0 by x_r and continue to do so for the other points until the maximum of iterations or the minimum tolerance is reached.

Usually the parameters are fixed such $\rho = .50, \gamma = 2$ and $\alpha = 1$. Moreover, we omitted one step, which is shrinkage in our case. Shrinkage are usually a redefined simplex and it is very unlikely to happen in practice. In addition, all efficient Nelder-Mead algorithms do not use it. A drawback of the Nelder-Mead algorithm is that it can't handle neither bounds nor constraints. As a result, During our implementation we used penalties for either bounds' or constraints' violations. On our quick implementation we omitted to input penalties for brevity. Otherwise each step would have to carry an *if* condition, making the pseudo-code too large to display. The implementation of the Nelder-Mead algorithm is given on the pseudo-code 3.

	Summary
MSE	669.304
$AARE(in\%)$	6.862
$RMSE$	25.871
$IVMSE$	4.430×10^{-4}
N	210
ρ	-0.878
σ	0.635
θ	0.042
κ	4.745
v_0	0.014
λ	0.024
m	-0.106
ν	0.088

Table 9.5: SVJ Heston calibration results

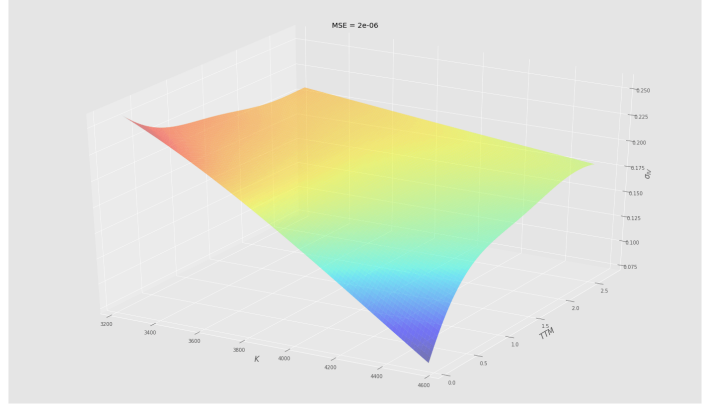


Figure 9.4: SVJ surface generated by Gaussian process interpolation

Algorithm 3 Nelder-Mead Algorithm

```

procedure NM HESTON OPTIMIZER (self,  $x_0$ )
     $\rho = .50; \gamma = 2; \alpha = 1; j = 0; tol = 1$ 
     $n = len(x_0), M = \{\}, pen = \{\}$ 
    Generate  $n$  points such  $IVMSE(x_0) < IVMSE(x_1) \dots < IVMSE(x_n)$ .
    while  $tol > e^{-9}$  do
         $X = argmax(IVMSE); c = \frac{\sum_{i=0}^{n-1} x_i}{n-1}$ 
         $x_r = X + \alpha(c - X)$ 
        if  $IVMSE(X) < IVMSE(x_r)$  then
             $x_{r'} = x_r + \gamma(x_r - c)$ 
            if  $IVMSE(x_{r'}) < IVMSE(x_r)$  then
                 $X = x_{r'}$ 
            else
                 $X = x_r$ 
            end if
        else
             $x_l = X + \rho(X - c)$ 
             $X = x_l$ 
        end if
         $M[j] = X; j = j + 1$ 
         $argmin_{\Theta} IVMSE = M[j]$ 
         $tol = IVMSE(M[j]) - IVMSE(M[j - 1])$ 
    end while
end procedure

```

▷ We need an initial guess
 ▷ Reflection
 ▷ Expansion
 ▷ Contraction

9.4 SVI calibration

For the SVI calibration, we used an optimization scheme based on Zeliad's white paper [29]. Recall the raw SVI parameterization of the total implied variance :

$$w_{\text{imp}}^{\text{SVI}}(x) = a + b \left(\rho(x - m) + \sqrt{(x - m)^2 + \sigma^2} \right)$$

We could do a change of variables such :

$$y(x) = \frac{x - m}{\sigma}$$

Under this change of variables, the total implied variance in the raw parameterization becomes equal to:

$$\begin{aligned} w_{\text{imp}}^{\text{SVI}}(x) &= a + b\sigma \left(\rho y(x) + \sqrt{y(x)^2 + 1} \right) \\ &= \hat{a} + dy(x) + cz(x) \end{aligned}$$

Where :

$$\begin{aligned} \hat{a} &= a, & d &= \rho b\sigma \\ c &= b\sigma, & z(x) &= \sqrt{y(x)^2 + 1}. \end{aligned}$$

We just created a linear formula of the total implied variance, depending on 3 terms \hat{a} , d and c given m and σ . Therefore, we employ a quasi-explicit (QE) parameterization for the convex linear problem. But first of all, we need to fix our parameters' domains

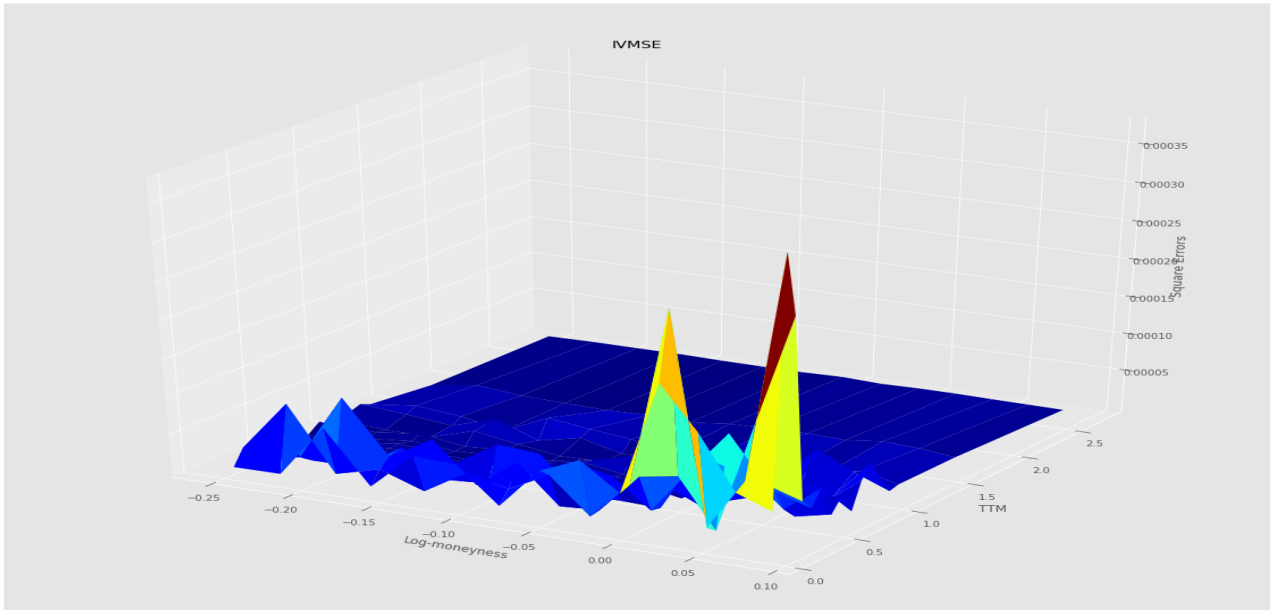


Figure 9.5: IVMSE errors of the SSVI surface

correctly such in domain \mathbb{R}^3 we have :

$$\mathcal{D} = \left\{ (\hat{a}, d, c) : \begin{cases} 0 \leq c \leq 4\sigma \\ |d| \leq c \\ |d| \leq 4\sigma - c \\ 0 \leq \hat{a} \leq \max_i \{w_i\} \end{cases} \right\}$$

For a fixed set (σ, m) and respectively an empirical set of log moneyness and implied total volatility $x_i, w_i \forall i \in 0, \dots, N$, we are facing the optimization problem :

$$\operatorname{argmin}_{(\hat{a}, d, c) \in \mathcal{D}} C_1(\hat{a}, d, c, x_i, w_i, m', \sigma')$$

where $C_1(\hat{a}, d, c)$ is the first quadratic cost function for m and σ , the given iterated guesses.

$$C_1(\hat{a}, d, c, x_i, w_i, m, \sigma) = \sum_{i=0}^n (\hat{a} + dv(x_i) + cz(x_i) - w_i)^2$$

After that we could implement the whole optimization for σ and m using a standard cost function minimisation such :

$$\operatorname{argmin}_{(m, \sigma) \in \mathcal{E}} C_2(m, \sigma, x_i, w_i, \hat{a}^*, d^*, c^*, x_i, w_i)$$

where $C_2(m, \sigma, x_i, w_i, \hat{a}^*, d^*, c^*, x_i, w_i)$ is the second quadratic cost function viewed for \hat{a}^*, d^*, c^* constant for each iterated guess m, σ .

$$C_2(m, \sigma, x_i, w_i, \hat{a}^*, d^*, c^*, x_i, w_i) = \sum_{i=0}^n (w(m, \sigma)_i - w_i)^2$$

With the parameter domain being this time :

$$\mathcal{E} = \left\{ (m, \sigma) : \begin{cases} \min\{x_i\} \leq m \leq \max\{x_i\} \\ \sigma > 0 \end{cases} \right\}$$

In order to have parameters that satisfy Durrelman condition 5.21 so as the price of a cliquet option doesn't become negative and provokes butterfly arbitrage opportunities, we need to rewrite our parameters using the SVI Jump-Wing reparametrization 5.22.

Chapter 10

Interpolations

10.1 Multivariate normal distribution

In order to interpolate the obtained implied volatility, so as to make a smooth surface, we decided to use Gaussian process regressions. In order to show that, let's first define the most basic properties of the Gaussian distribution : A Gaussian random variable $X \sim \mathcal{N}(\mu, \Sigma)$ where μ is the mean and Σ is the covariance matrix has the following probability density function:

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|} e^{-\frac{1}{2}((x-\mu)^\top \Sigma^{-1}(x-\mu))} \quad (10.1)$$

Where $|\Sigma|$ is the determinant of the covariance matrix

Proof of 10.1. A multivariate probability distribution function (PDF) can be originally written as a product of multiple Gaussians' PDFs of different dimensions such :

$$P(x_1; x_2; \mu, \Sigma) = \frac{1}{G} e^{-\frac{1}{2} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{1}{2} \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)} \quad (10.2)$$

With G , a normalization term.

If $|\Sigma|$ is a positively defined matrix, then, using eigenvalue decomposition we can write it under the form :

$$\begin{aligned} \Sigma &= Q \Lambda Q^T \\ \rightarrow \Sigma^{-1} &= Q \Lambda^{-1} Q^T \end{aligned}$$

With Λ , a diagonal matrix that contains all the positive eigenvalues.

Remark 11. The reason behind the choice of the Gaussian distribution is that it tends to occur very often in real world data (we saw it for log volatility increments under the rough Bergomi's model) for a good reason: the Central Limit Theorem (CLT). The CLT states that the arithmetic mean samples is approximately normal distributed - independent of the original sample distribution (provided it has finite mean and variance).

We then set a new variable z and y , respectively such :

$$\begin{aligned} z &= \Lambda^{-\frac{1}{2}} y \\ y &= Q^T(x - \mu) \end{aligned}$$

We can now reconstitute the original probability distribution function normalized by an unknown term G for a d -dimensional multivariate Gaussian :

$$\begin{aligned} P(x; \mu, \Sigma) &= \frac{1}{G} e^{-\frac{1}{2}(x-\mu)^T Q \Lambda^{-1} Q^T (x-\mu)} \\ &= \frac{1}{G} e^{-\frac{1}{2} y^T \Lambda^{-1} y} \\ &= \frac{1}{G} e^{-\frac{1}{2} \sum_i^d \frac{y_i^2}{\lambda_i}} \\ &= \frac{1}{G} \prod_{i=1}^d e^{-\frac{y_i^2}{2\lambda_i}} \end{aligned}$$

We know that a PDF always integrate to 1 and therefore we deduce G from that such:

$$\begin{aligned} G &= \int_{y_d} e^{-\frac{y_d^2}{2\lambda_d}} \cdots \int_{y_1} e^{-\frac{y_1^2}{2\lambda_1}} dy_1 \cdots dy_d \rightarrow \text{By applying Fubini's theorem} \\ &= \int_{y_d} e^{-\frac{y_d^2}{2\lambda_d}} \cdots \int_{y_2} e^{-\frac{y_2^2}{2\lambda_2}} \sqrt{2\pi\lambda_1} dy_2 \cdots dy_d \rightarrow \text{By applying polarization} \\ &= \prod_{i=1}^d \sqrt{2\pi\lambda_i} \\ &= (2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \end{aligned}$$

□

10.2 Properties of multivariate normal distribution

In one sentence : Gaussian once , Gaussian forever .In reality linear operations on Gaussian variables such integration , addition , subtraction , division and multiplication don't generally modify their normal properties(unless if we multiply the i.i.d variable by 0 or subtract by itself).Before we continue,some important properties are listed such :

- **Marginalization:** The marginal distributions $p(y_A) = \int_{y_B} p(y_A, y_B; \mu, \Sigma) dy_B$ and $p(y_B) = \int_{y_A} p(y_A, y_B; \mu, \Sigma) dy_A$ are Gaussian.
- **Summation:** If $y \sim \mathcal{N}(\mu, \Sigma)$ and $y' \sim \mathcal{N}(\mu', \Sigma')$, then

$$y + y' \sim \mathcal{N}(\mu + \mu', \Sigma + \Sigma') .$$

- Conditioning: The conditional distribution of y_A on y_B

$$p(y_A | y_B) = \frac{p(y_B | y_A)p(y_A)}{p(y_B)} = \frac{p(y_A, y_B; \mu, \Sigma)}{\int_{y_A} p(y_A, y_B; \mu, \Sigma) dy_A}$$

10.3 Covariance matrix estimation

The most crucial distribution's parameter to estimate is the covariance matrix in our case, of course the mean is important but the parameter that assesses uncertainty is without doubt the covariance matrix. In order to evaluate the covariance matrix, one would first need to estimate it by using some kernel functions types. In our case we used a radial basis function kernel defined such:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma_l^2}\right)$$

With σ_l a constant, this kernel function leads us to our $\mathbb{R}_+^{d \times d}$ covariance matrix constructed such :

$$\Sigma = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots & K(x_1, x_d) \\ K(x_1, x_2)^T & \ddots & & \ddots & K(x_2, x_d) \\ \vdots & & \ddots & & \ddots & \vdots \\ K(x_1, x_d)^T & K(x_2, x_d)^T & \dots & K(x_d, x_d) \end{bmatrix}$$

The main advantage of using the "super" RBF kernels here is to grasp the whole inter-dimensional connections in and between our data towards the infinite dimensional space thanks to an extremely efficient inner product transformation. Isn't that nice? For better understanding let's define the multidimensional matrix based on the 2-dimensional coordinates vectors x_1 and x_2 of dimensions $n \times 1$ each, we will be working on :

$$\begin{bmatrix} A \\ x_1 \\ x_2 \\ x_1 x_2 \\ x_1^2 \\ x_2^2 \\ x_1^2 x_2 \\ \vdots \\ \infty \end{bmatrix}$$

With A , an $n \times 1$ matrix equal to $A = [1, 1, \dots, 1]^T$. More details on the RBF kernels are available on the paper [30].

10.4 Gaussian Process Regression

First, in order to use a Gaussian process regression, we need to separate our data in two, the train data $f(\text{strike}, \text{ttm}) = \sigma(\text{strike}, \text{ttm})_{IV}$ and the test data $f(\text{strike}', \text{ttm}') = \sigma(\text{strike}', \text{ttm}')_{IV}$. Through Bayesian's rule our samples are then distributed such :

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}' \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}') \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) & K(\mathbf{X}, \mathbf{X}') \\ K(\mathbf{X}', \mathbf{X}) & K(\mathbf{X}', \mathbf{X}') \end{bmatrix} \right)$$

Please note the following notation definitions below:

- $m(\cdot)$ represents a mean function
- \mathbf{X} is a $(N, 2)$ matrix of training features
- \mathbf{X}' is a $(N', 2)$ matrix of test points
- \mathbf{f} is a $(N, 1)$ vector of realizations of the Gaussian Process on training features \mathbf{X}
- \mathbf{f}' is a $(N', 1)$ vector of realizations of the Gaussian Process on test points \mathbf{X}'

As the exact implied volatility is found through the well optimization $\sigma(K, \tau)_{IV}$ for a given forward price F_t we can keep our covariance matrix as originally set. However, because our data are sparse in time and mostly for the long maturities, we can add uncertainty to the train covariance matrix. It also allows us to avoid the case where the variance becomes equal to 0, which is impossible. This implementation is possible by setting :

$$y_{i,j} = f(\text{strike}_i, \text{ttm}_j) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

such our data become finally distributed as:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}' \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}') \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma^2 I & K(\mathbf{X}, \mathbf{X}') \\ K(\mathbf{X}', \mathbf{X}) & K(\mathbf{X}', \mathbf{X}') \end{bmatrix} \right)$$

We now rewrite our main equation. In addition, we add a Gaussian variable \mathbf{z} that will help us derive our conditional mean and covariance matrix :

$$\begin{aligned} \mathbf{y} &= \mathbf{f} + \varepsilon \\ \mathbf{f}' &= m(\mathbf{X}') + \varepsilon_1 \\ \mathbf{z} &= \mathbf{f}' - K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} \mathbf{y} \end{aligned}$$

We set $A = -K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1}$ and ε_1 a standard normal variable. This is a linear combination of normal distributed random variables and as such itself normal distributed with conditional expectation of \mathbf{f}' given \mathbf{y} derived such :

$$\begin{aligned} E(\mathbf{f}'|\mathbf{y}) &= E(\mathbf{z} - \mathbf{A}\mathbf{y}|\mathbf{y}) \\ &= m(\mathbf{X}') + A(\mathbf{y} - m(\mathbf{X})) \end{aligned}$$

And the covariance matrix is derived such :

$$\begin{aligned} var(\mathbf{f}'|\mathbf{y}) &= var(\mathbf{z} - \mathbf{A}\mathbf{y}|\mathbf{y}) \\ &= var(\mathbf{z}|\mathbf{y}) + \mathbf{var}(\mathbf{A}\mathbf{y}|\mathbf{y}) - \mathbf{A} \text{cov}(\mathbf{z}, -\mathbf{y}) - \text{cov}(\mathbf{z}, -\mathbf{y})\mathbf{A}^T \\ &= var(\mathbf{z}|\mathbf{y}) \\ &= var(\mathbf{z}) \end{aligned}$$

for $var(\mathbf{f}'|\mathbf{y})$ derived as :

$$\begin{aligned} var(\mathbf{f}'|\mathbf{y}) &= var(\mathbf{z}) = var(\mathbf{f}' + A\mathbf{y}) \\ &= var(\mathbf{f}') + Avar(\mathbf{y})A^T + A \text{cov}(\mathbf{f}', \mathbf{y}) + \text{cov}(\mathbf{f}', \mathbf{y})A^T \\ &= K(\mathbf{X}', \mathbf{X}') + K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} (K(\mathbf{X}, \mathbf{X}) \\ &\quad + \sigma^2 I) (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} K(\mathbf{X}', \mathbf{X}) - 2K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) \\ &\quad + \sigma^2 I)^{-1} K(\mathbf{X}', \mathbf{X}) \\ &= K(\mathbf{X}', \mathbf{X}') - K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} K(\mathbf{X}', \mathbf{X}) \end{aligned}$$

Therefore now, for each possible point on a surface we are able to predict our mean conditionally to the 2 coordinates dimensions. Moreover, it is now possible for us to have a confidence interval for each prediction. For a more complete understanding about the Gaussian process regression one may try to read [31], which covers a huge amount of material concerning the Gaussian process regression.

Conclusion

During this thesis, we compared the different properties of standard stochastic volatility models by analyzing in details their skews, their term structures and their accuracy with replicating original markets data.

We then showed that stochastic volatility models and the original at-the-money terms structures are not correctly fitting the market's empirical skews and we therefore tried to implement some solutions such as the stochastic volatility model with jumps but its major drawback is the extreme hardness for first, the optimizer to converge to a determined set of solution, and second, to reproduce prices correctly with a reasonable amount of errors.

As the standard stochastic volatility approaches weren't efficient enough to fit the models to our empirical data we decided to adopt a model lead by a non-Markovian process using volatility swaps. This model, which was originally derived by Bergomi had a huge economic sens as we were deriving the realized volatility through a time-weight of the volatility swaps for different maturities. However, the Bergomi's model originally used a semi-Martingale approach for its filtration, impeding us to reproduce a likely simulation of realized volatility using Monte-Carlo's technique.

The results yielded by the rough Bergomi model using a semi-stationary fraction Brownian motions are however satisfactory as we were able to fit all the skews and term structures correctly using our models. The advantage of such model resides in its tractability and easy understanding. The rough Bergomi model has only 3 parameters, highly easy to understand, to optimize. This tractability allows us to remove the pain of adding and calibration jump processes from our mind.

Nevertheless, the lack of existence of closed or semi-closed form formula forces us to use extremely computational expensive Monte-Carlo simulations. In addition, we need to use a high enough number of simulations so as our simulated option prices converge.

Chapter 11

Appendix

11.1 Black and Scholes

11.1.1 Call option price derivation

$$\begin{aligned} C(S_t, K, r, T, t) &= E(e^{-r(T-t)}(S_T - K)^+) \\ &= e^{-r(T-t)} \int_{-\infty}^{\infty} \max[S_t e^z - K, 0] f(z) dz \\ &= e^{-r(T-t)} \left(\int_{-\infty}^{\ln \frac{K}{S_t}} 0 \cdot f(z) dz + \int_{\ln \frac{K}{S_t}}^{\infty} (S_t e^z - K) f(z) dz \right) \\ &= e^{-r(T-t)} \int_{\ln \frac{K}{S_t}}^{\infty} (S_t e^z - K) f(z) dz \\ &= e^{-r(T-t)} \left(S_t \int_{\ln \frac{K}{S_t}}^{\infty} e^z f(z) dz - K \int_{\ln \frac{K}{S_t}}^{\infty} f(z) dz \right) \\ &= \frac{e^{-r(T-t)}}{2\sqrt{\pi(T-t)}\sigma} \left(S_t \int_{\ln \frac{K}{S_t}}^{\infty} e^z e^{\frac{-\frac{1}{2}(z-(r-\frac{\sigma^2}{2})(T-t))^2}{\sigma^2(T-t)}} dz - K \int_{\ln \frac{K}{S_t}}^{\infty} e^{\frac{-\frac{1}{2}(z-(r-\frac{\sigma^2}{2})(T-t))^2}{\sigma^2(T-t)}} dz \right) \end{aligned}$$

Replacing z by the standard normal variable y as :

$$z = \left(r - \frac{\sigma^2}{2} \right) (T - t) + \sigma\sqrt{T - t}y$$

Inserting y in the equation such we end up with 2 standard normal probabilities:

$$\begin{aligned} &= S_t \Phi \left(-\frac{\ln \frac{K}{S_t} - \left(r - \frac{\sigma^2}{2} \right) (T - t)}{\sigma\sqrt{T - t}} + \sigma\sqrt{T - t} \right) \\ &\quad - K e^{-r(T-t)} \Phi \left(\frac{\ln \frac{S_t}{K} + \left(r - \frac{\sigma^2}{2} \right) (T - t)}{\sigma\sqrt{T - t}} \right) \end{aligned}$$

$$= S_t \Phi \left(\frac{\ln \frac{S_t}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma \sqrt{T-t}} \right) - K e^{-r(T-t)} \Phi \left(\frac{\ln \frac{S_t}{K} + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma \sqrt{T-t}} \right)$$

11.1.2 Greeks' derivation

The Greek's were derived for Call options. For Put options, the derivations are very similar

Vega Derivation

$$\frac{\partial C}{\partial \sigma} = S_t \frac{\partial N(d_1)}{\partial \sigma} - K \frac{\partial N(d_2)}{\partial \sigma}$$

Applying the Chain Rule and finally obtain :

$$\begin{aligned} &= S_t \frac{\partial N(d_1)}{\partial d_1} \frac{\partial d_1}{\partial \sigma} - K \frac{\partial N(d_2)}{\partial d_2} \frac{\partial d_2}{\partial \sigma} \\ &= \frac{S_t}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}} \left(\frac{-1}{\sigma} \right) (d_2) - \frac{K e^{-r\tau}}{\sqrt{2\pi}} e^{-\frac{d_2^2}{2}} \left(\frac{-1}{\sigma} \right) (d_1) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{d_1^2}{2}} \left[-\frac{S_t d_2}{\sigma} + \frac{K e^{-r\tau} d_1}{\sigma} e^{\frac{d_1^2}{2} - \frac{d_2^2}{2}} \right] \\ &= N'(d_1) \left[-\frac{S_t d_2}{\sigma} + \frac{K e^{-r\tau} d_1}{\sigma} e^{\frac{1}{2}(d_1 - d_2)(d_1 + d_2)} \right] \\ &= N'(d_1) \left[-\frac{S_t d_2}{\sigma} + \frac{K e^{-r\tau} d_1}{\sigma} e^{\frac{1}{2}\sigma\sqrt{\tau} \frac{2 \ln \frac{S_t}{K} + 2r\tau}{\sigma\sqrt{\tau}}} \right] \\ &= N'(d_1) \left[-\frac{S_t d_2}{\sigma} + \frac{S_t d_1}{\sigma} \right] \\ &= S_t N'(d_1) \sqrt{\tau} \end{aligned}$$

We pursue similarly for the other Greeks :

Gamma derivation

$$\begin{aligned} \frac{\partial^2 C}{\partial S_t^2} &= \frac{\partial N(d_1) + S_t \frac{\partial N(d_1)}{\partial S_t} - K e^{-r\tau} \frac{\partial N(d_2)}{\partial S_t}}{\partial S_t} \\ &= 2N'(d_1) \frac{\partial d_1}{\partial S_t} + S_t \frac{\partial^2 N(d_1)}{\partial d_1^2} \frac{\partial^2 d_1}{\partial S_t^2} - K e^{-r\tau} \frac{\partial^2 N(d_2)}{\partial d_1^2} \frac{\partial^2 d_2}{\partial S_t^2} \\ &= N'(d_1) \frac{\partial d_1}{\partial S_t} \\ &= \frac{N'(d_1)}{S_t \sigma \sqrt{\tau}} \end{aligned}$$

Rho derivation

$$\frac{\partial C}{\partial r} = S_t \frac{\partial N(d_1)}{\partial d_1} \frac{\partial d_1}{\partial r} - K \frac{\partial N(d_2)}{\partial d_2} \frac{\partial d_2}{\partial r}$$

$$\begin{aligned}
&= \tau K e^{-r\tau} N(d_2) + S_t \frac{\partial N(d_1)}{\partial d_1} \frac{\partial d_1}{\partial r} - K e^{-r\tau} \frac{\partial N(d_2)}{\partial d_2} \frac{\partial d_2}{\partial r} \\
&= \tau K e^{-r\tau} N(d_2) + S_t N'(d_1) \frac{\partial d_1}{\partial r} - K e^{-r\tau} N'(d_2) \frac{\partial d_2}{\partial r} \\
&= \tau K e^{-r\tau} N(d_2) + \frac{\partial d_1}{\partial r} [S_t N'(d_1) - K e^{-r\tau} N'(d_2)] \\
&\approx \tau K e^{-r\tau} N(d_2)
\end{aligned}$$

Theta derivation

$$\begin{aligned}
\frac{\partial C}{\partial \tau} &= S_t \frac{\partial N(d_1)}{\partial d_1} \frac{\partial d_1}{\partial \tau} - K \frac{\partial N(d_2)}{\partial d_2} \frac{\partial d_2}{\partial \tau} \\
&= S_t N'(d_1) \frac{\partial d_1}{\partial \tau} - K e^{-r\tau} N'(d_2) \frac{\partial d_2}{\partial \tau} + r K e^{-r\tau} N(d_2) \\
&= S_t N'(d_1) \frac{\partial (d_1 - d_2)}{\partial \tau} + r K e^{-r\tau} N(d_2) \\
&= S_t N'(d_1) \frac{\partial (\sigma \sqrt{\tau})}{\partial \tau} + r K e^{-r\tau} N(d_2) \\
&= S_t N'(d_1) \frac{\sigma}{2\sqrt{\tau}} + r K e^{-r\tau} N(d_2)
\end{aligned}$$

11.2 Heston Model

11.2.1 Gil Pelaez Proof(based on the paper [1])

$$\begin{aligned}
M(x) &= \frac{1}{2\pi} \int_{\psi}^{\lambda} \frac{e^{iux} \phi(-u) - e^{-iux} \phi(u) du}{iu} \\
&= \frac{1}{2\pi} \int_{\psi}^{\lambda} \int_{-\infty}^{\infty} \frac{e^{iu(x-y)} - e^{-iu(x-y)} f_y dy du}{iu} \\
&= \frac{1}{2\pi} \int_{\psi}^{\lambda} \int_{-\infty}^{\infty} \frac{\cos -u(y-x) + i \sin -u(y-x) - (\cos u(y-x) + i \sin u(y-x)) f_y dy du}{iu}
\end{aligned}$$

By the property of Conjugate and symmetry :

$$M(u) = \frac{1}{2\pi} \int_{\psi}^{\lambda} \int_{-\infty}^{\infty} \frac{-2 \sin u(y-x) f_y dy du}{iu}$$

Rewriting it as :

$$\begin{aligned}
M(x) &= \frac{1}{2} \left(\frac{1}{\pi} \int_{\psi}^{\lambda} \int_{-\infty}^{\infty} \frac{-2 \sin u(y-x) f_y dy du}{iu} \right) \\
M(x) &= -\frac{1}{2} \text{sign}(u, y)
\end{aligned}$$

With $\text{sign}(u, y)$, a sort of continuous Heaviside function available on the figure 11.1 :

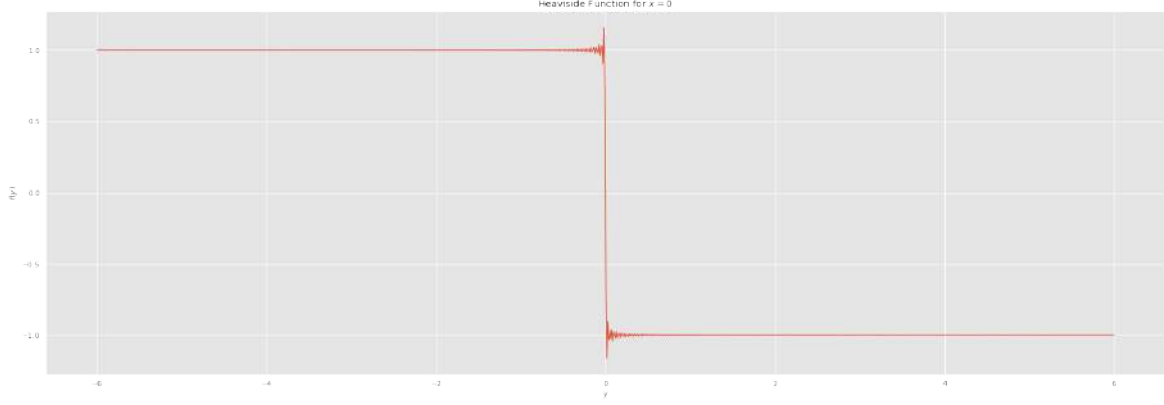


Figure 11.1: Heaviside Function

Afterward, we straightforwardly continue the derivation such :

$$\begin{aligned} M(x) &= -\frac{1}{2} \left(\int_{-\infty}^x -f_y dy + \int_x^{\infty} f_y dy \right) \\ &= -\frac{1}{2} (1 - 2F(x)) \end{aligned}$$

We set $\lambda \rightarrow \infty$ and $\psi \rightarrow 0$ as we only know the Dirichlet integral such $\int_0^{\infty} \left| \frac{\sin(x)}{x} \right| dx = \frac{\pi}{2} < \infty$ resulting that we can apply Fubini's theorem as previously :

$$\begin{aligned} F(x) - \frac{1}{2} &= \frac{1}{2\pi} \int_0^{\infty} \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)du}{iu} \\ F(x) &= \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)du}{iu} \end{aligned}$$

By the property of conjugates and symmetry of the characteristic function. The imaginary part of the integral will sum to 0. In addition, the real part is even and will then add up on both x -axis. For these reasons, we can say that :

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{iux}\phi(-u) - e^{-iux}\phi(u)du}{iu} \iff \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\mathbb{I}(e^{-iux}\phi(u))du}{u}$$

Bibliography

- [1] J. G. Wendel. The Non-Absolute Convergence of Gil-Pelaez' Inversion Integral. *The Annals of Mathematical Statistics*, 32(1):338 – 339, 1961. pages vi, 20, 83
- [2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973. pages 5
- [3] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976. pages 9
- [4] M.S. Torres and J.M.A. Figueiredo. Probability amplitude structure of fokker–plank equation. *Physica A: Statistical Mechanics and its Applications*, 329:68–80, 2003. pages 12
- [5] Bruno Dupire. Pricing with a smile. 1994. pages 13
- [6] Steven L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 2015. pages 16
- [7] D.S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, 9(1):69–107, 1996. pages 20
- [8] I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960. pages 21
- [9] Jim Gatheral. Perturbative analysis of volatility smile. *Presentation at the Columbia practitioners conference on the mathematics of finance, New York*, 2000. pages 25
- [10] Jim Gatheral. *The volatility surface, a practitioner's guide*. 2006. pages 25
- [11] Jim Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives Risk Management, Madrid*, 2004. pages 25
- [12] Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167–179, 1976. pages 27

- [13] Jim Gatheral and Antoine Jaquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, pages 14–16, 2013. pages 35, 36
- [14] Lorenzo Bergomi. *Stochastic volatility modeling*. Chapman & Hall/CRC financial mathematics series. CRC Press, Boca Raton, 2016. pages 37, 39
- [15] Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968. pages 42
- [16] Ryan McCrickerd and Mikko S. Pakkanen. Turbocharging monte carlo pricing for the rough bergomi model. *Quantitative Finance*, 18(11):1877–1886, Apr 2018. pages 44
- [17] CR Dietrich and Garry Neil Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4):1088–1107, 1997. pages 49
- [18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, pages 933–949, 2018. pages 51, 53, 55, 61
- [19] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016. pages 51
- [20] P. Lévy. Random functions: General theory with special references to laplacian random functions. *University of California Publications in Statistics*, pages 331–390, 1953. pages 52
- [21] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323, 1998. pages 55
- [22] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15:635–654, 2011. pages 58
- [23] D. Lien. Option valuation under stochastic volatility with mathematica code: Alan I. Lewis (2000). newport beach, california: Finance press. isbn: 0-9676732-0-1. *International Review of Economics Finance*, 11:331–333, 2002. pages 64
- [24] Peter Carr and Dilip B. Madan. Option valuation using the fast fourier transform, 1999. pages 64
- [25] Pospíšil J. Mrázek M. and Sobotka T. On calibration of stochastic and fractional stochastic volatility models. *European J. Oper. Res.* 254(3), pages 1036–1046, 2016. pages 66
- [26] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer, 2006. pages 68

- [27] Rainer Martin and Storn Kenneth Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11(4), pages 341–359, 1997. pages 69
- [28] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965. pages 69
- [29] Zeliade White. Quasi-explicit calibration of gatheral’s svi model. 2009. pages 73
- [30] Hsieh Cho-Jui Chang Kai-Wei Ringgaard Michael Lin Chih-Jen Chang, Yin-Wen. Training and testing low-degree polynomial data mappings via linear svm. *Journal of Machine Learning Research*, page 1471–1490. pages 77
- [31] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. pages 79