

UNIVERSITY OF LAUSANNE

DEPARTMENT OF FINANCE

Rough Volatility Modeling

Author:
Elyes Mahjoubi

Supervisor:
Dimitrios Karyampas

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

September 2021

Abstract

Financial mathematics began to appear at the beginning of the 20th century by Louis Bachelier. However, theories and models about options pricing only began to emerge after 1970, especially with the breakthrough of the Black Scholes (BS) model , developed by Fischer Black , Myron Scholes, and Robert C. Merton. Since then, research in the field of option pricing didn't stop evolving, with the goal of evaluating options accurately and efficiently. The principal flawed assumption of the BS model was that the volatility was assumed to be constant across strike prices and maturities. Moreover, in the BS model, it is assumed that log-returns are normally distributed. These assumptions were proven wrong by empirical studies. Indeed, the studies demonstrate the existence of a skew in the volatility term structure across strikes and maturities, and that log-equity returns' distribution is generally not normal, but rather leptokurtic and negatively skewed. To fix these flaws, in 1993, Steven L. Heston asserted in his paper that the volatility depends on a mean-reverting stochastic process, allowing to price options using a non-constant volatility. Meanwhile, a significant drawback of the Heston model is that it fails to fit the volatility skew of very short-term maturities correctly. Additionally, the model is less intuitive than the BS model, and accounts for many parameters, making the optimization hard and tedious. In 1996, David S.Bates introduced the stochastic volatility jump model, which aimed at helping fix the underestimation of the stochastic volatility model for the very short term maturity. Meanwhile, even if empirical results back the jump models, adding jumps increases the number of parameters and , then, the calibration time.A new generation of stochastic volatility models emerged in early 2010, developed and later termed by Jim Gatheral, Thibault Jaisson and Mathieu Rosenbaum as "Rough volatility models". These new models are backed by empirical studies demonstrating that the volatility dynamic is neither semi-Martingale, nor mean-reverting. Furthermore, the usefulness of the "rough volatility" models reside in their easiness to be understood for common traders.

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 Dimitrios Karyampas for providing invaluable insight, support, and guidance throughout the thesis's research and writing process. Furthermore, I would like to offer my gratitude to Süleyman Ceran for the support and assistance in 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	9
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	The Heston Model under the \mathbb{P} measure	15
5.1.2	The Heston Model under the \mathbb{Q} measure	15
5.1.3	The semi-closed solution of the Heston model	16
5.2	The Bates (SVJ) 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 motions	45
7.3.1	Fraction Brownian motion using Cholesky decomposition:	45

7.3.2	FBm using spectral decomposition :	46
7.3.3	FBm using efficient spectral decomposition :	46
8	The rough Bergomi model	52
8.1	The unconventional Bergomi model	52
8.2	Variance forecast	54
8.3	FSV or RFSV under the \mathbb{P} measure	56
8.4	Rough Bergomi model under the \mathbb{P} measure	56
8.5	Rough Bergomi under the \mathbb{Q} measure	57
8.6	Volatility skew	59
8.7	The smoothness of volatility log increments	60
8.8	Volatility skew of the rough Bergomi model	63
9	Calibration	65
9.1	Cumulative Distribution functions approximation for SV models : . . .	65
9.2	Calibration of the stochastic volatility models	66
9.3	Optimizations	69
9.4	SVI calibration	74
10	Interpolations	76
10.1	Multivariate normal distribution	76
10.2	Properties of multivariate normal distribution	77
10.3	Covariance matrix estimation	78
10.4	Gaussian Process Regression	79
11	Appendix	82
11.1	Black and Scholes	82
11.1.1	Call option price derivation	82
11.1.2	Greeks' derivation	83
11.2	Heston Model	84
11.2.1	Gil Pelaez Proof(based on the paper [1])	84

List of Figures

2.1 Empirical implied volatility	4
3.1 Simulation of a Delta coefficient of an option expiring the 29 October 2021 with an initial date, the 21 June 2021 using a risk free rate of 3%.	10
3.2 Monte Carlo simulation of stock prices following the dynamic 3.1	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, where: $r = 0, q = 0, V_{dt} = 0.09, dt = \frac{1}{252}$ 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, in the case of the Heston model	20
5.6 Heatmap of the AARE errors for the Bates model	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 16th April 2021.	35
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 The Koch curve is a self-similar and expresses regularity.	42
7.2 Fractional Brownian motions with different Hurst parameters	47
7.3 Covariance matrix with daily lags	49
7.4 Generated fractional Gaussian noises	51
8.1 Forecast and realized variance for $\Delta = \frac{1}{252}$ given different parameters ν	55
8.2 A log-moment comparison between the RFSV and FSV models	57

8.3	Monte Carlo simulation of the rough Bergomi model under the \mathbb{P} measure	58
8.4	Term structure of S&P500 at-the-money-forward volatility skew and its alternative fit	61
8.5	Log-volatility increments distribution, under different lags Δ	61
8.6	Linear fit between $\log m(q, \Delta)$ and qH	62
8.7	Log-moments of the log-volatility increments	62
8.8	Rough volatility skews	64
9.1	ϑ weights	68
9.2	IVMSE scatter-plot for the calibration of a simple Heston surface	69
9.3	IVMSE scatter-plot for the calibration of a surface SVI	69
9.4	SVJ surface generated by a Gaussian process interpolation	72
9.5	IVMSE errors of the SSVI surface	73
11.1	Heaviside Function	85

List of Tables

8.1	Comparison of norm-errors between the realized and the forecasted variance for a slate of parameters ν	55
8.2	Result of the rough Bergomi model calibration using a differential evolution algorithm	63
9.1	Differences in the execution time between the Gil Peleaz integral and the fast Fourier algorithm	66
9.2	Difference in the accuracy of the Gil Peleaz integral and the fast Fourier algorithm	66
9.3	SSVI calibration results	69
9.4	Simple Heston calibration results	69
9.5	SVJ Heston calibration results	72

Chapter 1

Introduction

The stochastic volatility models enhanced in the past show important limitations: unfamiliar, not tractable, and hard to calibrate for an average trader. Today, the financial industry needs simpler and tractable models that reflect the actual dynamic of the volatility. The rough volatility models perfectly fit that role. Indeed, these models contain very few intuitive parameters. Throughout the thesis, we will first recall the properties of the main past option pricing models. In addition, we will show what their main flaws and advantages are for each of these models. Finally, we will see why the 'rough' volatility models are among the most realistic models, replicating the term structures very accurately. To deliver complete insights into the advantages and drawbacks of the "Rough" volatility modeling in option pricing, we will lean on the "Rough Bergomi" model, which intends to replicate option prices and their dynamics. In this regard, the main objective of the thesis is to compare the newly developed volatility model to the old ones.

Chapter 2

Preliminaries

2.1 Keywords and definitions

We used specific appellations during the whole thesis 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
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 $\log(K)$
$\sigma_{IV}(\tau, s, k)$	Implied volatility function
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 at the stock market opening on the 16th of April 2021 . We choose to study the call options of the S&P500 ,as it is considered as one of the primary benchmarks 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%

Throughout the thesis:

- The risk-free rate r corresponds to a 3-month treasury bill rate.
- the underlying price S_0 is the one of the S&P500 at the opening of the trading session, the April 16,2021.

A Gaussian interpolation of the volatility surface gives us the following surface:

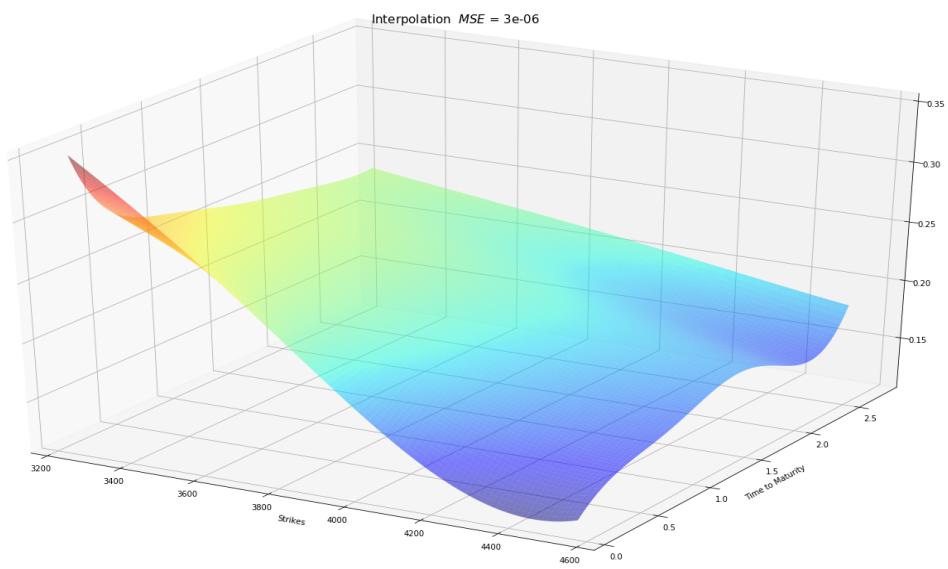


Figure 2.1: Empirical implied volatility

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 option pricing model considers an economy with two assets: the available underlying and a zero-coupon bond. The originality of the Black and Scholes model [2] is that it allows for the payoff of an option to be replicated using a basket of two available assets:

Assumption 1. *Before telling more about the hedging techniques, we need to cite the most critical 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 arbitrages).*
- *There are no transaction fees in buying the option.*
- *The risk-free rate and the underlying volatility are known and constant.*
- *The returns on the underlying asset are log-normally distributed.*

We write the option value of a call under the BS theory:

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

Where:

- S_t is the spot price of the underlying.
- Σ is the quantity of bonds sold.
- Δ is the quantity of underlying bought.

- S_t is the spot price of the underlying.

We use this relation to derive the price of an option with $Ke^{-r(T-t)}$, symbolizing the price of a zero-coupon bond. While, $\Delta = N(d_1)$ is the probability for the underlying to exceed the strikes price K at maturity with a basis of e^z such $z = \log \frac{K}{S_t}$, $\Sigma = -N(d_2)$ is the very same probability with a base of 1. Although the Black and Scholes assumptions don't seem very realistic, they allowed far more advanced models and hedging strategies to be developed.

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 is 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 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 .*

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

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 $0 = t_0 < t_1 < \dots < t_m$ with increments defined such:*

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

These increments are normally distributed with expectation and variance:

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

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. A Martingale means is that the future change is expected to be 0, having all information until 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 a simple process 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 (SDEs) is represented by:

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

Where:

- μ is the drift term.

- σ the standard deviation of the process.
- W_t a standard Brownian motion.

Using a Taylor Approximation along with the Ito's lemma, we find:

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

Where:

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

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

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

Where:

- r is 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 $C(S_t, K, t, T)^1 \forall T \geq t$:

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

Where:

- Φ is 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.

¹Derivations are available on the appendix

3.2.1 The Greeks

The Greeks are very used tools in finance. Generally, traders use them to hedge and adjust their positions in risky and safe assets, limiting potential 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	$\frac{\partial V}{\partial \sigma}$

Where:

- Δ is the sensibility of the option's value to a change in the underlying price.
- Γ is the sensibility of the option's value to the change of the change the price in the underlying price.
- ρ is the sensibility of the option's value to a change in the interest rate.
- Θ is the sensibility of the option's value to a change in the volatility of the underlying.
- ϑ is the sensibility of the option's value to the time to maturity.

In the appendix, we attached the derivation of these coefficients in the case of a call option. An example of the evolution of the coefficient Δ according to the movement of the asset price S_t is shown in the figure 9.1.

3.3 The Merton Jump Model

Jump stochastic models were initially created by Robert C. Merton [3]. The main goal of adding jumps to the dynamic of the underlying price is to make assets movements discontinuous. It is empirically shown that asset prices are not continuous, and thus, we can justify the addition of a Poisson process in the following SDEs:

The modified SDEs takes the form of:

$$dS_t = \mu S_t dt + \sigma S_t dW_t^{\mathbb{P},1} + S_t dJ_t \quad (3.1)$$

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

Where:

- $dN(t)$ is a Poisson process distributed with a probability of λdt .

- S_{t^-} is the price before a jump happen.
- H_i is the log-normal intensity of the jump.

After an appropriate twice differentiation using a Taylor serie, we obtain a smooth function:

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

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

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

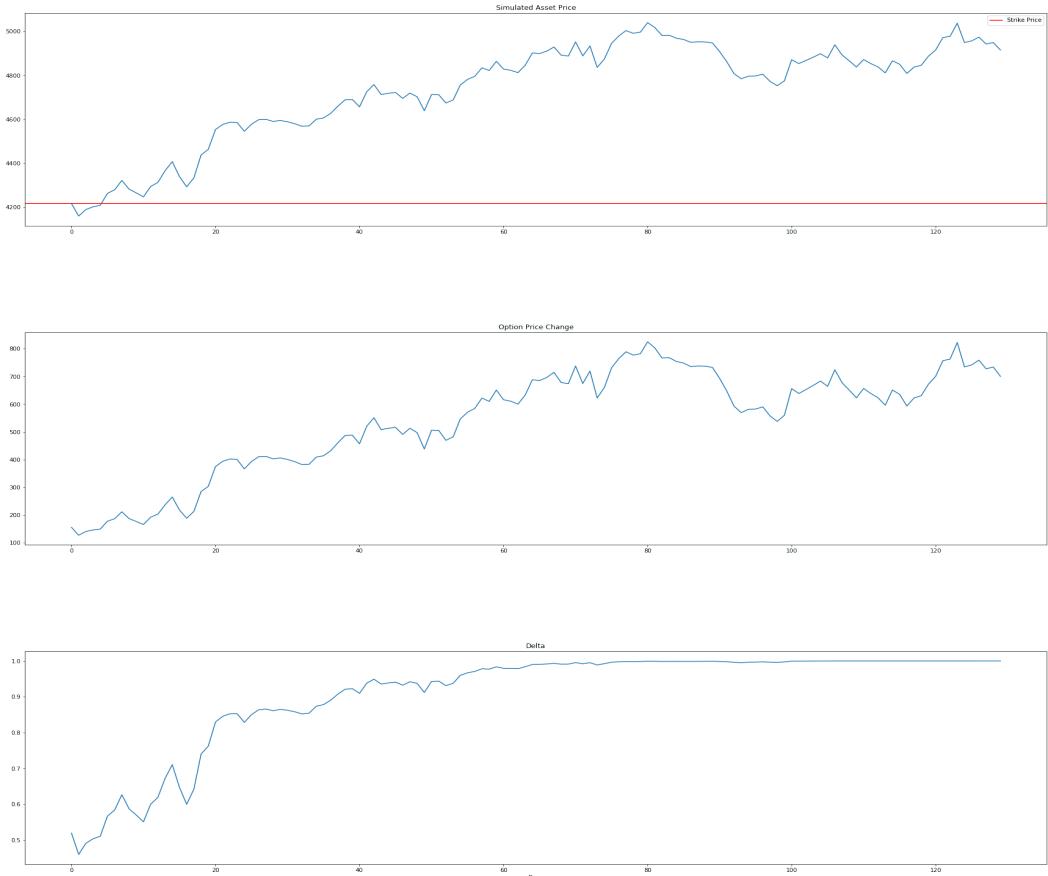


Figure 3.1: Simulation of a Delta coefficient of an option expiring the 29 October 2021 with an initial date, the 21 June 2021 using a risk free rate of 3%.

In a risk-neutral world, we obtain:

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

This Yields to a deformation of the Black and Scholes formula:

$$C = \sum_{n=0}^{\infty} \frac{e^{-\lambda T} (\lambda T)^n}{n!} BS(S_0, \sigma_n, r_n, T, K) \quad (3.2)$$

With :

$$\sigma_n = \sqrt{\sigma^2 + n\nu/T} \quad (3.3)$$

$$r_n = r - \lambda(e^{m+\frac{1}{2}\nu} - 1) + \frac{1}{T}(\frac{n\nu}{2} + nm) \quad (3.4)$$

Knowing that (3.3) and (3.4) are the transformed drift and standard deviation of the asset price, we can derive the non-actualized option price C (3.2). To conclude, the Merton-jump model was a good try to fix the flaws of having a constant volatility and a non-leptokurtic distribution in the Black and Scholes model.

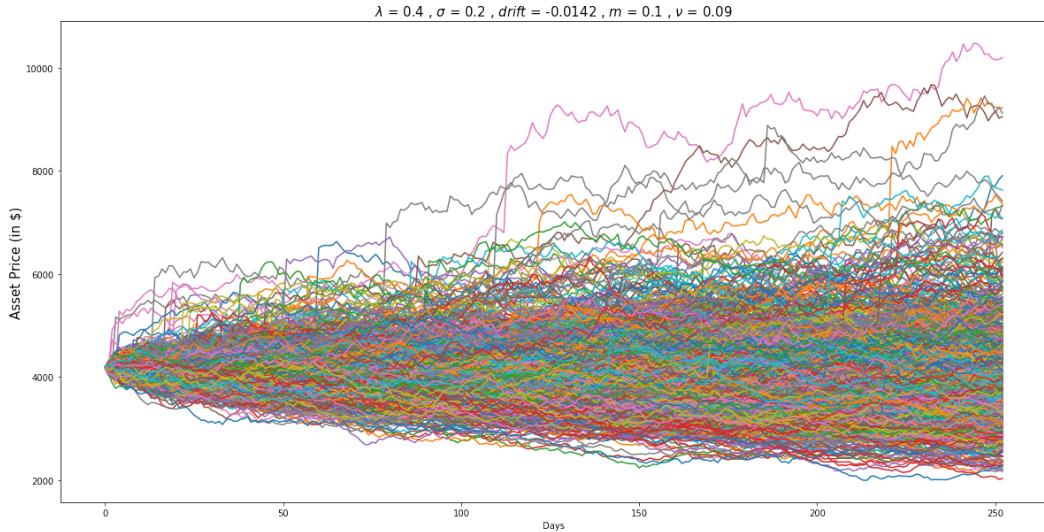


Figure 3.2: Monte Carlo simulation of stock prices following the dynamic 3.1

Chapter 4

Local Volatility Model

Bruno Dupire, in his research paper, suggested the volatility could be computed directly from a combination of strike and maturity. Dupire ingeniously used the Fokker-Planck equation, which describes the dynamics of a known distribution.

4.1 The Fokker Plank equation

We define a probability space $(\Omega, \mathcal{F}_t, \mathbb{P}) \forall t \in \mathbb{R}_+$, where $p \exists \forall t \in \mathbb{R}$. The dynamic of a random variable $W_t \in \mathbb{R}$ is ascribed in the physical probability space:

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

However, as the dividend yield rate q is introduced, working in a risk-neutral environment would require the use of the forward F_t such :

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

Where:

- $F_t(S_t, \tau) = S_t e^{(r-q)\tau}$
- τ is the time to maturity.

Because we want to work in a risk-neutral \mathbb{Q} space, we will use the forward underlying price, defining the dynamic of our probability density function such $p : \mathbb{R} \rightarrow \mathbb{R}_+$:

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

A proof of 4.1 is given on the article [4]. Now, we define $D(S_t, t) = \sigma^2 F_t^2$ as the variance of our distribution. Hence, a call is expressed such:

$$C(S_t, K, T, t) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(F_t - K)^+]$$

$$\begin{aligned}
&= e^{-r(T-t)} \int_K^\infty (F_t - K) p dS \\
&= e^{-r(T-t)} \int_K^\infty F_t p dF - K e^{-r(T-t)} \int_K^\infty p dF
\end{aligned}$$

Setting $\tau = T - t$, the precedent formula equation becomes:

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

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

4.2 Computation of the Local Volatility

We now have a direct relationship between the volatility, the strikes and the time to maturity:

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

Additionally, for an implementation purpose, we can derive the variance using the formulas below:

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

An exemplified interpolated surface coming from the local volatility model is shown in the figure 4.1. As we can compare with 2.1, the main advantage of the local volatility models is that we can fit the data extremely well without pain. However, its main drawback locates in the fact that its dynamic is wrong and 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.

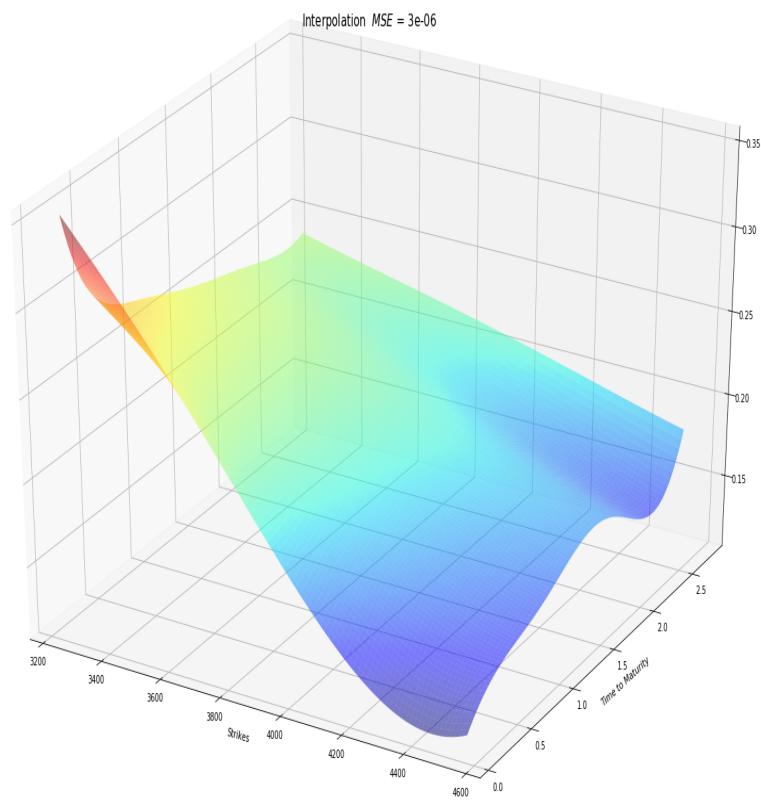


Figure 4.1: Local volatility surface of the S&P500, the 16th April 2021

Chapter 5

Stochastic Volatility Models

5.1 The Heston Model

5.1.1 The Heston Model under the \mathbb{P} measure

Empirically it has been shown that the volatility tends to be mean revertible. Then, a better approach to price options would be to model the volatility using an Ornstein-Uhlenbeck (OU) process, describing the dynamic of the variance in a probabilistic space $(\Omega, \mathcal{F}_t, \mathbb{P})$:

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

Where:

- κ is the mean reverting speed of the variance.
- θ is the long-term variance.
- ξ is the volatility of the variance.

A particularity of this model is that it can be multifactorial. Indeed, the volatility process and the stock prices tend to be correlated but not monofactorial, 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 The Heston Model under the \mathbb{Q} measure

To keep our κ or θ fixed, we set the market premium of the volatility to 0. Thus, our risk-neutral 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 solution of the Heston model

In 1993, Steve L.Heston [6] developed a semi-closed formula to price European options, for $[S_t, V_t] \in [\mathbb{R}_+, \mathbb{R}_+]$. Meanwhile, to derive the semi-closed formula, we need to use a Taylor serie approximation and the Ito's lemma:

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

Applying a change of variable:

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

In the meantime, 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:

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

However, to get our result, we first need to rewrite the formula of an option value:

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

Substituting $P(x_t, V_t, \tau)$ by 5.2:

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

Where :

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

Remark 3. A tip to pass from 5.1 to 5.3 is to express D_1 using $K = 0, S_t = 1$, and D_0 using $K = -1, S_t = 0$.

Where:

- $D_1(x_t, V_t, \tau)$ is the probability of exercise of the option under the base e^{x_T} .
- $D_1(x_t, V_t, \tau)$ is the probability of exercise of the option under the base e^{x_T} .
- $D_0(x_t, V_t, \tau)$ is the natural probability of exercise of the option.
- $D_j \in \mathbb{R}_+ \forall t \in \mathbb{R}_+$.

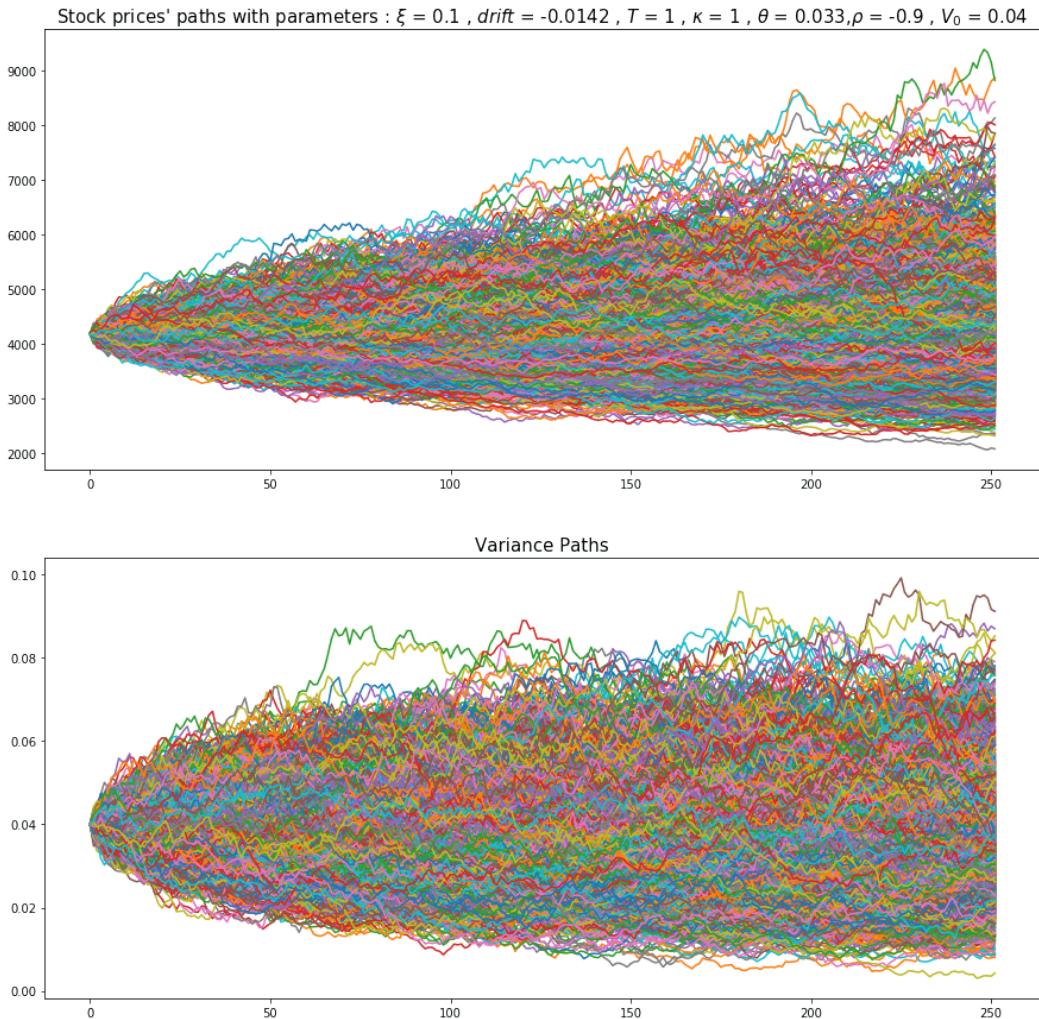


Figure 5.1: Monte Carlo Simulation of stock and variance paths using a Stochastic Volatility Model

Convolution While several distributions don't have a probability distribution function, many possess a characteristic function. In this regard, the convolution property of characteristic functions allows us to write a characteristic function of a sum of two continuous random i.i.d variables A and B such:

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

Where:

- ϕ_A and ϕ_B are two characteristic functions.
- A and B are Lebesgue measurable in \mathbb{R} .
- $\phi_i: \mathbb{R} \rightarrow \mathbb{C}$.

The essential properties of the characteristic functions are:

- $\phi(0) = 1$.
- ϕ is uniformly continuous.
- $\phi(t)$ is Hermitian: $\phi(t) = \overline{\phi(-t)}$.
- $|\phi(t)|$ is bounded such as $|\phi(t)| < 1$.
- The real part of ϕ is even and the imaginary part is odd.

Subsequently, it is natural to use the inverse Fourier transform to retrieve the cumulative distribution function. To do so, we transform our density function written $d_j(x_t, V_t, \tau) = \partial_{x_t} D_j(x_t, V_t, \tau)$ into a signal through our Fourier transform:

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

By Applying the inverse Fourier transform, we retrieve the original probability density function:

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

By integrating both side over x_t , we finally get our cumulative distribution function:

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

The obtained result is Lebesgue integrable and satisfies 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 $\int_{R_1 \times R_2} \left| e^{-iux_t} \frac{\tilde{d}_j(u, V_t, \tau)}{2\pi} \right| d(x_t, u) < \infty$. Moreover, the presented figure 5.2 gives the intuition why the measure is Lebesgue integrable. This property allows us to apply the Fubini's theorem:

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

To continue the derivation of the semi-closed European options formula, we need $D_j(x_t, V_t, \tau)$ to be tractable to compute its derivatives. For that reason, we assume that $\tilde{D}_j(x_t, V_t, \tau)$ has an 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:

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

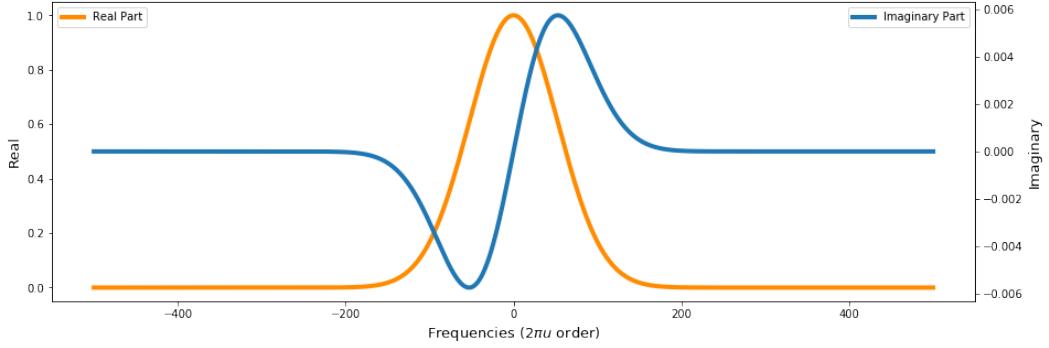


Figure 5.2: Visualization of the characteristic function of a stock price S_{dt} following a log-normal distribution, where: $r = 0$, $q = 0$, $V_{dt} = 0.09$, $dt = \frac{1}{252}$ and $S_0 = 4360.30$.

We define:

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

Through the derivation of 5.6, we get:

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

Finally, plugging our results from 5.7 in 5.6, leads to:

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

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

Upon obtaining a Riccati equation, the solutions of our system become:

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

with :

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

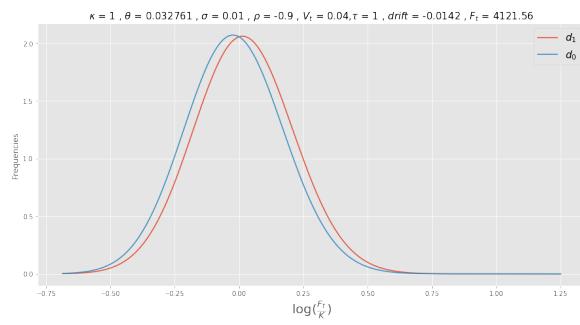


Figure 5.3: Heston probability density functions for d_1 and d_2

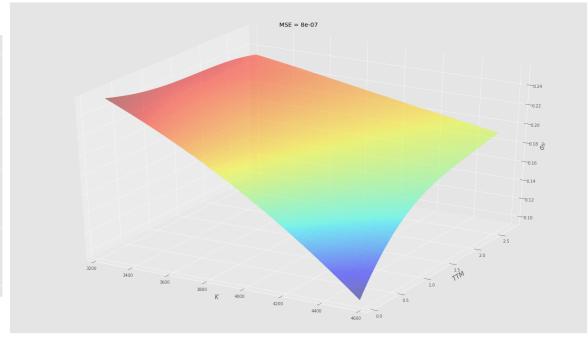


Figure 5.4: Stochastic volatility surface of a simple Heston model

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 [1].

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

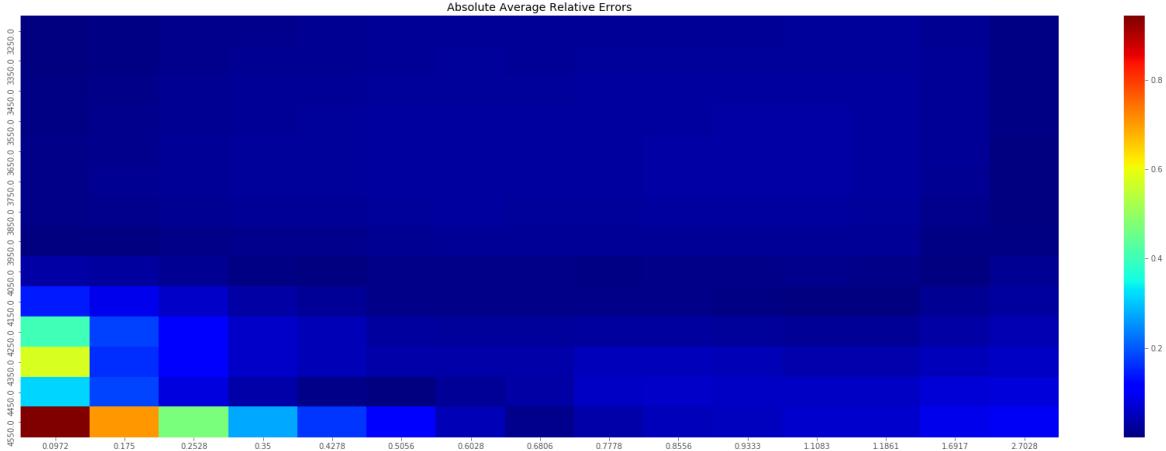


Figure 5.5: Heatmap of the AARE errors, in the case of the Heston model

5.2 The Bates (SVJ) 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 adding jumps to the natural Standard differential equation of the price,

explaining the occurrence of crisis or major events (Black Monday in 1987, Lehman Brothers collapse in 2008, Coronavirus pandemic in 2020). In order to implement such models, we set a new standard differential equation with the same physical probability space defined previously:

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

By rewriting our Taylor serie, we include our observable discrete jumps:

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

However we are still in the physical risk world. Therefore, to translate in the risk-neutral world \mathbb{Q} , we apply the Girsanov theorem [8] and our new risk-neutral SDEs become :

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

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 sudddenly 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 $X_1 + X_2 + \dots + X_{k+1}$ 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 :

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

We can notice that the significant change to evaluate the price of an option under the SVJ Heston resides in adding poisson terms in the SDE. As a result , by reproducing the steps in 5.1 and adding the discrete term expressed in 5.11, we find out that the characteristic function of the Heston Model becomes equal to 5.13:

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

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

$$\widehat{\Xi}_{\lambda}(u, \tau) = \mathbb{E}^{\mathbb{Q}}(E(e^{iu \sum_{i=1}^{N_t} \log(H_i)} | N_t = k)) \iff \sum_{k=1}^{\infty} E(e^{iu \log(H_i)} | N_t = k)^k P(N_t = k) \\ = \sum_{k=1}^{\infty} \frac{e^{-\lambda\tau}}{k!} \left(\lambda\tau \int_{-\infty}^{\infty} e^{iu \log(H_i)} p(\log(H_i)) d\log(H_i) \right)^k$$

Set $j = \log(H_i)$

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

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

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

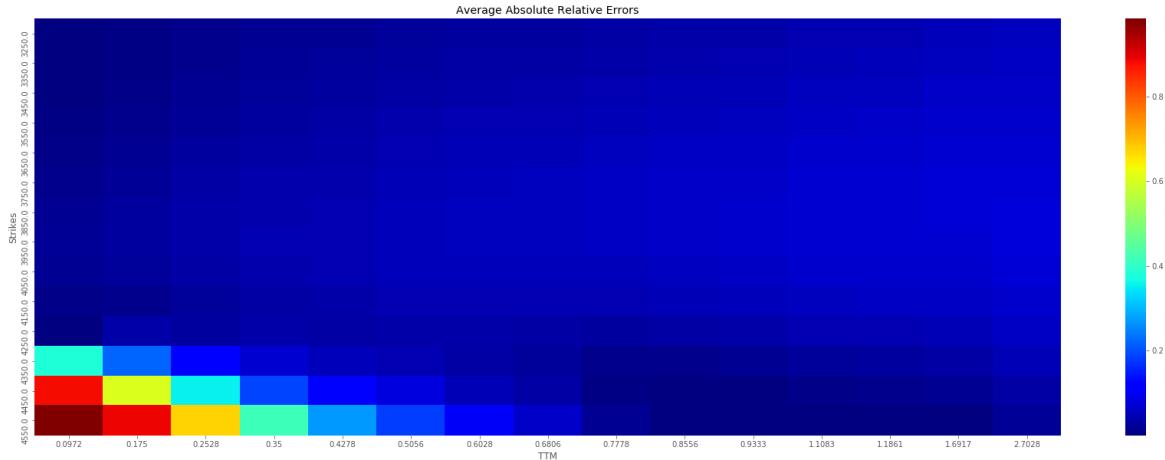


Figure 5.6: Heatmap of the AARE errors for the Bates model

5.3 The SVJJ Model

The Heston stochastic model with jumps allows us to replicate accurately the volatility skew of options available in the market for very short term maturities. The model was initially set to correct a major drawback of Bate's model, where the correlation between the volatility and the log stock price is assumed to be 0. In other words, the SVJJ model assumes that a shock on the price won't have any consequences on the volatility of the underlying. Meanwhile appreciable, it is empirically untrue as shown on the figure 5.7. To implement the SVJJ model, we ,first, need to reset our main SDEs such they transform to :

$$\begin{aligned} dS_t &= (r - q)S_t dt + \sigma S_t dW_t^{\mathbb{P},1} + S_t dJ_t^1 \\ dV_t &= \kappa(\theta - V_t) dt + \xi \sqrt{V_t} dW_t^{\mathbb{P},2} + dJ_t^2 \\ dJ_t^1 &= \prod_{i=1}^{dN_t} H_i - 1, \mathbb{E}(dW_t^{\mathbb{P},1} dW_t^{\mathbb{P},2}) = \rho dt, dJ_t^2 = d\left(\sum_{i=1}^{N_t} B_i\right) \end{aligned}$$



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

$$\log(H_i) \sim N(m, \nu), N_t \sim \text{Poisson}(\lambda t), B_i \sim \exp\left(\frac{1}{\gamma_v}\right)$$

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

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

As usual we translate our SDEs into a risk-neutral world by applying the Girsanov theorem, we obtain :

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

Where κ' and θ' are risk-neutral adjusted parameters. Afterward, we apply our change of variable such: $x_t = \log \frac{F_t(S_t, \tau)}{K}$. Rewriting the option price C , as showed in 5.3 and taking the expectation of the obtained PDE, leads 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] links the jump endured by the stock price to the one experienced by the volatility. To make it quick, a jump in the stock

price simultaneously provokes a jump in volatility. This It also allows the option price to be derived, I such :

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

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

Where $I(u, \tau) = \frac{1}{T} \int_0^\tau e^{\gamma_v L(u, t)} dt$. In the meantime, the interpretation of the presented formulas is that a jump in the stock price will provoke instantaneously a leap in the variance, expressed by $V_t = V_{t-} + \gamma_v$. However, multiple papers (On of them is available at [10]) show that the SVJJ model doesn't improve substantially the volatility skew by the simple SVJ model. In reverse, the model extension adds more time to calibrate, increasing the difficulty 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 a for stochastic implied volatility. It aims to provide a simple function that calibrates the implied volatility following our tenor. Jim Gatheral had disclosed the first SVI model in [11]. Its main advantage resides in its easy-to-use application for traders, as only the parameters to plug need to be remembered. The SVI model exists for a probability space $(\Omega, \mathcal{F}_t, P)$, where F is the possible σ -algebra sets available at time t , $\forall t \in \mathbb{R}_+$. Indeed, the SVI 'Raw' model is expressed 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 a volatility skew fitter than a market price replicator. Meaning that the originality of the parametrization comes from its dependency between τ and the total implied variance, expressed as $w_{IV} = \sigma(K, \tau)_{IV}^2 \tau$. Moreover, for the SVI to be free of arbitrage opportunities so as to avoid free trading gains and negative option prices, it requires to, primarily, operate a variable change on the option price defined as :

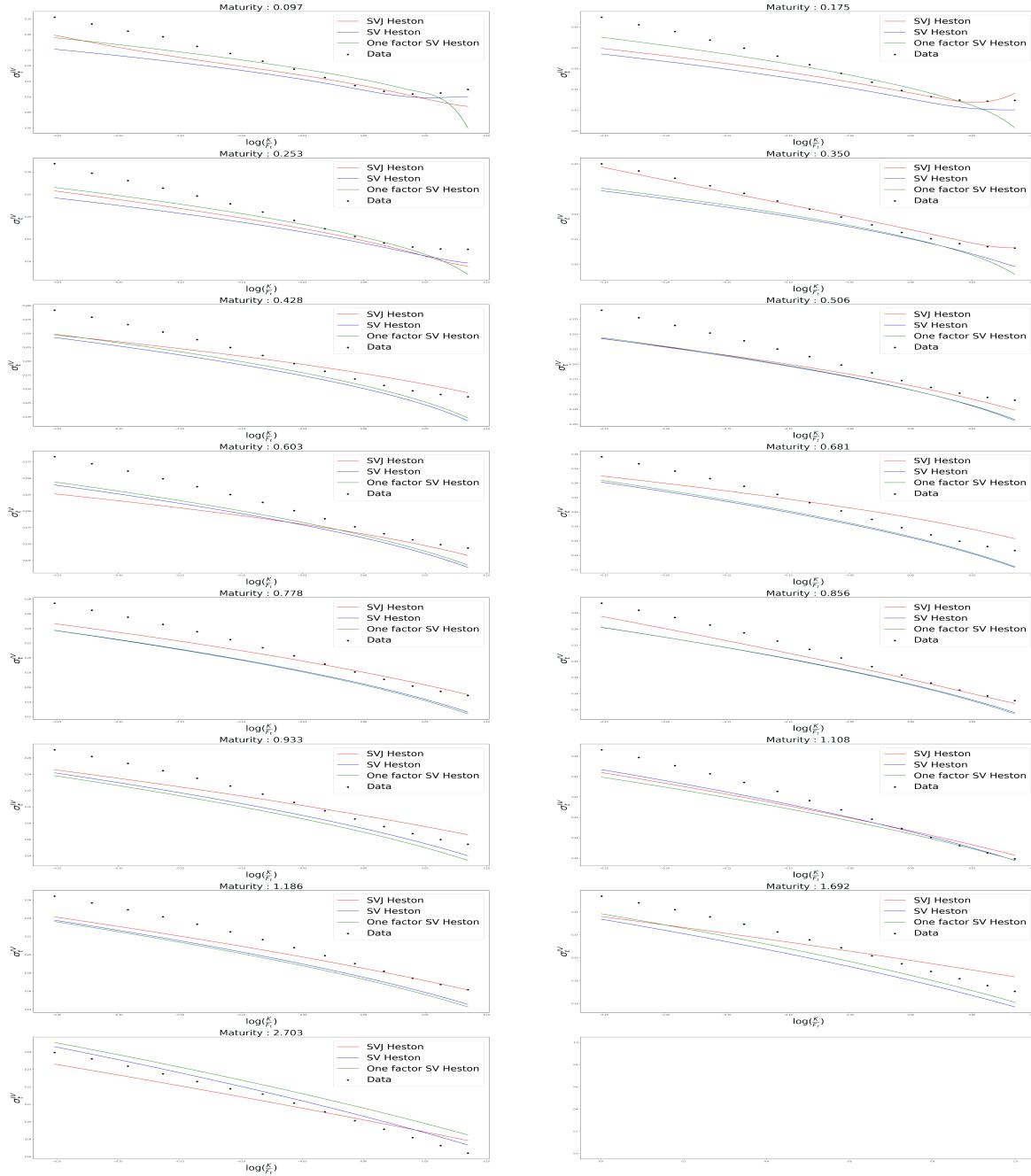


Figure 5.8: Stochastic volatility model skews, the one factor Heston is created using a constrained argument such $\rho = -1$

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

Where:

- $x = \log \frac{K}{F_0}$.

Introducing the total implied variance into the Black Scholes equation, we need to rewrite the terms d_1 and d_2 in a way Black did in 1976 on his paper [12].

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

Resultantly, by applying the chain rule we shall easily find the first two required derivatives, such as:

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

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

Lastly, the third derivative is straightforward to derive:

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

Thus, substituting 5.15, 5.16, 5.17 in the original Dupire's formula gives the subsequent formula :

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

However, we still need to express the local volatility in terms of the total variance w . This is achievable by doing some derivations to express $\frac{\partial^2 C_{BS}}{\partial x^2}$ in terms of total implied volatility.

The first derivative is expressed as :

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

The second derivative is expressed as :

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

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

The third derivative is expressed as :

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

Finally, by replacing all our derivatives in the initial formula, we find 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 derive the last two most needed derivatives:

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

Continuing through the calculations, we end-up with a deterministic formula to price options such as proposed by Dupire 5.4:

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

By factoring and removing $\frac{\partial C}{\partial w}$ on both side of the equation, we reformulate our equation 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, the variance 5.18 becomes local:

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

Needing our local variance to be positive to eliminate the risks of having a calendar spreads and a butterfly arbitragers, we implement some constraints defined as:

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

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

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

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

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

At this stage, one might ask why we are talking about local volatility in the stochastic volatility chapter. A simple answer for that is that local volatility models originally inspired the current stochastic one.

What also makes the SVI model suitable, compared to stochastic volatility models, is the existence of the constrain 5.19, precluding traders from taking advantage of calendar spreads. Moreover, This condition is easily verified and very intuitive to grasp for traders. However, another constrain, in 5.20, is suitable as it restricts the probability density function $\frac{\partial^2 C}{\partial K^2}$ from being negative, prevailing negative volatility premiums to exist. In this regard, the latter condition is sometimes called the Durleman's condition.

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

$$\lim_{x \rightarrow \infty} d_2 = -\frac{x}{\sqrt{w}} - \frac{\sqrt{w}}{2}$$

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

However, even if the SVI allows to have good fits easily and rapidly. A central problem persists : the 'raw' SVI is not intuitive for traders. Indeed, The parameters wear mathematical understandings because the whole process is centered around total implied variance w , an unfamiliar metric for traders. Meanwhile, a reparametrization, called SVI-JW, makes a lot more sense for traders.

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

The main 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 :

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

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

$$\begin{aligned}
- \lim_{x \rightarrow -\infty} \sqrt{\frac{w}{x\tau}} &= \lim_{x \rightarrow -\infty} \sqrt{\frac{\frac{1}{x} \left(a + b(\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}) \right)}{\tau}} \\
&= \lim_{x \rightarrow -\infty} \sqrt{\frac{1}{\tau} \left(\frac{a}{x} + b(\rho(1 + \frac{m}{x}) + \sqrt{(1 + \frac{m}{x})^2 + \frac{\sigma^2}{x^2}}) \right)}
\end{aligned}$$

We can notice for $x < 0$, $(1 + \frac{m}{x}) = (-1 - \frac{m}{|x|})$, then:

$$= \sqrt{\frac{1}{\tau} \left(b(1 - \rho) \right)} = \sqrt{\frac{p\sqrt{w}}{\tau}}$$

$$p = \frac{b}{\sqrt{w}} (1 - \rho)$$

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

$$- \lim_{x \rightarrow \infty} \sqrt{\frac{w}{x\tau}} = \lim_{x \rightarrow -\infty} \sqrt{\frac{\frac{1}{x} \left(a + b(\rho(x-m) + \sqrt{(x-m)^2 + \sigma^2}) \right)}{\tau}}$$

$$= \lim_{x \rightarrow \infty} \sqrt{\frac{\frac{1}{x} \left(a + b(\rho(1 - \frac{m}{x}) + \sqrt{(1 - \frac{m}{x})^2 + \frac{\sigma^2}{x^2}}) \right)}{\tau}}$$

We can notice for $x > 0$, $(1 - \frac{m}{x}) = (1 - \frac{m}{|x|})$, then:

$$= \sqrt{\frac{1}{\tau} \left(b(1 + \rho) \right)} = \sqrt{\frac{c\sqrt{w}}{\tau}}$$

$$c = \frac{b}{\sqrt{w}} (1 + \rho)$$

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

$$- 0 = \frac{(x^* - m)}{\sqrt{(x^* - m)^2 + \sigma^2}} + \rho$$

$$(x^* - m)^2 = \frac{\sigma^2 \rho^2}{1 - \rho^2}$$

$$x^* = -\frac{\sigma \rho}{\sqrt{1 - \rho^2}} + m$$

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

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

In addition to the intuitive parameters, the provided fit are as good as using the 'raw' model. However, to be 100% sure that our SVI is free of arbitrage Gatheral and Jaquier re-derived the minimum implied volatility and the deep out-of-the-money call's slope :

$$c' = p + 2\psi$$

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

In the figure 5.9, we can see that the left tail on the Durrleman's condition, $g(k)$, the very short maturity appears to stabilize very quickly using the arbitrage-free SVI in contrary to the unconstrained SVI. Later in time, Gatheral and Jaquier developed the SSVI, called the Surface SVI, model. The specificity of the latter 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, starting 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 :

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

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

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

Therefore, using analogous calculation previously, the SSVI-JW parameters are exemplified as :

$$v = \frac{\theta_\tau}{\tau}$$

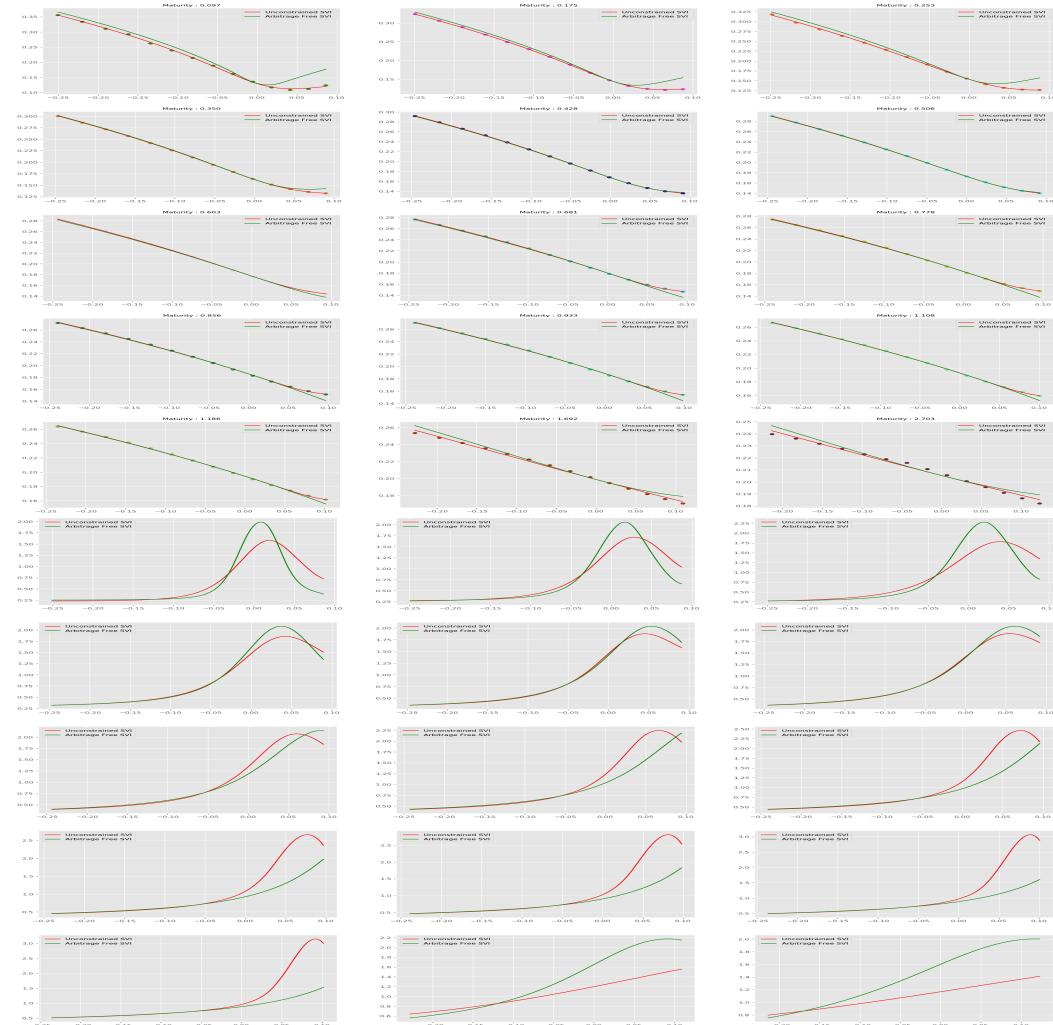


Figure 5.9: Stochastic volatility "inspired" of the S&P500, the 16th April 2021 and its Durrleman's condition g)

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

To express a volatility surface without calendar spread and butterfly arbitrage, we need to implement the two following conditions:

First, the SSVI surface is free of calendar-spread arbitrage if and only if:

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

Interestingly, we reformulated only the original condition 5.4 using the chain rule, as defined:

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

Proof of 5.23.

$$\begin{aligned}\frac{\partial w}{\partial\theta_\tau} &= \frac{1}{2} \left(1 + \rho\phi(\theta_\tau)x + \sqrt{(\phi(\theta_\tau)x)^2 + 2\rho\phi(\theta_\tau)x + 1} + \frac{\partial\phi(\theta_\tau)}{\partial\theta_\tau}\rho x\theta_\tau \right. \\ &\quad \left. + x\frac{\partial\phi(\theta_\tau)}{\partial\theta_\tau} \frac{\theta_\tau(\phi(\theta_\tau)x + \rho)}{\sqrt{(\phi(\theta_\tau)x)^2 + 2\rho\phi(\theta_\tau)x + 1}} \right) \\ \text{we can rewrite } \theta_\tau \frac{\partial\phi(\theta_\tau)}{\partial\theta_\tau} &\text{ as being equal to : } \frac{\partial\theta_\tau\phi(\theta_\tau)}{\partial\theta_\tau} - \phi(\theta_\tau) \text{ and } k = x\phi(\theta_\tau) \\ &= \frac{1}{2} \left(1 + k\rho + \sqrt{k^2 + 2k\rho + 1} + \left(\frac{\partial\theta_\tau\phi(\theta_\tau)}{\partial\theta_\tau} \frac{x}{\phi(\theta_\tau)} - x \right) \left(\frac{x + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right) \\ &= \frac{1}{2} \left(1 + k\rho + \frac{k^2 + 2\rho k + 1 - k^2 - k\rho}{\sqrt{k^2 + 2k\rho + 1}} - k\rho + k \frac{\partial\theta_\tau\phi(\theta_\tau)}{\partial\theta_\tau} \frac{1}{\phi(\theta_\tau)} \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right) \\ &= \frac{1}{2} \left(1 + \frac{k\rho + 1}{\sqrt{k^2 + 2k\rho + 1}} + \frac{\partial\theta_\tau\phi(\theta_\tau)}{\partial\theta_\tau} \frac{k}{\phi(\theta_\tau)} \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right) \right)\end{aligned}\tag{5.24}$$

obtaining a linear derivative ,in 5.24, that can be expressed as:

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

With :

$$\psi_0(k, \rho) = 1 + \frac{k\rho + 1}{\sqrt{k^2 + 2k\rho + 1}}$$

$$\psi_1(k, \rho) = k \left(\frac{k + \rho}{\sqrt{k^2 + 2k\rho + 1}} + \rho \right)$$

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

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

We know that the square root denominator cannot be negative and only the numerator could. Thus, to show that the numerator is strictly positive, we need to use a valid identity such:

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

Resultantly:

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

As we known that $\psi_1(k, \rho)$ is strictly positively defined on the intervals \mathcal{M}_ρ , then:

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

Here, the intervals are straightforward to find because we are in presence of a linear and a quadratic equation :

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

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 :

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

□

The supremum for positive $\psi_1(x, \rho)$ is equal to 0, meaning that the function keeps tending towards 0 without reaching it for $|x| \rightarrow \infty$. Finally, the last two constraints made to avoid the SSVI to have butterfly arbitrages are listed such :

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

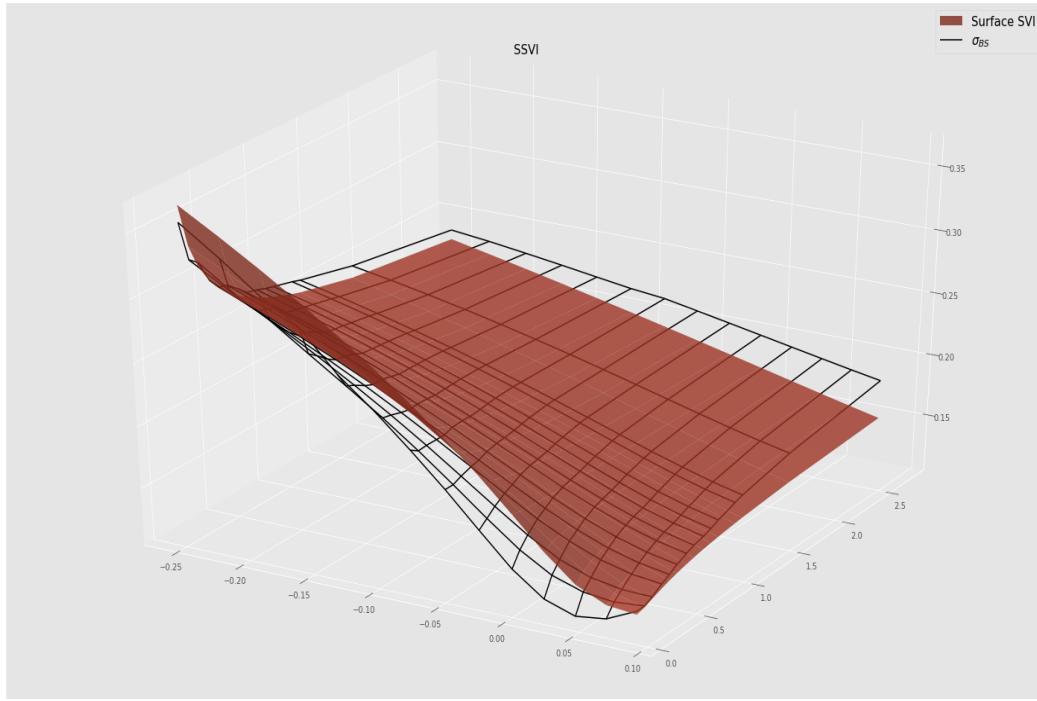


Figure 5.10: Comparison between SSVI of the S&P500 and the market's implied volatility surface ,the 16th April 2021.

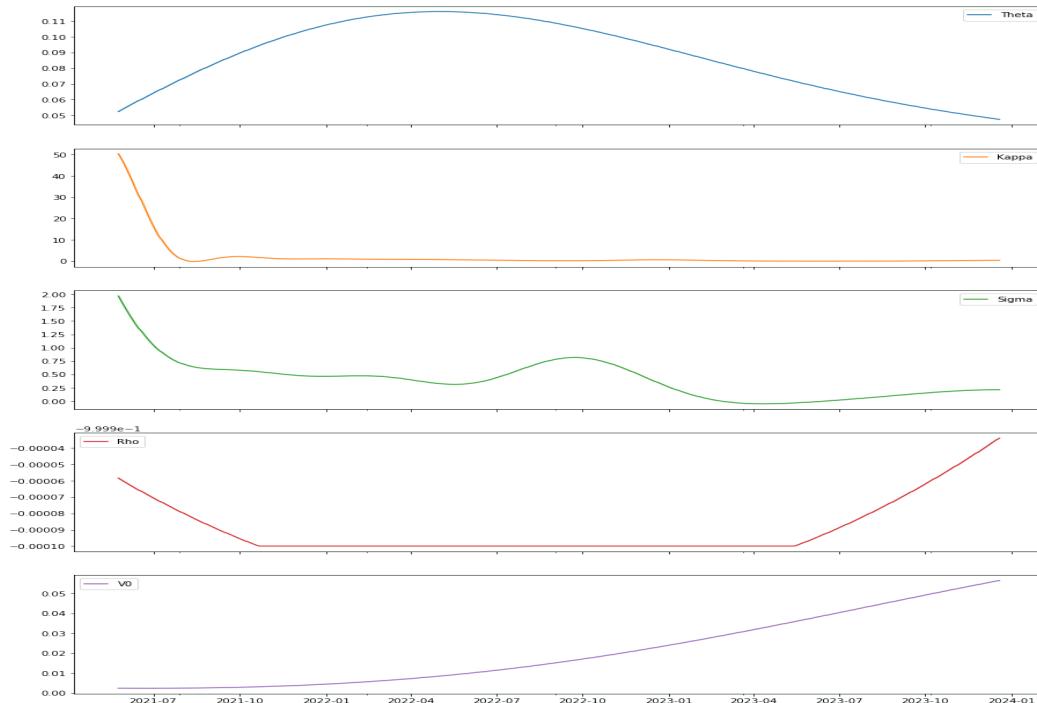


Figure 5.11: Heston Parameters across maturities ,the 16th April 2021.

These two conditions are obtained using Durrleman's condition 5.20. Proofs are voluntarily omitted and available at [13]. Meanwhile, the SSVI and SVI with free

of static arbitrages are inspired initially from the Heston stochastic volatility model, which fails to accommodate the extreme tails of returns' by treating the volatility as a diffusion process, which is wrong.

Finally, stochastic volatility models have strength because their parameters have a clear economic meaning. For example, across maturities, we can determine the anticipation of the market presented in the figure 5.11. Nevertheless, as cited previously stochastic volatility models fail to reproduce accurate options' term structures and are hard to fit to market prices because of the high number of parameters. Additionally, the very short maturity term structure is hard to reproduce using a basic stochastic volatility model without incorporated jumps.

Indeed, a significant drawback of adding jumps to an SV model is that it makes the calibration harder due to the presence of more parameters. Finally, the Stochastic Volatility 'Inspired' model is easy to calibrate and can be very intuitive to understand for traders (SV-JW model). Although, in SVI models, prices are not upheld to static arbitrage opportunities, the resultant volatility surface doesn't give a realistic dynamic insight. In this regard, it is not surprising that the SVI dynamic is not very relevant in evaluating volatility surfaces because Jim Gatheral showed in his paper [13], that the SVI is just a limit case of a standard Heston model.

Chapter 6

Bergomi's Model

The Bergomi model, presented by L.Bergomi in [14], offers an exciting way of pricing options using forward variance. An advantage of such a model is that it uses forward variance, traded on secondary markets. Nevertheless, a secondary market forces us to carry the position on the realized variance, unlike in a standardized one. In this regard, to set up the Bergomi model, we need to issue a crucial assumption, which is that the forward variance only depends on the time to maturity of options. Before deepening on how the model is constructed, 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 to maturity of $T - t$ such:

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

Thus, we can deduce the direct formula of the forward variance ξ_t^T :

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

Definition 6 (Instantaneous Variance). *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 remaining of the thesis, we will note the Instantaneous volatility by the symbol $\xi_t^{T=t}$.*

In the two following sections, we will discuss the technical insights the Bergomi model is offering to u:

6.1 One factor model

First, a one-factor Bergomi model assumes that the volatility follows such SDE:

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

- ξ_t^T is the instantaneous forward variance.

- $\omega(\tau)$ is a weighted average of the future volatility forwards at time t , for an horizon of T .
- W'_t is a standard Brownian path at time 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 omitted because it is trivial.

Moreover, our randomness is still constructed by the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ $\forall t \in \mathbb{R}_+$, as exemplified on chapter 1. At the same time, to have the possibility of hedging our position using variance swaps we need a mean reversible process from 0 to our maturity t , explaining the dynamic of our variance evolution. A solution to implement our mean reversible process is to make dW'_t , a mean reverting process:

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

Doesn't it seem like an OU process of the form:

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

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

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

With $X_t = \omega(\tau)dW'_t$.

Afterward, we obtain the solution of our SDE and can now control the volatility of the volatility skew of our model such :

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

Proof of 6.1. First, we use a Taylor serie of degree 1 such:

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

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

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

By Ito's isometry, we derive X_t^2 :

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

□

It is now trivial to derive the relation 6.4, defining ω as the log-normal volatility of $\xi_t^{T=t}$. Thus, as shown by 7.4 , each forward curve depends solely on its standard Brownian motion, precluding us to find a semi-closed or closed formula.

6.2 N factors model

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

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

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

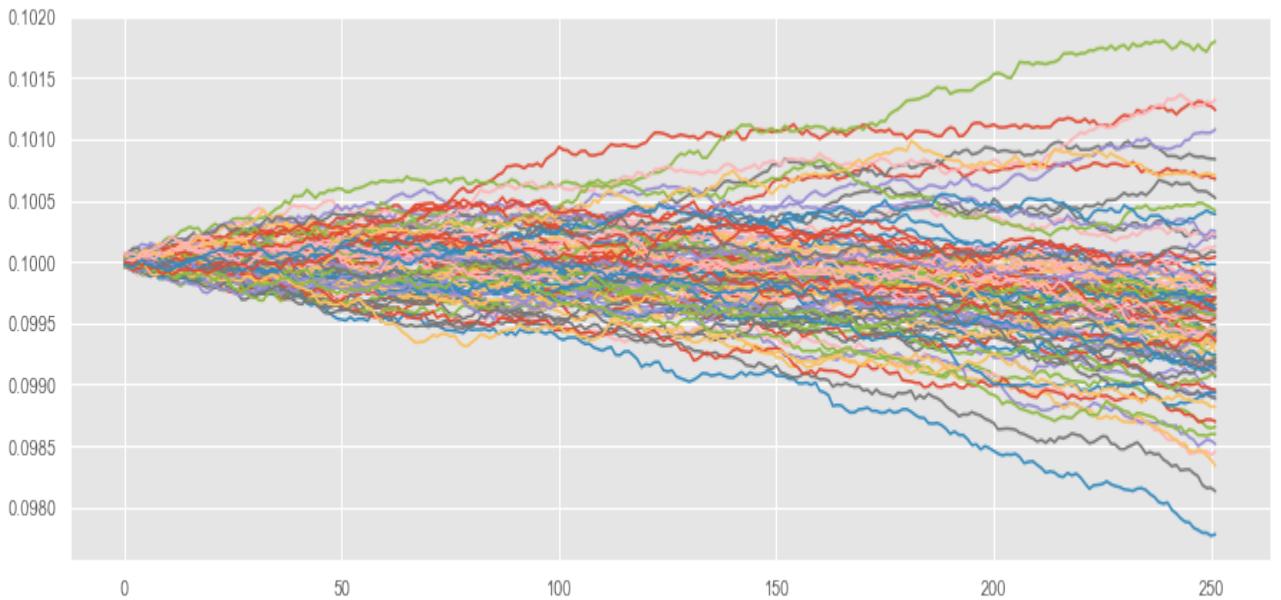


Figure 6.1: Monte Carlo simulations of 100 forward curves under Bergomi's model with parameters $\kappa = 1$, $T = 2$, $dt = \frac{1}{252}$, $\xi_0^T = 0.10$ and $w = 0.01$.

- $W_t^{\mathbb{P},i}$ is the i th correlated Brownian motion

In the case of an N factor model in 6.4, with a 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 take our SDE equation 6.3 and integrate it 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 :

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

We integrate and 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, we obtain those two SDEs:

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

Where W_t is a standard Brownian Motion.

Chapter 7

The Fractional Brownian Motions

Throughout this chapter, we will study processes defined in a probability space such $(\Omega, \mathcal{F}_t, \mathbb{P})$, for $t \in R_+$.

7.1 Definitions

Definition 7 (Non independence). *Fractional Brownian motion is neither Markovian nor Martingale, its covariance matrix is time-dependent and is expressed 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. The Hurst parameter 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 determined times $(s, t) \in (M_s \cup J_t)$, such:*

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

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. In a 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 for any $h > 0$ and any $t \in \mathbb{R}$:*

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

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

Definition 11 (Regularity). W_t^H is almost nowhere differentiable. However, all its paths are locally Hölder continuous, meaning that for any $H > 0$, $t < T$, $s > 0$ and $\epsilon \rightarrow 0$, there is 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 $N \in \mathbb{R}_+$:

$$M = cm^d \leq km^d$$

Where :

- c is a constant such $c \leq k$.
- m is the scaling factor such $m = \frac{1}{N} > 0$.
- d is the Minkowski–Bouligand dimension of the Koch fractal. In our case, d is equal to 1.262.

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:

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

Intuition can be given by looking at the rough figure 7.1. The more we iterate over the segments, the more it becomes smooth on the interval $|t - s| < r$, $\forall r \rightarrow 0_+$.

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

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

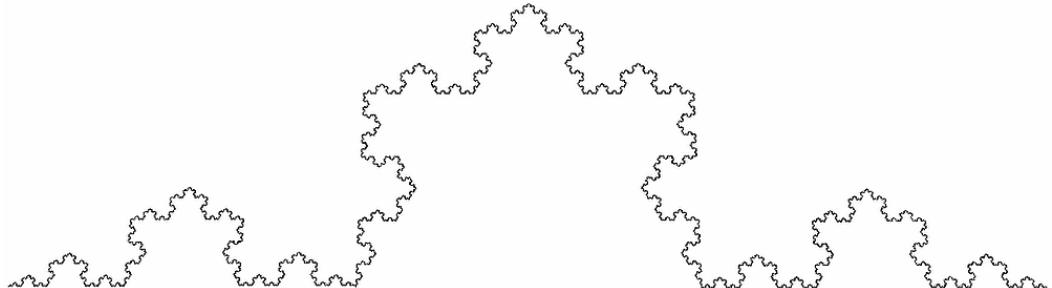


Figure 7.1: The Koch curve is a self-similar and expresses regularity.

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^{1/2} law" of the Standard Brownian motion comes.

Proof of 7.4.

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

Through Ito's isometry we have :

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

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

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

Finally, we get:

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

It is not what we were originally looking for. In this regard, we want the variance $\mathbb{E}((W_{t+}^H - W_{t-}^H)^2)$ to be exclusively dependent on dt^{2H} . For this reason, we rewrite our equation such:

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

Where C_H is equal to:

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

Resultantly, the covariance is expressed as:

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

Remark 6. *The original fractional standard Brownian Motion used by Paul Levis was represented such:*

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

Meanwhile, the repeated Cauchy integral displayed uses a local minimum $a = 0$, which puts too much weight on the origin. Hence, it becomes ill-suited for many applications. In this regard, we prefer to use the Mandelbrot expression 7.3.

7.2 Volterra Process

The properties of the 'rough' stochastic processes are desirable. However, these processes need to be expressed in discrete forms to be tractable. For instance, the Volterra serie can approximate a 'rough' stochastic process by using a kernel $K(t-s)$ that convolutes with a Brownian motion increment dW_s :

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

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

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

Where :

- $t > s > 0$
- $H \in (0, 1)$
- ${}_2F_1$ is an hyper-geometric function

This derivation leads us to a simpler and efficient formula to derive fractional Brownian using a Gaussian quadrature ${}_2F_1$ function:

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

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

7.3 Simulation of Fractional Brownian motions

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

7.3.1 Fraction Brownian motion using Cholesky decomposition:

To use the Cholesky decomposition method in order to generate Fractional Brownian motion, 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} (similar to 9). The terms in the covariance matrix are expressed such:

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

As we said, Γ is a Hermitian matrix. Therefore, we can rewrite 7.6 using the Cholesky Triangulation method :

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

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

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

The last step to obtain our desired FBm is to multiply, using the dot-product, our Lower triangular A 7.7 by an i.i.d random vector 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 positive definition:

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

Where :

- Q is the eigenvector of Γ
- Θ is the diagonal eigenvalues' matrix for $Q \in \mathbb{R}^{n \times n}$
- Θ is defined $\in \mathbb{R}_+^{n \times n}$

Subsequently, to produce a Fractional Brownian path, we have to multiply our eigenvector matrix (Q) to our eigenvalues' matrix at the power $\frac{1}{2}$ such:

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

Consequently, we can generate a fractional Brownian motion by multiplying, using the dot-product, $\Gamma^{\frac{1}{2}}$ to a standard normally distributed matrix Z of dimensions $n \times L$ (for L different discrete random paths):

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

7.3.3 FBm using efficient spectral decomposition :

We can efficiently compute a fractional Brownian motion using the Fast Fourier algorithm. To do so, we need to express our covariance matrix in a circulant way, satisfying the following shift-ward to the bottom property illustrated as :

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

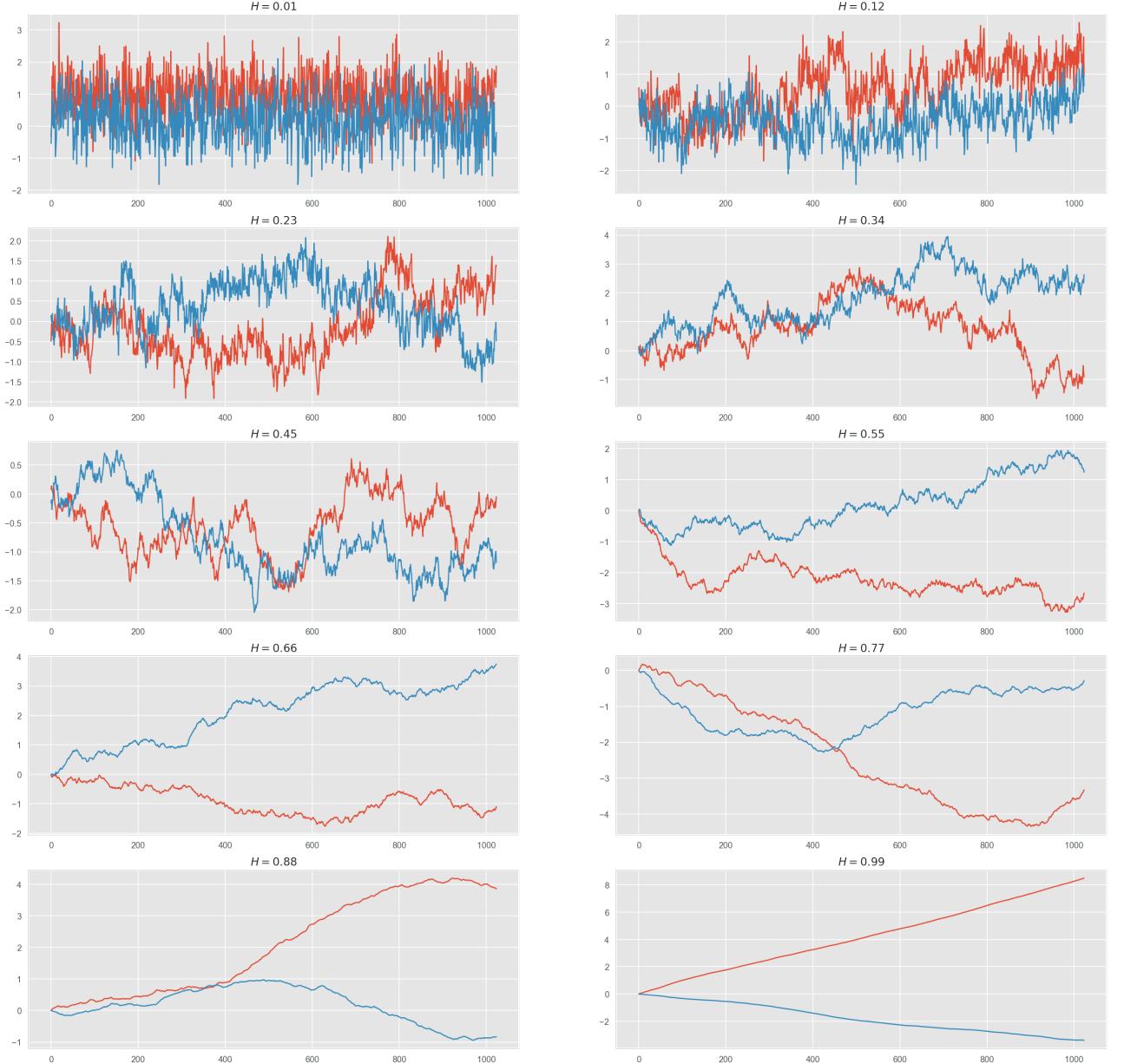


Figure 7.2: Fractional Brownian motions with different Hurst parameters

Because our matrix is not circulant, we cannot directly apply the Fast Fourier algorithm. Instead, a way to circumvent this constrain is to transform our covariance matrix to a topaliz matrix. Where $\gamma(t, s)$ represents the covariance of the Fractional Gaussian noise increment:

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

Proof. We first our covariance matrix such:

$$\mathbb{E}(\delta W_{t,s}^H \delta W_{u,m}^H) = \frac{1}{2}(|t|^{2H} + |u|^{2H} - |t - u|^{2H} - |t|^{2H} - |m|^{2H} + |t - m|^{2H} - |s|^{2H} - |u|^{2H})$$

$$+|s-u|^{2H}+|s|^{2H}+|m|^{2H}-|s-m|^{2H}\Big)$$

Which simplifies to :

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

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

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

Rewriting the expression using $k = t - s$, we obtain our $\gamma(t, s)$:

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

In our case as we are working in a defined amount of days per year. Thereupon, we want to standardize the time to be realistic:

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

However, we still have a Toeplitz matrix that is not circulant (7.8):

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

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 quickly fix the problem. Primarily, we need to embed the covariance matrix (7.9) by a circulant matrix M :

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

Afterward, our matrix becomes circulant, and we can now use the fast Fourier transform algorithm, allowing us to reduce the complexity of our operation from $O(n^3)$ for the Cholesky decomposition and the spectral decomposition to $O(n \log(n))$. This

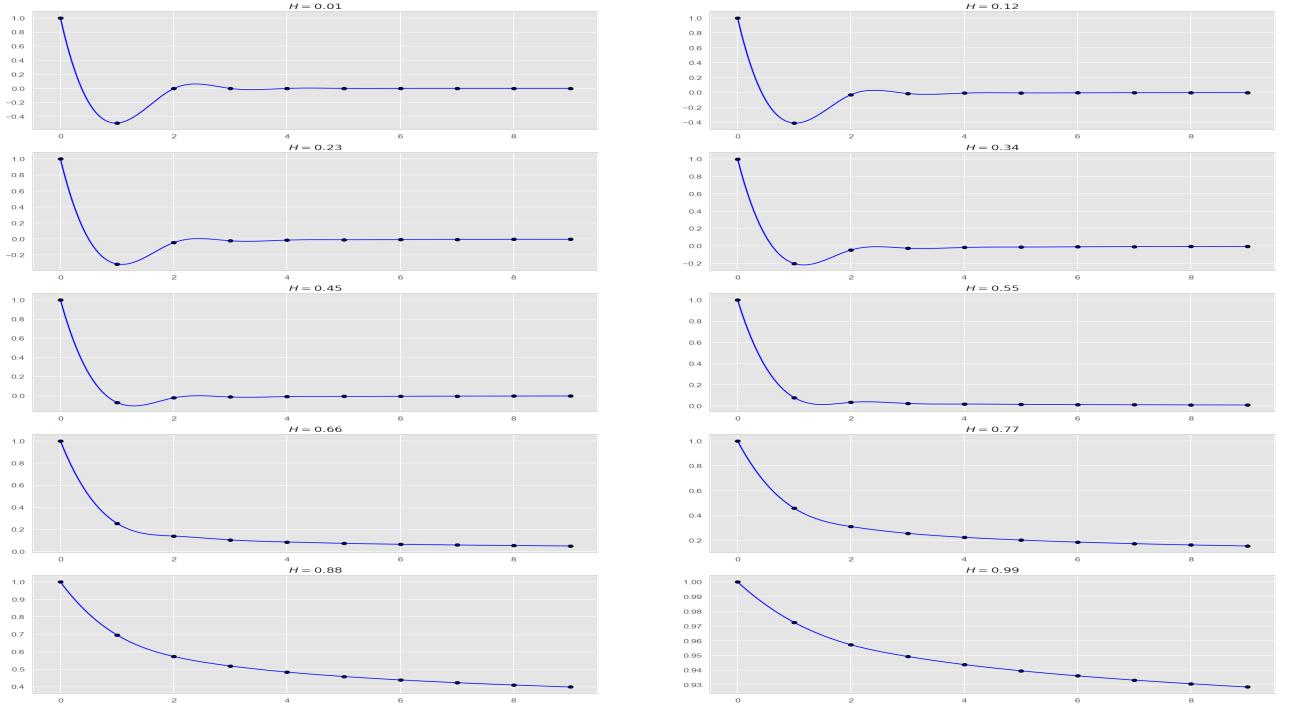


Figure 7.3: Covariance matrix with daily lags

complexity reduction is very nice, especially when dealing with many Monte Carlo simulations and/or time increments.

The last step to generate our Gaussian noises would be to multiply our orthonormal inverse $2(n - 1) \times 2(n - 1)$ discrete Fourier transform matrix, using the dot product, to the square root of our diagonal eigenvalues' matrix. After that, we multiply the obtained matrix by a $2(n - 1) \times L$ matrix that contains our L standard normal vectors. Lastly, we take the real part of the matrix and multiply it by $\sqrt{2}$ such as showed on the following equation:

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

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

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 $\sqrt{(2n - 2)}$ ifft(F) and multiply by $\sqrt{2}$, calling the result Q.Take the real part of Q such $\text{Re}(Q)$.**end procedure**

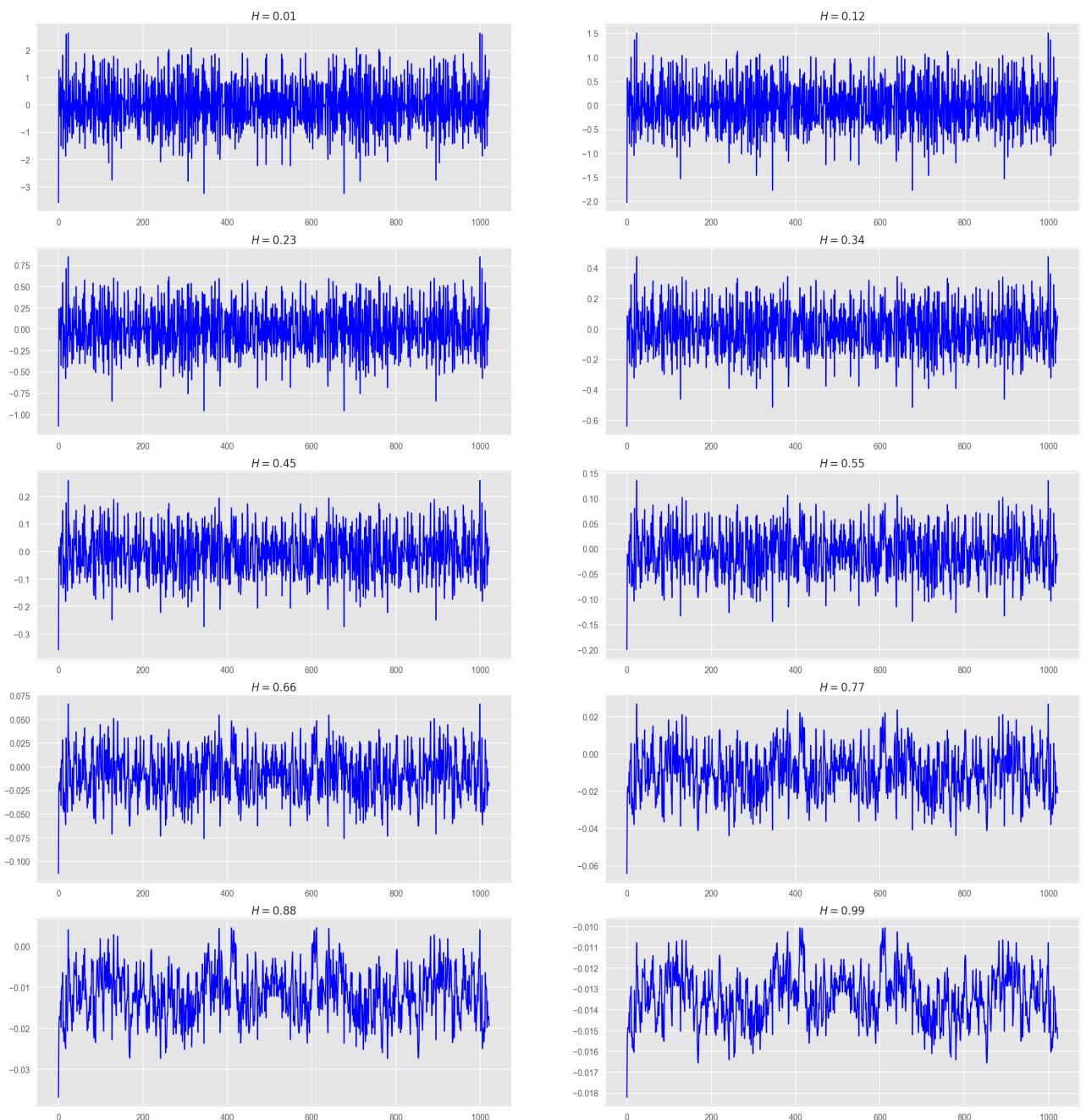


Figure 7.4: Generated fractional Gaussian noises

Chapter 8

The rough Bergomi model

8.1 The unconventional Bergomi model

An interesting fact is that the realized variance distribution of most traded equity indexes is Gaussian 8.5. Furthermore, 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 most straightforward RFSV model [18], that we will cover later:

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

Where:

- $v_{t+\Delta}$ is the variance of the asset at time $t + \Delta$.

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

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

Where:

- X_t , a variable following an OU process with a mean reversal slope $\kappa \rightarrow 0$.

Gatheral, Bayer and Friz [19] found that the Gaussian distribution of realized variance held for all 21 equity indexes, bond futures, crude oil futures, and gold futures in their data-set. Now, considering the original expression of the Fractional Brownian motion 7.3 derived by Mandelbrot under the physical measure \mathbb{P} :

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

In this expression, the left integral $M_t(u)$ is independent of the filtration \mathcal{F}_t . However, the right integral $Z_t(u)$ is \mathcal{F}_t measurable because of its sole dependence on the past events.

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:

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

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

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

Subsequently, it is 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:

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

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

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

Expressing our equation in terms of Wick exponential integral:

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

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 $\mathcal{E}(Z)_t = \exp(drift_t \mathbf{1}_{\mathcal{M}} + Z_t - Z_0 - \frac{1}{2}[Z, Z]_{0,t})$, 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 essential formula to forecast log-volatility is derived by Gatheral, Jaisson and Rosenbaum in [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 fractional Brownian motion with $H < 1/2$ and \mathcal{F}_t

By construction, over any reasonable and equally spaced time scale, we may approximate the fractional OU volatility process in the RFSV model:

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

For some constants ν and C . The forecast formula for log-variance is:

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

The Riemann sum approximation of the 8.2 formula, knowing that the variance is assumed easily observable in the market (our data are cleaned), is used to forecast the log-variance for the periods of 1,5 and 20 days ahead ($\Delta = 1, 5, 20$). Thereby, recall the relation 8.1 and formulate the variance forecast such :

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

It is now straightforward to compute the variance $\sigma_t^2 \mid \mathcal{F}_t$ for a determined lags such:

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

Using the Riemann Sum:

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

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

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

We compared different fits of variance forecasts given some values ν . In this regard, 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 slate of parameters ν

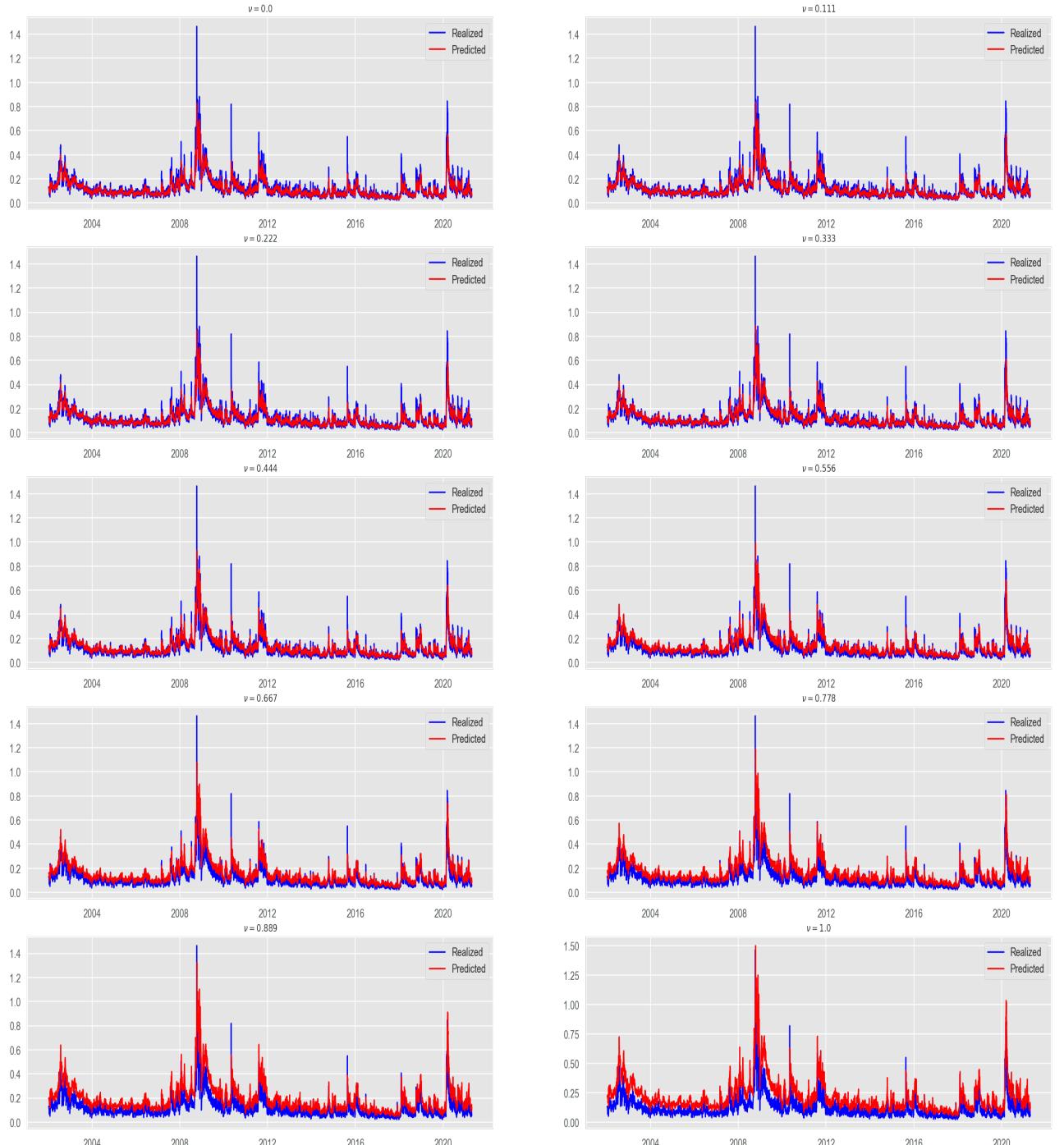


Figure 8.1: Forecast and realized variance for $\Delta = \frac{1}{252}$ given different parameters ν

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

Comte and Renault expressed the first rough volatility model in [21]. Their main idea was to replace the standard Brownian motion expressed on the SDEs with a mean revertible one. Indeed, it is this mean revertible Brownian motion that will later be called fractional Brownian Motion. As usual, we define the probability space: $(\Omega, \mathcal{F}_t, \mathbb{P}) \forall t \in \mathbb{R}_+$.

The SDEs solved by the FSV model are:

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

While solving the first equation is trivial, the second equation is not harder because it underscores the dynamic of an OU process. Consequently, the solutions of our SDEs are:

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

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. But why did the authors use Fractional Brownian motions with long-term memory? The answer resides from the fact that in the past, it has been the widely accepted that the volatility exhibits long-term memory. Never heard the myth 'Periods of high volatility follow up'.

Later on, Gatheral, Jaisson, and Rosenbaum [18] showed that long memory models don't hold with empirical data because of the nonlinearity of their log-moments compared to the one of the original data. Moreover, in [18], it is shown that the RFSV holds with empirical data (8.2). To make it quick, an RFSV model is just an FSV model that exhibits short-term memory on its fractional Brownian motion, 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 :

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

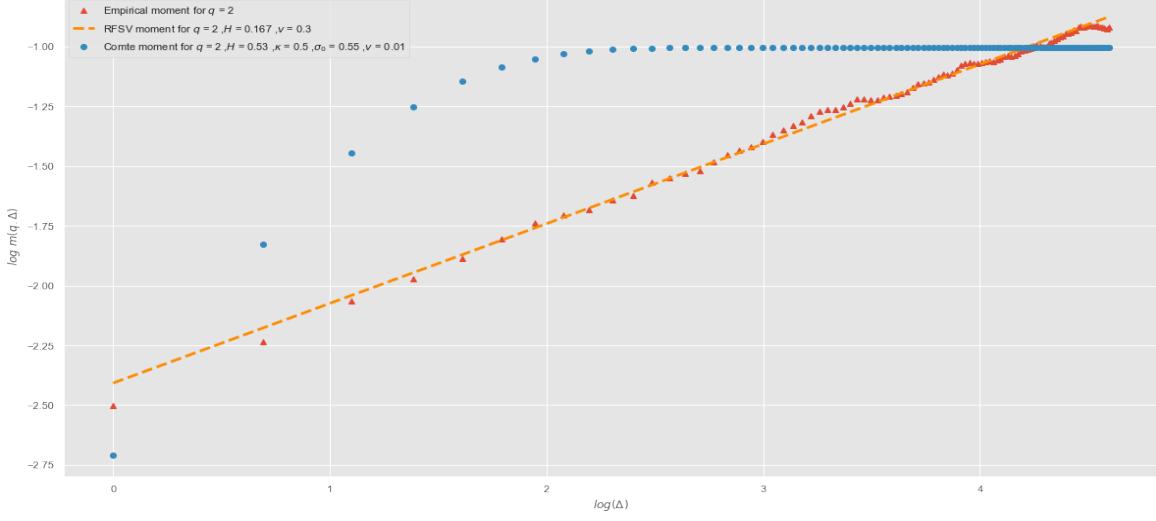


Figure 8.2: A log-moment comparison between the RFSV and FSV models

$$\mathbb{E}\left(dW_u^{\mathbb{P},1}dW_u^{\mathbb{P},2}\right) = \rho du$$

Knowing that the log-volatility is usually normal (8.5 and [19]), we could easily find a way to solve the SDEs of both the asset price and the variance such that our final result is given by:

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

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

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

A considerable drawback of the rough Bergomi model is the absence of tractable closed formula. This drawback results from the fact that a fractional Brownian motion is neither a Martingale nor a 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:

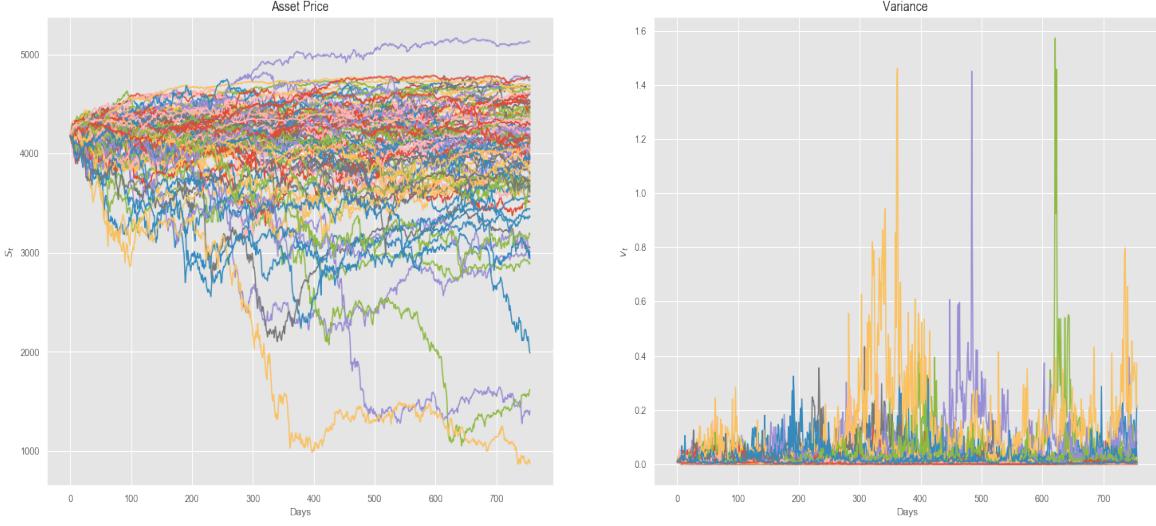


Figure 8.3: Monte Carlo simulation of the rough Bergomi model under the \mathbb{P} measure

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

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 $[t, T]$:

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

The correlation factor of $W_u^{\mathbb{P},1}$ and $W_u^{\mathbb{P},2}$ is equal to ρ , which is empirically negative for the equity [19]. 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 $Z_u^{\mathbb{P}}$

After that, 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. In this regard, the variance process can be expressed via the pricing measure \mathbb{Q} so that, assuming:

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

The risk-neutral forward variance is 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, it is practically impossible to evaluate λ without altering the log-normal property of the variance. As a result, we should just set the parameter $\lambda_s = 0$, and move forward. In this regard, the Rough Bergomi model is non-Markovian in the instantaneous variance v_u and the risk-neutral stochastic differential equations in \mathbb{Q} are:

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

8.6 Volatility skew

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

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

Where:

- x is the log moneyness of the option.
- τ is the time to maturity.

The volatility skews across maturities are very sensitive to the choice of dynamics in a stochastic model. For the simplest models, such as shown in the figure 8.4, the volatility skew stays flat even for very short-term maturities.

Although adding jumps to stochastic volatility models can improve the volatility skew across maturities, it requires understanding involved models and time-consuming

calibrations. Moreover, the simplest stochastic volatility models are inconsistent with forming an accurate at-the-money term structure skews. Indeed, this is especially true for very short-term maturities.

Empirical studies [22] find that the ATM term structure skew can be well approximated by the power-law $\psi(\tau) \sim A\tau^{-\alpha} \quad \forall \alpha \in (0.3, 0.5)$, $A \in (0, 1)$ and $\alpha = H - \frac{1}{2}$. 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 . This is possible if we use 8.4, a variable change $x = \log \frac{K}{S_t}$, and the chain rule.

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

In terms of strikes K , we obtain:

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

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

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

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 :

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

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

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

8.7 The smoothness of volatility log increments

It is not cryptic that the log-increments of volatility are normally distributed for slate of traded instruments. However, to assess this finding, we need to study the evolution of the log-increments of volatility $\delta \log \sigma$ at time $t \in \mathbb{R}_+$. In addition, the log-increments is studied given a specific time lag Δ , where $\min\{t_i\} = \Delta \forall i \in \{1, 2, \dots, N\}$ and $N\Delta = T = \max\{t_i\}$:

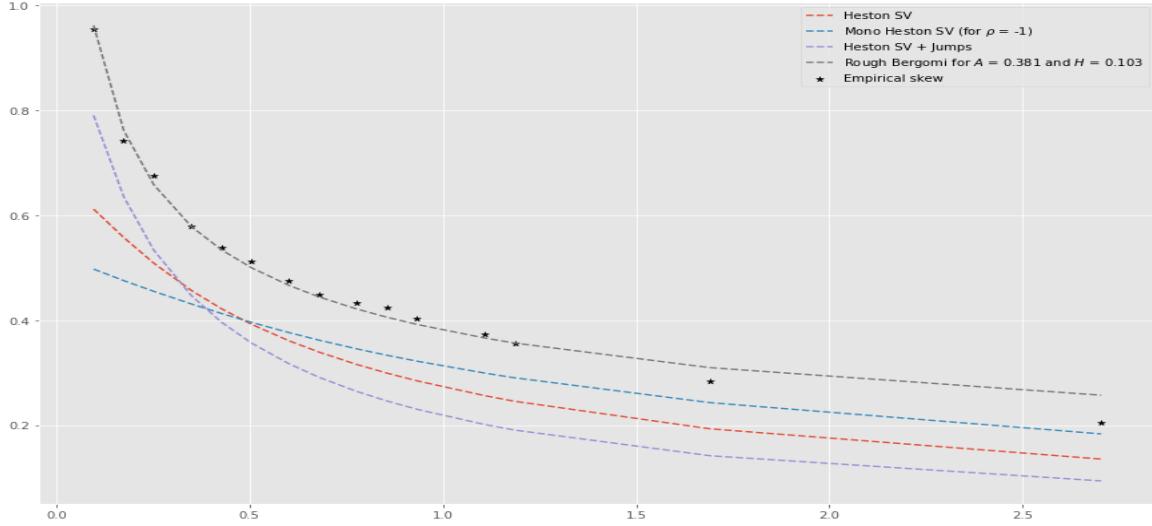


Figure 8.4: Term structure of S&P500 at-the-money-forward volatility skew and its alternative fit

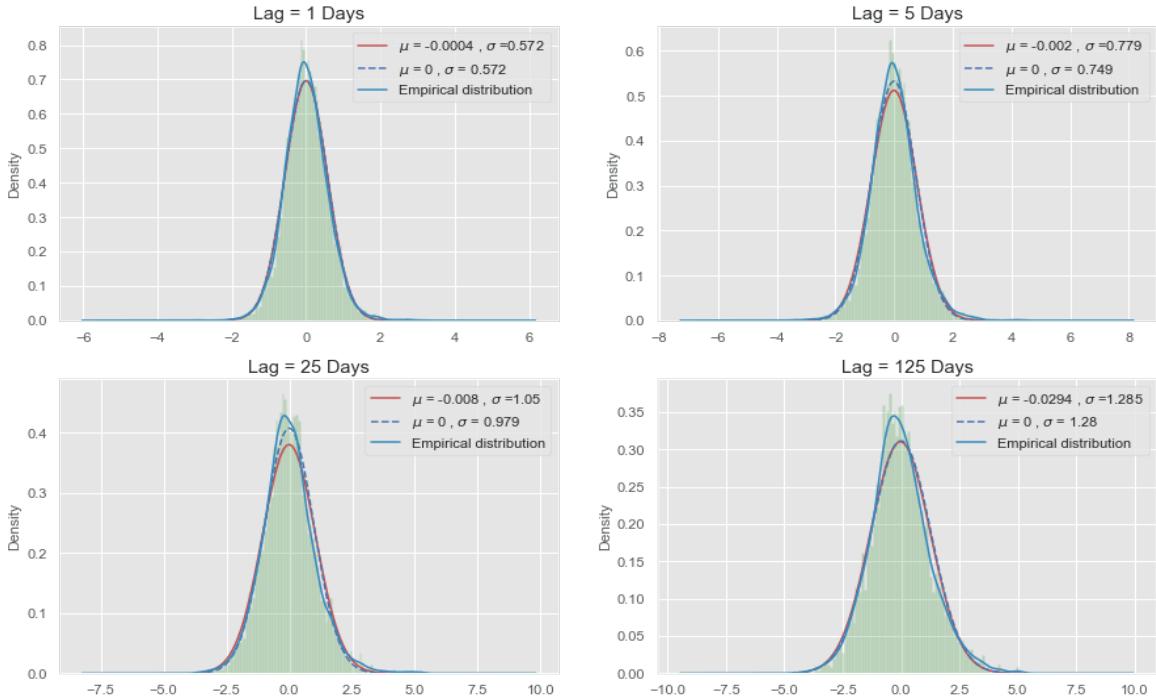


Figure 8.5: Log-volatility increments distribution, under different lags Δ

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

However, a big question is still unanswered. How to find the optimal Hurst H parameter simulate our Monte Carlo paths? One possible solution for that problem is to use sample moments. To do so, we fix our q th moment and our time lags Δ :

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

By looking to the generated moment graph we can approximate each of the q th moments by the power-law 8.7: $\log m(q, \Delta) \sim \Delta^{\zeta_q}$. Thereafter, we try to estimate the fractal scaling relationship:

$$\zeta_q = qH$$

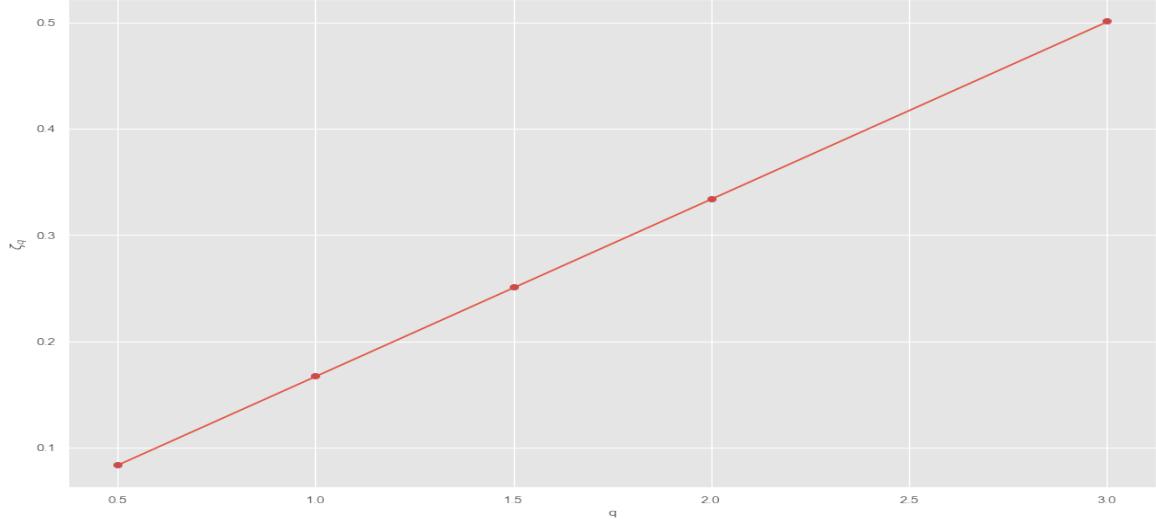


Figure 8.6: Linear fit between $\log m(q, \Delta)$ and qH

Where $H \in (0, \frac{1}{2})$

A noticeable drawback in this estimation is the over-dependency on the empirical data, which can be very noisy. Indeed, Gatheral empirically found that our estimator $H = 0.166$ is biased high because the log-volatility increments are equally weighted, even if the mean squared error of our linear estimator is equal to 1.95×10^{-8} for 5 evaluations.

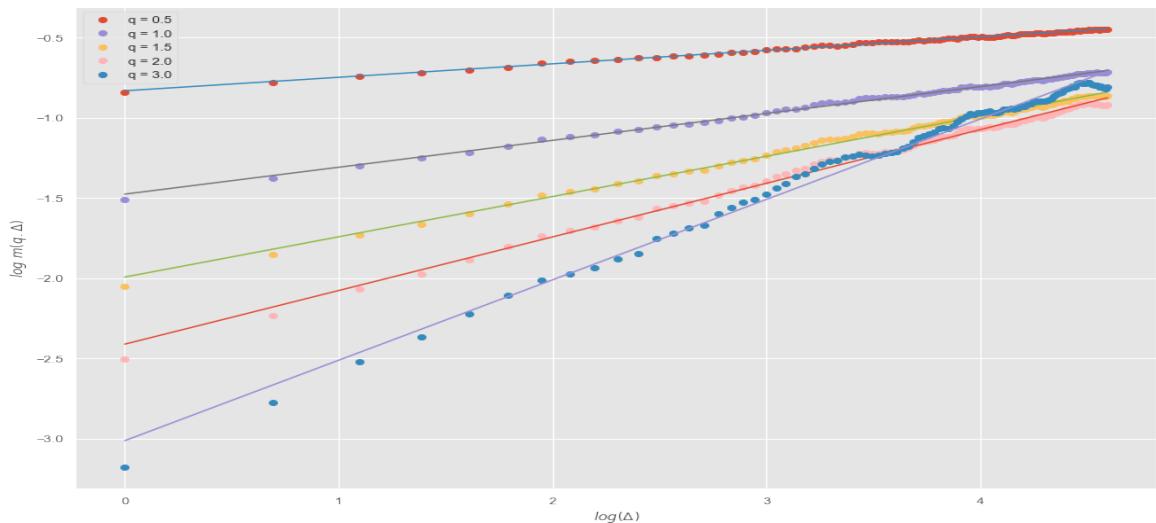


Figure 8.7: Log-moments of the log-volatility increments

Lastly, Gatheral, Jaisson, and Rosenbaum [18] demonstrated that neither of the two fitting methods delivers a satisfactory Hurst parameter to price options. For this reason, we might try to estimate all our parameters using Monte Carlo simulations.

8.8 Volatility skew of the rough Bergomi model

The rough Bergomi model's volatility skew fits the data extremely well. A significant drawback, for now, is the nonexistence of a rapid-solving method, forcing us to use either global optimizers or local optimizers with very good initial guesses. In both cases, there is no reason for the parameters to be stable. The differential algorithm is a global optimizer giving satisfactory results at the expense of time: the algorithm can take up to several hours to produce results. In this regard, to speed-up the computations, we can use the Nelder-Mead algorithm in combination with the following initial guess:

\tilde{H}	$\tilde{\eta}$	$\tilde{\rho}$	$\tilde{\xi}_0^T$
0.102	2.173	-0.871	0.0209

For the concern of the thesis, we used the differential evolution algorithm to calibrate our model. All the price structures were calibrated in a time frame of 3 hours, using 30000 Monte Carlo simulations. In our case, we calibrated 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.

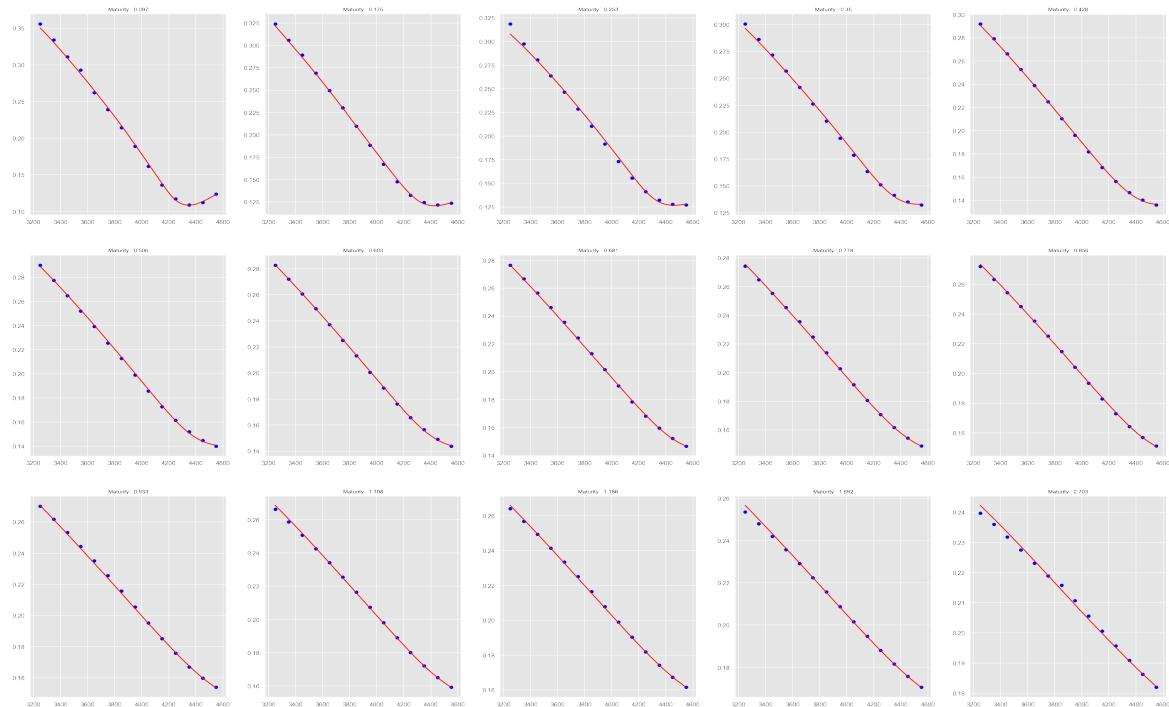


Figure 8.8: Rough volatility skews

Chapter 9

Calibration

9.1 Cumulative Distribution functions approximation for SV models :

A huge downside of the Gil Peleaz approximation 5.10 is the inefficiency of the computation. Hence, 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 :

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

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

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

Where $\alpha > 0$

Then, 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 using the Fourier transform of $c(\tau, s, k)$:

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

By using the inverse Fourier transform, we retrieve our initial function $C(\tau, s, k)$:

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

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

Finally, we obtain:

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

We can retrieve the original function using the fast Fourier transform algorithm for an N power of 2. This operation allows us to gain fast computing power (FFT divides and conquers 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: Differences in the execution time between 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 stochastic volatility models

The stochastic volatility models' calibrations are extremely cumbersome. The non-convexity across the parameters of the models makes the optimization hard to converge and requires the use of stochastic optimizers such as the differential evolution

or other genetic algorithms. The calibration problem is formulated as an optimization problem, where we aim to minimize the pricing errors between the model's prices and the market prices for an exhaustive set of traded options. A common approach to measure this error is to use the squared difference between the real and model's implied volatility. This approach leads us to the nonlinear least square problem:

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

Where Θ is a vector parameter.

Making the cost function $IVMSE$ convex enough would require to change the cost function to a convergent one , by, for instance, using a damping factor G such:

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

The lack of liquidity in some options creates a considerable bid and ask spread that affects the option's middle price by slowing the calibration. Thereby, a slight change to improve the convergence speed of the optimizer is to add weights in function of the bid ask spread.

Now, Through the remaining of the chapter, we express the parameters of stochastic volatility models by $\Theta = (v_0, \kappa, \theta, \sigma, \rho, \lambda, m, \nu)$, and those of the SSVI surface by $\Theta = (\sigma, \eta, \rho, \gamma)$.

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

With determined weights:

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

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

- First, ϑ gives a change in price with respect to a change in volatility. So when we obtain a bid-ask fluctuations, we can multiply by ϑ to get the bid-ask in terms of volatility.
- The second reason is that bid-ask spreads are wider for higher implied volatility, in general.

Weights A, B, and C are decreasing functions of the spread. Meanwhile, it has been shown [25] that ϑ weighting underperforms every other weighting in the calibration process. Lately, the best results were found using the weight A in our case.

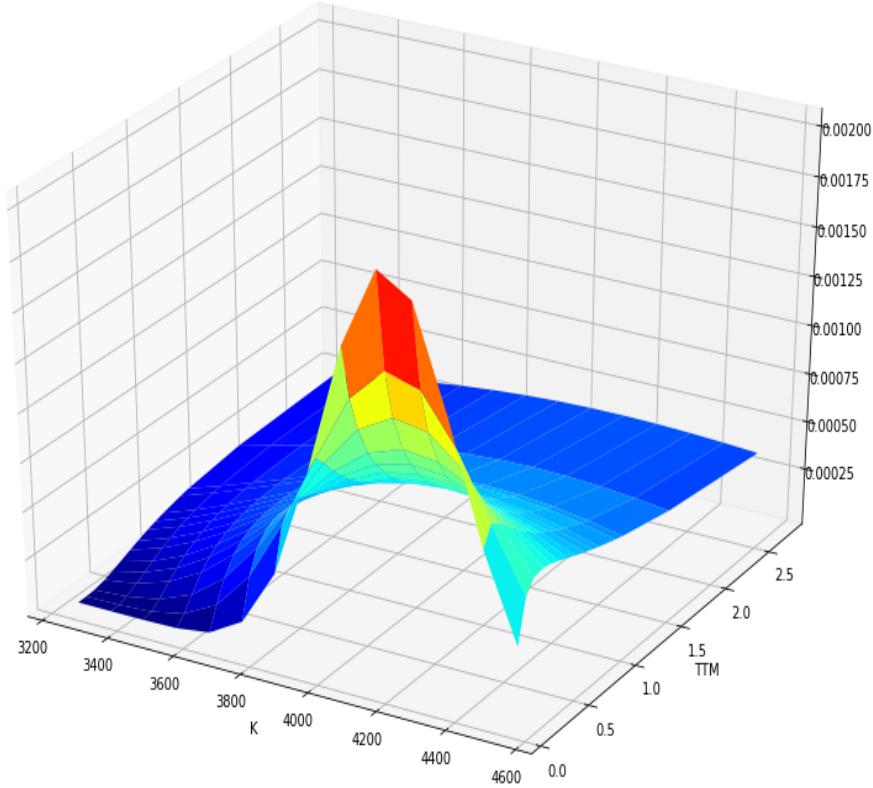


Figure 9.1: ϑ weights

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

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

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

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

On the 4 dimensions, IVMSE error surfaces 9.3, the mispricing error surface is not convex at all and is highly sensitive to slight changes of parameters. As a result, multiple combinations can give great fits, making the optimizer fall on a local optimum quickly.

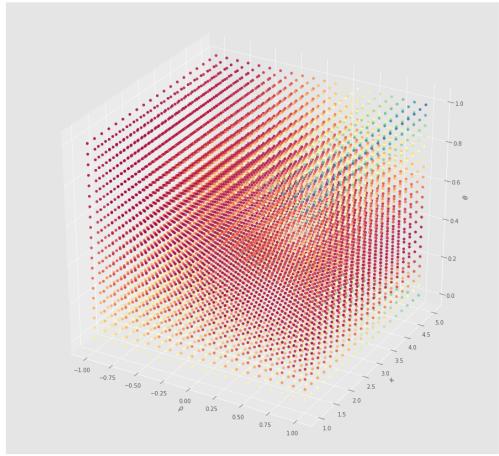


Figure 9.2: IVMSE scatter-plot for the calibration of a simple Heston surface

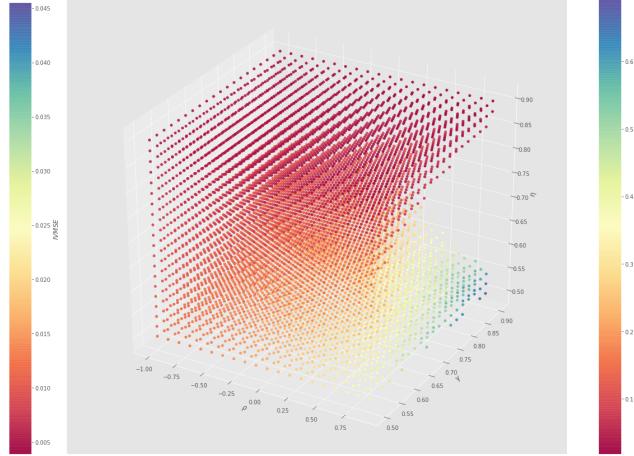


Figure 9.3: IVMSE scatter-plot for the calibration of a surface SVI

	Summary
MSE	852.393
$AARE(\text{in}\%)$	7.838
$RMSE$	29.196
$IVMSE$	2.130×10^{-4}
N	210
σ	0.428
γ	1.404
η	0.172
ρ	-0.712

Table 9.3: SSVI calibration results

	Summary
MSE	153.992
$AARE(\text{in}\%)$	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 shown on figure 9.3, the error cost function is not linear across the parameters. A gradient descend algorithm ($\Theta_{k+1} = \Theta_k - \eta \nabla f(\Theta)$) suggested, in [26], won't certainly converge towards a global solution unless a very good guess is given. Instead, we can use a global optimizer based on Metaheuristic methods to find a good solution.

Definition 15 (Metaheuristic methods). *Metaheuristics methods are algorithms used to solve complex problems, generally involving high dimensions. While these algorithms do not ensure finding a global optimum, they make very few assumptions on the optimization problems and provide typically good enough solutions.*

Differential Evolution Algorithm To fit all our surfaces, we used a stochastic optimizer developed by Storn and Price [27]. The steps performed by the optimizer are

as follows:

- Define the objective cost function, in our case $IVMSE(\Theta)$ for $IVMSE : \mathbb{R}_{1 \setminus \rho}^n \rightarrow \mathbb{R}_+$
- 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 that $M \in (0, 2)$. 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 of the large search radius.
- The algorithm now mutates each candidate solution by randomly adding 2 candidates, generating a trial candidate. To do that, we used a well-known strategy called "*best1bin*", which determines the new candidate:

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

- We must choose a recombination number such $C \in (0, 1)$. The recombination dictates the stochasticity of the research 9.1. The lower the parameter, the more random the search is. Usually, we want a fixed and converging solution, making the C parameter crucial for us. In our case, we fixed it as $C = 0.70$.

The differential evolution solved numerous scientific and industrial problems during the last decade. This is because of the stochastic nature that prevails these algorithms to be trapped very easily on local optimum.

However, constraints such as the feller condition need to be checked 2 each time we select a new population of candidates. A simple but "unhealthy" way of adding this constrain will be to add a penalty if the constrains are not respected. The penalty was set as a multiplier equal to 10^{50} in our case.

Nelder-Mead Algorithm To polish the results, we also used a Nelder-Mead optimizer and used as an initial guess the solution given by 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 reverse to the differential equation algorithm, the Nelder-Mead's algorithm is not stochastic, that's why we are using it only to polish our results. The search method of the Nelder-Mead algorithm is explained on the original paper of J.Nelder and R.Mead [28]. The steps performed by the optimizer are as follows.

- Define the N dimension calibration problem and set an initial parameter x_{init} . After that, we generate n points to rank the outcomes $IVMSE(x_0) < IVMSE(x_1) \dots < IVMSE(x_n)$.
- Using the $n + 1$ points we set a simplex in $n + 1$ dimensions.

Algorithm 2 Differential Evolution Algorithm

```

procedure DE HESTON OPTIMIZER (self)            $\triangleright$  We don't need an initial guess
   $t = 0; j = 0; x = 0; tol = 1; CR = 0.7; M = 1; popsize = 50$ 
  Set dimension  $N = 5$                                  $\triangleright N = 8$  for SVJ Heston
  Set bounds bnds for the different parameters       $\triangleright$  for instance,  $|\rho| \leq 1$ 
  Set constraints cons for the different parameters    $\triangleright$  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
         $pen[j] = 10^{50}$ 
      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
           $pen[j] = 10^{50}$ 
        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

```

- We compute the centroid c of all the $IVMSE(x_i)$ $\forall i \in \{0, \dots, n - 1\}$, such $c = \frac{\sum_{i=0}^{n-1} x_i}{n}$.
- We first implement the reflection method, which consists of choosing a point from a basis of the worst one (x_n in our case):

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

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

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

- We use the contraction method, trying to find a point from the centroid towards the interior of the simplex:

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

- We replace the worst point x_0 by x_r and keep doing the same until the maximum of iterations or the minimum tolerance is reached.

Usually the parameters are fixed such $\rho = .50$, $\gamma = 2$ and $\alpha = 1$. In our algorithm, we omitted to include a step: the shrinkage. Shrinkage is usually a redefined simplex, which is very unlikely to happen in practice. In this regard, all efficient Nelder-Mead algorithms do not use it. However, a drawback of the Nelder-Mead algorithm is that it can handle neither bounds nor constraints, pushing us to implement penalties to avoid either bounds' or constraints' violations. On our implementation we omitted to input penalties for brevity. Otherwise each step would have to carry an additional *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
$AARE(\text{in}\%)$	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

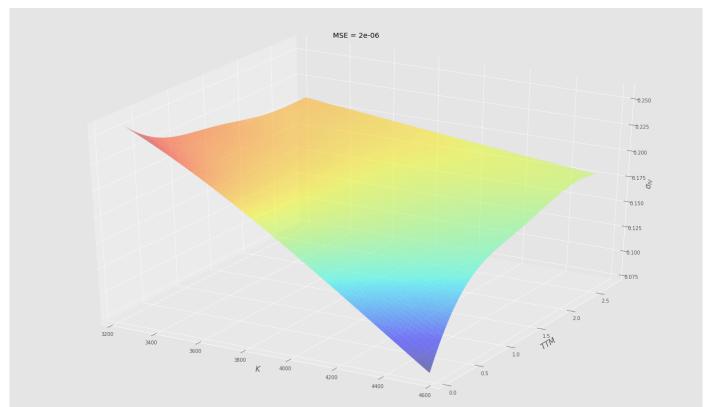


Figure 9.4: SVJ surface generated by a Gaussian process interpolation

Table 9.5: SVJ Heston calibration results

Algorithm 3 Nelder-Mead Algorithm

```

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

```

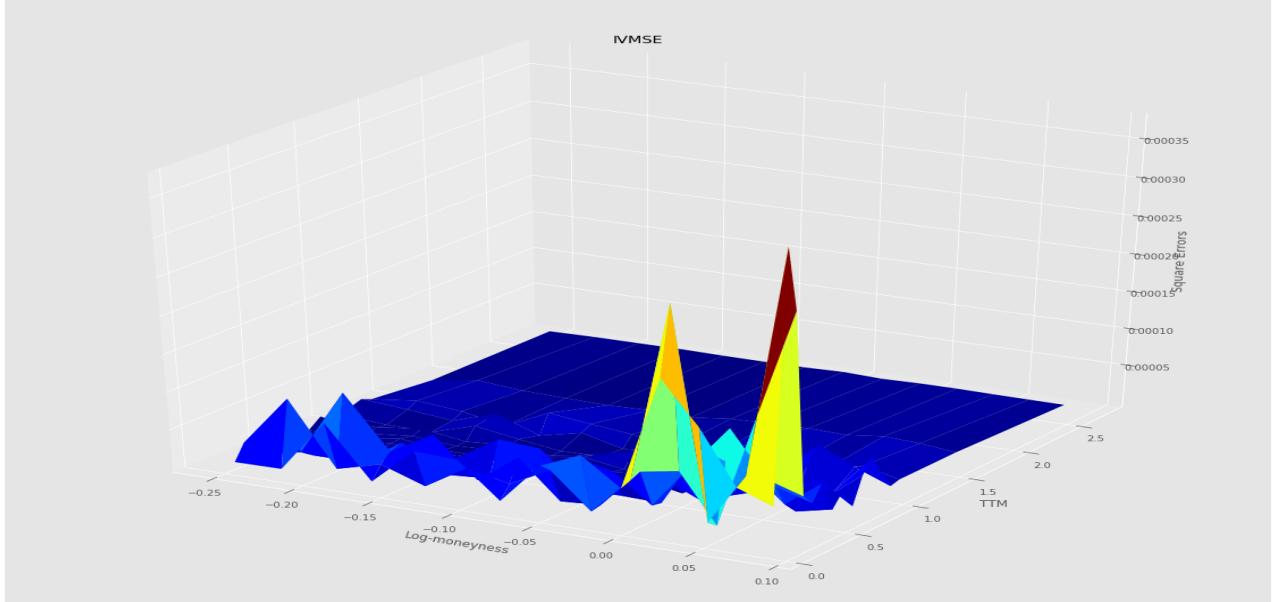


Figure 9.5: IVMSE errors of the SSVI surface

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:

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

A change of variable is done:

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

Under this change of variables, the total implied variance in the raw parameterization is:

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

Where :

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

To sum up, we just created a linear formula of the total implied variance. The total implied variance effectively depends on 3 terms: \hat{a} , d and c . Therefore, we can employ a quasi-explicit (QE) parameterization for the convex linear problem. We define the domains of our parameters in \mathbb{R}^3 :

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

For a fixed set (σ, m) , a set of options' log-moneyness x_i and implied total volatility $w_i \forall i \in 0, \dots, N$, we face the following 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 a first quadratic cost function.
- m and σ are iterated guesses.

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

Resultantly, we add an optimization for the parameters σ and m , using a standard cost function:

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

Where:

- $C_2(m, \sigma, x_i, w_i, \hat{a}^*, d^*, c^*, x_i, w_i)$ is a quadratic cost function.
- \hat{a}^*, d^*, c^* are constant.

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

With the parameter's domain being:

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

To have the parameters satisfying the Durrelman condition 5.21 is essential. It allows the price of cliquet options to not fall under 0, provoking an arbitrage opportunity. To do that, we inspire ourselves from the SVI Jump-Wing reparametrization 5.22.

Chapter 10

Interpolations

10.1 Multivariate normal distribution

We decided to use a Gaussian process regression to interpolate the obtained implied volatility surfaces. We will first Recall the most fundamental properties of the Gaussian distribution.

A multidimensional Gaussian variable $X \sim \mathcal{N}(\mu, \Sigma)$ has as a mean a vector μ and as a covariance, a matrix Σ . The probability density function of the variable X is:

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

- $|\Sigma|$ 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:

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

Where:

- G is a normalization term.

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

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

Where:

- Λ is a diagonal matrix that contains all the positive eigenvalues.

Remark 11. The motivation behind our choice to use the Gaussian distribution is that a lot of data seem to be normally distributed in the real world (we saw it was the case for the log-volatility increments of many traded instruments) for a good reason: the Central Limit Theorem (CLT). The CLT states that the arithmetic mean of infinite samples are approximately normal distributed, provided samples have finite mean and variance.

We then set a new pair of variables z and y :

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

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 the CDF is always capped to 1. Therefore, we deduce G 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 like integration, addition, subtraction, division, and multiplication don't modify their normal properties (unless we multiply the i.i.d variable by 0 or subtract by itself). Before we continue, some important properties are recalled:

- 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:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma_l^2}\right) \quad (10.3)$$

Where:

- σ_l is a smoothing parameter.

The kernel function 10.3 leads us to our $\mathbb{R}_+^{d \times d}$ covariance matrix:

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

The main advantage of using the "super" RBF kernels 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. For better understanding let's define the multidimensional matrix based on the 2-dimensional coordinates vectors x_1 and x_2 of the same dimension $n \times 1$:

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

Where:

- $A = [1, 1, \dots, 1]^T$

More details on the RBF kernels are available on the paper [30].

10.4 Gaussian Process Regression

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

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

For the remaining of the thesis, we adopt the following notations:

- $m(\cdot)$ 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 :

$$y_{i,j} = f(strike_i, ttm_j) + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2)$$

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:

$$\begin{aligned} \mathbf{y} &= \mathbf{f} + \varepsilon \\ \mathbf{f}' &= m(\mathbf{X}') + \varepsilon_1 \end{aligned}$$

$$\mathbf{z} = \mathbf{f}' - K(\mathbf{X}, \mathbf{X}') (K(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} \mathbf{y}$$

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

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

The covariance matrix is derived such:

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

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

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

Now, we can predict our mean conditionally to the two coordinates dimensions for each possible point on a surface. Additionally, it is now possible for us to evaluate a confidence interval for each prediction. For a more complete understanding of the Gaussian process regression, one may try to read [31], which covers a big chunk of material concerning the Gaussian process regression.

Conclusion

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

We then showed that stochastic volatility models do not correctly reproduce the market's empirical skews. Hence, we tried to implement some solutions to the stochastic volatility model, by adding jumps. However, the major drawback remains the extreme difficulty of converging the optimizer to a set of solutions. In addition, the stochastic volatility models with jumps still do not fit accurately the term structures of options.

As the stochastic volatility models weren't efficient to fit the empirical data, we decided to adopt a model lead by a non-Markovian process. This model, which Bergomi originally derived, had a tractable economic sens. Indeed, We were deriving the realized volatility by discounting the volatility swaps using time-to-maturity weights. However, the Bergomi's model originally used a semi-Martingale approach to model the volatility, which fails in reality.

Meanwhile, the results yielded by the rough Bergomi model using a fractional Brownian motion are satisfactory as we were able to fit all the skews and term structures of sample data correctly. The advantage of such a model resides in its intuitiveness and easy understanding. The rough Bergomi model has only three parameters that are easy to understand and optimize. Additionally, this model allows us to remove the pain of adding and calibration jump processes.

Nevertheless, the lack of existence of closed or semi-closed solutions forces us to use highly computational expensive Monte Carlo simulations. In addition, we need to use a high enough number of paths to allow our optimizer to converge to a set of solutions.

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

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

Rho derivation

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

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

Theta derivation

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

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 Conjugates 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 :

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

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

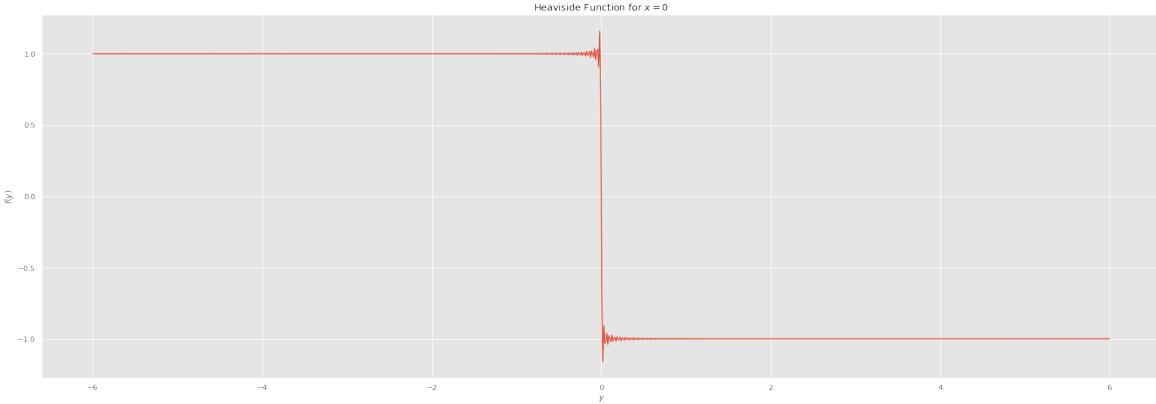


Figure 11.1: Heaviside Function

Afterward, we continue the derivation:

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

We set $\lambda \rightarrow \infty$ and $\psi \rightarrow 0$, only knowing the Dirichlet integral such $\int_0^\infty \left| \frac{\sin(x)}{x} \right| = \frac{\pi}{2} < \infty$. Resultantly, we can apply Fubini's theorem:

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

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 the x -axis.

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

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 v, 20, 84
- [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 24
- [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 Jacquier. 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
- [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 52, 54, 56, 63
- [19] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016. pages 52, 57, 58
- [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 53
- [21] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323, 1998. pages 56
- [22] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15:635–654, 2011. pages 60
- [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 65
- [24] Peter Carr and Dilip B. Madan. Option valuation using the fast fourier transform, 1999. pages 65
- [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 67
- [26] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer, 2006. pages 69

- [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 70
- [29] Zeliade White. Quasi-explicit calibration of gatheral’s svi model. 2009. pages 74
- [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 79
- [31] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. pages 80