Calibration and simulation of the Heston model

Martin Duguey

Supervisors: Gautier Poursin, Alexis Sanandedji, Paul Vialard.

Master 2 Modélisation aléatoire
UFR de Mathématiques
Université Paris Cité
France

Abstract

The scope of this master's thesis it to propose detailed methods and comparative results to add the Heston model in an industrial quantitative library. We review the model effects as well as mathematical properties. The analytical pricing formula for european vanilla options and its formal computation are crucial in this work. We develop numerical methods for its computation, from quadrature rule to fast fourier transform. We propose calibration results on market data as well as various perspectives depending on purposes. In a performance outlook, we present several calibration approaches. We obtain efficient results as soon as we define a target surface near the money, but we get also competitive outcomes when we use a larger interpolated surface as a target. However, we notice the poor performance of the variance Swaps calibration method. Finally, we also introduce issues related to model inputs sensitivities.

Key words: Heston model, Option Pricing, Diffusion, Calibration, Stochastic Volatility.

Acknowledgements

I would like to thank Exiom Partners for the opportunity to work on a such interesting subject. Particularly I would like to thank my supervisors Gautier Poursin, Alexis Sanandedji and Paul Vialard as well as Paulin Aubert. Your investment and your interest in this project help me to be efficient and creative. Moreover your kindness helps me to become intergated with ease.

I am also very grateful to people whom accept to spend time discussing with me to improve my knowledge and discover new things. A special thank to Iliasse Bourchid, Youssouf Harouna Kerzika and Taha Boutouba. You have deeply participated to the creation of this thesis and opened my mind on different aspects that I haven't thought about.

I would like to also thank people that I have discussed with everyday and whom have turn this internship into a fun experience. I am thinking to my fellow interns, members of the quantitative team and colleagues.

Contents

List of Figures

Introduction

Stochastic volatility is a crucial concept in the field of quantitative finance, particularly in the context of options pricing and risk management. It refers to a financial market phenomenon where the volatility of an asset is not constant over time but rather follows a random or stochastic process. Directly related to implied volatility, these two notions are defined and largely discussed in this thesis. We present implied volatility smile and the wrong assumption of constant volatility models in a pricing and hedging perspective.

The Heston model is a well-known stochastic volatility model proposed by Samuel L. Heston in 1993. Mostly used for equities related asset, this model provides a parametrization that can match the market volatility behaviour. Despite the complexity trade off that we incur with this model, we still can derived indispensable content for portfolio managers in the model framework, from partial differential equation to semi-closed pricing formula for vanilla options. The purpose of this work is to present an overview of this model from a practical perspective. To proceed, we have decided to cover the implementation of several interesting axis related to the model from diffusion to sensitivities computation.

What means stochastic volatility modeling and why the Heston model is well suited for this purpose? In the first chapter we try to answer this question with a comparison to the *Black-Scholes* model. How can we explain market dynamic with the Heston model? This is the topic of the second chapter, as well as the presentation of tools and calibration methods directly derived from the model framework. Numerical results and relevant issues that meet the model's limits are presented in the third chapter. We have also been interested in further work about volatility parametrization and how the market can reflect volatility dependency. This is discussed in the fourth chapter. Finally, we can also wonder what changes are related to portfolio management with the Heston model? To this end, we propose some results from the literature on greeks and sensitivities computation in chapter five.

Besides the results produced and presented in this thesis, the purpose of this internship was to develop codes and documentation on the Heston model in order to create a usable tool in the pricing and valuation plateform of Exiom Partners named **FinX**.

Exiom Partners is a consulting firm that provides, to corporate and investment banking institutions and insurance companies, tools and analytics for risk management. Founded in 2019, the firm employs about 180 collaborators specialized in law, project management or actuarial sciences. Its clients are mostly in France but the activity has been recently diversified with abroad interventions in South America, in Morocco and in the United Kingdom.

Chapter 1

Heston model

In this chapter we recall and review in details the upsides of using the Heston model, particularly in contrast with the *Black-Scholes* model [1]. The mathematical framework and theoretical explanation developed by the author Samuel L. Heston in his article (see [2]) are the core of the first section. We also want to point out convenient discretization of the stochastic differential equation. We discuss about the implementation and the comparison of different schemes. In a final section, we provide a complete developement to meet Heston's partial differential equation for vanilla options.

1.1 Stochastic volatility model

Volatility in the *Black-Scholes* framework is fixed and therefore is independent with respect to time, strike and maturity. This family of models rely on the hypothesis that market's volatility is constant.

As a matter of fact, market's implied volatility changes depending on strike and maturity. It turns out that we observe an implied volatility smile, a curvature along strike values that evolves with maturity. Since implied volatility is defined as the numerical value of volatility in order to obtain the observed market price with the *Black-Scholes* formula for vanilla options, it bears out that this value can't be constant.

Now in a pricing perspective we reach *Black-Scholes* model's boundaries as we aren't able to replicate observed market prices for liquid options. For a given maturity, the dependency in strike of implied volatility can be understood by risk-aversion arguments that we will underline further. The main assumption of Heston's model is that volatility is defined as a random variable.

In the following, we denote as $(\Omega, \mathcal{F}, \mathbb{P})$, the probabilistic space of interest with $\mathbb{P}$ the physical measure of probability.

We consider a specific market with a risky asset denoted as $\{S_t\}_{t\geq 0}$ and a risk-free asset denoted as $\{S_t^0\}_{\{t\geq 0\}}$. For convenient purpose we also consider a constant risk-free interest rate r, such that we can define the discount factor B and the expression of the risk-free asset:

$$B_t := e^{-rt} \quad \text{and} \quad S_t^0 = S_0^0 e^{rt}.$$

The diffusion of $\{S_t\}_{t\geq 0}$ in the case of the Heston model could be written as:

$$\begin{cases} dS_t = \mu S_t dt + \sqrt{\nu(t)} S_t dB_t^{(1)} \\ d\sqrt{\nu(t)} = -\beta\sqrt{\nu(t)}dt + \delta dB_t^{(2)} \\ d < B^{(1)}, B^{(2)} >_t = \rho dt \end{cases} . \tag{1.1}$$

where μ, β et δ are constant parameters. As a matter of fact, problem (1.1) could be transformed into an equivalent and more common form:

$$\begin{cases} dS_t = \mu S_t dt + \sqrt{\nu(t)} S_t dB_t^{(1)} \\ d\nu(t) = \kappa[\theta - \nu(t)]dt + \sigma\sqrt{\nu(t)} dB_t^{(2)} \\ d < B^{(1)}, B^{(2)} >_t = \rho dt \end{cases} . \tag{1.2}$$

This parity is a direct application of Ito's lemma to the square function associated to the process $\{\sqrt{\nu(t)}\}_{t\leq 0}$, which is defined as an Ito process in problem (1.1). Hence we can define constant parameters such that $\kappa := 2\beta$, $\theta := \delta/2\beta$ and $\sigma := 2\delta$. We recognize that in problem (1.2), the variance process $\{\nu(t)\}_{t\leq 0}$ is exactly a CIR process or as mentionned previously a square-root process, thus this process is positive.

Feller's condition defined as $2\kappa\theta > \sigma^2$, ensures strict positivity as well as the existence and uniqueness of the continuous solution of this stochastic differential equation (see [3]). We distinguish five parameters to this model: ν_0 the initial point of the variance process for the reason that it is not directly observable on the market, κ, θ, σ and ρ. Further details will be provided on the influence of each parameter.

As pointed out in litterature (see [4],[5]), since the variance process is well defined, the only difficulty for option pricing is the two-dimensionnal randomness. Considering a portfolio with an available asset of our market and an option on $\{S_t\}_{t\leq 0}$, we can prevent from this risky asset or spot process movements by buying or selling this asset but we can't provide a solution for variance movements as it is not defined as an asset. Therefore, in our market, Heston's model is unusable since hedging an option on $\{S_t\}_{t\leq 0}$ is impossible.

Now that we have introduced Heston's model, we are interested in the effects that can be created with stochastic volatility diffusion. This is a major difference with *Black-Scholes*'s model, and we want to highlight benefits of this representation.

A good start could be to point out shared properties of Heston's model with more classical Geometric Brownian Motion model (GBM). Options prices are affected in the same way with volatility. High values of variance process increase prices whereas low values reduce them.

Let's focus on the diffusion of ν under a risk-neutral framework:

$$d\nu(t) = \kappa^*[\theta^* - \nu(t)]dt - \sigma\sqrt{\nu(t)}dW_t^2, \quad \text{with} \quad \kappa^* := \kappa + \lambda, \quad \theta^* := \frac{\kappa\theta}{\kappa + \lambda}.$$

We identify a drift term that is similar to an Ornstein-Ulhenbeck process, so we can capture the mean reversion phenomenon. As soon as ν is over θ^*, the drift is negative and in return when ν is lower than θ^*, the drift is positive. The drift term helps the process to stay close to θ^* therefore it can be presented as the mean of the process and κ^* can be seen as the speed at which the process ν goes back to its mean. Particularly when θ^* increases, prices also increase as explained before since θ^* has a direct influence on ν. Note that λ shouldn't be considered as a parameter, it is the market price of volatility risk and we refer to [4] for more details. We are interested in finding directly proper κ^* and θ^* that fit market data in order to use risk-neutral tools to price derivatives.

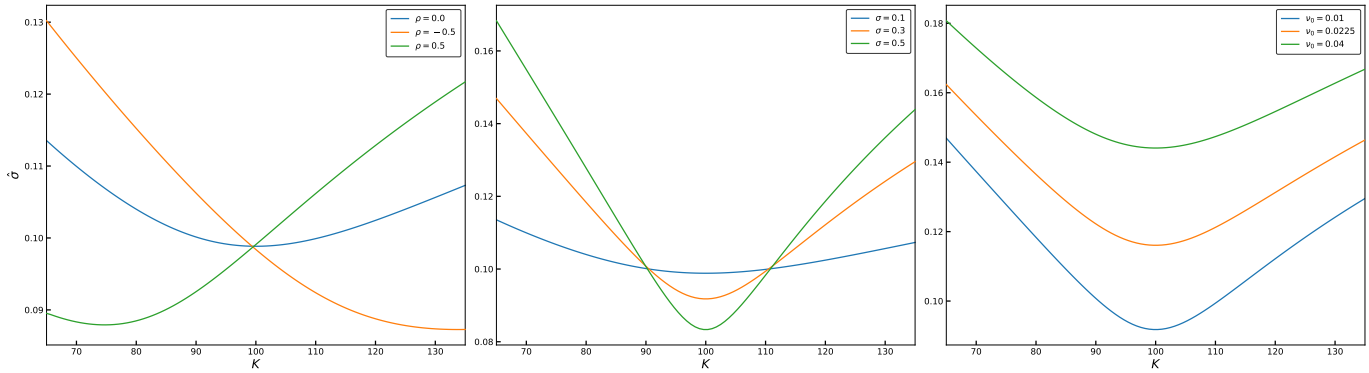

Figure 1.1: Influence of ρ, σ and ν_0 *(left to right)* parameters on implied volatility smile.

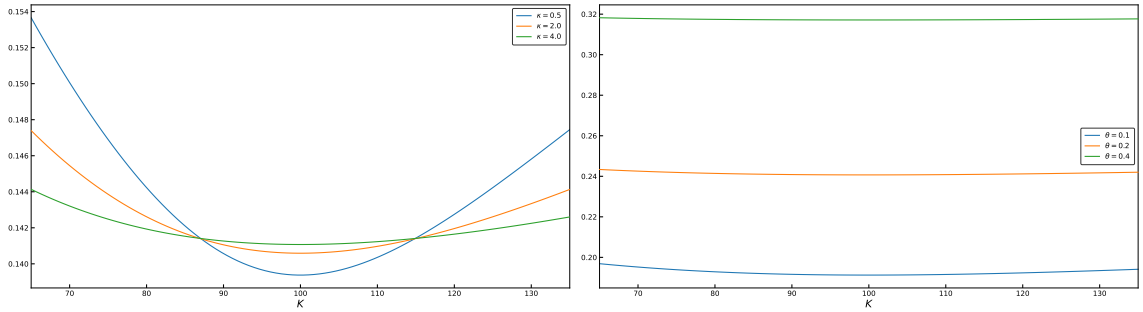

Figure 1.2: Influence of κ and θ parameters on implied volatility smile.

Parameters ν_0, σ and ρ generate effects that change the behavior of the model from a GBM. They allow a reproduction of the implied volatility smile of the market, which isn't possible with a constant volatility model. Figure [1.1] presents the impacts of those parameters on the volatility smile for a specific maturity. Whereas ρ clearly influences skewness of the smile, σ and ν_0 change respectively the slope and the global level of the curve.

The value of σ is meaningful, the higher σ is, the more uncertain the volatility will be. Uncertainties on volatility reflects uncertainties on prices, therefore, and according to observations that we can make on the middle figure [1.1], at-the-money (ATM) options prices are reduced. It is a simple consequence that uncertainty on volatility implies a decreasing probability of the event that the option reaches a positive intrinsic value.

As mentionned before parameters κ and θ have direct influence on variance dynamic, but they also have an influence on volatility smile, see figure [1.2]. We notice that as κ values increase convexity of the smile decrease. Which is the opposite effect of increasing σ value. Parameter θ is similar to parameter ν_0 since we see that we mainly operate on level of implied volatility by tuning these parameters.

Also ρ values and their effects can be easily understood. As a matter of fact, negative values increase in-the-money (ITM) Call and out-of-the-money (OTM) Put prices. This phenomenon has a comprehensive interpretation on equity market: a negative value of ρ implies that a strong rise in volatility will induce a fall on the underlying. Thus when the underlying is decreasing, prices of OTM Put rise because sell's options tend to become in-the-money options. And when the underlying is stable, OTM Put can be a protection to prevent a strong downgrade of the underlying. Symmetrical arguments can be underlined for positive values of ρ, and this is the case of commodities market: in a period of high volatility, a preference for insurance tends to increase OTM Call since we prefer to pay a

higher price for a supply than having any supply at all.

1.2 Discrete forms and specificities

If the benefits of using Heston's model are clear in a modeling point of view, we can wonder what are the impacts of this model choice especially on diffusion. The *Black-Scholes* model defines a log-normal distribution of the underlying. Parameters are constant, so there is no need to approximate paths to have a projected value of the underlying at a specific future date. The Heston model is different in a sense that it is path-dependent: the underlying's value at a future date depends on the variance's value at this future date. Therefore behavior of the spot process is much more complex, and the law unknown.

In this paragraph we are interested in approximating the theoretical law of our underlying. It could be useful for pricing and risk management to use Monte-Carlo methods.

In fact, the conditional law of the variance process is known and could be defined as a noncentral chi-squared distribution, but we still need to estimate behavior of the spot process with a discrete scheme. This is the cost to pay in order to have more market consistency in our modeling.

A particularity of a CIR process is that we know the conditional law of the solution. It is the consequence of expressing CIR diffusion's solution as a Bessel process, which leads to the following proposition:

Proposition 1. *$\forall s < t$, we have that $\nu(t)|\nu(s)$ follows a chi-squared distribution of $\frac{4\kappa\theta}{\sigma^2}$ degrees of freedom and a non-centrality parameter of $\nu(s)n(s,t)$, i.e*

$$\nu(t)|\nu(s) \sim \chi^2\left(\frac{4\kappa\theta}{\sigma^2}, \nu(s)n(s,t)\right),$$

with $n(s,t)$ defined as follow:

$$n(s,t) := \frac{4\kappa e^{-\kappa(t-s)}}{\sigma^2(1-e^{\kappa(s-t)})}.$$

We will consider in the following, Leif Andersen's method to compute efficiently random paths according to this property (see [6]). However we can compute paths that are consistant with CIR marginal distribution, we will see that this method could be slow compared to a discrete scheme. A major discussion that we will have in this section is: Is this method useful in practice? Or can we do the same with approximations and therefore less computational time?

We are looking for answers by studying four different schemes. These schemes are part of literature and especially the article of Milan Mrazek and Jan Pospisil (see [7]). This paper provides us with theoretical and numerical knowledge for practical implementation. We aim at validating our code and our understanding by reproducing results of this article. Therefore convenient schemes such as Alfonsi's implicit and explicit schemes for variance process (see [8]) and their combination with spot process schemes haven't been studied.

In the following, we consider a grid of time $\pi := \{t_0 < t_1 < ... < t_N\}$, $\Delta_n := t_{n+1} - t_n$ and we denote as $h := \max_{n\in\{0,...,N-1\}} \Delta_n$ its time step. We also introduce a useful notation $\Delta W_n := W_{t_n} - W_{t_{n+1}}$, which is the increment of a Brownian motion on the grid π. For convenient purpose we also use a Cholesky decomposition in order to obtain correlated path: $\Delta W_n^C := \rho\Delta W_n^\nu + \sqrt{1-\rho^2}\Delta W_n^S$.

The first scheme of interest is the Euler scheme. Both processes of the Heston model are discretized with a first order approximation. A correction is added to the variance process to guarantee positivity, notation $\cdot_+$ refers to $\max(\cdot, 0)$: this is an absorption correction.

$$\begin{cases} S^\pi_{t_{n+1}} = S^\pi_{t_n} \exp\left((r - \frac{\nu^\pi(t_n)_+}{2})\Delta_n + \sqrt{\nu^\pi(t_n)_+}\Delta W^C_n\right) \\ \nu^\pi(t_{n+1}) = \nu^\pi(t_n) + \kappa[\theta - \nu^\pi(t_n)_+]\Delta_n + \sigma\sqrt{\nu^\pi(t_n)_+}\Delta W^\nu_n \end{cases}. \tag{1.3}$$

We use log-spot process in our scheme because, as it is mentionned by Gatheral [4], in this case its prevents from the influence of higher order terms as we introduce an intermediate step using Ito calculus. The second scheme used in our work is the Milstein scheme where both processes of the Heston model are discretized with a second order approximation.

$$\begin{cases} S^\pi_{t_{n+1}} = S^\pi_{t_n} + S^\pi_{t_n}\left((r - \frac{\nu^\pi(t_n)_+}{2})\Delta_n + \sqrt{\nu^\pi(t_n)_+}\Delta W^C_n + \frac{\nu^\pi(t_n)_+}{2}(\Delta W^C_n)^2\right) \\ \nu^\pi(t_{n+1}) = \nu^\pi(t_n) - \frac{\sigma^2}{4}\Delta_n + \kappa[\theta - \nu^\pi(t_n)_+]\Delta_n + \sigma\sqrt{\nu^\pi(t_n)_+}\Delta W^\nu_n + \frac{\sigma^2}{4}(\Delta W^\nu_n)^2 \end{cases}. \tag{1.4}$$

A way to find a convenient discretization for the spot process often starts with the continuous formulation:

$$\ln(S_t) = \ln(S_u) + r(t-u) - \frac{1}{2}\int_u^t \nu(s)ds + \rho\int_u^t \sqrt{\nu(s)}dW^\nu_s + \sqrt{1-\rho^2}\int_u^t \sqrt{\nu(s)}dW_s,$$

$$\nu(t) = \nu(u) + \kappa\theta(t-u) - \kappa\int_u^t \nu(s)ds + \sigma\int_u^t \sqrt{\nu(s)}dW^\nu_s.$$

Based on this formulation the idea of Broadie and Kaya (see [9]) is to replace the formulation of the stochastic term given by the continuous variance process into the expression of the continuous spot process:

$$\int_u^t \sqrt{\nu(s)}dW^\nu_s = \frac{1}{\sigma}\left(\nu(t) - \nu(u) - \kappa\theta(t-u) + \kappa\int_u^t \nu(s)ds\right).$$

Therefore we only need to approximate one deterministic term and one stochastic term. According to Riemann's integral, a convenient approximation for the deterministic term would be:

$$\int_u^t \nu(s)ds \approx (t-u)\left[\gamma_1\nu(u) + \gamma_2\nu(t)\right],$$

where γ_1 and γ_2 are arbitrary weights that sum to one. We recall that if we choose $\gamma_1 = 1.0$ and $\gamma_2 = 0.0$, we are in the case of the Euler scheme. By default, we have considered in the following that both values are equal to 0.5. We can notice that doing this approximation on the deterministic term leads to simplification on the stochastic one. By construction, we know that:

$$\int_u^t \sqrt{\nu(s)}dW_s \sim \mathcal{N}\left(0, \int_u^t \nu(s)ds\right).$$

Since we have approximated the variance we could have an approximation of the stochastic term. A scheme denoted as E+M, uses this spot process dsicretisation. E for exact scheme on spot process and M for the Milstein scheme on variance process:

$$\begin{cases} S^\pi_{t_{n+1}} = S^\pi_{t_n} \exp\left(K^n_1 + K^n_2\nu^\pi(t_n)_+ + K^n_3\nu^\pi(t_{n+1})_+ + \sqrt{K^n_4\nu^\pi(t_n)_+ + K^n_5\nu^\pi(t_{n+1})_+}Z\right) \\ \nu^\pi(t_{n+1}) = \nu^\pi(t_n) - \frac{\sigma^2}{4}\Delta_n + \kappa[\theta - \nu^\pi(t_n)_+]\Delta_n + \sigma\sqrt{\nu^\pi(t_n)_+}\Delta W^\nu_n + \frac{\sigma^2}{4}(\Delta W^\nu_n)^2 \end{cases}. \tag{1.5}$$

In (1.5), Z is a standard gaussian random variable, and we define $K_1^n, K_2^n, K_3^n, K_4^n$ and K_5^n as:

$$K_1^n := \Delta_n \left(r - \frac{\rho\kappa\theta}{\sigma} \right), \quad K_2^n := \Delta_n \gamma_1 \left(\frac{\kappa\rho}{\sigma} - \frac{1}{2} \right) - \frac{\rho}{\sigma}, \quad K_3^n := \Delta_n \gamma_2 \left(\frac{\kappa\rho}{\sigma} - \frac{1}{2} \right) + \frac{\rho}{\sigma},$$

$$K_4^n := \Delta_n \gamma_1 \left(1 - \rho^2 \right), \quad K_5^n := \Delta_n \gamma_2 \left(1 - \rho^2 \right).$$

Schemes (1.3), (1.4) and (1.5) all use a discretization of the variance process. But as it is mentionned above, we know the conditional law of the variance process. Moreover a method to sample variance process paths from this exact law is presented in [6]. This method exploits the two-way behavior of the variance process.

$$\nu(t) = a(b + Z)^2, \text{ and for low variance } \nu(t) = \psi^{-1}(U; p, \beta). \tag{1.6}$$

where Z and U are respectively standard gaussian and standard uniform random variables and ψ is a cumulative distribution function (CDF) based on an asymptotic density:

$$\psi : x \mapsto p + (1-p)(1 - e^{-\beta x}), \text{ so } \psi^{-1} : u \mapsto \begin{cases} 0, & \text{if } u \leq p \\ \frac{1}{\beta} \ln\left(\frac{1-p}{1-u} \right), & \text{if } u > p \end{cases}.$$

This expression implies that p parameter is in $[0, 1]$. Finally all parameters a, b, p and β are computed by using first two moments that are known for $u < t$:

$$\mathbb{E}\left[\nu(t)|\nu(u)\right] := \theta + (\nu(u) - \theta)e^{-\kappa(t-u)},$$

$$\mathbb{V}\left[\nu(t)|\nu(u)\right] := \frac{\nu(u)\sigma^2 e^{-\kappa(t-u)}}{\kappa} \left(1 - e^{-\kappa(t-u)} \right) + \frac{\theta\sigma^2}{2\kappa} \left(1 - e^{-\kappa(t-u)} \right)^2.$$

We refer the interested reader to [6] to explicitly see formulas that link these moments with the above parameters. It remains to define a threshold which will help us to use one of the previous sampling expressions (1.6). Leif Andersen advises to pick a level in the interval $[1, 2]$, and specifies that this choice doesn't infer in the simulation. In the following we have fixed this threshold to 1.5. Expression (1.6) and the exact scheme used in (1.5) are part of a studied scheme that is denoted as Quadratic Exponential:

$$\begin{cases} S_{t_{n+1}}^\pi = S_{t_n}^\pi \exp\left(K_0^n + K_2^n \nu(t_n) + K_3^n \nu(t_{n+1}) + \sqrt{K_4^n \nu(t_n) + K_5^n \nu(t_{n+1})} Z \right) \\ \nu(t_{n+1}) = a^{(n)}(b^{(n)} + Z)^2, \text{ and for low variance } \nu(t_{n+1}) = \psi^{-1}(U; p^{(n)}, \beta^{(n)}) \end{cases}. \tag{1.7}$$

We precise that K_2^n, K_3^n, K_4^n and K_5^n are the same variables defined previously, however K_0^n is a modified term to take into account the martingale correction recommended in [6]. Dependency in n remains as this term is computed differently according to the sampling formula used to compute $\nu(t_n)$. We also underline the fact that parameters $a^{(n)}$, $b^{(n)}$, $p^{(n)}$ and $\beta^{(n)}$ depend on the value of $\nu(t_n)$ through the expression of the first two moments.

Now that we have seen several ways to approximate the law of our spot process, we are interested in quantifying approximations that have been done. The purpose of strong convergence and weak convergence is to measure the error introduced with our scheme.

Definition 1. *(Weak convergence) An approximating process Y^π of Y converges in the weak sense with order $\beta > 0$ if for each $g \in \mathcal{C}_p^{2\beta+1}$ there exists a finite constant K and a positive constant h_0 such that:*

$$\varepsilon^W(\pi) := |\mathbb{E}\left[g(Y_T^\pi) - g(Y_T)\right]| \leq Kh^\beta$$

for any time discretisation with maximum step size $h \in (0, h_0)$ (see [10]).

Definition 2. *(Strong convergence) An approximating process Y^π of Y converges in the strong sense with order $\gamma > 0$ if there exists a finite constant K and a positive constant h_0 such that:*

$$\varepsilon^S(\pi) := \mathbb{E}\left[|Y_T^\pi - Y_T|\right] \leq Kh^\gamma$$

for any time discretisation with maximum step size $h \in (0, h_0)$ (see [10]). We often use $\mathbb{E}\left[\sup_{r\in\pi} |Y_r^\pi - Y_r|\right]$ as an approximation of Kh^γ.

It is important to mention that it is very complicated to evaluate strong and weak convergences of our schemes of interest (1.3). (1.4), (1.5) and (1.7) to the theoretical law. This is a consequence of our dependency in square root variance in our spot process and our variance process for schemes (1.3), (1.4) and (1.5). Due to the fact that the square root function isn't lipschitz, estimation of errors induced with our spot process schemes are far more difficult.

Nevertheless, we can consider the following lemma from Alfonsi (see [8]), to compute numerically strong convergence for each scheme.

Lemma 1. *Let us consider an approximating process Y^π that converges towards Y in the sense:*

$$\mathbb{E}\left[\sup_{r\in\pi} |Y_r^\pi - Y_r|\right] \underset{h\to 0}{\to} 0$$

where h is the step of the discrete grid π. Then for any $\alpha > 0$, $K \geq 0$,

$$\mathbb{E}\left[\sup_{r\in\pi} |Y_r^\pi - Y_r|\right] \leq Kh^\gamma \Leftrightarrow \mathbb{E}\left[\sup_{r\in\pi} |Y_r^\pi - Y_r^{2\pi}|\right] \leq Kh^\gamma$$

where 2π grid is twice thiner than π.

Several points need to be underlined from the above lemma in order to compare various spot process schemes for the Heston model. First we need the schemes to be computed with the same randomness since we want to estimate an average behavior under the same probability measure. Furthermore, we need to compare two vectors of same lenght, so for our spot process $S_r^{2\pi}$ for $r \in \pi$ means that we take only values computed at a discrete time with an even index. We are thus able to approximate strong convergence order for each schemes. Results can be seen on figure [1.3], where the average is computed on 10000 paths for each time-step h.

Figure [1.3] shows that as long as h is thin enough, schemes tend to converge towards the same order. However means are estimated and should be presented with a confidence interval, we can't identify a scheme which provides a better approximation spot process law in the Heston model. But we can also be interested in evaluating their performance in a pricing perspective.

We present here a second test which is detailed in [7]. For a given range of vanilla call options, we can estimate prices of these options using a Monte-Carlo method over 1000

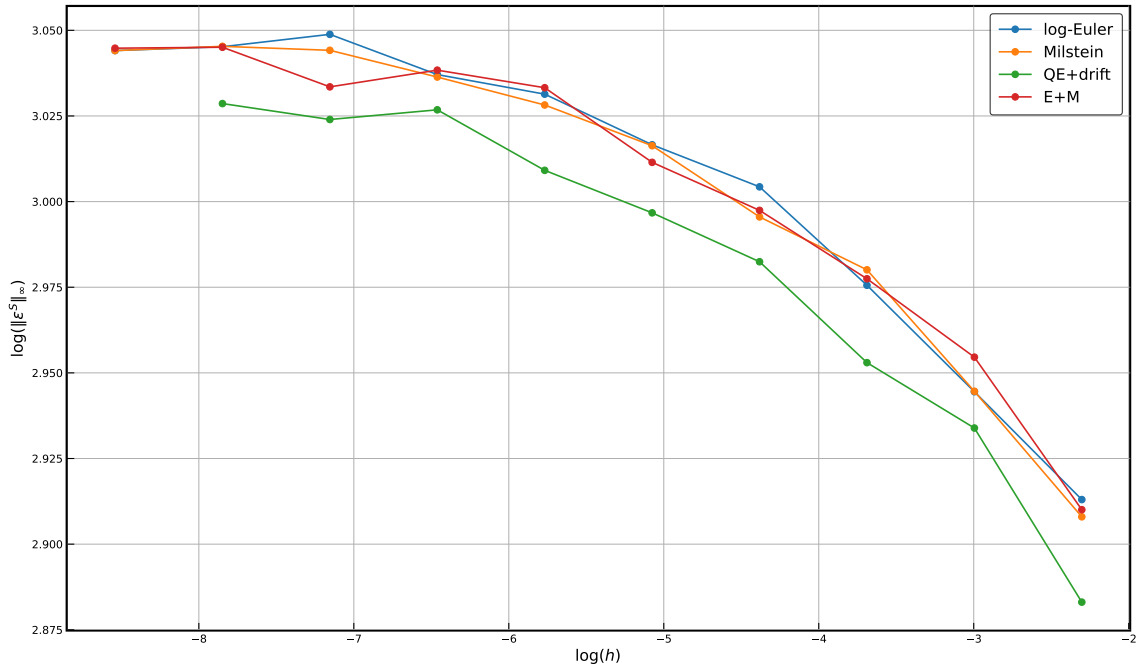

Figure 1.3: Evolution of $\|\varepsilon^S\|_\infty := \mathbb{E}\left[\sup_{r\in\pi} |S_r^\pi - S_r^{2\pi}|\right]$ for different time-step h of π.

paths with a time-step h. Prices computed can then be compared to a reference price. In the case of the Heston model we have an explicit formula for Call options that we will detail in §2.1, and for a given set of parameters of this model we can use this formula as a reference. Therefore we can define an error ε:

$$\varepsilon := \frac{1}{N}\sum_{n=1}^{N}\left|\frac{C_n - \hat{C}_n}{C_n}\right|.$$

where N is the number of options, C_n and $\hat{C}_n$ denote respectively the reference price and the Monte-Carlo computed price. However our Monte-Carlo computed price is sensitive to the generated paths, which makes ε a random variable. We can compute a confidence interval by repeating this computation of ε a hundred times with different sets of paths. The above method is repeated with different values for the time-step h in order to evaluate the accuracy of the computed prices.

The set of parameters is the same that is used in the paper [7]. We generate paths and reference prices according to values presented in table 1.1.

ν_0	κ	θ	σ	ρ
0.02497	1.22136	0.06442	0.55993	−0.66255

Table 1.1: Parameters of the Heston model, $r = 0.00207$, $S_0 = 7962.31$.

Results can be seen on figure [1.4]. We remark that Quadratic Exponential scheme produces the poorest results especially when h is small. More precisely results are more or less constant, and there is no convergence of this scheme according to h. As mentionned before this scheme can be slow and doesn't seem to be relevant from a pricing perspective. Contrary to Quadratic Exponential scheme, other schemes of interest are associated to decreasing values as long as h is decreasing, and moreover the schemes converge towards

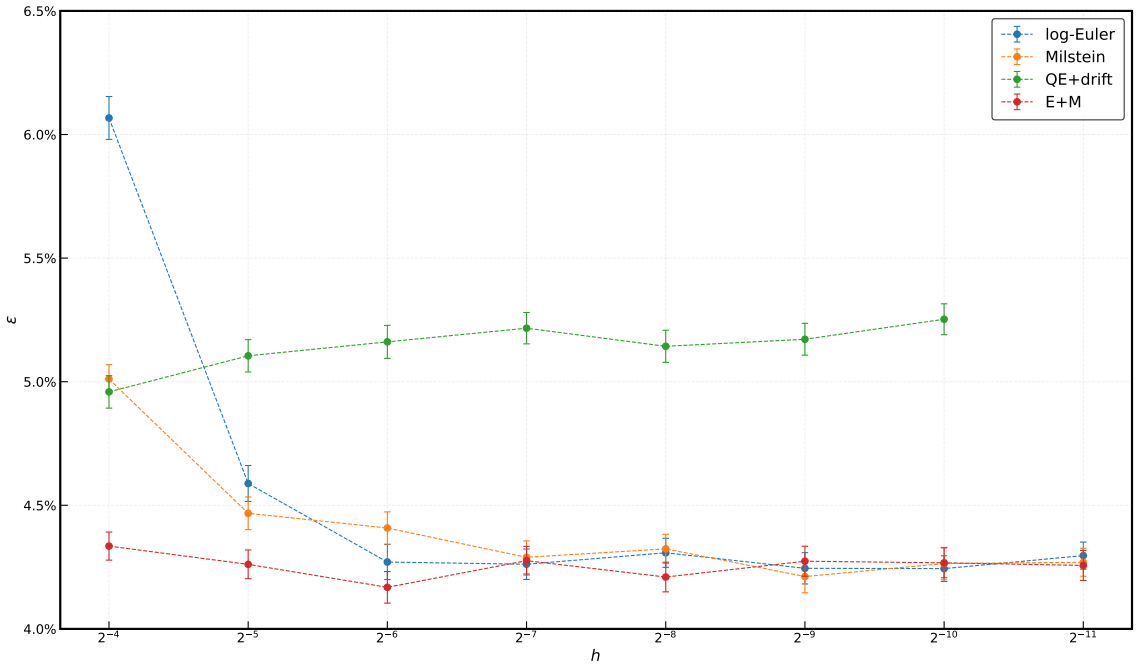

Figure 1.4: Evolution of ε with a confidence level of 95% for different time-step h of π.

the same error value. We note also that for large values of h, the log-Euler scheme provides the largest error whereas E+M scheme provides the smallest error.

This result differs from results observable in paper [7]. In the article Quadratic Exponential is the scheme that leads to the smallest pricing error in average. Note also that the authors use a computed price over 100000 paths whereas we compute prices with 1000 paths. An observation that should be made is that these schemes operate the same way as they are presented in the article, and E+M scheme should be used for large values of h as it leads to the lowest value of ε.

We conclude that results presented in this section don't allow us to clearly see the benefits of using Leif Andersen's method to simulate variance process.

1.3 Valuation Partial Differential Equation

We have seen that in comparison with the *Black-Scholes* model, it is more complicated to simulate the Heston model and to evaluate approximations quality due to a the lack of results on law and convergence. In this section we are interested in meeting the Heston valuation partial differential equation, and we will see that there is also some difference with GMB model.

The first problem which was underlined in §1.1 is that our market is incomplete. In order to give a price C to options with (1.2), an alternative is to complete our market with another asset of price P. Under a risk-neutral measure $\mathbb{Q}$ we know how to define the price at a time t of an option of maturity T and payoff ϕ (see [11]):

$$C_t := \mathbb{E}^{\mathbb{Q}}\left[\frac{B_T}{B_t}\phi(S_T)|\mathcal{F}_t\right]. \tag{1.8}$$

As long as S and ν are driven by Brownian motions, the assumption of a Markovian framework is plausible. Thus we can write:

$$C_t = \frac{B_T}{B_t}\mathbb{E}^{\mathbb{Q}}\left[\phi(S_T)|S,\nu\right] =: U(S,\nu,t),$$

with U a smooth function with respect to all arguments. We also know that processes $\tilde{S}_t := B_t S_t$ and $\tilde{C}_t := B_t C_t$ are $\mathbb{Q}$-martingales. In the following we consider that the market is arbitrage-free, it allows us to justify the existence of at least one risk-neutral probability measure $\mathbb{Q}$.

The price of an option is exactly the cost to replicate this option or the price to generate $\phi(S_T)$ at time T. To evaluate this cost, we can consider a portfolio Π with our option of interest C, the asset S and another option P that depends on volatility. As specified above, P will denote the price of the option used to complete the market. We want to evaluate changes in this portfolio during a time dt (see Gatheral [4]). At any date t we consider that our portfolio value is exactly C_t. We suppose that $\{\alpha_t^0\}_{t\in[0,T]}$, $\{\alpha_t^1\}_{t\in[0,T]}$ and $\{\alpha_t^2\}_{t\in[0,T]}$ are proportions hold for the assets $\{C_t\}_{t\in[0,T]}$, $\{S_t\}_{t\in[0,T]}$ and $\{P_t\}_{t\in[0,T]}$ respectively. Then the portfolio value and self-financing condition could be written as:

$$\Pi_t = \alpha_t^0 C_t - \alpha_t^1 S_t - \alpha_t^2 P_t. \tag{1.9}$$

$$d\Pi_t = \alpha_t^0 dC_t - \alpha_t^1 dS_t - \alpha_t^2 dP_t. \tag{1.10}$$

Additionally C and P should be interpreted as smooth functions according to (1.8), and since S and ν are solution to problem (1.2) we can apply Ito's lemma. Stochastic coefficients in diffusion aren't impacted by change of measure, hence we can write:

$$d<S,S>=\nu S^2 dt, \quad d<\nu,\nu>=\nu\sigma^2 dt \quad \text{and} \quad d<S,\nu>=\sigma\nu S\rho dt.$$

$$dC_t = \mathcal{L}C dt + \frac{\partial C}{\partial S}dS + \frac{\partial C}{\partial \nu}d\nu, \quad dP_t = \mathcal{L}P dt + \frac{\partial P}{\partial S}dS + \frac{\partial P}{\partial \nu}d\nu.$$

with $\mathcal{L}$ the differential operator associated with finite differential term in Ito's lemma formulas. To create a riskless replication strategy, terms in dS and in $d\nu$, so randomness, in equation (1.10) need to balance, thus we have:

$$\frac{\partial C}{\partial S} = \alpha_t^1 + \alpha_t^2\frac{\partial P}{\partial S}, \quad \frac{\partial C}{\partial \nu} = \alpha_t^2\frac{\partial P}{\partial \nu}.$$

Given relation (1.9) and by considering that α_t^1 and α_t^2 are defined as above, we can write:

$$d\Pi_t = r\left[C_t - \left(\frac{\partial C}{\partial S} - \alpha_t^2\frac{\partial P}{\partial S}\right)S - \alpha_t^2 P_t\right]dt.$$

All these arguments can be used to write a new formulation of equation (1.10):

$$\frac{\mathcal{L}C - rC + rS\frac{\partial C}{\partial S}}{\frac{\partial C}{\partial \nu}} = \frac{\mathcal{L}P - Pr + rS\frac{\partial P}{\partial S}}{\frac{\partial P}{\partial \nu}}.$$

As we know that C and P are prices of two different assets, quantities above can't be equal as prices differ. It proves the existence of a function f which depends on S, ν and t and satisfies the equation above. Because we are interested in giving the price C of the option, we consider the expression:

$$\mathcal{L}C - rC = -f(S,\nu,t)\frac{\partial C}{\partial \nu} - rS\frac{\partial C}{\partial S}. \tag{1.11}$$

A possible interpretation could be that because f is the coefficient of the term $\frac{\partial C}{\partial \nu}$ and rS is the coefficient of the term $\frac{\partial C}{\partial S}$, f is the drift function of the process ν under a risk-neutral measure. We can try to be more precise by writting, in a risk-neutral framework, the diffusion of S as we know that $\tilde{S}_t$ is a martingale:

$$\frac{dS_t}{S_t} = \left(\mu - \frac{\mu - r}{\sqrt{\nu(t)}}\sqrt{\nu(t)}\right) dt + \sqrt{\nu(t)} dW_t^1.$$

where $\{W_t^1\}_{t\geq 0}$ is a Brownian motion under a risk-neutral measure. Note that the drift term is effectively equal to r. We deliberately keep μ, the drift under the physical measure, and the volatility term in $\sqrt{\nu(t)}$ to find the equivalent in the diffusion of ν. Intuitively a proper form to define f would be:

$$f(\nu, t) := \kappa[\theta - \nu(t)] - \lambda\sigma\sqrt{\nu(t)}, \quad \lambda \in \mathbb{R}.$$

Therefore we find the partial differential equation (1.12) of the original Heston article. A more mathematical approach using Girsanov's theorem can also be consider. In fact, by using Cholesky decomposition it is possible to develop the calculations and define an equivalent of (1.2) under a risk-neutral measure. The same partial differential equation can be found by using Ito's lemma and the fact that $\tilde{S}_t$ is a martingale. We refer the reader to §A.1 for details.

$$\frac{1}{2}\nu S^2 \frac{\partial^2 U}{\partial S^2} + \rho\sigma\nu S \frac{\partial^2 U}{\partial S \partial \nu} + \frac{1}{2}\sigma^2 \nu \frac{\partial^2 U}{\partial \nu^2} + rS\frac{\partial U}{\partial S} + f(\nu, t)\frac{\partial U}{\partial \nu} - rU + \frac{\partial U}{\partial t} = 0. \tag{1.12}$$

Chapter 2

Calibration, context and implementation

In this chapter, we provide formulas and numerical methods with the objective to turn our theoretical model in an usable tool. If we recall our objective from the previous section, we want to find values of our parameters that are consistent with the market. In fact we can evaluate how our model differs from the market, especially european vanilla options, with a semi-closed formula, that we will introduced in the first section. Because of the semi-closed nature of this formula, it is not directly computable. In the second section, we present numerical methods to efficiently approximate this formula. As this section will be more practical we will also present the code architecture. Finally, we will add market consistency to our work by introducing our calibration procedure through literature discussions.

2.1 Pricing of European option: a semi-closed formula

Let us consider an underlying with the Heston risk-neutral dynamic:

$$\begin{cases} dS_t = rS_t dt + \sqrt{\nu(t)} S_t dW_t^{(1)} \\ d\nu(t) = \kappa[\theta - \nu(t)]dt + \sigma\sqrt{\nu(t)} dW_t^{(2)} \\ d < W^{(1)}, W^{(2)} >_t = \rho dt \end{cases} . \tag{2.1}$$

We can in fact introduce a formula for Call options.

Proposition 2. *Price $C_0(T,K)$ at time $t = 0$ of a european Call of strike K and maturity T verifies:*

$$C_0(T,K) = S_0 \tilde{\mathbb{Q}}(S_T > K) - Ke^{-rT}\mathbb{Q}(S_T > K). \tag{2.2}$$

Proof. Since the definition of $\mathbb{Q}$-price is $C_0(T,K) := \mathbb{E}\left[e^{-rT}(S_T - K)^+\right]$ as mentionned in (1.8), with $\mathbb{E}$ the expected value with respect to $\mathbb{Q}$. We can introduce a new measure $\tilde{\mathbb{Q}}$ such as :

$$\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}} = \frac{S_T}{\mathbb{E}[S_T]} = \frac{S_T}{S_0 e^{rT}}, \quad L_T := \frac{S_T}{S_0 e^{rT}}$$

where last equality holds from the fact that process $\{\tilde{S}_t\}_{t\in[0,T]}$ is a $\mathbb{Q}$-martingale so $e^{-rT}\mathbb{E}[S_T] = S_0$. We denote as $\tilde{\mathbb{E}}$ the expected value with respect to $\tilde{\mathbb{Q}}$.

We can develop the following expression:

$$\begin{aligned}\mathbb{E}\left[(S_T-K)_+\right] &= \mathbb{E}\left[S_T 1_{S_T>K}\right] - K\mathbb{Q}(S_T>K),\\ &= \mathbb{E}\left[\frac{S_T}{L_T}1_{S_T>K}L_T\right] - K\mathbb{Q}(S_T>K),\\ &= \tilde{\mathbb{E}}\left[\frac{S_T}{L_T}1_{S_T>K}\right] - K\mathbb{Q}(S_T>K) \quad \textit{(Bayes formula)},\\ &= S_0 e^{rT}\tilde{\mathbb{Q}}(S_T>K) - K\mathbb{Q}(S_T>K).\end{aligned}$$

Multiplying with the discount factor we get that $C_0(T,K)$ satisfies (2.2). □

Since Call-Put parity (2.3) is a model free relationship, we can reduce our research to a formula of the analytical computation of (2.2).

$$S_0 - Ke^{-rT} = C_0(T,K) - P_0(T,K). \tag{2.3}$$

Fourier pricing formula helps defining an explicit formulation of the probabilities according to the log-spot caracteristic function.

Proposition 3. *Denote $\phi_T(z) := \mathbb{E}[e^{izX_T}]$, $X_T = \ln(S_T)$ and $k = \ln(K)$, we have:*

$$\begin{cases}\mathbb{Q}(S_T>K) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathcal{R}e\left[\frac{e^{izk}\phi_T(z)}{iz}\right]dz\\ \tilde{\mathbb{Q}}(S_T>K) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathcal{R}e\left[\frac{e^{izk}\phi_T(z-i)}{iz\phi_T(-i)}\right]dz\end{cases}. \tag{2.4}$$

Then our last step to define a semi-closed formula is about determining the expression of the log-spot caracteristic function in the case of the Heston model. Model's diffusion is explicitly affine and Markovian, and in this case we know the form of the log-spot caracteristic function.

Proposition 4. *For the Heston model, ϕ_T is such that $\forall z \in \mathbb{R}_+$:*

$$\phi_T(z) = e^{C(\tau,z)\theta + D(\tau,z)\nu(t) + iz(X_t + rT)}$$

with $\tau := T - t$ where t denotes the valuation date, which involves that $\nu(t)$ and X_t are known. Particularly, by defining:

$$\omega := -\frac{1}{2}z(i+z), \quad y := \kappa - \rho\sigma iz, \quad c := \frac{\sigma^2}{2}, \quad p := \sqrt{y - 4\omega c}, \quad y_\pm := \frac{y \pm p}{2c}, \quad g := \frac{y_-}{y_+}$$

we have that:

$$C(\tau,z) := \kappa\left[\tau y_- - \frac{1}{c}\ln\left(\frac{1-ge^{-p\tau}}{1-g}\right)\right], \quad D(\tau,z) := \frac{1-e^{-p\tau}}{1-ge^{-p\tau}}y_-.$$

We note that this expression differs form the one presented in [2]. Heston's formulation incurs a discontinuity as it is presented in [12]. The above expression presents an advantage as it is numerically more robust, and is equivalent to the one presented by Gatheral.

On one hand, proposition 3 is often written with one intergral as it prevents from adding another numerical uncertainty with a second intergral. In fact $\phi_T(z)$ is a caracteristic function associated to a risk-neutral density which can be defined by inverse Fourier transform (IFT):

$$\phi_T(\omega) = \int_{-\infty}^{+\infty} e^{i\omega x}q(x)dx \quad \overset{(IFT)}{\Rightarrow} \quad q(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega x}\phi_T(\omega)d\omega.$$

On the other hand we have by recalling definition (1.8) and denoting as x the random part of spot process that:

$$C_0(T,K) = e^{-rT}\mathbb{E}_{\mathbb{Q}}\left[S_0 e^{rT+x} | S_0 e^{rT+x} \geq K\right] - e^{-rT}\mathbb{E}_{\mathbb{Q}}\left[K | S_0 e^{rT+x} \geq K\right].$$

Since we know the expression of the risk-neutral density, by defining $l := \frac{Ke^{-rt}}{S_0}$, we have by identification that:

$$\tilde{\mathbb{Q}}(S_T > K) = \int_l^{+\infty} e^x \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega x}\phi_T(\omega)d\omega\right)dx.$$

Using Fubini's theorem and dissociating case with possible values of l, one can show that we have an expression that can be computed with residual theorem (see [Appendix A.3]). We thus obtain:

$$\tilde{\mathbb{Q}}(S_T > K) = 1 + \frac{1}{2\pi}\int_{-\infty}^{+\infty} \phi_T(\omega)\frac{e^{-i(\omega+i)l}}{i(\omega+i)}d\omega.$$

Because we need real values, we have that the expression of the $\mathbb{Q}$-price can be rewritten as :

$$C_0(T,K) = S_0 - \frac{e^{-rT}K}{2} - e^{-rT}K\left[\frac{1}{\pi}\int_0^{+\infty}\left(A(\omega)+B(\omega)\right)\frac{1}{1+\omega^2}d\omega\right], \tag{2.5}$$

$$A(\omega) := \mathcal{R}e(\phi_T(\omega)) + \frac{\mathcal{I}m(\phi_T(\omega))}{\omega}, \quad B(\omega) := \mathcal{I}m(\phi_T(\omega)) - \frac{\mathcal{R}e(\phi_T(\omega))}{\omega}.$$

In the following, combination of (2.2) and (2.4) will be denoted as Heston's formula, whereas (2.5) will be denoted as Attari's formula. A crucial part of the implementation will be how to compute integral in each formula and this will be discussed in the next section.

A such technique implies that according to a continuous relationship, we introduce a computing error using discrete points. A possibility is also to consider directly a discrete approximated formulation of the problem. In our case this is convenient using the Fast Fourier Transform algorithm, a method that has been developed by Carr and Madan (see [13]). The particularity of this method is to compute prices on a whole range of strikes with a time complexity that is linearithmic $\mathcal{O}(n\log(n))$. It is also a well-known algorithm that can be used in many languages standard library by giving the discrete Fourier transform of the problem. Formulas provided by Carr and Madan are the following approximations $\hat{C}_0$ of C_0:

$$\hat{C}_0(T,e^{k_n}) = \frac{e^{-\alpha k_n}}{\pi}\mathcal{R}e\left[\sum_{j=0}^{N-1} e^{-2i\pi\frac{jn}{N}}e^{-i\underline{k}z_j}\psi_T^{\alpha}(z_j)w_j\right], \tag{2.6}$$

$$\hat{C}_0(T,e^{k_n}) = \frac{1}{\sinh(\alpha k_n)\pi}\mathcal{R}e\left[\sum_{j=0}^{N-1} e^{-2i\pi\frac{jn}{N}}e^{-i\underline{k}z_j}\gamma_T^{\alpha}(z_j)w_j\right]. \tag{2.7}$$

The grid k_n is defined as $k_n := \underline{k} + \frac{2\pi n}{Nh}, \forall n \in \{0,...,N-1\}$ with $\underline{k} = \ln(S_0) + rT - \frac{\pi}{h}$ and $z_j = jh$. Weights w_j refer to Simpson's rule for quadrature: $w_j := \frac{h}{3}(3 + (-1)^{j+1} - \mathbf{1}_{\{j=0\}\cup\{j=N-1\}})$.

In the article, Carr and Madan provide numerical values for parameter $N = 4096$, the dumping factor $\alpha = 1.5$ and the step $h = 0.25$ and will be used in our implementation.

Functions ψ_T and γ_T are both dependent of ϕ_T but aren't the same. In fact, formula (2.7) is supposed to outperfom formula (2.6) especially for the computation of OTM option price. Therefore, we will often refer to (2.6) as the FFT formula and (2.7) as the FFTOTM formula.

By considering that we have an efficient method to compute Heston's formula and Attari's formula, we have then four ways to give a price to european vanilla options based on our Heston model parameters, through ϕ especially.

2.2 Implementation issues and overview

We have discussed previously the relationship between our Heston model and the price of an european vanilla option. We have also seen that methods such as FFT and FFTOTM are ready to use since it can be computed, in the case of the FFT formula for example, by providing the entry vector:

$$\left(e^{-i\underline{k}z_0}\psi_T^{\alpha}(z_0)w_0, \quad e^{-2i\pi\frac{n}{N}}e^{-i\underline{k}z_1}\psi_T^{\alpha}(z_1)w_1, \cdots, \quad e^{-2i\pi\frac{(N-1)n}{N}}e^{-i\underline{k}z_{N-1}}\psi_T^{\alpha}(z_{N-1})w_{N-1}\right)^T.$$

However, in order to use the Heston and Attari formulas, we need to precise their discrete approximation. The particularity of these formulas is that it involves integrals over an unbounded interval. In literature, this kind of calculation are often done by using a factor that ensures integrability of our function of interest on large values of our semi-axis. This is the case of the Gauss-Laguerre method where the function integrated is e^{-x} times our initial integrand of interest.

But this is not the only way to proceed, in fact Jäckel, Kahl and Lord (see [12] and [14]) show that we can use a change of variable which helps us to change the integral over the positive semi-axis into an integral over the unit interval $[0; 1]$. By doing this modification, we are able to use a wide range of quadrature rules. According to the work of Schmelzle (see [15]), we have decided to study the adaptative Gauss-Kronrod procedure mainly because it is presented as the best compromise between computation time and accuracy.

In this section, we will introduced Gauss-Laguerre and Gauss-Kronrod quadrature methods and discussed about code issues. To add some context, we are interested in computing:

$$I := \int_0^{+\infty} f(x)dx,$$

where f depends on the characteristic function of our Heston model and doesn't differ that much whereas we consider the Heston formula or the Attari formula. As mentionned before, Gauss-Laguerre rule allows us to compute:

$$\int_0^{+\infty} e^{-x}g(x)dx \approx \sum_{i=1}^{N} w_i g(x_i),$$

where $(w_i)_{1\leq i\leq N}$ and $(x_i)_{1\leq i\leq N}$ are respectively weights and nodes of the quadrature. We can therefore compute I:

$$I := \int_0^{+\infty} f(x)dx = \int_0^{+\infty} e^{-x}h(x)dx \approx \sum_{i=1}^{N} w_i h(x_i),$$

where $h(x) := e^x f(x)$ given values for weights and nodes. But by definition, x_i is the i-*th* root of the Laguerre polynom $L_N := \frac{e^x}{N!}\frac{d^N}{dx^N}(e^{-x}x^N)$. Weights are then computed by the following formula:

$$w_i = \frac{x_i}{[(N+1)(L_{N+1}(x_i))]^2}, \quad \forall i \in \{1, ..., N\}.$$

We note that we reach small values of w_i, around 10^{-100}, as soon as N is larger than a hundred. With our code we use *Python*'s package *NumPy* to obtain numerical values of weights and nodes according to the value of N wanted. In our case, given different numerical values of Heston model parameters, we estimate the number N needed by fixing a tolerance on the residual: we increase N as long as the absolute difference between the computation with $N-1$ and N is above a fixed tolerance threshold. We observed, for the Heston formula, that in some cases we reach $N = 186$ without being under the threshold, but we can't compute the integral with a larger value of N as *NumPy* doesn't allow it. Unless otherwise stated, we will use the Gauss-Laguerre quadrature rule with $N = 180$. This means that in the case of the Heston formula and because two approximations are computed, we have 360 calls to the characteristic function of our model to compute a price, given values for maturity T and strike K.

We have also been interested in exponentially-fitted quadrature rule (see [16]) in order to reduce number of calls to ϕ_T and speed-up the procedure. Benefits from this method come from the fact that nodes and weights are computed according to the oscillatory nature of the integrand. So nodes and weights depend of the function of interest whereas with Gauss-Laguerre rule nodes and weights are the same independently from the function to be intergated. But this method implies to solve an ill-conditionned system to obtain weights values. Therefore we can't compute values for N larger than 6 which leads to poor approximations in terms of accuracy.

Gauss-Kronrod adaptive procedure is a little bit different. As it is adaptive, integration points are selected depending where information is lost, so where it is difficult to integrate. It is usually done by subdivising the original interval and this is the case for Gauss-Kronrod quadrature rule. Literature often present Gauss-Kronrod quadrature rule on bounded interval, and even if this kind of procedure can be used for semi-infinite interval, as it is done in *QUADPACK* (see [17]), we present here a change of variable in order to modify I directly. In their article Kahl and Jäckel [12] proposed the following change:

$$I := \int_0^{+\infty} f(x)dx = \int_0^1 f(x(y))dy, \quad x(y) = -\frac{\ln(y)}{C_\infty}, \quad C_\infty = \frac{\sqrt{1-\rho^2}}{\sigma}\left(\nu_0 + \kappa\theta T\right).$$

Note that C_∞ constant only depends on model parameters ν_0, κ, θ, σ and ρ as well as maturity option T.

Now regarding the method, Gauss-Kronrod quadrature rule is the combination of two quadrature results. We first evaluate I with a Gauss quadrature rule of n points, it could be Gauss-Legendre or Gauss-Jacobi for example. Then we evaluate I with the first n points and with $n+1$ Kronrod's rule points. So both approximations have n common nodes:

$$I \approx G^{(n)} := \sum_{i=1}^{n} w_i^G f \circ x(y_i^G), \quad I \approx K^{(2n+1)} := \sum_{i=1}^{2n+1} w_i^K f \circ x(y_i^K).$$

Then if the absolute difference between $G^{(n)}$ and $K^{(2n+1)}$ is above a pre-defined tolerance threshold, we split the intergation interval in two equal parts. We can do the computation

FFT methods		Quadrature methods		
FFT	FFTOTM	HESTON	ATTARI	
FFT algorithm		GAUSS-LAG.	GAUSS-LAG.	GAUSS-KRON.
NumPy		*NumPy*	*NumPy*	*SciPy*
fft()		*laggauss()*	*laggauss()*	*quad_vec()*
$N = 4096$ $\alpha = 1.5$ $h = 0.25$		$N = 180$	$N = 180$	$2n + 1 = 21$ tol. $= 1e - 9$

Table 2.1: Different numerical methods implemented for each formula.

of $G^{(n)}$ and $K^{(2n+1)}$ on both parts and so on until we reach the threshold. This confirmed the adaptive nature of this method. In our case we have decided to use a Gauss-Kronrod adaptive 21-points rule, which means that we compute during each iteration $G^{(10)}$ and $K^{(21)}$. We have also chosen an arbitrary absolute tolerance of 10^{-9}. We have used *Python*'s package *SciPy* to use *quad_vec()* method as it is already implemented and in a convenient way that we will discuss later. We propose the table 2.1 to recap all numerical methods implemented with parameter choice and according to our theoretical formula. The method *laggaus()* allows us to have access to nodes and weights of Gauss-Laguerre quadrature rule, we then compute the sum as it is presented above. Whereas with *fft()* and *quad_vec()* methods we used built-in functions to obtain an approximation by either giving a vector which corresponds to discrete Fourier transform for *fft()* or a callable *Python*'s function for *quad_vec()*.

However, we need to underline the fact, as stated before, that FFT methods compute prices for a range of strikes predetermined by our numerical value of N. Issues are met as soon as we try to answer to the question: how can we compare FFT and quadrature methods?

In fact, the first problem is that comparison is impossible at this step because of the methods themselves: for a call to our pricing method with Heston formula we have a numerical value whereas with FFT formula we have a set of prices that matchs log-strike's calculation grid of the method. So both methods don't return the same thing, and consequently aren't comparable at least in the algorithmic and complexity point of view.

On one hand, to tackle this problem we have decided to use a vectorize approach of our code. The idea is that given a set of maturities, a set of strikes and eventually a term structure of the risk-free interest rate of the market, we are able, independently from the method, to return a surface of european vanilla options prices. For example with FFT methods, the entry vector is replaced by a matrix with each row i refering to a discrete Fourier transform according to maturity T_i. In the case of quadrature methods we can also perform numerical intergation simultaneously for each element (T_i, K_j) of the matrix, which is very simple by using *quad_vec()*.

On the other hand, now that we compute a surface of prices we still can't compare numerical results as FFT algorithm has his own (log-) strikes grid to perform well. This incurs to compare two surfaces of different dimensension. Our approach was to build a linear dependence between each prices of the FFT grid in order to compute interpolated price associated to any grid of strikes. With this solution by providing the same set of option's caracteristics, we obtain a well-sized surface of prices either with FFT or

quadrature methods.

Since diffusion and vanilla pricing methods have been presented, we introduce, in this more practical section, the code environment where all these methods have been developed. Exiom's quantitative library named **FinX** is a *Python*'s module developed in *Python* by the quantitative team. Various models and tools have already been implemented to cope with several issues from pricing derivatives to curve stripping for example. A large part of the library is usable, and we aim through this work to add the Heston model and its tools as a new content for users. We propose a representation of the library's architecture with figure [2.1].

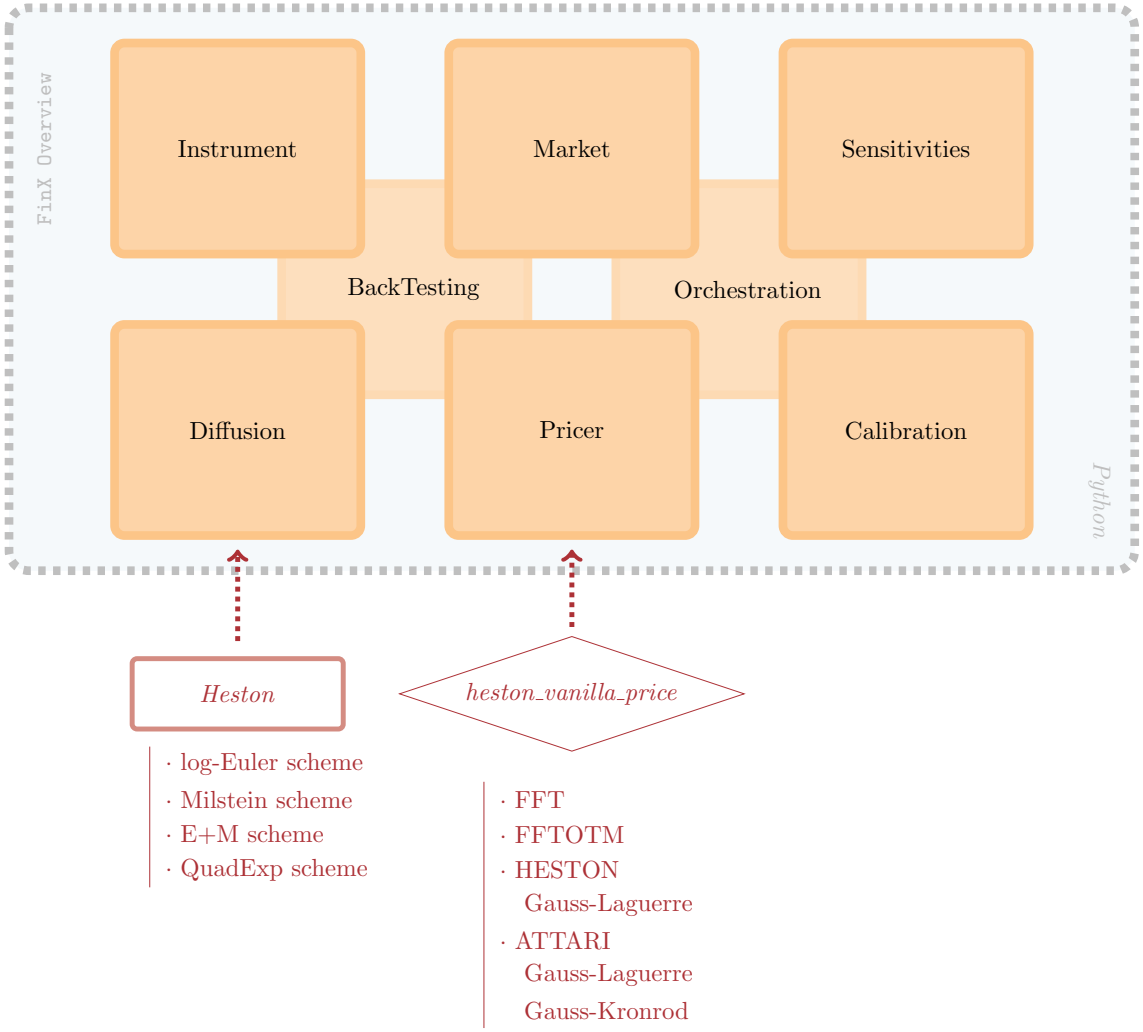

Figure 2.1: Overview of Exiom's quantitative library *FinX*.

A qualitative test is then to compare our code for the four formulas to results of well-known open source quantitative library *Quantlib* ([18]). We create surfaces of prices for four expiries and for each formula. We are able through an inversion method to obtain surfaces of implied volatilities. Through *Quantlib* we can also obtain a surface of implied volatilities associated to a Heston model. We repeat this for different numerical values of our model's parameters and we compare the surface on four maturity slices: 1 month, 6 months, 1 year and 2 years. Results are presented in figure [2.2].

On figure [2.2] only Gauss-Laguerre quadrature rule is used. We observe on this figure that FFT methods are less robust on small expiries, where the convexity is higher. We can also notice some irregularities due to the linear reconstruction. Results are in overall

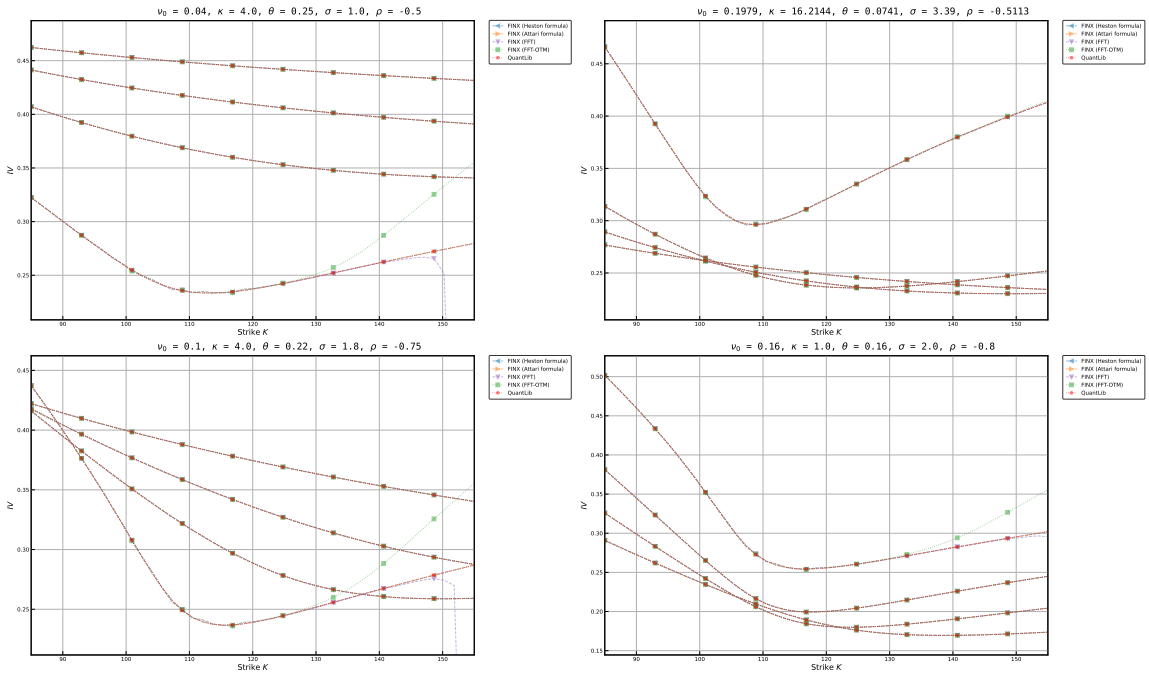

Figure 2.2: Comparison between **FinX** implemented formulas and *Quantlib* with $r = 0.0$ and $S_0 = 100$.

satisying as we succeed in matching *Quantlib* surface.

2.3 Calibration methods

Generating prices of european vanilla options with our model's parameters is feasible and it has been developed in sections above. Then again if we are concerned by modeling questions, we aim to find numerical values for our model's parameters that reflect market behaviour. This is done by using quoted and observed market prices of options and by comparing with model's prices in an optimization procedure.

According to literature several methods can be used, and we have been interested in results obtain by Mikhailov and Nögel (see [19]) and Cui *and al.* (see [20]), as well as heuristics proposed in [21]. Among particularities exposed in these references, we can remark that authors preference is to create an optimization procedure on implied volatilities instead of using directly market's and model's prices. Note that price and implied volatility are directly linked with *Black-Scholes* formula so both problems are equivalent. An answer to this choice lies in the fact that a small difference between two implied volatilities induces a larger gap between the two associated prices. We have also done this choice in such a way that in this section we aim to solve a minimization problem with the following objective function:

$$\min_{\Theta \in \mathcal{S}} \mathcal{J}(\Theta) = \sum_t \sum_k \vartheta_{BS}(K_k, T_t) \left(\sigma^{market}_{implied}(K_k, T_t) - \sigma^{model}_{implied}(\Theta; K_k, T_t) \right)^2, \tag{2.8}$$

where Θ is the vector of model's parameters, ϑ_{BS} is the *Black-Scholes* Vega, $\sigma^{market}_{implied}$ is the implied volatility of a market price, so the target and $\sigma^{model}_{implied}$ is the implied volatility of a model price. The *Black-Scholes* Vega is used as a weighting function. As we aim at replicating liquid options, in our case european vanilla options, we prefer to reduce uncertainty of the price given by our model for near ATM options than to well replicate

far OTM options. This is exactly the behaviour of the Vega as it is never higher than the Vega of the related ATM option. Also, unless otherwise stated, we will define $\mathcal{J}$ on the whole surface. It is a choice that could be criticized as long as we know with [4] that it is hard to fit the observed surface for short expiries. It is also mentionned in [21] that we can often fix a maturity, in the definition of $\mathcal{J}$ to obtain better results for calibration, and this is an alternative that can be explored. This discussion is the consequence of the lack of knowledge about $\mathcal{J}$ convexity on the space of all possible Θ. For this reason, we can't justify the existence and uniqueness of a global minimum Θ^*.

In the absence of theoretical results to solve problems like (2.8), it could be complicated to find in practice one or several local minimums that could be great candidates to be a global minimum. A possibility is to start by restraining the space of search $\mathcal{S}$ and this is discussed in [20] and in [21]. If we consider that $\Theta := (\nu_0, \kappa, \theta, \sigma, \rho)^T$ then $\mathcal{S}$ is defined as $]0;1] \times \mathbb{R}_+ \times]0;1] \times \mathbb{R}_+ \times [-1;0]$ for equity options for example. A heuristic proposed in [21] is to estimate ν_0 parameter by the square of the implied volatility of the most ATM observable options. It is an interesting estimation because we have seen in §1.1, that ν_0 influences the level of the smile, and this is an alternative choice that we have made. Note that other estimations are mentionned in [20], but it can highly depend on market and data. Thus in practice in the following $\mathcal{S}$ is a subset of $\mathbb{R}^4$ with the mentionned estimation of ν_0, and by defining $\tilde{\Theta} := (\kappa, \theta, \sigma, \rho)^T$ it leads to the following problem:

$$\min_{\tilde{\Theta} \in \mathbb{R}_+ \times]0;1] \times \mathbb{R}_+ \times [-1;0]} \mathcal{J}(\tilde{\Theta}) = \sum_t \sum_k \vartheta_{BS}(K_k, T_t) \left(\sigma_{implied}^{market}(K_k, T_t) - \sigma_{implied}^{model}(\tilde{\Theta}; K_k, T_t) \right)^2. \tag{2.9}$$

Now that we have introduced theoretical issues, we want to detail our calibration procedure in practice. We agreed on following the choice of Mikhailov and Nögel which is to use a stochastic optimization algorithm to find a good initial guess of an optimal candidate and then use this candidate as the initial guess of a local optimization algorithm. This is done to tackle one of the main drawbacks of a local algorithm: it is highly dependent of the starting value. As an example for some very oscillating functions starting from two diffent values can leads to two different local minimum by gradient descent. A stochastic optimization helps in order to start at a better initial point, however it costs a lot of computational time as it explores a large number of paths and points.

FinX library has a dependency with *SciPy* so we have decided to use built-in algorithm of this module. Our stochastic optimization will be performed by an implementation of *Differential Evolution* algorithm. For the local optimization we will provide results based on two different algorithm: *Nelder-Mead* algorithm and *Trust-region Reflective* algorithm. We provide a pseudo-code algorithm that explains our calibration procedure.

Algorithm 1 (Calibration of the Heston model)

1. Estimation of parameter ν_0
2. $\Theta^{DE} \leftarrow$ Differential_Evolution over the bounded problem (2.9)
3. $\Theta^* \leftarrow$ Nelder_Mead or Trust_Region_Reflective over the bounded problem (2.9) with initial starting value Θ^{DE}.

Chapter 3

Calibration numerical results

In this chapter we present calibration results. The first section is about testing calibration procedure on observed data generated from a Heston model. Bias of our model are presented in the second section. As soon as we introduce real market data, we have to cope with phenomena that incur a loss of precision in our calibration procedure.

3.1 Tests cases with a generated surface of price

For this test cases, we generate a surface of prices with a Heston model and the Attari formula with a Gauss-Kronrod quadrature. This means that $\sigma_{implied}^{market}$ are in this case $\sigma_{implied}$ of a target model. In fact we have built surfaces of prices for two sets of parameters:

dataset 1	$\nu_0 = 0.05$	$\kappa = 0.5$	$\theta = 0.05$	$\sigma = 0.2$	$\rho = -0.4$
dataset 2	$\nu_0 = 0.05$	$\kappa = 3.0$	$\theta = 0.05$	$\sigma = 0.4$	$\rho = -0.57$

For each model, we generate a surface with the same caracteristics: time and strike grids are fixed to create a surface of 150 options and market parameters, spot $S_0 = 100$ and risk-free rate $r = 0.01$, are constant. Our purpose is then to find out which computation method is better to meet by the calibration procedure the above parameters. Note that, in this section and because we use generated data, we set $\vartheta_{BS}(K_k, T_t) = 1, \forall k, \forall t$.

We have decided to compare the Heston formula computed with Gauss-Laguere quadrature rule, the Attari formula computed with Gauss-Kronrod quadrature rule and both FFT methods. To evaluate the performance of our formulas, we have defined three measures of interest:

$$
\begin{aligned}
MSE(\Theta^*) &:= \frac{1}{N_k N_t} \sum_{t=1}^{N_t} \sum_{k=1}^{N_k} \left(\sigma_{implied}^{market}(K_k, T_t) - \sigma_{implied}^{model}(\Theta^*; K_k, T_t) \right)^2, \\
RMSE(\Theta^*) &:= \sqrt{MSE}, \quad \text{and} \\
MAE(\Theta^*) &:= \max_{(k,t)} |\sigma_{implied}^{market}(K_k, T_t) - \sigma_{implied}^{model}(\Theta^*; K_k, T_t)|.
\end{aligned}
$$

We present results with standard deviation, based on a twenty values set, as stochastic algorithm may provide different local minimums between each execution. Unless stated otherwise this will be the case for each calibration results presented. To see differences obtain with each local optimization algorithm, we decide to execute calibration with *Trust Region Reflective* algorithm for the market generated with the first dataset and we use

Nelder-Mead algorithm for the market generated with the second dataset. A methodology of our tests is proposed with algorithms 2 and 3.

Algorithm 2 (Test case with **dataset 1**)

for p = 1,...,20 **do**
1. $\Theta^{*,p} \leftarrow$ Calibration procedure with Trust_Region_Reflective and $\vartheta_{BS}(K_k, T_t) = 1$
2. $MSE^p, RMSE^p, MAE^p \leftarrow$ Compute measures with $\Theta^* = \Theta^{*,p}$

end for

$\Theta^* \leftarrow \frac{1}{20}\sum_{p=1}^{20} \Theta^{*,p}$

$MSE \leftarrow \frac{1}{20}\sum_{p=1}^{20} MSE^p$

$RMSE \leftarrow \frac{1}{20}\sum_{p=1}^{20} RMSE^p$

$MAE \leftarrow \frac{1}{20}\sum_{p=1}^{20} MAE^p$

Algorithm 3 (Test case with **dataset 2**)

for p = 1,...,20 **do**
1. $\Theta^{*,p} \leftarrow$ Calibration procedure with Nelder_Mead and $\vartheta_{BS}(K_k, T_t) = 1$
2. $MSE^p, RMSE^p, MAE^p \leftarrow$ Compute measures with $\Theta^* = \Theta^{*,p}$

end for

$\Theta^* \leftarrow \frac{1}{20}\sum_{p=1}^{20} \Theta^{*,p}$

$MSE \leftarrow \frac{1}{20}\sum_{p=1}^{20} MSE^p$

$RMSE \leftarrow \frac{1}{20}\sum_{p=1}^{20} RMSE^p$

$MAE \leftarrow \frac{1}{20}\sum_{p=1}^{20} MAE^p$

For the market generated with **dataset 1**, results are presented in tables 3.1 and 3.2, and also through figure [3.1]. We recall that target prices are generated with the Attari formula and Gauss-Kronrod quadrature rule. According to table 3.1, it is with this formula that we obtain optimal parameters that are the closest from target values. However, we note that Heston formula provides the best results in table 3.2. We can also underline poor performance of FFT methods. Parameters obtained with FFT formula largely differ from market values and even if numerical values obtained with FFTOTM are closer, the formula isn't as robust as formulas computed with a quadrature rule (see table 3.2).

The overall behaviour that we observe with calbiration results over **dataset 1** is that MAE value is reached, for each formula, for points with the smallest maturity, except for FFTOTM formula whose associated point has the greatest maturity.

For the market generated with **dataset 2**, results are presented in tables 3.3 and 3.4, and also through figure [3.2]. Remarks made about **dataset 1** are still valid for **dataset 2**. Attari's formula provides the closest set of parameters but it needs to be balance with the fact that market prices have been generated with this formula. And as a matter of fact we still have better results on our measures with the Heston formula. We still show the lack of robustness of FFT methods, but we can underline the fact that standard deviation is lower in tables 3.3 and 3.4 than in tables 3.1 and 3.2. Further development will be focused on determining whether this values stability is a consequence of *Nelder-Mead* algorithm or a model dependent result.

A satisying observation is that in both cases we have exploitable results. Particularly with a quadrature rule, we are able to estimate a target surface with a mean-square error from 28 to 112 basis points over 150 option prices and with a maximum absolute error

	HES-GL	ATT-GK	FFT	FFTOTM
ν_0	0.04940	0.04940	0.04940	0.04940
κ	0.39956	0.39993	1.0145	0.37873
(s.d.) κ	$2.6703e-3$	$2.5537e-3$	$3.230e-2$	$8.8173e-2$
θ	0.05184	0.05184	0.04968	0.05600
(s.d.) θ	$6.8424e-5$	$6.3788e-5$	$2.4134e-4$	$2.0739e-2$
σ	0.19375	0.19381	0.29736	0.18380
(s.d.) σ	$2.6902e-4$	$3.6653e-4$	$1.5490e-3$	$7.1816e-3$
ρ	-0.39375	-0.39373	-0.36475	-0.40304
(s.d.) ρ	$2.2849e-4$	$2.0591e-4$	$1.7627e-3$	$5.5450e-3$

Table 3.1: Optimal parameters values for **dataset 1**.

	HES-GL	ATT-GK	FFT	FFTOTM
MSE	28.18	28.19	1370	82.75
(s.d.) MSE	0.026	0.031	6.823	80.17
RMSE	5.309	5.309	37.01	8.577
(s.d.) RMSE	0.002	0.003	0.092	3.031
MAE	14.50	14.48	107.7	22.40
(s.d.) MAE	0.154	0.187	0.275	7.041

Table 3.2: Values of different measures for **dataset 1** in basis points (bps).

from 14 to 35 basis points. Now we can wonder if we can obtain satisying results with these procedure on real market data?

	HES-GL	ATT-GK	FFT	FFTOTM
ν_0	0.04827	0.04827	0.04827	0.04827
κ	2.6169	2.6178	4.9999	2.4061
(s.d.) κ	$1.921e-3$	$2.161-3$	$1.02e-4$	$2.446e-2$
θ	0.05075	0.05075	0.05075	0.05061
(s.d.) θ	$4e-6$	$4e-6$	$6.4e-5$	$2.4e-5$
σ	0.37834	0.37847	0.68520	0.34313
(s.d.) σ	$1.20e-4$	$1.36-4$	$2.752e-3$	$2.103e-3$
ρ	-0.56003	-0.56004	-0.53483	-0.57206
(s.d.) ρ	$3.0e-5$	$3.0e-5$	$1.325e-3$	$3.94e-4$

Table 3.3: Optimal parameters values for **dataset 2**.

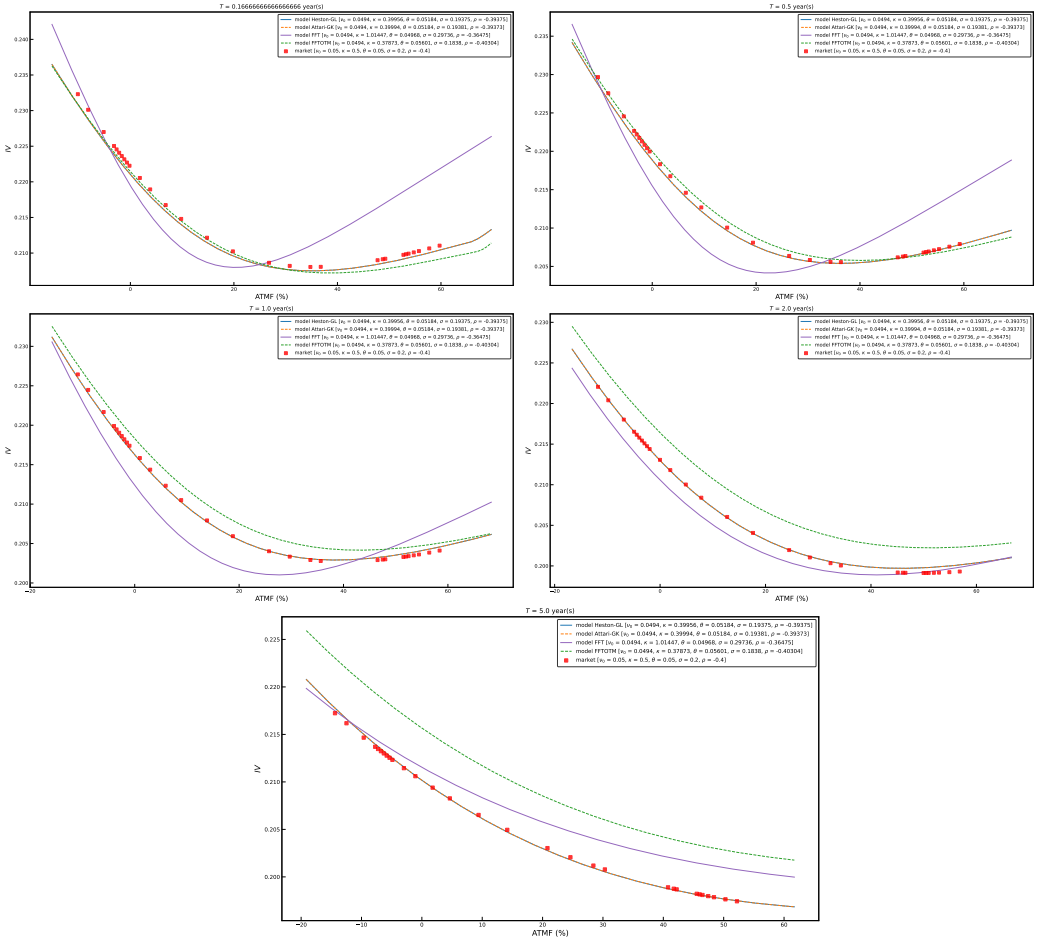

Figure 3.1: Comparison between formulas for calibration procedure over **dataset 1**.

	HES-GL	ATT-GK	FFT	FFTOTM
MSE	112.7	112.7	3832	247.7
(s.d.) MSE	0.003	0.003	1.302	8.882
RMSE	10.62	10.62	61.90	15.73
(s.d.) RMSE	$1.3e-4$	$1.5e-4$	0.0105	0.2821
MAE	34.87	34.82	221.6	51.36
(s.d.) MAE	0.050	0.051	0.027	1.607

Table 3.4: Values of different measures for **dataset 2** in basis points (bps).

3.2 Results on $CAC40$ Index

Due to results observable in §3.1, we want to evaluate how calibration performance is affected when we use real market data. In this section, $\sigma_{implied}^{market}$ is the implied volatility of $CAC40$ Index from the 21st July 2023. Data are provided by *Bloomberg* as well as strikes and maturities of the surface. We have also access to spot $S_0 = 7425.94$ and its forward values with respect to each maturity. This allows us to implicit the term-structure of the risk-free rate.

Thus our surface is composed with 16 maturities, from 1 month to 10 years, and 40 strikes. Initially we have access to 18 maturities, from 1 week to 10 years but we have

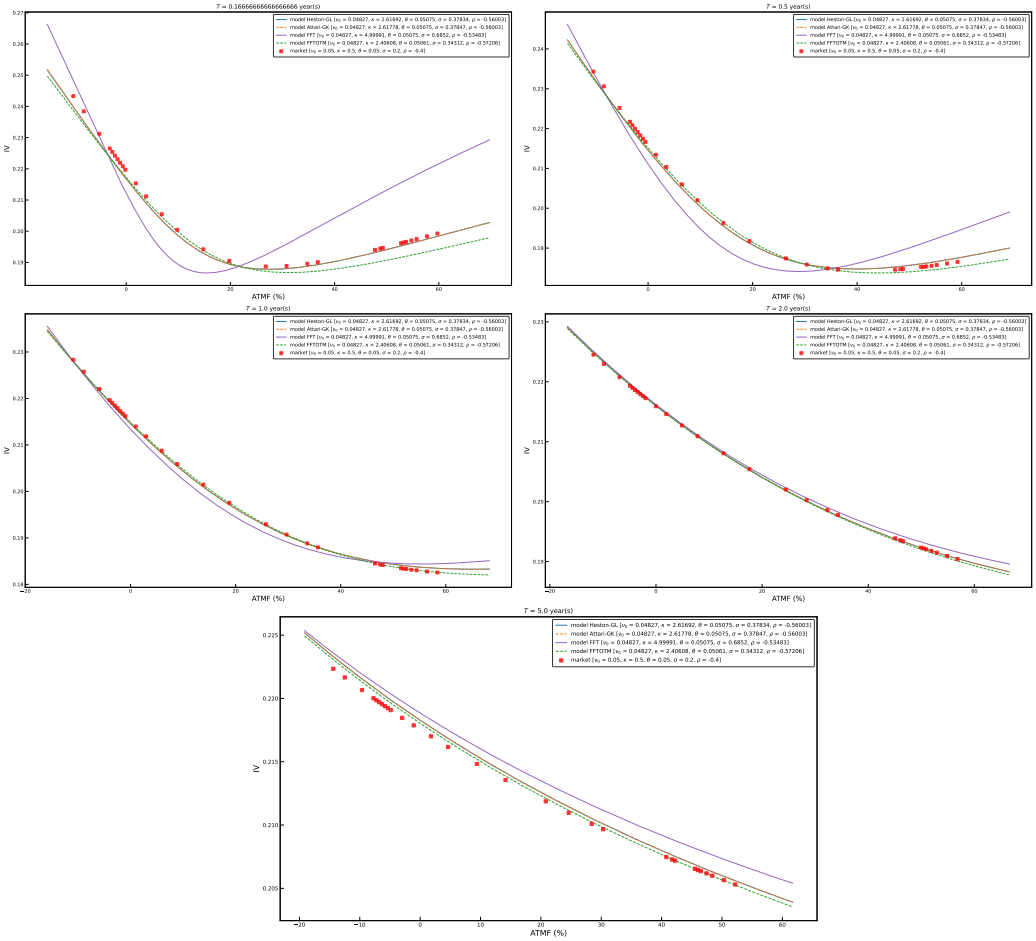

Figure 3.2: Comparison between formulas for calibration procedure over **dataset 2**.

decided to delete 1 and 2 weeks data from our calibration procedure because of numerical instability in our computation of model's vanilla price.

We use the same procedure as mentionned in algorithms 2 and 3 to compare both local optimization methods over the same dataset but with this time a real computation of $\vartheta_{BS}(K_k, T_t)$. Results are presented in tables 3.6, 3.7, 3.8 and 3.7. We also show the absolute relative error of each surface with the market surface with figures [3.3] and [3.4].

We remark that with *Nelder-Mead* algorithm we have better approximations of parameters. By better we mean that according to tables 3.6 and 3.7 standard deviation of values are lower with *Nelder-Mead* algorithm. Moreover this is also with this algorithm that we obtain the lowest error values according to tables 3.8 and 3.9 and figures [3.3] and [3.4].

Regression can also be evaluated with the computation of the coefficient of determination of each model using the mean values previously computed. Results are presented in table 3.5. Although absolute relative error is higher in average with FFT methods, according to figures [3.3] and [3.4], R^2 values are very close especially for *Nelder-Mead* algorithm.

Overall, calibration results confirm observation made in §3.1. Formulas using a quadrature rule outperform FFT methods. It is also indisputable that *Nelder-Mead* algorithm's performance is better than *Trust Region Reflective*. Nevertheless we can wonder if a mean-square error of nearly 653 basis points over the whole surface and a maximum absolute error of 134 basis points are acceptable?

Several approximations can be made in order to build an interpolated surface on a thiner grid in order to avoid granularity and to increase calibration performance. This is the object of the next chapter.

	HES-GL	ATT-GK	FFT	FFTOTM
R^2 with *Trust Region Reflective*	0.9768	0.9766	0.9661	0.9668
R^2 with *Nelder-Mead*	0.9869	0.9869	0.9830	0.9830

Table 3.5: Coefficient of determination with optimal parameters values over $CAC40$ Index data.

	HES-GL	ATT-GK	FFT	FFTOTM
ν_0	0.01880	0.01880	0.01880	0.01880
κ	1.2197	1.2215	1.2152	1.2051
(s.d.) κ	$2.394e-2$	$2.495e-2$	$2.605e-2$	$3.253e-2$
θ	0.05023	0.05023	0.05054	0.05057
(s.d.) θ	$7.631e-5$	$8.626e-5$	$1.112e-4$	$1.041e-4$
σ	0.67932	0.68075	0.70585	0.70024
(s.d.) σ	$1.323e-2$	$1.391e-2$	$1.539e-2$	$2.134e-2$
ρ	-0.62940	-0.62891	-0.60449	-0.60514
(s.d.) ρ	$1.972e-3$	$2.289e-3$	$3.515e-3$	$3.467e-3$

Table 3.6: Optimal parameters values for $CAC40$ Index with *Trust Region Reflective* algorithm.

	HES-GL	ATT-GK	FFT	FFTOTM
ν_0	0.01880	0.01880	0.01880	0.01880
κ	1.0822	1.0816	1.0365	1.0368
(s.d.) κ	$2.931e-3$	$3.278e-3$	$4.402e-3$	$3.368e-3$
θ	0.05137	0.05138	0.05195	0.05194
(s.d.) θ	$2.374e-5$	$1.6016e-5$	$2.513e-5$	$2.688e-5$
σ	0.61776	0.61743	0.62053	0.62063
(s.d.) σ	$1.263e-3$	$1.713e-3$	$2.346e-3$	$1.609e-3$
ρ	-0.66082	-0.66084	-0.63914	-0.63915
(s.d.) ρ	$1.46285e-4$	$1.9535e-4$	$1.833e-4$	$1.447e-4$

Table 3.7: Optimal parameters values for $CAC40$ Index with *Nelder-Mead* algorithm.

	HES-GL	ATT-GK	FFT	FFTOTM
MSE	1091	1101	1540	1511
(s.d.) MSE	45.10	54.06	108.3	122.0
RMSE	33.02	33.17	39.22	38.85
(s.d.) RMSE	0.683	0.811	1.373	1.532
MAE	181.1	181.9	215.1	213.2
(s.d.) MAE	3.380	4.172	6.789	7.619

Table 3.8: Values of different measures for $CAC40$ Index with *Trust Region Reflective* algorithm in basis points (bps).

	HES-GL	ATT-GK	FFT	FFTOTM
MSE	653.4	653.1	817.2	817.1
(s.d.) MSE	1.452	1.838	4.124	2.812
RMSE	25.56	25.56	28.59	28.58
(s.d.) RMSE	0.028	0.036	0.072	0.049
MAE	134.3	134.2	160.9	160.9
(s.d.) MAE	0.333	0.474	0.578	0.388

Table 3.9: Values of different measures for $CAC40$ Index with *Nelder-Mead* algorithm in basis points (bps).

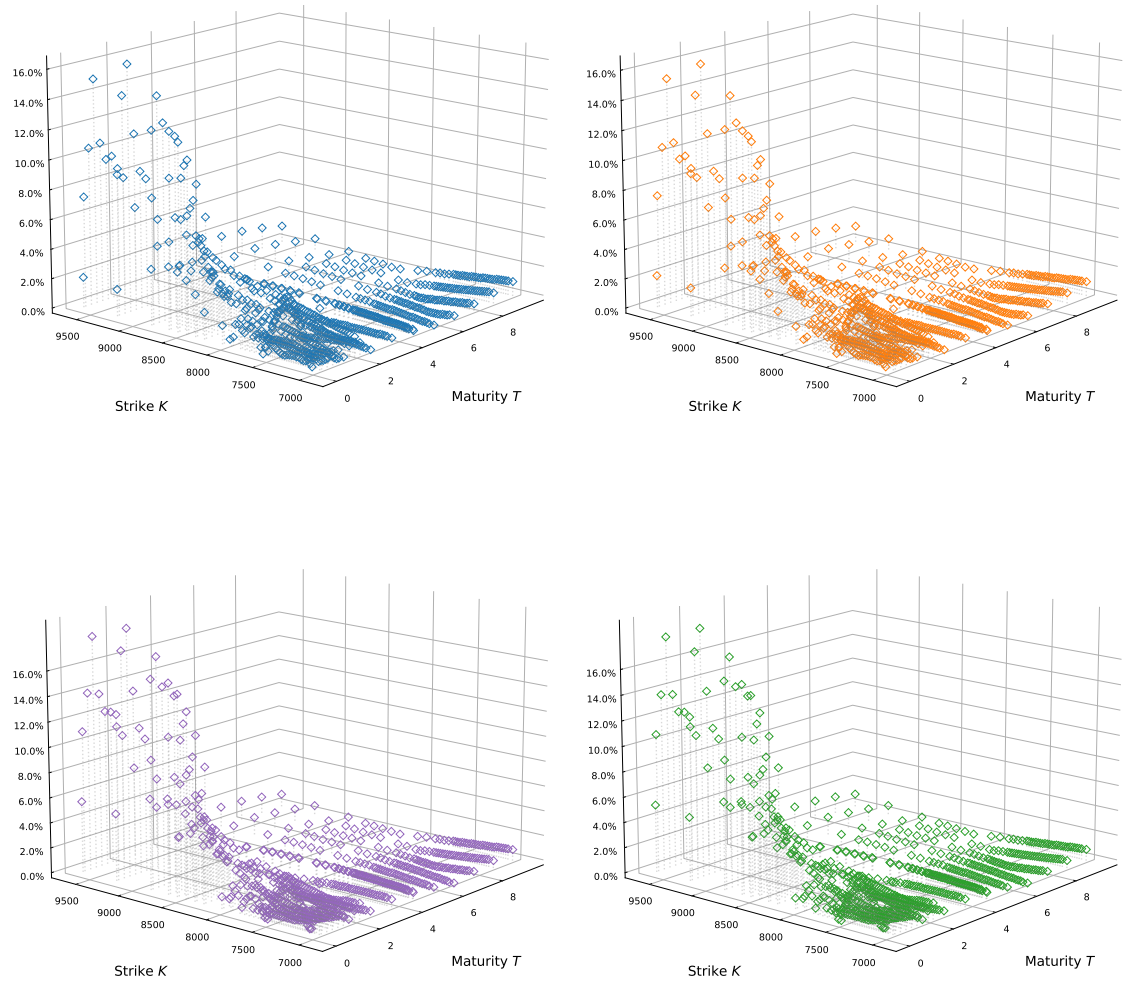

Figure 3.3: Absolute relative error of each point for each formula *Heston-GL, Attari-GK, FFT and FFTOTM* from left to right with *Trust Region Reflective* algorithm.

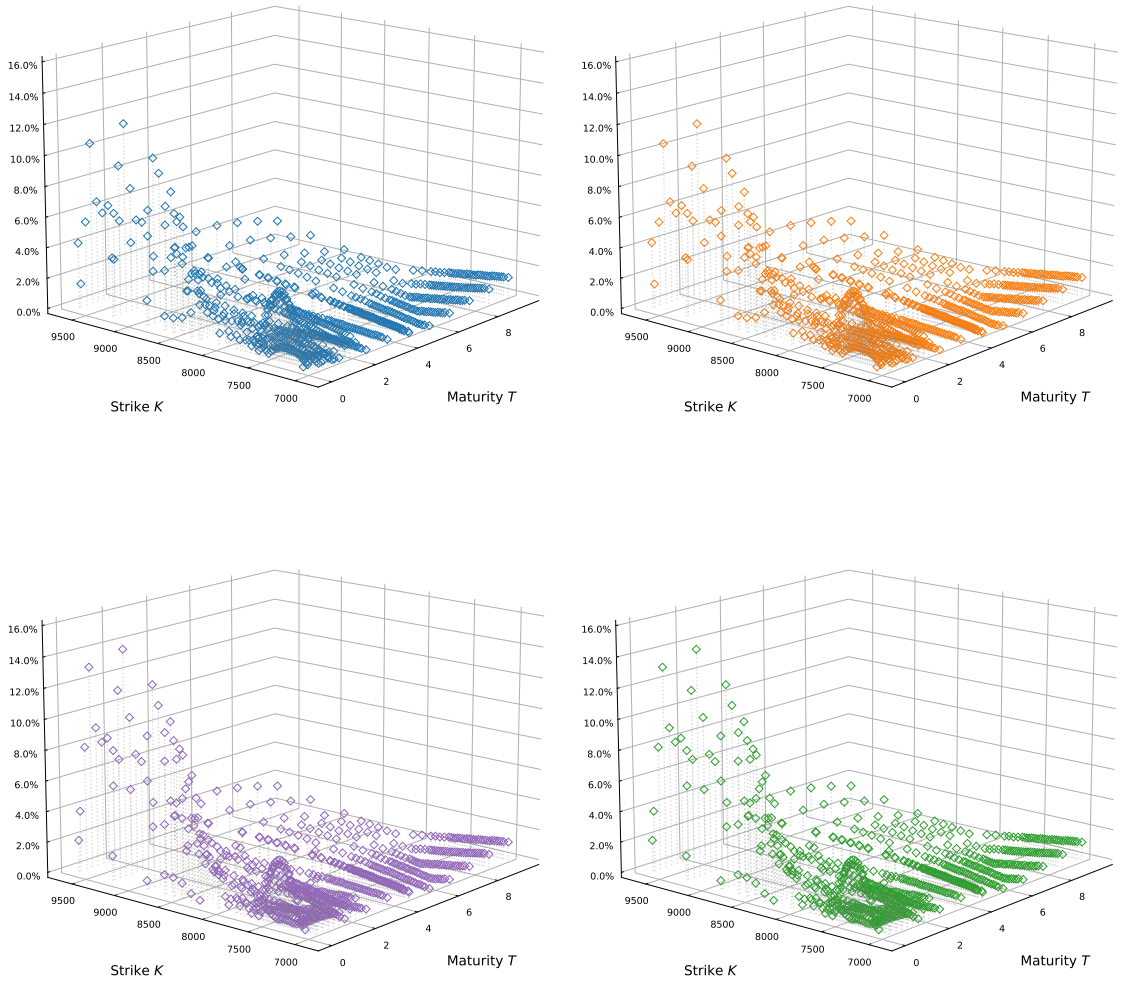

Figure 3.4: Absolute relative error of each point for each formula *Heston-GL, Attari-GK, FFT and FFTOTM* from left to right with *Nelder-Mead* algorithm.

Chapter 4

Improvement of calibration results

In this chapter, we aim at investigating the benefits of new calibration procedures. We first present a methodology to interpolate prices from market data. The idea is that by adding points we can create a surface which is smoother and may be easier to fit. The calibration procedure is modified and results are presented in the second section. We have also investigated the use of variance swaps. Although these product aren't vanilla options, it helps to get information on market's future variance. A future variance that can be used as a target for the future variance of our model. It induces also a new calibration procedure which is presented in details as well as results in the third and the fourth sections of this chapter.

4.1 Stochastic volatility inspired (SVI)

In this section we present the Stochastic Volatility Inspired parametrization. First introduced in [4], this model aims at explaining the volatility smile for each quoted maturities, with the formula:

$$\sigma^2_{implied}(k) := a + b\left(\rho(k-m) + \sqrt{(k-m)^2+\sigma^2}\right), \tag{4.1}$$

with k the log-strike value and a, b, ρ, σ and m are maturities dependent parameters. With some change of variables, this formula is equivalent to parametrization presented in [22]. The crucial point in our work with equation (4.1) is that given market's fitted parameters, we can get the value of the implied volatility of an unquoted option. We can therefore create a market-like smile with more points. The question now is the following: does a generated point from (4.1) respect market's behaviour? Or more generally, does SVI model prevent from static arbitrage? This question is highly sensitive in our work as we want to use an interpolated SVI surface as a target for our model in a further calibration procedure.

In the following, we introduce some key points established in [22] in order to answer that question. We can start by recalling the definition of a surface which is free of static arbitrage given by Gatheral and Jacquier.

Definition 3. *A volatility surface is free of static arbitrage if and only if the following conditions are satisfied:*

(i) It is free of calendar spread arbitrage.

(ii) Each time slice is free of butterfly arbitrage.

Definition 4. *A volatility surface σ is free of calendar spread arbitrage if*

$$\partial_t \left(\sigma^2_{implied}(k,t)t\right) \geq 0, \quad \forall (k,t) \in \mathbb{R} \times \mathbb{R}^*_+.$$

where k is the log-moneyness.

Definition 5. *A smile is said to be free of butterfly arbitrage if the corresponding density is non-negative.*

In order to observe whether there is butterfly arbitrage or not, we prefer to study, as mentionned in [22], the following function g with the result given by the lemma 2:

$$g(k) := \left(1 - \frac{kw'(k)}{2w(k)}\right)^2 - \frac{w'(k)^2}{4}\left(\frac{1}{w(k)} + \frac{1}{4}\right) + \frac{w''(k)}{2}, \tag{4.2}$$

with

$$w(k) := t\sigma^2_{implied}(k,t),\ w'(k) = t\left(b\rho + \frac{k-m}{\sqrt{(k-m)^2+\sigma^2}}\right),\ w''(k) = \frac{t\sigma^2}{((k-m)^2+\sigma^2)^{\frac{3}{2}}}.$$

Lemma 2. *A smile is free of butterfly arbitrage if and only if $g(k) \geq 0 \quad \forall k \in \mathbb{R}$ and $\lim_{k\to+\infty} d_+(k) = -\infty$.*

However the work of Jacquier and Gatheral doesn't lead to theoretical bounds on parameters that prevent from butterfly arbitrage, we can still estimate the associated density to conclude about positivity. We can also observe the absence of calendar spread arbitrage by showing that the quantity $\sigma^2_{implied}(k,t)t$ increases with t. Typically we take $\sigma_{implied}$ like the SVI model that fit market data and we can draw curves for each quoted maturities t.

We can now introduce the adjustment procedure, that we will have to repeat. In fact, if we have 8 quoted maturities, we will have 8 SVI model, one for each smile. We also use a mix between *Differential Evolution* and *Nelder-Mead* algorithm to fit each mid smile, which can be rewritten as:

$$\min_{\Lambda\in\mathcal{P}} \mathcal{J}(\Lambda) = \sum_k \left(\sigma^{market}_{implied}(K_k,T_t) - \sigma^{model}_{implied}(\Lambda;K_k,T_t)\right)^2, \quad \forall t. \tag{4.3}$$

where $\Lambda := (a,b,\rho,\sigma,m)$ and the square of $\sigma^{model}_{implied}$ can be written as (4.1). With our previous market data on $CAC40$ index from the 21st July 2023, we can adjust each SVI smile with mid data along each quoted maturities. Results are presented in figure [4.1].

Note that we have also specified bid and ask implied volatilities. A prerequisite for our model is to behave like a mid price, since we use mid implied volatilities, so our curve needs to lay between bid and ask values. We observe that it is easier to respect this constraint with large maturity. We can also remark that we don't use maturities larger than a year and a half due to liquidity consideration. We took the last quoted maturity where prices are still certain and aren't interpolated. Interestingly, even if our fitting procedure use a stochastic algorithm, on this dataset, each parametrized smile is very robust, in the sense that parameters values differ from 10^{-7} from an iteration to another.

Moreover, we want to investigate whereas our SVI smiles allow arbitrage or not. We have represented along log-moneyness the total implied variance and the function g for each smile (see figure [4.2]). We can now use previous definition and results to confirm that our SVI smiles don't provide calendar spread arbitrage as no line of total implied variance is crossed. On the contrary our smiles allow butterfly arbitrage as the positivity

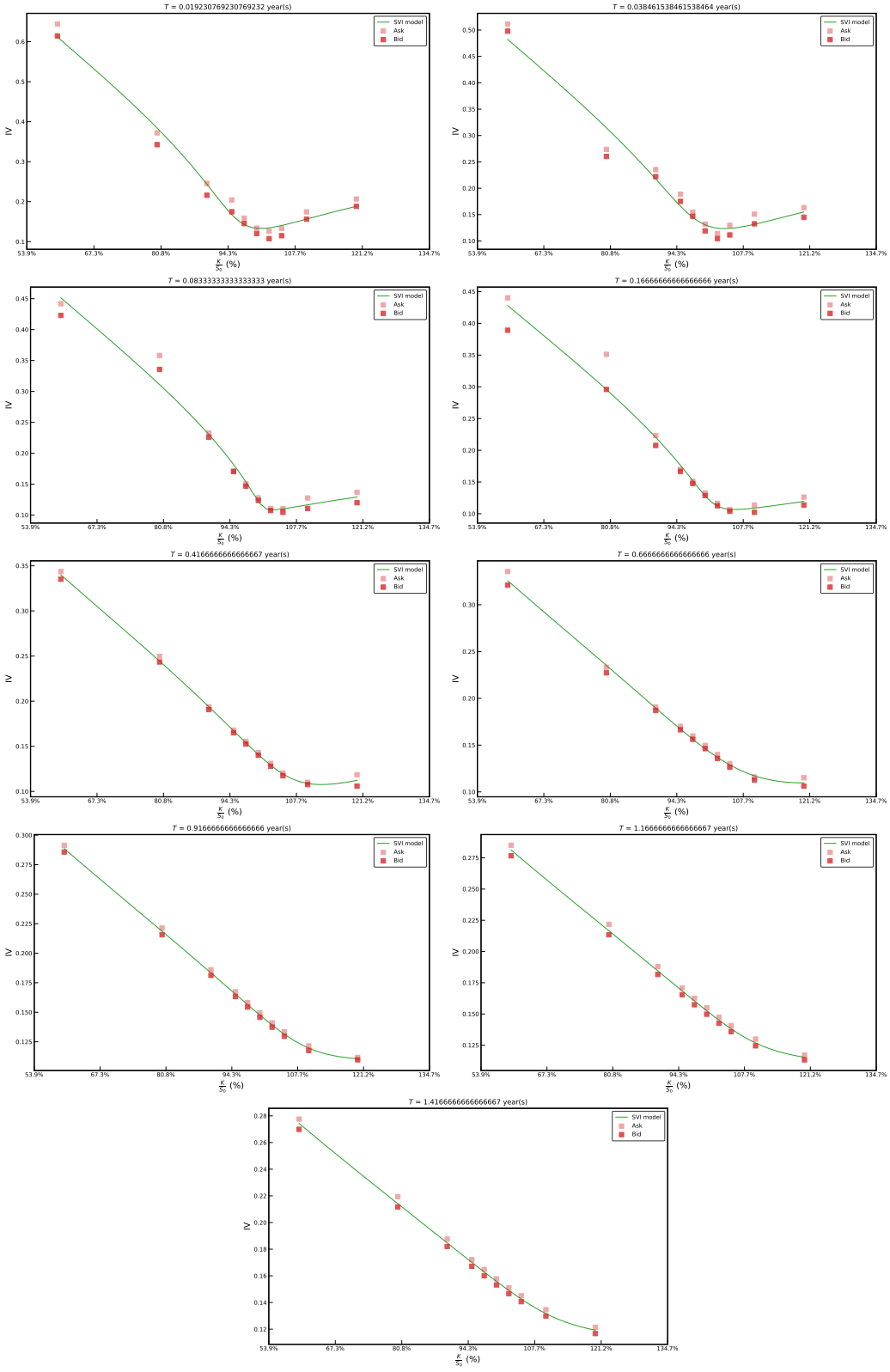

Figure 4.1: Adjustment of the SVI smiles over mid price data on $CAC40$ index from the 21st July 2023.

of g depends on maturity: for larger maturities we cross the horizontal axis. This an expected result in light of our previous remark about the work of Gatheral and Jacquier and the abscence of theoretical bounds for parameters to preclude butterfly arbitrage.

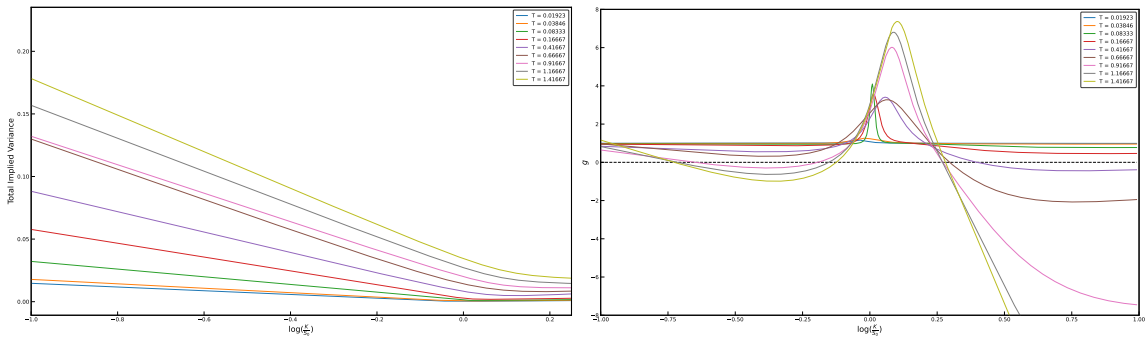

Figure 4.2: Total implied variance of SVI model (left) and g function (right) for each maturities along log-moneyness.

4.2 An interpolated surface: a new methodology for calibration over the $CAC40$ Index data.

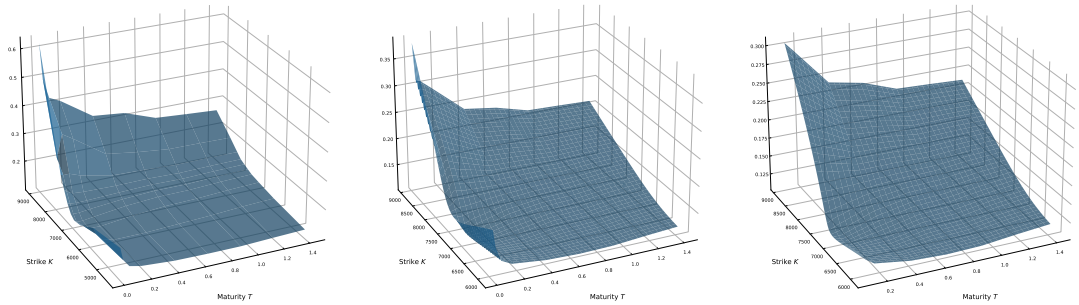

Figure 4.3: From left to right the building process to create a smoother surface for calibration.

In the section before, we have found nine different sets of numerical values, one for each maturity, that allow us to produce values of $\sigma^2_{implied}$ for unquoted log-strike k. These values respect at least, the absence of calendar spread arbitrage. In the following we have decided to use formula (4.1) to create a discrete grid between the two extreme quoted strike values with a step of 25.0. We have then a thin discretization along strike values, we can now wonder how to interpolate along maturities in order to create a surface?

This is not a real problem due to the fact quoted maturities will respect arbitrage-free property through SVI parametrization. Between the two extreme quoted maturities, we can build a dependency between each smile on a discrete grid with a step of one week. In literature this dependency is often linear, this is also the choice that we have made. On figure [4.3], we can see that these modifications help creating a smoother surface. To get the last surface, we have only deleted the first week in order to adjust our surface model on an implied surface with a first expiration date from one month long. This choice is explained by two arguments. First we know from previous computation and from literature (see [4]) that it is pointless to use short maturities in the calibration procedure. Second, we have used the one month maturity as the first maturity of interest in §3.2, so it is convenient to make this choice in order to compare results.

We can now present our results from the calibration procedure. We use the same technique as the one presented in §3.2, unless our target surface is now an interpolated surface. Since we have shown in last chapter the superiority of quadrature methods over FFT methods we have been interested in computing this calibration procedure with Heston and Attari formulas with the *Nelder-Mead* algorithm. Results are presented in tables 4.1, 4.2 and 4.3. We have also represented the absolute relative error between our interpolated surface and our model calibrated surface on figure [4.4].

	HES-GL	ATT-GK
R^2 with *Nelder-Mead*	0.9914	0.9914

Table 4.1: Coefficient of determination with optimal parameters values over $CAC40$ Index data using SVI interpolation.

	HES-GL	ATT-GK
ν_0	0.016995	0.016995
κ	4.7012	4.6992
(s.d.) κ	$3.777e-3$	$8.630e-3$
θ	0.03569	0.03570
(s.d.) θ	$4.2421e-6$	$1.2662e-5$
σ	1.3274	1.3270
(s.d.) σ	$8.0747e-4$	$1.4181e-3$
ρ	-0.6900	-0.6901
(s.d.) ρ	$3.4291e-5$	$5.4931e-5$

Table 4.2: Optimal parameters values for $CAC40$ Index with *Nelder-Mead* algorithm using SVI interpolation.

	HES-GL	ATT-GK
MSE	1474	1474
(s.d.) MSE	0.40	1.41
RMSE	38.38	38.38
(s.d.) RMSE	$5e-3$	$1.8e-2$
MAE	222	222
(s.d.) MAE	0.267	0.328

Table 4.3: Values of different measures for $CAC40$ Index with *Nelder-Mead* algorithm using SVI interpolation.

A striking difference comparing these results with results from section §3.2 is that we more than doubled our mean-square error but we gain one percent on the coefficient of determination (see tables 4.1 and 4.3). We have also increased the maximum absolute error of about a hundred basis points. This could be explained by the fact that we used more OTM strikes for this calibration procedure than in §3.2. As written before, for short maturities and far OTM or far ITM strikes our model dependent pricing formulas lack of

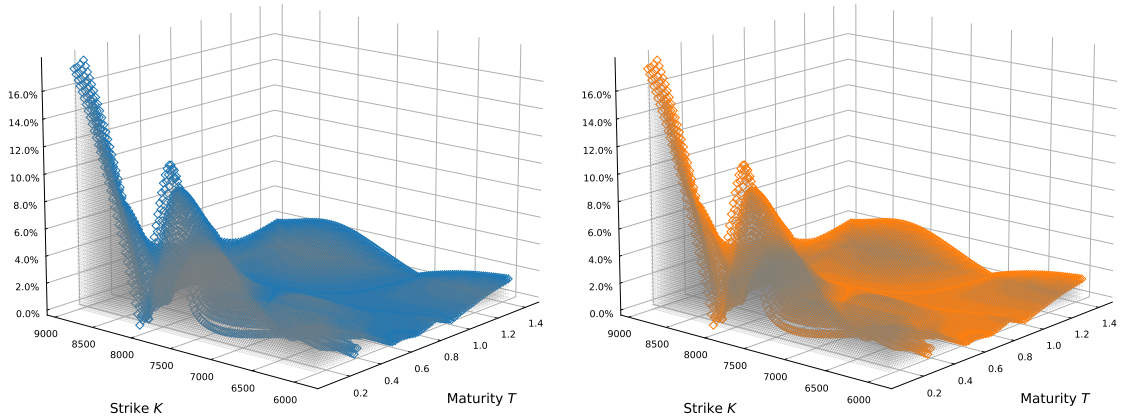

Figure 4.4: Absolute relative error of each point for formulas *Heston-GL* (left) and *Attari-GK* (right) with *Nelder-Mead* algorithm.

robustness.

Moreover in this case, Heston's formula and Attari's formula lead to similar results. We can still notice that Heston's formula leads to an optimal set of parameters that is less sensitive according to the standard deviation on table 4.3.

In the computational point of view, the procedure to interpolate market data doesn't have a significant impact on the calibration procedure. We can build the new surface outside regression algorithm and once for all. It increases memory complexity since the target surface is thiner so the associated matrix is larger, and it also increases time complexity of about 1 minute.

Even though we improve the coefficient of determination, the absolute relative error maps are still consequent locally (see figure [4.4]). In order to reduce the absolute relative error, we have been interested in changing the way of finding optimal parameters for our model by using variance Swaps.

4.3 Variance Swaps

In this section, we are interested in an alternative calibration procedure. Literature often suggest to decompose the calibration process using variance Swaps before vanilla options. The aim of this section is to present the market information provided by this volatility product and to show how suitable is its use for the Heston model calibration. We begin by giving a definition of a variance Swap:

Definition 6. *We define as variance Swap the variance product of payoff:*

$$N(V_{realized} - V_K),$$

where N is the nominal, $V_{realized}$ is the market realized annual variance and V_K is the fixed variance of the Swap.

Therefore variance Swaps express views on future market's variance, because the idea is to trade the realized market's variance for a fixed variance. If we suppose that market variance will increase in the future, we want to buy a variance Swap, which means buying

V_K with the idea that at the swap date $V_{realized} - V_K > 0$. Symmetrically if we suppose that market variance will decrease in the future, we rather want to sell a variance Swap at price V_K in order to have at the swap date $V_K - V_{realized} > 0$.

We haven't provided details on the fact that the price of the variance Swap is exactly V_K. It comes from the fact that the Swap value is equal to zero at pricing date according to arbitrage free assumption. As a consequence, we have that:

$$V_K = \mathbb{E}\left[V_{realized}\right]. \tag{4.4}$$

We then focus on computing V_K. In the following, we consider that N is equal to one and that our pricing date is now $t = 0$. We also consider a market with an underlying wich follows a geometrical brownian motion diffusion with a time-dependent volatility parameter σ_t:

$$dS_t = \sigma_t S_t dW_t.$$

In this case the realized variance over $[0;T]$ is exactly the average of σ_t^2 values over $[0;T]$ which leads to:

$$V_{realized} = \frac{1}{T}\int_0^T \sigma_t^2 dt.$$

Using Ito lemma to the function $\ln(S_t)$ leads to an expression for σ_t^2 that allows us to write:

$$V_{realized} = \frac{2}{T}\left[\int_0^T \frac{dS_t}{S_t}dt + \int_0^T d\ln(S_t)\right] = \frac{2}{T}\left[\int_0^T \frac{dS_t}{S_t}dt + \ln\left(\frac{S_T}{S_0}\right)\right]. \tag{4.5}$$

We denote as $\mathbb{E}$ the expected value under the risk-neutral measure. With the expression (4.4), we can write:

$$V_K = \frac{2}{T}\mathbb{E}\left[\int_0^T \frac{dS_t}{S_t}dt + \ln\left(\frac{S_T}{S_0}\right)\right].$$

Using Fubini-Lebesgue's theorem and the expected returns under the risk-neutral measure we finally have:

$$V_K = \frac{2}{T}\left[rT + \mathbb{E}\left[\ln\left(\frac{S_T}{S_0}\right)\right]\right].$$

The price of the variance Swap is therefore obtained by computing the expected value term. This is often done in literature by using log-contract products. In fact the first right hand side term can be expressed as $-\mathbb{E}\left[-\ln(e^{rT})\right]$ and we can expressed V_K as the price of a log-contract using the payoff decomposition of Carr and Madan (see [4]):

$$V_K = -\frac{2}{T}\mathbb{E}\left[\ln\left(\frac{S_T}{F_T}\right)\right] = \frac{2}{T}\left[\int_{-\infty}^0 p(k)dk + \int_0^{+\infty} c(k)dk\right], \tag{4.6}$$

$$\text{with} \quad c(k) := \frac{C(F_T e^k)}{F_T e^k}, \quad \text{and} \quad p(k) := \frac{P(F_T e^k)}{F_T e^k}.$$

where $k := \ln(\frac{K}{F_T})$ and C and P denote respectively Call and Put prices. In the following, we consider that this price is already observable on the market. A well known marker of this price is the VIX index. This is the volatility index associated to movements of log-contracts over the $S\&P500$ index (noted also SPX). It is defined as the straightforward discretization of equation (4.6):

$$VIX^2 = \frac{2}{T}\sum \frac{\Delta K_i}{K_i} Q_i(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2, \tag{4.7}$$

$$\Delta K_i = \begin{cases} K_2 - K_1 \text{ if } i = 1, \\ K_N - K_{N-1} \text{ if } i = N, \\ (K_{i+1} - K_{i-1})/2, \text{ else.} \end{cases} \qquad Q_i(K_i) = \begin{cases} P(T, K_i) \text{ if } K_i < K_0, \\ C(T, K_i) \text{ if } K_i > K_0, \\ (C(T, K_0) + P(T, K_0))/2, \text{ if } K_i = K_0. \end{cases}$$

with $K_0 := \max_{1 \leq i \leq N} \{K_i | K_i <= F_T\}$.

4.4 A calibration methodology using future market variance

In this section, we show how we can use previous development about variance Swaps to match future market's variance with our model. We use market data on SPX and VIX from the 16th August 2023. In the Heston model we know the law of the variance process as stated in [6] and in [4]:

$$\mathbb{E}\left[\nu_T\right] = \theta T + \frac{1 - e^{-\kappa T}}{\kappa} \left(\nu_0 - \theta\right). \tag{4.8}$$

Remark that with this expression the expected variance of the model only depends on variance process parameters: the initialization value ν_0, the mean reversion speed κ and the long-term variance mean θ. Since the variance Swap is quoted in annualized variance, we are therefore interested in comparing the annualized version of the expression (4.8) to the expression (4.7). And by finding optimal parameters $(\nu_0^*, \kappa^*, \theta^*)$ we want to replicate future market variance with our model through relationship (4.4):

$$\frac{1}{T}\left(\theta^* T + \frac{1 - e^{-\kappa^* T}}{\kappa^*}(\nu_0^* - \theta^*)\right) = VIX(T)^2. \tag{4.9}$$

Relationship (4.9) can be seen as a major change in terms of calibration. In fact we have now a way to estimate optimal parameters of the Heston model in two distinct steps: we first look for values that satisfy equation (4.9), and then we look for σ and ρ values with the implied volatility surface of SPX as we have done previously on $CAC40$ data.

We can apply the same procedure as the one presented in §2.3 but we expect a speed up in computational time since we can now split our calibration problem over $\mathbb{R}^5$ in two calibration problems over respectively $\mathbb{R}^3$ and $\mathbb{R}^2$. We then use the values obtained as starting values for a local optimization on the whole parametrized surface, which means solving problem (4.12). We solve successively problems (4.10), (4.11), and (4.12) as presented in algorithm 4.

$$\min_{(\nu_0,\kappa,\theta)\in]0;1]\times\mathbb{R}_+\times]0;1]} \mathcal{J}_1(\nu_0, \kappa, \theta) = \sum_t \left[tVIX(t)^2 - \theta t - \frac{1 - e^{-\kappa t}}{\kappa}\left(\nu_0 - \theta\right)\right]^2. \tag{4.10}$$

$$\min_{(\sigma,\rho)\in\mathbb{R}_+\times[-1;0]} \mathcal{J}(\sigma, \rho) = \sum_t \sum_k \frac{1}{s^2}\left(\sigma_{implied}^{market}(K_k, T_t) - \sigma_{implied}^{model}(\nu_0^*, \kappa^*, \theta^*, \sigma, \rho; K_k, T_t)\right)^2. \tag{4.11}$$

$$\min_{(\nu_0,\kappa,\theta,\sigma,\rho)} \mathcal{J}(\nu_0, \kappa, \theta, \sigma, \rho) = \sum_t \sum_k \frac{1}{s^2}\left(\sigma_{implied}^{market}(K_k, T_t) - \sigma_{implied}^{model}(\nu_0, \kappa, \theta, \sigma, \rho; K_k, T_t)\right)^2. \tag{4.12}$$

where s denote the bid-ask spread between each option. This modification is similar as using the vega *Black-Scholes*. In fact, we can observe that the spread is smaller for liquid options than for illiquid options, thus by weighting with an inverse function of the spread we insure to give more weight to near the money options than far in our out the money options.

Algorithm 4 (Calibration of the Heston model using VIX and SPX data)

1. $\nu_0^*, \kappa^*, \theta^* \leftarrow$ Nelder_Mead over the bounded problem (4.10).
2. $\sigma^{DE}, \rho^{DE} \leftarrow$ Differential_Evolution over the bounded problem (4.11).
3. $\nu_0^\star, \kappa^\star, \theta^\star, \sigma^\star, \rho^\star \leftarrow$ Nelder_Mead over the two bounded problems (4.12) with starting set $(\nu_0^*, \kappa^*, \theta^*, \sigma^{DE}, \rho^{DE})$.

We can note that the first calibration procedure is independent from the choice of computation of european vanilla options. With a formal calculus software, we can also note that the objective function of problem (4.10) is convex along all of its variable since the Hessian matrix is positive semidefinite on the specified domain. Therefore there exist a unique set of parameters that solves the above problem. This justifies the use of only one local minimizer *Nelder-Mead* since by gradient descent we can only find the global optimum. The adjusted model that matches future market's variance is presented in figure [4.5], as well as future model's variance after minimizing problem (4.12). Since SPX

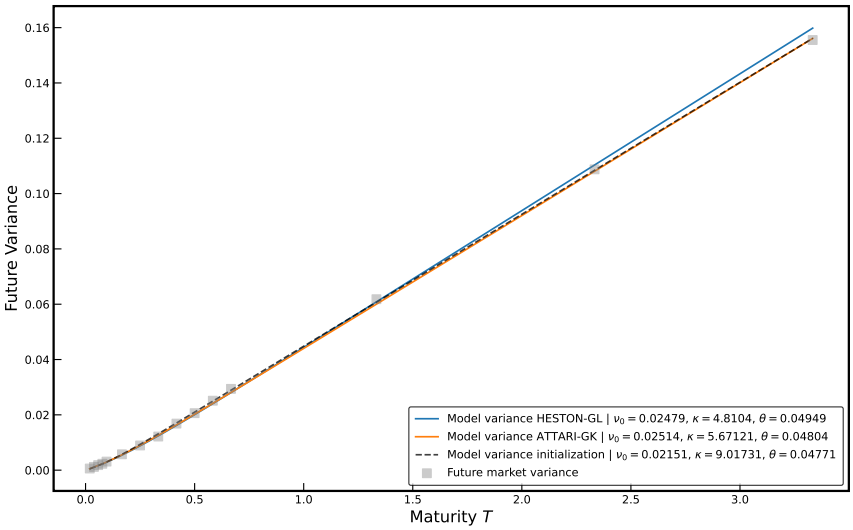

Figure 4.5: Adjusted model on market future variance ($tVIX^2$).

index is very liquid, we have quoted prices for options far out-of-the-money or far in-the-money. This result in calibrating our model where the analytical formula numerically fails to give a price, and this problem is presented in figure [4.6]. Therefore we have decided to delete strikes above and below an absolute value of the at-the-money forward of 50. We present the results in tables 4.4, 4.5 and 4.6 as well as the absolute relative error in figure [4.7].

	HES-GL	ATT-GK
R^2 with *Nelder-Mead*	0.9446	0.9572

Table 4.4: Coefficient of determination with optimal parameters values over SPX Index data using variance Swaps.

The first observation that could be made is the obvious difficulties to get efficient results. A mean-square error of 28000 or 22000 basis points is beyond compare with the results obtain with $CAC40$ data. This helps also to reconsider the coefficient of determination as a criterion of decision concerning model choice: it is still very high whereas mean-square error as well as maximum absolute error have increased a lot comparing to SVI results on

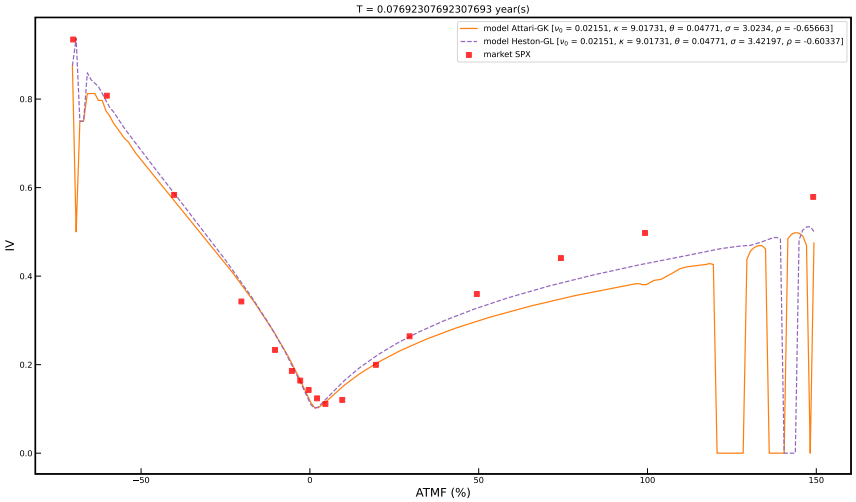

Figure 4.6: Illustration of the lack of robustness of the Heston and Attari formulas.

	HES-GL	ATT-GK
ν_0	0.02479	0.02514
κ	4.8104	5.6712
(s.d.) κ	0.0402	0.0387
θ	0.04948	0.04804
(s.d.) θ	$1.06e-4$	$6.26e-5$
σ	1.8465	2.0997
(s.d.) σ	$6.722e-3$	$8.566e-3$
ρ	-0.6917	-0.6741
(s.d.) ρ	$2.82e-4$	$2.83e-4$

Table 4.5: Optimal parameters values for SPX Index with *Nelder-Mead* algorithm using variance Swaps.

	HES-GL	ATT-GK
MSE	28017	22154
(s.d.) MSE	63.315	88.773
RMSE	167.38	148.84
(s.d.) RMSE	0.1891	0.2982
MAE	1034	889.7
(s.d.) MAE	2.039	2.976

Table 4.6: Values of different measures for SPX Index with *Nelder-Mead* algorithm using variance Swaps.

$CAC40$ data. However, against what we have seen in the above results, it is the Attari formula that provides the best fit to our market surface. We can see on table 4.6, a range of about 6000 basis points between each model's surface which is consequent. Note also that we have a better coefficient stability with this formula according to table 4.5.

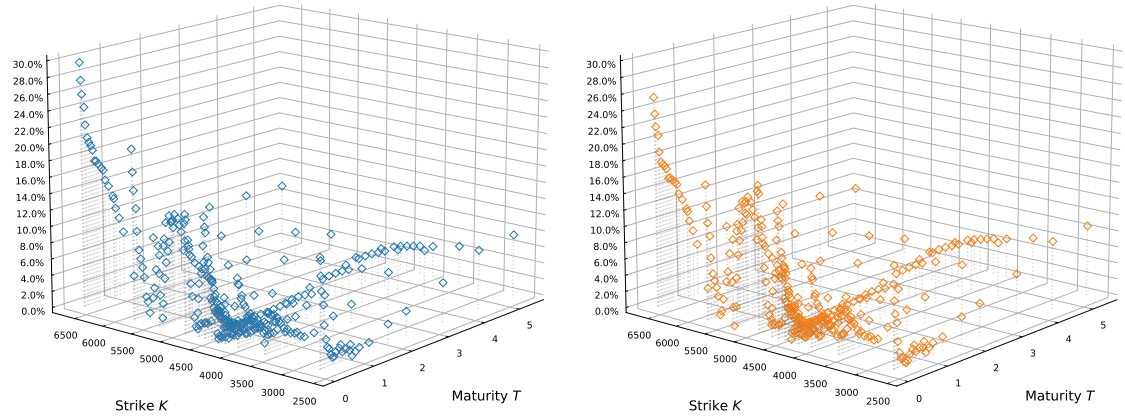

Figure 4.7: Absolute relative error of each point for formulas *Heston-GL* (left) and *Attari-GK* (right) with *Nelder-Mead* algorithm.

Moreover, on figure [4.5], this is also the formula that leads to the best fit according to market data and initialization values. These results may be also a consequence of using the market's spread as a weighting function in the calibration procedure and this could be an axis to investigate. As a conclusion on this variance Swaps methodology, we can say that the real upside of this method is the weak computational cost. As a reference the mean computational time is about 12 times shorter than the mean computational time of the SVI methodology.

Chapter 5

Sensitivities computation in the Heston model

In this chapter, we try to introduce issues linked to greeks and sensitivities computation in the Heston model. Our work here is mainly driven by Paul Glasserman's book (see [23]) as well as the work of Mark Broadie and Özgür Kaya (see [24]). We aim also at distinguishing option Greeks under a *Black-Scholes* framework and sensitivities of the Heston model inputs. In the first section we introduce general Monte-Carlo methods for the computation of sensitivities, as well as their downsides and benefits. The second section provides details on the computation of option Greeks for the Heston model and the studied scheme. Finally, in the last section, we present the issues that we have to cope with in the Heston model in order to compute sensitivities with respect to the model inputs.

5.1 A review of Monte-Carlo methods for the computation of sensitivities

We are interested in introducing numerical methods to differentiate prices with respect to some variable x:

$$\partial_\theta P\left(t=0,x;\phi(\theta)\right) \quad \text{where, for simplicity, } P \text{ is defined as} \quad P\left(t,x;\phi(\theta)\right) := \mathbb{E}_t\left[\phi(S_T^{x,\theta})\right].$$

In the above formulation, ϕ is the discounted payoff function and it could depend on several intermediate values S_t, with $t \in [0;T]$ for path-dependent options especially. Independently from specific methods associated to payoff's properties, we can identify in literature three different methods based on Monte-Carlo estimations: the Bump method, the Pathwise approach and the Likelihood Ratio approach. We present here the difference between each method and their applicability. In the following we define $P(\theta) := P\left(t=0,x;\phi(\theta)\right)$ for conveniency.

The Bump method

The Bump method is an approach based on finite difference, and we aim at differentiating Monte-Carlo estimators. It suggests a relative smoothness of the price function with respect to θ input since we want to use a Taylor's approximation. As a direct consequence we can quantify the bias introduced by the finite difference estimator.

To proceed, we suppose that we have computed the Monte-Carlo estimators $\hat{P}(\theta-\varepsilon)$ and $\hat{P}(\theta+\varepsilon)$ from two samples of equal size. Through smoothness property, we can write:

$$P(\theta+\varepsilon) = P(\theta) + \varepsilon\partial_\theta P(\theta) + \frac{\varepsilon^2}{2}\partial^2_{\theta\theta} + \mathcal{O}(\varepsilon^3),$$

$$P(\theta-\varepsilon) = P(\theta) - \varepsilon\partial_\theta P(\theta) + \frac{\varepsilon^2}{2}\partial^2_{\theta\theta} + \mathcal{O}(\varepsilon^3).$$

By denoting by Δ_θ the centered approximation of $\partial_\theta P(\theta)$, we therefore have that:

$$\Delta_\theta(\varepsilon) := \frac{1}{2\varepsilon}\left(\hat{P}(\theta-\varepsilon) + \hat{P}(\theta+\varepsilon)\right).$$

Then we can wonder if this estimator is suited for the computation of $\partial_\theta P(\theta)$? From the above relation we clearly have that $\Delta_\theta(\varepsilon)$ converges to $\partial_\theta P(\theta)$ when ε goes towards 0. We are interested in taking ε as small as possible, but considering the bias-variance trade off specific to Monte-Carlo methods we are constrained in our choice. In fact, variance of $\Delta_\theta(\varepsilon)$ is proportional to $4\varepsilon^{-2}$ and it is clear that we increase uncertainty of our estimator by taking smaller values. Note that if we build our estimators $\hat{P}(\theta-\varepsilon)$ and $\hat{P}(\theta+\varepsilon)$ not from two samples but from one sample

$$\left((\phi(S_T^{x,\theta-\varepsilon}), \phi(S_T^{x,\theta+\varepsilon}))_i\right)_{0\geq i\geq n},$$

so we compute both terms from the same set of m observations of $\phi(S_T^{x,\theta})$, we get a variance that is proportional to n^{-1} (see [10] and [23]). We have only presented here the methodology for first order derivatives but we can do the same for second order derivatives.

The Pathwise approach

Contrary to the Bump method, with the Pathwise approach we use the parametrization of the diffusion process to differentiate with respect to parameters. Particularly this method supposes that expectation and derivative are interchangeable. We present in the following a result introduced in Glasserman's book ([23]).

Theorem 1. *If for all $t \in [0;T]$ the function $\theta \mapsto S_t^{x,\theta}$ is in $\mathcal{C}^1$ a.s., ϕ is Lipschitz and $\mathbb{E}\left[\sup_{t\leq T}|\partial_\theta S_t^{x,\theta}|\right] < \infty$ then*

$$\partial_\theta P(\theta) = \mathbb{E}\left[\phi'(S_T^{x,\theta})\partial_\theta\left(S_T^{x,\theta}\right)\right].$$

The proof of this theorem involves the application of the dominated convergence theorem. Through this result we notice that if we can compute the derivatives of the right hand side term, we can build an unbiased estimator of $\partial_\theta P(\theta)$. We can remark that the upper result can be generalized to second order derivatives, and it is sufficient to consider the Lipschitz criterion on ϕ'. However as an example, a digital option doesn't satisfy this condition since its first order derivative isn't defined in the classical sense.

The Likelihood Ratio approach

Since the Pathwise approach is dependent of the payoff continuity, it might not be applicable to all options. An alternative is the Likelihood Ratio approach and it is based on the following computation:

$$\partial_\theta\mathbb{E}\left[\phi(S_T^x)\right] = \int_\Omega \phi(s)\partial_\theta(p_\theta(s))ds = \int_\Omega \phi(s)\frac{\partial_\theta(p_\theta(s))}{p_\theta(s)}p_\theta(s)ds = \mathbb{E}\left[\phi(s)\frac{\partial_\theta(p_\theta(s))}{p_\theta(s)}\right]\Big|_{s=S_T^x}.$$

We note that this approach also relies on the interchange of derivative and expectation and if its holds we keep the unbiased property. As stated by Glasserman, this relationship is often verified due to the smooth overall behaviour of density function. But we might need an argument of absolute continuity on p_θ to conclude that the following equalities for first and second order derivatives hold:

$$\partial_\theta P(\theta) = \mathbb{E}\left[\phi \frac{\partial_\theta p_\theta}{p_\theta}\right], \quad \partial^2_{\theta\theta} P(\theta) = \mathbb{E}\left[\phi \frac{\partial^2_{\theta\theta} p_\theta}{p_\theta}\right].$$

As a final word on these methods we can say that even if the Bump methods can be easily implemented because it requires no further computation, the bias induced by its estimator can be a downside. As a consequence we do not cover finite difference implementation in the next section.

5.2 Greeks in the Heston model

Through the above computation methods we aim at computing options Greeks as they are defined in the *Black-Scholes* model. Since we approximate the law of the underlying of our process, we can identify log-normal distribution conditionnally to the variance values. Our work is largely inspired by the article of Broadie and Kaya (see [24]) but adapted to the studied schemes introduced in section §1.2. In this section only we consider that P is the fair price at time $t = 0$ of a Call option of maturity T and we want to compute the below quantities:

$$\Delta := \frac{\partial P}{\partial S}, \quad \Gamma := \frac{\partial^2 P}{\partial S^2}, \quad \Theta := -\frac{\partial P}{\partial T}, \quad \vartheta := \frac{\partial P}{\partial \sigma}.$$

In a GMB model, S denotes the spot value of the underlying and σ is the volatility parameter. The methodology proposed in [24] identifies the equivalent of the σ parameter in a stochastic volatility framework and it can be extend to log-Euler, Quadratic Exponential and E+M schemes. In fact, each scheme can be written like:

$$S_T = S\xi e^{\left(r - \frac{\sigma^2}{2}\right)T + \bar{\sigma}\sqrt{T}Z} \tag{5.1}$$

$$\xi := e^{-\frac{\rho^2}{2}\int_0^T \nu_s ds + \rho \int_0^T \sqrt{\nu_s} dW_s^\nu} \quad \text{and} \quad \bar{\sigma}^2 := \frac{(1-\rho^2)\int_0^T \nu_s ds}{T}.$$

If we consider that variance process values are known on a discrete time set $t_0 = 0 < t_1 < \cdots < t_M = T$ the variables ξ and $\bar{\sigma}$ are known and constant and the approximation of the underlying follows a log-normal distribution. Particularly, $\bar{\sigma}$ will be the volatility parameter used in the computation of ϑ. Note that the above expression and comments can't be done in the case of the Milstein scheme since we have a further development. We would have the square of a standard gaussian variable and the approximated law would no longer be log-normal. Therefore this methodology won't be applied to the Milstein scheme.

With the parametrized formula (5.1), it is then very convenient to apply the Pathwise approach. And since we obtain a log-normal distribution the Likelihood Ratio approach can also be used on the approximated law. In fact the distribution density g in the case of (5.1) could be written (see appendix §A.4):

$$g(x) := \frac{f(k(x))}{x\bar{\sigma}\sqrt{T}} \quad \text{with} \quad k(x) := \frac{\ln(\frac{x}{S\xi}) - (r - \frac{\bar{\sigma}^2}{2})T}{\bar{\sigma}\sqrt{T}}.$$

where f is the probability density function of a standard gaussian variable. Now we can wonder what are the differences between each scheme in the light of formula (5.1)? The answer is in the computation of ξ and $\bar{\sigma}$. Schemes as it has been introduced in §1.2 differ from the approximation:

$$\textit{(Euler)} \int_{t_i}^{t_{i+1}} \nu_s ds \approx \nu_{t_i} T, \quad \textit{(Quadratic Exponential, E+M)} \int_{t_i}^{t_{i+1}} \nu_s ds = \frac{T}{2}\left[\nu_{t_i} + \nu_{t_{i+1}}\right].$$

An other difference is that the variance process is sample directly from the conditional law in the Quadratic Exponential scheme whereas it is approximated by a Milstein scheme in the E+M scheme. Now considering that we have a discrete collection $(\nu_0, \cdots, \nu_M)$, we can compute an approximation of terms ξ and $\bar{\sigma}$ through:

$$\textit{(Euler)} \int_0^T \nu_s ds = \sum_{i=1}^{M} \int_{t_{i-1}}^{t_i} \nu_s ds \approx \sum_{i=1}^{M} \nu_{t_{i-1}}(t_i - t_{i-1}),$$

$$\textit{(Quadratic Exponential, E+M)} \int_0^T \nu_s ds = \sum_{i=1}^{M} \int_{t_{i-1}}^{t_i} \nu_s ds \approx \sum_{i=1}^{M} \left[\nu_{t_{i-1}} + \nu_{t_i}\right] \frac{(t_i - t_{i-1})}{2}.$$

$$\textit{(Euler)} \int_0^T \sqrt{\nu_s} dW_s^{\nu} \sim \mathcal{N}\left(0, \sum_{i=1}^{M} \nu_{t_{i-1}}(t_i - t_{i-1})\right),$$

$$\textit{(Quadratic Exponential, E+M)} \int_0^T \sqrt{\nu_s} dW_s^{\nu} \sim \mathcal{N}\left(0, \sum_{i=1}^{M} \left[\nu_{t_{i-1}} + \nu_{t_i}\right] \frac{(t_i - t_{i-1})}{2}\right).$$

Now with (5.1) and the above details we can compute estimators for Pathwise and Likelihood Ratio approaches. Calculation details are provided in appendix (see appendix §A.4), and the estimators, in the absence of risk-free interest rate, are presented in the table 5.1.

	Pathwise	Likelihood Ratio
Δ	$\mathbb{E}\left[\frac{S_T}{S} 1_{S_T \geq K}\right]$	$\mathbb{E}\left[(S_T - K)_+ \frac{\partial_S g(S_T)}{g(S_T)}\right]$
		$\frac{\partial_S g(S_T)}{g(S_T)} := \frac{k(S_T)}{S\bar{\sigma}\sqrt{T}}$
Γ	$\mathbb{E}\left[\frac{Kk(S_T)}{S^2\bar{\sigma}\sqrt{T}} 1_{S_T \geq K}\right]$	$\mathbb{E}\left[(S_T - K)_+ \frac{\partial_S^2 S g(S_T)}{g(S_T)}\right]$
		$\frac{\partial_{SS}^2 g(S_T)}{g(S_T)} := \frac{k(S_T)^2 - k(S_T)\bar{\sigma}\sqrt{T} - 1}{S^2\bar{\sigma}^2 T}$
Θ	$\mathbb{E}\left[-\left(r - \frac{\bar{\sigma}^2}{2} + \frac{\bar{\sigma}Z}{2\sqrt{T}}\right) S_T 1_{S_T \geq K}\right]$	$\mathbb{E}\left[-(S_T - K)_+ \frac{\partial_T g(S_T)}{g(S_T)}\right]$
		$-\frac{\partial_T g(S_T)}{g(S_T)} := \frac{1}{2T} - \frac{k(S_T)\left[\ln(\frac{S_T}{S\xi}) + (r - \frac{\bar{\sigma}^2}{2})T\right]}{\bar{\sigma}2T\sqrt{T}}$
ϑ	$\mathbb{E}\left[\left(-\bar{\sigma}T + \sqrt{T}Z\right) S_T 1_{S_T \geq K}\right]$	$\mathbb{E}\left[(S_T - K)_+ \frac{\partial_{\bar{\sigma}} g(S_T)}{g(S_T)}\right]$
		$\frac{\partial_{\bar{\sigma}} g(S_T)}{g(S_T)} := \frac{k(S_T)\left[\ln(\frac{S_T}{S\xi}) - (r + \frac{\bar{\sigma}^2}{2})T\right]}{\bar{\sigma}^2\sqrt{T}} - \frac{1}{\bar{\sigma}}$

Table 5.1: Monte-Carlo estimators for *Pathwise* and *Likelihood Ratio* approaches.

Since we have discussed about problems induced by the computation of digital options with the Pathwise approach, we note that we have also a problem for the computation of Γ. Hence, we compute this variable by differentiating with the Pathwise methodology the Δ Likelihood Ratio estimator. We have then computed estimators for several spot

values with the following model $\nu_0 = 0.04$, $\kappa = 2.0$, $\theta = 0.04$, $\sigma = 0.6$ and $\rho = -0.69$ and with 100000 observations of S_T. Results are presented in figures [5.1] and [5.2] for several expiries and with a confident interval of 95%.

We observe that both methods allow us to obtain curves that match greeks in the *Black-Scholes* model. We also notice that variance increases when the spot increases, even if curves are smoother and with less variance with the Pathwise approach.

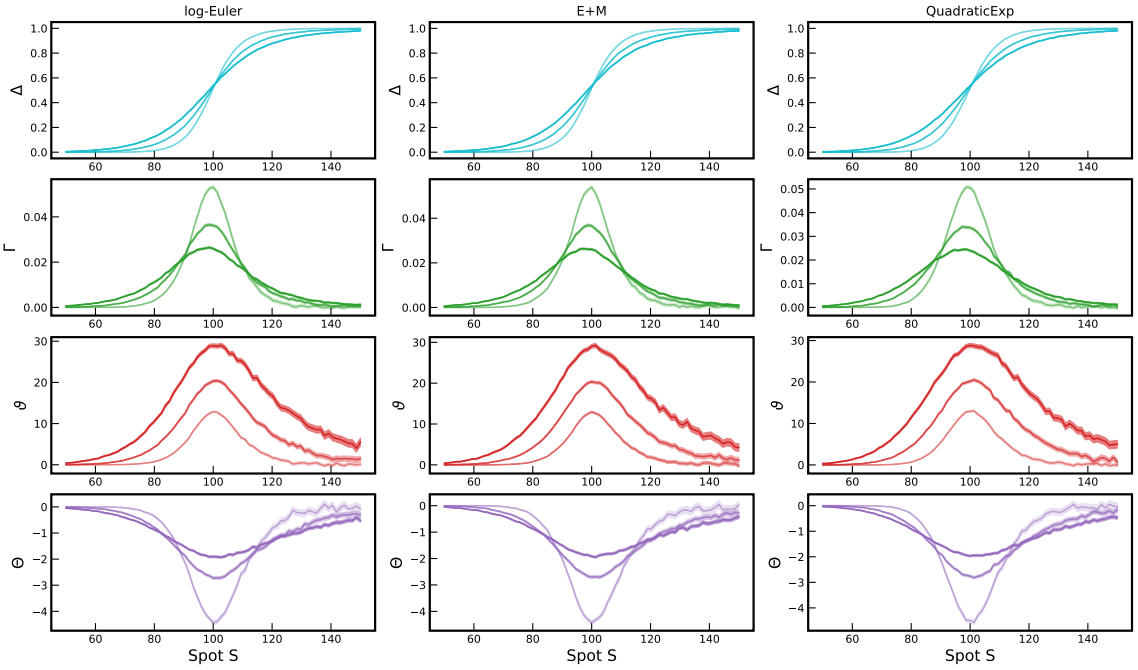

Figure 5.1: Pathwise Greeks computation along spot values with $K = 100$, $r = 0.0$ and for 3 different expiries with a confident interval of 95%.

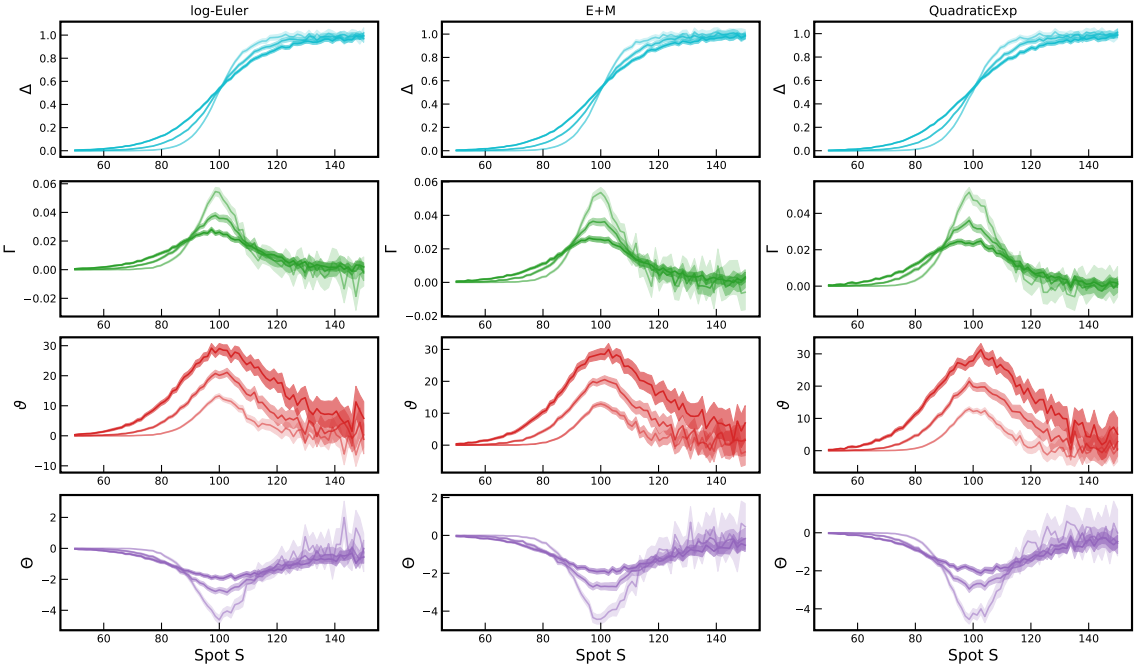

Figure 5.2: Likelihood Ratio Greeks computation along spot values with $K = 100$, $r = 0.0$ and for 3 different expiries with a confident interval of 95%.

5.3 Issues relative to model inputs sensitivities

However in terms of risk management we prefer to estimate model sensitivities with respect to its inputs. In the case of the Heston model it means computing also the partial derivatives $\partial_{\nu_0} P, \partial_\kappa P, \partial_\theta P, \partial_\sigma P$, and $\partial_\rho P$.

Even if the methods introduced in this chapter can be used, their application can be more challenging. For example if we refer to the absolute continuity required for the Likelihood Ratio approach, we note that variance process isn't continuous with respect to its parameters κ, θ, σ and ρ due to the switching rule in the sampling.

We can also note the fact that computations of section §5.2, don't hold since the variance process needs to be known. Therefore analytical computation as it has been done above can be slightly different and challenging. Moreover this was already compromised by the fact that it can't be adapted to the Milstein scheme.

Problems encountered and presented above have been tackled in the article written by Jiun Chan, Mark Joshi and Dan Zhu (see [25]). The solution proposed use automatic adjoint differentiation (AAD) and rely on the mapping between the model inputs and the scheme, but the implementation of this method exceeds the purpose of this work.

Conclusion

To conclude we have discussed about the Heston model and its suitability to replicate market's volatility behaviour. We have also highlighted the similarities with the *Black-Scholes* framework. On the diffusion implementation we have more or less succeeded in matching the results presented in the article [7].

We also compared methods using a quadrature rule and Fast Fourier transform methods and in the light of our developement, we have observed the poor performance of the Fast Fourier methods. Another interesting point concerns Attari and Heston formulas. Even if we increase the bias of the approximation in Heston's formula by using a second integral, we haven't perceived major differences with Attari's formula.

Based on vanilla options prices we have proposed different methods to calibrate the numerical value of each parameter. These methods present benefits and inconvenients that need to be taken into account depending on the purpose. We have been interested in having efficient numerical results on calibration but an in-depth study of the computational complexity could also be a factor of interest for further developements. A possibility often proposed in literature is to use a deep calibration procedure. In fact the lack of knowledge on the objective function could be exploited by neural network and may outperform classical methodologies presented in this report.

We have also presented issues related to sensitivities computation. Even if we succeeded in finding adapted results for the computation of option Greeks, the computation of sensitivities with respect to model inputs requires advanced techniques such as automatic adjoint differentiation (AAD). The major upside of using AAD is to tackle payoff related issues as well as parameters continuity.

Through this work we have also been able to show the limits of the Heston model. Particularly we have seen that, due to numerical issues on the analytical formula, it might be hard to obtain a calibrated model that fit the market for short expiries. Among tools that can be used to correct this behavior we can suggest local volatility function or more generally local stochastic volatility model, but we can wonder what is the cost to pay in order to use this kind of model?

From a personnal point of view this internship helped me to improve my knowledge in modeling volatility and issues related to model's calibration. I enjoyed working on this master's thesis and discussing it with my supervisors and colleagues, whom, through their knowledge and skills, turned this internship and work into a concrete continuity of this year's courses. This was also the opportunity to face a wide range of mathematical modeling and quantitative finance related issues, and to reinforce my desire to work in the industry. I am therefore very grateful to Exiom Partners for this opportunity.

Bibliography

[1] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, pp. 637–654, 1973.

[2] S. L. Heston, "A closed-form solution for options with stochastic volatility with application to bonds and currency options," *The Review of Financial Studies*, vol. 6, no. 2, pp. 327–343, 1993.

[3] M. Jeanblanc, M. Yor, and M. Chesney, *Mathematical Methods for Financial Markets*. Springer, 2009.

[4] J. Gatheral, *The Volatility Surface, A Practitioner's Guide*. Wiley Finance, 2006.

[5] P. Wilmott, *Paul Wilmott on Quantitative Finance*. John Wiley and Sons, 2000.

[6] L. Andersen, "Efficient Simulation of the Heston Stochastic Volatility Model," 2007.

[7] M. Mrazek and J. Pospisil, "Calibration and simulation of Heston model," *De Gruyter Open, Open Mathematics*, vol. 15, 2017.

[8] A. Alfonsi, "High order discretization schemes for the cir process: application to affine term structure and heston models," *Mathematics of Computation*, pp. 209–237, 2010.

[9] M. Broadie and Özgür Kaya, "Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes," *Operation Research*, 2006.

[10] J.-F. Chassagneux, *Monte-Carlo methods in Finance, Notes de cours*. Master 2, Modélisation aléatoire, 2022-2023.

[11] S. Crépey, *Produits Dérivés, Notes de cours*. Master 2, Modélisation aléatoire, 2022-2023.

[12] C. Kahl and P. Jäckel, "Not-so-complex logarithms in the Heston model," *Wilmott Magazine*, pp. 94–103, 2005.

[13] P. Carr and D. B. Madan, "Option valuation using the fast fourier transform," *Journal of Computational Finance*, pp. 463–520, 1999.

[14] R. Lord and C. Kahl, "Optimal fourier inversion in semi-analytical option pricing," *Tinbergen Institute Discussion Paper*, 2006.

[15] M. Schmelzle, "Option Pricing Formulae using Fourier Transform: Theory and Application." 2010.

[16] D. Conte, L. G. Ixaru, B. Paternoster, and G. Santomauro, "Exponentially-fitted gauss-laguerre quadrature rule for integrals over an unbounded interval," *Journal of Computational and Applied Mathematics*, pp. 725–736, 2014.

[17] R. Piessens, E. de Doncker-Kapenga, C. W. Überhuber, and D. K. Kahaner, *QUADPACK: A Subroutine Package for Automatic Integration*. Springer Verlag, 1983.

[18] F. Ametrano and L. Ballabio, "Quantlib - a free/open-source library for quantitative finance," 2003.

[19] S. Mikhailov and U. Nögel, "Heston Stochastic Volatility Model Implementation, Calibration and Some Extensions," *Wilmott Magazine*, pp. 74–79, 2009.

[20] Y. Cui, S. del Bano Rollin, and G. Germano, "Full and fast calibration of the Heston stochastic volatility model," *European Journal of Operation Research*, pp. 625–638, 2017.

[21] C. W. Oosterle and L. A. Grzelak, *Mathematical Modeling and Computation in Finance*. World Scientific Publishing Europe Ltd., 2020.

[22] J. Gatheral and A. Jacquier, "Arbitrage-free SVI volatility surfaces," *Quantitative Finance*, vol. 14, no. 1, 2014.

[23] P. Glasserman, *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

[24] M. Broadie and Özgür Kaya, "Exact simulation of option greeks under stochastic volatility and jump diffusion models," *Proceedings of the 2004 Winter Simulation Conference*, vol. 2, 2004.

[25] J. H. Chan, M. S. Joshi, and D. Zhu, "First and second order greeks in the Heston model," 2014.

Appendix A

Calculation details

A.1 Pricing european options - Girsanov's theorem application

In this section, notations used are independent. To justify Girsanov's theorem we set an arbitrage-free framework and therefore justify the existence of at least one risk-neutral measure. Problem (1.2) using Cholesky's decomposition is equivalent to:

$$\begin{cases} dS_t = \mu S_t dt + \sqrt{\nu(t)} S_t \rho dW_t^{(2)} + \sqrt{\nu(t)} S_t \sqrt{1-\rho^2} dW_t^{(1)} \\ d\nu(t) = \kappa[\theta - \nu(t)]dt + \sigma\sqrt{\nu(t)} dW_t^{(2)} \\ d < W^{(1)}, W^{(2)} >_t = 0 \end{cases} . \tag{A.1}$$

Let $\mathbb{Q}_\lambda$ be a risk neutral measure. Transition from physical probability to risk-neutral probability can be realized by reducing drift term from a factor $\lambda\sigma\sqrt{\nu(t)}$. Girsanov's theorem set that:

$$dW_t^{\mathbb{Q}_\lambda,2} = dW_t^{(2)} + \lambda dt.$$

with $\{W_t^{\mathbb{Q}_\lambda,2}\}_{\{t\geq 0\}}$ a $\mathbb{Q}_\lambda$ brownian motion. $\tilde{S}$ is a $\mathbb{Q}_\lambda$-martingale, so necessarly with Girsanov's theorem we have:

$$dW_t^{\mathbb{Q}_\lambda,1} = dW_t^{(1)} + \frac{\mu - r - \lambda\rho\sqrt{\nu(t)}}{\sqrt{\nu(t)}\sqrt{1-\rho^2}} dt.$$

with $\{W_t^{\mathbb{Q}_\lambda,1}\}_{\{t\geq 0\}}$ a $\mathbb{Q}_\lambda$ brownian motion independent de $\{W_t^{\mathbb{Q}_\lambda,2}\}_{\{t\geq 0\}}$. The added term is obtained by using $dW_t^{(2)}$ expression in the diffusion of S in (A.1). Under risk-neutral measure $\mathbb{Q}_\lambda$ we have:

$$\begin{cases} dS_t = rS_t dt + \sqrt{\nu(t)} S_t \rho dW_t^{\mathbb{Q}_\lambda,2} + \sqrt{\nu(t)} S_t \sqrt{1-\rho^2} dW_t^{\mathbb{Q}_\lambda,1} \\ d\nu(t) = \left(\kappa[\theta - \nu(t)] - \lambda\sigma\sqrt{\nu(t)}\right) dt + \sigma\sqrt{\nu(t)} dW_t^{\mathbb{Q}_\lambda,2} \\ d < W^{\mathbb{Q}_\lambda,1}, W^{\mathbb{Q}_\lambda,2} >_t = 0 \end{cases} . \tag{A.2}$$

Or with correlated brownian motions:

$$\begin{cases} dS_t = rS_t dt + \sqrt{\nu(t)} S_t dB_t^{\mathbb{Q}_\lambda,1} \\ d\nu(t) = \left(\kappa[\theta - \nu(t)] - \lambda\sigma\sqrt{\nu(t)}\right) dt + \sigma\sqrt{\nu(t)} dB_t^{\mathbb{Q}_\lambda,2} \\ d < B^{\mathbb{Q}_\lambda,1}, B^{\mathbb{Q}_\lambda,2} >_t = \rho dt \end{cases} . \tag{A.3}$$

We set the risk-neutral framework, so we can apply Ito's lemma with diffusions under $\mathbb{Q}_\lambda$ as we do for a *Black-Scholes*'s model. Partial differential equation of Heston's article appears with a martingale argument on $\tilde{C}$. We could also use Feynman-Kac with the equation (1.8) and $\mathbb{Q}_\lambda$ the measure of calculus. Both cases leads to:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\nu S\frac{\partial^2 C}{\partial S^2} + \frac{1}{2}\nu\sigma^2\frac{\partial^2 C}{\partial \nu^2} + \nu S\sigma\rho\frac{\partial^2 C}{\partial S\partial\nu} + rS\frac{\partial C}{\partial S} + [\kappa(\theta-\nu) - \lambda\sigma\sqrt{\nu}]\frac{\partial C}{\partial \nu} - rC = 0.$$

A.2 Semi-closed price formulation for european options

Proof. (Proposition 3)

Lemma 3. *Considering f as the probability density function of X, Fourier transform is exactly caracteristic function $\phi(z) := \mathbb{E}[e^{izX}]$ and $\forall x \in \mathbb{R}$:*

$$\mathbb{Q}(X > x) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathcal{R}e\left[\frac{e^{izk}\phi_T(z)}{iz}\right] dz.$$

Proof. (Lemma 1) First point is trivial with the definition of Fourier transform:

$$\mathcal{F}f(z) := \int_{-\infty}^{+\infty} e^{izx} f(x)dx.$$

Using probability density function, this is exactly the definition of caracteristic function, so $\phi(z) = \mathcal{F}f(z)$. Note also that inverse Fourier transform gives:

$$f(x) := \frac{1}{2\pi}\int_{-\eta i-\infty}^{-\eta i+\infty} e^{-izx}\mathcal{F}f(z)dz = \frac{1}{2\pi}\int_{-\eta i-\infty}^{-\eta i+\infty} e^{-izx}\phi(z)dz.$$

Hence we can develop the following term using inverse Fourier transform:

$$\begin{aligned}
\mathbb{Q}(X > x) &= \int_x^\infty f(y)dy, \\
&= \int_x^\infty \left(\frac{1}{2\pi}\int_{-\eta i-\infty}^{-\eta i+\infty} e^{-izy}\phi(z)dz\right) dy, \\
&= \frac{1}{2\pi}\int_{-\eta i-\infty}^{-\eta i+\infty} \phi(z)\left(\int_x^\infty e^{-izy}\phi(z)dy\right) dz \quad \textit{(Fubini-Lebesgue)}, \\
&= \frac{1}{2\pi}\int_{-\eta i-\infty}^{-\eta i+\infty} \frac{e^{-izx}}{iz}\phi(z)dz.
\end{aligned}$$

Cauchy's residue theorem, for any $\eta > 0$ leads to:

$$\int_{-\eta i-\infty}^{-\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz - \int_{\eta i-\infty}^{\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz = 2\pi i.$$

Or also:

$$\frac{1}{2\pi i}\int_{-\eta i-\infty}^{-\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz = 1 + \frac{1}{2\pi i}\int_{\eta i-\infty}^{\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz.$$

By principal values (see [11]) we also have that:

$$\lim_{\eta\to 0^+}\left(\int_{-\eta i-\infty}^{-\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz + \int_{\eta i-\infty}^{\eta i+\infty} \frac{e^{-izx}}{z}\phi(z)dz\right) = 2\lim_{\epsilon\to 0^+}\int_{(-\infty,-\epsilon)\cup(+\infty,+\epsilon)} \frac{e^{-izx}}{z}\phi(z)dz.$$

Thus we get that:

$$\frac{1}{\pi i}\int_{-\eta i-\infty}^{-\eta i+\infty}\frac{e^{-izx}}{z}\phi(z)dz = 1+\frac{1}{2\pi i}\left(\int_{-\eta i-\infty}^{-\eta i+\infty}\frac{e^{-izx}}{z}\phi(z)dz+\int_{\eta i-\infty}^{\eta i+\infty}\frac{e^{-izx}}{z}\phi(z)dz\right).$$

And we can write:

$$\begin{aligned}\mathbb{Q}(X>x) &= \frac{1}{2}+\frac{1}{2\pi i}\lim_{\epsilon\to 0^+}\int_{(-\infty,-\epsilon)\cup(+\infty,+\epsilon)}\frac{e^{-izx}}{z}\phi(z)dz,\\ &= \frac{1}{2}+\frac{1}{\pi}\int_0^{+\infty}\mathcal{R}e\left[\frac{e^{-izx}}{iz}\right]\phi(z)dz.\end{aligned}$$

□

Applying previous lemma, first probability is clear since logarithm function is monotonic. For the second probability, we still have: $\tilde{\mathbb{Q}}(S_T>K)=\tilde{\mathbb{Q}}(X_T>K)$. Applying directly the previous lemma helps to get an expression with a caracteristic function $\tilde{\mathbb{Q}}$-dependent. Nevertheless we can find a relationship between the caracteristic function with respect to $\tilde{\mathbb{Q}}$ and the caracteristic function with respect to $\mathbb{Q}$. This is a simple application of Bayes formula:

$$\begin{aligned}\tilde{\phi}_T(z) &= \tilde{\mathbb{E}}\left[e^{izX_T}\right],\\ &= \tilde{\mathbb{E}}\left[\nu_T e^{izX_T}\frac{1}{\nu_T}\right],\\ &= \mathbb{E}\left[\nu_T e^{izX_T}\right],\\ &= \frac{1}{\mathbb{E}\left[e^{X_T}\right]}\mathbb{E}\left[e^{X_T}e^{izX_T}\right],\\ &= \frac{\phi_T(z-i)}{\phi_T(z)}.\end{aligned}$$

□

For convenient purpose, we rewrite here the expression of the log-spot caracteristic function:

$$\phi_T(z) = e^{C(\tau,z)\theta+D(\tau,z)\nu(t)+iz(X_t+rT)}$$

with $\tau := T-t$ where t denotes the valuation date, and

$$\omega := -\frac{1}{2}z(i+z),\quad y := \kappa-\rho\sigma iz,\quad c := \frac{\sigma^2}{2},\quad p := \sqrt{y-4\omega c},\quad y_\pm := \frac{y\pm p}{2c},\quad g := \frac{y_-}{y_+}$$

$$C(\tau,z) := \kappa\left[\tau y_- - \frac{1}{c}\ln\left(\frac{1-ge^{-p\tau}}{1-g}\right)\right],\quad D(\tau,z) := \frac{1-e^{-p\tau}}{1-ge^{-p\tau}}y_-.$$

Proof. (Proposition 4) Let $F_t := S_t e^{-rT}$. By definition:

$$\begin{aligned}\phi_T(z) := \mathbb{E}[e^{izX_T}] &= \mathbb{E}[S_T^{iz}]\quad (X_T=\ln(S_T)),\\ &= \mathbb{E}[e^{-rTiz}S_T^{iz}]e^{rTiz},\\ &= \mathbb{E}[F_T^{iz}]e^{rTiz},\\ &= \Phi(0,F,\nu)e^{rTiz},\end{aligned}$$

with $\Phi(t, F_t, \nu(t)) := \mathbb{E}[F_T^{iz}|F_t = F, \nu(t) = \nu]$. Note that $\Phi(t, F_t, \nu(t))$ as a process satisfies a partial differential equation associated to its martingale behaviour. An application of Ito's lemma to $\Phi(t, F_t, \nu(t))$ and a use of expectancy helps to detect it. Martingales terms disapear, then the Tower formula and definition of Φ allow us to ensure that $\Phi(t_1, F_{t_1}, \nu(t_1)) = \Phi(t_2, F_{t_2}, \nu(t_2))$ independently of the choice of t_1 and t_2. Thus we finally have that:

$$\partial_t \Phi + \mathcal{A}_{\nu,F}\Phi = 0,$$

with $\mathcal{A}_{\nu,F} = \kappa(\theta - \nu)\partial_\nu + \frac{1}{2}\nu F^2 \partial^2_{F^2} + \frac{1}{2}\sigma^2\nu\partial^2_{\nu^2} + \rho\sigma\nu F\partial^2_{\nu F}$. We are looking for Φ with an expression:

$$\Phi(t, F, \nu) = F_t^{iz} e^{C(\tau,z)\theta + D(\tau,z)\nu},$$

since our Heston model is an affine Markovian diffusion process, with $C(0,z)\theta + D(0,z)\nu = 0$ to satisfy final condition. Finding functions C and D comes from satisfying this final condition and the partial differential equation. As it is done in [11], we can inject this form to get ordinary differential equation on C and D. We have that:

$$\begin{aligned} \partial_t\Phi = \left[-\partial_\tau C(\tau,z)\theta - \partial_\tau D(\tau,z)\nu\right]\Phi, \quad \partial_\nu\Phi = D(\tau,z)\Phi, \\ \partial^2_{\nu^2} = D(\tau,z)^2\Phi, \quad F\partial_F\Phi = iz\Phi, \quad F^2\partial_{F^2}\Phi = iz(iz-1)\Phi, \\ \text{and} \quad F\partial^2_{\nu,F}\Phi = Diz\Phi. \end{aligned}$$

Therefore we obtain the following equations:

$$\begin{cases} \partial_\tau C(\tau,z) = \kappa D(\tau,z) \\ \partial_\tau D(\tau,z) = c(D(\tau,z) - y_+)(D(\tau,z) - y_-) \end{cases}. \tag{A.4}$$

Note that the second one is a Riccati equation. □

A.3 Advanced formulation of semi-closed price

We have seen that the following expression can be verify:

$$\begin{aligned} \tilde{\mathbb{Q}}(S_T > K) &= \int_l^{+\infty} e^x \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega x}\phi_T(\omega)d\omega\right)dx, \\ &= 2\frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_l^{+\infty} e^{-i(\omega+i)x}dx\right)d\omega \quad \textit{(Fubini)}. \end{aligned}$$

We denote as $A := \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_l^{+\infty} e^{-i(\omega+i)x}dx\right)d\omega$ then we have, by decomposition *(Chasles)*:

$$\tilde{\mathbb{Q}}(S_T > K) = A + \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_{-\infty}^{+\infty} e^{-i(\omega+i)x}dx + \int_l^{-\infty} e^{-i(\omega+i)x}dx\right)d\omega.$$

As it is recall above we have that $\tilde{\mathbb{Q}}(S_T > K)|_{l=-\infty} = 2\frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\int_{-\infty}^{+\infty} e^{-i(\omega+i)x}dxd\omega$ but by definition of cumulative distribution function we have that $\tilde{\mathbb{Q}}(S_T > K)|_{l=-\infty} = 1$.

Therefore we can write that:

$$\begin{aligned}
\tilde{\mathbb{Q}}(S_T > K) &= A + \frac{1}{2} - \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_{-\infty}^{l} e^{-i(\omega+i)x}dx\right)d\omega, \\
&= \frac{1}{2} + \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_{l}^{+\infty} e^{-i(\omega+i)x}dx\right)d\omega - \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left(\int_{-\infty}^{l} e^{-i(\omega+i)x}dx\right)d\omega, \\
&= \frac{1}{2} + \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{1}{-i(\omega+i)} - \frac{e^{-i(\omega+i)l}}{-i(\omega+i)}\right]d\omega, \\
&- \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{e^{-i(\omega+i)l}}{-i(\omega+i)} - \lim_{R\to\infty}\frac{e^{i(\omega+i)R}}{-i(\omega+i)}\right]d\omega, \\
&= \frac{1}{2} + \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\lim_{R\to\infty}\frac{e^{-i(\omega+i)R}}{-i(\omega+i)} - \frac{e^{-i(\omega+i)l}}{-i(\omega+i)}\right]d\omega, \\
&- \frac{1}{4\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{e^{-i(\omega+i)l}}{-i(\omega+i)} - \lim_{R\to\infty}\frac{e^{i(\omega+i)R}}{-i(\omega+i)}\right]d\omega, \\
&= \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\frac{e^{-i(\omega+i)l}}{-i(\omega+i)}d\omega + \frac{1}{4\pi}\lim_{R\to\infty}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{e^{-i(\omega+i)R}}{-i(\omega+i)} + \frac{e^{i(\omega+i)R}}{-i(\omega+i)}\right]d\omega, \\
&= \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\frac{e^{-i(\omega+i)l}}{i(\omega+i)}d\omega - \frac{1}{4\pi}\lim_{R\to\infty}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{e^{-i(\omega+i)R} + e^{i(\omega+i)R}}{i(\omega+i)}\right]d\omega.
\end{aligned}$$

Let B defined as $B := \frac{1}{4\pi}\lim_{R\to\infty}\int_{-\infty}^{+\infty}\phi_T(\omega)\left[\frac{e^{-i(\omega+i)R}+e^{i(\omega+i)R}}{i(\omega+i)}\right]d\omega$, the last term of the above equation. We can compute B by residual theorem:

$$\begin{aligned}
B &= \frac{1}{2}\left(\frac{1}{2\pi}\lim_{R\to\infty}\int_{-\infty}^{+\infty}\phi_T(\omega)\frac{e^{i(\omega+i)R}}{i(\omega+i)} - \frac{1}{2\pi}\lim_{R\to\infty}\int_{-\infty}^{+\infty}\phi_T(-u)\frac{e^{i(u-i)R}}{i(i-u)}du\right), \\
&= \frac{-\phi_T(-i)}{2},
\end{aligned}$$

Thus by definition, we can write that

$$\begin{aligned}
\phi_T(-i) &= -2B, \\
&= \lim_{R\to\infty}\frac{1}{2\pi i}\int_{-\infty}^{\infty}\phi_T(\omega)\frac{e^{i(\omega+i)R} - e^{-i(\omega+i)}}{\omega+i}d\omega, \\
&= \tilde{\mathbb{Q}}(S_T > K)|_{l=-\infty}, \\
&= 1.
\end{aligned}$$

It comes, as it is mentionned in §2.1, that:

$$\tilde{\mathbb{Q}}(S_T > K) = 1 + \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_T(\omega)\frac{e^{-i(\omega+i)l}}{i(\omega+i)}d\omega.$$

Since both probabilities are expressed with ϕ_T we can regroup terms and find the expression given in (2.5).

A.4 Greeks computation in the Heston model

We recall here the formula which helps sampling S_T from the approximated law of each scheme:

$$S_T = S\xi e^{\left(r-\frac{\sigma^2}{2}\right)T+\bar{\sigma}\sqrt{T}Z} \tag{A.5}$$
$$\xi := e^{-\frac{\rho^2}{2}\int_0^T \nu_s ds+\rho\int_0^T \sqrt{\nu_s}dW_s^\nu} \quad \text{and} \quad \bar{\sigma}^2 := \frac{(1-\rho^2)\int_0^T \nu_s ds}{T}.$$

Defining g as the density of the random variable S_T, we have:

$$\begin{aligned}
g(x) &:= \frac{d}{dx}\mathbb{P}(S_T \le x),\\
&= \frac{d}{dx}\mathbb{P}(\ln(S_T) \le \ln(x)), \quad (\ln \textit{ is an increasing function})\\
&= \frac{d}{dx}\mathbb{P}\left(\ln(S) + \left(r-\frac{\bar{\sigma}^2}{2}\right)T + \bar{\sigma}\sqrt{T}Z \le \ln(x)\right),\\
&= \frac{d}{dx}\mathcal{N}(k(x)), \quad (\mathcal{N} \textit{ CDF of a standard gaussian r.v.})\\
&= \frac{d}{dk}\mathcal{N}(k)\frac{d}{dx}(k(x)),\\
&= \frac{f(k(x))}{x\bar{\sigma}\sqrt{T}}, \quad (f \textit{ PDF of a standard gaussian r.v.}).
\end{aligned}$$

We can then compute option greeks, for the Pathwise and Likelihood Ratio approaches.

Likelihood ratio

$$\begin{aligned}
\frac{\partial_S g(x)}{g(x)} &:= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{1}{x\bar{\sigma}\sqrt{T}}\partial_k(f(k))\partial_S(k(x))\right),\\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{1}{x\bar{\sigma}\sqrt{T}}(-k(x)f(k(x)))\frac{-1}{S\bar{\sigma}\sqrt{T}}\right),\\
&= \frac{k(x)}{S\bar{\sigma}\sqrt{T}}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2_{SS} g(x)}{g(x)} &:= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\partial_S\left(\frac{k(x)}{Sx\bar{\sigma}^2T}f(k(x))\right),\\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{1}{Sx\bar{\sigma}^2T}\left[\partial_S(k(x)f(k(x)))\right] - \frac{k(x)f(k(x))}{S^2x\bar{\sigma}^2T}\right),\\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{1}{Sx\bar{\sigma}^2T}\left[\frac{-f(k(x))}{S\bar{\sigma}\sqrt{T}} + \frac{k(x)^2f(k(x))}{S\bar{\sigma}\sqrt{T}}\right] - \frac{k(x)f(k(x))}{S^2x\bar{\sigma}^2T}\right),\\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{k(x)^2f(k(x)) - f(k(x))}{S^2x\bar{\sigma}^2T\bar{\sigma}\sqrt{T}} - \frac{k(x)f(k(x))}{S^2x\bar{\sigma}^2T}\right),\\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(f(k(x))\frac{k(x)^2 - k(x)\bar{\sigma}\sqrt{T} - 1}{S^2x\bar{\sigma}^2T\bar{\sigma}\sqrt{T}}\right),\\
&= \frac{k(x)^2 - k(x)\bar{\sigma}\sqrt{T} - 1}{S^2\bar{\sigma}^2T}.
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial_T g(x)}{g(x)} &:= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\partial_T\left(\frac{1}{x\bar{\sigma}\sqrt{T}}f(k(x))\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{2x\bar{\sigma}T\sqrt{T}}+\frac{1}{x\bar{\sigma}\sqrt{T}}\partial_k f(k)\partial_T k(x)\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{2x\bar{\sigma}T\sqrt{T}}+\frac{1}{x\bar{\sigma}\sqrt{T}}\left(-k(x)f(k(x))\right)\partial_T k(x)\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{2x\bar{\sigma}T\sqrt{T}}+\frac{1}{x\bar{\sigma}\sqrt{T}}\left(-k(x)f(k(x))\right)\left[-\frac{\ln(\frac{x}{S\xi})+(r-\frac{\bar{\sigma}^2}{2})T}{2\bar{\sigma}T\sqrt{T}}\right]\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(f(k(x))\frac{1}{x\bar{\sigma}\sqrt{T}}\left[\frac{k(x)\left(\ln(\frac{x}{S\xi})+(r-\frac{\bar{\sigma}^2}{2})T\right)-\bar{\sigma}\sqrt{T}}{2\bar{\sigma}T\sqrt{T}}\right]\right), \\
&= \frac{k(x)\left(\ln(\frac{x}{S\xi})+(r-\frac{\bar{\sigma}^2}{2})T\right)}{2\bar{\sigma}T\sqrt{T}}-\frac{1}{2T}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial_{\bar{\sigma}} g(x)}{g(x)} &:= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\partial_{\bar{\sigma}}\left(\frac{1}{x\bar{\sigma}\sqrt{T}}f(k(x))\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{x\bar{\sigma}^2\sqrt{T}}f(k(x))+\frac{1}{x\bar{\sigma}\sqrt{T}}\partial_{\bar{\sigma}} f(k(x))\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{x\bar{\sigma}^2\sqrt{T}}f(k(x))+\frac{1}{x\bar{\sigma}\sqrt{T}}(-k(x)f(k(x)))\partial_{\bar{\sigma}} k(x)\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(-\frac{f(k(x))}{x\bar{\sigma}^2\sqrt{T}}f(k(x))+\frac{1}{x\bar{\sigma}\sqrt{T}}(-k(x)f(k(x)))\left[-\frac{\ln(\frac{x}{S\xi})-(r+\frac{\bar{\sigma}^2}{2})T}{\bar{\sigma}^2\sqrt{T}}\right]\right), \\
&= \frac{x\bar{\sigma}\sqrt{T}}{f(k(x))}\left(\frac{f(k(x))}{x\bar{\sigma}\sqrt{T}}\frac{k(x)\left[\ln(\frac{x}{S\xi})-(r+\frac{\bar{\sigma}^2}{2})T\right]-\bar{\sigma}\sqrt{T}}{\bar{\sigma}^2\sqrt{T}}\right), \\
&= \frac{k(x)}{\bar{\sigma}^2\sqrt{T}}\left[\ln(\frac{x}{S\xi})-(r+\frac{\bar{\sigma}^2}{2})T\right]-\frac{1}{\bar{\sigma}}.
\end{aligned}
$$

Pathwise greeks estimators

$$
\begin{aligned}
\Delta &:= \frac{d}{dS}\mathbb{E}\left[(S_T-K)_+\right], \\
&= \mathbb{E}\left[\frac{d}{dS}\left(S\xi e^{\left(r-\frac{\sigma^2}{2}\right)T+\bar{\sigma}\sqrt{T}Z}-K\right)1_{S_T\geq K}\right], \\
&= \mathbb{E}\left[\frac{S_T}{S}1_{S_T\geq K}\right].
\end{aligned}
$$

$$
\begin{aligned}
\Gamma &:= \frac{d}{dS}\mathbb{E}\left[(S_T-K)_+\frac{k(S_T)}{S\bar{\sigma}\sqrt{T}}\right], \quad \textit{(from the Likelihood Ratio estimator)} \\
&= \mathbb{E}\left[\frac{d}{dS}\left(-\frac{Kk(S_T)}{S\bar{\sigma}\sqrt{T}}\right)1_{S_T\geq K}\right], \\
&= \mathbb{E}\left[\frac{Kk(S_T)}{S^2\bar{\sigma}\sqrt{T}}1_{S_T\geq K}\right].
\end{aligned}
$$

$$
\begin{aligned}
\Theta &:= -\frac{d}{dT}\mathbb{E}\left[(S_T-K)_+\right], \\
&= \mathbb{E}\left[-\frac{d}{dT}\left(S\xi e^{\left(r-\frac{\sigma^2}{2}\right)T+\bar{\sigma}\sqrt{T}Z}-K\right)1_{S_T\geq K}\right], \\
&= \mathbb{E}\left[-\left[r-\frac{\bar{\sigma}^2}{2}+\frac{\bar{\sigma}Z}{2\sqrt{T}}\right]S_T1_{S_T\geq K}\right].
\end{aligned}
$$

$$
\begin{aligned}
\vartheta &:= \frac{d}{d\bar{\sigma}}\mathbb{E}\left[(S_T-K)_+\right], \\
&= \mathbb{E}\left[\frac{d}{d\bar{\sigma}}\left(S\xi e^{\left(r-\frac{\sigma^2}{2}\right)T+\bar{\sigma}\sqrt{T}Z}-K\right)1_{S_T\geq K}\right], \\
&= \mathbb{E}\left[\left(-\bar{\sigma}T+\sqrt{T}Z\right)S_T1_{S_T\geq K}\right],
\end{aligned}
$$