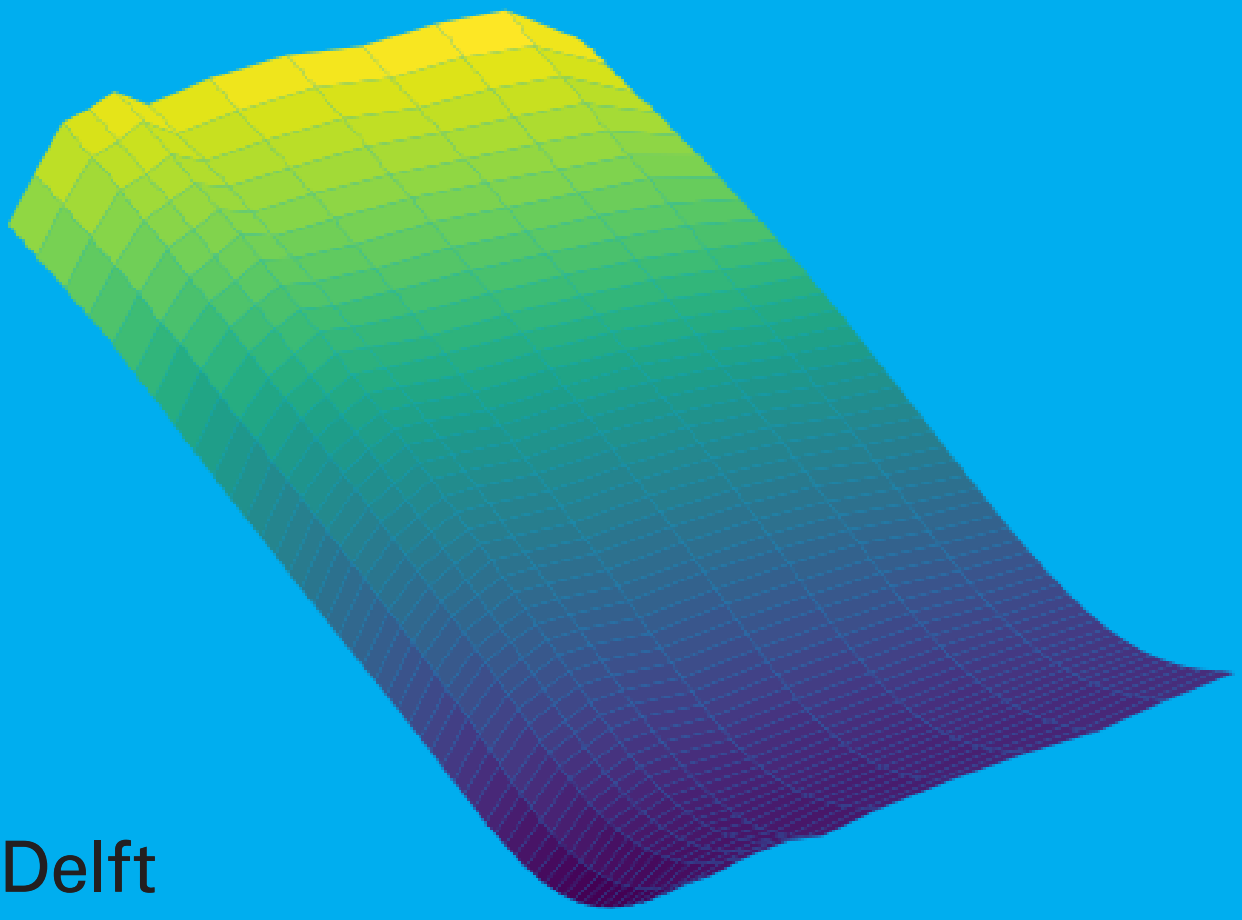
Comparing Rough Volatility Models

B.L. van Gisbergen



TU Delft

Comparing Rough Volatility Models

by

B.L. van Gisbergen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday June 23, 2022 at 14:30.

Student number:	4604253	
Project duration:	September 1, 2021 – June 23, 2022	
Thesis committee:	Prof. Dr. A. Papapantoleon,	TU Delft, supervisor
	Dr. F. Fang,	TU Delft
	Dr. M. Thul,	IMC Trading

An electronic version of this thesis is available at
http://resolver.tudelft.nl/uuid:cfc37681-3b2f-45dc-bf35-ea1ddc81eadb.

TU Delft

Abstract

Since the introduction of rough volatility there have been numerous attempts at combining it with existing models in order to better approximate the volatility surface with a low number of parameters. The drawback of rough volatility is usually the time needed to compute a volatility surface. We compare three major rough volatility models and compare their ability to fit to the market volatility surface. We implement the rough Bergomi, rough Heston and lifted Heston models and introduce a fourth model by taking a Nelson-Siegel parameterization for the instantaneous forward variance curve in the rough Bergomi model. We then minimize the volatility surfaces of the models to that of the market and compare these results through minimization time, fit and hedging possibilities. We find that for both Bergomi models, a single volatility surface is seven times faster to compute than for the lifted Heston model, which in turn is five times faster to compute than for the rough Heston model. Minimization times for the rough Heston models are comparable, however significantly higher than for the rough Bergomi models. We also find that the rough Bergomi model is under-parameterised, this is however fixed by the Nelson-Siegel parameterization, which has a minimization error in line with that of both Heston models. Finally, hedging specifically against a parameter rarely improves the outcome, although a hedge against the instantaneous forward variance curve is consistently good and can improve the error between only a delta hedge by 25-50% throughout all models.

Preface

The benefit of rough volatility is to be able to model the whole volatility surface with few parameter, which is promising in the search for models that are not over-parameterised. This work aims at comparing these rough volatility models both in how well they describe and fit the market as well as how fast they are to compute. The goal is to see if it is computationally feasible to implement rough volatility models in practice. It turns out that rough volatility models differ in performance, however all have their advantages and disadvantages. The choice for a single model relies strongly on the motivation for fitting and the context.

This thesis was done in collaboration with IMC Trading, who have been so kind to supply me with market data and their pricing infrastructure.

Mostly I would like to thank Matthias Thul for guiding me throughout this thesis by asking critical questions and sharing his knowledge of derivatives pricing with me. Secondly I would also like to thank Antonis Papapantoleon for helping me for the duration of this thesis and being an insightful supervisor. Next to this I would also like to thanks Fang Fang for being a part of my thesis committee. Finally, a big thank you to all my friends and family who have supported me throughout this time!

Bart van Gisbergen
Delft, June 2022

Contents

1. Introduction

Since the introduction of the Black-Scholes model [1] in 1973 there have been many attempts at modeling the derivatives market [2]. So far the most succesful models are those with a stochastic volatility, where both the stock process and the volatility process are driven by a Brownian motion [3]. Since *Volatility is Rough* [4] by Gatheral, Jaisson and Rosenbaum in 2014 there has been a lot of research towards models where the volatility is driven by a fractional Brownian motion, the difference being that fractional Brownian motion is a non-Markovian process. These models are called rough volatility models since the volatility paths are rough in the Hölder sense. Gatheral et al. show that a fractional Brownian motion would be better at describing the at-the-money skew of the volatility smile as well as the realized volatility time series than previous stochastic volatility models. However the fractional Brownian motion, being non-Markovian, induces little closed form solutions and an increase in computational complexity.

The first rough volatility model was the rough Bergomi model [5]. This model is based on the n-dimensional Bergomi model [2], but includes a fractional Brownian motion. Pricing under such a model is done using a Monte Carlo method, for which there has since been a lot of research to speed up convergence, see [6], [7] and [8]. This model enjoys the fact that it only has three parameters together with the forward variance curve as an input.

A second, major, rough volatility model introduced was the rough Heston model [9]. This is comparable to the classical Heston model [3], but where the Brownian motion is again replaced by a fractional Brownian motion. It is then shown in [10] that this model also only has three parameters together with the forward variance curve as an input. The affinity of the model has been proven [11], thus it is natural to use the COS method to price options. This, however still takes a lot of computation time.

This problem was attempted to be solved in [12], where the fractional kernel is replace by a sum of deltas, thus taking a Markovian approximation. The model was named the lifted Heston model. Computationally, it changes the problem of solving the characteristic function: from solving a fractional Riccati equation to solving an n-dimensional Riccati equation. However, this still needs numerical approximations. This model also makes an approximation of the forward variance curve instead of using it as an input, adding three parameters to the model, bringing it up to six.

The question now becomes, which of these models is most suitable for practical use? In an attempt to answer this question this work fits these models to the market volatility surface. We compare not only the fit to the market, but also how fast this fit can be computed and how well we can predict future surfaces using our fit.

We find that the rough Bergomi model is under-parameterized, but with a Nelson-Siegel parameterization of the forward variance curve, the fit is just as good as the rough Heston models, while also being faster in computation time. Next to that we find that the rough Heston model requires a large computation time for a single volatility surface, but this is compensated by the fact that it has few parameters to minimize. It does an impressive job at fitting the market volatility surface for only three parameters. The lifted Heston model improves the computation time for a single volatility surface by a factor of five, but this does not improve the total minimization time considerably because of the increased complexity of the optimization space. Also, the parameters in this model do not seem to be orthogonal. For a practical application a combination of these two models would be a good consideration. This report starts off with a background of all three models in chapter 2. This is followed by an explanation on how we price European options in the Bergomi model and Heston models in chapter 3 and 4, respectively. The following chapter 5 discusses how the instantaneous forward variance curve is determined form the market data. Chapter 6 discusses the method used for minimizing the model volatility surface to that of the market. Then, chapter 7 explains how we compare the obtained results. Subsequently, chapter 8 presents the obtained results, which is followed by a discussion of these results in chapter 9. Finally, chapter 10 compares the obtained results and draws conclusions.

2. Background

We start by introducing rough volatility. This is done by first discussing what a rough path is, leading to the definition of the fractional Brownian motion. After that there is a motivation for the usage of fractional Brownian motion to describe the path of the volatility process. Finally we discuss how this is incorporated into the Bergomi and Heston stochastic volatility models, leading to the rough Bergomi, rough Heston and lifted Heston models.

2.1. Rough Paths

The theory of rough paths was developed in order to develop the theory on stochastic differential equations in a pathwise setting. Rather than rely on probabilistic properties of processes (semimartingale structure, adaptedness, ...) instead the theory relies on the regularity conditions on the paths of a process. In order to define the regularity of a process we define α-Hölder continuity, see [13]

Definition 2.1. *We say $X : [0,T] \to E$ is α-***Hölder continuous** *for $\alpha > 0$ if $\|x\|_\alpha < \infty$, where*

$$\|x\|_\alpha := \sup_{t \neq s} \frac{|X_t - X_s|}{|t-s|^\alpha}.$$

$\|.\|_\alpha$ is a seminorm, in order to define a Banach space we define

Definition 2.2. $\mathcal{C}^\alpha = \{X : \|X\|_\alpha < \infty\}$ *with norm* $\|X\|_{\mathcal{C}^\alpha} = |X_0| + \|X\|_\alpha$

An α-Hölder rough path is a path which is α'-Hölder continuous for every $\alpha' > \alpha$ and where $\alpha \leq 1/2$. There is one version of a rough path that is applied to rough volatility, and that is the fractional Brownian motion, as defined by Mandelbrot and Van Ness [14].

Definition 2.3. *For a $H \in (0,1)$ and $t > 0$ we define the* **fractional Brownian motion (fBM)** *process*

$$B_H(t) = B_H(0) + \frac{1}{\Gamma(H+\frac{1}{2})} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} \right] dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s) \right\},$$

where $(B(s))_{0 \leq s \leq t}$ is a standard Brownian motion.

A property of fBM is (see [14])

$$\mathbb{E}[B_H(t) B_H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}). \tag{2.1}$$

From this it is clear that taking $H = \frac{1}{2}$ creates independent increments, which leads to the standard Brownian motion. In the case where $H \leq 1/2$ we see that $B_H(t)$ is an α-Hölder rough path for all $0 < \alpha < H$. H is called the Hurst index, from now on we only consider $H \leq 1/2$.
In order to give some intuition on fBM, Figure 1 shows a regular Brownian motion and a fractional Brownian motion using the same underlying sampling.

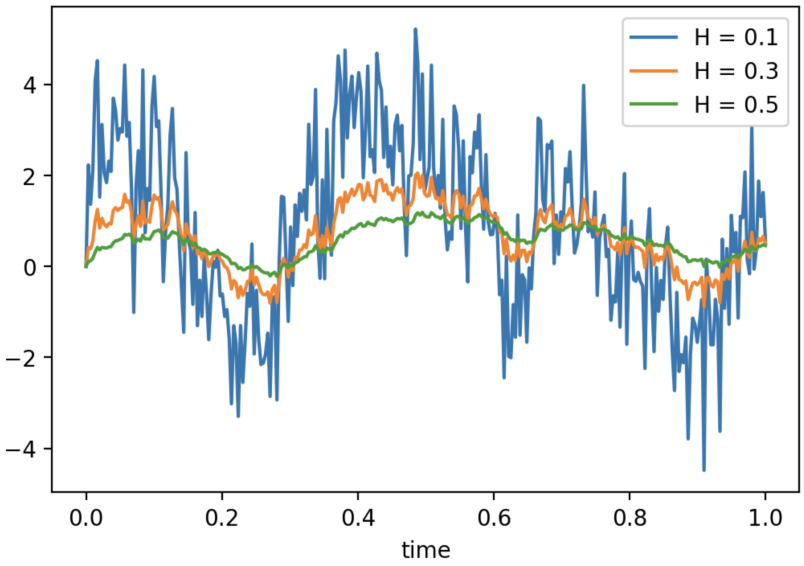


Figure 1: Paths of fractional Brownian motions for different Hurst indices with $\Delta t = 1/300$

As can be seen, the path of the fBM follows that of the Brownian motion. However, it is much rougher in the sense that the path is less well behaved.
Finally from equation (2.1) we note that $\mathbb{E}[B_H(t)^2] = |t|^{2H}$, this means that when comparing fBM to standard BM we get a larger variance for $t < 1$ and a smaller variance for $t > 1$. The distribution is Gaussian, meaning that compared to a Brownian motion, fBM has a much wider distribution for short times, but a narrower distribution for larger times.

Together with fractional Brownian motion we define the fractional derivative D^r and the fractional integral I^r for $r \in (0, 1]$ as in [14] as

$$\begin{aligned} D^r f(t) &= \frac{1}{\Gamma(r)} \int_0^t (t-s)^{r-1} f(s) ds \\ I^r f(t) &= \frac{1}{\Gamma(1-r)} \frac{d}{dt} \int_0^t (t-s)^{-r} f(s) ds. \end{aligned} \tag{2.2}$$

2.2. Rough Volatility

As first mentioned by Gatheral, Jaisson and Rosenbaum [4], when we artificially model the at-the-money (ATM) skew for different expiries we notice that we can fit a power function through it of the form $A \cdot T^{\alpha-1}$, where $\alpha - 1 < 0$. This can be seen by plotting the ATM skew ($\psi(T)$) defined as

$$\psi(T) := \left| \frac{\partial}{\partial k} \sigma_{\text{imp}}(k, T) \right|_{k=0}, \tag{2.3}$$

where k is log-moneyness, T is time to expiry and σ_{imp} is the implied volatility. Plotting this function and the fit gives the following figure

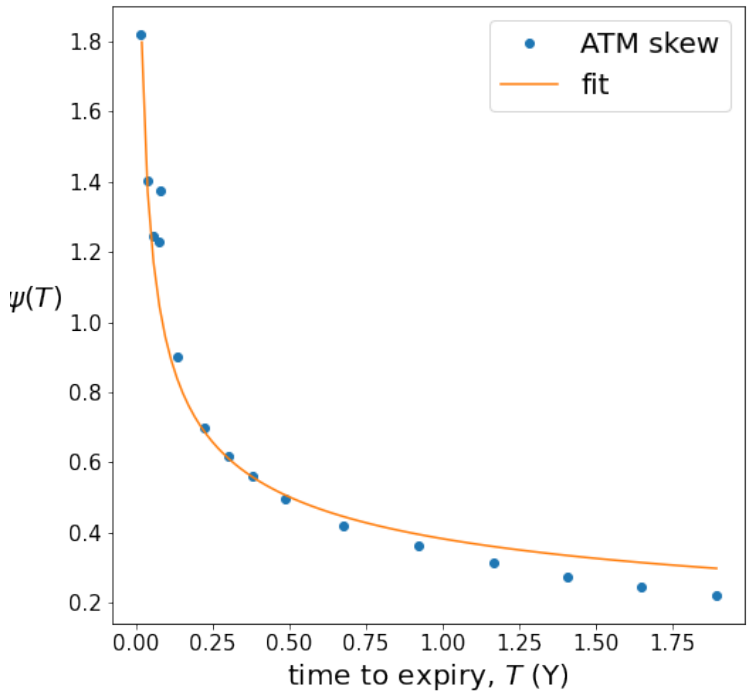


Figure 2: ATM skew and plot of $0.4 \cdot T^{-0.39}$. All data is taken from the Eurostoxx volatility curve on 1/11/2021. Approximation of the ATM skew is done using a numerical symmetric derivative. The fit excludes the somewhat strange 5^{th} observation.

In section 3.3 of [15], Fukasawa states that considering a simple stochastic volatility model where the volatility (vol) is driven by a fractional Brownian motion would result in an ATM volatility skew of order $T^{H-1/2}$ with H the Hurst parameter of the fBM. This is the main reason for considering stochastic volatility models where the vol follows a rough path, specifically a fBM.

2.3. Rough Bergomi

The first pricing model with rough volatility was introduced by Bayer, Friz and Gatheral [5]. They start by considering the n-factor Bergomi model [2] where the instantaneous forward variance curve $\xi_t(u)$ is given by

$$\xi_t(u) = \xi_0(u)\mathcal{E}\left(\sum_{i=1}^{n} \eta_i \int_0^t e^{-\kappa_i(u-s)} dW_s^i\right). \tag{2.4}$$

The instantaneous forward variance curve (FVC) observed at time t is defined as $\xi_t(u) := \mathbb{E}[V_u|\mathcal{F}_t]$ for $u > t$ with V_u being the instantaneous variance of the log-stock process. $\mathcal{E}(\cdot)$ is the stochastic exponential defined as $\mathcal{E}(Z_t) := \exp(Z_t - Z_0 - \frac{1}{2}[Z,Z]_{0,t})$.
Bergomi now states that the forward variance curve is determined by n Ornstein-Uhlenbeck type processes. For $n = 1$ this would translate to

$$\begin{cases} \xi_t(u) = \xi_0(u) \exp\left(\eta e^{-\kappa(u-t)} Y_t - \frac{1}{2}\eta^2 \exp^{-2\kappa(u-t)} \mathbb{E}[Y_t^2]\right) \\ dY_t = -\kappa Y_t dt + dW_t, \end{cases} \tag{2.5}$$

where $(W_s)_{0 \le s \le t}$ is a Brownian motion. The problem now becomes that, when calibrating to option prices, $n = 1$ gives an underparameterized model, while $n = 2$ is often overparameterized. The idea proposed in [5] is to replace the exponential kernel in (2.4) with a

fractional kernel. Thus

$$\xi_t(u) = \xi_0(u)\mathcal{E}\left(\eta \int_0^t \frac{dW_s}{(u-s)^{1-\alpha}}\right) \tag{2.6}$$

where $\alpha = H + 1/2$ with H being the Hurst index of the fractional Brownian motion. When assuming a constant initial instantaneous forward variance curve of $\xi_0(u) \coloneqq \sigma_0^2$, this translates to a vol process given by

$$\sigma_t = \sqrt{\xi_t(t)} = \sigma_0 \sqrt{\mathcal{E}\left(\eta\sqrt{2H}\int_0^t \frac{dW_s}{(t-s)^{\gamma}}\right)} = \sigma_0 \exp\left\{\frac{\eta}{2}V_t - \frac{\eta^2}{4}\mathbb{E}[V_t^2]\right\} \tag{2.7}$$

where $V_t = \sqrt{2H}\int_0^t \frac{dW_s}{(t-s)^{1-\alpha}}$, which is proportional to the Riemann-Liouville representation of fBM such that $\mathrm{Var}[V_t] = t^{2H}$
Translating this to a price process results in

$$\begin{cases} S_t = S_0\mathcal{E}\left(\int_0^t \sqrt{V_u}dB_u\right) \\ V_u = \xi_0(u)\mathcal{E}\left(\eta\sqrt{2H}\int_0^u \frac{1}{(u-s)^{1-\alpha}}dW_s\right) \end{cases} \tag{2.8}$$

where $\langle dB_t dW_t\rangle = \rho dt$. Throughout this report $(B_s)_{0\le s\le t}$ and $(W_s)_{0\le s\le t}$ are two Brownian motions with correlation coefficient ρ, unless otherwise stated.

This aim of this model it to use the forward variance curve as a base for the vol curve and models the vol by multiplying this with an exponential martingale driven by a fractional Brownian motion with Hurst index H. The fractional Brownian motion is multiplied by the volatility of volatility (vol-of-vol) parameter η and it is correlated to the Brownian motion of the stock process with correlation coefficient ρ.

In order to be unreliant on the initial instantaneous forward variance curve induced from the market, we introduce the Nelson-Siegel approximation of the forward variance curve. This is defined by

$$\xi_0(t) = \beta_0 + \beta_1\frac{1-e^{-t/\tau}}{t/\tau} + \beta_2\left(\frac{1-e^{-t/\tau}}{t/\tau} - e^{-t/\tau}\right).$$

This will add four parameters to the model ($\beta_0, \beta_1, \beta_2$ and τ), we name this the 7 parameter rough Bergomi model, abbreviated as rB7p. Note that we consider the initial instantaneous forward variance curve starting at $t = 0$, thus we do not specify a stochastic process for the parameters. This means we do not need to worry about the consistency of the FVC. Consistency meaning that the instantaneous variance V_t must always be a non-negative martingale. This then ensures that the price process of variance swaps is free of arbitrage. See [16] for more details on consistency.

2.4. Rough Heston

A second major model incorporating rough volatility is the rough Heston model introduced by El Euch and Rosenbaum [11]. The idea behind this model is to first consider the classical Heston model given by

$$\begin{cases} dS_t = S_t\sqrt{V_t}dB_t \\ dV_t = \lambda(\theta - V_t)dt + \nu\sqrt{V_t}dW_t \\ \langle dB_t dW_t\rangle = \rho dt. \end{cases} \tag{2.9}$$

El Euch and Rosenbaum now introduce a rough volatility process by introducing the kernel $(t-s)^{\alpha-1}$, where $\alpha = H + 1/2$. Doing this leads to the following volatity process

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\left(\lambda(\theta(s) - V_s)ds + \nu\sqrt{V_s}dW_s\right). \tag{2.10}$$

It is then observed by El Euch and Rosenbaum [10] that we can simplify the rough Heston model if we take into consideration the initial instantaneous forward variance curve: $\xi_0(t) = \mathbb{E}[V_t]$ for $t > 0$. They show that

$$\lambda\theta(s) = D^{\alpha}(\mathbb{E}[V_\cdot] - V_0)(t) + \lambda\mathbb{E}[V_t] \tag{2.11}$$

As observed in [4], smiles flatten consistently with the scaling properties of historical vol, thus there is no need for an extra mean reversion parameter λ, hence we can consider $\lambda \to 0$. Taking this assumption and plugging (2.11) back into equation (2.10), we can rewrite the rough Heston model as follows

$$\begin{cases} dS_t = S_t\sqrt{V_t}dB_t \\ V_t = \xi_0(t) + \frac{\nu}{\Gamma(H+1/2)}\int_0^t \frac{\sqrt{V_s}}{(t-s)^{1-\alpha}}dW_s \\ \langle dB_t dW_t\rangle = \rho dt, \end{cases} \tag{2.12}$$

where $\xi_0(t)$ can be obtained from the market, as explained in Chapter 5. We will be using this version of the rough Heston model from now on.

Intuitively, this model aims to set a base vol implied by the market through the instantaneous forward variance curve. It then describes the vol process with the fractional Brownian motion, driven by the Hurst parameter H, and the vol-of-vol parameter ν. Finally we have a correlation coefficient, ρ, between the Brownian motions.

2.5. Lifted Heston

After observing that pricing under the rough Heston model was very computationally intenstive, Abi Jaber [12] introduced the lifted Heston model. In this model we replace the fractional kernel in the volatility process by a sum of delta peaks. This then results in the following system

$$\begin{cases} dS_t = S_t\sqrt{V_t}dB_t \\ V_t = g_0(t) + \sum_{i+1}^n c_i U_t^i \\ dU_t^i = (-x_i U_t^i - \lambda V_t)dt + \nu\sqrt{V_t}dW_t \\ \langle dB_t dW_t\rangle = \rho dt. \end{cases} \tag{2.13}$$

The idea here is that the forward variance curve is given by $g_0(t)$ and on top of this we add a weighted sum of zero mean processes U^i with weight c_i. These processes start at zero and are driven by the same Brownian motion, but mean revert at different speeds (x_i).
If we take expectations on both sides of the volatility process we obtain

$$\mathbb{E}[V_t] + \lambda\sum_{i=1}^n c_i \int_0^t e^{-x_i(t-s)}\mathbb{E}[V_s]ds = g_0(t). \tag{2.14}$$

We parameterize $g_0(t)$ by defining

$$g_0(t) \coloneqq V_0 + \lambda\theta\sum_{i+1}^n c_i\int_0^t e^{-x_i(t-s)}ds = V_0 + \lambda\theta\sum_{i=1}^n \frac{c_i}{x_i}(1-e^{-x_i t}), \tag{2.15}$$

where $V_0, \theta > 0$, this is in line with (2.9).

The idea behind the model is to set constants x_i, c_i such that we get a good Markovian approximation of a rough process, this is explained in section 4.4. Next to that we have the same H, η and ρ parameters as before. We have introduced the θ, λ and V_0 parameters and with these describe the instantaneous forward variance curve (FVC) $g_0(t)$. θ can be seen as the long term variance, λ as a measure of speed of mean reversion to this long term variance and V_0 as the base variance.

In this work we choose to let the FVC of the lifted Heston model be independent of the market induced FVC. This choice is based on the fact that this is an approximation of the rough Heston model. Since an analysis of the level of approximation is done in [12] it is interesting to see how these models compare with and without the use of the information of the instantaneous forward variance curve obtained from the market.

3. Pathwise Simulation

There is no closed form solution to the rough Bergomi model from (2.8), thus a path simulation is needed in order to price options. The idea of a path simulation is to do a Monte Carlo simulation of possible paths of the underlying and use these to price the option. This chapter starts by discussing an approximation scheme for integrals [6], followed by applying this technique to the rough integral in (2.8). This greatly decreases the amount of time steps needed in the simulation. The chapter concludes with an explanation of the Romano Touzi trick [17], a method which utilizes the full path in order to price products with a closed form solution to the Black-Scholes equation. Using this method decreases the amount of paths needed to price far out-of-the-money and in-the-money options.

3.1. Hybrid Scheme

The hybrid scheme was introduced by Bennedsen, Lunde and Pakkanen [6]. Using this scheme greatly improves the computation time as our approximation for the fractional integral around the origin is better. The goal of the scheme is to give an approximation of

$$X(t) = \int_{-\infty}^{t} g(t-s)\sigma(s)dW_s,$$

where $g(x) = x^\alpha L_g(x)$ for $x \in (0,1], \alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ and $L_g(x)$ is a slowly varying function at 0. Slowly varying meaning $\lim_{x\to 0} \frac{L(tx)}{L(x)} = 1$ for any $t > 0$. In our case we have $\sigma(s) = 1$. The idea behind the method then becomes to approximate the integral by a sum and split this sum into two parts

$$\begin{aligned}\int_0^t (t-s)^\alpha dW_s &\approx \sum_{k=1}^{\kappa} L_g\left(\frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k}{n}+\frac{1}{n}} (t-s)^\alpha dW_s + \sum_{k=\kappa+1}^{N_n} g\left(\frac{b_k^*}{n}\right)\left(W_{t-\frac{k}{n}+\frac{1}{n}} - W_{t-\frac{k}{n}}\right) \\ &:= \sum_{k=1}^{\kappa} L_g\left(\frac{k}{n}\right) W_{i-k,k} + \sum_{k=\kappa+1}^{N_n} g\left(\frac{b_k^*}{n}\right) W_{i-k,0}.\end{aligned} \tag{3.1}$$

By Proposition 2.8 of [6] we must take

$$b_k^* = \left(\frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha+1}\right)^{1/\alpha}, \quad k \geq \kappa + 1,$$

in order to minimize the asymptotic mean squared error (MSE), defined as

$$\lim_{n\to\infty} \mathbb{E}\left[\left|X(t) - \hat{X}_n(t)\right|^2\right],$$

where $\hat{X}_n(t)$ approximates $X(t)$. In equation (3.1) we define the terms $W_{i,j}$. We now note that we can simulate these terms by simulating an order-$(\kappa+1)$ i.i.d. multivariate Gaussian distribution with mean zero and covariance matrix Σ given by

$$\begin{aligned}\Sigma_{1,1} = \frac{1}{n}, \quad \Sigma_{1,j} = \Sigma_{j,1} = \frac{(j-1)^{\alpha+1} - (j-2)^{\alpha+1}}{(\alpha+1)n^{\alpha+1}}, \\ \Sigma_{j,j} = \frac{(j-1)^{2\alpha+1} - (j-2)^{2\alpha+1}}{(2\alpha+1)n^{2\alpha+1}}, \quad \text{for} \quad j = 2, \ldots, \kappa+1\end{aligned} \tag{3.2}$$

$$\begin{aligned}\Sigma_{j,k} = &\frac{1}{(\alpha+1)n^{2\alpha+1}}\left[(j-1)^{\alpha+1}(k-1)^\alpha \, {}_2F_1\left(-\alpha, 1, \alpha+1, \frac{j-1}{k-1}\right) - \right. \\ &\left. (j-2)^{\alpha+1}(k-2)^\alpha \, {}_2F_1\left(-\alpha, 1, \alpha+2, \frac{j-2}{k-2}\right)\right], \quad \text{for} \quad 1 < j < k \leq \kappa+1,\end{aligned} \tag{3.3}$$

where ${}_2F_1$ is the hypergeometric function, defined as

$$ {}_2F_1(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} $$

with

$$ (a)_n \coloneqq \begin{cases} 1, & n = 0, \\ a(a+1)\dots(a+n-1), & n > 0. \end{cases} $$

Another computational note is that we can express both terms in (3.1) as a discrete convolution, which can be efficiently calculated using the fast Fourier transform. This is less relevant for the first term, since κ is usually small ($\kappa \leq 3$). However, this is relevant for the second term in (3.1), which then becomes

$$ \sum_{k=\kappa+1}^{N_n} g\left(\frac{b_k^*}{n}\right) W_{i-k,0} = \sum_{k=1}^{N_n} \Gamma_k \Xi_{i-k} = (\Gamma * \Xi)_i, $$

where

$$ \Gamma_k \coloneqq \begin{cases} 0, & k = 1, \dots, \kappa \\ g\left(\frac{b_k^*}{n}\right), & k = \kappa+1, \dots, N_n \end{cases} $$
$$ \Xi_k \coloneqq W_k, \quad k = -N_n, \dots, nT-1. $$

3.2. Romano Touzi Trick

The goal of this trick, as introduced in [17], is to give a better estimation of the true value of a derivative when using a Monte Carlo simulation. The main idea is to, instead of assigning a single value to the path at maturity T, to assign a probability distribution to the final value at time T. This is done by taking into consideration the realized volatility of the path, thus utilizing the full path instead of only the value at maturity. The method is explained in the following proposition

Proposition 3.1. *Suppose we have a stochastic volatility process under the martingale measure* $\mathbb{Q}$

$$ \begin{cases} \frac{dF_t}{F_t} = \sigma(Y_t)\sqrt{1-\rho_t} dW_t^1 + \sigma(Y_t)\rho_t dW_t^2 \\ dY_t = \beta_t dt + \gamma_t dW_t^2. \end{cases} \tag{3.4} $$

Then the associated admissible arbitrage price of the European call option is given by

$$ V_t^{\mathsf{Call}} = \mathbb{E}_t^{\mathbb{Q}} \left[V_t^{\mathsf{Call,BS}} \left(F_t e^{Z_{t,T}}; \frac{1}{T-t} \int_t^T (1-\rho_u^2)\sigma_u^2 du \right) \right], $$

where $V_t^{\mathsf{Call,BS}}(F_t, \sigma_t^2)$ *is the Black-Scholes option price for a European call given the forward stock value* F_t *and (normalized) variance* σ_t^2 *and* $Z_{t,T} = \int_t^T \rho_v \sigma_u dW_u^2 - \frac{1}{2}\int_t^T \rho_u^2 \sigma_u^2 du$.

Remark 3.2. *Romano and Touzi give a more general framework for this in [17], but this formulation already takes into consideration some extra assumptions. Also note that since we do not work with a measure change, we assume that this is all under the same risk-neutral measure.*

Proof. Integrating (3.4) gives

$$ F_T = F_t e^{Z_{t,T}} \exp\left(-\frac{1}{2}\int_t^T \sigma_u^2 (1-\rho_u)^2 du + \int_t^T \sigma_u \sqrt{1-\rho_u^2} dW_u^1 \right). $$

Now realize that $Z_{t,T}$ is independent of W_t^1, so we can say that

$$ \log\left(\frac{F_T}{F_t e^{Z_{t,T}}}\right) \xrightarrow{\mathbb{Q}} \mathcal{N}\left(-\frac{1}{2}V_{t,T}^2, V_{t,T}^2\right), $$

where $V_{t,T} = \sqrt{\int_t^T (1-\rho_u^2)\sigma_u^2 du}$. Thus calculating the price for a European call option can be done using the Black-Scholes formula where the underlying asset price is replaced by $S_t e^{Z_{t,T}}$ and the volatility parameter is replaced by $V_{t,T}/\sqrt{T-t}$. □

Applying this to rough Bergomi means splitting the process into two parts with independent Brownian motions $\tilde{B},\ \tilde{W}$, obtaining

$$F_t = F_0 \mathcal{E}\left(\rho \int_0^t \sqrt{V_u} d\tilde{B}_u\right) \mathcal{E}\left(\sqrt{1-\rho^2} \int_0^t \sqrt{V_u} d\tilde{W}_u\right).$$

Thus applying the Romano Touzi trick results in

$$\log F_T \sim \mathcal{N}\left(\log \mathcal{E}\left(\rho \int_0^t \sqrt{V_u} d\tilde{B}_u\right) - \frac{1}{2}(1-\rho^2)\int_0^t V_u du\,,\, (1-\rho^2)\int_0^t V_u du\right).$$

Hence the final call price is given by

$$V_t^{\mathsf{Call,BS}}\left(\mathcal{E}\left(\rho \int_0^t \sqrt{V_u} d\tilde{B}_u\right), (1-\rho^2)\int_0^t V_u du\right).$$

4. Characteristic Function Pricing

4.1. COS Method

One of the recent developments of option pricing through the characteristic function has been to price European options using the COS method, introduced by Fang and Oosterlee [18]. This method utilizes an approximation of the inverse Fourier transform, in order to price options. In this method we try to approximate the density function $f(x)$ when given a characteristic function $\phi(\omega)$, using their relationship via the Fourier transform

$$\phi(\omega) = \int_{\mathbb{R}} e^{ix\omega} f(x)dx, \quad f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega x}\phi(\omega)d\omega. \tag{4.1}$$

The idea is to approximate this integral using the cosine expansion, given by

$$f(\theta) = {\sum_{k=0}^{\infty}}' A_k \cos(k\theta) \quad \text{with} \quad A_k = \frac{2}{\pi}\int_0^{\pi} f(\theta)\cos(k\theta)d\theta, \tag{4.2}$$

where ${\sum_{k=0}^{\infty}}'$ denotes the sum where the first term is multiplied by $\frac{1}{2}$. The above cosine expansion holds for functions supported on $[0,\pi]$, this can be changed to an arbitrary interval $[a,b]$ using a change of variables. This means we only take $\omega \in [a,b]$ in (4.1) and truncate the integral. Now our coefficients A_k become

$$A_k \approx \frac{2}{b-a}\Re\left\{\phi\left(\frac{k\pi}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}. \tag{4.3}$$

To determine the truncation range $[a,b]$ we use the method as suggested in [18] and take

$$[a,b] \coloneqq \left[c_1 - L\sqrt{c_2+\sqrt{c_4}}, c_1 + L\sqrt{c_2+\sqrt{c_4}}\right] \quad \text{with} \quad L \approx 10, \tag{4.4}$$

where c_i refers to cumulant i. How we determine these cumulants is explained in section 5.2. The value of N is set at 500, as this seems to be enough to get a good convergence of the characteristic function. In order to calculate $f(\theta)$ we need to truncate the sum from ${\sum_{k=0}^{\infty}}'$ to ${\sum_{k=0}^{N-1}}'$. This creates a second error in the method.

Eventually we would like to price a product at time t with pay-off $V(S_T,T)$, where S_T is the price of the underlying at T, which is the maturity. Following the computations in [18] we find that we can do this in the following way

$$e^{r(T-t)}V(S_t,t) \approx \frac{b-a}{2}{\sum_{k=0}^{N-1}}'\Re\left\{\phi\left(\frac{k\pi}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}V_k, \tag{4.5}$$

where

$$V_k \coloneqq \frac{2}{b-a}\int_a^b V(y,T)\cos\left(k\pi\frac{y-a}{b-a}\right)dy.$$

We note that the COS method makes use of the fact that density functions tend to be smooth, thus only few terms (N) will already give a good approximation. Also, the density function decays rapidly to zero as $y \to \pm\infty$, so we can also truncate $[a,b]$ at reasonable values without losing a lot of accuracy. Finally we note that the evaluation of the characteristic function is independent of the pay-off structure. In practice this means that we can use the same evaluations of the characteristic function for multiple option strikes.

Finally we calculate V_k for a put option. For this we will use the method introduced by Le Floc'h [19] and use the put-call parity to obtain call prices. Le Floc'h observed that for short maturities the original COS method was unsuitable to approximate far in-the-money call prices because the cosine coefficients for the put option were computed relatively to the strike K, but the truncation range was relative to the spot price (namely dependent on

the cumulants). In order to solve this he used cosine coefficients which were relative to the forward F. The coefficients then become

$$V_k^{\text{Put}} = \frac{2}{b-a}\left(-F\chi_k(a,z) + K\psi_k(a,z)\right), \tag{4.6}$$

where $z = \log\frac{K}{F}$, $y = \log\frac{S_T}{F}$

$$\chi_k(c,d) \coloneqq \frac{1}{1+\left(\frac{k\pi}{b-a}\right)^2}\Bigg[\cos\left(k\pi\frac{d-a}{b-a}\right)e^d - \cos\left(k\pi\frac{c-a}{b-a}\right)e^c + \frac{k\pi}{b-a}\sin\left(k\pi\frac{d-a}{b-a}\right)e^d - \frac{k\pi}{b-a}\sin\left(k\pi\frac{c-a}{b-a}\right)e^c\Bigg] \tag{4.7}$$

and

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

Do note that the functions χ and ψ need to now be evaluated for each strike (which wasn't the case in the original derivation in [18]).

Details of the derivations of (4.3), (4.5) and (4.6) are given in Appendix A

4.2. Affine Volterra Processes

Now that we have a pricing method through characteristic functions, we need to determine what the characteristic functions of the rough Heston and lifted Heston model are. This is done by utilizing the exponential-affine transform formula stated in Theorem 4.3 of [20]. An affine Volterra process is defined as

Definition 4.1. *Let $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$, $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, $b : \mathbb{R}^d \to \mathbb{R}^d$, $a = \sigma\sigma^\top$ and X_0 be deterministic. An affine Volterra process with state space E is a continuous E-valued solution X_t of*

$$X_t = X_0 + \int_0^t K(t-s)b(X_s)ds + \int_0^t K(t-s)\sigma(X_s)dW_s, \tag{4.8}$$

where

$$a(x) = A^0 + x_1A^1 + \ldots + x_dA^d$$
$$b(x) = b^0 + x_1b^1 + \ldots + x_db^d$$

with $A^i \in \mathbb{R}^{d\times d}$ and $b^i \in \mathbb{R}^d$.

The exponential-affine transform formula is then defined in the following theorem, see [20]

Theorem 4.1. *Let X be an affine Volterra process and fix some $T < \infty$, $u \in (\mathbb{C}^d)^*$ and let $f \in L^1([0,T],(\mathbb{C}^d)^*)$. Assume $\psi \in L^2([0,T],(\mathbb{C}^d)^*)$ solves the Riccati-Volterra equation*

$$\psi = uK + \left(f + \psi B + \frac{1}{2}A(\psi)\right) * K,$$

where $*$ denotes the convolution, $B = (b^1 \ldots b^d)$ and $A(\psi) = (\psi A^1\psi^\top, \ldots, \psi A^d\psi^\top)$. Then the process $\{Y_t, 0 \le t \le T\}$ defined by

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \psi(T-s)\sigma(X_s)dW_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^\top ds \\ Y_0 &= uX_0 + \int_0^T \left(f(s)X_0 + \psi(s)b(X_0) + \frac{1}{2}\psi(s)a(X_0)\psi(s)^\top\right)ds \end{aligned} \tag{4.9}$$

satisfies

$$Y_t = \mathbb{E}[uX_T + (f * X)_T|\mathcal{F}_t] + \frac{1}{2}\int_t^T \psi(T-s)a(\mathbb{E}[X_s|\mathcal{F}_t])\psi(T-s)^\top ds$$

for all $0 \le t \le T$. *The process* $\{\exp\{Y_t\}, 0 \le t \le T\}$ *is a local martingale and, if it is a true martingale, one has the exponential-affine transform formula*

$$\mathbb{E}\left[\exp\{uX_T + (f * X)_T\}|\,\mathcal{F}_t\right] = \exp\{Y_t\}, \quad t \le T.$$

We now set the following assumptions on kernel K and let $\Delta_h K(t) = K(t+h)$

$$\begin{gathered} K \in L^2_{\mathsf{loc}}(\mathbb{R}_+, \mathbb{R}) \text{ and there is } \gamma \in (0,2] \text{ such that } \int_0^h K(t)^2 dt = O(h^\gamma) \\ \text{and } \int_0^T (K(t+h) - K(t))^2 dt = O(h^\gamma) \text{ for every } T < \infty. \end{gathered} \tag{4.10}$$

$$\begin{gathered} \Delta_h K(.) \text{ is nonnegative, not identically zero, non-increasing and continuous on} \\ (0,\infty), \text{ and its resolvent of the first kind } L_{\Delta_h} \text{ is nonnegative and} \\ \text{non-increasing in that } s \mapsto L_{\Delta_h}([s, s+t]) \text{ is non-increasing for all } t \ge 0. \end{gathered} \tag{4.11}$$

Now considering the log-price process

$$\begin{cases} \log S_t = \log S_0 - \int_0^t \frac{V_s}{2} ds + \int_0^t \sqrt{V_s} dB_s \\ V_t = V_0 + \int_0^t K(t-s)\left(\lambda(\theta - V_s) + \nu\sqrt{V_s} dW_s\right). \end{cases}$$

This is an affine Volterra process with

$$\begin{aligned} A^0 = A^1 = 0, \quad A^2 &= \begin{pmatrix} 1 & \rho\nu \\ \rho\nu & \nu^2 \end{pmatrix} \\ b^0 = \begin{pmatrix} 0 \\ \lambda\theta \end{pmatrix}, \quad (b^1\ b^2) &= \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & -\lambda \end{pmatrix} \end{aligned}$$

Then we know the following theorem holds by Theorem 7.2(ii) in [20].

Theorem 4.2. *Assume* K *satisfies* (4.10) *and* (4.11) *for all* $h \in [0,1]$. *Let* $u \in (\mathbb{C}^2)^*$ *and* $f \in L^1_{loc}(\mathbb{R}_+, (\mathbb{C}^2)^*)$ *be such that*

$$\Re\psi_1 \in [0,1],\ \Re u_2 \le 0,\ \textit{and}\ , \Re f_2 \le 0,$$

where $\psi_1 = u_1 + \mathbf{1} * f_1$. *Then the process* $X_t = (\log S_t, V_t)$ *has a unique continuous weak solution for any initial condition* $(\log S_0, V_0) \in \mathbb{R} \times \mathbb{R}_+$. *Moreover the Riccati-Volterra equation*

$$\psi_2 = u_2 K + K * F(\psi_1, \psi_2)$$

has a unique global solution $\psi_2 \in L^2_{loc}(\mathbb{R}_+, \mathbb{C}^*)$, *which satisfies* $\Re\psi_2 \le 0$. *Moreover, the exponential-affine transform formula holds, given by*

$$\mathbb{E}\left[\exp(uX_T + (f * X)_T\right] = \exp\left\{\psi_1(T)\log S_0 + u_2 g_0(T) + \int_0^T F(\psi_1, \psi_2)(s) g_0(T-s) ds\right\}, \tag{4.12}$$

with

$$F(\psi_1, \psi_2) = f_2 + \frac{1}{2}(\psi_1^2 - \psi_1) + (\rho\nu\psi_1 - \lambda)\psi_2 + \frac{\nu^2}{2}\psi_2^2.$$

4.3. Characteristic Function of Rough Heston

The characteristic function of the rough Heston model follows straight from Theorem 4.2 as done in [11], we now need to make sure that the kernel $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ satisfies assumptions (4.10), (4.11).
Verifying the first assumption is done by

$$\begin{aligned}\int_0^h (K(t))^2 dt &= \int_0^t \frac{t^{2\alpha-2}}{\Gamma(\alpha)^2} dt \\ &= \frac{1}{\Gamma(\alpha)^2}\int_0^h t^{2\alpha-2} dt \\ &= \frac{1}{\Gamma(\alpha)^2}\frac{1}{2\alpha-1} h^{2\alpha-2} = \mathcal{O}(h^{2\alpha-1}).\end{aligned}$$

Noting that $2\alpha - 1 \in (0, 2]$ thus verifies the first part. The second part is verified by utilizing the Taylor expansion of $K(t+h) = t^{\alpha-1} + \frac{1}{\alpha-1}t^{\alpha-2}h + \mathcal{O}(h^2)$. Thus

$$\begin{aligned}\int_0^T (K(t+h) - K(t))^2 &= \frac{1}{\Gamma(\alpha)}\int_0^T ((t+h)^{\alpha-1} - t^{\alpha-1})^2 dt \\ &= \frac{1}{\Gamma(\alpha)}\int_0^T \frac{1}{(\alpha-1)^2} t^{2\alpha-2}h^2 + \mathcal{O}(h^4) dt = \mathcal{O}(h^2),\end{aligned}$$

verifying (4.10).
Secondly we show (4.11), this is due to an argument based on the monotonicity of $\Delta_h K$. Theorem 5.5.4 of [21] states that if a kernel is monotone, then it has a resolvent of the first kind, f, which is completely monotone, this means that $(-1)^k f^{(k)}(t) \geq 0, \forall t > 0, k \in \mathbb{N}$. Knowing that $\Delta_h K$ is monotone and positive concludes the reasoning.
Thus by Theorem 4.2 and taking $u_1 = ia$, $u_2 = 0$ and $f_1 = f_2 = 0$ we obtain

$$\mathbb{E}\left[\exp\{iaS_t\}\right] = \exp\left\{ia\log S_0 + \theta\lambda I^1 h(a,t) + V_0 I^{1-\alpha}h(a,t)\right\},$$

where h is the solution to the fractional Riccati equation

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

where D^α and $I^{1-\alpha}$ are defined in (2.2).

4.3.1. Numerical Approximation

The characteristic function is entirely defined through the fractional Riccati equation

$$\begin{cases} D^\alpha h(a,t) = F(a, h(a,t)), \quad I^{1-\alpha}h(a,0) = 0 \\ F(a,x) = \frac{1}{2}(-a^2 - ia) + \lambda(ia\rho\nu - 1)x + \frac{(\lambda\nu)^2}{2}x^2. \end{cases}$$

Applying the fractional integral operator I^α to both sides gives

$$h(a,t) = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1} F(a, h(a,s)) ds. \tag{4.13}$$

Numerically solving this equation is done using the Adams method, as explained in [22]. For this, define a time grid $(t_k)_{k\in\mathbb{N}}$ with mesh Δ, so $t_k = \Delta k$. Then estimate the integral in (4.13) by a trapezoidal approximation of F,

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

with

$$F(a,\hat{h}(a,t)) = \frac{t_{j+1} - t}{t_{j+1} - t_j} F(a, \hat{h}(a,t_j)) + \frac{t - t_j}{t_{j+1} - t_j} F(a, \hat{h}(a, t_{j+1})),\ t \in [t_j, t_{j+1}),\ 0 \leq j \leq k.$$

Calculating the integral results in

$$\hat{h}(a,t_{k+1}) = \sum_{0\le j\le k} a_{j,k+1}F(a,\hat{h}(a,t_j)) + a_{k+1,k+1}F(a,\hat{h}(a,t_{k+1})), \tag{4.14}$$

with

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

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

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

Now note that $\hat{h}(a,t_{k+1})$ is on both sides of (4.14), making this scheme implicit. Thus we use a pre-estimation of $\hat{h}(a,t_{k+1})$, called a predictor and denote it by $\hat{h}^P(a,t_{k+1})$. This predictor is calculated using a left-hand side integration of (4.13):

$$\hat{h}^P(a,t_{k+1}) = \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}F(a,\hat{h}(a,t_j))ds.$$

Therefore

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

where

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

So the final approximation is given by

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

4.4. Characteristic Function of Lifted Heston

For the lifted Heston model we choose $K(t) = \sum_i^n c_i e^{-x_i t}$ for c_i, x_i, where these are chosen to approximate the fractional kernel used in the rough Heston model as explained in subsection 4.4.2. Note that $K(t)$ is the Laplace transform of $\mu(dx) = \sum_i^n c_i\delta_{x_i}(dx)$.
Also set the initial instantaneous forward variance curve g_0 defined in (2.15) equal to

$$g_0(t) = V_0 + \lambda\theta\sum_{i=1}^n c_i\int_0^t e^{-x_i(t-s)}ds = V_0 + \lambda\theta\int_0^t e^{-x(t-s)}\mu(dx).$$

Now following the reasoning in section 4 from [23], consider the process

$$dU_t(x) = \left(-xU_t(x) - \lambda\left(g_0(t) + \int_0^\infty U_t(z)\mu(dz)\right)\right)dt + \nu\sqrt{g_0(t) + \int_0^\infty U_t(z)\mu(dz)}dW_t.$$

By Theorem 4.2, U_t has a unique continuous weak solution. Given the initial condition $U_0 = 0$ this solution is given by

$$U_t = \int_0^t e^{-x(t-s)}\left(-\lambda V_s ds + \nu\sqrt{V_s}dW_s\right)$$

where

$$V_t = g_0(t) + \int_0^\infty U_t(z)\mu(dz).$$

In order to apply Theorem 4.2, the conditions on K must be checked. (4.10) holds with $\gamma = 1$, since

$$\int_0^h \left(\sum_i c_i e^{-x_i t}\right)^2 dt = \sum_i \sum_j c_i c_j \int_0^h e^{-(x_i+x_j)t} dt$$
$$= \sum_i \sum_j \frac{c_i c_j}{x_i + x_j}[e^{-(x_i+x_j)h} - 1] = \mathcal{O}(h)$$

and

$$\int_0^h \sum_i c_i e^{-x_i(t+h)} - \sum_i c_i e^{-x_i t} dt = \sum_i c_i (e^{-x_i h} - 1) \int_0^T e^{-x_i t} dt = \mathcal{O}(h),$$

both following from the Taylor expansion of e^h. Secondly we show (4.11), which is due to the same argument as before. Again noting Theorem 5.5.4 of [21], which states that if a kernel is monotone, then it has a resolvent of the first kind which is completely monotone. Knowing that $\Delta_h K$ is monotone and positive concludes the reasoning.
We can now use (4.12) and take $u_1 = ia$, $u_2 = 0$ and $f_1 = f_2 = 0$. Thus we can now conclude that the characteristic function of the lifted Heston model is given by

$$\mathbb{E}\left[ia \log S_T\right] = \exp\left\{\log S_0 + \int_0^T F\left(u, \sum_i^n c_i \psi^i(s)\right) g_0(T-s) ds\right\} \tag{4.15}$$

where

$$\partial_t \psi^i = -x_i \psi^i + F\left(u, \sum_i^n c_i \psi^i(s)\right), \ \psi^i(0) = 0$$

and

$$F(u,v) = \frac{1}{2}(u^2 - u) + (\rho\nu u - \lambda)v + \frac{\nu^2}{2} v^2.$$

4.4.1. Numerical Approximation

In order to solve the characteristic function for the lifted Heston model we thus need to solve the following system of equations

$$\begin{cases} \partial_t \psi^i = -x_i \psi^i + F\left(u, \sum_i^n c_i \psi^i(s)\right), \ \psi^i(0) = 0, \ i = 1, \dots, n \\ F(u,v) = \frac{1}{2}(u^2 - u) + (\rho\nu u - \lambda)v + \frac{\nu^2}{2} v^2. \end{cases} \tag{4.16}$$

Since there is no closed form solution, we use a numerical scheme to approximate this n-dimensional Riccati equation. We will first look at the most simple approximation scheme, as described in [12], called the explicit-implicit Euler scheme. Define a time grid $(t_k)_{k \in \mathbb{N}}$ with mesh size Δt. Approximate the first equation of (4.16) by

$$\partial_t \psi^i = -x_i \psi^i, \quad i = 1, \dots, n, \tag{4.17}$$

for which the solution is given by

$$\psi^i(t) = \psi^i(0) e^{-x_i t}, \quad i = 1, \dots, n.$$

An explicit Euler scheme would take $\hat{\psi}^i(t_{k+1}) = (1 - x_i \Delta t)\hat{\psi}^i(t_k)$. The stability condition of this scheme would be $\Delta t \leq \frac{1}{x_i}, \forall i = 1, \dots, n$. Since we have $x_{20} \approx 6418$, this would require almost $13,000$ time steps for an expiry two years out. In order to get around this we consider the implicit Euler method

$$\hat{\psi}^i(t_{k+1}) = \hat{\psi}^i(t_k) - x_i \Delta t \hat{\psi}^i(t_{k+1}) \implies \hat{\psi}^i(t_{k+1}) = \frac{1}{1 + x_i \Delta t} \psi^i(t_k).$$

Thus the full scheme regarding the numerical approximation becomes

$$\begin{cases} \hat{\psi}^i(t_{k+1}) = \frac{1}{1+x_i\Delta t}\left(\hat{\psi}^i(t_k) + \Delta t F\left(u, \sum_{j=1}^n c_j \psi^j(t_k)\right)\right), \quad i = 1, \dots, n \\ \hat{\psi}^i(0) = 0. \end{cases} \tag{4.18}$$

Alternations to this method are replacing $\frac{1}{1+x_i\Delta t}$ by $e^{-x_i\Delta t}$, however this could result in numerical errors. One could also consider adding the terms constant and linear in ψ^i to (4.17)

$$\partial_t \psi^i = -x_i\psi^i + \frac{1}{2}(u^2 - u) + (\rho\nu u - \lambda)\sum_{j\neq i} c_j\psi^j + (\rho\nu u - \lambda)c_i\psi_i.$$

This differential equation has a closed form solution given by

$$\psi^i(t) = \psi^i(0)e^{a_i t} + \frac{b_i}{a_i}(1 - e^{-a_i t})$$

with

$$a_i = \frac{1}{2}(u^2 - u) + (\rho\nu u - \lambda)\sum_{j\neq i} c_j\psi^j$$
$$b_i = -x_i + (\rho\nu u - \lambda)c_i.$$

Finally, one could consider implementing higher order Runge-Kutta methods for solving (4.18), as suggested in [12].

4.4.2. Approximating the Rough Kernel

Finally we must define our $c_i, x_i,\ i = 1, \dots, n$. We want to choose x_i, c_i such that $K(t) = \sum_i^n c_i e^{-x_i t}$ best describes the fractional kernel $K^\alpha = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$. We now choose a measure $\mu(dx)$ such that the Laplace transform of this measure gives us K^α, this is done by taking

$$K^\alpha(t) = \int_0^\infty e^{-xt}\mu(dx), \quad \mu(dx) = \frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}dx.$$

We now approximate $\mu(t) \approx \sum_{i=1}^n c_i\delta_{x_i}(t)$ using methods from [24] section 3.1. They describe how the function $W(t) = \int_0^\infty U(x,t)f(x)dx$ can be approximated by $\hat{W}(t) = \sum_i c_i U(\eta_i, t)$ by choosing

$$c_i = \int_{\xi_i}^{\xi_{i+1}} f(x)dx, \quad \eta_i = \frac{\int_{\xi_i}^{\xi_{i+1}} xf(x)dx}{\int_{\xi_i}^{\xi_{i+1}} f(x)dx}.$$

Taking $f(x) = \frac{x^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)}$ and taking the geometric partition $\eta_i = r_n^{i-n/2}$ for some $r_n > 0$ results in

$$c_i = \frac{(r_n^{1-\alpha} - 1)r_n^{(\alpha-1)(1+n/2)}}{\Gamma(\alpha)\Gamma(2-\alpha)} r_n^{(1-\alpha)i},\ x_i = \frac{1-\alpha}{2-\alpha}\frac{r_n^{2-\alpha} - 1}{r_n^{1-\alpha} - 1} r_n^{i-1-n/2},\ i = 1, \dots, n.$$

In [12] they show that if $(r_n)_{n\geq 1}$ satisfies

$$r_n \downarrow 1, \quad \text{and} \quad \lim_{n\to\infty} n \log r_n = \infty,$$

then the lifted Heston model converges towards the rough Heston model as n goes to infinity. They also compare the lifted Heston model for different values of n and r_n and suggest taking $n = 20$ and $r_n = 2.5$ for a good approximation of the lifted Heston vol curve to that of the rough Heston. Thus we also take these values in our implementation.

5. Forward Variance Curve

An input to the rough Bergomi and rough Heston model is the forward variance curve. This chapter explains how to approximate the forward variance curve and the cumulants, as needed in (4.4), from market data. This is done using the Breeden-Litzenberger formula.

5.1. Volatility Derivatives

Let us consider a European call option with strike K and maturity T, then the price at current time t ($V^{\mathsf{Call}}(S_t, K)$) can be represented as

$$V^{\mathsf{Call}}(S_t, K) = e^{-rT} \int_K^\infty (x - K)\phi_{S_T}(x)dx,$$

where ϕ_{S_T} is the probability density function of the price of the underlying at maturity (S_T). Now differentiating $V^{\mathsf{Call}}(S_t, K)$ twice to K gives

$$\begin{aligned}
\frac{\partial^2}{\partial K^2} V^{\mathsf{Call}}(S_t, K) &= e^{-rT} \frac{\partial^2}{\partial K^2} \int_K^\infty (x - K)\phi_{S_T}(x)dx \\
&= e^{-rT} \frac{\partial}{\partial K} \left\{ -(K - K)\phi_{S_T}(K) + \int_K^\infty \frac{\partial}{\partial K}(x - K)\phi_{S_T}(x)dx \right\} \\
&= e^{-rT} \frac{\partial}{\partial K} \left\{ -\int_K^\infty \phi_{S_T}(x)dx \right\} \\
&= e^{-rT}\phi_{S_T}(K),
\end{aligned}$$

where we first apply the Leibniz integral rule and then the fundamental theorem of calculus. The same holds for a European put ($V^{\mathsf{Put}}(S_t, K)$), where we find that

$$\frac{\partial^2}{\partial K^2} V^{\mathsf{Put}}(S_t, K) = e^{-rT}\phi_{S_T}(K)$$

For notation sake we now introduce $\tilde{V}^{\mathsf{Call}}$ and $\tilde{V}^{\mathsf{Put}}$ as the undiscounted call and put prices respectively and denote the future of the underlying by F and consider them all at time $t < T$. The put-call parity now reads $\tilde{V}^{\mathsf{Call}}(K) - \tilde{V}^{\mathsf{Put}}(K) = F - K$. Using this we now consider a general pay-off $g(S_T)$, as done in chapter 11 of [25], this is given by

$$\begin{aligned}
\mathbb{E}\left[g(S_T)|S_t\right] &= \int_0^\infty g(K)\phi_{S_T}(K)dK \\
&= \int_0^F \frac{\partial^2 \tilde{V}^{\mathsf{Put}}}{\partial K^2} g(K)dK + \int_F^\infty \frac{\partial^2 \tilde{V}^{\mathsf{Call}}}{\partial K^2} g(K)dK \\
&= \frac{\partial \tilde{V}^{\mathsf{Put}}}{\partial K} g(K)\Big|_{K=0}^F - \int_0^F \frac{\partial \tilde{V}^{\mathsf{Put}}}{\partial K} g'(K)dK \\
&\quad + \frac{\partial \tilde{V}^{\mathsf{Call}}}{\partial K} g(K)\Big|_{K=F}^\infty - \int_F^\infty \frac{\partial \tilde{V}^{\mathsf{Call}}}{\partial K} g'(K)dK \\
&= \frac{\partial K - F}{\partial K}(F)g(F) - \int_0^F \frac{\partial \tilde{V}^{\mathsf{Put}}}{\partial K} g'(K)dK - \int_F^\infty \frac{\partial \tilde{V}^{\mathsf{Call}}}{\partial K} g'(K)dK \\
&= g(F) - \tilde{V}^{\mathsf{Put}}(K)g'(K)\Big|_{K=0}^F + \int_0^F \tilde{V}^{\mathsf{Put}}(K)g''(K)dK \\
&\quad - \tilde{V}^{\mathsf{Call}}(K)g'(K)\Big|_{K=F}^\infty + \int_F^\infty \tilde{V}^{\mathsf{Call}}(K)g''(K)dK \\
&= g(F) + \int_0^F \tilde{V}^{\mathsf{Put}}(K)g''(K)dK + \int_F^\infty \tilde{V}^{\mathsf{Call}}(K)g''(K)dK.
\end{aligned} \tag{5.1}$$

Thus for a general pay-off $g(S_T)$ which is twice differentiable we can determine a fair value by integrating over weighted out-of-the-money call and put prices. This is often referred to as the Breeden-Litzenberger formula.

5.2. Cumulants

The cumulants of $\log(S_t/F)$ are needed for two reasons. Firstly we want to use this to determine a truncation range for the COS method. In [18], equation (49) determines the truncation range as in (4.4). Secondly the cumulants are needed to determine the forward variance curve, which is described in section 5.3.
In order to determine the cumulants we take the function $g(S_t) = \log^n(S_T/F)$. First we note that for $n \geq 2$

$$\frac{\partial^2}{\partial K^2}\log^n\left(\frac{K}{F}\right) = \frac{\partial}{\partial K}\frac{n}{K}\log^{n-1}\left(\frac{K}{F}\right) = \frac{n}{K^2}\left[(n-1)\log^{n-2}\left(\frac{K}{F}\right) - \log^{n-1}\left(\frac{K}{F}\right)\right].$$

Now using equation (5.1) one can verify

$$\begin{aligned}\mathbb{E}\left[\log^n\left(\frac{S_T}{F}\right)\middle| S_t\right] = &\int_0^F \tilde{V}^{\mathsf{Put}}(K)\frac{n}{K^2}\left[(n-1)\log^{n-2}\left(\frac{K}{F}\right) - \log^{n-1}\left(\frac{K}{F}\right)\right]dK \\ &+ \int_F^\infty \tilde{V}^{\mathsf{Call}}(K)\frac{n}{K^2}\left[(n-1)\log^{n-2}\left(\frac{K}{F}\right) - \log^{n-1}\left(\frac{K}{F}\right)\right]dK.\end{aligned}$$

In the case of $n = 1$

$$\mathbb{E}\left[\log\left(\frac{S_T}{F}\right)\middle| S_t\right] = -\int_0^F \frac{\tilde{V}^{\mathsf{Put}}(K)}{K^2}dK - \int_F^\infty \frac{\tilde{V}^{\mathsf{Call}}(K)}{K^2}dK.$$

Usually in the COS method the cumulants are known from the model, however for the rough volatility models this is not the case. Hence we infer the cumulants from the market curve by taking a weighted integral over the out-of-the-money puts and calls.

5.3. Forward Variance Curve

The Breeden-Litzenberger formula can also be used to determine the discrete instantaneous forward variance curve, as done in chapter 11 of [25]. Let us assume that there are no interest rates and dividends, so $F = S_0$, then by Itô's lemma and working under the risk-neutral measure we find that

$$\mathbb{E}\left[\log\left(\frac{S_T}{S_0}\right)\right] = \mathbb{E}\left[\int_0^T d\log(S_t)\right] = \mathbb{E}\left[\int_0^T \frac{dS_t}{S_t} - \int_0^T \frac{\sigma_{S_t}^2}{2}dt\right] = -\mathbb{E}\left[\int_0^T \frac{\sigma_{S_t}^2}{2}dt\right].$$

Note that the last step holds under the risk-neutral measure. Hence the total variance of S_t over $[0,T]$ as follows

$$\mathbb{E}\left[\int_0^T \sigma_{S_t}^2 dt\right] = -2\mathbb{E}\left[\log\left(\frac{S_T}{S_0}\right)\right] = 2\left\{\int_0^F \frac{\tilde{V}^{\mathsf{Put}}(S_T,K)}{K^2}dK + \int_F^\infty \frac{\tilde{V}^{\mathsf{Call}}(S_T,K)}{K^2}dK\right\}, \tag{5.2}$$

Now the instantaneous forward variance (ξ_{T_1,T_2}) is defined as the annualized variance between times T_1 and T_2. Thus the total variance in terms of ξ_{T_{i-1},T_i} is given by

$$\mathbb{E}\left[\int_0^T \sigma_{S_t}^2 dt\right] = \lim_{N_T\to\infty}\sum_{i=1}^{N_T}\xi_{T_{i-1},T_i}(T_{i-1}-T_i) \approx \sum_{i=1}^{N_{\mathsf{exp}}}\xi_{T_{i-1},T_i}(T_{i-1}-T_i), \tag{5.3}$$

where N_{exp} is the number of expiries. The sum is taken to N_{exp} because a vol curve is needed from the market to determine ξ_{T_{i-1},T_i}. We can then approximate our forward variance by

combining equations (5.2) and (5.3) as follows

$$\begin{aligned}\xi_{T_{i-1},T_i} &= \frac{1}{T_{i-1}-T_i}\mathbb{E}\left[\int_{T_{i-1}}^{T_i}\sigma_{S_t}^2 dt\right] \\ &= \frac{2}{T_{i-1}-T_i}\left\{\int_0^F \frac{\tilde{V}^{\mathsf{Put}}(S_{T_i},K)-\tilde{V}^{\mathsf{Put}}(S_{T_{i-1}},K)}{K^2}dK + \right. \\ &\qquad \left. \int_F^\infty \frac{\tilde{V}^{\mathsf{Call}}(S_{T_i},K)-\tilde{V}^{\mathsf{Call}}(S_{T_{i-1}}K)}{K^2}dK\right\}.\end{aligned}$$

This then creates a step function for the forward variance curve as implied by the market. Note that this is the same as the discrete forward variance as defined in chapter 5 of [26].

6. Fitting

The goal of this research is to fit our models to the market data, there are many ways to approach such a problem. This chapter explains the fitting methods used and the reasons why. Fitting is done using a weighted root-mean-square error (RMSE) and is done in two stages. The first stage is a global search using differential evolution, the second stage then uses this result and finds a local minimum with a Nelder-Mead algorithm.

6.1. Weights

In order to fit the model vol surface to the market vol surface we add weights to different parts of the vol surface depending on how well the surface approximates the actual implied volatility. We add weights in two different ways, one according to the vega and another according to the market spread.

Multiplying our weights by vega would mean that we add a term $\partial V/\partial\sigma$ to the weights, where V is the price of an option. This essentially translates the objective from a minimization in vol space to a minimization in price space. On the other hand if we do not multiply by vega we only minimize in vol space. As a mid-way we therefore multiply our weights by the square-root of vega, thus we take into consideration the minimization in vol space and in price space equally.

Inspired by the solution in chapter 13.2 of [27], we also weigh the point on a vol surface according to the market spread. This way we take into consideration where along the vol curve the market is most sure about the vol. We do this by weighing the terms in the MSE with

$$w_{i,spread} = \frac{1}{V_i^{\text{Call/Put,ask}}(K) - V_i^{\text{Call/Put,bid}}(K)}.$$

Here $V^{\text{Call/Put}}$ refers to the out-of-the-money call or put price, dependent on the strike. Finally we normalize all the weights to sum to one in order to get a result in the units of vol-points.

6.2. Differential Evolution

For the global search we use a differential evolution algorithm as first introduced in [28]. Differential evolution is a method based on agents in search-space and is used because it does not use the gradient and does not rely on a convex objective function, which we know is not the case for volatility models.

The differential evolution algorithm requires three inputs:

1. N is the number of agents, usually set around 10 times the number of input parameters to the objective function.
2. Cr is the crossover rate, a constant which determines if an agent is accepted, usually set around 0.8.
3. F is the differential weight, a multiplication constant, usually set at 0.7.

The algorithm for minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$ is as follows

Distribute all agents in $X = \{x_i, i = 1, \ldots, N\}$ randomly in search space according to a uniform distribution in each parameter
while termination criterion is not met **do**
  for each agent $x_i \in X$ **do**
    Pick three random agents $a, b, c \in X$, where $a \neq b \neq c \neq x_i$
    Generate n uniformly distributed random numbers $r_j \sim U(0, 1), j = 1, \ldots, n$
    for $j = 1, \ldots, n$ **do**

Set

$$\begin{cases} y_i = a_i + F(b_i - c_i) & , \quad \text{if} \quad r_j < Cr \\ y_i = x_i & , \quad \text{if} \quad r_j \geq Cr \end{cases}$$

Define $y_i = [y_{i,1}, \ldots, y_{i,n}]$

end for

if $f(y_i) \leq f(x_i)$ **then**

$x_i \leftarrow y_i$

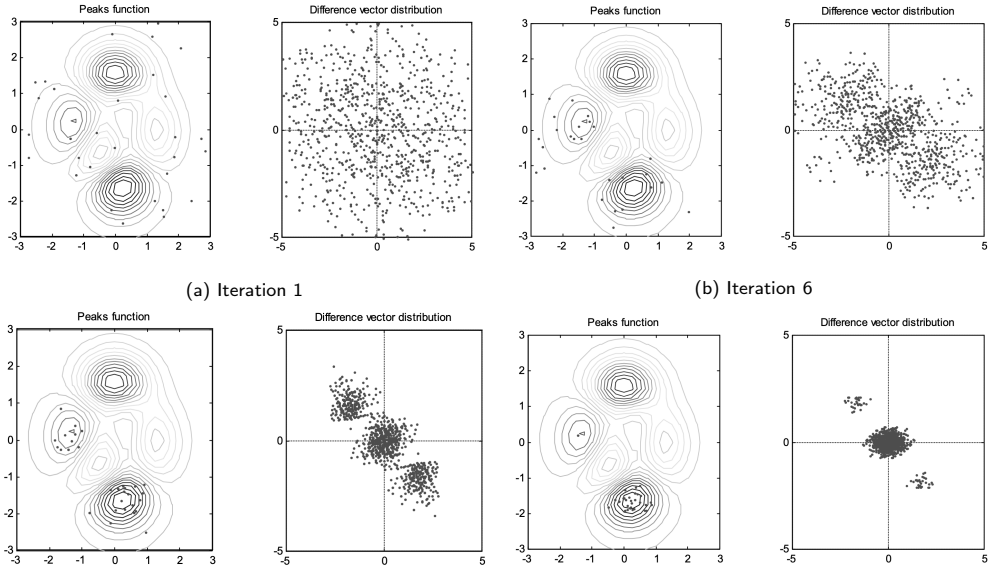
end if

end for

end while.

A modification to the algorithm as implemented by the `scipy` package is that F is a random variable distributed as $U(0.5, 1)$. The stopping criteria will depend on the computational feasibility of the model, but will always be a number of iterations.

The advantage of using a differential evolution algorithm on a non-convex surface is that the agents will first cluster together in local minima. Because of the combination of vectors formed, the agents can jump between these local minima and will eventually cluster together at the global minimum. An illustration of this phenomena is given in Figure 3, which is taken from section 2.1.7 of [29].



(a) Iteration 1

(b) Iteration 6

(c) Iteration 12

(d) Iteration 16

Figure 3: Illustration of clustering in differential evolution algorithm

In our implementation we use the standard `differential_evolution` function in `python`'s `scipy.optimize` library.

6.3. Nelder-Mead

The Nelder-Mead method (also known as the downhill simplex method) was first introduced in [30]. Just as the differential evolution algorithm it does not require a convex objective function, making it suitable for fitting our model.

The Nelder-Mead method requires a reflection, expansion, contraction and shrink coefficients as inputs, these are respectively denoted by α, γ, ρ and σ. Standard values would be $\alpha = 1$, $\gamma = 2$, $\rho = 1/2$ and $\sigma = 1/2$.

For a function $f : \mathbb{R}^n \to \mathbb{R}$ the algorithm works by considering $n + 1$ agents, sorted by their performance. We then look at the agents in space and take the middle of all but the worst performing agent and call this point the centroid. First we try reflecting the worst

performing agent through the centroid. If this yields a better result than the second-worst performing we replace our worst performing agent by our reflected point. However if this yields a better result than our best performing agent then we expand on this reflection by going further. We then test this expanded point, if it yields a better result we replace the reflected point by the expanded point.

On the other hand if the reflection did not improve the result on our second-worst performing agent but did improve on the worst preforming agent, we try a contraction with respect to the reflected point. This is a point between the centroid and the reflected point. If this yields a better result than the reflection we replace our worst performing agent with the contracted point.

If however the reflection did not improve on the worst performing agent then we try a contraction with respect to the worst performing agent. Again, if this yields a better result than the worst performing agent we replace it with the contracted agent.

Finally, if the contraction did not improve on the worst performing agent we shrink all agents but the best performing agent towards the best performing agent with a factor σ.

The pseudo-code for minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

Pick $n+1$ random agents in the search space according to a uniform distribution
while termination criterion is not met **do** Reflection
  Evaluate and order $f(x_1) \leq \ldots \leq f(x_{n+1})$
  Calculate x_o, the centroid of the best n agents ($x_o = \frac{1}{n}\sum_{i=1}^{n} x_i$)
  Determine reflection point $x_r = x_o + \alpha(x_o - x_{n+1})$ and evaluate $f(x_r)$
  if $f(x_1) \leq f(x_r) < f(x_n)$ **then**
    $x_{n+1} \leftarrow x_r$
  else if $f(x_r) < f(x_1)$ **then** Expansion
    Determine expansion point $x_e = x_o + \gamma(x_o - x_{n+1})$ and evaluate $f(x_e)$
    if $f(x_e) < f(x_r)$ **then**
      $x_{n+1} \leftarrow x_e$
    end if
  else if $f(x_r) \geq f(x_n)$ **then**
    if $f(x_r) < f(x_{n+1})$ **then** Contraction w.r.t. x_r
      Determine contraction point $x_c = x_o + \rho(x_r - x_o)$ and evaluate $f(x_c)$
      if $f(x_c) < f(x_r)$ **then**
        $x_{n+1} \leftarrow x_c$
      else Shrink
        for $i = 2, \ldots, n+1$ **do**
          $x_i \leftarrow x_1 + \sigma(x_i - x_1)$
        end for
      end if
    else if $f(x_r) \geq f(x_{n+1})$ **then** Contraction w.r.t. x_{n+1}
      Determine contraction point $x'_c = x_o + \rho(x_{n+1} - x_o)$ and evaluate $f(x'_c)$
      if $f(x'_c) < f(x_{n+1})$ **then**
        $x_{n+1} \leftarrow x'_c$
      else Shrink
        for $i = 2, \ldots, n+1$ **do**
          $x_i \leftarrow x_1 + \sigma(x_i - x_1)$
        end for
      end if
    end if
  end if
end while

A visualization of all the possible points on the surface is given by

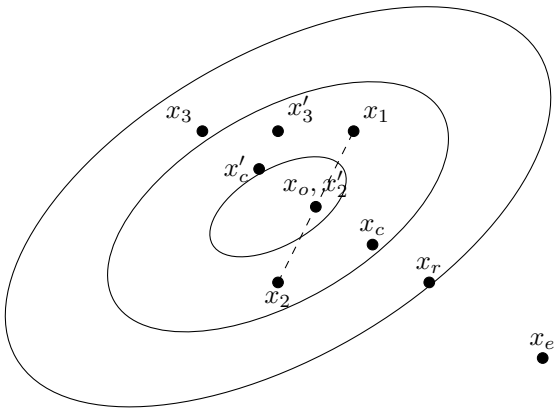


Figure 4: An illustration of the points in a Nelder-Mead iteration in two dimensions, with $f(x_1) < f(x_2) < f(x_3)$. The contour lines of the objective function are given by the solid lines. The dashed line is the plane for the centroid. Parameters are $(\alpha, \gamma, \rho, \sigma) = (1, 2, 1/2, 1/2)$.

The advantages of Nelder-Mead is that it requires relatively few function evaluations per iteration if we don't reach the shrink stage (which is rare in practice).

In this work we use the `minimize` function from the `scipy.optimize` package, this is an implementation of the method in [31]. For this function we can input the best $n + 1$ results of the differential evolution algorithm as a starting point for the Nelder-Mead method. This saves us iterations on searching the space if we already know roughly where the minimum is through our differential evolution.

7. Model Comparison

The goal is to compare the models and see if one is better than the other. There are many ways to compare these models, here we choose to compare the models based on three things: computation time, fitting error and hedging result.

7.1. Computation Time

There are two computation times we compare. One is for a single function evaluation, so for the computation of a single vol surface. The second is for the whole minimization.

The first gives a good metric on how easy it is to compute a vol surface.

The second computation time is the computation time of the whole minimization. Eventually the goal is to minimize the model to the market, so comparing the time it takes to minimize this is crucial in comparing the models. This however strongly depends on the parameters chosen in the minimization methods, which can also be optimized, although that is not part of this research.

7.2. Objective Function

Next we compare the models by their objective function value, so by their weighted RMSE. This gives an indication of how good the fit is to the market data.

7.3. Hedging

Finally we compare the models by their hedging result. What this means is that we compute a theoretical value of a call option at time T ($\hat{V}^{\text{Call}}(S_T)$), while knowing the market call price at $T = 0$ ($V^{\text{Call}}(S_0)$). We do this by hedging against the first order derivative in each parameter:

$$\hat{V}^{\text{Call}}(S_T) = V^{\text{Call}}(S_0) + \sum_{i=1}^{n} \frac{\partial V^{\text{Call}}_{\text{model}}(S_0)}{\partial \theta_0^i}(\theta_T^i - \theta_0^i),$$

where $V^{\text{Call}}_{\text{model}}(S_t)$ is the call price computed by the model at time t. We then compute the weighted RMSE of $\hat{V}^{\text{Call}}(S_T) - V^{\text{Call}}(S_T)$. The weights are the same as those in the minimization.

Next to this we always use a delta-hedge. Thus we compare the the hedging result for hedging the delta and all the parameters to only hedging the delta. This should then give an indication on how well the model describes what is actually happening in the market.

8. Results

This chapter will only be a presentation of the results, a discussion and conclusion of the results will be give in the following chapters.

8.1. Introduction

Firstly, the market conditions we are fitting to are introduced. This is followed by the parameters used for the minimization.

8.1.1. Market Conditions

For fitting vol surfaces we look at the month of November 2021 and fit the models to the vol surface of 12:00 on trading days. The vol surface is for Eurostroxx 50 options and is supplied by IMC Trading, which is the largest options trading firm for Eurostoxx. For the vol surface we considered all monthly expiries up to two years out, thus our vol surface is constructed using twelve vol curves. The strikes for all these vol curves are based on what is tradable on the Eurostoxx. The spot values of the Eurostoxx 50 for the considered vol surfaces are displayed in Figure 5

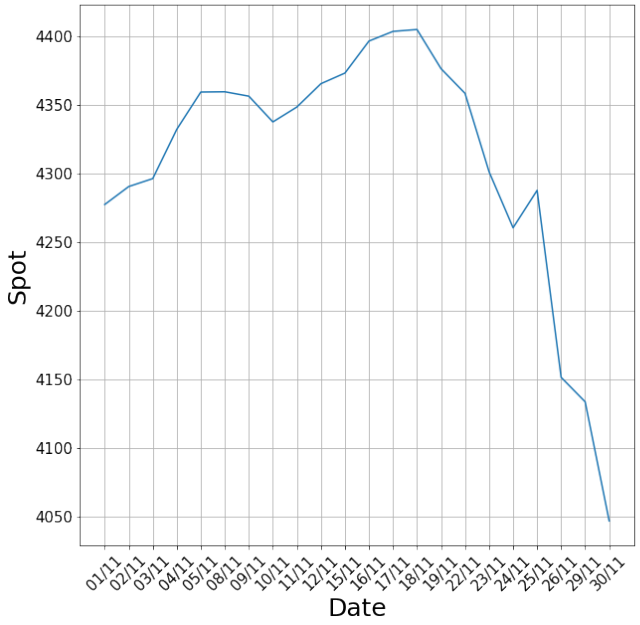
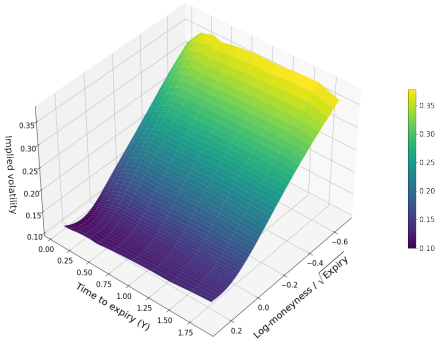
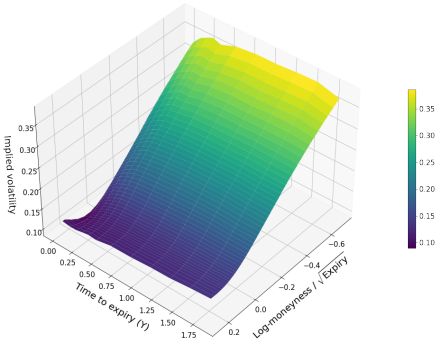


Figure 5: Spot values of Eurostoxx 50 futures from 1/11/2021 until 30/11/2021.

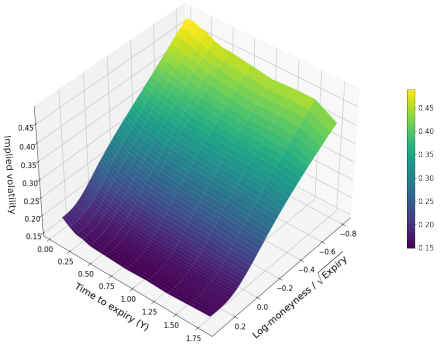
Interesting to note is that there is a calm period and also a period of sharp decline in this month. For this chapter we will mostly consider the vol surfaces from 1/11, 15/11 and 30/11. This should give an accurate representation of what the market behaved like throughout the month. The three vol surfaces for these dates are plotted here



(a) Market vol surface on 1/11



(b) Market vol surface on 15/11



(c) Market vol surface on 30/11

Figure 6: Plots of vol surfaces throughout November 2021

As can be seen in figure 6c, the dip in spot leads to an increase in the vol surface, while for the other two curves the vol is lower towards the short expiries. Also note that we scale the strike axis to log-moneyness / $\sqrt{\text{expiry}}$. The scaling to the expiry is to ensure that we do not extrapolate the short expiries too far out. The extrapolation is a linear extrapolation of log-moneyness versus σ^2_{imp} as per the moment formula in [32], this ensures no arbitrage in the tails.

8.1.2. Parameters

A big influence on the minimization time for all these models are the parameters chosen. For the Bergomi models we took $20,000$ paths and 100 time steps, in line with [33]. While for the rough Heston and lifted Heston we took 500 COS method evaluations, in the same order of magnitude as [12] and 500 and 400 time steps respectively. For the lifted Heston model we also took a 20 dimensional Riccati equation with $r_n = 2.5$, as suggested in [12].

For the minimization, the Bergomi models had 20 differential evolution stages, while the Heston models had 5. This difference was added because it requires a lot of evaluations, which was infeasible for the Heston models due to their long computation times. The Nelder-Mead optimization was terminated after a function evaluation difference of less than 10^{-4} was achieved.

8.2. Computation Time

The results for the computation time of a single realization together with the results for the minimization are given in Table 1.

model	single vol surface computation time (s)	total minimization time (h:m)	minimization function evaluations
rB	3.87 ± 0.009	$2{:}05 \pm 0{:}45$	946 ± 96
rB7p	4.89 ± 0.01	$4{:}25 \pm 1{:}26$	$2,321 \pm 433$
rH	177 ± 0.6	$26{:}48 \pm 12{:}29$	243 ± 45
LH	34.9 ± 0.09	$14{:}57 \pm 10{:}35$	$1,959 \pm 1.182$

Table 1: For the different models the times of vol surface evaluation and total minimization together with number of function evaluations needed for the minimization.

Do note that in this implementation we have not parallelized the computation of the vol curves, thus all twelve smiles are computed in a for-loop. Parallelizing this could theoretically speed up by time columns by a factor 12. The results for all the individual minimizations are given in the following figure

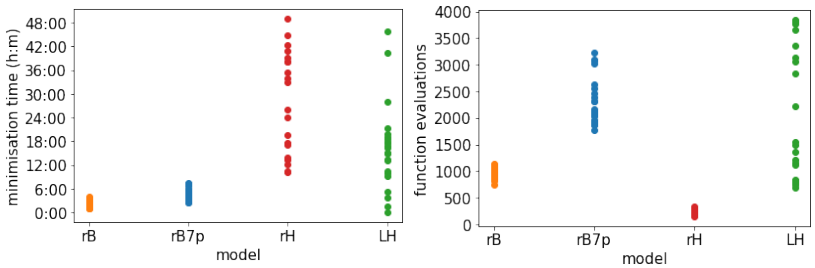


Figure 7: Minimization times and function evaluations

8.3. Volatility Surface

The different minimized weighted RMSE values are given in Figure 8

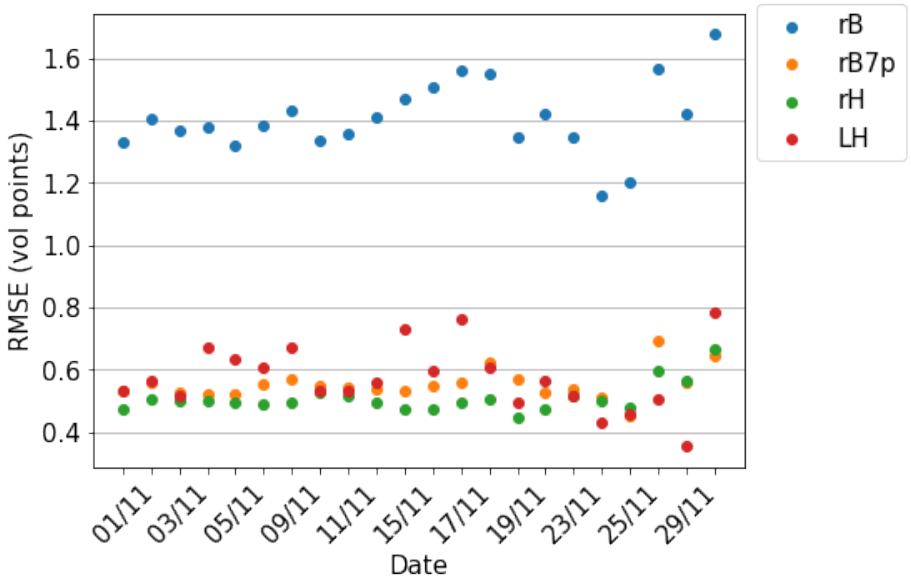


Figure 8: Weighted RMSE between fitted curve and market curve for different dates

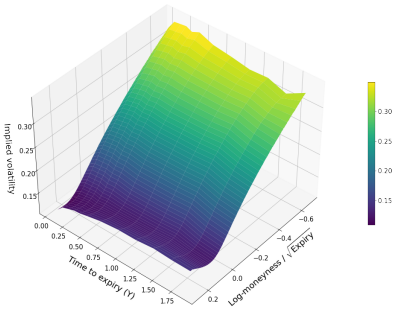
The means in Figure 8 with their errors are given in Table 2

model	weighted RMSE (vol points)
rB	1.41 ± 0.11
rB7p	0.55 ± 0.05
rH	0.51 ± 0.05
LH	0.57 ± 0.11

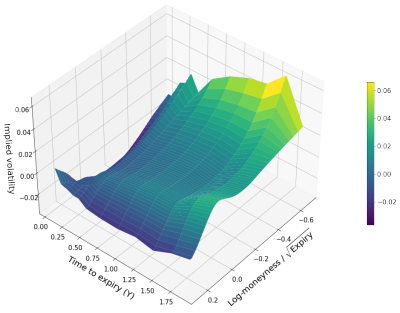
Table 2: For the different models the results for the weighted RMSE between the minimized model surface and the market surface.

Next we show the results of the minimized objective function by plotting the vol surfaces and their differences between the market surface for the same dates as in Figure 6. The difference surface is calculated by subtracting the model surface from the market surface. We also plot the market and model vol curves for a short, middle and long expiry for these three dates for all models as this might give a more intuitive insight into how the models behave.

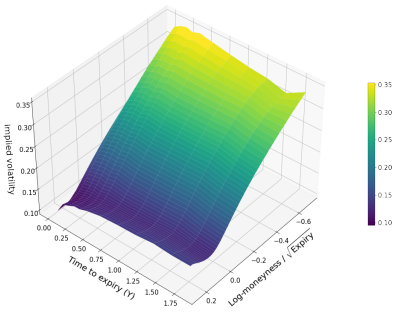
8.3.1. Rough Bergomi



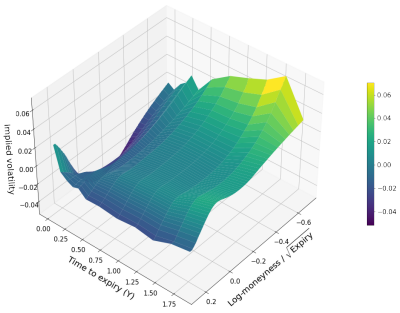
(a) Rough Bergomi vol surface on 1/11



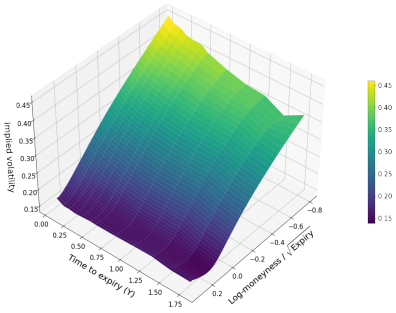
(b) Rough Bergomi difference surface on 1/11



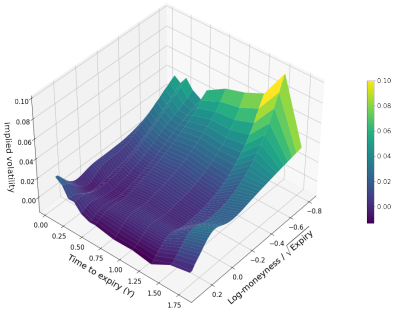
(c) Rough Bergomi vol surface on 15/11



(d) Rough Bergomi difference surface on 15/11



(e) Rough Bergomi vol surface on 30/11



(f) Rough Bergomi difference surface on 30/11

Figure 9: Rough Bergomi vol surfaces and difference between market surfaces

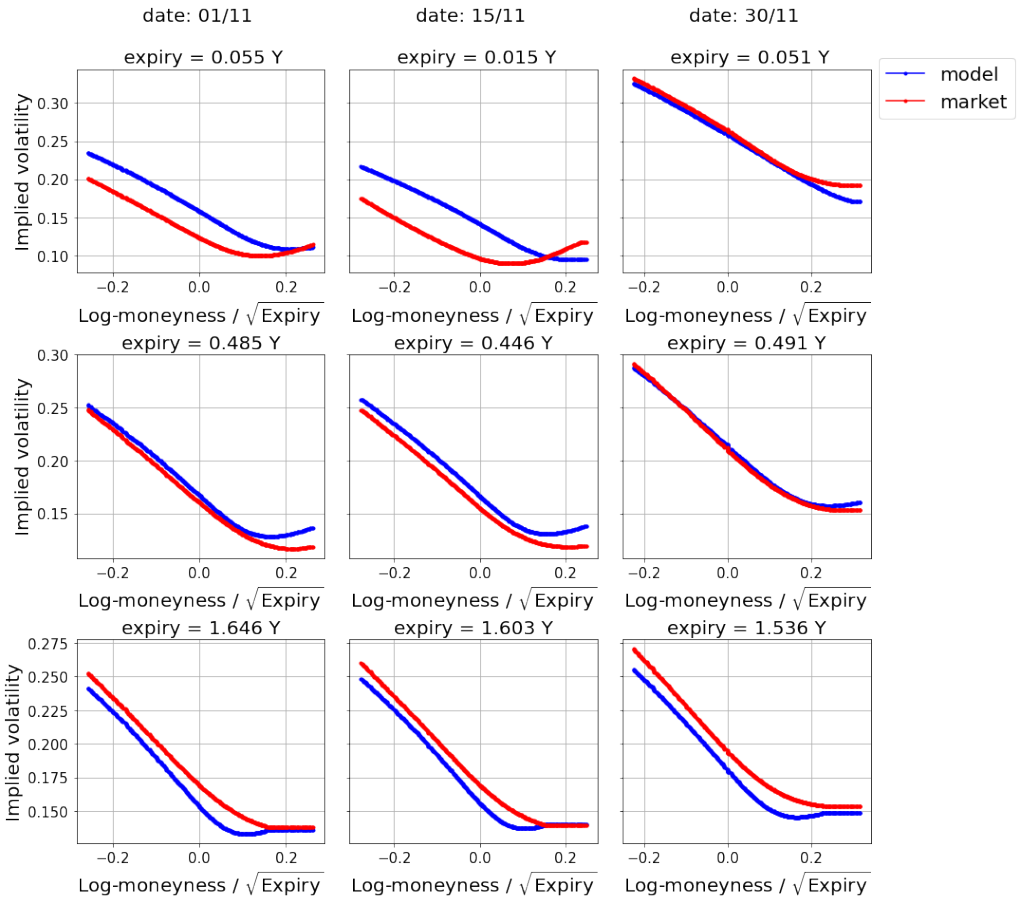
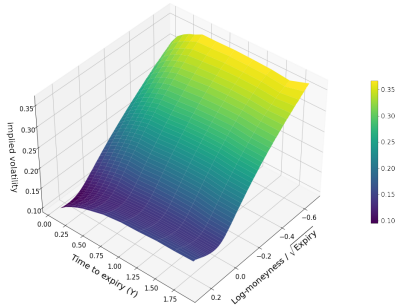
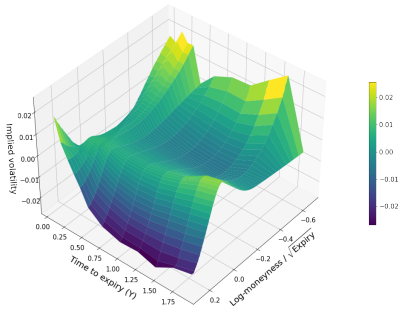


Figure 10: Market and rough Bergomi vol curves for short, middle and long expiries

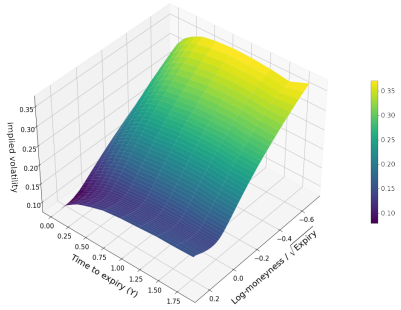
8.3.2. Rough Bergomi 7 parameters



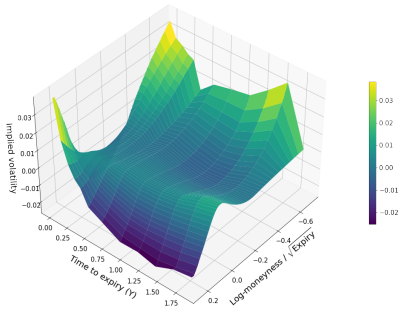
(a) Rough Bergomi 7 parameters vol surface on 1/11



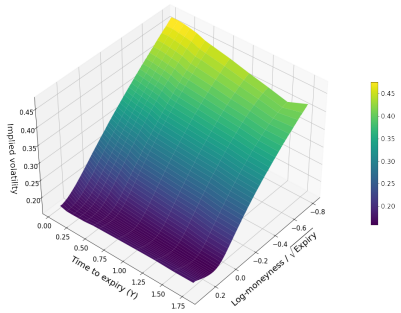
(b) Rough Bergomi 7 parameters difference surface on 1/11



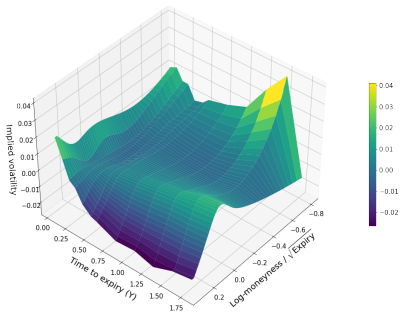
(c) Rough Bergomi 7 parameters vol surface on 15/11



(d) Rough Bergomi 7 parameters difference surface on 15/11



(e) Rough Bergomi 7 parameters vol surface on 30/11



(f) Rough Bergomi 7 parameters difference surface on 30/11

Figure 11: Rough Bergomi 7 parameters vol surfaces and difference between market surfaces

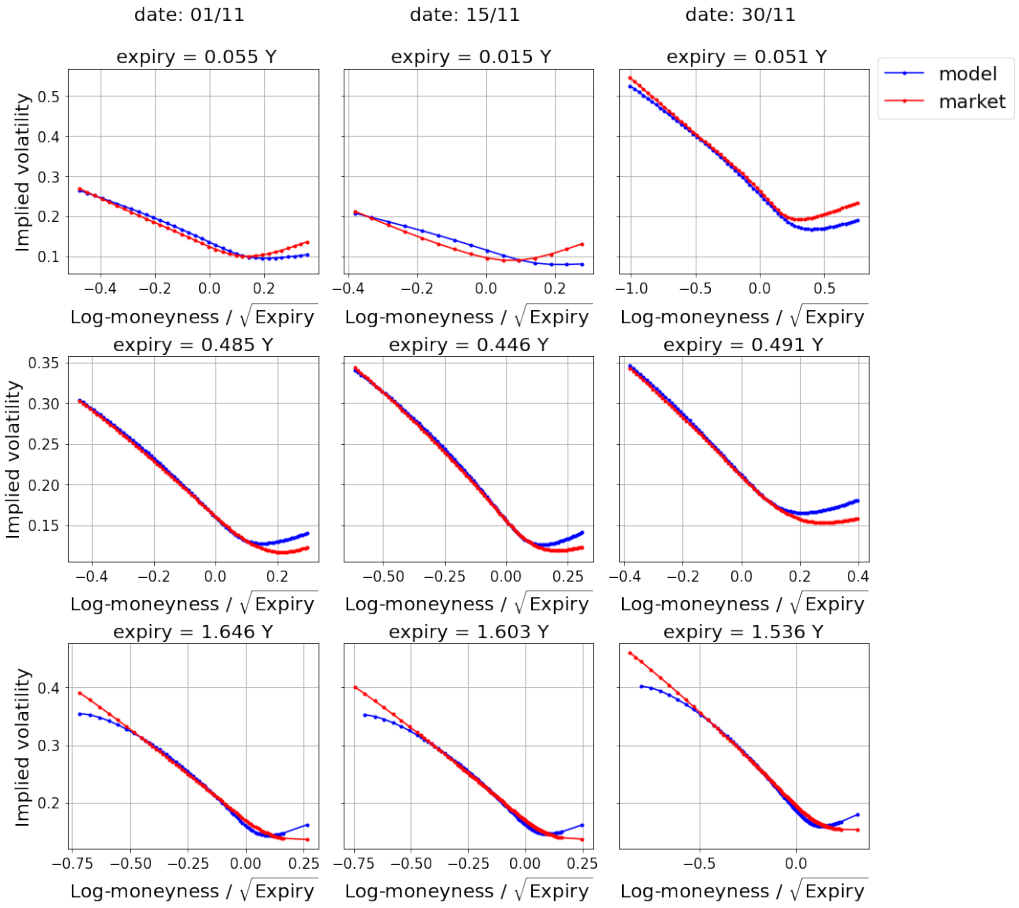


Figure 12: Market and rough Bergomi 7 parameters vol curves for short, middle and long expiries

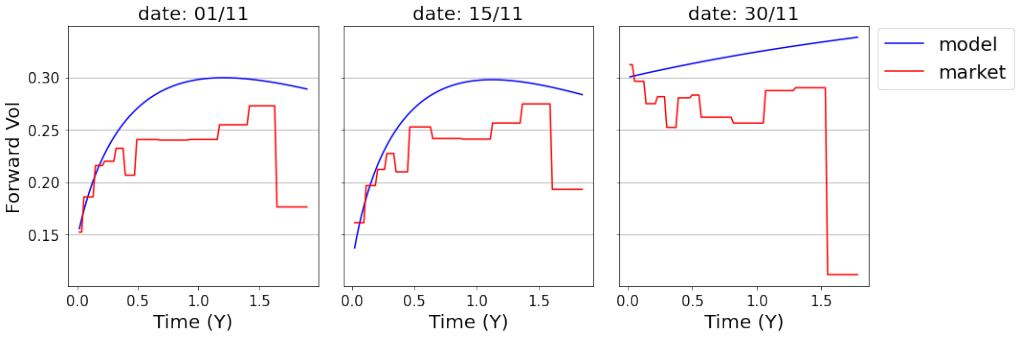
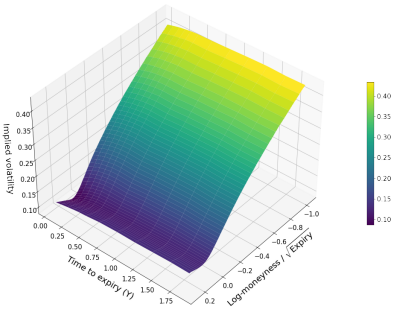
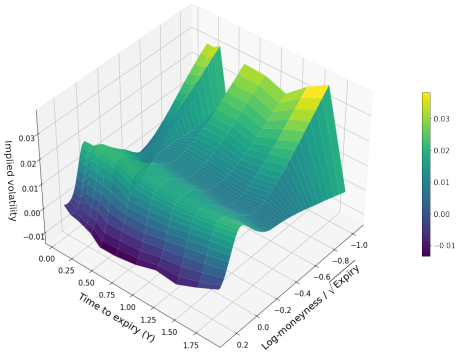


Figure 13: Market and model fitted forward volatility curves.

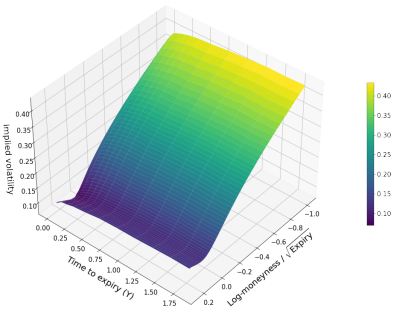
8.3.3. Rough Heston



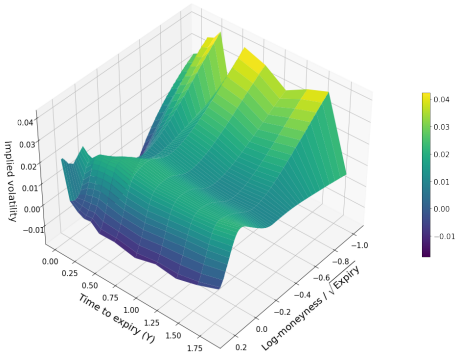
(a) Rough Heston vol surface on 1/11



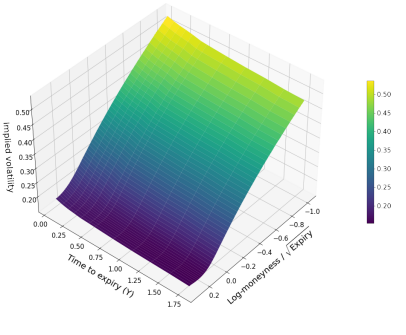
(b) Rough Heston difference surface on 1/11



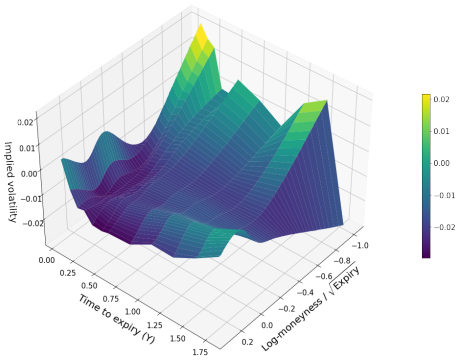
(c) Rough Heston vol surface on 15/11



(d) Rough Heston difference surface on 15/11



(e) Rough Heston vol surface on 30/11



(f) Rough Heston difference surface on 30/11

Figure 14: Rough Heston vol surfaces and difference between market surfaces

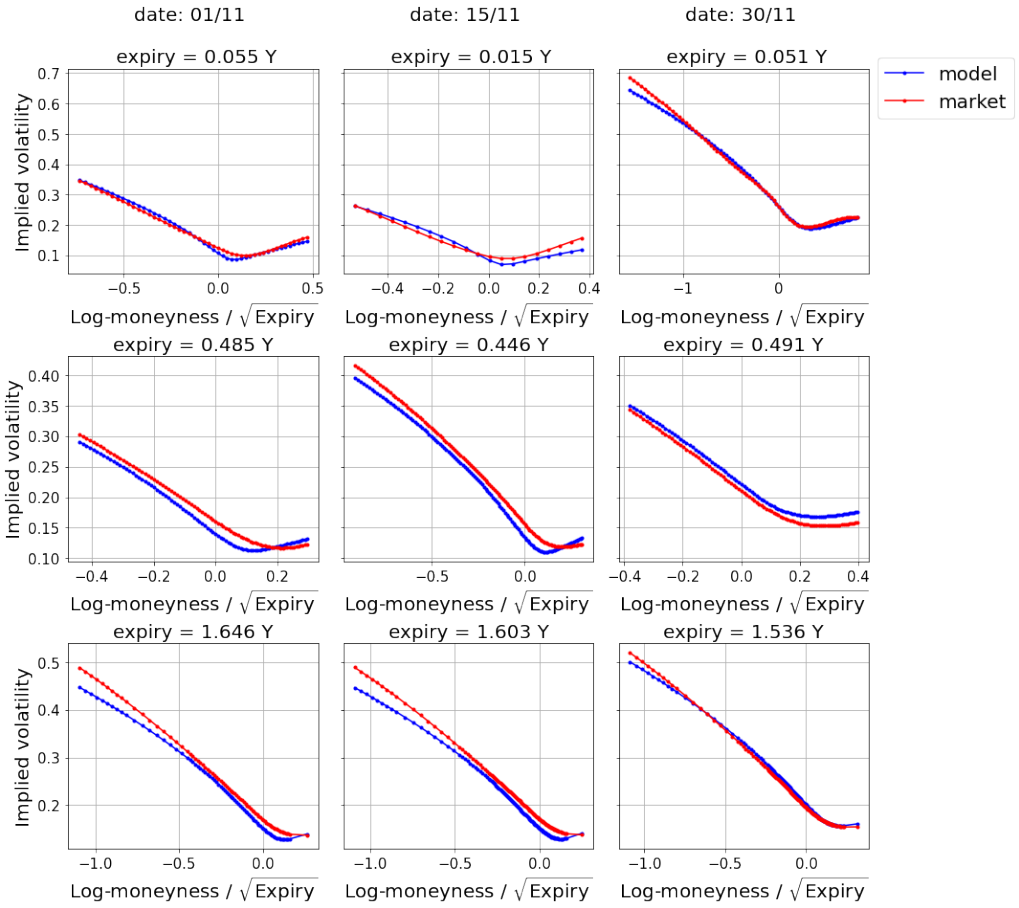
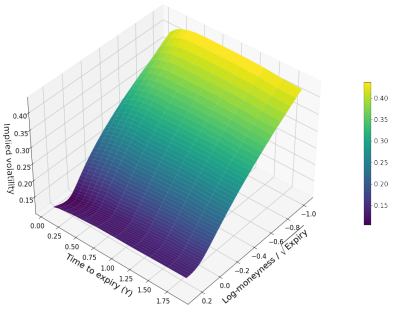
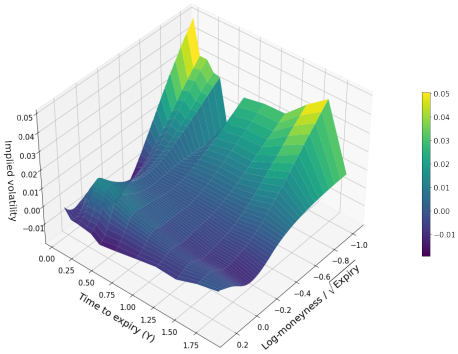


Figure 15: Market and rough Heston vol curves for short, middle and long expiries

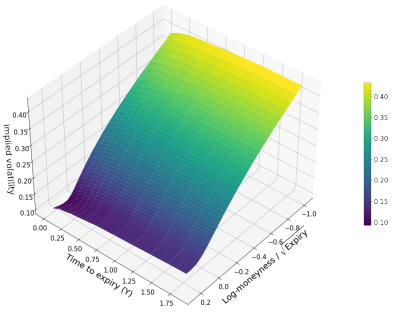
8.3.4. Lifted Heston



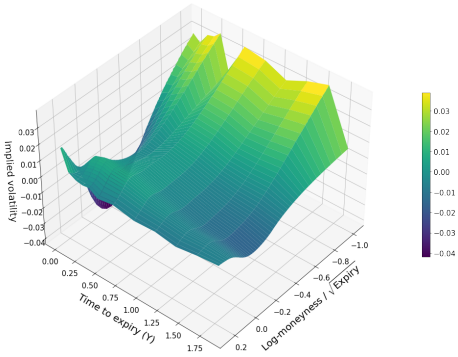
(a) Lifted Heston vol surface on 1/11



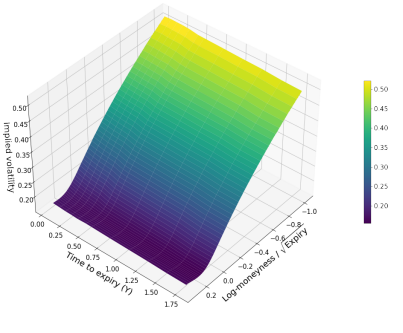
(b) Lifted Heston difference surface on 1/11



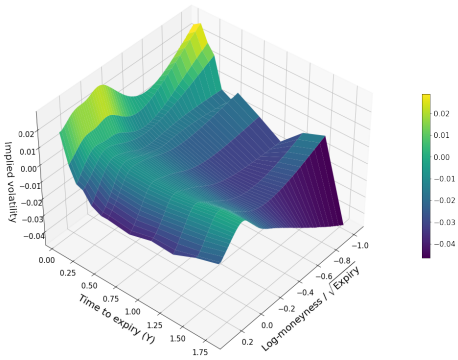
(c) Lifted Heston vol surface on 15/11



(d) Lifted Heston difference surface on 15/11



(e) Lifted Heston vol surface on 30/11



(f) Lifted Heston difference surface on 30/11

Figure 16: Lifted Heston vol surfaces and difference between market surfaces

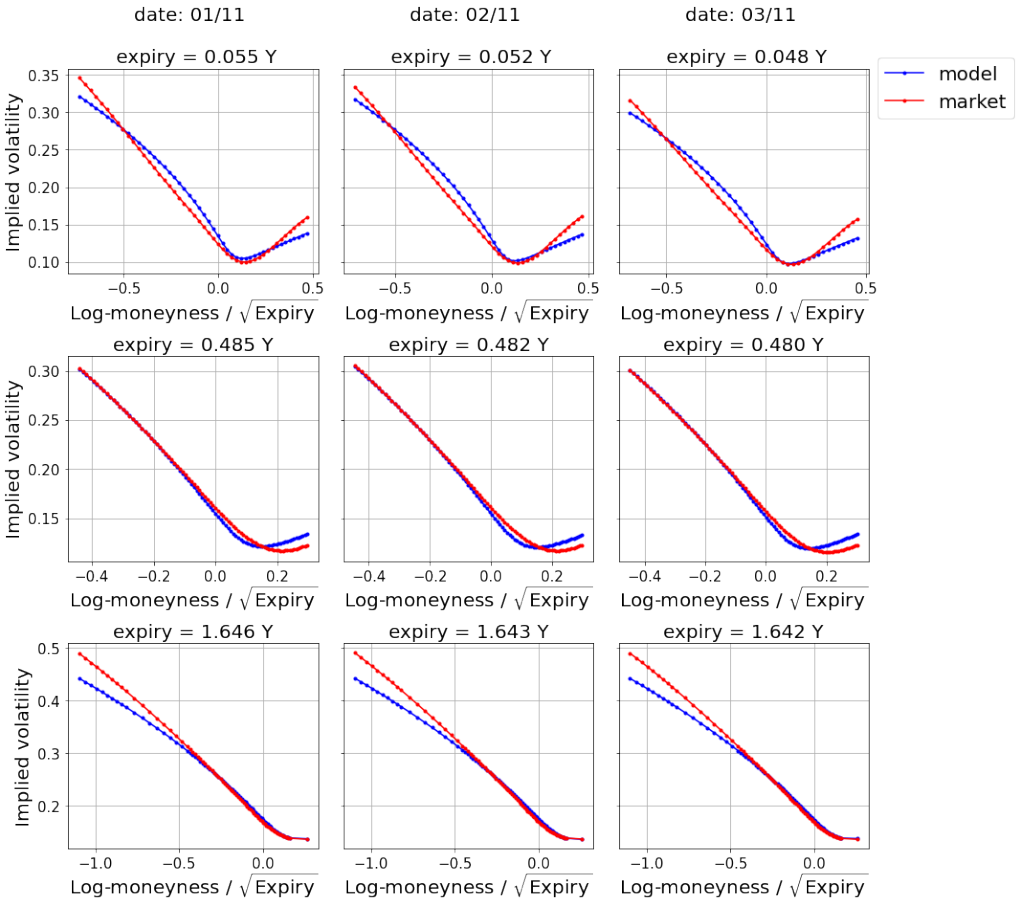


Figure 17: Market and lifted Heston vol curves for short, middle and long expiries

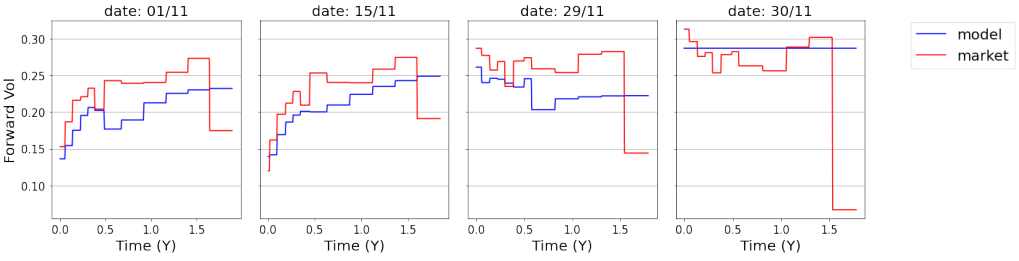


Figure 18: Market and model fitted forward volatility curves.

8.4. Hedging

In order to give a clear representation of the hedging performance of each model we consider the following plots

- Firstly we consider the time series of the optimized parameters throughout November 2021. This will give a representation of the fluctuation of a parameter, indicating if a hedge against it will be useful. A highly fluctuating parameter while the market is stable would indicate the parameter does not represent anything in the market, thus making it unnecessary to hedge against.
- Secondly we consider a sample of the hedging results. In order to do this we consider the hedging results for three different dates (2/11, 16/11 and 30/11) and again for a short, middle and long expiry. Since we are already considering a delta hedge, the plots will include a delta hedged price difference as well as the magnitude of the delta

hedge. We also consider a hedge against the FVC. The derivative of this is calculated by shifting the FVC up by $\varepsilon = 10^{-3}$ and dividing the difference in call prices by $\varepsilon = 10^{-3}$. The hedging factor is then calculated by multiplying this derivative by the area change in forward variance curves.
A note on this plot is that if one were to calculate the price difference for the hedge against a parameter, the plot should be inverted and added to the delta hedged price difference. Thus a positive parameter hedge curve results in a downward shift in the delta hedged price difference.

- Finally, putting all these results together will lead to the last plot. Here we plot the difference in weighted MSE of the delta hedge vs the delta+parameter hedge as a time series. The weights are the same as those used in the minimization. Since the underlying dropped from the 26^{th} to the 27^{th}, this will lead to an outlier. Thus we will also plot the time series up until the 26^{th}. The title of these plots will have the mean of its time series. If these time series are negative, it means that hedging against the parameter decreased the MSE of the price difference, indicating a hedge against the parameter would be an improvement over just a delta hedge. A plot with only the delta hedged MSE is provided at the top, this indicates how well the model delta does and gives an indication on the ratio of improvement for the parameter hedges.
Next to this we also plot how well the model would do if we were to hedge against the delta and all the parameters. The MSE of this is then again compared to the MSE of the delta hedge, as before.
Finally we again consider the hedge against the FVC and plot its weighted MSE time series.

8.4.1. Rough Bergomi

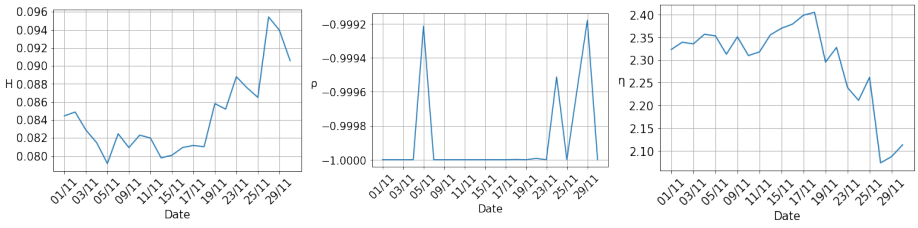


Figure 19: Parameter evolution for rough Bergomi

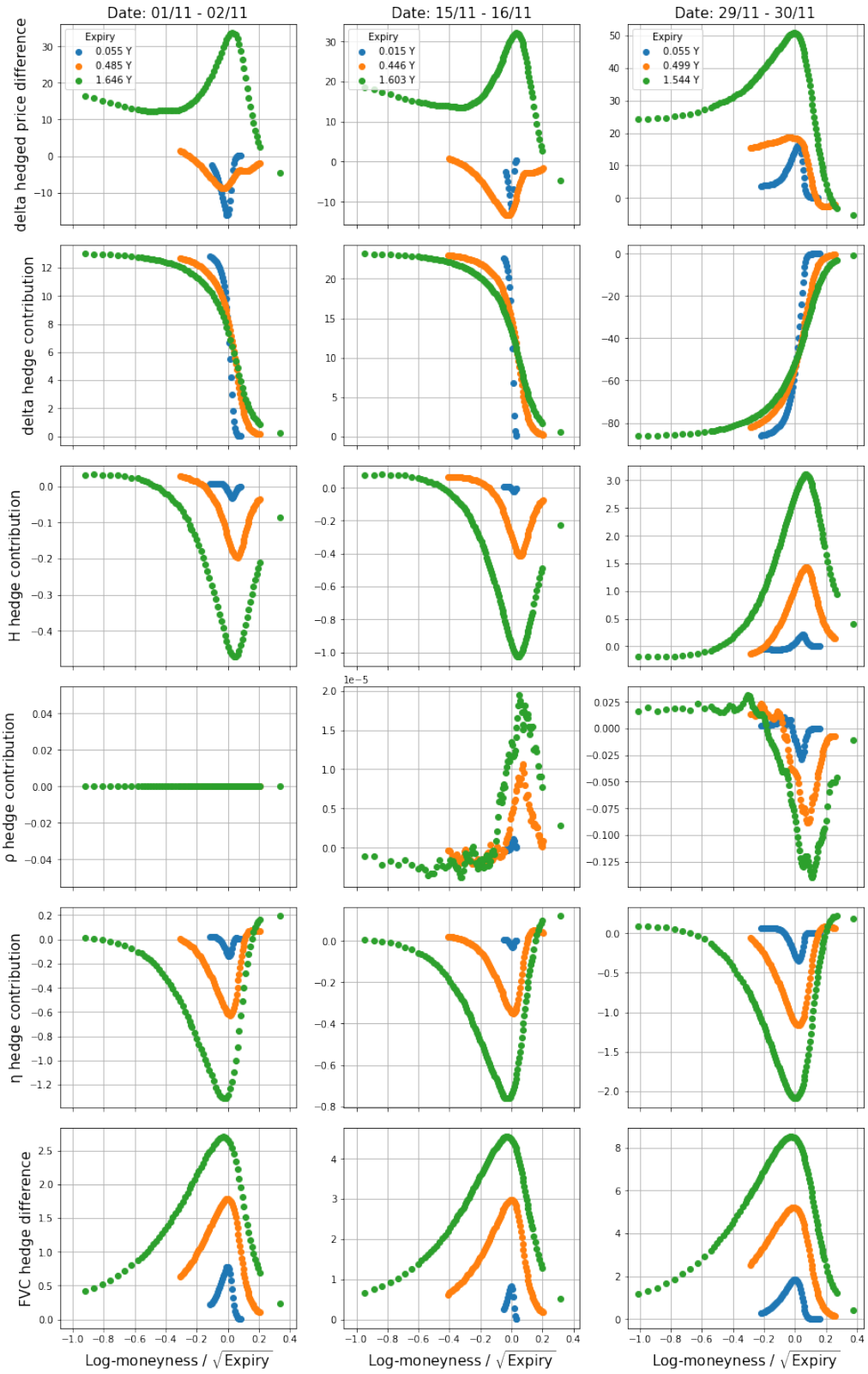
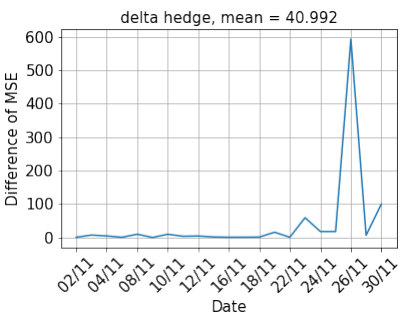
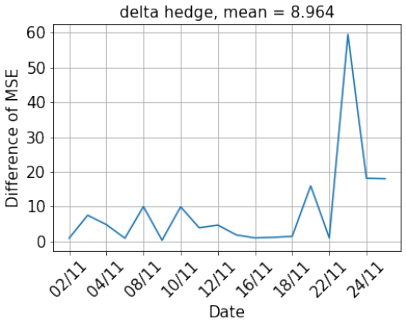
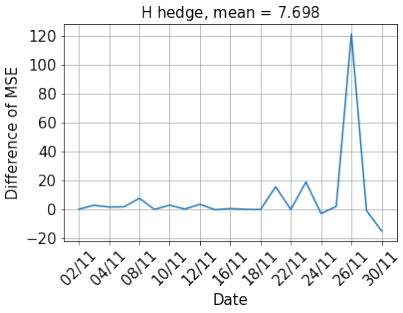
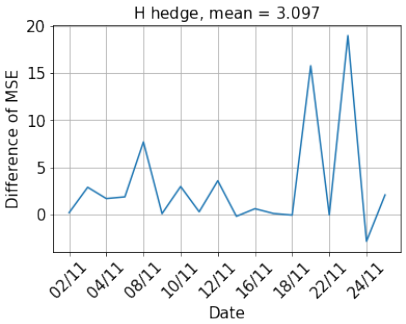


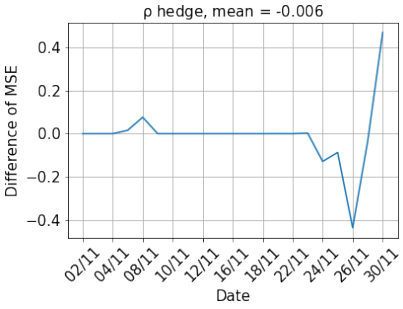
Figure 20: Hedging results of hedging against the parameters of the rough Bergomi model

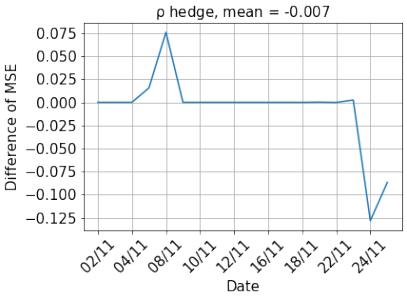

delta hedge, mean = 40.992
Difference of MSE
Date

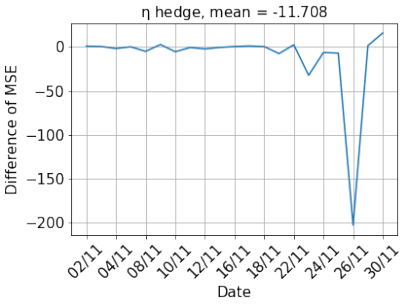

delta hedge, mean = 8.964
Difference of MSE
Date

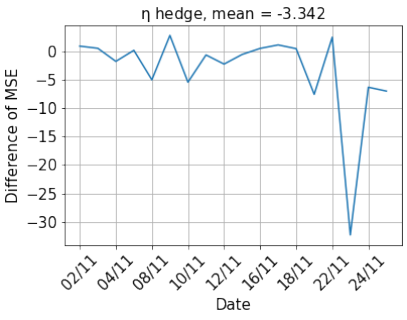

H hedge, mean = 7.698
Difference of MSE
Date

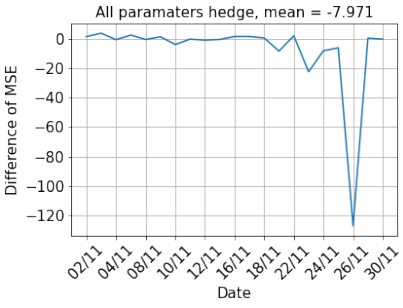

H hedge, mean = 3.097
Difference of MSE
Date

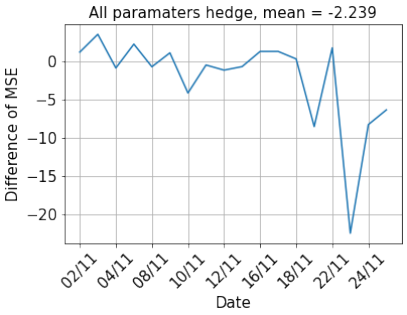

ρ hedge, mean = -0.006
Difference of MSE
Date


ρ hedge, mean = -0.007
Difference of MSE
Date


η hedge, mean = -11.708
Difference of MSE
Date


η hedge, mean = -3.342
Difference of MSE
Date


All paramaters hedge, mean = -7.971
Difference of MSE
Date


All paramaters hedge, mean = -2.239
Difference of MSE
Date

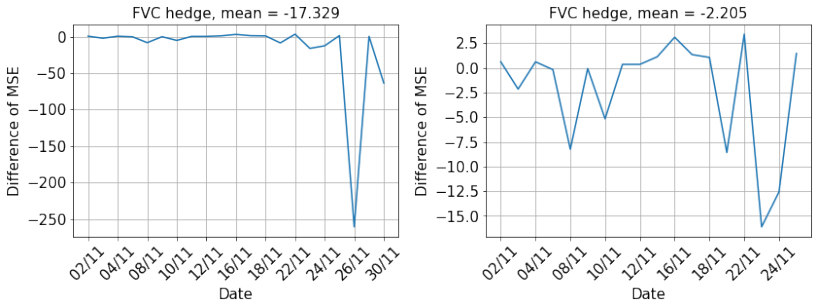


Figure 21: Time series of delta hedged weighted MSE and the difference between the hedged result MSE and the delta hedged MSE.

8.4.2. Rough Bergomi 7 parameters

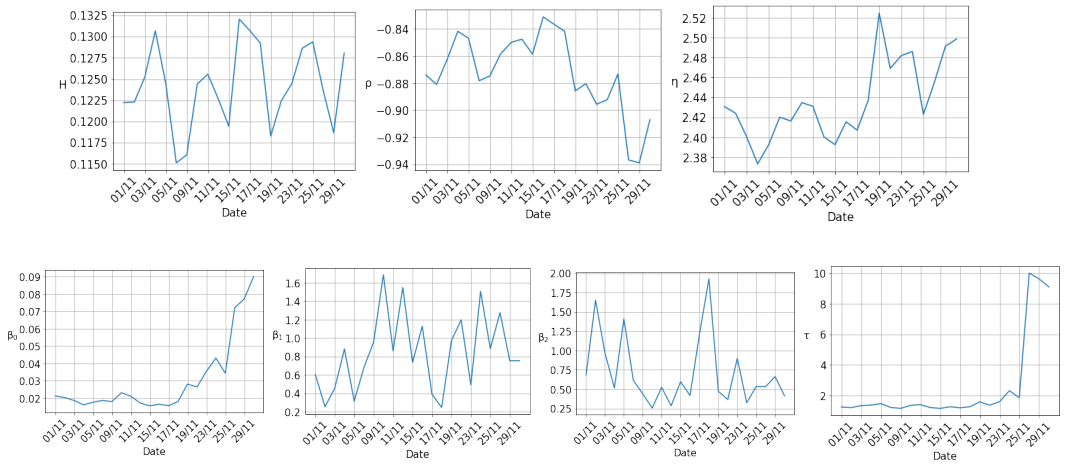
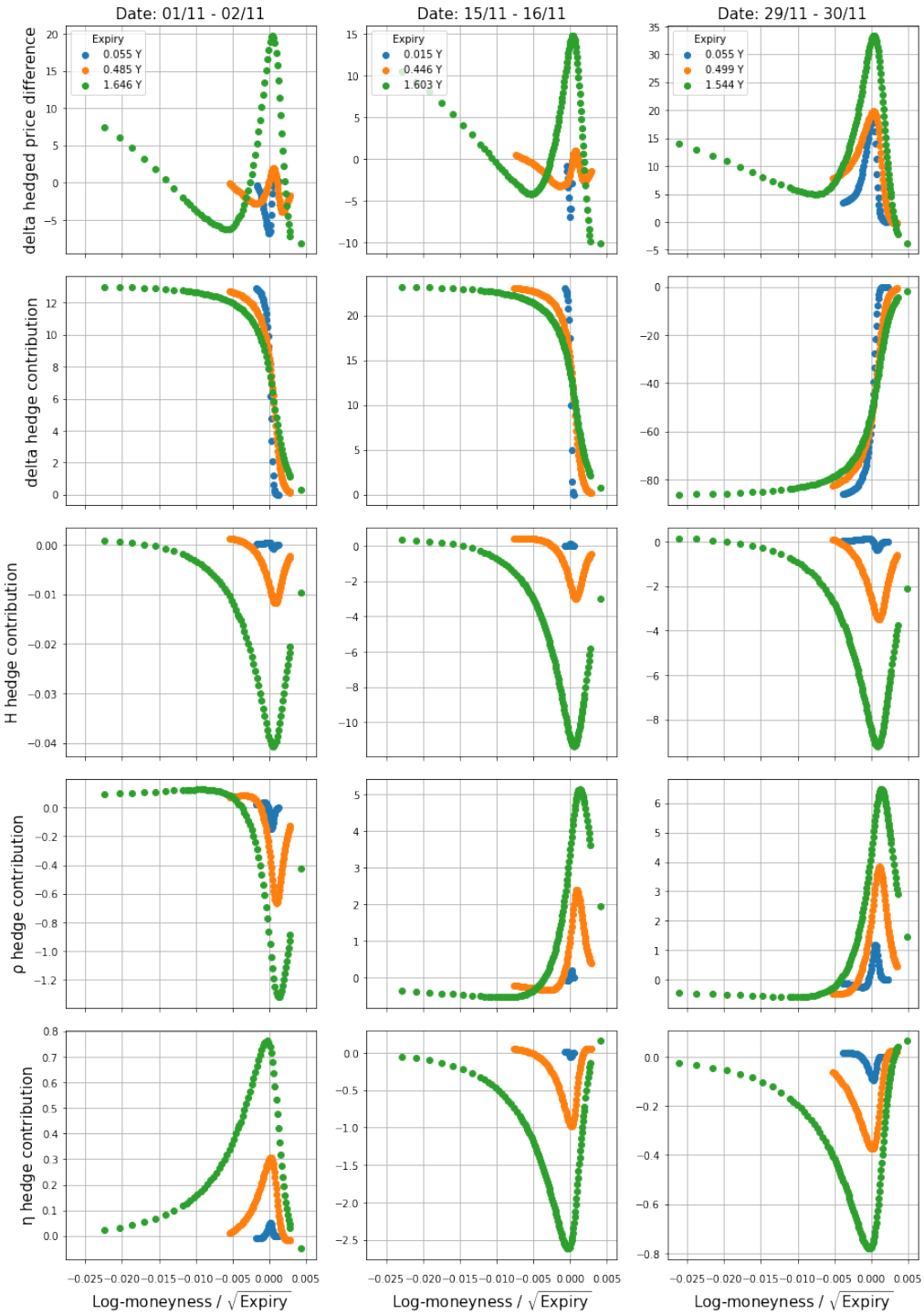


Figure 22: Parameter evolution for rough Bergomi 7 parameter


Date: 01/11 - 02/11
Date: 15/11 - 16/11
Date: 29/11 - 30/11
Expiry
0.055 Y
0.485 Y
1.646 Y
Expiry
0.015 Y
0.446 Y
1.603 Y
Expiry
0.055 Y
0.499 Y
1.544 Y
delta hedged price difference
delta hedge contribution
H hedge contribution
ρ hedge contribution
η hedge contribution
Log-moneyness / √Expiry
Log-moneyness / √Expiry
Log-moneyness / √Expiry

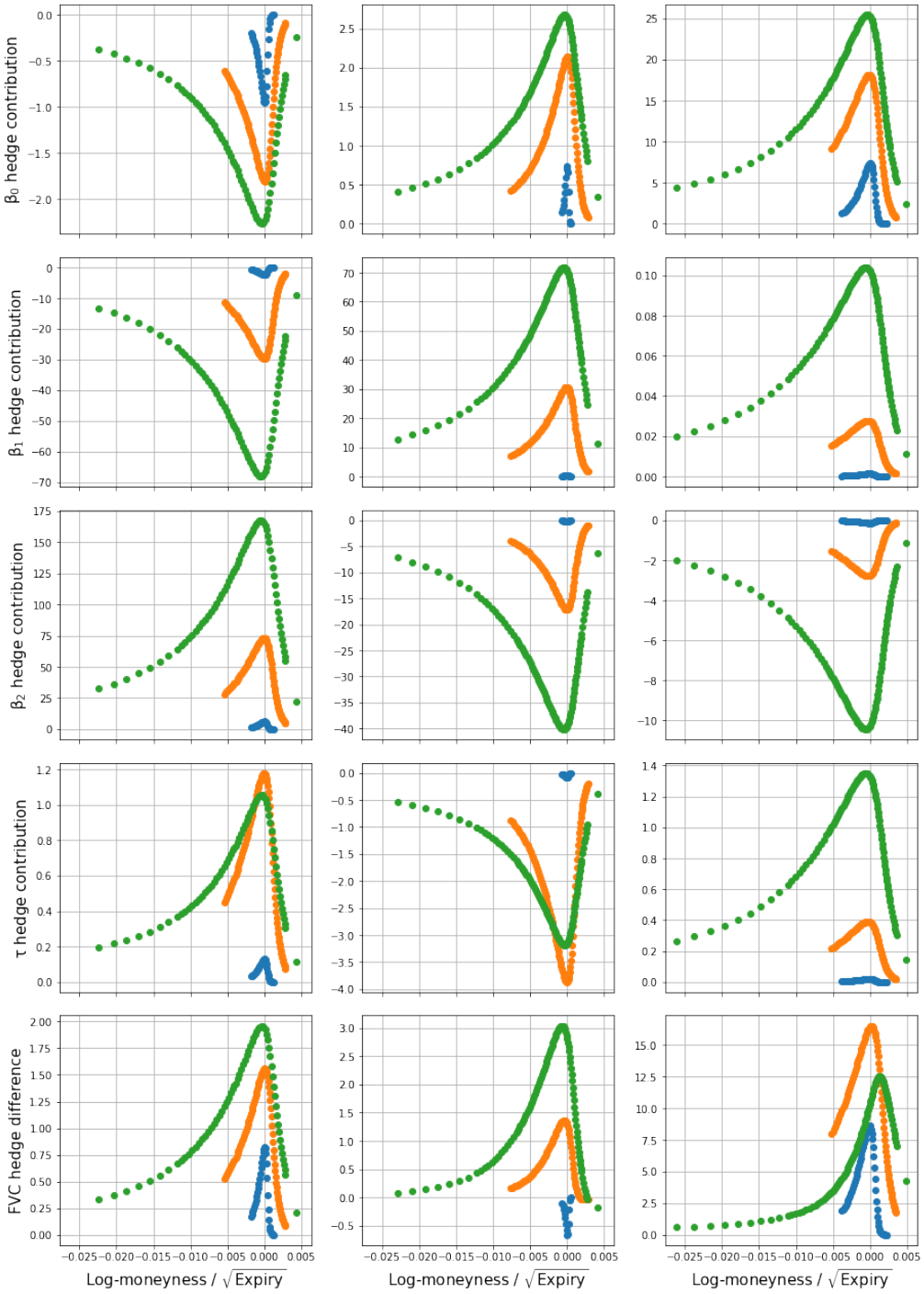
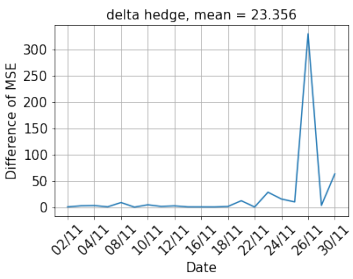
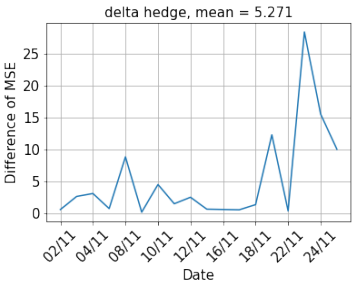
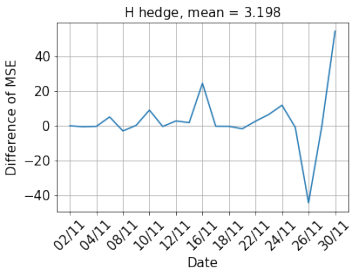
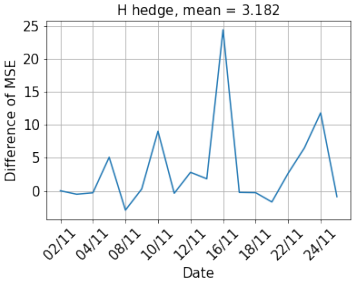


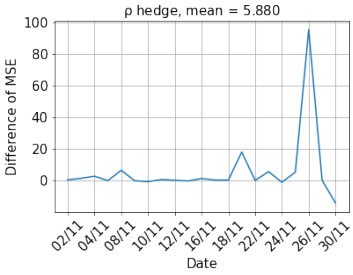
Figure 22: Hedging results of hedging against the parameters of the rough Bergomi 7 parameter model

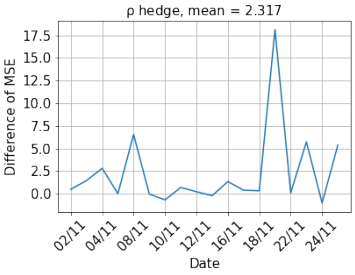

delta hedge, mean = 23.356
Difference of MSE
Date

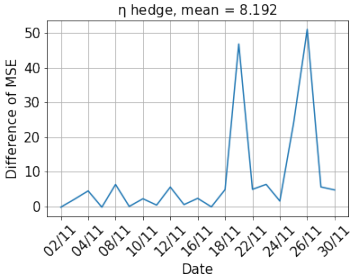

delta hedge, mean = 5.271
Difference of MSE
Date

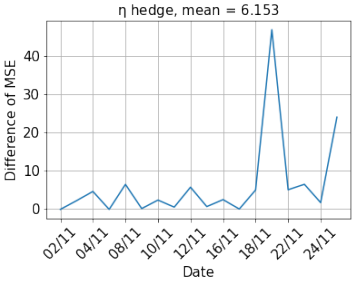

H hedge, mean = 3.198
Difference of MSE
Date

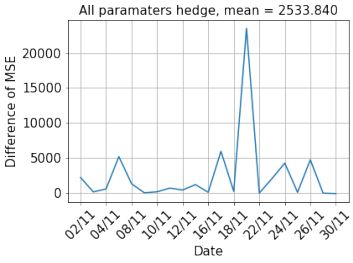

H hedge, mean = 3.182
Difference of MSE
Date

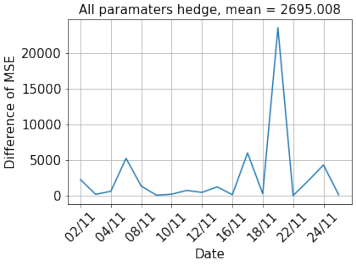

ρ hedge, mean = 5.880
Difference of MSE
Date


ρ hedge, mean = 2.317
Difference of MSE
Date


η hedge, mean = 8.192
Difference of MSE
Date


η hedge, mean = 6.153
Difference of MSE
Date


All paramaters hedge, mean = 2533.840
Difference of MSE
Date


All paramaters hedge, mean = 2695.008
Difference of MSE
Date

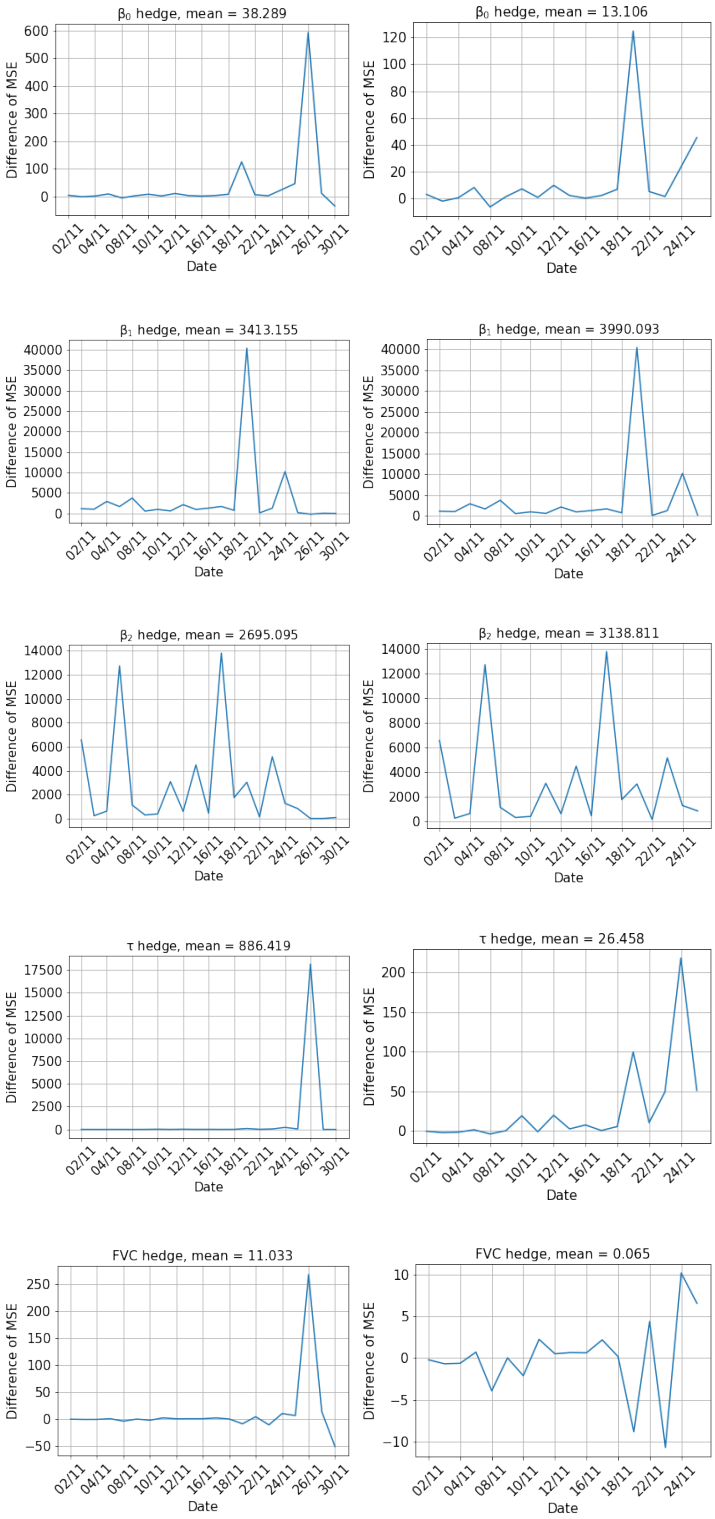


Figure 23: Time series of delta hedged weighted MSE and the difference between the hedged result MSE and the delta hedged MSE.

8.4.3. Rough Heston

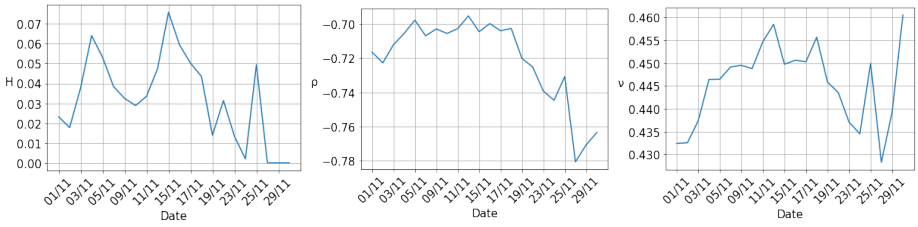


Figure 24: Parameter evolution for rough Heston

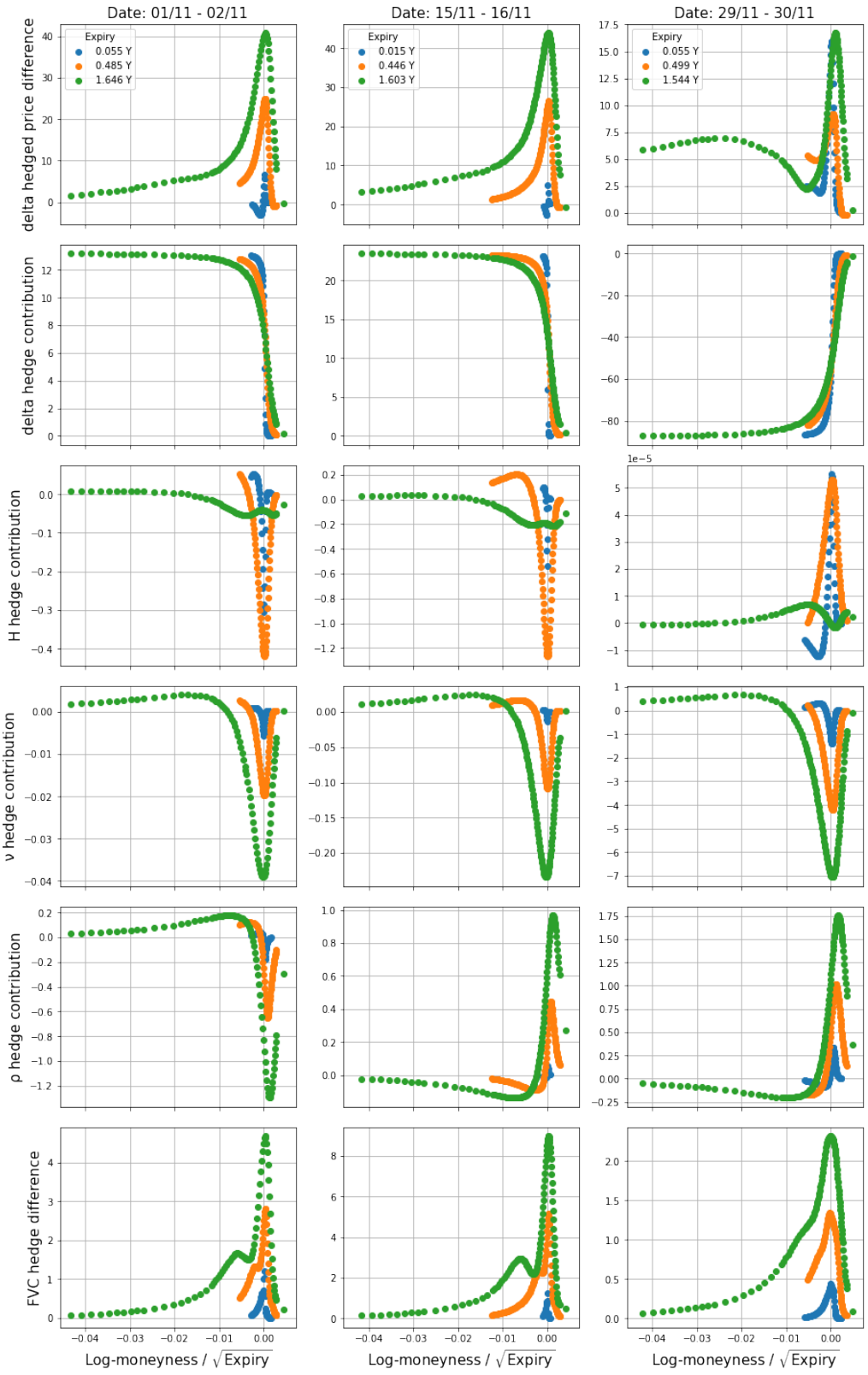
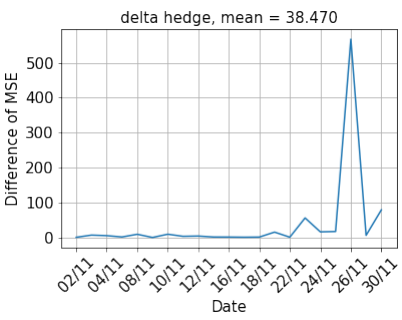
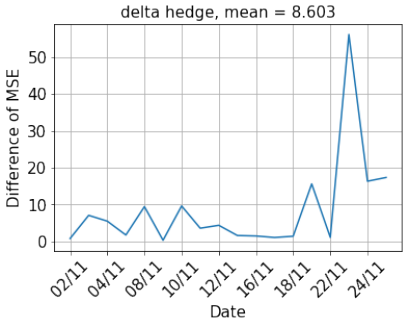
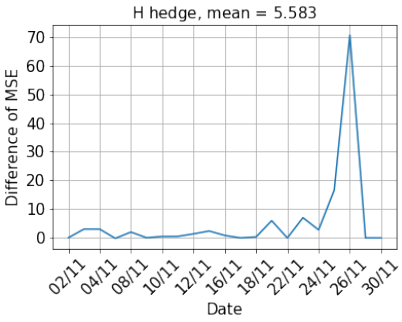
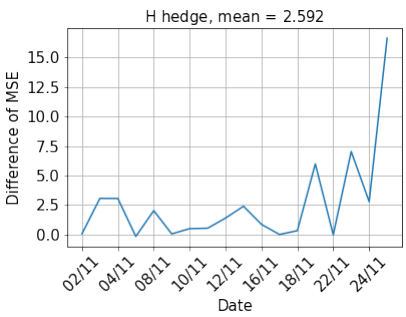


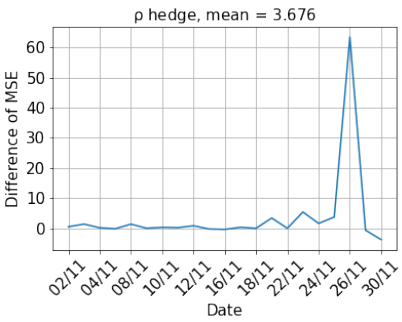
Figure 25: Hedging results of hedging against the parameters of the rough Heston model

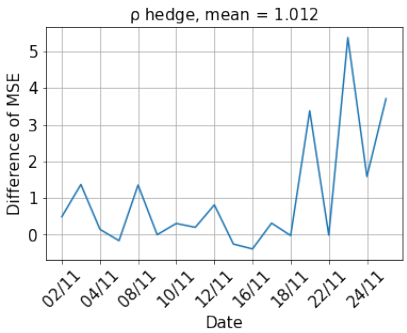

delta hedge, mean = 38.470
Difference of MSE
Date

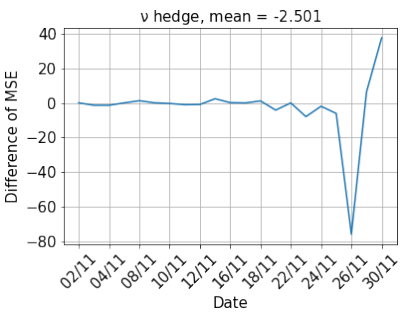

delta hedge, mean = 8.603
Difference of MSE
Date

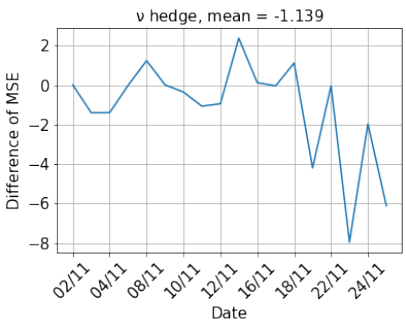

H hedge, mean = 5.583
Difference of MSE
Date

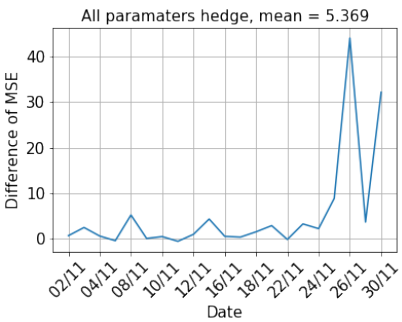

H hedge, mean = 2.592
Difference of MSE
Date

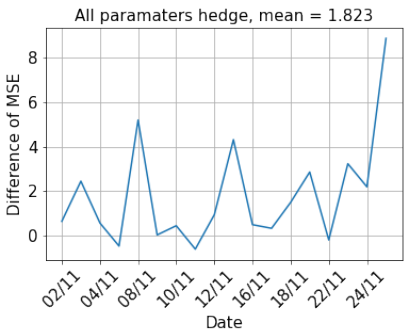

ρ hedge, mean = 3.676
Difference of MSE
Date


ρ hedge, mean = 1.012
Difference of MSE
Date


ν hedge, mean = -2.501
Difference of MSE
Date


ν hedge, mean = -1.139
Difference of MSE
Date


All paramaters hedge, mean = 5.369
Difference of MSE
Date


All paramaters hedge, mean = 1.823
Difference of MSE
Date

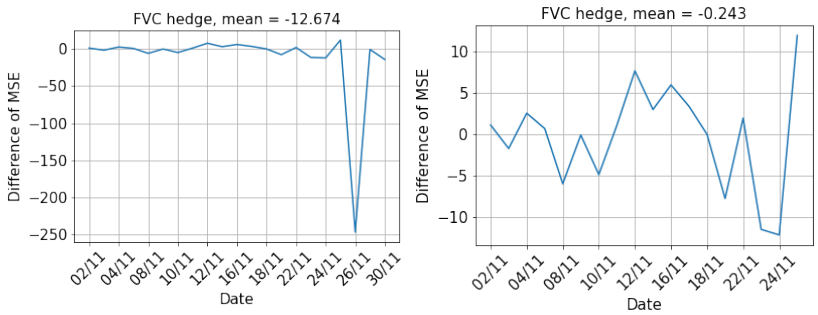


Figure 26: Time series of delta hedged weighted MSE and the difference between the hedged result MSE and the delta hedged MSE.

8.4.4. Lifted Heston

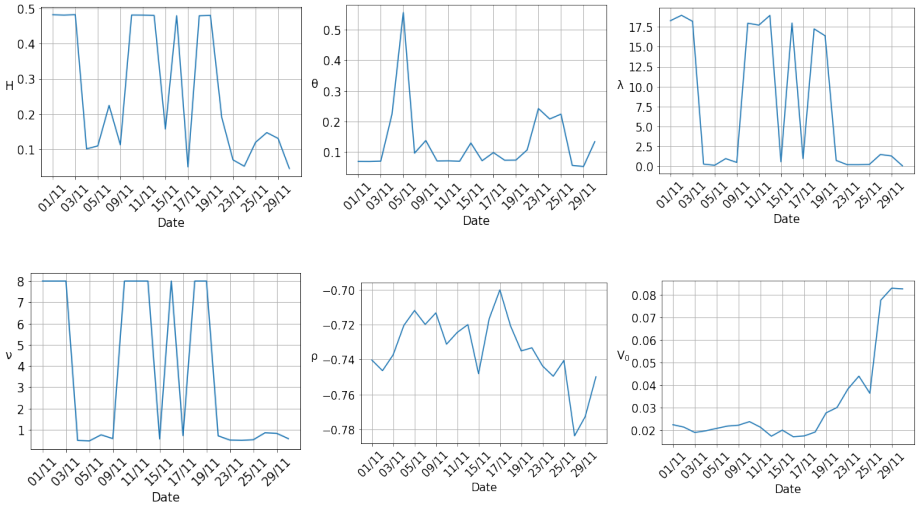
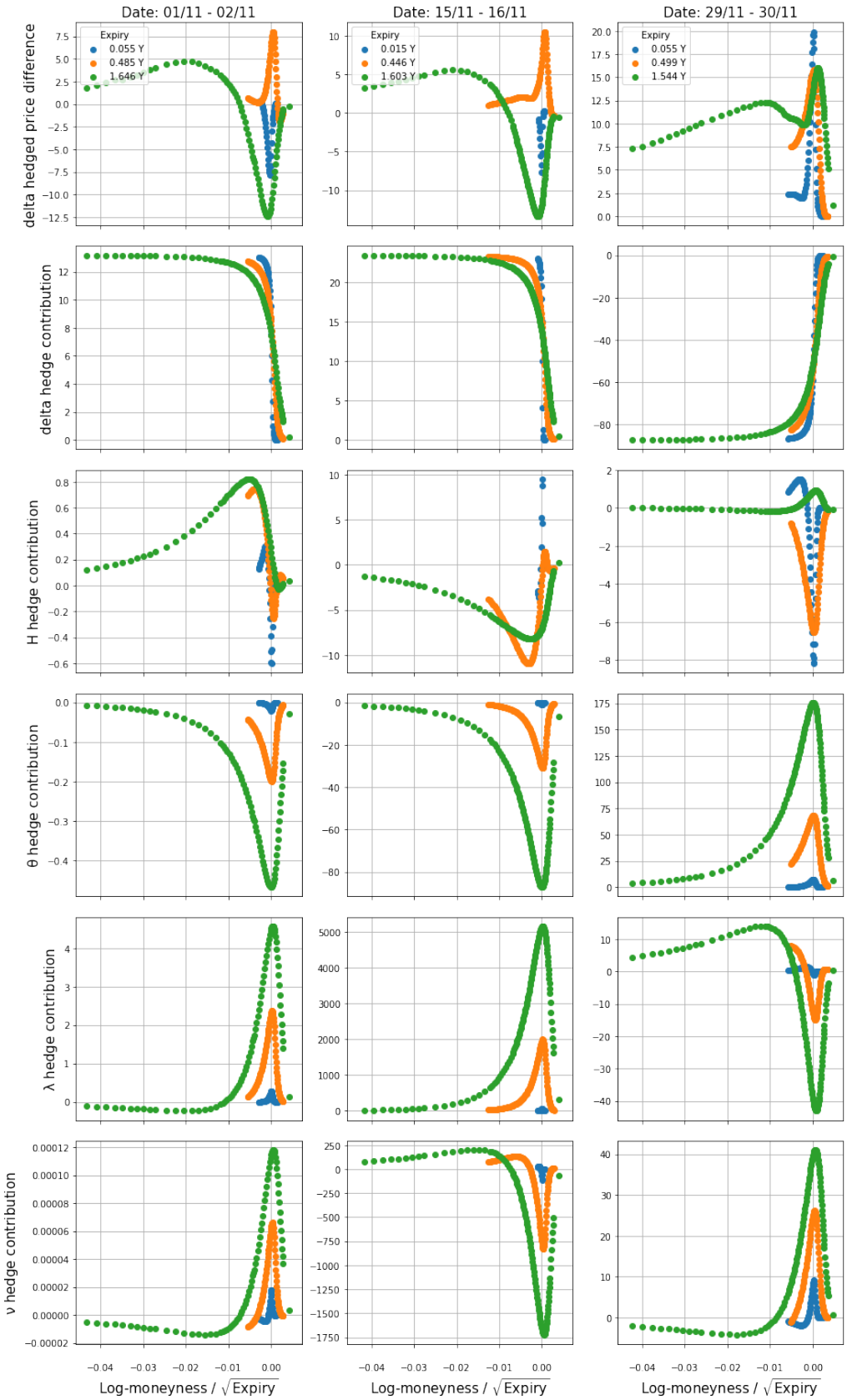


Figure 27: Parameter evolution for lifted Heston


Date: 01/11 - 02/11
Date: 15/11 - 16/11
Date: 29/11 - 30/11
Expiry
0.055 Y
0.485 Y
1.646 Y
Expiry
0.015 Y
0.446 Y
1.603 Y
Expiry
0.055 Y
0.499 Y
1.544 Y
delta hedged price difference
delta hedge contribution
H hedge contribution
θ hedge contribution
λ hedge contribution
ν hedge contribution
Log-moneyness / √Expiry
Log-moneyness / √Expiry
Log-moneyness / √Expiry

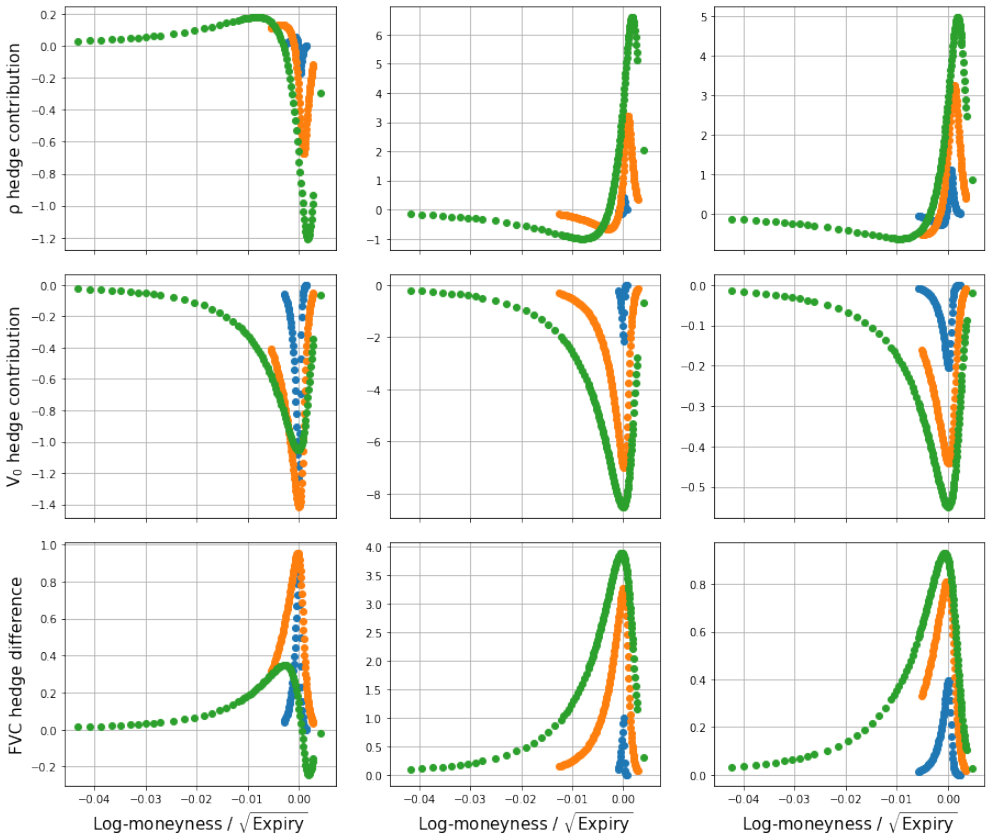
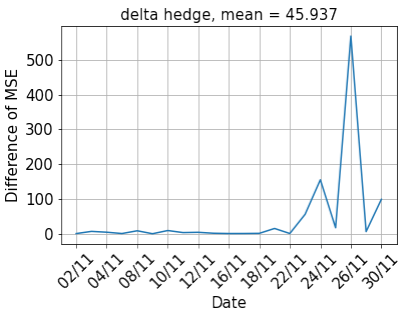
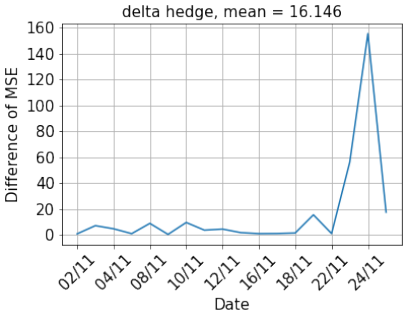
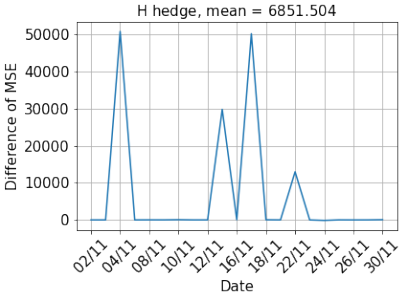
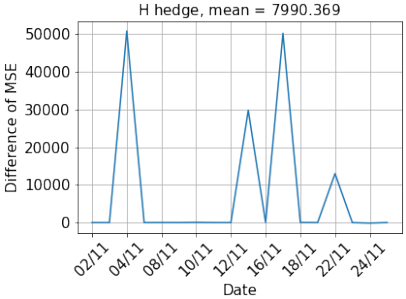


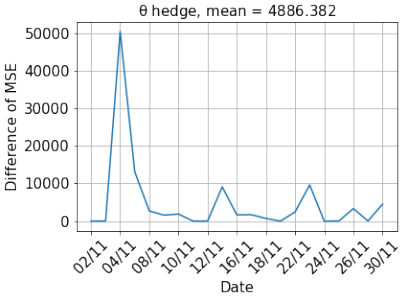
Figure 27: Hedging results of hedging against the parameters of the lifted Heston model

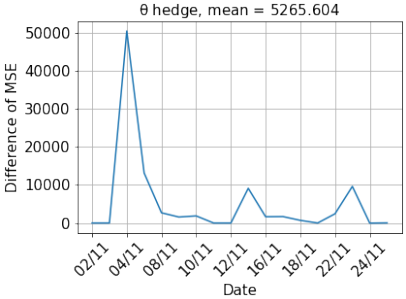

delta hedge, mean = 45.937
Difference of MSE
Date

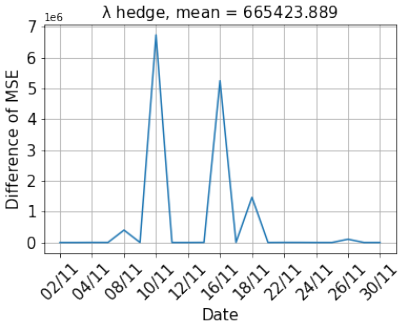

delta hedge, mean = 16.146
Difference of MSE
Date

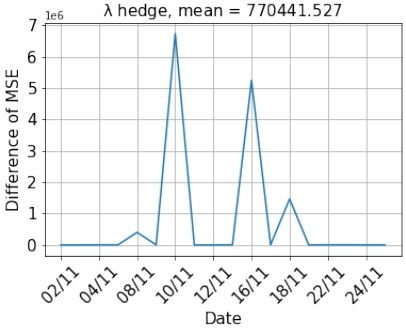

H hedge, mean = 6851.504
Difference of MSE
Date

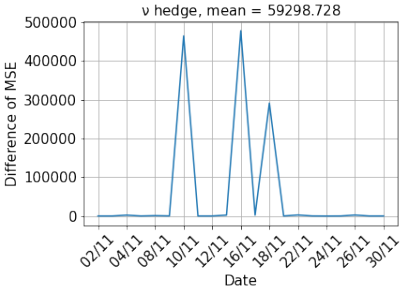

H hedge, mean = 7990.369
Difference of MSE
Date

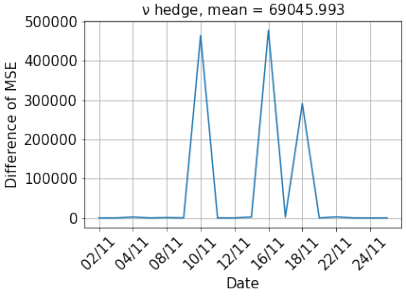

θ hedge, mean = 4886.382
Difference of MSE
Date


θ hedge, mean = 5265.604
Difference of MSE
Date


λ hedge, mean = 665423.889
1e6
Difference of MSE
Date


λ hedge, mean = 770441.527
1e6
Difference of MSE
Date


ν hedge, mean = 59298.728
Difference of MSE
Date


ν hedge, mean = 69045.993
Difference of MSE
Date

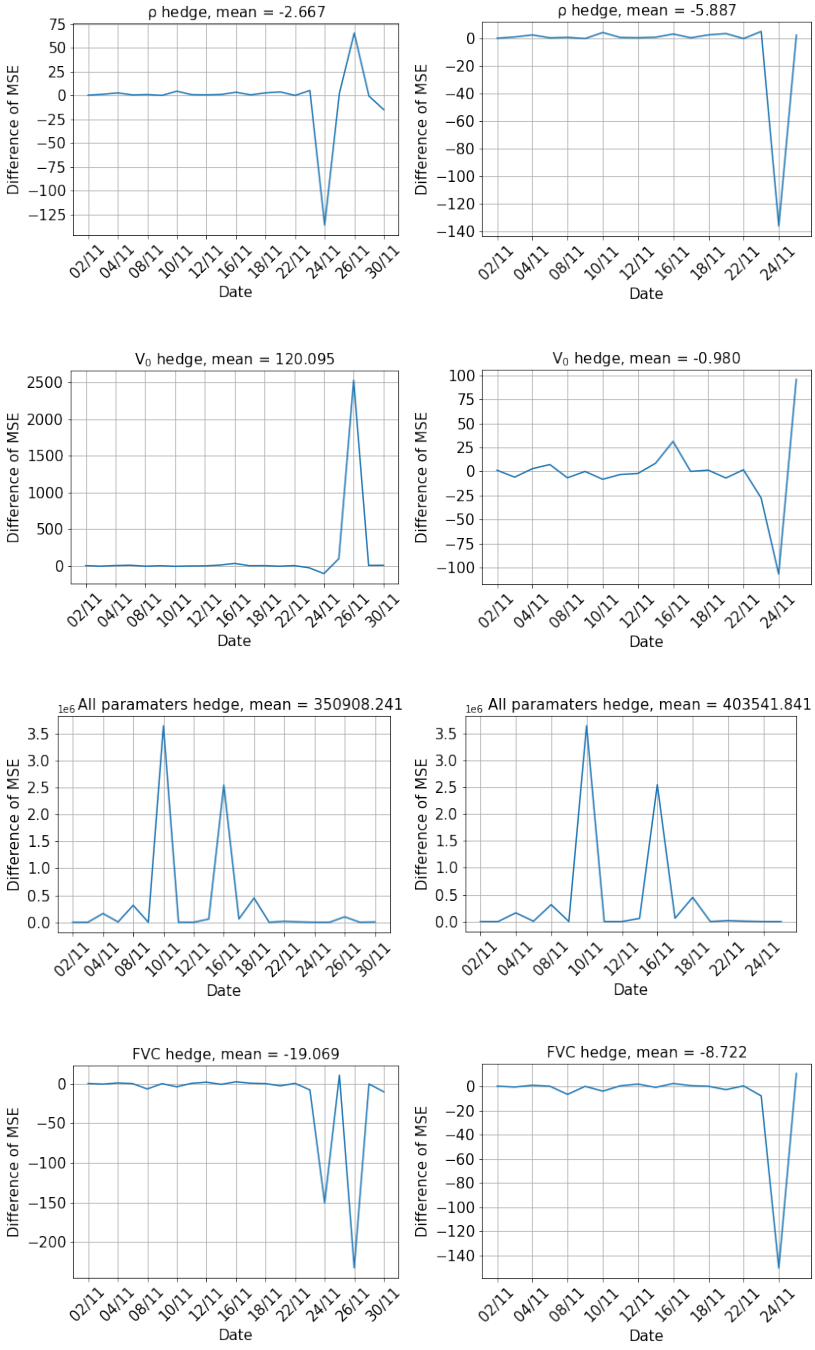


Figure 28: Time series of delta hedged weighted MSE and the difference between the hedged result MSE and the delta hedged MSE.

9. Discussion

9.1. Computation Time

As can be seen from Table 1 and Figure 7, the computation times for a single evaluation differ drastically. The rough Bergomi model seems to do the best, followed closely by the rough Bergomi 7 parameter model. Both have a computation time of a few seconds for the full vol surface. We do indeed see that the rough Heston has a very large computation time, which is to be expected and that the lifted Heston Markovian approximation cuts down the computation time of a single surface by a factor of five compared to rough Heston.

We observe that the three parameter models have an easier time converging, which is also to be expected. In the rough Bergomi case this leads to a reasonably short time for convergence, while the rough Heston model still takes a long time due to the nature of the computation. We also see that the lifted Heston model has a wide range of function evaluations needed and that it might take a long time to converge. This convergence mainly comes from the fact that we need to optimize six parameters instead of three, which is more difficult and thus requires more function evaluations.

9.2. Volatility Surface

Firstly note that Figure 8 and Table 2 tell us that the rough Bergomi 7 parameter, rough Heston and lifted Heston model perform roughly equal in how well they fit the market curve. Meanwhile the rough Bergomi model has trouble with the fit and has an error of roughly three times the size.

From Figures 9 and 10 we note that the rough Bergomi gives a good approximation of the middle expiry, but fails on the earlier and later expiries. We also note that since this is a path simulation, it is often difficult to correctly encompass the heavy tail of the distribution on the lower strikes, this becomes even more clear for the long expiry where this is much more prevalent.

The extra parameters for the instantaneous forward variance curve in the rough Bergomi 7 parameter model give more room to fit to the vol surface, as can be seen in Figures 11 and 12. This ensures a good fit around the skew. We again have the heavy tail problem for the lower strikes, however it does seem to be less than in the rough Bergomi model. Although the skew gives a good fit, there do seem to be some inconsistencies with the smile. One could possibly fix this by giving more weight to the smile. When looking at Figure 13, it can be seen that the FVC roughly follows the shape of the market FVC, but is always above it. We do note that in some scenarios it differs by more than 5 vol points and it is strongly driven by the first maturity, as can be seen in the final figure.

Looking at Figures 14 and 15, the rough Heston model seems to do a very good job at modeling the whole surface. Although the fit for 15/11 would have probably been better had the surface been shifted upwards, this is not possible because of the fixed FVC. The model seems to do a good job in the low vol and the high vol regime and throughout all expiries, especially considering it only has three parameters.

Finally, considering Figures 16 and 17 we note that the fit for the lifted Heston model is good. Although for the short expiries, the skew does not follow the same trajectory as the market but instead is more concave. We also see that the model does well in lower and higher vol regimes and does particularly well around the ATM skew for all expiries, which is the most liquid region. Considering Figure 18 we note that the lifted Heston model does well at approximating the forward variance curve, although the predicted FVC is consistently lower than the market FVC. Despite the outlier of the last expiry in the FVC, the predicted FVC seems to have a lot of flexibility and seems to always have roughly the same shape as the one from the market. We observe that for 30/11 the curve does seem strange, where the model predicts a constant curve which looks like a linear regression of the market curve.

9.3. Hedging

9.3.1. Rough Bergomi

The first thing that should be noted from Figure 19 is that $\rho \approx -1$ for every minimization. This seems strange and indicates an under-parameterized model. This also explains why the hedging for ρ in Figure 20 is either zero or very small.

Looking further at the contribution of the vol-of-vol parameter η, we can try to make sense of the curve. An increase in the vol-of-vol parameter will widen the distribution of the vol, thus decreasing the ATM vol and increasing the tails, as explained in section 2.3 of [34]. If we multiply this by a vega curve (shaped like a normal distribution) we obtain something similar; note the change in sign here. On all these dates there are positive differences in the η parameter, which leads to a positive contribution to the expected price around the ATM. Now taking the realized price and subtracting the expected price will lead to a negative curve, which is in line with Figure 20. Thus the sign of these curves also makes sense. Moving on to the Hurst parameter H we note that $\mathbb{E}[W_H(t)^2] = t^{2H}$, thus an increase in the Hurst parameter will decrease the variance of the short term vol (if expiry $< 1Y$), while increasing the variance of the long term vol (if expiry $> 1Y$). Since we apply this to the exponential martingale we obtain a correction to this change in variance by expiry. Thus the effect of an increase in H is the same as an increase in η. This seems to be the case when looking at Figure 20, except for on 30/11, which might be because the model is under-parameterized. Finally looking at the FVC hedge, an increase in the FVC means and increase in vol and thus also in price. It also has the shape of a vega curve, which is to be expected.

Finally we discuss Figure 21. The hedge against H seems to do consistently worse than delta only hedging. The ρ hedge often does not give an improvement, this is when $\rho = -1$ for those days. Even on the other days the improvement is not substantial. The η parameter does give an improvement over the only delta hedge, this is especially notable during a big decrease in spot price, while it gives a roughly equal result on the other days. The hedge improvement can go up to half of the delta hedge. Finally the FVC hedge seems to consistently outperform the delta-only hedge and can improve this hedge up to roughly 20%.

9.3.2. Rough Bergomi 7 parameters

Note that we do not have $\rho = -1$ consistently as in the rough Bergomi model. However, from Figure 22 it is apparent that the parameters $H, \beta_0, \beta_1, \beta_2$ and τ are not stable over time, indicating that our chosen model for the FVC is not representative of the market.

Moving on to Figure 22 we see comparable results for hedging against η. H also follows a similar reasoning as above and follows the same shape and sign as η, which is in line with almost all dates. In the first date we see a flip in sign, however the contribution for the H hedge is small here. We also observe that indeed trying to hedge against any of the parameters of the FVC does not yield and improved hedge, it seems like the Nelson-Siegel function for the FVC is over-parameterized. Finally hedging against the FVC also does not do a good job as the difference between the FVC's can be quite large, as can be seen in Figure 13.

Moving on to Figure 23, we see that none of the hedges outperform the delta hedge consistently. This indicates that the model does not describe reality, which is what we would expect since the FVC is probably not realistic. What is interesting to note is that the delta-hedge seems to be roughly a factor two better in this model than in any of the other models.

9.3.3. Rough Heston

Looking at Figure 24 we see lower values of H than expected. Where we would expect $H \approx 0.1$, all values of H are all under 0.08 and some even zero.

The H hedge contribution seems to have a different shape than the rough Bergomi models. This is because in the Heston models we do not get a correction for the change in vol-of-vol over the expiry. Hence we clearly see in Figure 25 that an increase in the Hurst parameter will decrease the variance of the short term vol (if expiry $< 1Y$), while increasing the variance of the long term vol (if expiry $> 1Y$). Thus we get a flip in sign depending on the expiry, while the shape is the same as the vol-of-vol (ν) hedge. Also, the ν parameter

seems in line with what was discussed earlier, however, FVC hedge is not smooth in this model, which is strange. Finally, increasing the ρ parameter tilts the vol curve around the ATM, as can be seen in Figure 2.14 of [34]. However the tilt has a higher effect on the smile part of the curve, hence multiplying this by vega we get a curve which has the shape of the one in Figure 25. Finally considering Figure 26, we observe that hedging against η leads to improvements, although not consistently, and hedging against the FVC seems to do a good job and can again correct up to one-third of the delta hedge during the high vol days.

9.3.4. Lifted Heston

The first thing to note is that Figure 27 shows a large variance in the H, θ, λ and ν parameters. This indicates that it is going to be less meaningful to hedge against these parameters. This is then also supported by the very high peaks for these parameters in Figure 28. Not only do they have a very large variance, they also seem to be correlated. The correlation seems to be positive for H, λ and ν and negative for θ. This is can also be seen when plotting the time series of the parameters for the minimization

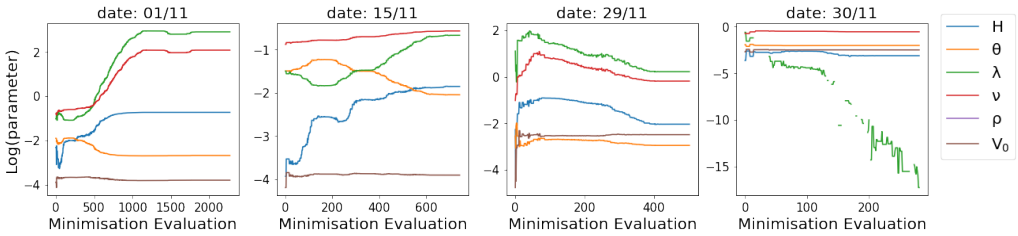


Figure 29: Time series of the evaluations in the differential evolution minimization for lifted Heston.

For all but the last date we see that indeed the parameters seem correlated. This means they are probably stuck in a valley in the optimization surface. This can then take a long time to minimize, as seen on $01/11$ in Figure 29. It also means they lose an economic interpretation. Another interesting thing to note from this figure is that the minimization terminates relatively quickly on $30/11$, when λ seems to go to zero. This supports the claims in [10] and [9] that $\lambda \to 0$ is a reasonable assumption and that with this assumption the parameters are not correlated, since the minimization does not seem to be stuck in a valley. Therefore we can indeed include the FVC as an input to the model.

Moving on to 27 we can explain the shape of the θ, λ and V_0 curves. θ is the long term vol, which mainly has a big contribution on the far expiries and is shaped like vega. λ is the speed of mean reversion, when this increases we get less vol-of-vol (ν), this is in line with expectations for the two later dates, however is not the case for the $1/11$-$2/11$ hedge, although the values for the ν hedge are very small here. Finally an increase in the base vol V_0 leads to a vega-shaped curve, which is in line with the figure.

When looking at the H, θ, λ and ν hedges we see that they are poor, this is in line with what we expected from Figure 28. Even V_0 suffers from this phenomenon from the 25^{th} to the 26^{th}. We can however note that hedging against the generated FVC seems to do well throughout the month, especially when the underlying drops. This makes sense considering the delta hedge peaks around this time. Hedging against the FVC can improve the delta hedge by roughly 40%.

10. Conclusion

In conclusion we find that the rough Bergomi is an under-parameterized model, however does a reasonable job at fitting the vol surface. In an attempt to give more degrees of freedom to the model we introduced the Nelson-Siegel function for the forward variance curve, this improves the fit. Interesting to note is that the delta hedges from the Bergomi models outperform the delta hedges from the Heston models.

On the other hand, the rough Heston model seems to do a good job at approximating the vol surface with only three parameters. It does come with the drawback that it takes a long time to compute the surface through the characteristic function with the COS method. One could try to run a Monte Carlo simulation similar to that of the rough Bergomi model on the rough Heston model in an attempt at speeding up pricing.

Finally the lifted Heston model does improve computation time considerably from the rough Heston model, however is still a lot slower than the rough Bergomi models. Also, the parameters for the forward variance curve seem to be correlated, which can lead to a lot of function evaluations in the minimization and also means a loss in economic interpretation.

In general the process could be sped up by improving the parameters of the minimization, but this was beyond the scope of the project. Besides that a major improvement can be made by parallelizing the computation of the vol curves, since our implementation depends on a for-loop. This could speed up the process, however since the later expiries have the hardest time converging this speed up will be less than a factor of 12.

If one were to apply these methods in practice, one would have to choose a model depending on the goal. For an accurate representation of the vol smile one could stick with the rough Bergomi 7 parameter model, possibly choose a different forward variance approximation. One could also consider this model if the goal is to obtain an accurate delta hedge. On the other hand if the goal is to hedge against the forward variance curve one you consider the rough Heston model or even consider using the Markovian approximation together with the market forward variance curve in order to improve fitting time. Finally if the goal is to have a model which simulates the vol surface as well as the forward variance curve accurately, the lifted Heston model should be considered.

In further research one could try to run the Heston models using a path simulation. Since the difference in computation time of a single vol surface is at least a factor of seven between the path simulated models and the characteristic function model. This might improve computation time.

One could also try to find a consistent parameterization of the forward variance curve in the rough Bergomi model. As seen, this model is under-parameterized, however does a good job at fitting the market smile when given additional degrees of freedom for the instantaneous forward variance curve in the rough Bergomi 7 parameter model.

Finally one would of course expand on this work by fitting the vol surface on more dates and considering more models. Another expansion one could do is take into account the trading costs or compare the delta hedge of the rough volatility models to other models. Lastly, could also implement possible improvements to the models such as those mentioned in [6], [7] and [8].

Bibliography

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

[2] Lorenzo Bergomi. Smile dynamics ii. *Available at SSRN 1493302*, 2005.

[3] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[4] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative finance*, 18(6):933–949, 2018.

[5] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[6] Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen. Hybrid scheme for brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965, 2017.

[7] Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. Hierarchical adaptive sparse grids and quasi-monte carlo for option pricing under the rough bergomi model. *Quantitative Finance*, 20(9):1457–1473, 2020.

[8] Qinwen Zhu, Grégoire Loeper, Wen Chen, and Nicolas Langrené. Markovian approximation of the rough bergomi model for monte carlo option pricing. *Mathematics*, 9(5):528, 2021.

[9] Omar El Euch, Jim Gatheral, and Mathieu Rosenbaum. Roughening heston. *Risk*, pages 84–89, 2019.

[10] Omar El Euch and Mathieu Rosenbaum. Perfect hedging in rough heston models. *The Annals of Applied Probability*, 28(6):3813–3856, 2018.

[11] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[12] Eduardo Abi Jaber. Lifting the heston model. *Quantitative Finance*, 19(12):1995–2013, 2019.

[13] Giovanni Zanco. A brief introduction to rough paths, 2016.

[14] Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

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

[16] Hans Buehler. Consistent variance curve models. *Finance and Stochastics*, 10(2):178–203, 2006.

[17] Marc Romano and Nizar Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997.

[18] Fang Fang and Cornelis W Oosterlee. A novel pricing method for european options based on fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2009.

[19] Fabien Le Floc'h. More robust pricing of european options based on fourier cosine series expansions. *arXiv preprint arXiv:2005.13248*, 2020.

[20] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

[21] Gustaf Gripenberg, Stig-Olof Londen, and Olof Staffans. *Volterra integral and functional equations*. Number 34. Cambridge University Press, 1990.

[22] Gerhard Wanner and Ernst Hairer. *Solving ordinary differential equations II*, volume 375. Springer Berlin Heidelberg, 1996.

[23] Eduardo Abi Jaber and Omar El Euch. Markovian structure of the volterra heston model. *Statistics & Probability Letters*, 149:63–72, 2019.

[24] Philippe Carmona, Laure Coutin, and Gérard Montseny. Approximation of some gaussian processes. *Statistical inference for stochastic processes*, 3(1):161–171, 2000.

[25] Jim Gatheral. *The volatility surface: a practitioner's guide*. John Wiley & Sons, 2011.

[26] Lorenzo Bergomi. *Stochastic volatility modeling*. CRC press, 2015.

[27] Peter Tankov. *Financial modelling with jump processes*. Chapman and Hall/CRC, 2003.

[28] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[29] Kenneth Price, Rainer M Storn, and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[30] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[31] Fuchang Gao and Lixing Han. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1):259–277, 2012.

[32] Roger W Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 14(3):469–480, 2004.

[33] Blanka Horvath, Antoine Jack Jacquier, and Aitor Muguruza. Functional central limit theorems for rough volatility. *Available at SSRN 3078743*, 2017.

[34] Joerg Kienitz and Daniel Wetterau. *Financial modelling: Theory, implementation and practice with MATLAB source*. John Wiley & Sons, 2013.

A. COS Method Derivations

In order to approximate (4.2) on an arbitrary interval $[a, b]$ we use the change of variables

$$\theta \coloneqq \frac{x-a}{b-a}\pi, \quad x \coloneqq \frac{b-a}{\pi}\theta + a.$$

We can then determine the coefficients A_k through

$$\begin{aligned}
A_k &= \frac{2}{b-a}\int_a^b f(x)\cos\left(k\pi\frac{x-a}{b-a}\right)dx \\
&= \frac{2}{b-a}\int_a^b \Re\left\{f(x)e^{ik\pi\frac{x-a}{b-a}}\right\}dx \\
&= \frac{2}{b-a}\Re\left\{\int_a^b f(x)e^{ik\pi\frac{x}{b-a}}dx e^{ik\pi\frac{-a}{b-a}}\right\} \\
&\approx \frac{2}{b-a}\Re\left\{\phi\left(\frac{k\pi}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\},
\end{aligned}$$

where the approximation sign is because of the truncation of the integral.

For pricing a product with pay-off $V(S_T, T)$ at time $t < T$, where T is the maturity and S_T the price of the underlying at time T, we define

$$A_k(x) \coloneqq \frac{2}{b-a}\int_a^b f(y|S_t)\cos\left(k\pi\frac{y-a}{b-a}\right)dy,$$

where $f(y|S_t)$ is the conditional probability density function from (4.1) given the price of the underlying S_t at time t. Computing the pay-off from (4.5) is done in the following way

$$\begin{aligned}
e^{r(T-t)}V(S_t,t) &= \int_{\mathbb{R}} V(y,T)f(y|S_t)dy \\
&\approx \int_a^b V(y,T)f(y|S_t)dy \\
&= \int_a^b V(y,T)\sum_{k=0}^{\infty}{}'A_k(x)\cos\left(k\pi\frac{y-a}{b-a}\right)dy \\
&\approx \int_a^b V(y,T)\sum_{k=0}^{N-1}{}'A_k(x)\cos\left(k\pi\frac{y-a}{b-a}\right)dy \\
&= \frac{b-a}{2}\sum_{k=0}^{N-1}{}'A_k(x)V_k \\
&\approx \frac{b-a}{2}\sum_{k=0}^{N-1}{}'\Re\left\{\phi\left(\frac{k\pi}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}V_k,
\end{aligned}$$

where

$$V_k \coloneqq \frac{2}{b-a}\int_a^b V(y,T)\cos\left(k\pi\frac{y-a}{b-a}\right)dy.$$

Finally computing the coefficients V_k^{Put} in (4.6) is done by

$$\begin{aligned}
V_k^{\text{Put}} &= \frac{2}{b-a}\int_a^b F\left|\frac{K}{F}-e^y\right|^+\cos\left(k\pi\frac{y-a}{b-a}\right)dy \\
&= \frac{2F}{b-a}\int_a^z (e^z-e^y)\cos\left(k\pi\frac{y-a}{b-a}\right)dy \\
&= \frac{2F}{b-a}\left(-\chi_k(a,z)+e^z\psi_k(a,z)\right) \\
&= \frac{2}{b-a}\left(-F\chi_k(a,z)+K\psi_k(a,z)\right),
\end{aligned}$$

where $z = \log \frac{K}{F}$, $y = \log \frac{S_T}{F}$ and χ and ψ are given in (4.7).