

Bayesian framework to quantify rough volatility calibration risk

Alessandro Nodari

A thesis presented for the
Master's degree in Mathematics



Università degli Studi di Milano
Dipartimento di Matematica
19 July 2023

Acknowledgement

Index

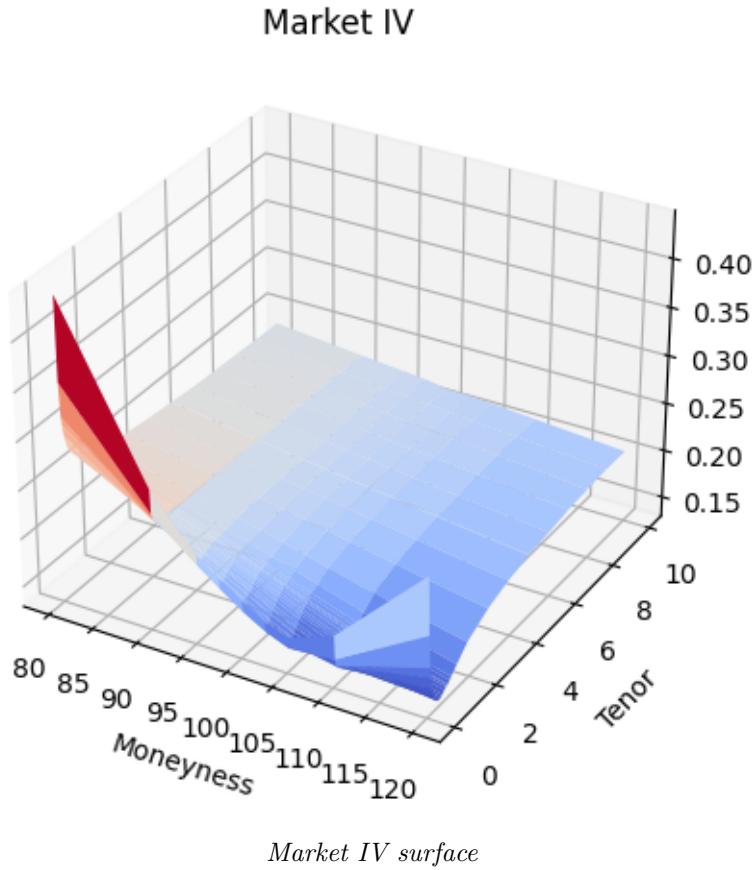
1	Introduction	4
2	Dataset	5
2.1	Implied drift term	5
2.2	Forward Variance Curve	7
3	Fractional Brownian Motion	10
3.1	Definition and Properties	10
3.2	Simulation Methods	11
3.2.1	Cholesky Decomposition	11
3.2.2	Hybrid Scheme	12
3.2.3	Simulation results	15
4	Rough Bergomi Model	18
4.1	The Model	18
4.1.1	N-Factor Model	18
4.1.2	One-Factor Model	19
4.1.3	2-Factors Model	20
4.2	The Realized Variance	21
4.3	The probability measure change	21
4.4	Pricing	23
4.5	Calibration	24
4.5.1	Objective Function	24
4.5.2	Numerical results	25
4.6	Volatility Skew	26
4.6.1	Numerical results	27
4.7	Forward-Start Options	29
4.7.1	Implied Volatility	30
4.8	Parameters stability	31
4.9	Bayesian Inverse Problem	34
4.9.1	Sliced Wasserstein distance	35
4.9.2	Numerical results	35
5	Quintic Ornstein-Uhlenbeck model	39
5.1	The Model	39
5.2	SPX derivatives	40
5.2.1	Numerical results	41
5.3	VIX derivatives	42
5.3.1	Numerical results	44
5.4	Joint calibration	46
5.4.1	Numerical results	47
5.5	SPX Parameters stability	47
5.6	Bayesian Inverse Problem	51

5.6.1	Numerical results	51
6	Other models	56
6.1	Heston	56
6.2	rHeston	60
7	Bayesian Risk Quantification	66
7.1	rBergomi	66
7.2	Quintic OU	69
7.3	Heston	70
7.4	rHeston	71
8	Conclusion	74

1 Introduction

2 Dataset

As our dataset we decided to use the SPX options data of the 23rd January 2023 (source: Bloomberg). We took into consideration options with moneyness from 80% to 120% and tenor from 2 weeks to almost 10 years, they are 32 in total. The spot reference for that day is 4019.81.



2.1 Implied drift term

The options on the SPX index are very liquid and rich in structure. Whenever we transform our model equations into the risk-free world to price some contingent claims we have to estimate the drift term which is defined as the difference between the risk-free rate and the yield (continuously compounded) of the underlying. In order to estimate this term we recall the put-call parity. We will employ the following notation:

- $C_{t,K}$ is the value of a call at time t with strike K and tenor T ;

- $P_{t,K}$ is the value of a put at time t with strike K and tenor T ;
- S_t is the spot value of the underlying;
- r is the risk-free rate (continuously compounded);
- q is the yield of the underlying (continuously compounded).

Then the following hold:

$$C_{t,K} - P_{t,K} = S_t e^{-q(T-t)} - K e^{-r(T-t)}$$

solving for q we obtain:

$$q = \frac{1}{T-t} \log \left(\frac{C_{t,K} - P_{t,K} + K e^{-r(T-t)}}{S_t} \right)$$

The above relation holds true also for $t = 0$. At time 0 we are able to observe:

- S_0 that is the spot market price of the underlying;
- $\hat{C}_{0,K}$ which is the call market-price;
- $\hat{P}_{0,K}$ which is the put market-price.

Thus, at time 0, the relation simplifies to:

$$q = -\frac{1}{T} \log \left(\frac{\hat{C}_{0,K} - \hat{P}_{0,K} + K e^{-rT}}{S_0} \right) \quad (2.1)$$

One could argue that q can depend on the strike, but this is false. Indeed, let's suppose that it is true and consider two strikes $K_1 < K_2$ with $q_{K_1} > q_{K_2}$. If at time $t = 0$ we buy the following portfolio:

$$\frac{1}{S_0} (\hat{C}_{0,K_1} - \hat{P}_{0,K_1} + K_1 e^{-rT}) - \frac{1}{S_0} (\hat{C}_{0,K_2} - \hat{P}_{0,K_2} + K_2 e^{-rT})$$

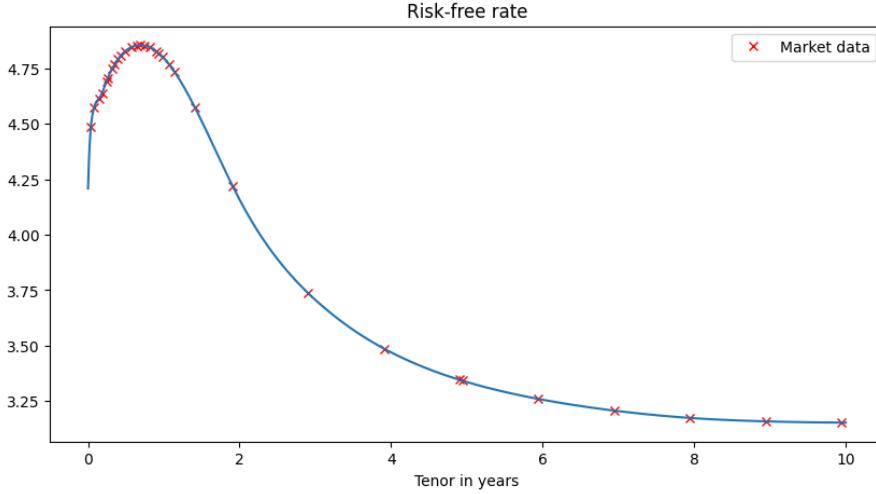
due to the put-call parity this gives us money. In fact, at time 0 its value is given by:

$$\begin{aligned} \frac{1}{S_0} (\hat{C}_{0,K_1} - \hat{P}_{0,K_1} + K_1 e^{-rT}) - \frac{1}{S_0} (\hat{C}_{0,K_2} - \hat{P}_{0,K_2} + K_2 e^{-rT}) \\ = e^{-q_{K_1} T} - e^{-q_{K_2} T} < 0 \end{aligned}$$

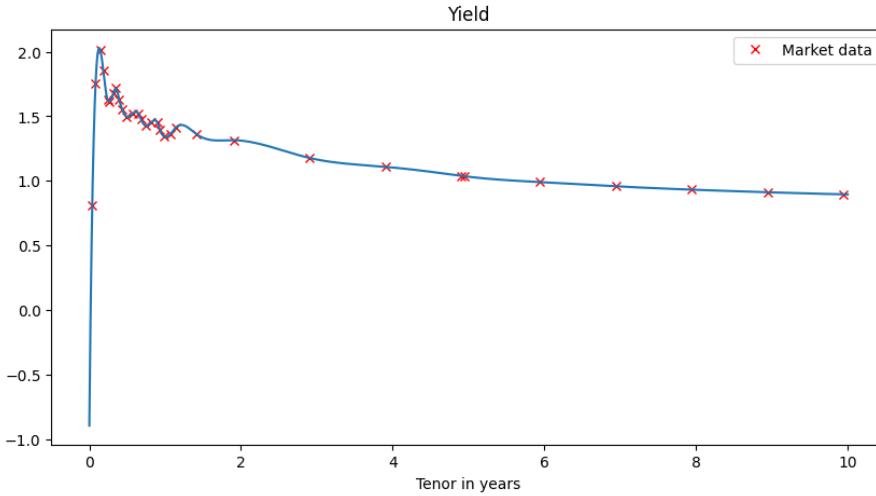
At maturity the value of the portfolio is 0:

$$C_{T,K_1} - P_{T,K_1} + K_1 = S_T = C_{T,K_2} - P_{T,K_2} + K_2$$

Thus, this strategy is an arbitrage. The same, with obvious modifications, applies if $q_{K_1} < q_{K_2}$. So, in conclusion, the term q depends only on the tenor T and not on the strike K . To estimate the risk-free rate we used the values of the zero coupon USD OIS Swap rates, interpolated with cubic splines for the missing tenors. The resulting curve that we obtained is represented in the next figure.



Once we have the risk-free rate curve we can infer, using formula (2.1), the value of the yield. The next figure represents the yield curve that we obtained.



2.2 Forward Variance Curve

The forward variance curve is associated with the fair strike of a variance swap (VS). A VS with maturity T is a contract which pays out the realized variance of a financial underlying, computed as the sum of the squares of daily log-returns, in exchange for a fixed strike V_0^T called the variance swap variance that is determined in such a way that the initial value of the contract is zero. The market, instead of quoting the rate V_0^T of a VS, quotes its volatility which is the strike K such that:

$$\frac{V_0^T}{T} - K^2 = 0$$

Therefore we define the volatility of a VS with maturity T as:

$$\hat{\sigma}_0^T := \sqrt{\frac{V_0^T}{T}}$$

This relates to the forward variance curve since we have:

$$\hat{\sigma}_t^2(T) = \frac{1}{T-t} \int_t^T \xi_t(u) du$$

or equivalently:

$$\xi_t(T) = \frac{d}{dT} [(T-t)\hat{\sigma}_t^2(T)]$$

And for the initial forward variance curve, that is when we consider $t = 0$, it simplifies to:

$$\xi_0(t) = \frac{d}{dt} [t\hat{\sigma}_0^2(t)]$$

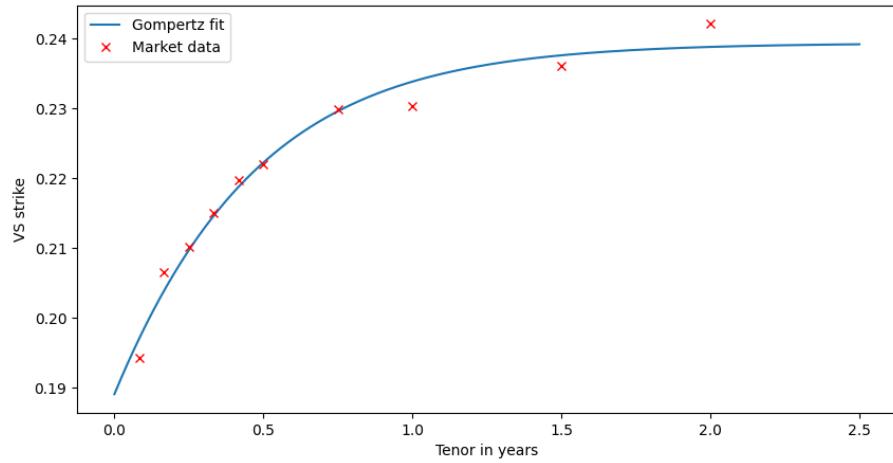
In order to compute the initial forward variance curve $\xi_0(t)$ we have to find a parametrization for the volatility of the VS. It is reasonable to choose a parameterization which has an asymptotic line, our choice is the Gompertz function:

$$\hat{\sigma}_0^2(t) = z_1 e^{-z_2 e^{-z_3 t}}$$

where $z_1 > 0$ is the asymptote, $z_2 > 0$ sets the displacement along the x -axis, that is the time to maturity, and $z_3 > 0$ sets the growth rate. We used the least squares method to fit the parameters and we obtained:

$$z_1 = 0.2393444556 \quad z_2 = 0.2355916740 \quad z_3 = 2.3126258447$$

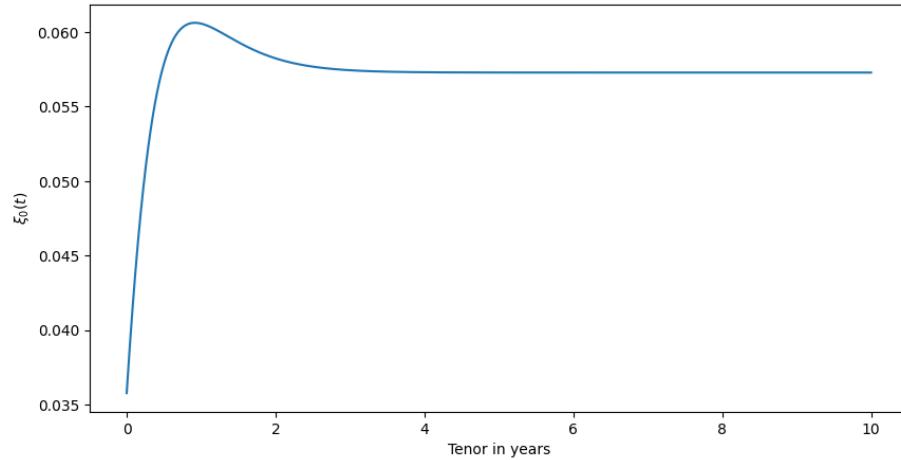
and the next figure is the resulting fit.



The initial forward variance curve $\xi_0(t)$ can be obtained as:

$$\xi_0(t) = \hat{\sigma}_0^2(t) + t \frac{d}{dt} [\hat{\sigma}_0^2(t)]$$

and since we have parametrized $\hat{\sigma}_0^2(t)$ we can easily compute its derivative. Doing so we obtained the following figure:



3 Fractional Brownian Motion

The fractional Brownian Motion (fBM) was first introduced within a Hilbert space framework by Kolmogorov in 1940 in [1], where it was called Wiener Helix. It was further studied by Yaglom in [2]. The name fractional Brownian Motion is due to Mandelbrot and Van Ness, who in 1968 provided in [3] a stochastic integral representation of this process in terms of a standard Brownian motion. From now on we will consider a probability space $(\Omega, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, where \mathbb{F} is the natural filtration generated by a Brownian Motion.

3.1 Definition and Properties

We will use the definition of fBM given in [4].

Definition 1. A fractional Brownian Motion W_t^H of Hurst index $H \in (0, 1)$ is a continuous and centered Gaussian process with covariance function

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

By **Definition 1** a fBM W_t^H has the following properties:

1. $W_0^H = 0$;
2. $W_t^H \sim \mathcal{N}(0, t^{2H})$, $\forall t \geq 0$;
3. W^H has stationary increments:

$$W_{t+s}^H - W_t^H = W_s^H \quad s \in \mathbb{R}$$

4. W_t^H has \mathbb{P} a.s. continuous trajectories.

Remark 1. The fBm is divided into three very different families:

- $0 < H < \frac{1}{2}$ where two increments of the form $(W_{t+h}^H - W_t^H)$ and $(W_{t+2h}^H - W_{t+h}^H)$ are negatively correlated;
- $H = \frac{1}{2}$ then the fBM is actually a standard Brownian Motion and the increments are independent;
- $\frac{1}{2} < H < 1$ where two increments of the form $(W_{t+h}^H - W_t^H)$ and $(W_{t+2h}^H - W_{t+h}^H)$ are positively correlated.

Proposition 1. A fBM W_t^H admits the following stochastic integral representation:

$$W_t^H = C_H \left(\int_{-\infty}^0 [(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] dW_s + \int_0^t (t-s)^{H-\frac{1}{2}} dW_s \right)$$

where

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

Where Γ is the usual gamma function and W_t is a two-sided Brownian Motion defined on \mathbb{R} as:

$$W_t = \begin{cases} W^1(t) & t \geq 0 \\ W^2(-t) & t < 0 \end{cases}$$

where W^1 and W^2 are two independent Brownian Motion.

Proposition 2. *A fBM W_t^H with Hurst parameter $H \in (0, 1)$ is a self-similar process such that, for any $c \geq 0$, it holds:*

$$W_{ct}^H \stackrel{d}{=} c^H W_t^H$$

Proposition 3. *The paths of a fBM W_t^H with Hurst parameter $H \in (0, 1)$ are almost surely locally $(H - \varepsilon)$ -Hölder continuous for $\varepsilon \in (0, H)$.*

Proposition 4. *A fBM W_t^H has a monofractal scaling property:*

$$\mathbb{E}[|W_{t+\Delta}^H - W_t^H|^q] = \mathbb{E}[|W_\Delta^H|^q] = K_q \Delta^{Hq}$$

where

$$K_q = \int_{-\infty}^{\infty} |x|^q \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

We also highlight that a fBM is not a Markov process nor a semi-martingale.

3.2 Simulation Methods

A great number of methods have been developed to simulate the paths of a fBM. Some of them are exact methods, which are more demanding from a computational standpoint, and other are approximations. We will present an exact method, the Cholesky decomposition, and the hybrid scheme approximation. More methods are presented and analyzed in depth in [4].

3.2.1 Cholesky Decomposition

This method is based on the so-called Cholesky decomposition of the covariance matrix. We will analyze the case in which we will operate that is when we want to simulate the fBM W_t^H for $t \in [0, T]$. First we discretize the interval using an equi-spaced grid of $n + 1$ points $0 = t_0 < t_1 < \dots < t_n = T$ with time-step $h = \frac{T}{n}$. The covariance structure of our discretization is:

$$\mathbb{E}[W_{t_i}^H W_{t_j}^H] = \frac{1}{2}(t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H}) = \frac{h^{2H}}{2}(i^{2H} + j^{2H} - |i - j|^{2H})$$

The covariance matrix C is defined element wise as $C_{i,j} = \mathbb{E}[W_{t_i}^H W_{t_j}^H]$ for $i, j = 1, \dots, n$. C is a symmetric and positive semi-defined matrix in $\mathbb{R}^{n \times n}$.

$$C = \begin{bmatrix} \mathbb{E}[W_{t_1}^H W_{t_1}^H] & \cdots & \mathbb{E}[W_{t_1}^H W_{t_n}^H] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[W_{t_n}^H W_{t_1}^H] & \cdots & \mathbb{E}[W_{t_n}^H W_{t_n}^H] \end{bmatrix} \quad (3.2.1)$$

Hence the usual Cholesky decomposition reads:

$$C = LL^T$$

where L is a lower triangular matrix with real and positive diagonal entries. Now we will draw n independent samples Z_i from a normal distribution $\mathcal{N}(0, 1)$. The vector $(0, LZ)$ of size $n + 1$ yields a sample path of W_t^H . To summarize, the steps of the method are:

1. choose an equi-spaced grid $\{t_i\}_{i=1,\dots,n}$ for the interval $[0, T]$;
2. compute the covariance matrix as in (3.2.1);
3. use the Cholesky decomposition to find the matrix L such that $C = LL^T$;
4. construct a vector Z of n independent realization of a standard normal distribution $\mathcal{N}(0, 1)$;
5. compute the path of the fBM as the vector $(0, LZ)$.

We note that the complexity of this method is of the order $\mathcal{O}(n^3)$ which is quite demanding.

3.2.2 Hybrid Scheme

The following method is an approximation method that was proposed in [5] as a scheme to simulate a Brownian semi-stationary (BSS) process. The class of BSS processes are studied extensively in [6]. For our purpose we will define a BSS process $(Y_t)_{t \in \mathbb{R}}$ as:

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

where W_t is a two-sided Brownian motion, g is a deterministic non-negative weight function and σ is a so called càdlàg process. In order to use the hybrid scheme we have to assume that:

1. for some $\alpha \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ it holds:

$$g(x) = x^\alpha L_g(x) \quad x \in (0, 1]$$

where $L_g : (0, 1] \rightarrow [0, \infty)$ is continuously differentiable, slowly varying at 0 and bounded away from 0. Moreover, there exists a constant $d > 0$ such that:

$$|L_g(x)| \leq d(1 + x^{-1}) \quad \vee \quad |L'_g(x)| \leq d(1 + x^{-1})$$

where L'_g it's the derivative of L_g ;

2. the function g is continuously differentiable in $(0, \infty)$;
3. for some $\beta \in (-\infty, \frac{1}{2})$ it holds:

$$g(x) = \mathcal{O}(x^\beta)$$

Since we are only interested in the time intervals that start at 0 we will use a Truncated Brownian semi-stationary (TBSS) process X_t defined as:

$$X_t = \int_0^t g(t-s)\sigma(s)dW_s$$

We will use as the discretization grid $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{\lfloor nT \rfloor}{n}\}$ and we will assume that σ can be taken constant on each interval of the grid. Doing so the TBSS can be approximated as:

$$X_t \simeq \sum_{k=1}^{\lfloor nT \rfloor} \sigma\left(t - \frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k-1}{n}} g(t-s)dW_s =: X_n(t)$$

For small values of k , say $k \leq \kappa$ for a given κ , we can approximate g as in the assumptions:

$$g(t-s) \simeq (t-s)^\alpha L_g\left(\frac{k}{n}\right) \quad (t-s) \in \left[\frac{k-1}{n}, \frac{k}{n}\right]$$

For larger values of k , say $k > \kappa$, we can approximate g as:

$$g(t-s) \simeq g\left(\frac{b_k}{n}\right)$$

where in [6] the optimal b_k is shown to be:

$$b_k^* = \left(\frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha+1} \right)^{\frac{1}{\alpha}}$$

Thus, we have that our approximation of the TBSS is composed of two parts:

$$X_n(t) = X_n^1(t) + X_n^2(t)$$

where:

$$\begin{aligned} X_n^1(t) &:= \sum_{k=1}^{\kappa} L_g\left(\frac{k}{n}\right) \sigma\left(t - \frac{k}{n}\right) \int_{t-\frac{k}{n}}^{t-\frac{k-1}{n}} (t-s)^\alpha dW_s \\ X_n^2(t) &:= \sum_{k=\kappa+1}^{\lfloor nT \rfloor} g\left(\frac{b_k^*}{n}\right) \sigma\left(t - \frac{k}{n}\right) \left(W_{t-\frac{k-1}{n}} - W_{t-\frac{k}{n}}\right) \end{aligned}$$

This decomposition tells us that we have to simulate on the grid points $\{\frac{i}{n}\}$:

$$\begin{aligned} W_{i,j}^n\left(\frac{i}{n}\right) &= \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left(\frac{i+j}{n} - s\right)^\alpha dW_s \quad j = 1, \dots, \kappa \\ W_i^n\left(\frac{i}{n}\right) &= \int_{\frac{i}{n}}^{\frac{i+1}{n}} dW_s \end{aligned} \quad (3.2.2.1)$$

We will explicit some of the properties of these two processes:

$$\begin{aligned} \text{Var}[W_i^n] &= \frac{1}{n} \\ \text{Var}[W_{i,j}^n] &= \frac{j^{2\alpha+1} - (j-1)^{2\alpha+1}}{(2\alpha+1)n^{2\alpha+1}} \\ \mathbb{E}[W_{i,j}^n W_j^n] &= \frac{j^{\alpha+1} - (j-1)^{\alpha+1}}{(\alpha+1)n^{\alpha+1}} \delta_{i,j} \end{aligned}$$

where $\delta_{i,j}$ id the Kronecker delta. Thus, simulating a fBM can be seen as simulating a Volterra process of the form:

$$V(t) = \sqrt{2\alpha+1} \int_0^t (t-s)^\alpha dW_s$$

Defining $\tilde{V}(t) = \frac{V(t)}{\sqrt{2H}}$ and taking:

$$g(s) = s^{H-\frac{1}{2}} \quad \sigma(s) = 1 \quad L_g(s) = 1 \quad s \in (0, T)$$

we have that $\alpha = H - \frac{1}{2} \in (-\frac{1}{2}, \frac{1}{2}) \setminus \{0\}$ and so the Gaussian Volterra process $\tilde{V}(t)$ is a TBSS process that satisfies the assumptions for the use of the hybrid scheme. Choosing $\kappa = 1$ the process is simulated, in the grid points, as:

$$V_n\left(\frac{i}{n}\right) = \sqrt{2\alpha+1} \left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\frac{i}{n} - s\right)^\alpha dW_s + \sum_{k=2}^i \left(\frac{b_k^*}{n}\right)^\alpha (W_{\frac{i-(k-1)}{n}} - W_{\frac{i-k}{n}}) \right)$$

using the covariance structure:

$$\Sigma = \begin{pmatrix} \frac{1}{n} & \frac{1}{(\alpha+1)n^{\alpha+1}} \\ \frac{1}{(\alpha+1)n^{\alpha+1}} & \frac{1}{(2\alpha+1)n^{2\alpha+1}} \end{pmatrix} \quad (3.2.2.2)$$

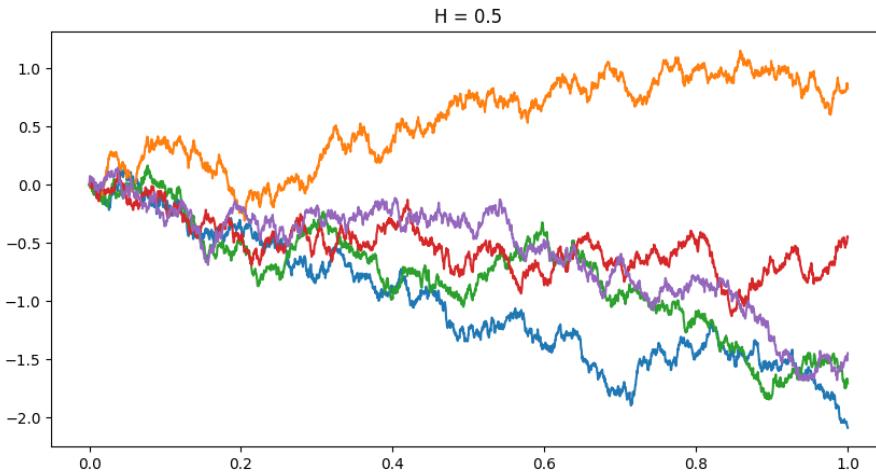
To summarize, the steps of the hybrid scheme are:

1. compute the covariance matrix Σ as expressed in (3.2.2.2);
2. generate a multivariate normal variable $Z = (Z_1, Z_2)$ with mean $\mu = (0, 0)$ and covariance Σ ;
3. estimate the first component (the integral one) of V_n using Z_2 since we have $\int_{\frac{i-1}{n}}^{\frac{i}{n}} (\frac{i}{n} - s)^\alpha dW_s \sim \mathcal{N}(0, \frac{1}{(2\alpha+1)n^{2\alpha+1}})$;

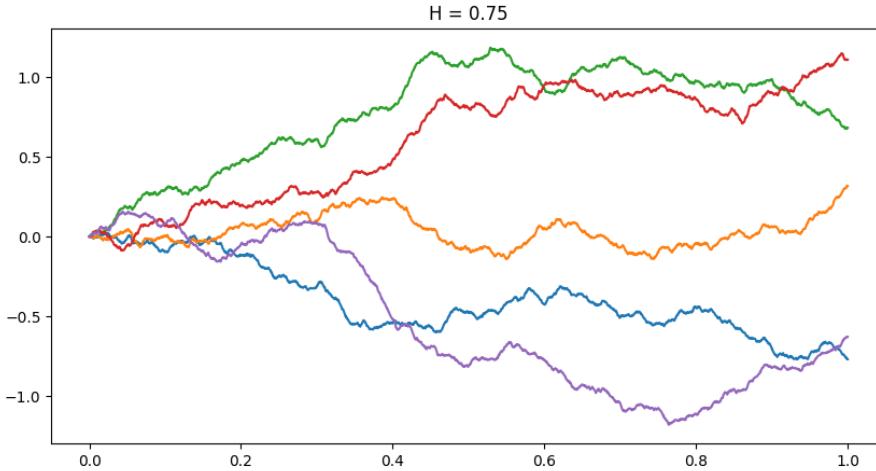
4. estimate the second component (the discrete sum) computing $(\frac{b_k^*}{n})^\alpha$ and compute the convolution with Z_1 ;
5. sum the two components and multiply by the factor $\sqrt{2\alpha + 1}$.

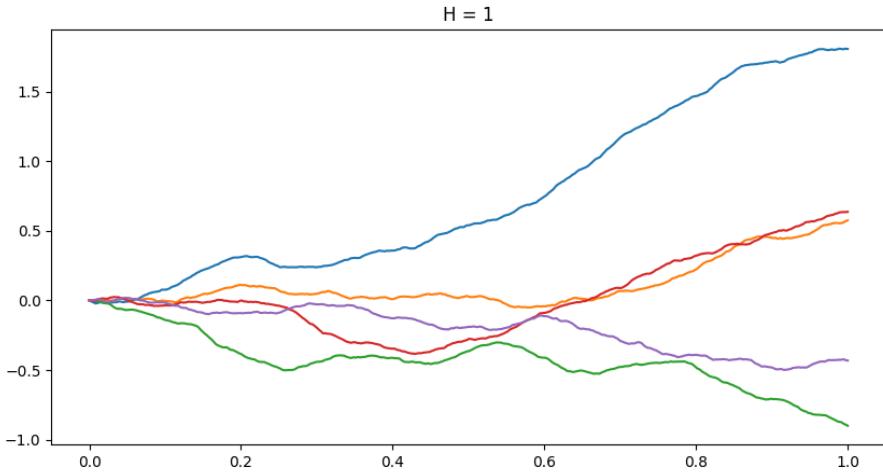
3.2.3 Simulation results

In this section we present a graphical representation of the effect of the Hurst parameter H on the paths. For all the simulations we have used as the time final time $T = 1$ and as the number of points in the grid we used $n = 2500$. First we present the paths of a standard Brownian Motion.

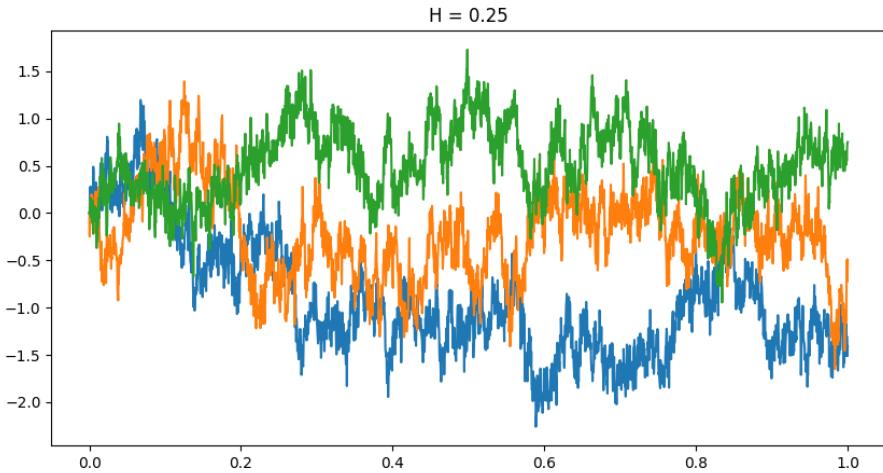


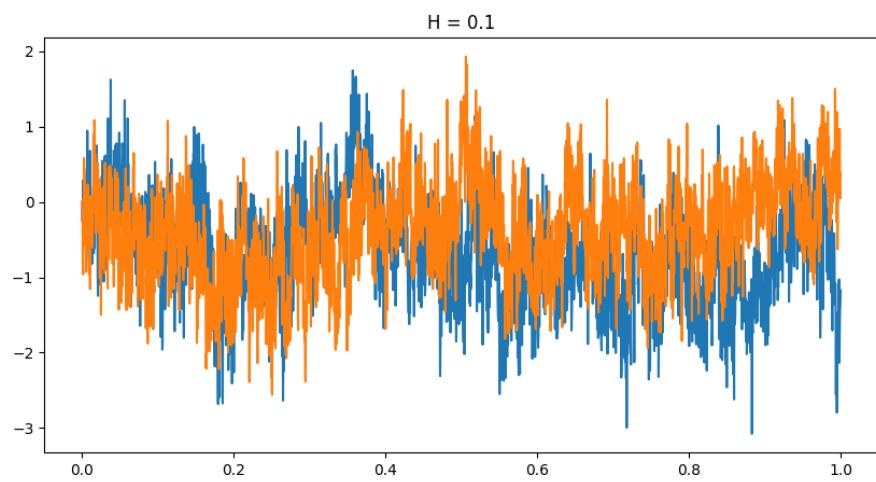
If we take $H > \frac{1}{2}$ we will have smoother trajectories as H keeps increasing as highlighted by the next two figures





While if we take $H < \frac{1}{2}$ we will have progressively rougher paths as H decreases as highlighted by the next two figures.





4 Rough Bergomi Model

The rough Bergomi (rBergomi) model is a Rough Fractional Stochastic Volatility (RFSV) model. The model has only three, time-independent, parameters and is able to replicate accurately the implied volatility surface dynamics. We will see that we need to use simulation methods to generate option prices as there is no closed form solution and the non-Markovian property of the model doesn't allow a PDE approach. From now on let $(\Omega, \mathbb{F} = \{(F_t)_{t \geq 0}\}, \mathbb{P})$ be a complete filtered probability space, let \mathbb{P} be the *physical-measure* and $T < \infty$ be the right limit of our time horizon.

4.1 The Model

The rBergomi model is known as a market model, that is to say a financial model consistent with market data. The idea of Bergomi, proposed in [7], is to model the dynamics of the forward variance instead of modelling instantaneous volatility. We will denote the forward variance curve observed at time t with maturity T as $\xi_t(T)$. The forward variance curve observed at time t with maturity T is associated with the fair strike of a variance swap, observed in the same instant t and with the same maturity T , that we will denote as $\sigma_t^2(T)$ and can be expressed via:

$$\sigma_t^2(T) = \frac{1}{T-t} \int_t^T \xi_t(u) du$$

equivalently we have:

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

4.1.1 N-Factor Model

In a general N-dimensional setting dictated by the N-dimensional Brownian motion $(W_t^i)_{i=1}^N$ the forward variance $\xi_t(T)$ dynamics are governed by the following SDE:

$$d\xi_t(u) = \frac{\omega}{\sqrt{\sum_{i,j=1}^N \omega_i \omega_j \rho_{i,j}}} \xi_t(u) \sum_{i=1}^N \omega_i e^{k_i(u-t)} dW_t^i \quad (4.1.1)$$

where we have that $d[W_t^i W_t^j]_t = \rho_{i,j} dt$ and $\omega_i, k_i > 0$. We also have that $\omega > 0$ is the instantaneous volatility of $\xi_t(t)$. In this general setting the solution is given by:

$$\xi_t(T) = \xi_0(T) \exp \left\{ \omega \sum_{i=1}^N \omega_i e^{-k_i(T-t)} X_t^i - \frac{\omega^2}{2} \sum_{i,j=1}^N \omega_i \omega_j e^{-(k_i+k_j)(T-t)} \mathbb{E}[X_t^i X_t^j] \right\}$$

where the N driven Ornstein-Uhlenbeck (OU) processes $(X_t^i)_{t \geq 0}$ are defined by:

$$\begin{cases} dX_t^i = -k_i X_t^i dt + dW_t^i \\ X_0^i = 0 \end{cases}$$

The instantaneous volatility of $\xi_t(T)$, thanks to (4.1.1), is:

$$\omega(T-t) = \frac{2\nu}{\sqrt{\sum_{i,j} \omega_i \omega_j \rho_{i,j}}} \sqrt{\sum_{i,j} \omega_i \omega_j \rho_{i,j} e^{-(k_i+k_j)(T-t)}}$$

Where ν is the log-normal volatility of a Variance Swap with vanishing maturity which can be related to the instantaneous volatility of $\xi_t(t)$ by:

$$\omega = 2\nu$$

4.1.2 One-Factor Model

Now we will restrict our analysis to the mono-dimensional case that is dictated by the Brownian motion $(W_t)_{t \geq 0}$. Now we have, for the forward variance curve, the following dynamics:

$$d\xi_t(T) = \omega e^{-k(T-t)} \xi_t(T) dW_t \quad (4.1.2)$$

The choice of an exponential decaying volatility function is equivalent to letting an OU process $(X_t)_{t \geq 0}$ dictate the dynamics of the forward variances. The process X_t has to satisfy the following SDE system:

$$\begin{cases} dX_t = -k X_t dt + dW_t \\ X_0 = 0 \end{cases}$$

We can solve the system and find the solution:

$$X_t = \int_0^t e^{-k(t-s)} dW_s$$

We can also calculate its expected value, its variance and the expected value of the square of the process:

$$\mathbb{E}[X_t] = 0 \quad \mathbb{V}[X_t] = \frac{1 - e^{-2k}}{2k} \quad \mathbb{E}[X_t^2] = \mathbb{V}[X_t] = \frac{1 - e^{-2k}}{2k}$$

Then the solution to the (4.1.2) SDE is:

$$\xi_t(T) = \xi_0(T) \exp \left\{ \omega e^{-k(T-t)} X_t - \frac{\omega^2}{2} e^{-2k(T-t)} \mathbb{E}[X_t^2] \right\}$$

This model is still not flexible enough to capture simultaneously the forward volatilities term structure and the forward skew. Thus, we should need more factors and the next step is the 2-Factors model.

4.1.3 2-Factors Model

Here we will present the 2-factors model in which we can achieve greater flexibility in the term-structure of volatilities of variances that can be generated. We will also introduce a mixing parameter $\theta \in [0, 1]$. in this context the dynamics become:

$$\begin{cases} d\xi_t(T) = \omega \alpha_\theta \xi_t(T) [(1-\theta)e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2] \\ \alpha_\theta = 1/\sqrt{(1-\theta)^2 + \theta^2 + 2\rho_{12}(1-\theta)\theta} \end{cases} \quad (4.1.3)$$

Where ρ is the correlation between W^1 and W^2 and we have defined the two OU processes X^1 and X^2 given by:

$$\begin{cases} dX_t^i = -k_i X_t^i dt + dW_t^i \\ X_0^i = 0 \end{cases}$$

We will also introduce the auxiliary Gaussian drift-less process:

$$dx_t^T = \alpha_\theta [(1-\theta)e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2]$$

whose quadratic variation is given by:

$$d\langle x^T \rangle_t = \eta^2(T-t)dt$$

where we have defined:

$$\eta(s) := \alpha_\theta \sqrt{(1-\theta)^2 e^{-2k_1 s} + \theta^2 e^{-2k_2 s} + 2\rho_{12}\theta(1-\theta)e^{-(k_1+k_2)s}}$$

Thus, substituting in the SDE (4.1.3) we obtain the simplified form:

$$d\xi_t(T) = \omega \xi_t(T) dx_t^T$$

Now we can obtain the solution that is given by:

$$\begin{cases} \xi_t(T) = \xi_0(T) e^{\omega x_t^T - \frac{\omega^2}{2} f(t, T)} \\ f(t, T) = \int_{T-t}^T \eta^2(u) du \end{cases}$$

We can explicit the value of $f(t, T)$ that is:

$$\begin{aligned} f(t, T) &= \alpha_\theta^2 \left[\frac{(1-\theta)^2}{2k_1} e^{-2k_1(T-t)} (1 - e^{-2k_1 t}) + \frac{\theta^2}{2k_2} e^{-2k_2(T-t)} (1 - e^{-2k_2 t}) \right. \\ &\quad \left. + 2\theta(1-\theta)\rho_{12}e^{-(k_1+k_2)(T-t)} \frac{1 - e^{-(k_1+k_2)t}}{k_1 + k_2} \right] \end{aligned}$$

4.2 The Realized Variance

We will use the Mandelbrot-Vann Ness representation of the fractional Brownian motion to express the increments of the logarithm of realized variance $v = \sigma^2$ under the *physical measure* \mathbb{P} as:

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

We note that $\mathbb{E}[M_t(u)|\mathcal{F}_t] = 0$ and that $Z_t(u)$ is \mathcal{F}_t -measurable. If we define $\tilde{W}_t^{\mathbb{P}}$ as:

$$\tilde{W}_t^{\mathbb{P}}(u) := \sqrt{2H} \int_t^u |u-s|^{H-\frac{1}{2}} dW_s^{\mathbb{P}}$$

it has the same properties of $M_t(u)$ and defining $\eta := \frac{2\nu C_H}{\sqrt{2H}}$ we have that:

$$\log(v_u) - \log(v_t) = \eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u)$$

This gives us:

$$v_u = v_t \exp \{ \eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u) \}$$

Thanks to the properties, gaussianity in this case, of $\tilde{W}_t^{\mathbb{P}}$ we have:

$$\tilde{W}_t^{\mathbb{P}}(u) \sim \mathcal{N}(0, (u-t)^{2H})$$

which gives us that $v_u|\mathcal{F}_t$ is log-normal and thus entails that:

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

Now we can express the realized variance as:

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

where $\mathcal{E}(\cdot)$ is the Doléans-Dade exponential.

4.3 The probability measure change

As observed in [8] the model with two factors is already over-parameterized so we will use the 1-factor model which, even if in theory is not flexible enough, in practice can achieve good quality results. From what we have said up to now we have that, under the *physical probability* \mathbb{P} , the model is expressed as:

$$\begin{cases} dS_u = \mu_u S_u du + \sqrt{v_u} S_u dZ_u^{\mathbb{P}} \\ v_u = v_t \exp \{ \eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u) \} \end{cases} \quad (4.3.1)$$

Now, in order to price options, we want to change the *physical* probability measure \mathbb{P} with an equivalent martingale measure \mathbb{Q} in the interval $[t, T]$. To do that we use Girsanov theorem and obtain:

$$dZ_u^{\mathbb{Q}} = dZ_u^{\mathbb{P}} + \frac{\mu_u - (r - q)}{\sqrt{v_u}} du$$

When we change from \mathbb{P} to \mathbb{Q} we also have to remember that the Brownian motion $W_u^{\mathbb{P}}$, that is used to construct the Volterra-type process $\tilde{W}_u^{\mathbb{P}}$, is correlated with $Z_u^{\mathbb{P}}$ with correlation factor ρ :

$$dW_u^{\mathbb{P}} = \rho dZ_u^{\mathbb{P}} + \sqrt{1 - \rho^2} d(Z_u^{\perp})^{\mathbb{P}}$$

where $(Z_u^{\perp})^{\mathbb{P}}$ is independent of $Z_u^{\mathbb{P}}$. A general change of measure for $(Z_u^{\perp})^{\mathbb{P}}$ is of the form:

$$(Z_u^{\perp})^{\mathbb{Q}} = (Z_u^{\perp})^{\mathbb{P}} + \gamma_u du$$

where γ_u is a suitable process that can be seen as the market price of volatility risk. Now we can express the change in measure also for $W_u^{\mathbb{Q}}$:

$$\begin{aligned} dW_u^{\mathbb{Q}} &= \rho dZ_u^{\mathbb{Q}} + \sqrt{1 - \rho^2} d(Z_u^{\perp})^{\mathbb{Q}} \\ &= dW_u^{\mathbb{P}} + \left(\frac{\mu_u - (r - q)}{\sqrt{v_u}} \rho + \gamma_u \sqrt{1 - \rho^2} \right) du \\ &= dW_u^{\mathbb{P}} + \lambda_u du \end{aligned}$$

We may now rewrite, assuming that the filtration generated by $W^{\mathbb{P}}$ is the same as the one generated by $W^{\mathbb{Q}}$, the dynamics of the realized variance:

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

where we have

$$\tilde{W}_t^{\mathbb{Q}}(u) := \sqrt{2H} \int_t^u (u-s)^{H-\frac{1}{2}} dW_s^{\mathbb{Q}}$$

Thus, the model (4.3.1) under \mathbb{Q} is expressed as:

$$\begin{cases} dS_u = (r - q)S_u du + \sqrt{v_u} S_u dZ_u^{\mathbb{Q}} \\ v_u = \xi_t(u) \mathcal{E}(\eta \tilde{W}_t^{\mathbb{Q}}(u)) \end{cases} \quad (4.3.2)$$

This model is a non-Markovian generalization of the Bergomi model. Specifically, this model is non-Markovian in the instantaneous variance v_t :

$$\mathbb{E}^{\mathbb{Q}}[v_u | \mathcal{F}_t] \neq \mathbb{E}^{\mathbb{Q}}[v_u | v_t]$$

but is Markovian in the infinite-dimensional state vector:

$$\mathbb{E}^{\mathbb{Q}}[v_u | \mathcal{F}_t] = \xi_t(u)$$

4.4 Pricing

Under the pricing measure \mathbb{Q} , given the starting time $t_0 = 0$, the scheme to simulate the model is:

$$\begin{cases} S_t = S_0 \exp \left\{ (r - q)t - \frac{1}{2} \int_0^t v_u du + \int_0^t \sqrt{v_u} dZ_u^{\mathbb{Q}} \right\} \\ v_t = \xi_0(t) \exp \left\{ 2\eta C_H \int_0^t (t-u)^{H-\frac{1}{2}} dW_u^{\mathbb{Q}} - \frac{\eta^2 C_H^2}{H} t^{2H} \right\} \\ d[Z^{\mathbb{Q}}, W^{\mathbb{Q}}]_t = \rho dt \end{cases}$$

First we need to simulate the Volterra process using the hybrid scheme introduced in **Section 3.2.2**. Then we need to extract the Brownian Motion $W^{\mathbb{Q}}$ that drives the Volterra process and correlate it with $Z^{\mathbb{Q}}$ by the parameter ρ . Lastly, we simulate the stock price process S using the forward Euler scheme. To sum up, in order to simulate the stock price process we have to:

1. fix an equispaced grid $\mathcal{G} = \{t_0 = 0, t_1 = \frac{1}{n}, \dots, t_{\lfloor nT \rfloor} = \frac{\lfloor nT \rfloor}{n}\}$;
2. simulate the Volterra process $\mathcal{V}_t = \int_0^t (t-u)^{H-\frac{1}{2}} dW_u^{\mathbb{Q}}$, $t \in \mathcal{G}$, using the hybrid scheme;
3. compute the variance process v using the previously computed Volterra process:

$$v_t = \xi_0(t) \mathcal{E}(2\eta C_H \mathcal{V}_t) \quad t \in \mathcal{G}$$

4. extract the path of the Brownian Motion $W^{\mathbb{Q}}$ that drives the Volterra process:

$$\begin{aligned} W_{t_i}^{\mathbb{Q}} &= W_{t_{i-1}}^{\mathbb{Q}} + n^{H-\frac{1}{2}} (\mathcal{V}_{t_i} - \mathcal{V}_{t_{i-1}}) & i = 1, \dots, \kappa \\ W_{t_i}^{\mathbb{Q}} &= W_{t_{i-1}}^{\mathbb{Q}} + W_{i-1}^n & i > \kappa \end{aligned}$$

where W^n is defined as in formula (3.2.1) in **Section 3.2.2**;

5. correlate the stock price process, driven by $Z^{\mathbb{Q}}$, and the variance process, driven by $W^{\mathbb{Q}}$ through the Volterra process, as:

$$Z_{t_i}^{\mathbb{Q}} - Z_{t_{i-1}}^{\mathbb{Q}} = \rho(W_{t_i}^{\mathbb{Q}} - W_{t_{i-1}}^{\mathbb{Q}}) + \sqrt{1-\rho^2}(W_{t_{i-1}}^{\mathbb{Q},\perp} - W_{t_{i-1}}^{\mathbb{Q},\perp})$$

where $W^{\mathbb{Q},\perp}$ is a standard Brownian Motion independent of $W^{\mathbb{Q}}$;

6. simulate the stock price process S using the forward Euler scheme:

$$S_{t_i} = S_{t_{i-1}} + (r - q)S_{t_{i-1}}(t_i - t_{i-1}) + \sqrt{v_{t_{i-1}}}S_{t_{i-1}}(Z_{t_i}^{\mathbb{Q}} - Z_{t_{i-1}}^{\mathbb{Q}})$$

To price an option at time $t < T$, where T is the maturity, that has payoff $f(S_T)$ we have to calculate the discounted payoff given by:

$$P_t = \mathbb{E}^{\mathbb{Q}}[e^{-(r-q)(T-t)}f(S_T)|\mathcal{F}_t]$$

To compute this quantity we will use a Monte-Carlo simulation. In practice we have adapted to our use the implementation devised by McCricked and Pakkanen in [9].

4.5 Calibration

Since we are using the one-factor rBergomi model when we are calibrating it to market data we are finding the values for the parameters used to simulate the price paths: H , η and ρ . These parameters have direct interpretation: H controls the decay of the term structure of volatility skew for very short tenors whereas the product $\rho\eta$ sets the level of the ATM skew for longer tenors. Keeping the product $\rho\eta$ roughly constant but decreasing ρ (so as to make it more negative) has the effect of pushing the minimum of each smile towards higher strikes. As the initial forward variance curve we will use the parametrization found in **Section 2.2**.

4.5.1 Objective Function

To calibrate the model we want to minimize the difference between the market implied volatility and the rBergomi implied volatility. To calculate the rBergomi IV we have to set the parameters, simulate the price paths, compute the price of the option (given a certain strike k and time to maturity τ) and then invert the B&S formula to find the IV given by our simulation, in formula:

$$\sigma_{rB}(k, \tau, H, \eta, \rho) = P_{BS}^{-1}(k, \tau, H, \eta, \rho)$$

We indicate the set of parameters that we have to calibrate as $\Theta := \{H, \eta, \rho\}$ and the set of strikes for a given maturity as K . The problem of minimization is thus:

$$\underset{\Theta}{\operatorname{argmin}} f(K, \tau) := \underset{\Theta}{\operatorname{argmin}} \|\sigma_{rB}(k, \tau) - \sigma_{mkt}(k, \tau)\|^2$$

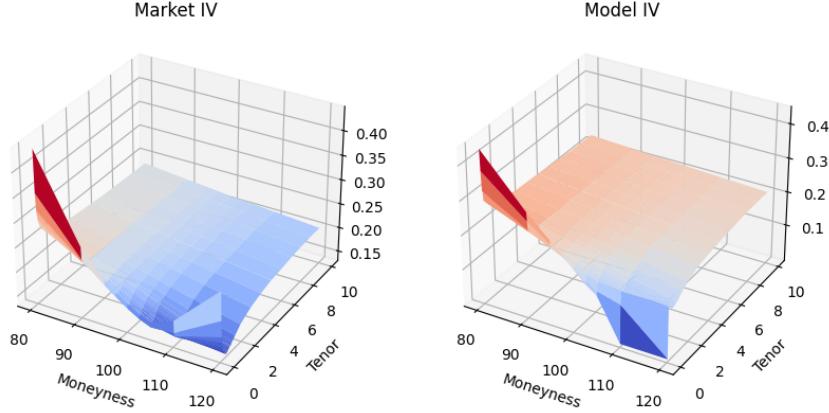
where σ_{mkt} represents the market IV. We calibrate the model for every tenor using the sequential least squares programming algorithm proposed in [10]. We will also calibrate the model in a global way finding just a set of parameters for all the tenors.

4.5.2 Numerical results

We report some of the calibrated parameters in the following table.

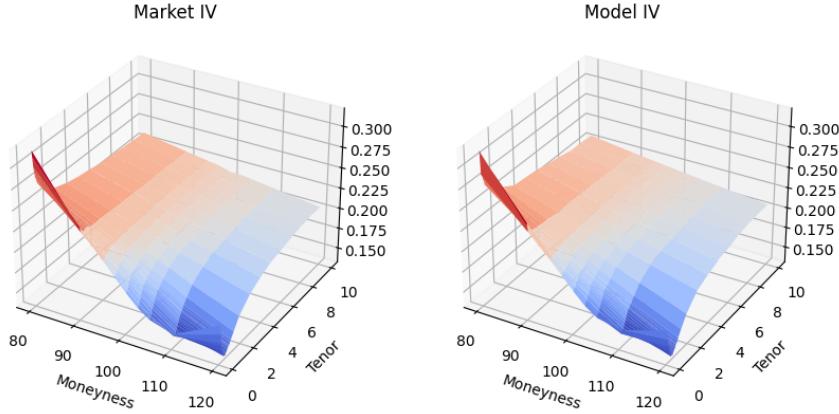
Tenor	H	η	ρ
2 weeks	0.0279	1.9077	-0.8583
1 month	0.0328	2.0162	-0.8441
6 months	0.0811	1.9189	-0.8923
1 year	0.0958	1.7628	-0.9368
10 years	0.0466	1.4197	-0.9575

Comparing our results with the market we obtained a mean relative percentage error of 2.2799%.



Comparison between market and model IV, local approach

As we can see the model fails for the first two tenors in the high moneyness range. In fact, if we exclude the first two tenors we obtain a mean relative percentage error of 0.7832%.

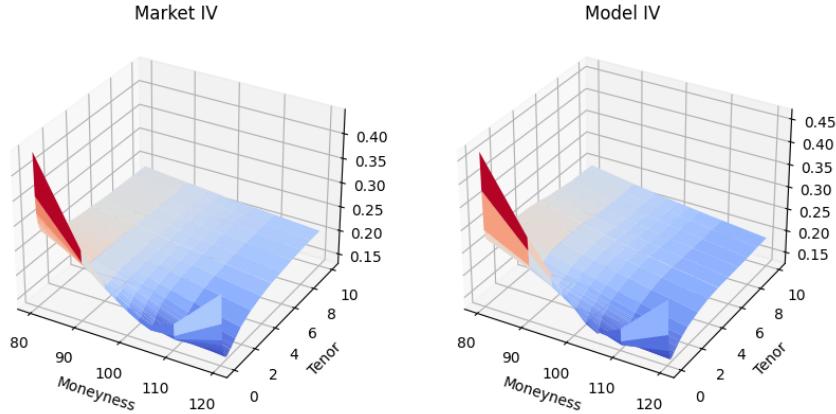


Comparison between market and model IV, local approach restricted

In the global approach the calibrated parameters are:

\bar{H}	η	ρ
0.0856	1.8906	-0.8978

and the mean relative percentage error that we obtained is 3.1008%.



Comparison between market and model IV, global approach

4.6 Volatility Skew

The at-the-money (ATM) volatility skew ψ is defined as:

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \hat{\sigma}(k, \tau) \right|_{k=0}$$

where k is the log-moneyness, τ is the time to maturity and $\hat{\sigma}(k, \tau)$ is the implied volatility. The term structure of the market data at-the-money-forward (ATMF) volatility skew Ψ is defined as:

$$\Psi(\tau) := \left| \frac{\partial}{\partial K} \sigma_{mkt}(K, \tau) \right|_{K=F}$$

where K is the strike, F is the forward price and σ_{mkt} is the market implied volatility. It is easy to express the log-moneyness strike k in terms of actual strike K as:

$$k = \log \left(\frac{K}{F} \right)$$

Equating the two implied volatilities $\hat{\sigma}(k, \tau) = \sigma_{mkt}(K, \tau)$ we obtain:

$$\frac{\partial}{\partial K} \sigma_{mkt}(K, \tau) = \frac{\partial}{\partial k} \hat{\sigma}(k, \tau) \frac{\partial k}{\partial K} = \frac{1}{K} \frac{\partial}{\partial k} \hat{\sigma}(k, \tau)$$

Therefore we have mapped data expressed in terms of strike K to data expressed in terms of log-moneyness k . Thus, the ATMF volatility skew can be expressed as:

$$\psi(\tau) = F \left| \frac{\partial}{\partial K} \sigma_{mkt}(K, \tau) \right|_{K=F}$$

4.6.1 Numerical results

In order to compute the market ATMF volatility skew we will fit a stochastic volatility inspired (SVI) model to the market implied volatility data. The SVI model, presented in [11] and further expanded in [12], is calibrated to the market implied volatility surface using a set of parameters $\lambda = \{a, b, \rho, m, \sigma\}$ such that the total implied variance is expressed, for $k \in \mathbb{R}$, as:

$$\sigma_{imp}^2(k; \lambda) = a + b[\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}]$$

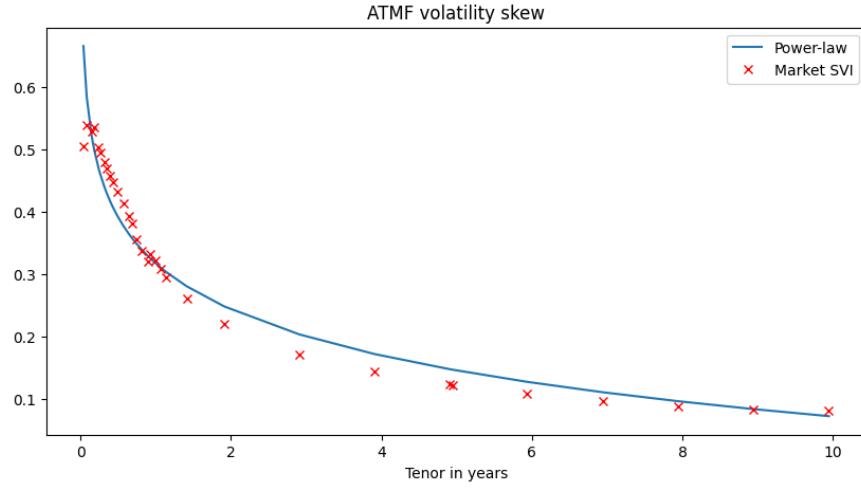
where $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$ and $\sigma > 0$. In addition, if $a + b\sigma\sqrt{1 - \rho^2} \geq 0$ we have that the condition $\sigma_{imp}^2(k; \lambda) > 0$ is obtained. Every parameter has distinct effect:

- a controls the general level of variance being a vertical translation of the smile;
- b controls the slopes of both puts and calls, increasing b we will tighten the smile;
- increasing ρ we will have a counter-clockwise rotation of the smile;
- m controls the x-axis position of the smile, increasing m will translate the smile to the right;

- σ controls the ATM curvature of the smile, increasing ρ will reduce the curvature.

It has been shown, for example in [13], that the term structure of the ATMF volatility skew can be approximated by a power-law function of the time to maturity. In this context the function $\tau \rightarrow A\tau^{-\alpha}$ is fitted to the market data obtaining the following values:

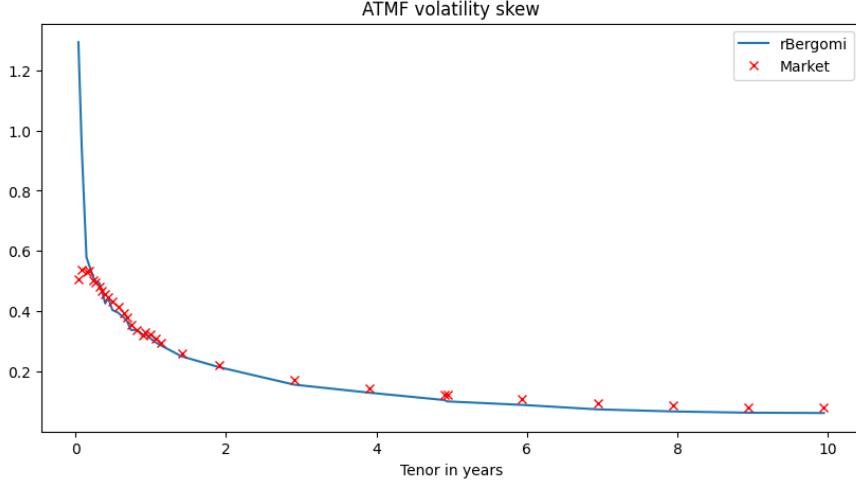
$$A = 0.30287671 \quad \alpha = 0.26347397$$



To compute the ATMF volatility skew given by the rBergomi model we will first calculate the implied volatility given by the model for log-moneyness in the neighbourhood of $k = 0$ and then use the central finite difference approximation. Letting $h > 0$ be the difference step and τ be one of the tenors, the approximation of ψ is given by:

$$\psi(\tau) \approx \left| \frac{\hat{\sigma}_{rB}(h, \tau) - \hat{\sigma}_{rB}(-h, \tau)}{2h} \right|$$

By construction, since the rBergomi model smile should fit the market data also the skew generated by it should fit the market skew obtained by the SVI model. In practice, using $h = 10^{-3}$ for the approximation scheme, we obtained the following fit.



We note that apart from the first two tenors the fit is quite good.

4.7 Forward-Start Options

Many options can be used to study the sensitivity to forward-smile risk, which is defined as the risk coming from the market future implied volatility and its uncertainty. These options, of which the forward-start options are an example, are priced given the distribution of forward returns in the model, as described in [14]. We remember that the payoff of a forward-start option involves the price of the underlying in two different dates T_1 and T_2 with $T_1 < T_2$. The payoff Φ can be expressed as:

$$\Phi = (S_{T_2} - kS_{T_1})^+$$

where $k > 0$ is the moneyness of the option. The forward smile represents the expected future implied volatility for moneyness k : all possible realizations of future smiles are averaged to give $\hat{\sigma}_k^{T_1, T_2}$, that is the implied Black and Scholes volatility for the forward-start option with moneyness k . The instantaneous volatility is time-dependent, due to the forward-start option's nature, whose payoff depends on S_{T_1} and S_{T_2} . The Black and Scholes implied variance for maturity T is given by:

$$\hat{\sigma}_T^2 := \int_t^T \sigma^2(u) du$$

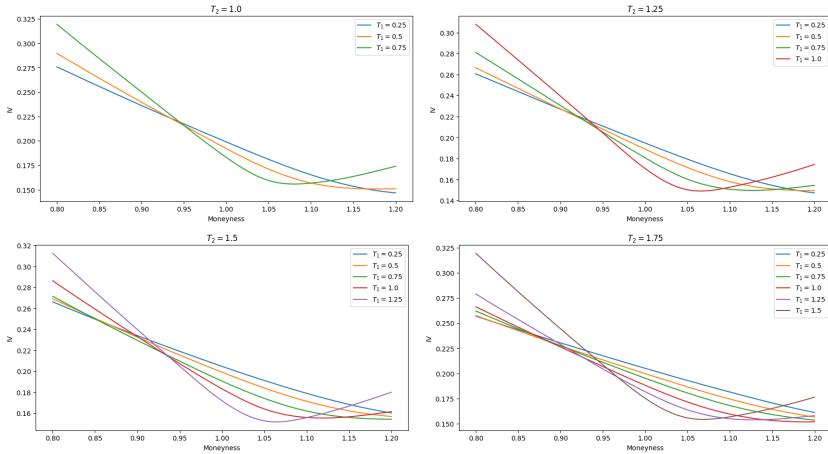
The price of a forward-start option is independent of the underlying price S , but depends on the forward volatility $\hat{\sigma}_{T_1, T_2}$, or the integrated variance over $[T_1, T_2]$, in such a way that:

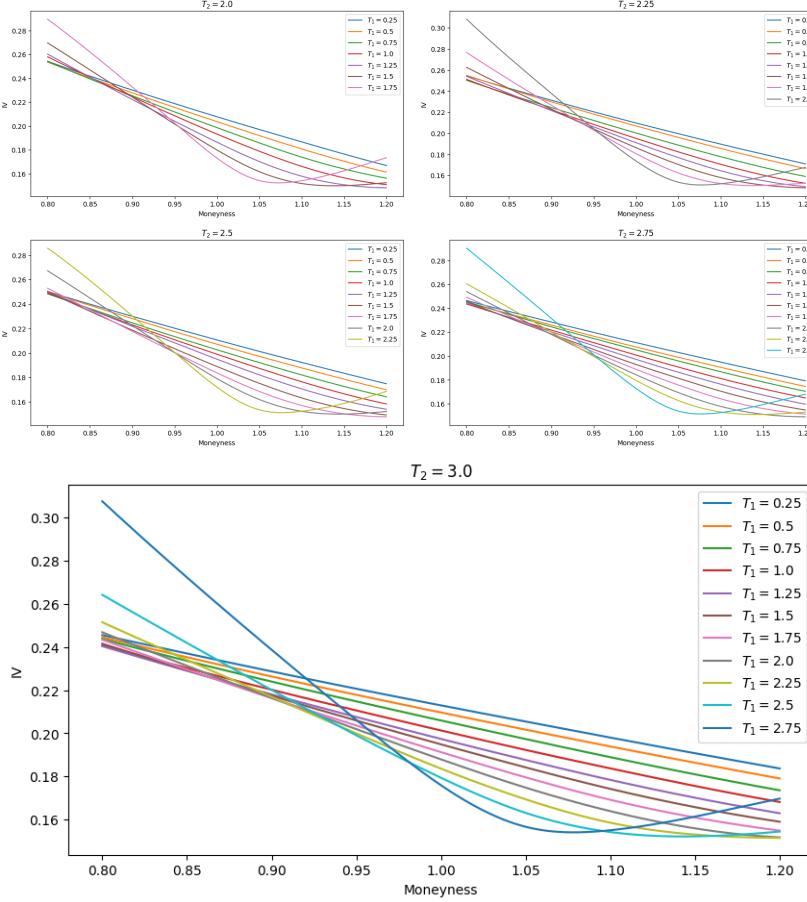
$$\begin{aligned}\hat{\sigma}_{T_1, T_2}^2 &:= \int_{T_1}^{T_2} \sigma^2(u) du \\ &= \frac{(T_2 - t)\hat{\sigma}_{T_2}^2 - (T_1 - t)\hat{\sigma}_{T_1}^2}{T_2 - T_1}\end{aligned}$$

4.7.1 Implied Volatility

As for a vanilla option to price a forward-start option in the rBergomi model we have to resort to a Monte-Carlo method. The generation of a price path is executed as explained in **Section 4.4** and then we extract the values at time T_1 and T_2 . The algorithm to compute the implied volatility of a forward-start option under rBergomi is the following:

1. set a range $\{1, 1.25, \dots, 3\}$ of step 0.25 for the maturity T_2 ;
2. set a range $\{0.25, \dots, T_2 - 0.25\}$ of step 0.25 for the starting date T_1 ;
3. simulate the price path under rBergomi;
4. set a range $\{0.8, 0.85, \dots, 1.2\}$ of step 0.05 for the moneyness k ;
5. compute the payoff of the forward-start option as $(S_{T_2} - kS_{T_1})^+$;
6. compute the price of the forward-start option as the discounted payoff over a Monte-Carlo simulation;
7. compute the corresponding Black and Scholes implied volatility using a root finding method, for example bisection.





4.8 Parameters stability

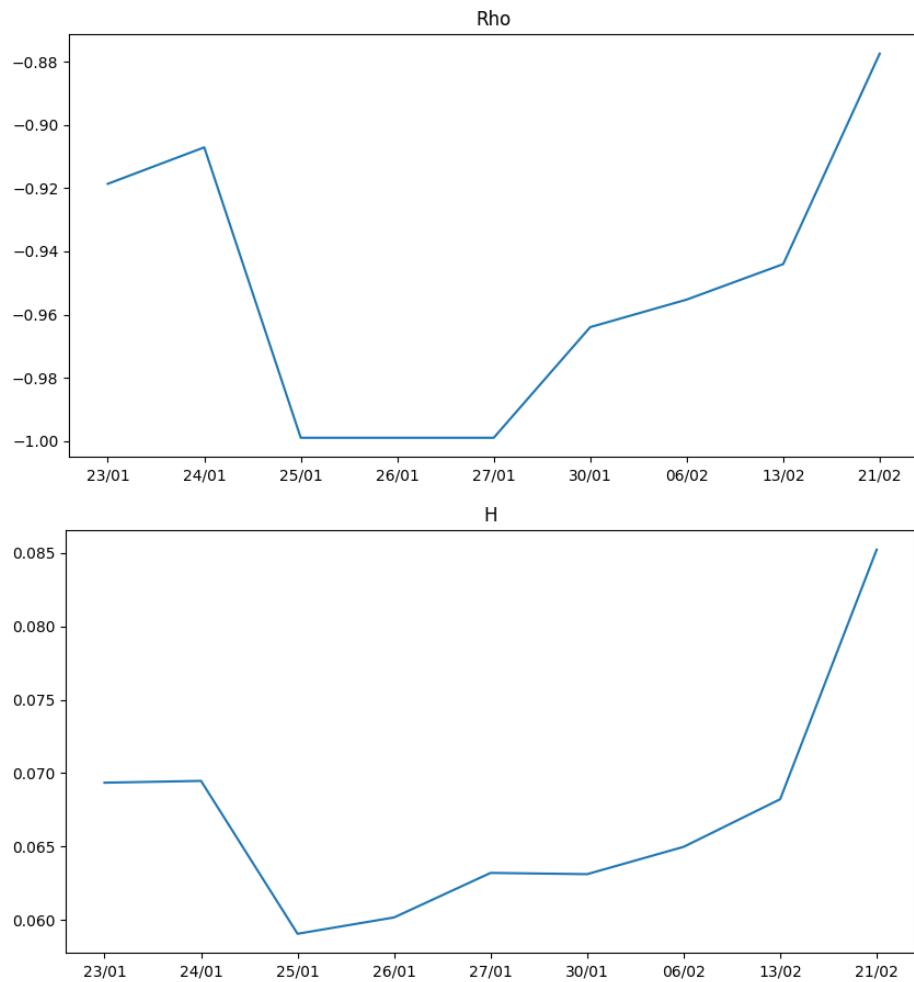
To check the stability of the model parameters we decided to test the model using data from other days than the principal one. The dates that we used are: 24/01/2023, 25/01/2023, 26/01/2023, 27/01/2023, 30/01/2023, 06/02/2023, 13/02/2023, 21/02/2023. In doing so we used only options available in all these days and they are 30 in total ranging from few months tenors to almost ten years. We did two tests:

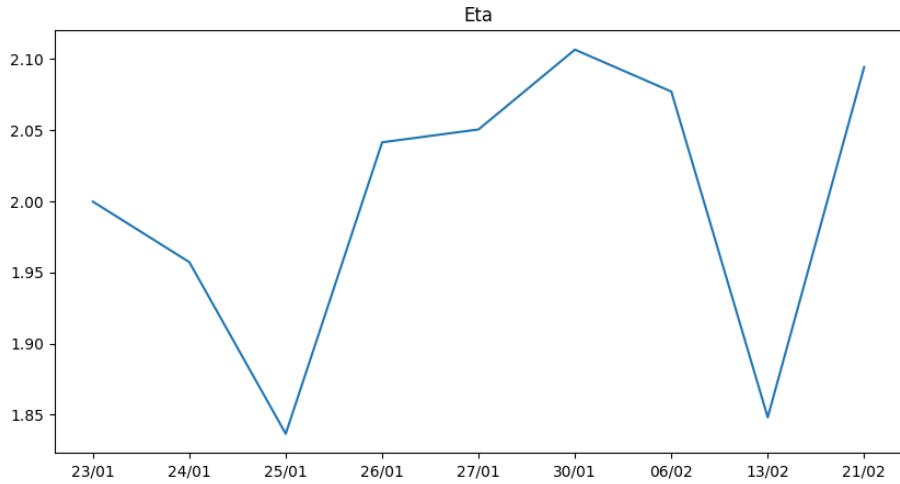
1. using the already calibrated parameters for the principal date we used the model to check the mean percentage error in all the other days comparing it with the error obtained calibrating the model for each day;
2. we calibrated the model in each day and then analyzed the standard deviation and variance of each parameter.

The next table contains the parameters and the error that we obtained for each day.

Day	ρ	H	η	Error
23/01	-0.9186	0.0694	1.9998	2.1776%
24/01	-0.9071	0.0695	1.9573	3.5448%
25/01	-0.9999	0.0591	1.8366	5.9528%
26/01	-0.9999	0.0602	2.0415	3.3960%
27/01	-0.9999	0.0632	2.0506	3.0902%
30/01	-0.9640	0.0631	2.1067	4.6805%
06/02	-0.9553	0.0650	2.0772	2.1697%
13/02	-0.9440	0.0682	1.8482	7.8451%
21/02	-0.8774	0.0852	2.0944	2.8109%

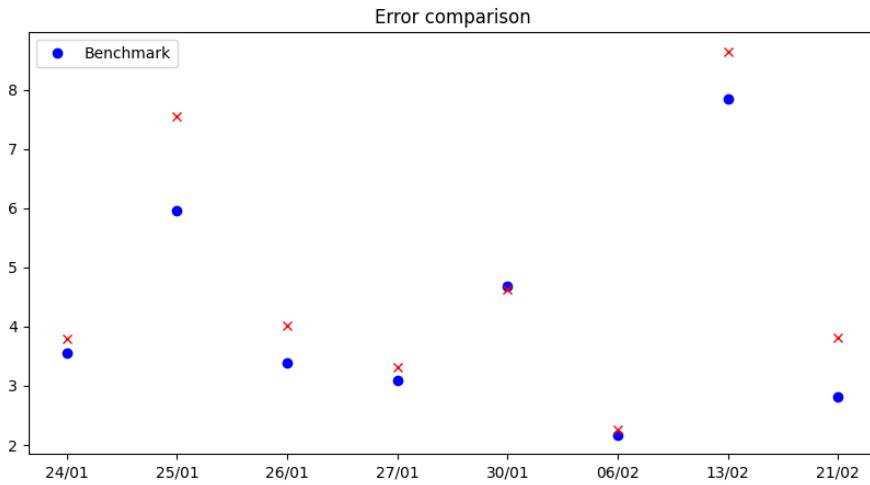
A graphical representation of the changes is presented in the next figures.





The next table contains the comparison between the benchmark error obtained before and the one obtained using as parameters the one calibrated for the first day (23/01).

Day	Benchmark	Error
24/01	3.5448%	3.7907%
25/01	5.9528%	7.5450%
26/01	3.3960%	4.0075%
27/01	3.0902%	3.3171%
30/01	4.6805%	4.6345%
06/02	2.1697%	2.2557%
13/02	7.8451%	8.6483%
21/02	2.8109%	3.8043%



Analyzing the standard deviation and variance of each parameter we obtained the following results.

Parameter	Standard Deviation	Variance
ρ	0.0415	0.0017
H	0.0074	0.0001
η	0.0955	0.0091

4.9 Bayesian Inverse Problem

To calibrate the model we used the Approximate Bayesian Computation (ABC) that is a sequential Monte-Carlo method. ABC methods (also called likelihood free inference methods), are a group of techniques developed for inferring posterior distributions in cases where the likelihood function is intractable or costly to evaluate. ABC comes useful when the model used contains unobservable random quantities, which make the likelihood function hard to specify, but data can be simulated from the model. These methods follow a general form:

1. Sample a parameter θ from a prior distribution $\pi(\theta)$.
2. Simulate a data set y using a function that takes θ and returns a data set of the same dimensions as the observed data set y_0 .
3. Compare the simulated dataset y with the experimental data set y_0 using a distance function d and a tolerance threshold ε .

In some cases a distance function is computed between two summary statistics $d(S(y^*), S(y_0))$, avoiding the issue of computing distances for entire datasets. As a result we obtain a sample of parameters from a distribution $\pi(\theta|d(y, y_0))$. If ε is sufficiently small this distribution will be a good approximation of the posterior distribution $\pi(\theta|y_0)$. Sequential monte carlo ABC is a method that iteratively morphs the prior into a posterior by propagating the sampled parameters through a series of proposal distributions $\phi(\theta^{(i)})$, weighting the accepted parameters $\theta^{(i)}$ like:

$$w^{(i)} \propto \frac{\pi(\theta^{(i)})}{\phi(\theta^{(i)})}$$

It combines the advantages of traditional SMC, that is the ability to sample from distributions with multiple peaks, but without the need for evaluating the likelihood function. For the implementation we will use the PyMC library. It is important to note that the role played by the summary statistic is a crucial one. In the next subsection we will present the Sliced Wasserstein distance proposed in [15].

4.9.1 Sliced Wasserstein distance

In order to avoid the possible loss of information due to the use of a summary statistic S , it has been proposed in [16] to operate on the full data by using low-variance Wasserstein distances in terms of empirical distributions of observed and synthetic data. These express distance via an optimal transport problem of minimizing, with respect to an underlying distance metric, the cost of transforming a given probability measure into another one. Consider $A \subset \mathbb{R}^n$, $\mathcal{P}(A)$ as the set of probability measure on A and $p \geq 1$. Then we can define the space

$$\mathcal{P}_p(A) = \left\{ \mu \in \mathcal{P}(A) : \int_A \|y - y_0\|^p d\mu(y) < \infty \right\}$$

for some $y_0 \in A$. The p -Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(A)$ is then

$$W_p^p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{A \times A} \|x - y\|^p d\gamma(x, y)$$

where $\Gamma(x, y)$ is the set of probability measures on $A \times A$ verifying that the marginal distributions are μ and ν respectively. Finding the solution of the optimal transport problem may be computationally challenging, especially for high-dimensional problems. To avoid this problem it has been suggested in [15] to project multi-dimensional distributions to one-dimensional ones via linear projections and then average the 1D Wasserstein distances, which can be efficiently calculated by sorting, across the various projections via a Monte-Carlo integral. Let \mathbb{S}^{d-1} be the d -dimensional unit sphere, $u \in \mathbb{S}^{d-1}$. u^* the linear form associated with u , such that $\forall a \in A$, $u^*(a) = \langle u, a \rangle$ where the inner product is the Euclidean one. The Sliced Wasserstein distance is then:

$$SW_p^p(\mu, \nu) = \int_{\mathbb{S}^{d-1}} W_p^p(u_*^\# \mu, u_*^\# \nu) d\sigma(u)$$

where σ is the uniform distribution on the unit sphere and $u_*^\# \mu$ denote the push-forward measure of μ by u^* . This is still a distance on $\mathcal{P}_p(A)$ and has significant lower computation requirement than the Wasserstein distance. Under some mild assumptions, it has been proved that the limiting posterior converges to the true posterior and also that if the number of samples tends to infinity then the Sliced Wasserstein distance between two empirical distributions converges to the Sliced Wasserstein distance between the two real distributions from which the observations are drawn.

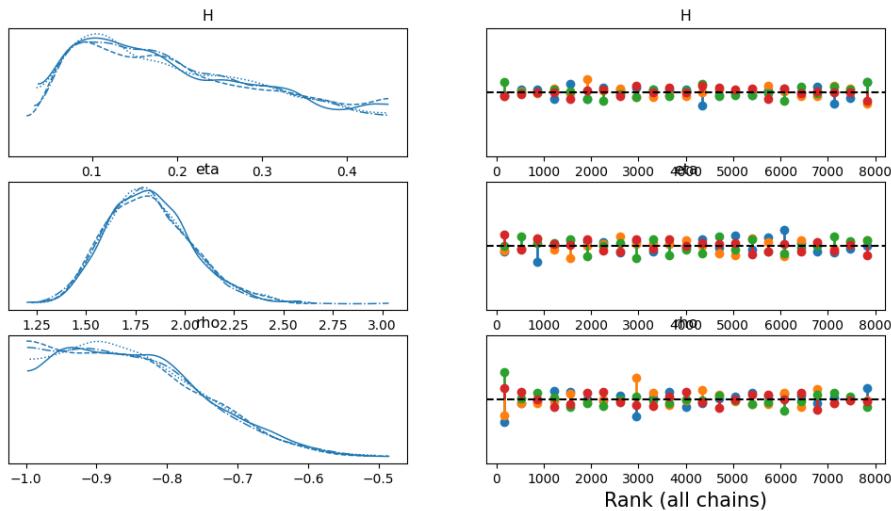
4.9.2 Numerical results

We decided to calibrate the model using only the tenors greater than 1 month and less than 6 months (9 in total). The set of parameters to calibrate is ρ , H and η . We used priors as uninformative as possible:

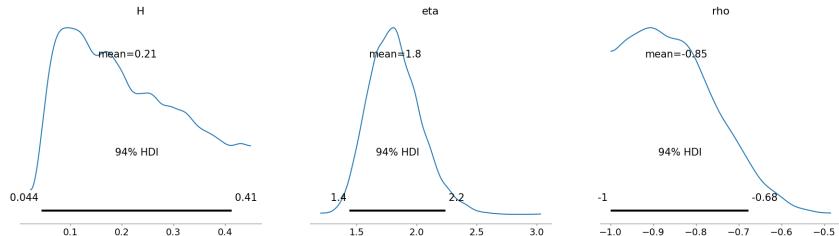
- $\rho \sim \mathcal{U}(-1, 0)$

- $H \sim \mathcal{U}(0, 0.45)$
- $\eta \sim \mathcal{U}(1, 5)$

We calibrated using the prices instead of the IV, we know that this is not the best, but for time and computational reasons we had to resort to this. We used 4 chains with 2000 samples each, also just for computational and time reasons. Then to test the quality of the calibration we used the maximum of the resulting distribution and plugged it in the model. This first figure represents on the left the distribution obtained from all the 4 chains superimposed, while in the right part we have a graphical representation of the rank of each chain: the vertical lines indicate deviation from the ideal expected value, which is represented with a black dashed line, if it is above we have more samples than expected and vice versa if it is below.



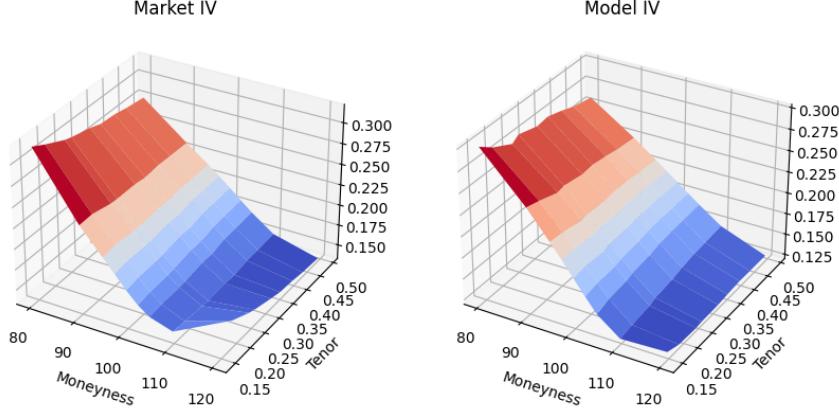
We note that the chains are coherent and the rank plots are good. The next figure represents the posterior density that we have obtained.



To compare the Bayesian calibration error we have to fix the parameters and then compute the error. We decided to use as parameters the maximum a posteriori of the densities that we have obtained, in this case they are:

ρ	H	η
-0.8967	0.1024	1.8348

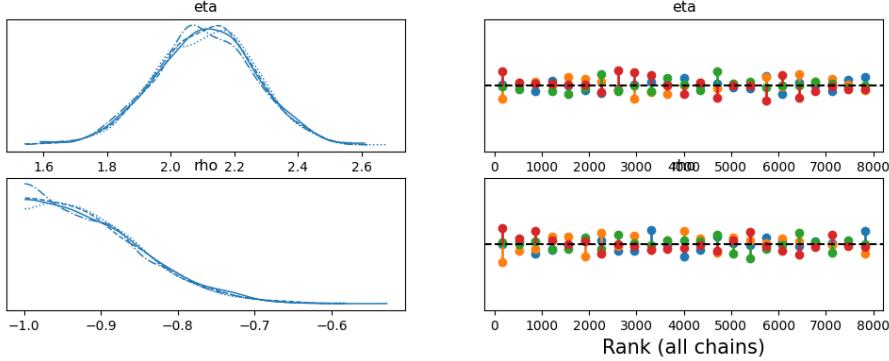
Using these parameters we obtained a mean percentage error of 4.0221%. The next figure contains the comparison between the market and model IV surfaces.



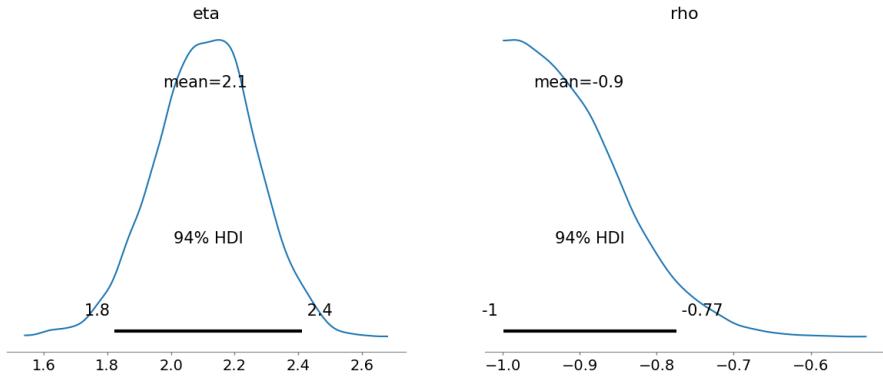
We also decided to do another test. We calibrated the model, with the global approach, using only the tenors that we will use in the Bayesian calibration and we obtained the following set:

ρ	H	η
-0.8831	0.0566	2.1398

obtaining a mean percentage error of 2.1518%. Then we fixed the H parameter and we calibrated the ρ and η parameters in the Bayesian framework.



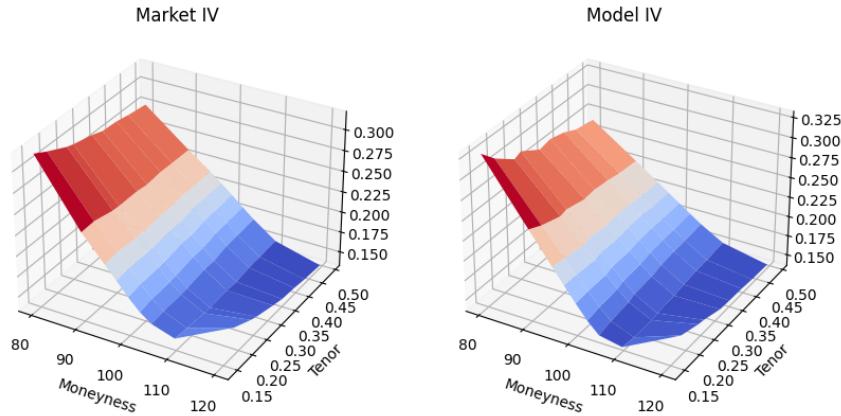
We note that the chains are coherent and the rank plots are good. The next figure represents the posterior density that we have obtained.



The maximum a posteriori of the densities that we have obtained are:

ρ	η
-0.9870	2.1484

Using these parameters and the fixed H we obtained a mean percentage error of 2.0087%. The next figure is the comparison between the market and model IV surfaces.



5 Quintic Ornstein-Uhlenbeck model

The Quintic Ornstein-Uhlenbeck (Quintic) volatility model, introduced in [19], is a stochastic volatility model where the volatility process is defined as a polynomial of degree five of a single Ornstein-Uhlenbeck (OU) process which has a fast mean reversion and a large volatility of volatility (vol-of-vol). This model tries to address the problem of the joint calibration of SPX derivatives and VIX derivatives.

5.1 The Model

Under the pricing measure \mathbb{Q} the dynamics of the stock price S are given by:

$$\begin{cases} dS_t &= (r - q)dt + \sigma_t S_t dB_t \\ \sigma_t &= \sqrt{\xi_0(t)} \frac{p(X_t)}{\sqrt{\mathbb{E}[p(X_t)^2]}} \\ dX_t &= -\left(\frac{1}{2} - H\right)\epsilon^{-1} X_t dt + \epsilon^{H-1/2} dW_t \quad X_0 = 0 \end{cases}$$

where B_t and W_t are two Brownian motions with correlation parameter $\rho \in [-1, 1]$. $\xi_0 \in L^2([0, T], \mathbb{R}^+)$ for any $T > 0$ is an input curve used to match certain term-structures observed in the market, we will use the forward variance curve since the normalization $\sqrt{\mathbb{E}[p(X_t)^2]}$ allows ξ_0 to match it:

$$\mathbb{E}\left[\int_0^t \sigma_s^2 ds\right] = \int_0^t \xi_0(s) ds \quad t \geq 0$$

The fifth grade polynomial $p(x)$ is defined as:

$$p(x) := \alpha_0 + \alpha_1 x + \alpha_3 x^3 + \alpha_5 x^5$$

with non-negative parameters $\alpha_0, \alpha_1, \alpha_3, \alpha_5 \geq 0$ ($\alpha_2 = \alpha_4 = 0$). The choice of a polynomial of degree five allows to reproduce the upward slope of the VIX smile, while restricting the coefficients α to be non-negative allows the sign of the ATM skew to be the same as ρ , as explained in more detail in [20]. We decided to set $\alpha_2 = \alpha_4 = 0$ in order to reduce the number of parameters to calibrate and this doesn't impact in a significant way the results as highlighted in [1]. The process X_t that drives the volatility is an OU process where the two parameters $H \in (-\infty, 1/2]$ and $\epsilon > 0$ control the mean-reversion speed through $(1/2 - H)\epsilon^{-1}$ and the vol-of-vol through $\epsilon^{H-1/2}$. For small values of ϵ we have a fast mean-reversion regime and a large vol-of-vol. These types of parametrizations can remind of the fast regimes studied in depth in [21] by Fouque. They can also be linked to jump models, studied for example in [22] and [23], when $H \leq -1/2$ and to rough volatility models, for example those presented in [20] and [23], where $H \in (0, 1/2)$ plays the role of the Hurst index.

We will restrict our analysis to this last case. The solution of the OU process is:

$$X_t = \epsilon^{H-1/2} \int_0^t e^{-(1/2-H)\epsilon^{-1}(t-s)} ds$$

Thus, the set of parameters to calibrate is:

$$\Theta := \{\rho, H, \epsilon, \alpha_0, \alpha_1, \alpha_3, \alpha_5\}$$

plus the input curve ξ_0 . As said before we will use the market initial forward variance curve parametrized using the Gompertz function as in **Section 2.2**.

5.2 SPX derivatives

To price SPX derivatives we have to resort to Monte Carlo simulations since there isn't a closed formula. Nevertheless, since X is a OU process it can be simulated exactly instead of approximating it using, for example, the Euler scheme which is often inaccurate in a fast mean-reversion regime. In order to simulate X we first define the auxiliary process \tilde{X} :

$$\tilde{X}_t := X_t e^{(1/2-H)\epsilon^{-1}t} = \epsilon^{H-1/2} \int_0^t e^{(1/2-H)\epsilon^{-1}s} dW_s$$

Thus, \tilde{X} can be simulated recursively by:

$$\tilde{X}_{t_{i+1}} = \tilde{X}_{t_i} + \sqrt{\frac{\epsilon^{2H}}{1-2H}} \left(e^{\frac{1-2H}{\epsilon} t_{i+1}} - e^{\frac{1-2H}{\epsilon} t_i} \right) Y_i$$

where Y_i are i.i.d. standard Gaussian. Naturally, to get back to $X_{t_{i+1}}$ we just have to divide $\tilde{X}_{t_{i+1}}$ by the factor $e^{\frac{1-2H}{\epsilon} t_{i+1}}$. This allows us to easily vectorize computations. Whereas to simulate the log-process $\log(S)$ we will use the Euler scheme paired with antithetic and control variates, that is the so called turbocharging method as outlined in [18] that we have also used in the rBergomi model. This means that we only need to simulate the part of $\log(S)$ that is $F^{\mathcal{W}}$ measurable, we call this $S^{\mathcal{W}}$, and this can be simulated as:

$$\log(S_{t_{i+1}}^{\mathcal{W}}) = \log(S_{t_i}^{\mathcal{W}}) - \frac{1}{2}(\rho\sigma_{t_i})^2(t_{i+1}-t_i) + \rho\sigma_{t_i}\sqrt{t_{i+1}-t_i}Y_i + \rho^2(r-q)(t_{i+1}-t_i)$$

We will use an equi-spaced grid so that calling the time step h the above formula reduces to:

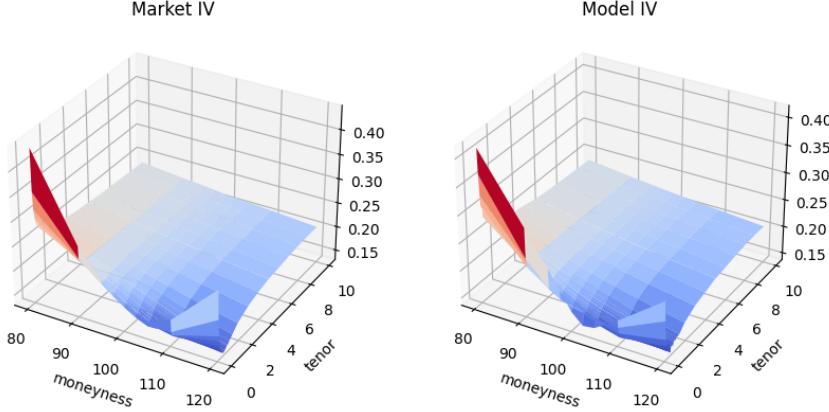
$$\log(S_{t_{i+1}}^{\mathcal{W}}) = \log(S_{t_i}^{\mathcal{W}}) + \left(r - q - \frac{1}{2}\sigma_{t_i}^2 \right) \rho^2 h + \rho\sigma_{t_i}\sqrt{h}Y_i$$

5.2.1 Numerical results

We decided to calibrate our model both in a local and global way, as in the rBergomi set. In the local case we calibrated a set of parameters for every tenor while in the global approach we calibrated the set of parameters so that it is the best fit considering all the tenors together. We report, for the local case, some of the calibrated parameters in the following table.

Tenor	ρ	H	ε	α_0	α_1	α_3	α_5
2 weeks	-0.5332	0.1034	0.1078	0.1353	0.6193	0.0954	0.1024
1 month	-0.6626	0.0915	0.0171	0.9777	0.0187	0.0343	0.1072
6 months	-0.6968	-0.0397	0.0083	1.2204	0.0035	0.2296	0.0462
1 year	-0.8565	0.1903	0.0234	1.0277	0.2463	0.0672	0.6016
10 years	-0.6658	0.1141	0.1173	0.9099	0.5726	0.1604	0.1229

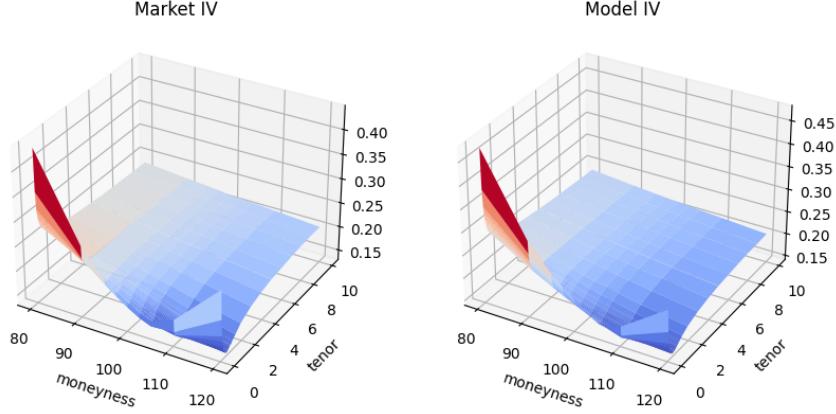
Comparing our results with the market we obtained a mean relative percentage error of 1.9299%. The next figure is the comparison between market and model IV surfaces.



Now we report, for the global approach, the set of the calibrated parameters.

ρ	H	ε	α_0	α_1	α_3	α_5
-0.9468	0.0305	0.1024	0.6101	0.3713	0.0054	0.0394

In this case we obtained a mean relative error of 3.5868%. The next figure is the corresponding comparison between market and model IV surfaces.



5.3 VIX derivatives

One major advantage of the Quintic OU model is that there is an explicit expression for the VIX. In a continuous time framework the VIX can be expressed as:

$$VIX_T^2 = -\frac{2}{\Delta} \mathbb{E} \left[\log \left(\frac{S_{T+\Delta}}{S_T} \right) \middle| \mathcal{F}_T \right] \cdot 100^2 = \frac{100^2}{\Delta} \int_T^{T+\Delta} \xi_T(u) du \quad (5.3.1)$$

with the usual $\Delta = 30$ days and $\xi_T(u) = \mathbb{E}[\sigma_u^2 | \mathcal{F}_T]$ the forward variance curve, that can be explicitly computed. In order to do that we fix $T \leq u$ and rewrite the process X as:

$$X_u = X_T e^{-(1/2-H)\epsilon^{-1}(u-T)} + \epsilon^{H-1/2} \int_T^u e^{-(1/2-H)\epsilon^{-1}(u-s)} dW_s =: Z_T^u + G_T^u$$

Thus, if we define:

$$g(u) := \mathbb{E}[p(X_u)^2]$$

we obtain the following formula:

$$\xi_T(u) = \mathbb{E}[\sigma_u^2 | \mathcal{F}_T] = \frac{\xi_0(u)}{g(u)} \mathbb{E} \left[\left(\sum_{k=0}^5 \alpha_k X_u^k \right)^2 \middle| \mathcal{F}_T \right]$$

Defining α the vector $[\alpha_0, \alpha_1, 0, \alpha_3, 0, \alpha_5, 0, 0, \dots]$ and indicating with $(\alpha * \alpha)$ the discrete convolution:

$$(\alpha * \alpha)_k = \sum_{j=0}^k \alpha_j \alpha_{k-j}$$

we have the following expression:

$$\xi_t(u) = \frac{\xi_0(u)}{g(u)} \mathbb{E} \left[\sum_{k=0}^{10} (\alpha * \alpha)_k X_u^k \mid \mathcal{F}_T \right]$$

Furthermore, making use of the binomial expansion we can improve the expression for $\xi_T(u)$ in terms of Z^u and G^u so that we have:

$$\xi_T(u) = \frac{\xi_0(u)}{g(u)} \sum_{k=0}^{10} \sum_{i=0}^k (\alpha * \alpha)_k \binom{k}{i} \left(X_T e^{-(1/2-H)\epsilon^{-1}(u-T)} \right)^i \mathbb{E}[(G_T^u)^{k-i}] \quad (5.3.2)$$

where we used both the fact that Z_T^u is \mathcal{F}_T -measurable and the independence of G_T^u from \mathcal{F}_T . Moreover, we know that G_T^u is actually a Gaussian random variable.

$$G_T^u \sim \mathcal{N} \left(0, \frac{\epsilon^{2H}}{1-2H} [1 - e^{-(1-2H)\epsilon^{-1}(u-T)}] \right)$$

We recall that the moments of a Gaussian variable $Y \sim \mathcal{N}(0, \sigma_Y^2)$ can be computed as:

$$\mathbb{E}[Y^n] = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \sigma_Y^n (n-1)!! & \text{if } n \text{ is even} \end{cases} \quad (5.3.3)$$

where $n!!$ is the double factorial. Therefore, we have an explicit expression for all the moments of G_T^u . Going back to (5.3.1) and plugging in expression (5.3.2) we have that the explicit expression of VIX_T^2 is polynomial in X_T and given by:

$$\begin{aligned} VIX_T^2 &= \frac{100^2}{\Delta} \sum_{k=0}^{10} \sum_{i=0}^k (\alpha * \alpha)_k \binom{k}{i} X_T^i \int_T^{T+\Delta} \frac{\xi_0(u)}{g(u)} \mathbb{E}[(G_T^u)^{k-i}] e^{-(1/2-H)\epsilon^{-1}(u-T)i} du \\ &= \frac{100^2}{\Delta} \sum_{i=0}^{10} X_T^i \sum_{k=i}^{10} (\alpha * \alpha)_k \binom{k}{i} \int_T^{T+\Delta} \frac{\xi_0(u)}{g(u)} \mathbb{E}[(G_T^u)^{k-i}] e^{-(1/2-H)\epsilon^{-1}(u-T)i} du \\ &= \frac{100^2}{\Delta} \sum_{i=0}^{10} \beta_i X_T^i \end{aligned} \quad (5.3.4)$$

where we have defined

$$\beta_i := \sum_{k=i}^{10} (\alpha * \alpha)_k \binom{k}{i} \int_T^{T+\Delta} \frac{\xi_0(u)}{g(u)} \mathbb{E}[(G_T^u)^{k-i}] e^{-(1/2-H)\epsilon^{-1}(u-T)i} du$$

We recall that thanks to formula (5.3.3) we can compute exactly every moment of G_T^u . We note, from formula (5.3.4), that VIX_T^2 is actually a polynomial in X_T that we will denote with $f(X_T)$. Since we have that X_T is Gaussian:

$$X_T \sim \mathcal{N}\left(0, \frac{\epsilon^{2H}}{1-2H} [1 - e^{-(1-2H)\epsilon^{-1}T}] \right)$$

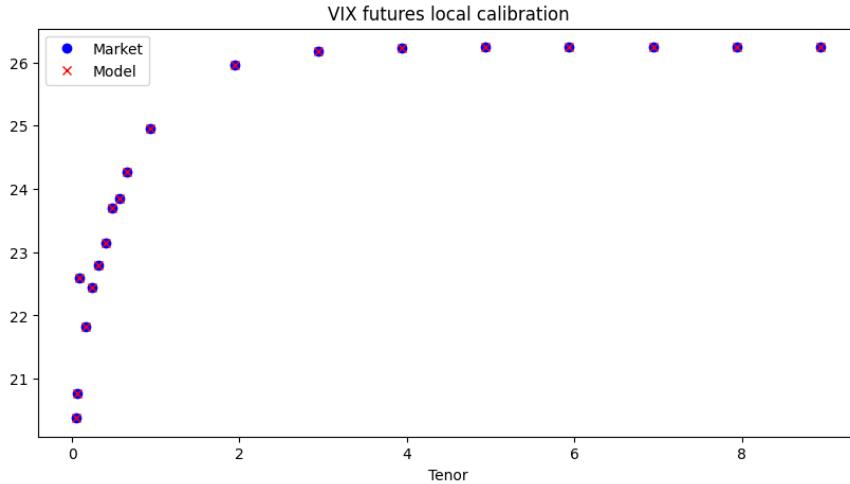
pricing VIX derivatives, with a general payoff function Φ , can be done integrating directly against the standard Gaussian density:

$$\mathbb{E}[\Phi(VIX_T)] = \mathbb{E}\left[\Phi(\sqrt{f(X_T)})\right] = \frac{1}{\sigma_{X_T}\sqrt{2\pi}} \int_{\mathbb{R}} \Phi(\sqrt{f(x)}) e^{-x^2/\sigma_{X_T}^2} dx$$

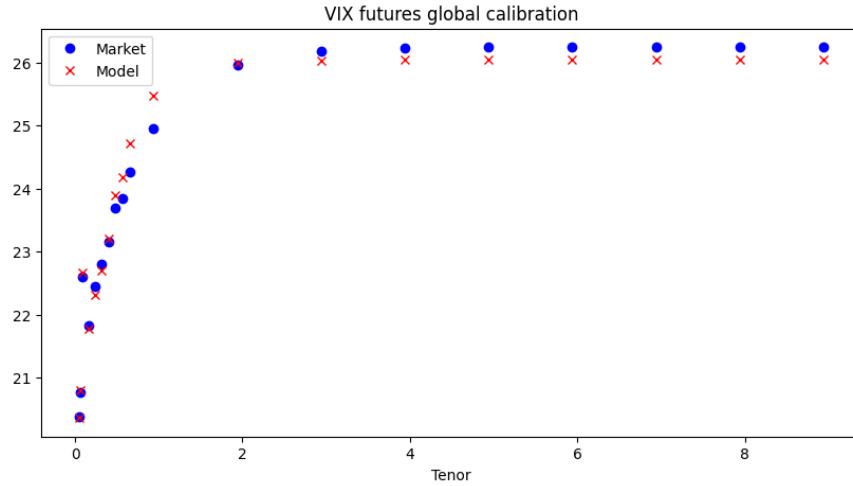
This integral can be computed efficiently using a variety of quadrature techniques.

5.3.1 Numerical results

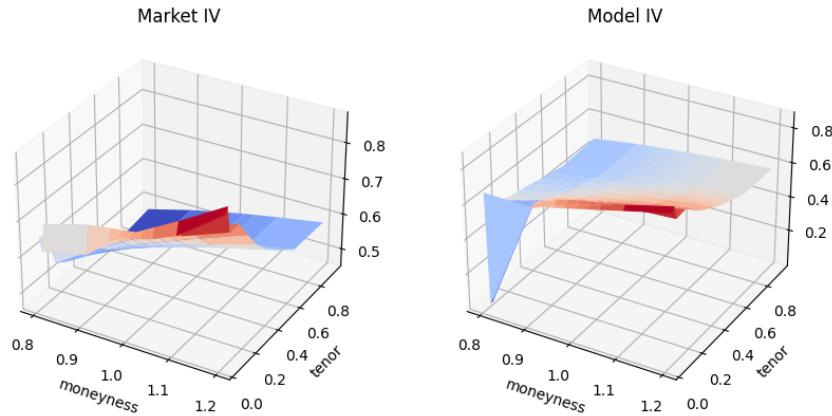
As for the SPX derivatives we used both a local and global approach for futures and options. First we report the future results. In the local case the calibration is almost perfect, while in the global case we obtained a mean percentage error of 0.6920%. The next figure is the comparison between model and market futures in the local approach.



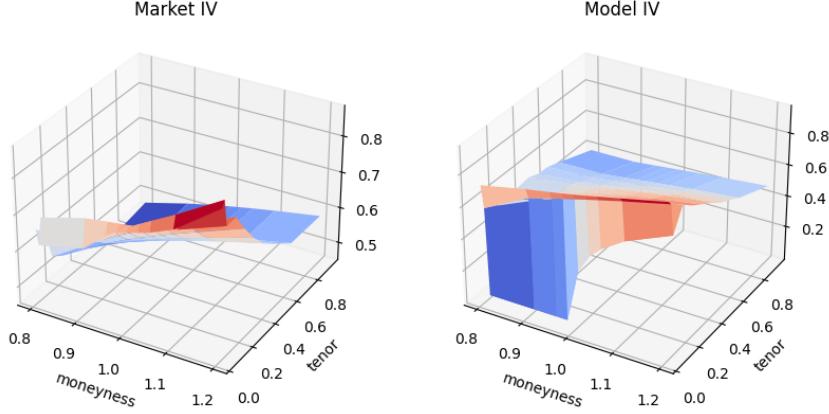
The next figure is the comparison between market and model futures in the global approach.



Calibrating the model to the IV of VIX options we discovered, during the local calibration, that the model isn't suitable for long term-options in fact we obtained a mean relative percentage error of 31.1995% if we use all the tenors, while if we use only the tenors that are less or equal to one year we obtained a mean relative percentage error of 2.4955%. The next figure is the comparison between market and model IV surfaces.



For the global approach using only the tenors that are less or equal to one year we obtained a mean relative percentage error of 8.7636%. The next figure is the comparison between market and model IV surfaces.



We note that the model tends to underestimate the IV surface.

5.4 Joint calibration

In this section we tackle the joint calibration problem that is the simultaneous calibration of the model to SPX European options, VIX European options and VIX futures across several tenors. To jointly calibrate the model we want to find the solution of the following optimisation problem:

$$\min_{\Theta} \{c_1 f_1(\Theta) + c_2 f_2(\Theta) + c_3 f_3(\Theta)\}$$

where Θ is the set of parameters and we have defined:

$$f_1(\Theta) := \sqrt{\sum_{i,j} (\sigma_{spx}^{\Theta}(K_i, T_j) - \sigma_{spx}^{mkt}(K_i, T_j))^2}$$

$$f_2(\Theta) := \sqrt{\sum_{i,j} (\sigma_{vix}^{\Theta}(K_i, T_j) - \sigma_{vix}^{mkt}(K_i, T_j))^2}$$

$$f_3(\Theta) := \sqrt{\sum_i (F_{vix}^{\Theta}(T_i) - F_{vix}^{mkt}(T_i))^2}$$

So that f_1 is the root mean squared error (RMSE) coming from the SPX options calibration, f_2 is the RMSE coming from the VIX options calibration and f_3 is the RMSE coming from the VIX futures calibration. The constants c_i are positive and reflect the weight that we want to give to each particular aspect. In our case we decided to give an equal weight to each part so that we have $c_1 = c_2 = c_3 = 1$.

5.4.1 Numerical results

For the joint calibration problem we decided to use only the global approach. The tenors that we used for the VIX, for both futures and options, are those which are less or equal to one year. Doing so we obtained the following mean relative percentage error:

- SPX options calibration: 7.7591%;
- VIX futures calibration: 0.4339%;
- VIX options calibration: 18.3786%.

What we can see is that the calibration gives us significantly worse results especially in the SPX and VIX options part. Thus, the model, at least with our data, doesn't seem to solve the joint calibration problem in a satisfying way.

5.5 SPX Parameters stability

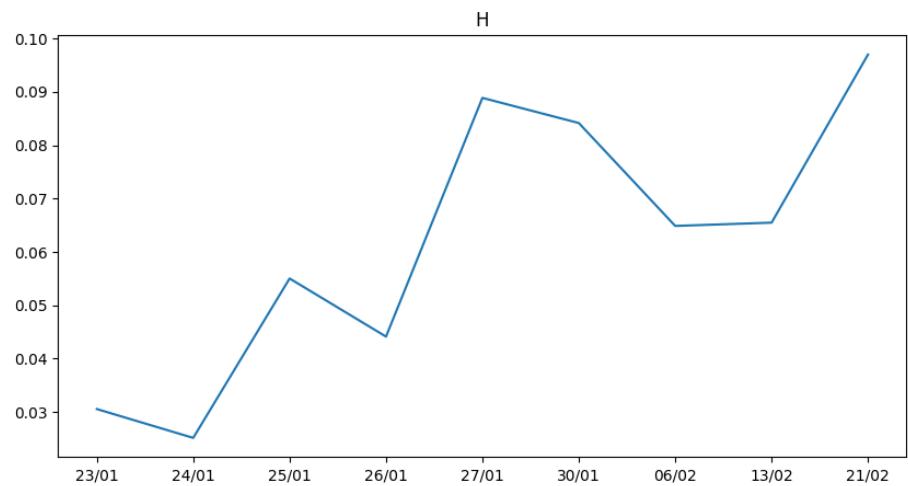
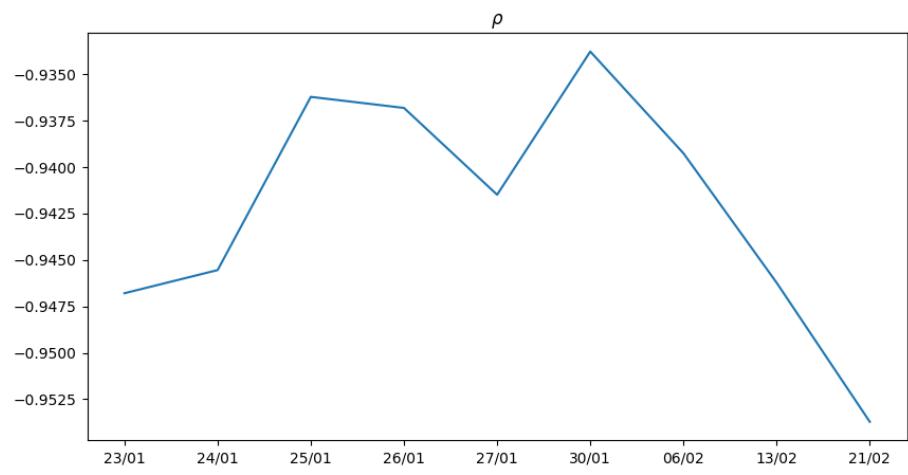
To check the stability of the model parameters we decided to test the model using data from other days than the principal one. The dates that we used are: 24/01/2023, 25/01/2023, 26/01/2023, 27/01/2023, 30/01/2023, 06/02/2023, 13/02/2023, 21/02/2023. In doing so we used only options available in all these days and they are 30 in total ranging from few months tenors to almost ten years. We did two tests:

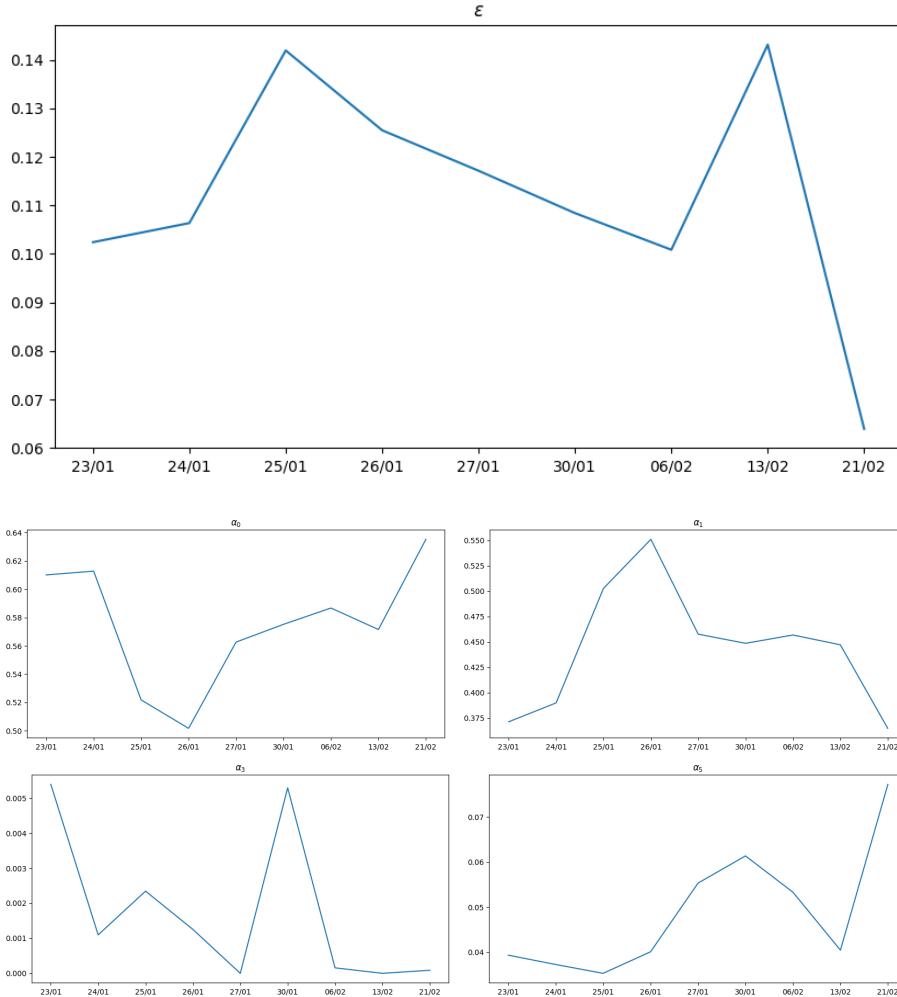
1. using the already calibrated parameters we used the model to check the mean percentage error in all these days comparing them with the error obtained calibrating the model for each;
2. we calibrated the model in each day and then analyzed the standard deviation and variance of the each parameter.

The next table contains the parameters that we obtained for each day.

Day	ρ	H	ε	α_0	α_1	α_3	α_5
23/01	-0.9468	0.0305	0.1024	0.6101	0.3713	0.0054	0.0394
24/01	-0.9455	0.0251	0.1064	0.6128	0.3899	0.0011	0.0373
25/01	-0.9362	0.0550	0.1419	0.5218	0.5026	0.0023	0.0353
26/01	-0.9368	0.0441	0.1255	0.5017	0.5509	0.0013	0.0401
27/01	-0.9415	0.0889	0.1171	0.5626	0.4575	1e-9	0.0553
30/01	-0.9338	0.0842	0.1084	0.5752	0.4485	0.0053	0.0614
06/02	-0.9393	0.0649	0.1009	0.5867	0.4567	0.0002	0.0533
13/02	-0.9462	0.0655	0.1431	0.5715	0.4471	1e-7	0.0405
21/02	-0.9537	0.0970	0.0640	0.6352	0.3648	0.0001	0.0772

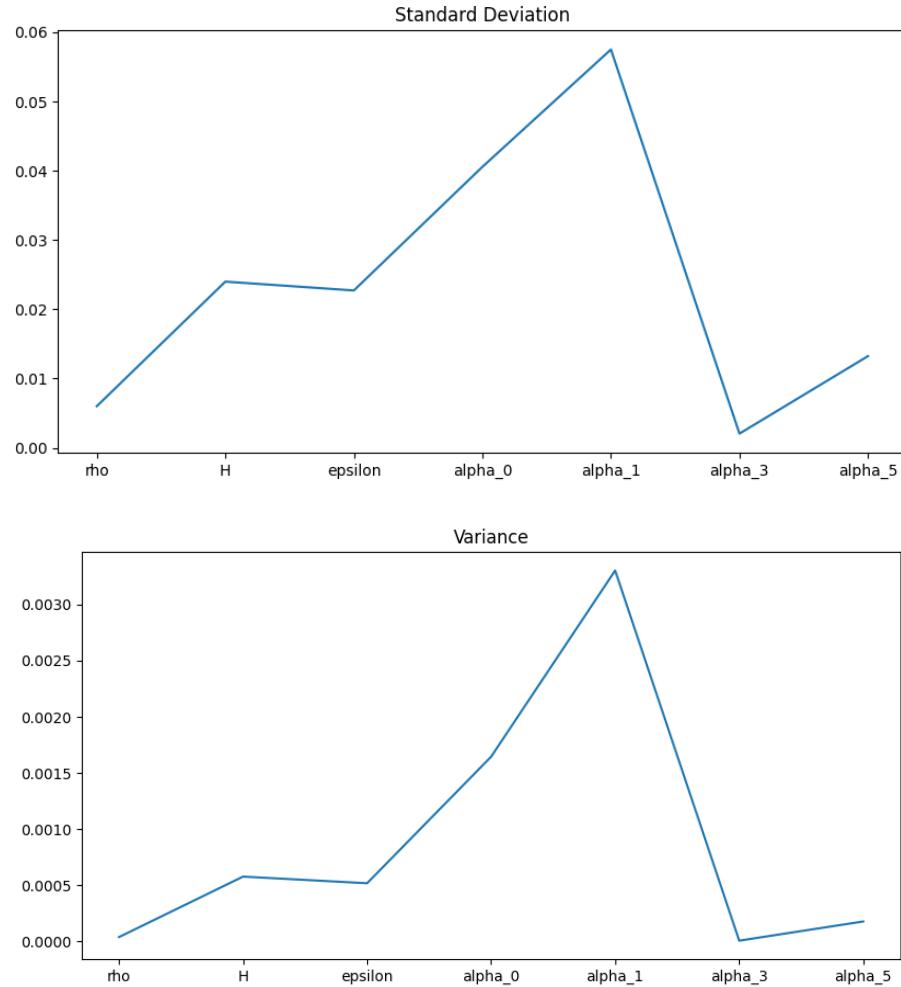
A graphical representation of the changes is presented in the next figures.





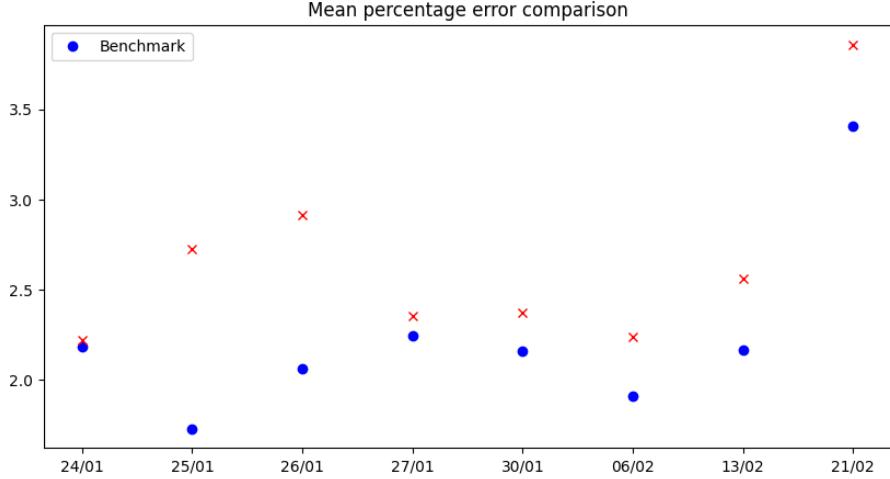
If we look at the standard deviation and variance of each parameter we obtain the following table.

	ρ	H	ε	α_0	α_1	α_3	α_5
Std	0.00602	0.02398	0.02271	0.04054	0.05747	0.00206	0.01324
Var	0.00004	0.00058	0.00052	0.00164	0.00330	4e-6	0.00018



The next table contains the comparison between the benchmark error obtained before and the one obtained using as parameters the one calibrated for the first day (23/01).

Day	Benchmark	Error
24/01	2.1833%	2.2226%
25/01	1.7296%	2.7234%
26/01	2.0642%	2.9135%
27/01	2.2429%	2.3555%
30/01	2.1602%	2.3735%
06/02	1.9085%	2.2389%
13/02	2.1663%	2.5612%
21/02	3.4100%	3.8596%



5.6 Bayesian Inverse Problem

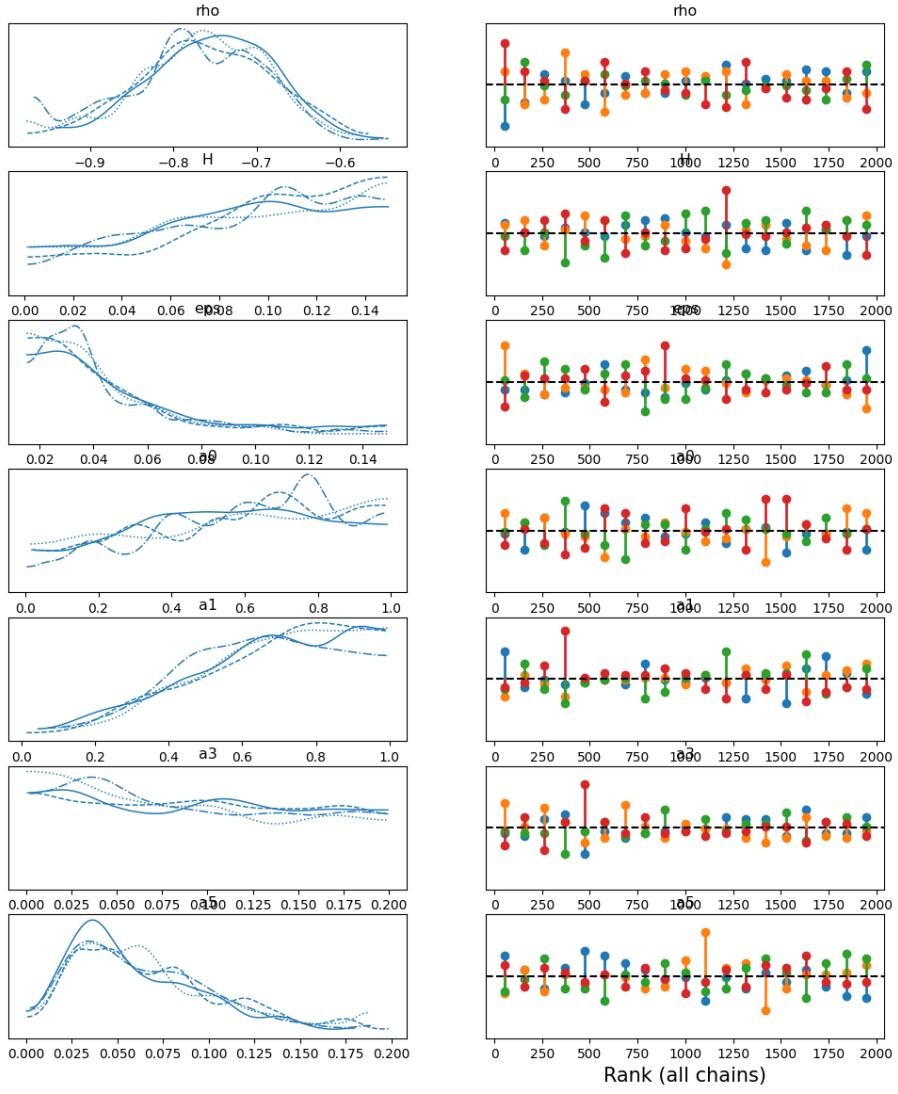
We decided to calibrate the Bayesian model only in the SPX case so that we can compare its results with the other models. The method that we used to do this calibration is again the ABC method. We decided to calibrate the model using only the tenors from 1 month to 6 months (10 in total) and we decided to approach the problem in two ways: calibrating all the parameters and calibrating only the set of parameters comprised of ρ , H and ϵ while using pre-calibrated values of the vector α . We calibrated taking as the reference data the market prices instead of the market IV and we utilized 4 chains with 2000 samples each, we resorted to this escamotage just for computational and time reasons (we are using just a standard laptop).

5.6.1 Numerical results

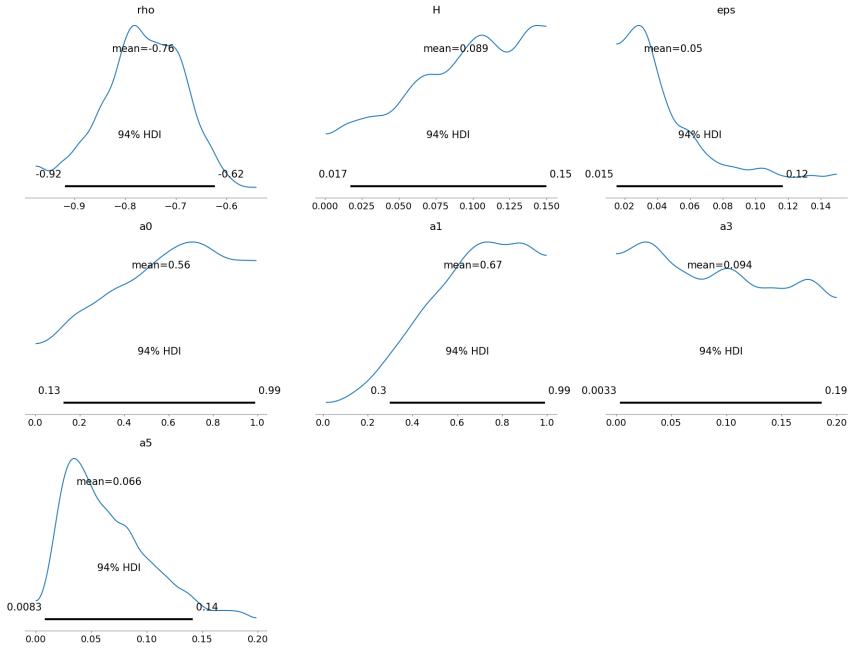
As priors for the parameters we employed:

- $\rho \sim \mathcal{U}(-1, -0.1)$
- $H \sim \mathcal{U}(0, 0.15)$
- $\epsilon \sim \mathcal{U}(0.015, 0.15)$
- $\alpha_0 \sim \mathcal{U}(0, 1)$
- $\alpha_1 \sim \mathcal{U}(0, 1)$
- $\alpha_3 \sim \mathcal{U}(0, 0.2)$
- $\alpha_5 \sim \mathcal{U}(0, 0.2)$

In the first case, where we calibrate all the parameters, we obtained the following rank plots.



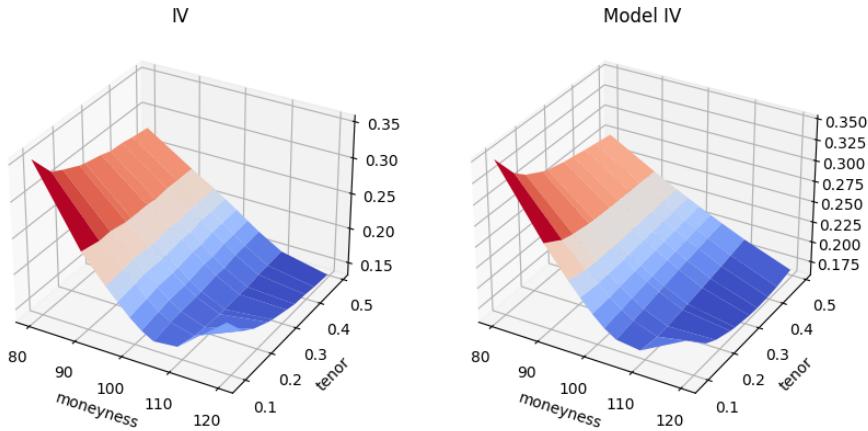
Looking at the chains plots we expect to obtain results that are not very good, since they are not very coherent, while looking at the rank plots we note that we have some problems when sampling. The next figure represents the posterior densities that we have obtained.



To compare the Bayesian calibration with the deterministic calibration we fixed the maximum a posteriori (MAP) values for each parameter:

ρ	H	ϵ	α_0	α_1	α_3	α_5
-0.7761	0.1090	0.0250	0.7630	0.6504	0.0095	0.0332

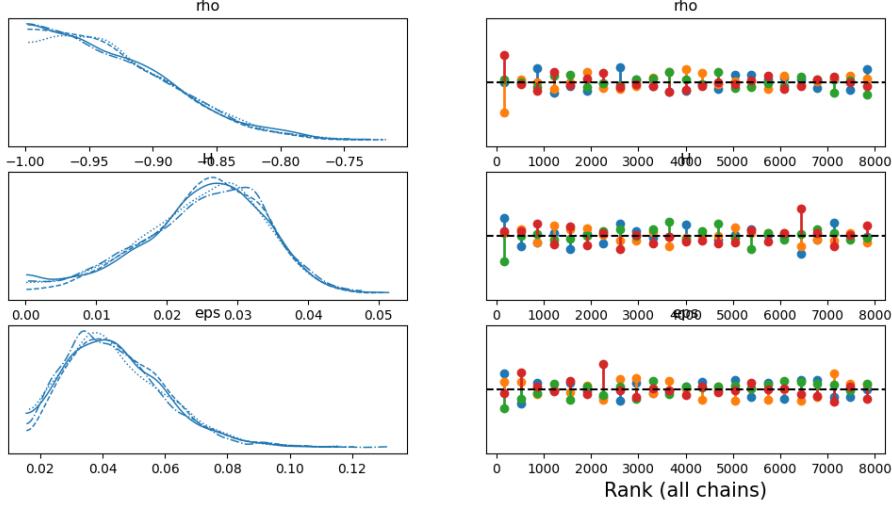
and calibrating the model we achieved a mean relative percentage error of 10.6836%. The next figure is a comparison between the market and model IV surfaces.



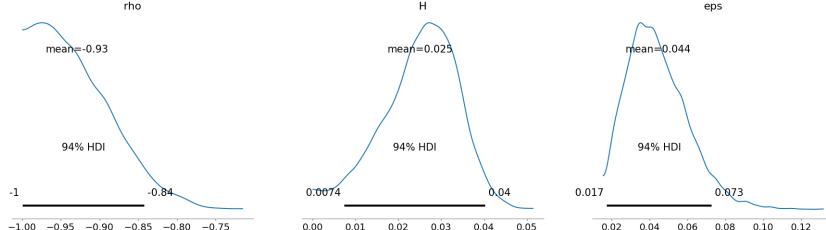
In the second case we fixed the value of the vector α obtained through a deterministic calibration:

$$\alpha_0 = 1.2237 \quad \alpha_1 = 0.4503 \quad \alpha_3 = 0.0361 \quad \alpha_5 = 0.0532$$

Then we calibrated the other three parameters and got the following chain and rank plots.



We note that this time the chains are more coherent and looking at the rank plots we have almost no problems in the sampling. The next figure presents the posterior densities that we got.

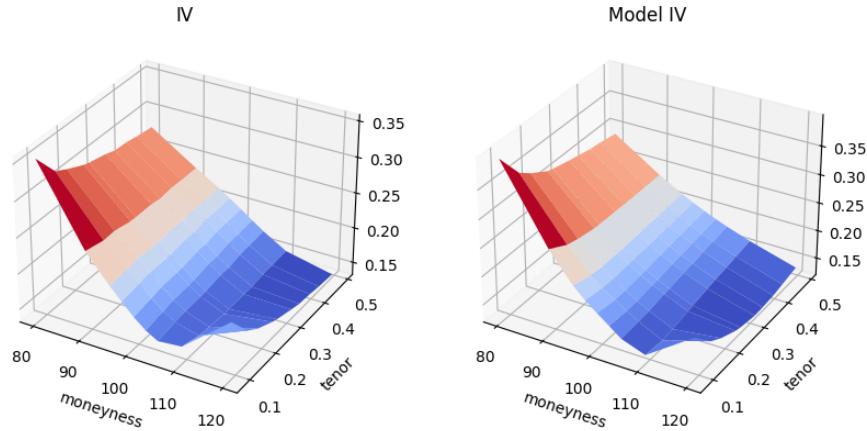


Posterior densities

As before, in order to compare the Bayesian calibration with the deterministic calibration we fixed the MAP values:

ρ	H	ϵ
-0.9635	0.0264	0.0333

and calibrating the model we achieved a mean relative percentage error of 3.7385%. The next figure is the comparison between market and model IV surfaces.



Thus, we can conclude that fixing the α parameter the Bayesian calibration is able to give us some good quality results. Whereas, if the α parameter has to be calibrated the Bayesian framework is not able to give us good results since the set of parameters is not completely independent.

6 Other models

Before the last chapter, in which we will develop a Bayesian framework to quantify rough volatility calibration risk, we decided to analyze two other widely used models: the Heston model and the rough Heston (rHeston) model. In the next sections we introduce the models and briefly present the results obtained, both for the deterministic and Bayesian calibration. To have a more in depth explanation of these two models you can refer to [24].

6.1 Heston

The Heston model is widely used in the industry due to its ability to capture some important features of low-frequency asset price movements:

1. volatility mean reversion;
2. volatility clustering and persistence;
3. negative correlation between prices and volatility;
4. skew and term structure of volatility.

Let $(\Omega, \{(F_t)_{t \geq 0}\}, \mathbb{P})$ be a complete filtered probability space and call \mathbb{P} the *physical measure*. Given a stock price process $S = (S_t)_{t \geq 0}$ the Heston model, under \mathbb{P} , is given as:

$$\begin{cases} dS_t = \mu S_t dt + S_t \sqrt{v_t} dW_t \\ dv_t = \kappa(\eta - v_t)dt + \theta \sqrt{v_t} d\tilde{W}_t \\ v_0 = \sigma_0^2 \end{cases}$$

where:

- μ is the drift of the stock returns;
- $W = (W_t)_{t \geq 0}$ and $\tilde{W} = (\tilde{W}_t)_{t \geq 0}$ are two correlated Brownian motions with $d\langle W, \tilde{W} \rangle_t = \rho dt$ and $\rho \in [-1, 1]$;
- $\sigma_0 > 0$ is the initial volatility;
- $\eta > 0$ is the long run variance;
- $\kappa > 0$ is the mean reversion rate;
- $\theta > 0$ is the volatility of the volatility.

The variance process is strictly positive if $2\kappa\eta > \theta^2$ (condition only sufficient, not necessary). This is known as *Feller condition*. To price derivatives we need to have to switch to an equivalent martingale measure \mathbb{Q} that we will call the *pricing measure*. We can choose \mathbb{Q} in a lot of different ways, but we will use the *minimal martingale measure* that solves the minimization problem:

$$\mathbb{Q}^M = \arg \min_{\mathbb{Q} \in \mathcal{M}} \mathbb{H}(\mathbb{Q} \parallel \mathbb{P})$$

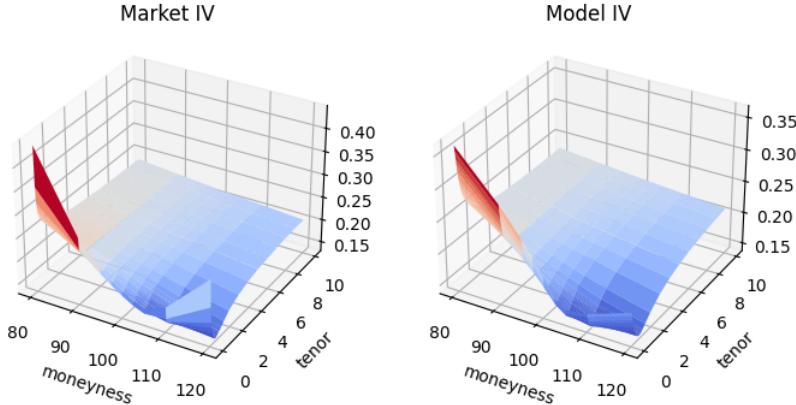
where \mathcal{M} is the set of equivalent martingale measures and \mathbb{H} is the reverse relative entropy. Thus, the Heston model under \mathbb{Q} is written as:

$$\begin{cases} dS_t = (r - q)S_t dt + S_t \sqrt{v_t} dW_t \\ dv_t = \kappa(\eta - v_t)dt + \theta \sqrt{v_t} d\tilde{W}_t \\ v_0 = \sigma_0^2 \end{cases}$$

with $W = (W_t)_{t \geq 0}$ and $\tilde{W} = (\tilde{W}_t)_{t \geq 0}$ are two correlated \mathbb{Q} Brownian motions with $d\langle W, \tilde{W} \rangle_t = \rho dt$. In order to calibrate the model we need a way to price vanilla options which usually requires integrating the probability density function. However, since we can find the characteristic function, which is its Fourier transform, we can leverage the *Fourier Cosine Expansion* method. The computational speed makes this integration method state-of-the-art for calibration at financial institutions. For more information on how to tackle this model refer to [THESIS PROSE]. In order to find the best set of parameters we used a least squared technique and obtained:

ρ	θ	κ	η	σ_0
-0.6766	1.3231	2.6523	0.0568	0.0442

The next figure is a comparison between the model and market IV.



The mean relative percentage error that we obtained is 4.5817%. We observe that the parameters do not satisfy the Feller condition which is just a sufficient condition to have the variance process always positive. To calibrate the Bayesian Heston model we used the ABC method. In this case, since the evaluation function is quite fast, we used all the tenors. We used 4 chains with 2000 samples each and as the reference data the matrix of the market options prices. First we tried to sample simultaneously all the parameters, but this didn't give good

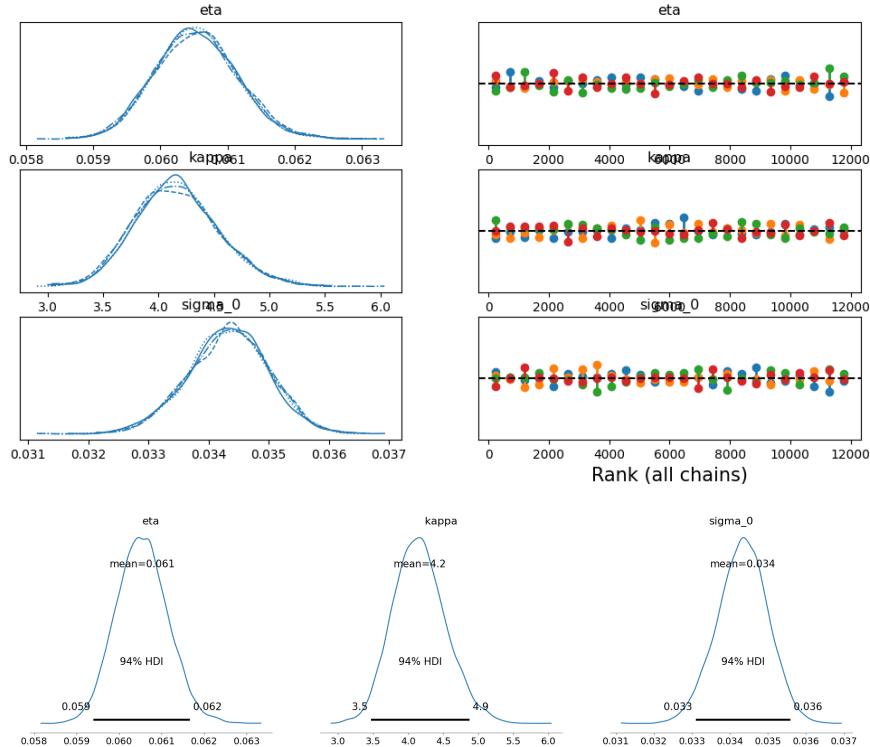
results. We expected so since the parameters aren't completely independent of one another. To address this problem we split the set of parameters in two. The first set is composed of η , κ and σ_0 , while in the second one there are θ and ρ . The first set is calibrated in order to match the market volatility $\hat{\sigma}_T$ of variance swaps with tenor T using:

$$\sigma_T = \sqrt{\eta + (\sigma_0 - \eta) \frac{1 - e^{-\kappa T}}{\kappa T}}$$

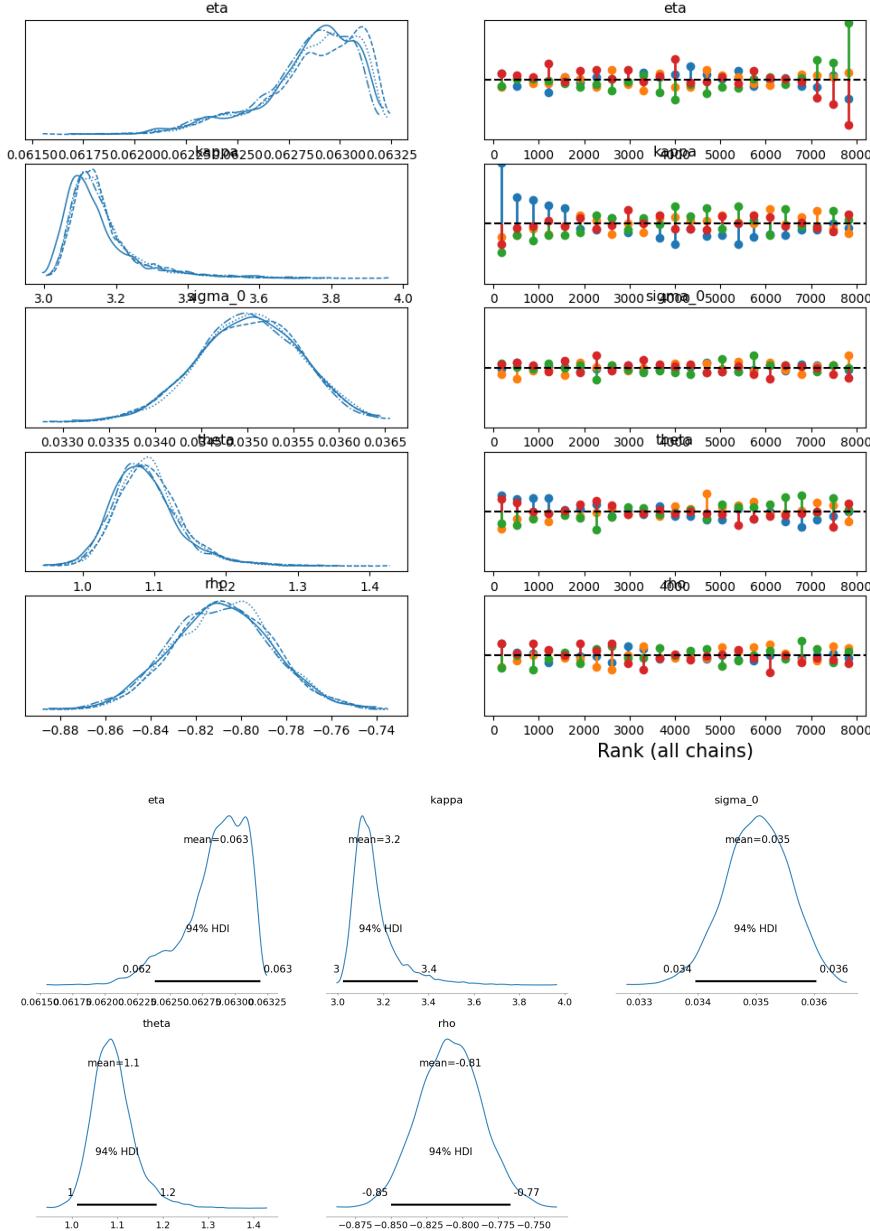
While in the second calibration we used the posterior densities obtained from the first calibration and two uninformative priors for the other two parameters. The priors that we used are:

- $\eta \sim \mathcal{U}(0, 2)$
- $\kappa \sim \mathcal{U}(1, 7)$
- $\sigma_0 \sim \mathcal{U}(0, 2)$
- $\theta \sim \mathcal{U}(0.5, 3)$
- $\rho \sim \mathcal{U}(-1, 0)$

After the first calibration we obtained the following rank plots and posteriors.



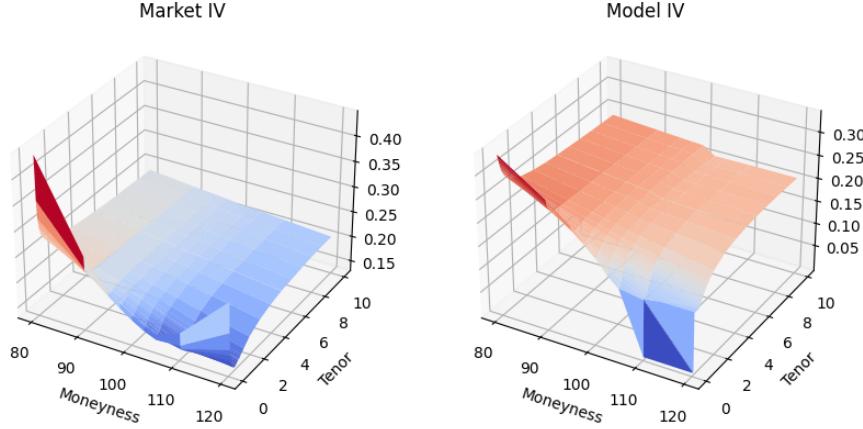
After the second calibration the final rank plots and posteriors are presented in the next two figures.



Fixing the MAP value for each parameter:

η	κ	σ_0	θ	ρ
0.0630	3.1113	0.0350	1.0864	-0.8099

and calibrating the model we obtained a mean relative percentage error of 4.5024%.



We note that the model fails for the first two tenors, especially in the high moneyness range, while it is quite good for the other cases.

6.2 rHeston

The Heston model can reproduce several important features of low frequency price data, but if we want to surpass that we have to build a model which can reproduce the stylized facts of modern electronic markets in the context of high frequency trading. There are 4 main stylized facts that we can observe in this type of market data:

- Markets are highly endogenous, which means that most of the orders have no real economic motivation, but are simply the reaction of algorithms to other orders.
- Markets at high frequency are much more efficient than at lower frequencies.
- There is some asymmetry in the liquidity on the bid and the ask side of the order book: a market-maker is more inclined to raise the price by less following a buy order than to lower the price following the same size sell order. This can be explained in a straight forward way: hedging the first type of position is easier than the second and market-makers usually have some inventory.
- Most of the transactions are due to big orders, called metaorders, which are not executed at once, but split in time. In fact, one of the most critical part of every trading strategy is to find the right way to execute it in large volumes without changing too much the state of the market.

Another important fact that has been empirically observed is that the volatility is rough. The model that tries to incorporate these stylized facts and the roughness of the volatility is the so called rough Heston (rHeston) model which is an evolution of the Heston model. In practice the rHeston model is what arise from taking the limit of Hawkes processes. Given a stock price process $S = (S_t)_{t \geq 0}$ the rHeston model, under the *risk-free probability measure* \mathbb{Q} , is expressed as:

$$\begin{cases} dS_t = (r - q)S_t dt + S_t \sqrt{v_t} d\tilde{W}_t \\ v_t = v_0 + \frac{\lambda}{\Gamma(H + \frac{1}{2})} \int_0^t \frac{\gamma(s) - v_s}{(t-s)^{\frac{1}{2}-H}} ds + \frac{\theta}{\Gamma(H + \frac{1}{2})} \int_0^t \frac{\sqrt{v_s}}{(t-s)^{\frac{1}{2}-H}} dW_s \end{cases}$$

where:

- r is the risk-free rate;
- q is the yield of the underlying;
- W and \tilde{W} are correlated Brownian motions with $d\langle \tilde{W}, W \rangle_t = \rho dt$;
- $H \in (0, 1/2)$ is the Hurst exponent of the fractional Brownian motion;
- θ is the volatility of the volatility;
- $\lambda \geq 0$ is a constant representing the "speed" of the mean reversion;
- $\gamma(\cdot)$ is a positive F_0 -measurable function representing the mean reversion level for the volatility.

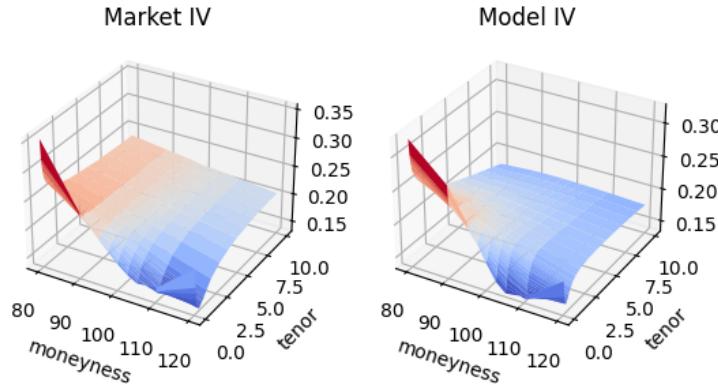
In *Perfect hedging in rough Heston models* El Euch and Rosenbaum showed that there is a link between $\lambda\gamma(\cdot)$ and the forward variance curve. Using this fact and assuming that λ is sufficiently small, we can rewrite the dynamic in a more compact way as:

$$\begin{cases} dS_t = (r - q)S_t dt + S_t \sqrt{v_t} \{ \rho dW_t + \sqrt{1 - \rho^2} dW_t^\perp \} \\ v_t = \xi_0(t) + \frac{\theta}{\Gamma(H + \frac{1}{2})} \int_0^t \frac{\sqrt{v_s}}{(t-s)^{\frac{1}{2}-H}} dW_s \end{cases}$$

the hypothesis that λ must be sufficiently small is sensible since the volatility should be slowly mean reverting. In order to price derivatives we will use again the characteristic function, of which we have a quasi-closed form, to obtain the solution to the fractional equation. Unfortunately, this solution is not known, so we resort to use the Padé approximants to obtain a fast and reliable approximation. Following the work of Gatheral and Radoicic in *Rational approximation of the rough Heston solution* we can derive an expansion for small and long times for the characteristic function and then the corresponding Padé rational expansion to match both expansions. Now that we have a good approximation of the characteristic formula we can apply the Lewis's formula to evaluate the price of an european call option. To find the set of parameters we will use a least squares approach fitting to the market IV. The set obtained this way is:

ρ	θ	H
-0.6995	0.2817	0.0010

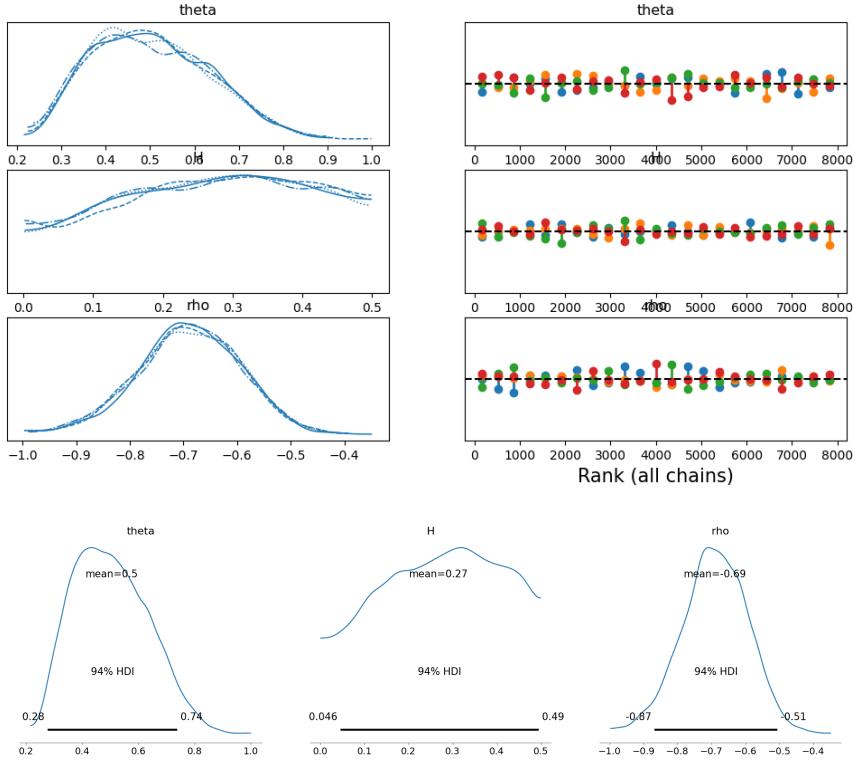
The next figure is the comparison between the market IV surface and the one obtained after the calibration of the model.



The mean relative percentage error that we obtained is of 6.4480%. As we can see we were able to have similar values for shorter tenors, meanwhile for longer tenors the model seems to be underestimating the IV. To calibrate the Bayesian rHeston model we use the ABC method, as before we used 4 chains with 2000 samples each. In this case, for computational reasons, we didn't use all the tenors, but only the ones greater than one month and less than six. Unlike in the Heston model here we calibrated all the parameters simultaneously. However, we decided to conduct two analyses as in the rBergomi model: the first calibrating all the parameters and the second in which we fixed the Hurst parameter H and calibrated the remaining two. The priors that we used are:

- $\rho \sim \mathcal{U}(-1, 0)$
- $\theta \sim \mathcal{U}(0, 3)$
- $H \sim \mathcal{U}(0, 0.5)$

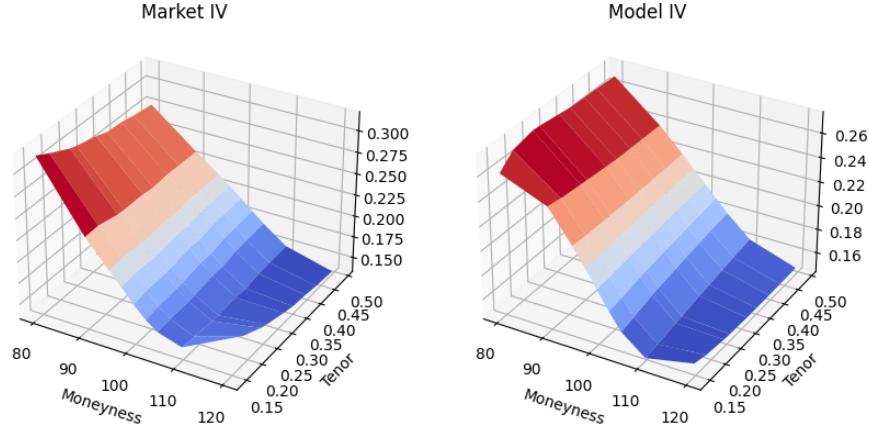
In the first case the rank plots and posterior densities that we obtained are as follow.



Looking at the rank plots we note that we don't suffer from excessive over/under sampling. Fixing the MAP value for each parameter:

ρ	θ	H
-0.7120	0.4354	0.3133

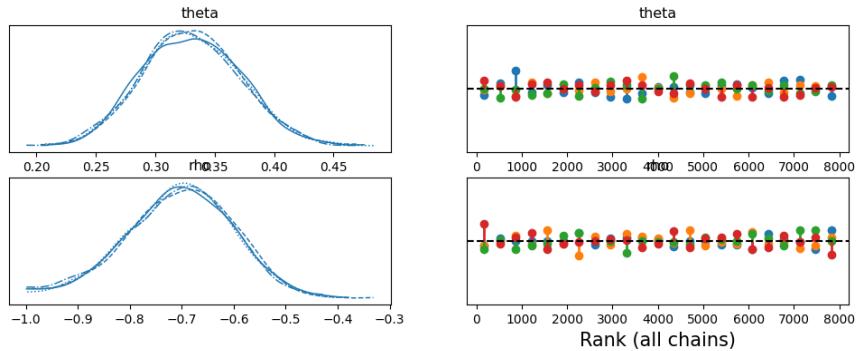
and calibrating the model we obtained a mean relative percentage error of 4.3099%. The next figure is the comparison between the market and model IV surfaces.

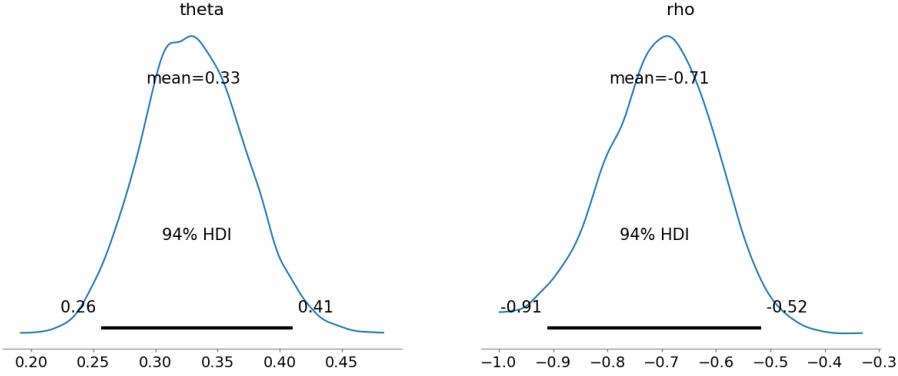


In the second case we first calibrated the deterministic model, with a least squares approach, using the same options that we will use in the Bayesian part. The resulting set of parameters is:

ρ	θ	H
-0.6780	0.3460	0.0460

that obtains a mean relative percentage error of 2.3024%. Fixing H and calibrating the other two parameters, using the same priors as before, we have the following rank plots and posteriors.

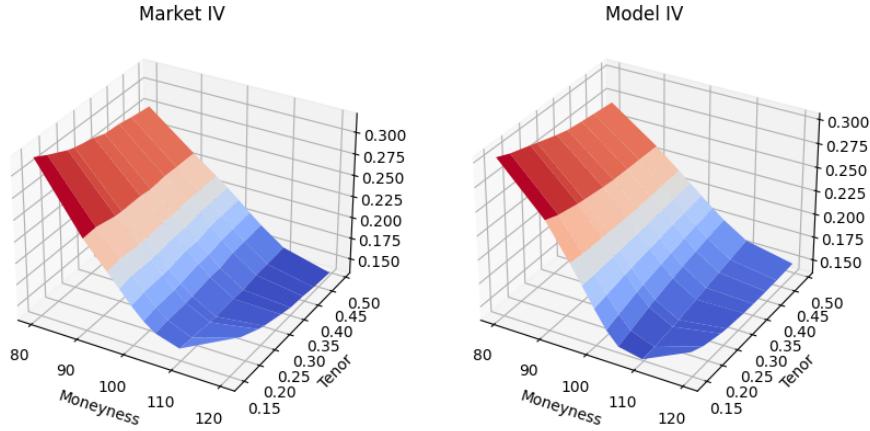




We note that also in this case we don't have problems in the sampling part.
Fixing the MAP values:

ρ	θ
-0.6874	0.3317

and calibrating the model we obtained a mean relative percentage error of 2.2626%. We note that both the error and the MAP parameters are very close to that obtained using the deterministic calibration. The next figure is the comparison between the market and model IV surfaces.

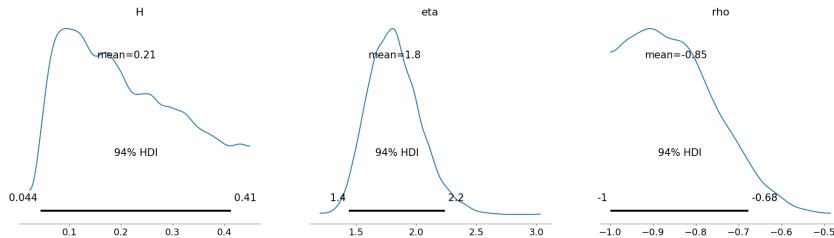


7 Bayesian Risk Quantification

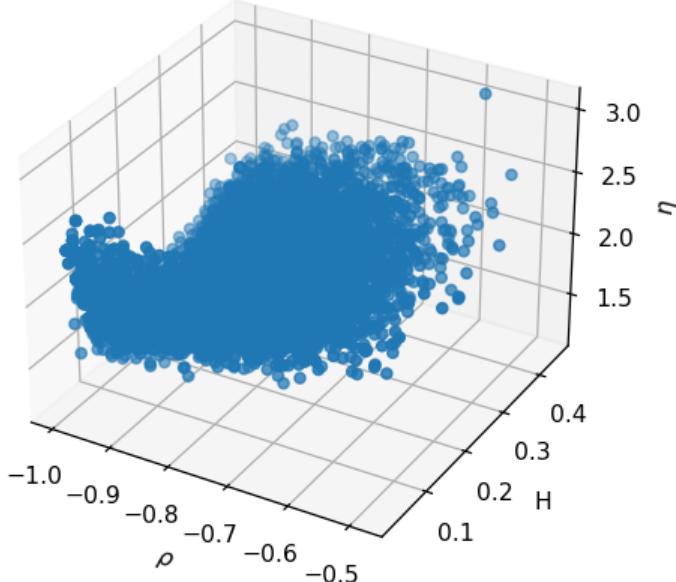
In this last chapter we will create a Bayesian framework to analyze the risk associated with the calibration of the models proposed. In order to do that, for each model, we extracted the samples produced by all the chains so that we will use only the sets of parameters that the Bayesian calibration deemed to be best and working. Then we calibrated the model using all these sets and computed the mean relative percentage error of all of these different calibrations, in our case, since we have 4 chains with 2000 samples each, we have a grand total of 8000 errors. To put all these results together we decided to weight each error with the likelihood of the parameters set that generated it. To find how likely each set is we tessellated the space of parameters, doing so we can approximate the probability of each area using the ratio of points inside over the total number of points. Thus, we weight the errors produced by samples in the same area in the same way. The problem is to find the right grid: if it is too fine then we will obtain just the mean of all the errors while if it is too loose we will obtain an abnormal error. In doing so what we will obtain is a sort of expected error of the Bayesian calibration. To quantify the model risk we will compare it with the error obtained from the deterministic calibration of the corresponding model, which should be lower, and use as an indicator the difference between the two. This risk indicator shouldn't be used as a mean to compare different models in an absolute way, but is just an indicator associated to the risk of calibrating the model in a Bayesian way instead of the classical deterministic calibration.

7.1 rBergomi

In the rBergomi model we did two analyses: one for the model in which we calibrated all the parameters and one for the model in which we fixed the Hurst parameter. We report the posterior densities that we obtained in the first case.



The next figure is the scatter plot of the parameter samples produced by the Bayesian calibration.



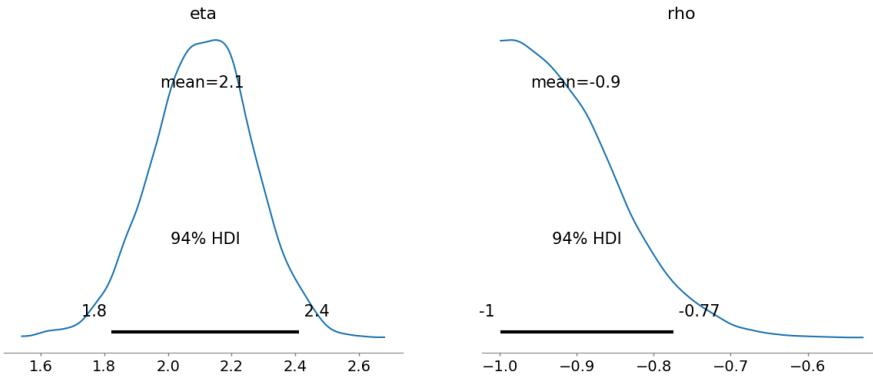
We calibrated the model using these samples and computed the mean relative percentage error. What we obtained is summed up in the next table.

Min	Max	Mean	Std	Var
1.5554%	15.4407%	5.2645%	2.1403	4.5807

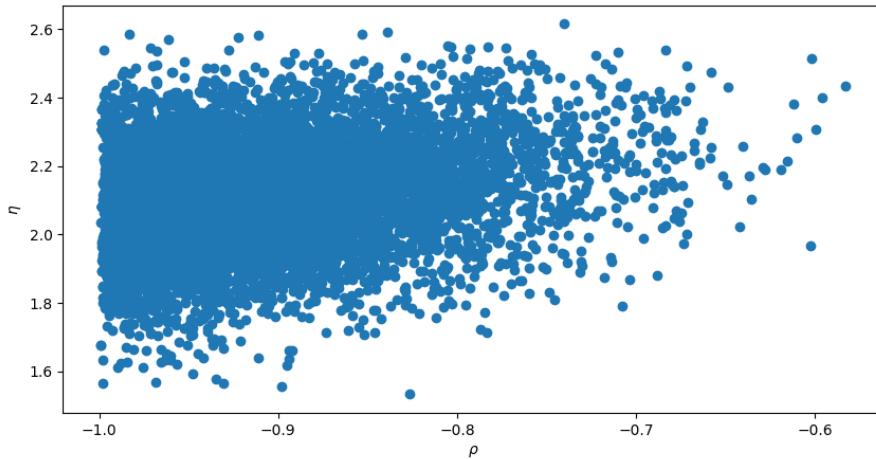
The set of parameters that minimize the error is:

ρ	H	η
-0.9313	0.0628	2.0303

In this case we divided each parameter in 150 bins. The weighted error that we obtained is of 5.4899% while the one from the deterministic calibration was of 2.0700%. Thus, the risk indicator associated with this model is 3.4199. In the second case, where we have fixed the Hurst parameter H , the posterior densities that we obtained are:



The next figure is the scatter plot of the parameter samples produced by the Bayesian calibration.



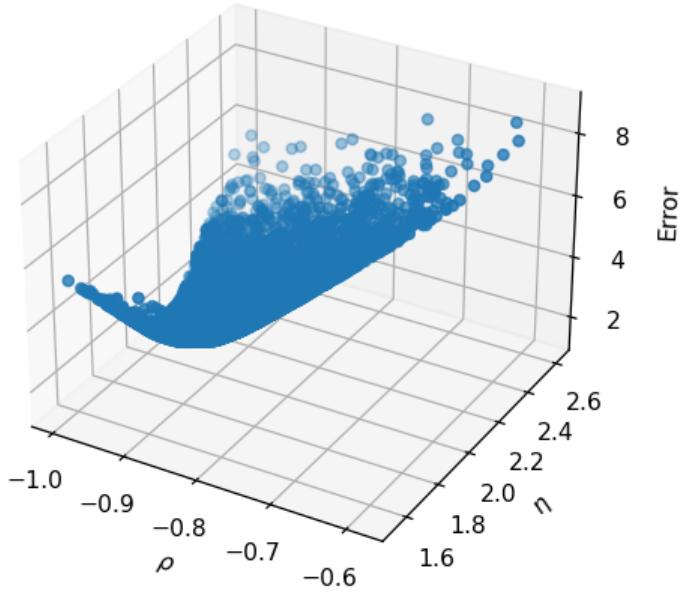
Calibrating the models we obtained the following error table.

Min	Max	Mean	Std	Var
1.5438%	9.1181%	2.6356%	1.0563	1.1157

The set of parameters that minimize the error is:

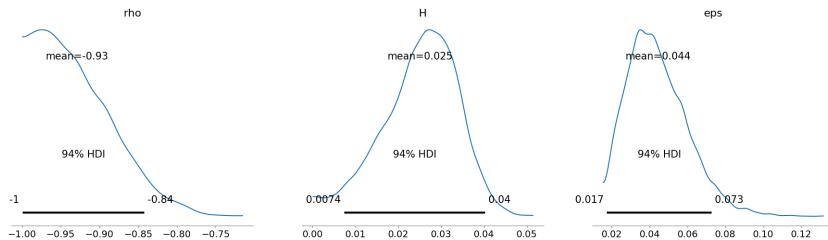
ρ	η
-0.9556	2.0981

In this case we divided each parameter in 400 bins. The weighted error that we obtained is of 3.1402%. Thus, the risk indicator associated with this model is of 1.0702. The next figure is the scatter plot of the error distribution.

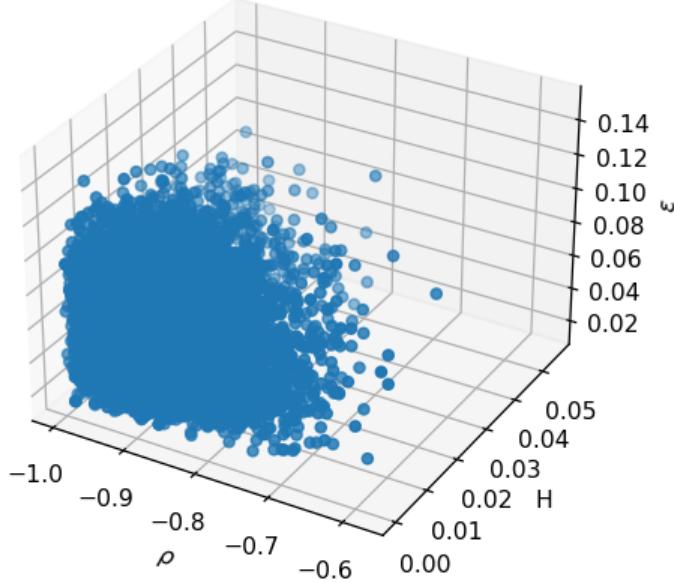


7.2 Quintic OU

For the Quintic OU model we analyzed only the case in which we fixed the α parameter. We report the posterior that we obtained via the Bayesian calibration.



The next figure is a scatter plot of the samples that we obtained.



Computing the errors we got the following table:

Min	Max	Mean	Std	Var
2.0847%	8.7843%	3.6687%	0.9051	0.8192

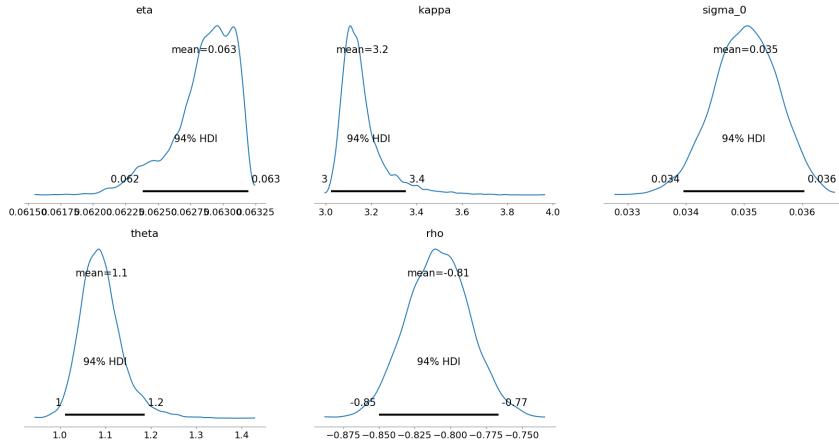
The set of parameters that minimizes the error is:

ρ	H	ε
-0.9074	0.0013	0.0324

In this case, since we have three parameters, we divided each parameter in 150 bins. The weighted error that we obtained is of 4.6547% while the one from the deterministic calibration was of 2.2917%. Thus, the risk indicator associated with this model is 2.3630.

7.3 Heston

We report the posterior densities that we obtained with the Bayesian calibration.



Computing the errors we obtained the following table:

Min	Max	Mean	Std	Var
3.2237%	6.3006%	4.4122%	0.4170	0.1739

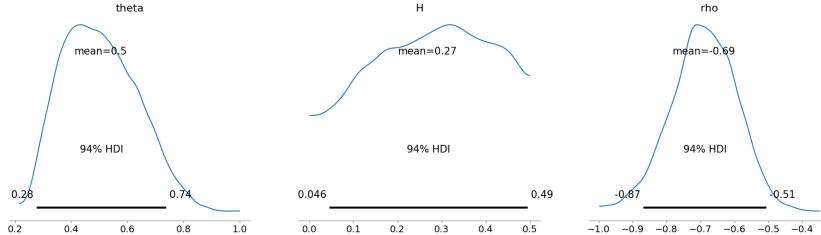
The set of parameters that minimizes the error is:

ρ	θ	κ	η	σ_0
-0.7352	1.1473	3.1583	0.0629	0.0360

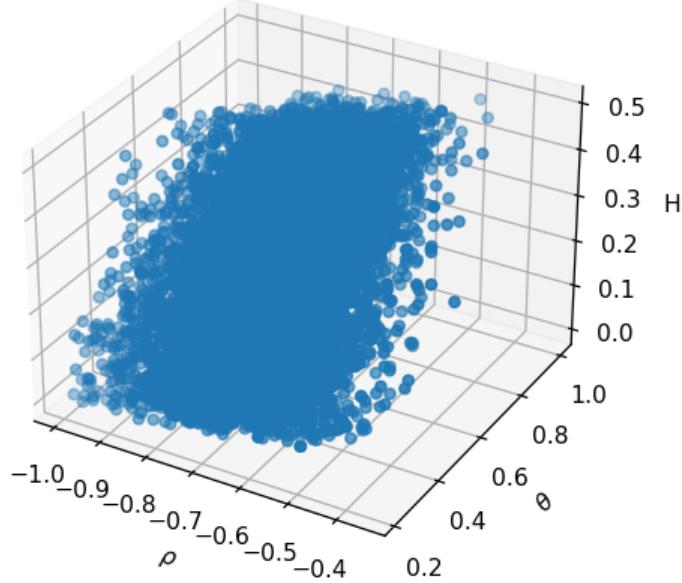
In this case we divided each parameter in 40 bins. The weighted error that we obtained is of 4.6026% while the deterministic one was 4.5817%. Thus, the risk indicator associated with this model is 0.0209.

7.4 rHeston

For the rHeston model, as for the rBergomi model, we analyzed both the case in which we calibrated all the parameters and the one in which we fixed the Hurst index H . We report the posterior densities that we obtained with the Bayesian calibration in the first case.



The next figure is a scatter plot of the samples that we are going to use.



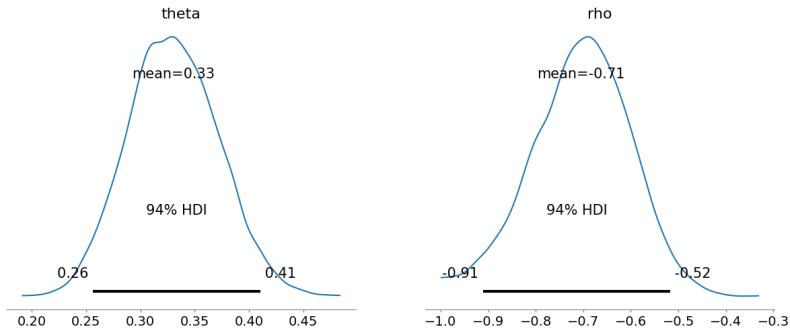
Computing the errors we obtained the following table:

Min	Max	Mean	Std	Var
1.9191%	18.2551%	3.6416%	1.6020	2.5663

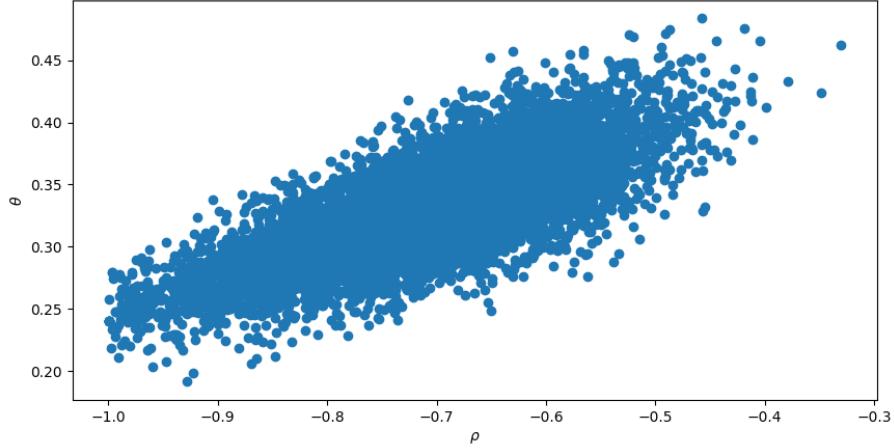
The set of parameters that minimizes the error is the following

ρ	θ	H
-0.6938	0.4938	0.2739

In this case we divided each parameter in 150 bins. The weighted error that we obtained is of 3.8465% while the deterministic one was of 2.3024%. Thus, the risk indicator associated with this model is of 1.5441. Now we report the posterior densities that we got in the second case when we fixed H .



The next figure is a scatter plot of the samples that we are going to use.



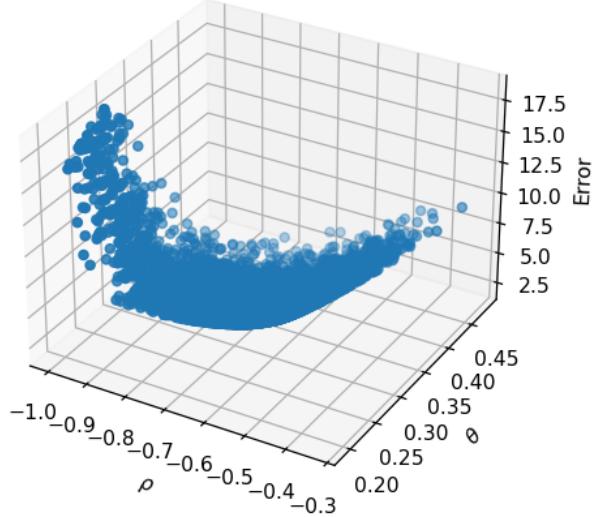
Computing the errors we obtained the following table:

Min	Max	Mean	Std	Var
2.2422%	18.2902%	3.8615%	2.0462	4.1868

The set of parameters that minimizes the error is the following

ρ	θ
-0.6762	0.3378

In this case we divided each parameter in 400 bins. The weighted error that we obtained is of 4.8054%. Thus, the risk indicator associated with this model is of 2.5030. The next figure is the scatter plot of the error distribution.



8 Conclusion

References

1. Kolmogorov, A.N.: Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C.R.(Doklady) Acad. URSS (N.S)* 26, 115–118, 1940.
2. Yaglom, A.M.: Correlation theory of processes with random stationary nth increments. *AMS Transl.* 2, 8, 87-141, 1958.
3. B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, Vol. 10, No. 4, pp. 422-437, Oct. 1968.
4. Dieker, T. (2004), Simulation of Fractional Brownian Motion, Master Thesis, Vrije Universiteit, Amsterdam.
5. M. Bennedsen, A. Lunde and M. S. Pakkanen. Hybrid scheme for Brownian semistationary process, 2015.
6. O.E. Barndorff-Nielsen and J. Schmiegel. Brownian semistationary processes and volatility/intermittency. *Advanced Financial Modelling*, de Gruyter, pp. 1-26, 2009.
7. L. Bergomi: Smile dynamics II. *Risk*, 10, pp. 67-73, 2005.
8. C. Bayer, P. Friz and J. Gatheral: Pricing under rough volatility. *Quantitative Finance* 16(6), 887–904, 2016.
9. R. McCrickerd, M. Pakkanen: Turbocharging Monte Carlo pricing for the rough Bergomi model, 2018.
10. D. Kraft. A software package for sequential quadratic programming. *Tech. Rep. DFVLR-FB*, Vol. 88-28, DLR German Aerospace Center, Institute for Flight Mechanics, Koln, Germany, 1988.
11. J. Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. Presentation at Global Derivatives, 2004.
12. J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1): 59–71, 2014.
13. J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is rough. Available at arXiv:1410.3394, 2014.
14. L. Bergomi. Stochastic Volatility Modeling. 2015.
15. Nadjahi et al., Generalized Sliced Wasserstein Distances, 2019
16. Bernton et al., On parameter estimation with the Wasserstein distance, 2019.

17. Toni T et al. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, 2008
18. Lintusaari J et al. Fundamentals and Recent Developments in Approximate Bayesian Computation, 2017
19. Eduardo Abi Jaber, Camille Illand, and Shaun Xiaoyuan Li. The quintic Ornstein-Uhlenbeck volatility model that jointly calibrates SPX & VIX smiles, 2023
20. Eduardo Abi Jaber, Camille Illand, and Shaun Xiaoyuan Li. Joint SPX VIX calibration with gaussian polynomial volatility models: deep pricing with quantization hints, 2022.
21. Jean-Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Solna. Multiscale stochastic volatility asymptotics. *Multiscale Modeling & Simulation*, 2(1):22–42, 2003.
22. Serguei Mechkov. Fast-reversion limit of the Heston model, 2015.
23. Eduardo Abi Jaber and Nathan De Carvalho. Reconciling rough volatility with jumps, 2023
24. Lorenzo Proserpio, Bayesian inference of stochastic volatility models used by market participants, 2023