ON VIX FUTURES IN THE ROUGH BERGOMI MODEL

ANTOINE JACQUIER, CLAUDE MARTINI, AND AITOR MUGURUZA

ABSTRACT. The rough Bergomi model introduced by Bayer, Friz and Gatheral [3] has been outperforming conventional Markovian stochastic volatility models by reproducing implied volatility smiles in a very realistic manner, in particular for short maturities. We investigate here the dynamics of the VIX and the forward variance curve generated by this model, and develop efficient pricing algorithms for VIX futures and options. We further analyse the validity of the rough Bergomi model to jointly describe the VIX and the SPX, and present a joint calibration algorithm based on the hybrid scheme by Bennedsen, Lunde and Pakkanen [4].

1. INTRODUCTION

Volatility, though not directly observed nor traded, is a fundamental object on financial markets, and has been the centre of attention of decades of theoretical and practical research, both to estimate it and to use it for trading purposes. The former goal has usually been carried out under the historical measure ($\mathbb{P}$) while the latter, through the introduction of volatility derivatives (VIX and related family), has been evolving under the pricing measure $\mathbb{Q}$. Most models used for pricing purposes (Heston [19], SABR [17], Bergomi [5]) are constructed under $\mathbb{Q}$ and are of Markovian nature (making pricing, and hence calibration, easier). Recently, Gatheral, Jaisson and Rosenbaum [14] broke this routine and introduced a fractional Brownian motion as driving factor of the volatility process. This approach (Rough Fractional Stochastic Volatility, RFSV for short) opens the door to revisiting classical pricing and calibration conundrums. They, together with the subsequent paper by Bayer, Friz and Gatheral, (see also [1, 12]) in particular showed that these models were able to capture the extra steepness of the implied volatility smile in Equity markets for short maturities, which continuous Markovian stochastic volatility models fail to describe. The icing on the cake is the (at last!!) reconciliation between the two measures $\mathbb{P}$ and $\mathbb{Q}$ within a given model, showing remarkable results both for estimation and for prediction.

One of the key issues in Equity markets is, not only to fit the (SPX) implied volatility smile, but to do so jointly with a calibration of the VIX (Futures and ideally options). Gatheral's [15] double mean reverting process is the leading (Markovian) continuous model in this direction, while models with jumps have been proposed abundantly by Carr and Madan [9] and Kokholm and Stisen [20]. This issue was briefly tackled by Bayer, Friz and Gatheral [3] for a particular rough model (rough Bergomi), and we aim here at providing a deeper analysis of VIX dynamics under this rough model and at implementing pricing schemes for VIX Futures and options. Our main contribution is a precise link between the forward variance curve $(\xi_T(\cdot))_{T \geq 0}$ and the initial forward variance curve $\xi_0(\cdot)$ in the rough Bergomi model. This in turn, allows us not only to provide simulation methods

for the VIX, but also to refine the log-normal approximation of [3] for VIX Futures, matching exactly the first two moments. Finally, we develop an efficient algorithm for VIX Futures calibration, upon which we build a joint calibration method with the SPX. As opposed to the Cholesky approach in [3], we adapt the hybrid-scheme by Bennedsen, Lunde and Pakkanen [4] with better complexity $\mathcal{O}(n \log n)$. Assuming the universality of the Hurst parameter H across VIX and SPX allows us to compute efficiently prices recursively with complexity $\mathcal{O}(n)$. In passing, we also investigate the joint consistency of VIX and SPX in the market. The organisation of the paper follows accordingly: we first introduce the rough Bergomi model and its main properties (Section 2), before presenting its pricing power for VIX Futures (Section 3), and finally develop the joint calibration algorithm in Section 4.

2. Rough volatility and the rough Bergomi model

Comte and Renault[7] were the first to propose a stochastic volatility model in which the instantaneous volatility is driven by a fractional Brownian motion, with a Hurst index restricted to be greater than $1/2$. Recently Gatheral, Jaisson and Rosenbaum [14] presented a new approach with a Hurst index smaller than $1/2$, producing extremely good fits to observed volatility data under the physical measure $\mathbb{P}$. These models form the so-called Rough Fractional Stochastic Volatility (RFSV) family that is understood as a natural extension of the classical volatility models driven by standard Brownian motion. Our work focuses on the pricing measure $\mathbb{Q}$ and we assume through this paper that the model presented by Gatheral, Jaisson and Rosenbaum [14] under $\mathbb{P}$ is a reasonable model. Finally, and most importantly, we follow the recent paper by Bayer, Friz and Gatheral [3], in order to extend the RFSV model to pricing schemes under the measure $\mathbb{Q}$. More precisely, Bayer, Friz and Gatheral [3] proposed the following model for the log stock price process $X := \log(S)$:

$$
\begin{aligned}
\mathrm{d}X_t &= -\frac{1}{2}V_t\mathrm{d}t + \sqrt{V_t}\mathrm{d}W_t, \quad X_0 = 0 \\
V_t &= \xi_0(t)\mathcal{E}(2\nu C_H \mathcal{V}_t), \qquad V_0 > 0,
\end{aligned}
\tag{2.1}
$$

with $\nu, \xi_0(\cdot) > 0$, $\mathcal{E}(\cdot)$ is the Doléans-Dade [10] stochastic exponential and $C_H := \sqrt{\frac{2H\Gamma(2-H_+)}{\Gamma(H_+)\Gamma(2-2H)}}$, where, for notational convenience (throughout the paper), we use the symbols $H_\pm := H \pm \frac{1}{2}$. All the processes are defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ supporting the two standard Brownian motions W and Z (see below). The initial forward variance curve is observed at inception, and we therefore assume without loss of generality that it is $\mathcal{F}_0$-measurable. The process $\mathcal{V}$, defined as

$$
\mathcal{V}_t := \int_0^t (t-u)^{H_-}\mathrm{d}Z_u,
\tag{2.2}
$$

is a centred Gaussian process with covariance structure

$$
\mathbb{E}(\mathcal{V}_t \mathcal{V}_s) = s^{2H}\int_0^1 \left(\frac{t}{s} - u\right)^{H_-}(1-u)^{H_-}\mathrm{d}u, \qquad \text{for any } s,t \in [0,1].
$$

We shall also introduce, for any $0 \leq T \leq t$ the notations

$$
\mathcal{V}_{t,T} := \int_T^t (t-u)^{H_-}\mathrm{d}Z_u \qquad \text{and} \qquad \mathcal{V}_t^T := \int_0^T (t-u)^{H_-}\mathrm{d}Z_u.
\tag{2.3}
$$

Note in particular that $\mathcal{V}_T^T = \mathcal{V}_T$. The two standard Brownian motions W and Z are correlated with correlation parameter $\rho \in (-1, 1)$. Here, $\xi_T(t)$ denotes the forward variance observed at time T for a maturity equal to t.

More precisely, if $\sigma_T^2(t)$ denotes the fair strike of a variance swap observed at time T and maturing at t, then

$$\sigma_T^2(t) = \frac{1}{t-T} \int_T^t \xi_T(u)\mathrm{d}u,$$

or equivalently $\xi_T(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left((t-T)\sigma_T^2(t)\right)$. For any fixed $t > 0$, the process $(\xi_s(t))_{s \leq t}$, is a martingale, i.e. $\mathbb{E}[\xi_s(t)|\mathcal{F}_u] = \xi_u(t)$, for all $u \leq s \leq t$. Furthermore, $\mathcal{V}_t^T$ is a centred Gaussian process with variance

(2.4)
$$\mathbb{V}(\mathcal{V}_t^T) = \frac{t^{2H} - (t-T)^{2H}}{2H}, \qquad \text{for } t \geq T,$$

and covariance structure

(2.5) $$\mathbb{E}\left(\mathcal{V}_t^T \mathcal{V}_s^T\right) = \int_0^T \left[(t-u)(s-u)\right]^{H_-} \mathrm{d}u = \frac{(s-t)^{H_-}}{H_+}\left\{t^{H_+}\mathrm{F}\left(\frac{-t}{s-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{s-t}\right)\right\},$$

for any $t < s$, where we introduce the function $\mathrm{F}: \mathbb{R}_- \to \mathbb{R}$ as

(2.6)
$$\mathrm{F}(u) := {}_2F_1\left(-H_-, H_+, 1 + H_+, u\right),$$

and ${}_2F_1$ is the hypergeometric function [2, Chapter 15]. Finally, the quadratic variation of $\mathcal{V}$ is given by

(2.7)
$$[\mathcal{V}]_t = \frac{t^{2H}}{2H}, \qquad \text{for } t \geq 0.$$

2.1. **Hybrid simulation scheme.** Bayer, Friz and Gatheral [3] present a Cholesky method to simulate the rough Bergomi model. Although exact, this method is very slow and other approaches need to be considered for calibration purposes. Recently, Bennedsen, Lunde and Pakkanen [4] presented a new simulation scheme for Brownian semistationary ($\mathcal{BSS}$) processes. This method, as opposed to Cholesky, is an approximate method. However, in [4] the authors show that the method yields remarkable results in the case of the rough Bergomi model. In addition, their approach leads to a natural simulation of both the Volterra process $\mathcal{V}$ and the stock price S and yields a computational complexity of order $\mathcal{O}(n \log n)$.

Definition 2.1. Let W be a standard Brownian motion on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. A truncated Brownian semistationary ($\mathcal{BSS}$) process is defined as $\mathcal{B}(t) = \int_0^t g(t-s)\sigma(s)\mathrm{d}W_s$, for $t \geq 0$, where σ is $(\mathcal{F}_t)_{t \geq 0}$-predictable with locally bounded trajectories and finite second moments, and $g: (0, \infty) \to [0, \infty)$ is Borel measurable and square integrable. We shall call it a $\mathcal{BSS}(\alpha, W)$ process if furthermore

(i) there exists $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}$ such that $g(x) = x^\alpha L_g(x)$ for all $x \in (0, 1]$, where $L_g \in \mathcal{C}^1((0, 1] \to [0, \infty))$, is slowly varying[1] at the origin and bounded away from zero. Moreover, there exists a constant $C > 0$ such that $|L_g'(x)| \leq C(1 + x^{-1})$ for all $x \in (0, 1]$;

(ii) the function g is differentiable on $(0, \infty)$.

Under this assumption, the hybrid scheme, proposed in [4] and recalled in Appendix A, provides an efficient way to simulate $\mathcal{BSS}$ processes. It applies in particular to the rough Bergomi model:

Proposition 2.2. *The Volterra process $\mathcal{V}$ in (2.2) is a truncated $\mathcal{BSS}(H_-, Z)$ process.*

Proof. From (2.2), $g(x) \equiv x^{H_-}$ and $\sigma(\cdot) \equiv 1$ as in Definition 2.1, so that V is a $\mathcal{BSS}$ process. Since $H_- \in \left(-\frac{1}{2}, 0\right)$, then $L_g \equiv 1$, and $\mathcal{V}$ satisfies Definition 2.1(i); Definition 2.1(i) trivially holds, and so does the corollary.   □

---

[1]A measurable function $L: (0, 1] \to [0, \infty)$ is slowly varying [6] at 0 if for any $t > 0$, $\lim_{x \downarrow 0} L(tx)/L(x) = 1$.

The corollary implies that we can apply the hybrid scheme to $\mathcal{V}$. In particular, for $\kappa = 1$ the matrix form representation of the scheme reads (recall that $n_T := \lfloor nT \rfloor$)

$$
\begin{pmatrix} \mathcal{V}\left(\frac{1}{n}\right) \\ \mathcal{V}\left(\frac{2}{n}\right) \\ \vdots \\ \mathcal{V}\left(\frac{n_T}{n}\right) \end{pmatrix} = \begin{pmatrix} \overline{Z}_{0,1} & 0 & \cdots & \cdots & 0 \\ \overline{Z}_{1,1} & \overline{Z}_0 & \ddots & \ddots & 0 \\ \overline{Z}_{2,1} & \overline{Z}_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \overline{Z}_{n_T-1,1} & \overline{Z}_{n_T-2} & \cdots & \overline{Z}_1 & \overline{Z}_0 \end{pmatrix} \begin{pmatrix} 1 \\ \left(\frac{1}{n}b_1^*\right)^{H_-} \\ \left(\frac{1}{n}b_2^*\right)^{H_-} \\ \vdots \\ \left(\frac{1}{n}b_{n_T-1}^*\right)^{H_-} \end{pmatrix},
$$

where the coefficients $\{b_i^*\}$ are defined in (A.1). This matrix multiplication is, by brute force, of order $\mathcal{O}(n^2)$, however using discrete convolution we may use FFT to reduce it to $\mathcal{O}(n \log n)$ as suggested in [4] .

3. Rough Bergomi and VIX

Bayer, Friz and Gatheral [3] briefly discuss the lack of consistency of the rough Bergomi model with observed VIX options data, leading to an incorrect term structure of the VIX. In this section, we investigate in detail the dynamics of the VIX, and propose a log-normal approximation. Additionally, we investigate the viability of the model in terms of VIX Futures and options, and compare it to the approximation in [3].

3.1. **VIX Futures in the rough Bergomi model.** From now on, we fix a given maturity $T \geq 0$, and define the VIX at time T via the continuous-time monitoring formula

$$
\text{VIX}_T^2 := \mathbb{E}\left(\frac{1}{\Delta} \int_T^{T+\Delta} \mathrm{d}\langle X_s, X_s \rangle \mathrm{d}s \,\middle|\, \mathcal{F}_T \right),
$$

where Δ is equal to 30 days. The risk-neutral formula for the VIX future $\mathfrak{V}_T$ with maturity T is then given by

$$
(3.1) \quad \mathfrak{V}_T := \mathbb{E}\left(\text{VIX}_T \middle| \mathcal{F}_0 \right) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left(\mathrm{d}\langle X_s, X_s \rangle | \mathcal{F}_T \right) \mathrm{d}s} \,\middle|\, \mathcal{F}_0 \right) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \xi_T(s) \mathrm{d}s} \,\middle|\, \mathcal{F}_0 \right).
$$

Note that, when $T > 0$, $\xi_T(s)$ is a market input which is not $\mathcal{F}_0$-measurable, and is hence difficult to interpret only knowing $\mathcal{F}_0$. We shall make repeated use of the following random variable defined for any $t \geq T$, by

$$
(3.2) \quad \eta_T(t) := \exp\left(2\nu C_H \mathcal{V}_t^T \right).
$$

Proposition 3.1. *The VIX dynamics are given by*

$$
\text{VIX}_T = \left\{ \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0(t) \eta_T(t) \exp\left(\frac{\nu^2 C_H^2}{H} \left[(t-T)^{2H} - t^{2H} \right] \right) \mathrm{d}t \right\}^{1/2}.
$$

Proof. Using Fubini's theorem and the instantaneous variance representation in (2.1), we can write

$$
\text{VIX}_T^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left(V_s | \mathcal{F}_T \right) \mathrm{d}s = \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left[\xi_0(t) \mathcal{E}\left(2\nu C_H \mathcal{V}_t \right) | \mathcal{F}_T \right] \mathrm{d}t
$$

$$
= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left[\xi_0(t) \eta_T(t) \exp\left(2\nu C_H \mathcal{V}_{t,T} - \frac{\nu^2 C_H^2 t^{2H}}{H} \right) \middle| \mathcal{F}_T \right] \mathrm{d}t,
$$

with $\mathcal{V}_{t,T}$ defined in (2.3). Since $\eta_T(t) \in \mathcal{F}_T$ and $\xi_0(t) \in \mathcal{F}_0$, this expression simplifies to

$$
\text{VIX}_T^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0(t) \eta_T(t) \mathbb{E}\left[\exp\left(2\nu C_H \mathcal{V}_{t,T} - \frac{\nu^2 C_H^2 t^{2H}}{H} \right) \middle| \mathcal{F}_T \right] \mathrm{d}t.
$$

The proposition follows since $\mathcal{V}_{t,T}$ is centred Gaussian, independent of $\mathcal{F}_T$, with variance given in (2.4), and $\mathbb{E}\left(e^{2\nu C_H \mathcal{V}_{t,T}}\big|\mathcal{F}_T\right) = \mathbb{E}\left(e^{2\nu C_H \mathcal{V}_{t,T}}\right) = \exp\left(\frac{\nu^2 C_H^2}{H}(t-T)^{2H}\right).$ □

The main challenge for simulation is $\eta_T(t)$. However, since the latter is independent of $\xi_0(\cdot)$, robustness of simulation schemes for the VIX will not be affected by the qualitative properties of the initial variance curve ξ_0.

Proposition 3.2. *The forward variance curve ξ_T in the rough Bergomi model admits the representation*

$$\xi_T(t) = \xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right), \qquad \text{for any } t \geq T.$$

Proof. Since $\mathbb{E}(V_t|\mathcal{F}_T) = \xi_T(t)$ by (3.1), the proposition follows from Proposition 3.1 and the equality

$$\mathbb{E}(V_t|\mathcal{F}_T) = \xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right).$$

□

Bayer, Friz and Gatheral [3] did not derive such a representation for ξ_T, and their approach for pricing VIX derivatives relies on an approximation which avoids the computations developed in this section. Proposition 3.2 allows for a better understanding of the process ξ_T, and for an innovative approach to price VIX derivatives.

3.2. **Upper and lower bounds for VIX Futures.**

Theorem 3.3. *The following bounds hold for VIX Futures:*

$$(3.3) \qquad \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t \leq \mathfrak{V}_T \leq \left\{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s\right\}^{1/2}.$$

Proof. The conditional Jensen's inequality gives

$$\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(s)\mathrm{d}s}\,\Bigg|\mathcal{F}_0\right) \leq \sqrt{\mathbb{E}\left(\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(s)\mathrm{d}s\,\Big|\mathcal{F}_0\right)}.$$

Furthermore, since ξ_0 is $\mathcal{F}_0$-adapted, Fubini's theorem along with the martingale property of ξ_T yield the upper bound $\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) \leq \sqrt{\Delta^{-1}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s}$. To obtain a lower bound we use the representation in Proposition 3.2, and Cauchy-Schwarz's inequality, and Fubini's theorem, so that

$$\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) = \mathbb{E}\left[\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t}\,\Bigg|\mathcal{F}_0\right]$$

$$\geq \mathbb{E}\left[\frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)\eta_T(t)}\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t\,\Bigg|\mathcal{F}_0\right]$$

$$= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\mathbb{E}\left(\sqrt{\eta_T(t)}\right)\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t$$

$$= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[t^{2H} - (t-T)^{2H}\right]\right)\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t$$

$$= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t,$$

since $\eta_T(t)^{1/2}$ is log-normal (Proposition 3.1), so that $\mathbb{E}(\sqrt{\eta_T(t)}) = \exp\left(\frac{\nu^2 C_H^2}{4H}\left[t^{2H} - (t-T)^{2H}\right]\right).$ □

We perform a numerical experiment to check the tightness of the bounds obtained in Proposition 3.3). For this analysis we consider three qualitative scenarios for the initial forward variance curve:

$$(3.4) \quad \text{Scenario 1}: \ \xi_0(t) = 0.234^2; \quad \text{Scenario 2}: \ \xi_0(t) = 0.234^2(1+t)^2; \quad \text{Scenario 3}: \ \xi_0(t) = 0.234^2\sqrt{1+t}.$$

Figures 1 suggest that the lower bound given in Proposition 3.3) is surprisingly tight for very different shapes
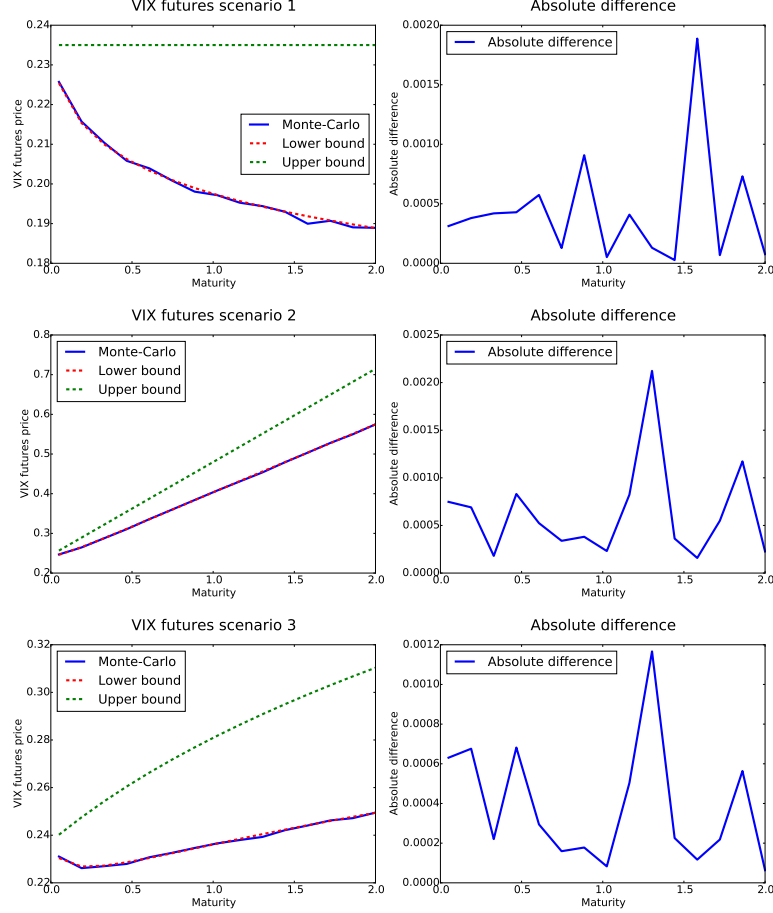

FIGURE 1. Bounds vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

of ξ_0. This can be explained with the following argument: consider a simplified and deterministic version of the VIX futures price in Proposition 3.1), denoted by

$$\phi(T) := \sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} f(t)\mathrm{d}t} = \sqrt{f(T) + \frac{\Delta}{2}f'(T) + \frac{\Delta^2}{6}f''(T) + \mathcal{O}(\Delta^3)},$$

for some strictly positive (deterministic) function $f \in \mathcal{C}^2(\mathbb{R})$. We further introduce

$$\psi(T) := \frac{1}{\Delta} \int_T^{T+\Delta} \sqrt{f(t)}\mathrm{d}t = \sqrt{f(T)} + \frac{\Delta}{4}\frac{f'(T)}{\sqrt{f(T)}} + \mathcal{O}(\Delta^2),$$

which is the corresponding lower bound by Cauchy-Schwarz's (or Jensen's) inequality, so that

$$\phi(T)^2 - \psi(T)^2 = \Delta^2\left(\frac{f''(T)}{6} - \frac{f'(T)^2}{16f(T)}\right) + \mathcal{O}(\Delta^3).$$

Hence, we observe that for small Δ, as it is the case in VIX futures, the lower bound is very close from the original value which explains (at least for the deterministic case) the behaviour observed in Figure 1.

3.3. Numerical implementation of VIX process.
In this section, we investigate different simulation schemes for the VIX in the rough Bergomi model.

3.3.1. Hybrid scheme and forward Euler approach. In order to simulate the process $(\mathcal{V}_t^T)_{t \in [T, T+\Delta]}$ it is important to notice from (2.3) that the kernel has a singularity only for $\mathcal{V}_T^T$, hence we may overcome this by simulating the process using the hybrid scheme from Section 2.1. Then, we may easily extract Z and simulate $(\mathcal{V}_t^T)_{t \in (T, T+\Delta]}$ using the forward Euler scheme with complexity $\mathcal{O}(n)$. A forward Euler scheme is chosen in this case due to the fact that the kernel in (2.3) no longer has a singularity for $t \in (T, T+\Delta]$ and this method is faster than the hybrid scheme. Once $(\mathcal{V}_t^T)_{t \in [T, T+\Delta]}$ is simulated, numerical integration routines may be used to simulate the VIX process using the expression in Proposition 3.1. It must be pointed out that this approach is computationally expensive and memory consuming since it involves to simulate the Volterra process using the hybrid scheme with complexity of $\mathcal{O}(n \log n)$ and additionally, each $\mathcal{V}_t^T$ by forward Euler. The simulation algorithm can be summarised as follows:

Algorithm 3.4 (VIX simulation in the rough Bergomi model)**.** Fix a grid $\mathcal{T} = \{t_i\}_{i=0,\ldots,n_T}$ and $\kappa \geq 1$.

(1) Simulate the Volterra process $(\mathcal{V}_t)_{t \in [0,T]}$ using the hybrid scheme in Appendix A, yielding a sample of the random variable $\mathcal{V}_T^T = \mathcal{V}_T$;

(2) extract the path of the Brownian motion Z driving the Volterra process.
$$\begin{aligned} Z_{t_i} &= Z_{t_{i-1}} + n^{H_-} \left(\mathcal{V}(t_i) - \mathcal{V}(t_{i-1}) \right), \quad \text{for } i = 1, \ldots, \kappa, \\ Z_{t_i} &= Z_{t_{i-1}} + \overline{Z}_{i-1}, \qquad\qquad\qquad\quad \text{for } i > \kappa; \end{aligned}$$

(3) fix a grid $\mathfrak{T} = \{\tau_j\}_{j=0,\ldots,N}$ on $[T, T+\Delta]$ and approximate the continuous-time process $\mathcal{V}^T$ by the discrete-time version $\widetilde{\mathcal{V}}^T$ defined via the following forward Euler scheme:
$$\widetilde{\mathcal{V}}_{\tau_0}^T := \mathcal{V}_T^T \qquad \text{and} \qquad \widetilde{\mathcal{V}}_{\tau_j}^T := \sum_{i=1}^{n_T} \frac{Z_{t_i} - Z_{t_{i-1}}}{(\tau_j - t_{i-1})^{-H_-}}, \qquad \text{for } j = 1, \ldots, N;$$

(4) compute the VIX process via numerical integration, for example using a composite trapezoidal rule:
$$\mathrm{VIX}_T \approx \left\{ \frac{1}{\Delta} \sum_{j=0}^{N-1} \frac{Q_{T,\tau_j}^2 + Q_{T,\tau_{j+1}}^2}{2} (\tau_j - \tau_{j-1}) \right\}^{1/2},$$
where $Q_{T,\tau_j}^2 := \xi_0(\tau_j) \exp\left(2 v C_H \widetilde{\mathcal{V}}_{\tau_j}^T \right) \exp\left(\frac{\nu^2 C_H^2}{H} \left((\tau_j - T)^{2H} - \tau_j^{2H} \right) \right).$

Remark 3.5. Step 4 may obviously be replaced by any available numerical integration routine, but one must then carefully choose the partition in Step 3.

3.3.2. Truncated Cholesky approach. Alternatively, one could use the more expensive, yet exact, Cholesky decomposition to simulate $\mathcal{V}^T$ on $[T, T+\Delta]$ since its covariance structure is known from (2.5). However, computational complexity aside, with the same grid $\mathcal{T}$ as in Algorithm 3.4, numerical experiments suggest that the determinant of the covariance matrix is equal to zero when using more than $n_T = 8$ discretisation points. Hence, although valid in theory, the Cholesky approach is not feasible numerically. This fact implies that there exists strong linear dependence. In fact, for any $\varepsilon > 0$, the strict inequality $\mathrm{corr}(\mathcal{V}_{t_1}^T, \mathcal{V}_{t_1+\varepsilon}^T) < \mathrm{corr}(\mathcal{V}_{t_2}^T, \mathcal{V}_{t_2+\varepsilon}^T)$ holds for all $T < t_1 < t_2$, as well as the following:

Proposition 3.6. *The limit* $\lim_{\varepsilon \downarrow 0} corr(\mathcal{V}_t^T, \mathcal{V}_{t+\varepsilon}^T) = 1$ *holds for any* $t \in [T, T + \Delta]$.

Proof. This follows readily from the continuity in L^2 of the map $t \mapsto \mathcal{V}_t^T$:

$$
E[(\mathcal{V}_{t+\varepsilon}^T - \mathcal{V}_t^T)^2] = \int_0^T [(t + \varepsilon - u)^{H_-} - (t - u)^{H_-}]^2 \mathrm{d}u
$$
$$
= \left[\frac{(T+\varepsilon)^{1+2H_+} - \varepsilon^{1+2H_+} + T^{1+2H_+}}{1 + 2H_+} - \frac{2T^{1+H_+}\varepsilon^{H_+}}{1 + H_+} {}_2F_1\left(-H_+, 1 + H_+, 2 + H_+, -\frac{T}{\varepsilon}\right) \right].
$$

Applying the identity [2, page 564]

$$
{}_2F_1(a, b, c; z) = \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-z)^{-a} {}_2F_1\left(a, a - c + 1, a - b + 1; \frac{1}{z}\right) + \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}\frac{1}{(-z)^b} {}_2F_1\left(b, b - c + 1, b - a + 1; \frac{1}{z}\right)
$$

to the case $(a, b, c) = (-H_+, 1 + H_+, 2 + H_+)$ and using properties of the Gamma function along with the fact that ${}_2F_1(a, 0, c, z) \equiv 1$, we obtain

$$
{}_2F_1\left(-H_+, 1 + H_+, 2 + H_+, -\frac{T}{\varepsilon}\right) = \frac{1 + H_+}{1 + 2H_+}\left(\frac{T}{\varepsilon}\right)^{H_+} {}_2F_1\left(-H_+, 2 + 2H_+, -2H_+, \frac{-\varepsilon}{T}\right).
$$

Finally, the series representation of ${}_2F_1$ [2, Chapter 15.1.1] implies that ${}_2F_1\left(-H_+, 2 + 2H_+, -2H_+, \frac{-\varepsilon}{T}\right)$ converges to 1 as ε tends to zero, and hence that $E[(\mathcal{V}_{t+\varepsilon}^T - \mathcal{V}_t^T)^2]$ tends to zero. $\square$

In light of Proposition 3.6, we model exactly the dependence structure on the first 8 grid points $t_1, \ldots, t_8$, then truncate the Cholesky decomposition and compute the correlations $\rho_i := \mathrm{corr}(\mathcal{V}_{t_{8+i}}^T, \mathcal{V}_{t_{8+i+1}}^T)$, for $i = 0, \ldots, n-9$ to approximate the process by adequately rescaling and correlating each pair of subsequent grid points. In contrast to the hybrid + forward Euler scheme, the computational complexity is much lower, since the Cholesky method is truncated with only 8 components. The VIX simulation algorithm therefore reads as follows:

Algorithm 3.7 (VIX simulation (truncated Cholesky)). Fix a grid $\mathfrak{T} = \{\tau_j\}_{j=0,\ldots,N}$ on $[T, T + \Delta]$,

(i) compute the covariance matrix of $(\mathcal{V}_{\tau_j}^T)_{i=j,\ldots,8}$ using (2.5);

(ii) generate $\{\mathcal{V}_{\tau_j}^T\}_{j=1,\ldots,N}$ by correlating and rescaling using (2.5):

$$
\mathcal{V}_{\tau_j}^T = \sqrt{\mathbb{V}(\mathcal{V}_{\tau_j}^T)}\left(\frac{\rho(\mathcal{V}_{\tau_{j-1}}^T, \mathcal{V}_{\tau_j}^T)\mathcal{V}_{\tau_{j-1}}^T}{\sqrt{\mathbb{V}(\mathcal{V}_{\tau_{j-1}}^T)}} + \sqrt{1 - \rho(\mathcal{V}_{\tau_{j-1}}^T, \mathcal{V}_{\tau_j}^T)^2}\mathcal{N}(0, 1)\right), \quad \text{for } j = 9, \ldots, N;
$$

(iii) compute the VIX via numerical integration as in Algorithm 3.4(4).

3.3.3. *Numerical experiment.* We compute the price of VIX Futures using the simulation algorithm introduced in the previous section. We set the same parameters as in [3] and [4]:

$$
(3.5) \qquad \xi_0 = 0.235^2, \qquad H = 0.07, \qquad \nu = 1.9\frac{C_H\sqrt{2H}}{2} \approx 1.2287, \qquad \kappa = 2.
$$

We perform 10^5 simulations for the hybrid scheme + forward Euler (HS+FE) method, while 10^6 simulations are used for the truncated Cholesky. Figure 2 suggests that both methods agree qualitatively and converge to a similar output. In particular, the truncated Cholesky approach seems to suffer from larger oscillations as maturity increases, even with 10^6 simulations. Nevertheless, Figure 2 indicates that the Monte-Carlo variance increases in T for both schemes, which is confirmed in Figure 3, where the error also increases with maturity. On the other hand, Figure 4 shows that the HS+FE methods is slower than the truncated Cholesky method, which is consistent with the computational complexities discussed in the previous section. In particular, the computational time of the Cholesky method is almost constant when using parallel computing. Figure 4 also

suggests that large simulations are needed to obtain precise prices. In light of this analysis, both methods seem to approximate the required output in a decent manner. Even if the truncated Cholesky approach gives a considerably fast output for each maturity, it is not considered for calibration, since its computational time grows linearly in the number of maturities, making the algorithm too slow for reasonable calibration. Instead, we will use the truncated Cholesky approach as a benchmark for the upcoming approximations.
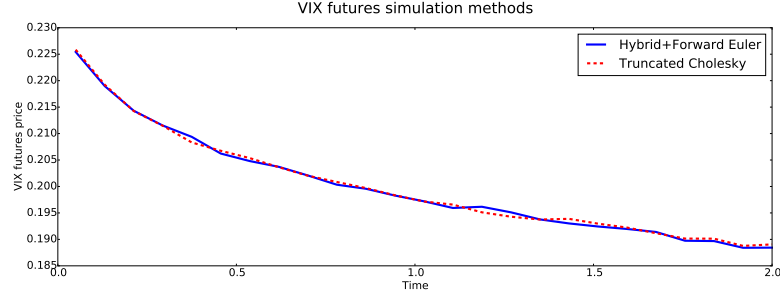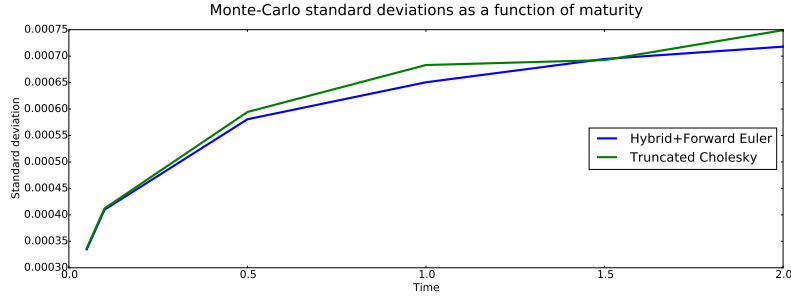

FIGURE 2. VIX Futures using HS+FE and truncated Cholesky

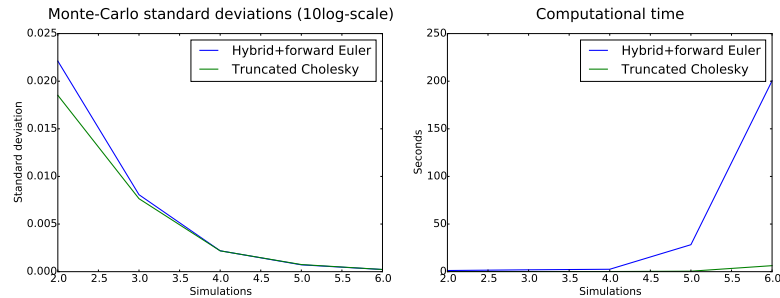FIGURE 3. Monte-Carlo standard deviations, with 10^5 simulations.

FIGURE 4. Monte-Carlo standard deviations and computational times for a fixed maturity ($T = 2$ years) for both methods using efficient parallel computing.

3.4. **The VIX process and log-normal approximations.** We now investigate approximate methods to price VIX futures and options. We define the $\mathcal{F}_T$-measurable random variable $\Delta\mathrm{VIX}_T^2 = \int_T^{T+\Delta} \xi_T(t)\mathrm{d}t$. In [3], Bayer, Friz and Gatheral assumed that $\log(\Delta\mathrm{VIX}_T^2)$ follows a Gaussian distribution, and computed directly its first and second moments, which hence fully characterise the distribution of $\Delta\mathrm{VIX}_T^2$. However, since the sum of log-normal random variables is not (in general) log-normal, the fact that ξ_T is log-normal does not imply that $\Delta\mathrm{VIX}_T^2$ is. In a different context (geometric Brownian motion and Asian options), Dufresne [11] proved that, under certain conditions, an integral of log-normal variables asymptotically converges to a log-normal. This approximation has been widely used for many applications [11], and Dufresne's result motivates Bayer-Friz-Gatheral's assumption. We provide here exact formulae for the mean and variance of this distribution, and compare them numerically to those by Bayer-Friz-Gatheral.

Proposition 3.8. *The following holds:*

$$\sigma^2 := \mathbb{V}(\log(\Delta\mathrm{VIX}_T^2)) = -2\log\mathbb{E}(\Delta\mathrm{VIX}_T^2) + \log\mathbb{E}\left((\Delta\mathrm{VIX}_T^2)^2\right),$$

$$\mu := \mathbb{E}(\log(\Delta\mathrm{VIX}_T^2)) = \log\mathbb{E}(\Delta\mathrm{VIX}_T^2) - \frac{\sigma^2}{2}.$$

Furthermore, $\mathbb{E}(\Delta\mathrm{VIX}_T^2) = \int_T^{T+\Delta} \xi_0(t)\mathrm{d}t$ and

$$(3.6)\quad \mathbb{E}\left[(\Delta\mathrm{VIX}_T^2)^2\right] = \int_{[T,T+\Delta]^2} \xi_0(u)\xi_0(t)\exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H}\right]\right\} \mathrm{e}^{\overline{\Theta}_{u,t}}\mathrm{d}u\mathrm{d}t.$$

where $\overline{\Theta}_{u,t}$ is equal to zero if $u = t$ and otherwise equal to $\Theta_{u\vee t, u\wedge t}$, where

$$\Theta_{u,t} := 2\nu^2 C_H^2 \left\{\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + 2\frac{(u-t)^{H_-}}{H_+}\left[t^{H_+}\mathrm{F}\left(\frac{-t}{u-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{u-t}\right)\right]\right\}.$$

Remark 3.9. Since all the integrals in the proposition are computed over compact intervals, they are finite as long as ξ_0 is well behaved on $[T, T+\Delta]$. Assuming this is indeed the case is not restrictive in practice as ξ_0 represents the initial forward variance curve; note in particular that continuity of ξ_0 is sufficient.

Remark 3.10. The reason why $\overline{\Theta}_{u,t}$ is defined that way is for numerical purposes. Indeed, when $u < t$, $\Theta_{u,t}$ is not well defined since $\mathrm{F}(x)$ only makes sense when $x \le 1$ (details on the radius of convergence of hypergeometric functions can be found in [2][Page 556]), even though the integral representation (3.6) is still well defined. However, most numerical packages implement hypergeometric functions via series expansions. The trick from $\Theta_{u,t}$ to $\overline{\Theta}_{u,t}$ allows us to bypass this issue.

Proof of Proposition 3.8. The expectation follows directly from the tower property and Fubini's theorem. For the second moment, we use the decomposition

$$\mathbb{E}\left((\Delta\mathrm{VIX}_T^2)^2\right) = \mathbb{E}\left(\int_T^{T+\Delta}\int_T^{T+\Delta} \xi_T(u)\xi_T(t)\mathrm{d}t\mathrm{d}u\right) = \int_T^{T+\Delta}\int_T^{T+\Delta} \mathbb{E}\left(\xi_T(u)\xi_T(t)\right)\mathrm{d}t\mathrm{d}u$$

where in the last step we use that $\xi_T(s)$ is $\mathcal{F}_T$-measurable in order to apply Fubini. Using the representation obtained in Proposition 3.2, we get

$$(3.7)\qquad \mathbb{E}\left(\xi_T(u)\xi_T(t)\right) = \xi_0(u)\xi_0(t)\exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} - u^{2H} + (t-T)^{2H} - t^{2H}\right]\right\}\mathbb{E}\left(\mathrm{e}^{\vartheta_{u,t}}\right),$$

where the random variable

$$\vartheta_{u,t} := 2\nu C_H \int_0^T \left[(u-s)^{H_-} + (t-s)^{H_-}\right]\mathrm{d}Z_s$$

is Gaussian with zero expectation and variance

$$\mathbb{V}(\vartheta_{u,t}) = 4\nu^2 C_H^2 \left(\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + 2 \int_0^T (u-s)^{H_-}(t-s)^{H_-} \mathrm{d}s \right)$$

$$= 4\nu^2 C_H^2 \left\{ \frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + \frac{2(u-t)^{H_-}}{H_+} \left[t^{H_+} \mathrm{F}\left(\frac{-t}{u-t} \right) - (t-T)^{H_+} \mathrm{F}\left(\frac{T-t}{u-t} \right) \right] \right\}.$$

Then,

$$\mathbb{E}\left(\xi_T(u)\xi_T(t) \right) = \xi_0(u)\xi_0(t) \exp\left(\frac{\nu^2 C_H^2}{H} \left((u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H} \right) \right) \exp\left(\frac{\mathbb{V}(\vartheta_{u,t})}{2} \right),$$

and the second moment follows from the immediate computation

$$\mathbb{E}\left((\Delta \mathrm{VIX}_T)^2 \right) = \int_T^{T+\Delta} \int_T^{T+\Delta} \mathbb{E}\left[\xi_T(u)\xi_T(t) \right] \mathrm{d}u \mathrm{d}t$$

$$= \int_T^{T+\Delta} \int_T^{T+\Delta} \xi_0(u)\xi_0(t) \exp\left\{ \frac{\nu^2 C_H^2}{H} \left[(u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H} \right] \right\} \mathrm{e}^{\frac{1}{2}\mathbb{V}(\vartheta_{u,t})} \mathrm{d}u \mathrm{d}t,$$

after defining $\Theta_{u,t} := \exp\left(\frac{1}{2}\mathbb{V}(\vartheta_{u,t}) \right)$. □

In order to provide closed-form expressions for VIX Futures and options, we enforce the following assumption:

Assumption 3.11. $\Delta \mathrm{VIX}_T^2$ is log-normal.

In fact, this is almost the same as the assumption by Bayer, Friz and Gatheral [3]; however, they did not compute the variance exactly as in Proposition 3.8, and instead considered the lognormal approximation

Assumption 3.12 (Bayer-Friz-Gatheral). $\log(\Delta \mathrm{VIX}_T^2)$ is Gaussian with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ given by

$$\tilde{\sigma}^2 = \frac{4\nu^2 C_H^2}{\Delta^2 H_+^2} \int_0^T \left[(T-s+\Delta)^{H_+} - (T-s)^{H_+} \right]^2 \mathrm{d}s \qquad \text{and} \qquad \tilde{\mu} = \log \int_T^{T+\Delta} \xi_0(t)\mathrm{d}t - \frac{\tilde{\sigma}^2}{2}.$$

Lemma 3.13. *A VIX future is worth*

$$\mathfrak{V}_T = \begin{cases} \Delta^{-1/2} \sqrt{\int_T^{T+\Delta} \xi_0(t)\mathrm{d}t} \, \exp\left(-\frac{\sigma^2}{8} \right), & \text{under Assumption 3.11,} \\ \Delta^{-1/2} \sqrt{\int_T^{T+\Delta} \xi_0(t)\mathrm{d}t} \, \exp\left(-\frac{\tilde{\sigma}^2}{8} \right), & \text{under Assumption 3.12.} \end{cases}$$

Proof. Since, by assumption $\Delta^{1/2}\mathrm{VIX}_T \overset{\Delta}{=} \exp\left(\frac{\mu + \sigma \mathcal{N}(0,1)}{2} \right)$, the price of a VIX future directly reads

$$\mathbb{E}\left(\mathrm{VIX}_T | \mathcal{F}_0 \right) = \Delta^{-1/2}\mathbb{E}\left(\Delta^{1/2}\mathrm{VIX}_T \right) = \Delta^{-1/2}\exp\left(\frac{\mu}{2} + \frac{\sigma^2}{8} \right),$$

and the second case follows analogously. □

As opposed to the simulation schemes, the log-normal approximation does depend on the qualitative proper-
ties of the initial forward variance curve $\xi_0(\cdot)$. Hence, different curves should be analysed to check the robustness of the method. We now exploit the approximation in Assumption 3.11 to obtain a closed-form Black-Scholes type formulae for European options on the VIX.

Lemma 3.14. *For $0 \leq t \leq T$, let $\mathfrak{V}_T(t) := \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_t)$ denote the price at time t of a VIX future maturing at T. Then, under Assumption 3.11, a European Call option on a VIX future maturing at T is worth*

$$\mathcal{C}_T^{\mathcal{V}} := \mathbb{E}\left[(\mathfrak{V}_T(T) - K)_+|\mathcal{F}_0\right] = \Delta^{-1/2}\sqrt{\int_T^{T+\Delta} \xi_0(t)\mathrm{d}t} \exp\left(-\frac{\sigma^2}{8}\right)\Phi(d_1) - K\Phi(d_2),$$

where $\widetilde{K} := [\log(K^2\Delta) - \log\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t + \sigma^2/2]/\sigma$, $d_1 := -\widetilde{K} + \frac{1}{2}\sigma$ and $d_2 := -\widetilde{K}$, with σ^2 as in Proposition 3.8

Proof. Under Assumption 3.11, the lemma follows directly from the following trivial computations:

$$\mathbb{E}\left(\mathfrak{V}_T(T) - K\right)_+ = \int_{\widetilde{K}}^{\infty}\left[\Delta^{-1/2}\exp\left(\frac{\mu}{2} + \frac{\sigma z}{2}\right) - K\right]_+\phi(z)\mathrm{d}z = \frac{1}{\sqrt{\Delta}}\exp\left(\frac{\mu}{2} + \frac{\sigma^2}{8}\right)\int_{\widetilde{K}}^{\infty}\phi\left(z - \frac{\sigma}{2}\right)\mathrm{d}z - K\Phi(d_2).$$

$\square$

3.4.1. *Numerical tests of the log-normal approximation.* We perform a numerical analysis of the approximation by pricing VIX Futures using the parameters in (3.5). We consider 10^6 simulations and the three qualitative scenarios introduced in (3.4) for the forward variance curve. Figures 5 suggest that the approximation is accurate
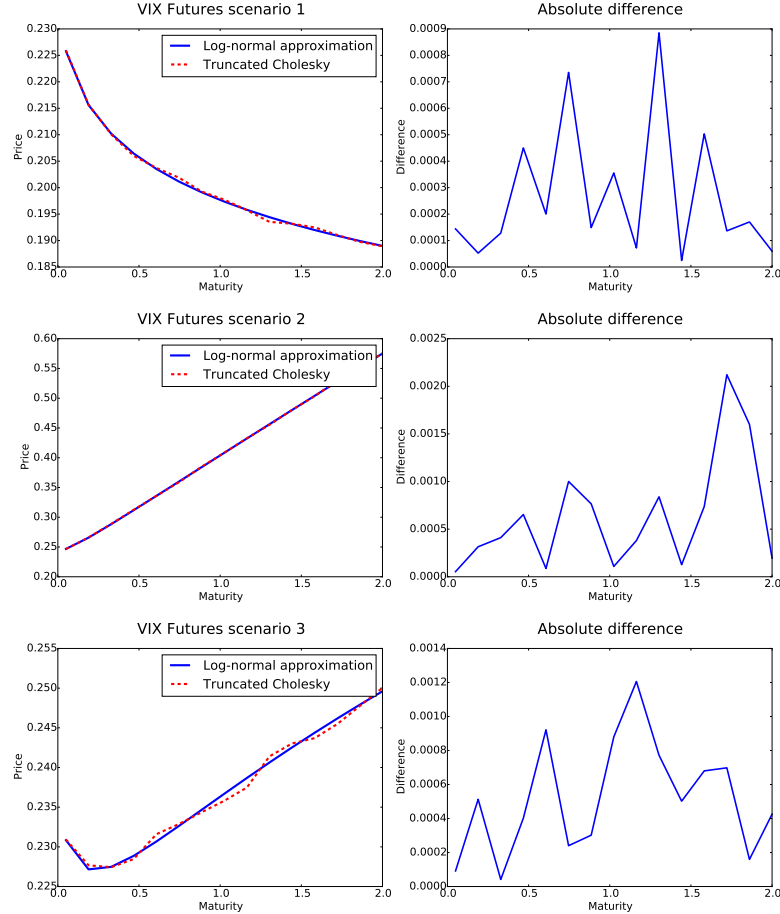

FIGURE 5. Log-normal approximations vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

being the difference of order 10^{-3} or less. The oscillating nature of the difference is also a good sign since it is probably caused by the Monte Carlo error and does not show any monotonicity. Furthermore, the method

shows to be robust for different type of curves $\xi_0(\cdot)$. Moreover, the log-normal approximation seems to converge to the true mean, avoiding the oscillations of the Truncated Cholesky method. Therefore, we may conclude that the log-normal approximation produces a good output, with desired smoothness properties. On the other hand, we also analyse European VIX Call options repeating the previous parameters and considering the flat forward variance curve from Scenario 1. Figure 6 shows that the log-normal approximation is accurate for at-the-money
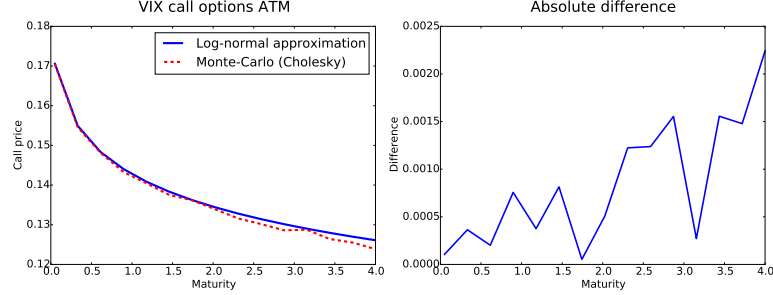

FIGURE 6. Log-normal approximation vs. Monte Carlo (Truncated Cholesky) with 10^6 simulations

options. Nevertheless, the difference increases with maturity, and hence one must be careful when using this approximation to price options with long maturities. However, in practice, on Equity markets, liquid maturities are only up to two years.

3.4.2. *Numerical tests.* We benchmark Bayer-Friz-Gatheral's approximation against ours with the parameters in (3.5). Figures 7 and 8 suggest that both are similar. In particular, we observe that the variance deviates as maturity increases. Nevertheless, for practical purposes, both approximations agree over a four-year horizon, long enough to cover available Futures data. Many functional forms of $\xi_0(t)$ were also tested giving similar results as the ones shown in Figures 7 and 8. We conclude that the approximation by Bayer, Friz and Gatheral is accurate enough for practical purposes and yields and reduces significantly the computational costs since the computation of $\tilde{\sigma}^2$ involves a single integral while our approximation requires a double integral (for σ^2). In particular, in order to generate Figure 7 our approximation was 40 times slower than Bayer, Friz and Gatheral's.
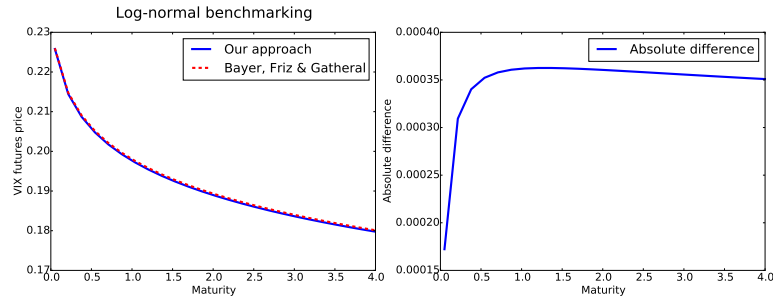

FIGURE 7. VIX Futures prices following Assumption 3.11 and Assumption 3.12.

3.5. **VIX Futures calibration in rBergormi.** In light of the promising results obtained in the previous sections, we create a calibration algorithm based on the log-normal approximation by Bayer, Friz and Gatheral [3]. Even if our approximation seems to be more accurate the computational cost is much larger and the difference
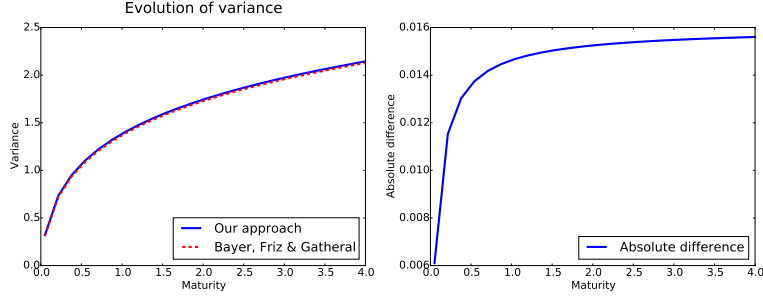
FIGURE 8. Comparison of the variance following Assumption 3.11 (σ^2) and Assumption 3.12 ($\tilde{\sigma}^2$).

between both approaches has been shown to be very small. Moreover, the approach by Bayer, Friz and Gatheral allows to compute the gradient of the objective function in a semi-explicit form, which in terms of optimisation and calibration is extremely useful.

3.5.1. *Objective function.* To calibrate the model to VIX Futures, we first define the objective function

$$(3.8) \qquad \mathcal{L}^{\mathfrak{F}}(\nu, H) := \sum_{i=1}^{N} (\mathfrak{V}_{T_i} - \mathfrak{F}_i)^2,$$

which we minimise over (ν, H). Here $(\mathfrak{F}_i)_{i=1,\ldots,N}$ are the observed Futures prices on the time grid $T_1 < \ldots < T_N$, $\mathfrak{V}_{T_i} = \Delta^{-1/2}\sqrt{\int_{T_i}^{T_i+\Delta} \xi_0(t)\mathrm{d}t} \exp\left(-\frac{\tilde{\sigma}_i^2}{8}\right)$. and

$$\tilde{\sigma}_i^2 = \frac{4\nu^2 C_H^2}{H_+^2 \Delta^2}\left[\frac{(T_i+\Delta)^{1+2H_+} - \Delta^{1+2H_+} + T_i^{1+2H_+}}{1+2H_+} - 2\frac{T_i^{1+H_+}\Delta^{H_+}}{1+H_+}{}_2F_1\left(-H_+, 1+H_+, 2+H_+, -\frac{T_i}{\Delta}\right)\right],$$

which is the closed-form expression of the variance in Assumption 3.12. The gradient of the objective function is an important source of information in many optimisation algorithms. To compute it, we differentiate the objective function in (3.8) with respect to ν and H and apply the chain rule:

$$\frac{\partial \mathcal{L}^{\mathfrak{F}}}{\partial \nu}(\nu, H) = -\frac{1}{4}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\frac{\partial\tilde{\sigma}_i^2}{\partial\nu} = -\frac{1}{2\nu}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\tilde{\sigma}_i^2,$$

where

$$\frac{\partial\tilde{\sigma}_i^2}{\partial\nu} = \frac{8\nu C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left((T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right)^2 \mathrm{d}s = \frac{2\tilde{\sigma}_i^2}{\nu}.$$

On the other hand, $\dfrac{\partial \mathcal{L}^{\mathfrak{F}}}{\partial H}(\nu, H) = -\dfrac{1}{4}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\dfrac{\partial\tilde{\sigma}_i^2}{\partial H}$, with

$$\frac{\partial\tilde{\sigma}_i^2}{\partial H} = \frac{4\nu^2\frac{\partial C_H^2}{\partial H}H_+ - 8\nu^2 C_H^2}{\Delta^2 H_+^3}\int_0^{T_i}\left((T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right)^2\mathrm{d}s$$

$$+ \frac{8\nu^2 C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left[(T_i - s + \Delta)^{H+\frac{1}{2}} - (T_i - s)^{H_+}\right]\left[(T_i - s + \Delta)^{H_+}\log(T_i - s + \Delta) - (T_i - s)^{H_+}\log(T_i - s)\right]\mathrm{d}s$$

$$= \frac{\tilde{\sigma}_i^2(K_H - 2)}{H_+}$$

$$+ \frac{8\nu^2 C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left[(T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right]\left[(T_i - s + \Delta)^{H_+}\log(T_i - s + \Delta) - (T_i - s)^{H_+}\log(T_i - s)\right]\mathrm{d}s,$$

where

$$\frac{\partial C_H^2}{\partial H} = \frac{2\Gamma(2 - H_+)}{\Gamma(H_+)\Gamma(1 - 2H_-)}\left\{1 + [2\psi(2 - H_+) - \psi(H_+) - \psi(1 - 2H_-)]\,H\right\}$$

and $K_H = \frac{H_+}{H}\left\{1 + H\psi\left(2 - H_+\right) - H\left(\psi(H_+) + \psi(1 - 2H_-)\right)\right\}$, where ψ is the digamma function.

3.5.2. *Obtaining the initial forward variance curve.* The initial forward variance curve plays a crucial role in both the Bergomi [5] and the rough Bergomi models [3], since it is a market input. In particular it depends on the current term structure of variance swaps. Even if variance swaps are not traded in standard exchange markets, they are actively traded over-the-counter (OTC). This in turn means that there is no observable data and we must establish a valuation method for variance swaps. The celebrated static replication formula by Carr and Madan [8], applicable here since the underlying process is a continuous semimartingale, allows us to price any variance swap. Nevertheless, out-of-money Call and Put option prices are needed for all possible strikes. Since this information is not available in practice, one can either adopt a model-free valuation formula or directly propose a parameterisation for the implied volatility surface. The first approach involves a discretisation of the static replication formula, which is how the VIX index is computed by the Chicago Board Options Exchange. It is not easy, however, to extend this computation for large maturities (VIX is a 30-day ahead index), where liquidity of options may play a major role. On the other hand, the second approach allows to calibrate a model using available data and additionally allows to interpolate / extrapolate available option data to all strikes and maturities. In this work we follow the latter approach, with the eSSVI parameterisation [18] for the implied volatility surface, which is a refinement of the SSVI parametrisation introduced in [16]:

$$(3.9) \qquad \sigma_{\mathrm{BS}}^2(t, k)t = w(t, k) := \frac{\theta_t}{2}\left\{1 + \rho(\theta_t)\varphi(\theta_t)k + \sqrt{(\varphi(\theta_t)k + \rho(\theta_t))^2 + 1 - \rho(\theta_t)^2}\right\},$$

where θ_t is the observed ATM variance curve, and where the shape function $\varphi(\cdot)$ takes the form $\varphi(\theta) = \eta\theta^{-\lambda}(1 + \theta)^{\lambda-1}$. For the correlation parameter $\rho(\cdot)$ we restrict it to the following functional form:

$$(3.10) \qquad \rho(\theta) = (A - C)\mathrm{e}^{-B\theta} + C, \qquad \text{for } (A, C) \in (-1, 1)^2, B \geq 0,$$

ensuring that $|\rho(\cdot)| \leq 1$. We shall indistinctly refer to the total implied variance by $\sigma_{\mathrm{BS}}^2(\cdot)t$ or $w(\cdot)$, where $\sigma_{\mathrm{BS}}(\cdot)$ represents the implied volatility. Gatheral and Jacquier [16] found (sufficient and almost necessary) conditions on the parameters ρ, $\varphi(\cdot)$, η and λ preventing arbitrage. In the eSSVI formulation, the correlation has a term structure (3.10), and care must be taken in order to preclude arbitrage. Concretely, following [18], the restriction

$$(3.11) \qquad |\theta\partial_\theta(\rho(\theta)) + \rho(\theta)\gamma| \leq \gamma,$$

where $\gamma := \partial_\theta(\theta\varphi(\theta))/\varphi(\theta)$, is a necessary condition to preclude calendar spread arbitrage. To prevent butterfly arbitrage, exactly as in the SSVI parameterisation, it is sufficient [16, Theorem 4.2] to check that the inequality $\theta_t\varphi^2(\theta_t)(1 + |\rho(\theta_t)|) \leq 4$ holds for all maturity t. In this parameterisation, variance swaps can be computed in closed form, as proved by Martini [21], based on earlier works by Gatheral [13] and Fukasawa [12]:

Proposition 3.15. *The fair strike (in total variance) of a variance swap in the eSSVI model reads*

$$\sigma_0(t)^2 t := -2\mathbb{E}\log\left(\frac{S_t}{S_0}\right) = \frac{b_t^2 + 2a_t(c_t + \theta_t)}{2a_t^2},$$

where $\chi_t := \frac{1}{4}[1 - \rho(\theta_t)^2]\theta_t\varphi(\theta_t)$, and

$$a_t = 1 + \frac{\theta_t\varphi(\theta_t)}{2}\left(\rho(\theta_t) - \frac{\chi_t}{2}\right), \qquad b_t = \theta_t\varphi(\theta_t)\left[\chi_t - \rho(\theta_t)\right], \qquad c_t = \theta_t\varphi(\theta_t)\chi_t.$$

Recalling the relation between variance swaps and the forward variance curve, we have

$$\xi_0(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(t\sigma_0^2(t)\right) = \sigma_0^2(t) + t\frac{d}{\mathrm{d}t}\sigma_0^2(t).$$

Remark 3.16. In order to interpolate/extrapolate the eSSVI, it is necessary to also interpolate/extrapolate θ_t for all t. However, θ_t is only observed on a discrete set of maturity dates, and consequently, cubic splines are used to interpolate/extrapolate all other maturities.

3.5.3. *Calibration algorithm and numerical results.* We first calibrate the eSSVI parameterisation (3.9) on the SPX implied volatility surface on the 4th of December 2015. Figure 9 shows the fit for the shortest, medium and longest maturities available in the data set. For short maturities the eSSVI is not able to fully capture the volatility smile, however as maturity increases the fit improves remarkably. For VIX Futures, the calibration
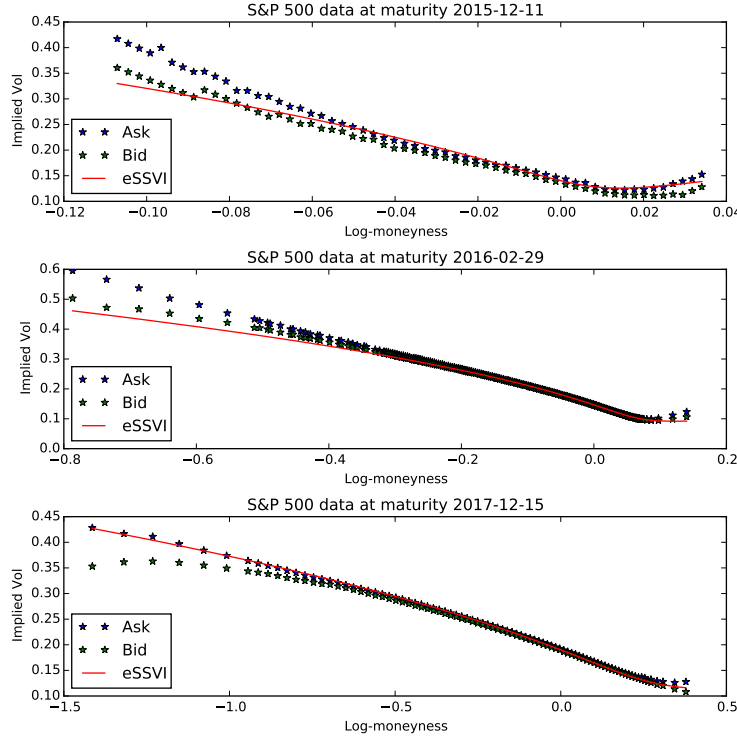

FIGURE 9. eSSVI calibration results on 4/12/2015 using traded SPX options.

algorithm reads as follows:

Algorithm 3.17 (VIX Futures calibration algorithm in the rough Bergomi model)**.**

(i) Calibrate eSSVI to available SPX option data;

(ii) compute the variance swap term structure $(\sigma_0(t)^2)_{t\geq 0}$ using Proposition 3.15;

(iii) extract the initial forward variance curve, $\xi_0(\cdot)$ via $\xi_0(t) \approx \sigma_0^2(t) + \frac{\sigma_0^2(t+\varepsilon)-\sigma_0^2(t-\varepsilon)}{2\varepsilon}t$ (with $\varepsilon = 1E-8$);

(iv) minimise (over ν, H) the objective function in (3.8).

Figures 10-12 suggest that the model fits very well the observed VIX Futures term structure for different dates. Moreover, we notice that both the model and the observed data are qualitatively equal in terms of

convexity/concavity. In the rough Bergomi model this information is obtained from option prices through ξ_0, which suggests a correspondence in the market between VIX futures and SPX options. However, we also observe that in all three cases the error is greater for short maturities, mimicking the calibration limits of eSSVI for short maturities, as detailed in Section 3.5.2.
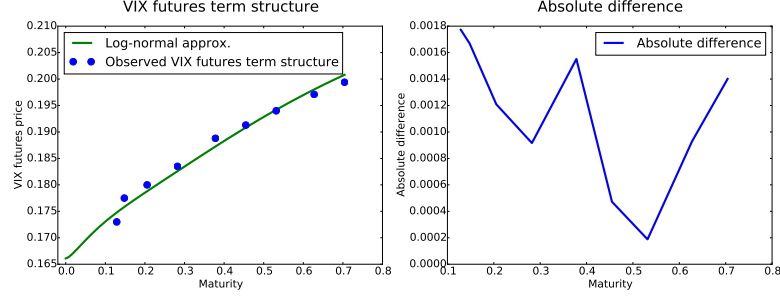

FIGURE 10. VIX Futures calibration on 4/12/2015. Optimal parameters: $(H, \nu) = (0.09237, 1.004)$.
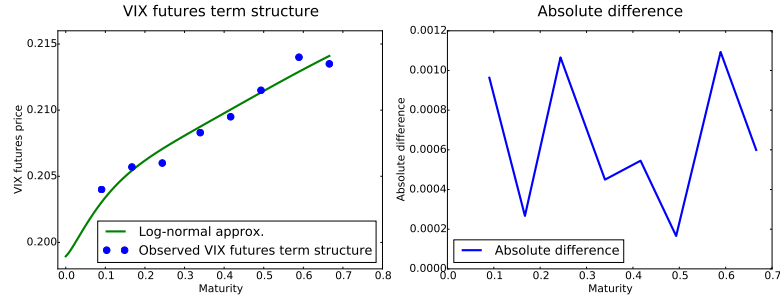

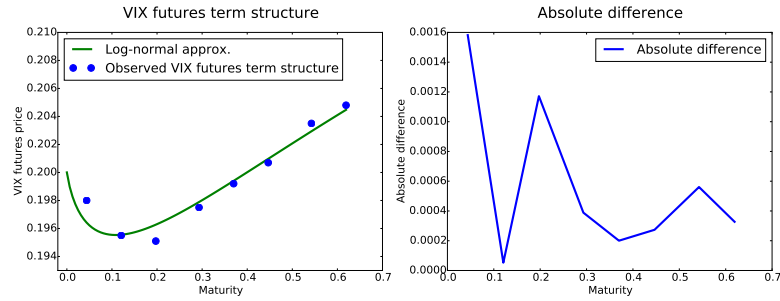FIGURE 11. VIX Futures calibration on 22/2/2016. Optimal parameters: $(H, \nu) = (0.10093, 1.00282)$.

FIGURE 12. VIX Futures calibration on 4/1/2016. Optimal parameters: $(H, \nu) = (0.0509, 1.2937)$.

Remark 3.18. The reader should recall the importance of the initial forward variance curve $(\xi_0(t))_{t \geq 0}$ in the VIX Futures process, since $(\mathfrak{V}_t)_{t \geq 0}$ depends on the whole path of ξ_0 up to time t. Therefore, even if ξ_0 is only misspecified for short maturities (as is the case of the eSSVI), this affects the whole term structure of the VIX process (for details we refer the reader to Section 3). Therefore, an improved ξ_0 estimation would not only increase the accuracy of the model for short maturities, but also for the whole term structure.

Remark 3.19. Our calibration involves two different data sets, SPX options and VIX Futures: the Vanilla quotes are extracted from the CBOE delayed option quotes page[2] and the VIX Futures from the CBOE VIX futures historical data page[3]. We perform an aggregation of the different Future quotes, since they are quoted on a Future per Future basis, and check the consistency between the two data sets by comparing the left end extrapolation of the Futures curve with the theoretical VIX computed from option prices.

4. From VIX Futures to SPX options

In this final chapter we assess whether the Hurst parameter H obtained through the VIX Futures calibration algorithm is consistent with SPX options. For this purpose, we calibrate the rough Bergomi model to SPX option data by fixing the parameter H and letting the algorithm calibrate ν and ρ. One of the main reasons to fix H is that the hybrid scheme introduced in Section 2.1 remarkably reduces its complexity to $\mathcal{O}(n)$, since the $\mathcal{O}(n \log n)$ complexity of the Volterra is computed only once and reused afterwards. Therefore, by fixing H the pricing scheme is much faster when several valuations are performed, as is the case of a calibration algorithm.

4.1. **Pricing in the rough Bergomi model.** We present a pricing scheme, where the Volterra process $\mathcal{V}$ is simulated using a hybrid scheme, while a standard Euler scheme generates the paths of the stock process:

Algorithm 4.1 (Simulation of the rough Bergomi model). Consider the grid $\mathcal{T} := \{t_i\}_{i=0,\ldots,n_T}$, and fix $\kappa \geq 1$.

 (i) Simulate the Volterra process $\mathcal{V}$ on the grid $\mathcal{T}$ using the hybrid scheme;
 (ii) simulate the variance process as $V_t = \xi_0(t)\mathcal{E}(2\nu C_H \mathcal{V}_t)$, for $t \in \mathcal{T}$ and where $([\mathcal{V}]_t)_{t \geq 0}$ is given in (2.7);
 (iii) extract the path of the Brownian motion Z driving $\mathcal{V}$:

$$\begin{aligned} Z_{t_i} &= Z_{t_{i-1}} + n^{H-}\left(\mathcal{V}(t_i) - \mathcal{V}(t_{i-1})\right), \quad \text{for } i = 1, \ldots, \kappa, \\ Z_{t_i} &= Z_{t_{i-1}} + \overline{Z}_{i-1}, \quad \text{for } i > \kappa; \end{aligned}$$

 compute $\{Z^\perp\}_{i=0}^{n_T-1}$ where $Z_i^\perp \overset{\Delta}{=} \mathcal{N}(0, 1/n_T)$ is an independent standard Gaussian sample;
 (iv) correlate the two Brownian motions via $W_{t_i} - W_{t_{i-1}} = \rho \overline{Z}_{i-1} + \sqrt{1-\rho^2}\overline{Z}_{i-1}^\perp$;
 (v) simulate $S_{t_i} = \exp(X_{t_i})$ using a forward Euler scheme:

$$X_{t_{i+1}} = X_{t_i} - \frac{1}{2}V_{t_i}(t_{i+1} - t_i) + \sqrt{V_{t_i}}\left(W_{t_{i+1}} - W_{t_i}\right), \qquad \text{for } i = 0, \ldots, n_T - 1;$$

 (vi) compute the expectation by averaging the payoff over all terminal values of each path.

4.2. **Calibration of SPX options via VIX Futures.** We first follow the calibration algorithm 3.17 to obtain H and ξ_0, and we then aim at minimising, over (ν, ρ), the objective function

$$(4.1) \qquad\qquad \mathcal{L}^{\mathrm{C}}(\nu, \rho) := \sum_{j=1}^{L}\sum_{i=1}^{N}(\mathrm{C}_{T_{i,j}} - \mathrm{C}_{i,j}^{\mathrm{obs}})^2,$$

where $\mathrm{C}_{T_{i,j}}$ is the Call price given by the rough Bergomi model, computed using the scheme introduced in Section 4.1, with maturity T_i and strike $K^{(j)}$. On the other hand, $(\mathrm{C}_{i,j}^{\mathrm{obs}})_{i,j}$ is the set observed Call prices in the time grid $T_1 < \ldots < T_N$ and strike grid $K^{(1)} < \ldots < K^{(L)}$. In order to optimise the calibration algorithm, we first compute the Volterra process $\mathcal{V}$, which will then be used in a forward Euler simulation at each calibration step:

---

[2]CBOE delayed option quotes: http://www.cboe.com/delayedquote/quotetable.aspx

[3]CBOE VIX futures historical data: http://cfe.cboe.com/data/historicaldata.aspx

Algorithm 4.2 (Calibration algorithm for SPX options via VIX Futures).

(i) Calibrate H and ξ_0 using the VIX Futures;

(ii) compute M paths of the Volterra process, $\{\mathcal{V}^{(u)}\}_{u=1}^M$ and extract the Brownian motions $\{Z^{(u)}\}_{u=1}^M$ driving each process. Also, compute independent Brownian motions $\{Z^{\perp(u)}\}_{u=1}^M$;

(iii) evaluate the Call prices in each calibration step:

$$
\begin{aligned}
V_t^{(u)} &= \xi_0(t)\mathcal{E}\left(2\nu C_H \mathcal{V}_t^{(u)}\right), & u &= 1,\dots,M, \\
W^{(u)} &= \rho Z^{(u)} + \sqrt{1-\rho^2}Z^{\perp(u)}, & u &= 1,\dots,M, \\
S_{t+\Delta}^{(u)} &= S_t^{(u)} + S_t^{(u)}\sqrt{v_t^{(u)}}\left(W_{t+\Delta}^{(u)} - W_t^{(u)}\right), & u &= 1,\dots,M;
\end{aligned}
$$

(iv) compute the Call price for each available maturity $\{T_1,\dots,T_N\}$ and set of strikes $\{K^{(1)},\dots,K^{(L)}\}$:

$$
\mathrm{C}_{T_i,j} = \frac{1}{M}\sum_{u=1}^M (S_{T_i}^{(u)} - K^{(j)})_+, \quad \text{for } i=1,\dots,N \text{ and } j=1,\dots,L;
$$

(v) minimise over (ν,ρ) the objective function $\mathcal{L}^{\mathrm{C}}(\nu,\rho)$ in (4.1).

Remark 4.3. Item (v) in the algorithm above may change the optimal values for ν, which was initially calibrated in (i) from the VIX Futures. Backtesting however shows that the calibration in (i) is not really affected by this.

4.3. **Results.** We calibrate the model on December 4, 2015, fixing $H = 0.09237$ obtained previously through VIX, and plot the fit in Figure 13. The model is not fully consistent for short maturities, which may follow from the inability of ξ_0 to fully capture the smiles for these maturities, but the fit greatly improves with maturity. Interestingly, we observe a 20% difference between the the parameter ν obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?). Nevertheless, we emphasise the importance of an accurate ξ_0 curve to improve the fit to SPX and to provide an efficient joint calibration.
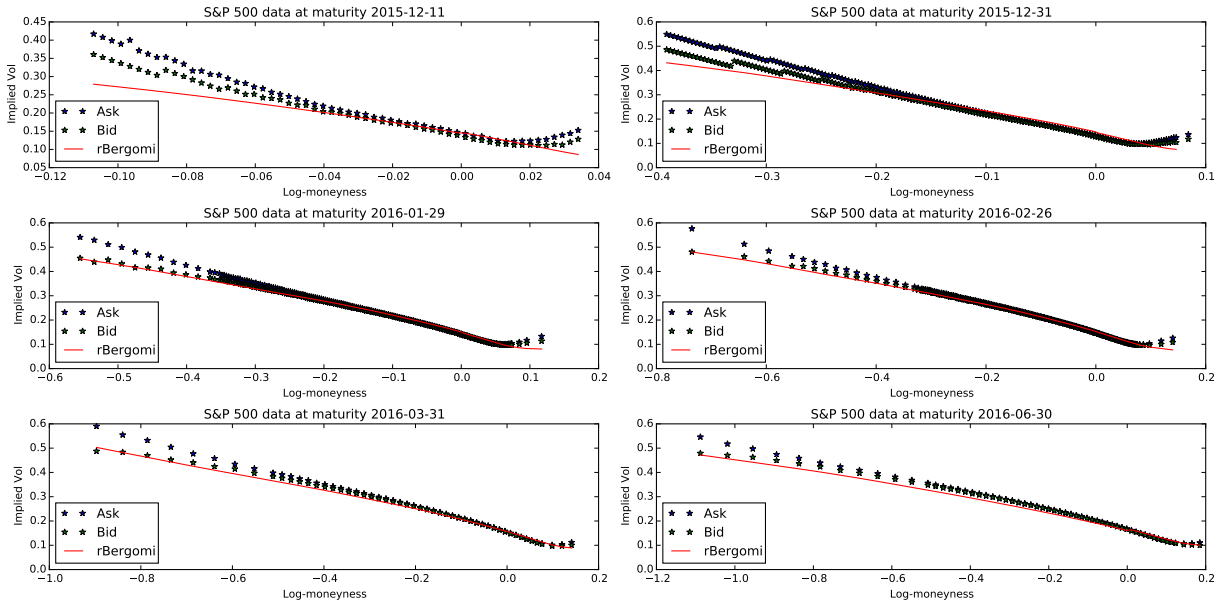

FIGURE 13. Calibration of SPX smiles on 4/12/2015. Calibrated parameters: $(\nu,\rho) = (1.19, -0.999)$.

Conclusion

Following the path set by Bayer, Friz and Gatheral [3], we developed here a relatively fast algorithm to calibrate VIX Futures and the VIX smile, consistently with the SPX smile, in the rough Bergomi model. The clear strength of this model is that only a few parameters are needed, making the (re)calibration robust and stable. From a trader's point of view, we highlight some potential market discrepancy between the VIX and the SPX, and leave a refined analysis thereof for future research.

Appendix A. The hybrid scheme

We briefly recall the hybrid scheme developed in [4]. Following the notation in Definition 2.1, we consider a (truncated) Brownian semistationary process $\mathcal{B}(\alpha, W)$, and introduce the truncation parameter $\kappa \in \mathbb{N}$. On an equidistant grid $\mathcal{T} := \{t_i = i/n\}_{i=0,\ldots,n_T}$, with $n_T := \lfloor nT \rfloor$, for $n \geq 2$, under Definition 2.1, the hybrid scheme for the BSS process $\mathcal{B}$ is approximated by $\mathcal{B}_n(t_i) = \widetilde{\mathcal{B}}_n(i) + \widehat{\mathcal{B}}_n(i)$ with

$$\widetilde{\mathcal{B}}_n(i) = \sum_{k=1}^{i \wedge \kappa} L_g\left(\frac{k}{n}\right) \sigma\left(\frac{i-k}{n}\right) \overline{W}_{i-k,k} \qquad \text{and} \qquad \widehat{\mathcal{B}}_n(i) = \sum_{k=\kappa+1}^{i} g\left(\frac{b_k^*}{n}\right) \sigma\left(\frac{i-k}{n}\right) \overline{W}_{i-k},$$

with L_g introduced in Definition 2.1, and where

$$(\text{A.1}) \quad \overline{W}_i := \int_{t_i}^{t_{i+1}} \mathrm{d}W_s, \quad \overline{W}_{i,k} := \int_{t_i}^{t_{i+1}} (t_{i+k} - s)^\alpha \, \mathrm{d}W_s, \quad b_k^* = \left(\frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha+1}\right)^{1/\alpha}, \text{ for } k \geq \kappa+1.$$

For any i, k, the random variables $\overline{W}_i$ and $\overline{W}_{i,k}$ are centred Gaussian with the following covariance structure:

$$\mathbb{E}\left(\overline{W}_{i,k}\overline{W}_i\right) = \frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{n^{\alpha+1}\alpha + 1}, \qquad \text{and} \qquad \mathbb{E}\left(\overline{W}_{i,k}\overline{W}_j\right) = 0, \quad \text{for } k \neq j,$$

$$\mathbb{E}\left(\overline{W}_{i,k}\overline{W}_{i,j}\right) = \int_0^{1/n} \left(\frac{k}{n} - u\right)^\alpha \left(\frac{j}{n} - u\right)^\alpha \mathrm{d}u, \text{ for } k \neq j, \quad \mathbb{V}\left(\overline{W}_{i,k}\right) = \frac{k^{2\alpha+1} - (k-1)^{2\alpha+1}}{n^{2\alpha+1}2\alpha + 1}, \quad \mathbb{V}\left(\overline{W}_i\right) = \frac{1}{n}.$$

References

[1] E. Alòs, J. León and J. Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, **11**(4), 571-589, 2007

[2] M. Abramowitz and I.A. Stegun. Handbook of mathematical functions. *Washington, D.C.*, 1965.

[3] C. Bayer, P. Friz and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*: 1-18, 2015.

[4] M. Bennedsen, A. Lunde and M. S. Pakkanen. Hybrid scheme for Brownian semistationary processes. arXiv:1507.03004, 2016.

[5] L. Bergomi. Smile dynamics II. *Risk*: 67-73, October 2005.

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. Regular variation. *Cambridge University Press*, 1989.

[7] F. Comte and E. Renault. Long memory continuous time models. *Journal of Econometrics*, **73**(1): 101-149, 1996.

[8] P. Carr and D. Madan. Towards a theory of volatility trading. *Risk Publications*: 417-427, 1998.

[9] P. Carr and D. Madan. Joint modeling of VIX and SPX options at a single and common maturity with risk management applications. *IIE Transactions*, **46**(11): 1125-1131, 2014.

[10] C. Doléans-Dade. Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **16**: 181-194, 1970.

[11] D. Dufresne. The log-normal approximation in financial and other computations. *Adv. in App. Prob*, **36**: 747-773, 2004.

[12] M. Fukasawa. The normalizing transformation of the implied volatility smile. *Mathematical Finance*, **22**(4): 753-762, 2012.

[13] J. Gatheral. The volatility surface: a practitioner's guide. *Hoboken, NJ: John Wiley and Sons*, 2006.

[14] J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is rough. arXiv: 1410.3394, 2014.

[15] J. Gatheral. Consistent modelling of SPX and VIX options. *Presentation, Bachelier Congress, London*, 2008.

[16] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, **14**(1): 59-71, 2014.

[17] P.S. Hagan, D. Kumar, A. Lesniewski and D.E. Woodward. Managing smile risk. *Wilmott Magazine*: 84-108, 2002.

[18] S. Hendriks and C. Martini. The eSSVI volatility surface. *'At the frontiers of quantitative finance', ICMS, Edinburgh, 2016.*

[19] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2): 327-342, 1993.

[20] T. Kokholm and M. Stisen. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance*, 16: 1, 27-48, 2015.

[21] C. Martini. An explicit formula for the VIX in eSSVI. *Slides 'At the frontiers of quantitative finance', ICMS, Edinburgh, 2016.*

Department of Mathematics, Imperial College London
E-mail address: `a.jacquier@imperial.ac.uk`

Zeliade Systems, Paris
E-mail address: `c.martini@zeliade.com`

Department of Mathematics, Imperial College London
E-mail address: `aitor.muguruza-gonzalez15@imperial.ac.uk`