Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

On VIX futures in the rough Bergomi model

Antoine Jacquier, Claude Martini & Aitor Muguruza

To cite this article: Antoine Jacquier, Claude Martini & Aitor Muguruza (2017): On VIX futures in the rough Bergomi model, Quantitative Finance, DOI: 10.1080/14697688.2017.1353127

To link to this article: http://dx.doi.org/10.1080/14697688.2017.1353127

Published online: 25 Aug 2017.

Quantitative Finance, 2017
https://doi.org/10.1080/14697688.2017.1353127

On VIX futures in the rough Bergomi model

ANTOINE JACQUIER*†, CLAUDE MARTINI‡ and AITOR MUGURUZA†

†Department of Mathematics, Imperial College London, London, UK
‡Zeliade Systems, Paris, France

(*Received 16 January 2017; accepted 30 June 2017; published online 25 August 2017*)

The rough Bergomi model introduced by Bayer *et al.* [*Quant. Finance*, 2015, 1–18] 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 *et al.* [*Finance Stoch.*, forthcoming].

Keywords: Implied volatility; Fractional Brownian motion; Rough Bergomi; VIX futures; VIX smile

JEL Classification: 91G20, 91G99, 91G60, 91B25

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 1993, SABR 2002, Bergomi 2005) are constructed under $\mathbb{Q}$ and are of Markovian nature (making pricing, and hence calibration, easier). In contrast, Alòs *et al.* (2007) and Fukasawa (2011), and more recently, Gatheral *et al.* (2014) 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 Alòs *et al.* 2007, Fukasawa 2011), 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 (2008) 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 (2014) and Kokholm and Stisen (2015). This issue was briefly tackled by Bayer *et al.* (2015) 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 Bayer *et al.* (2015) 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 Bayer *et al.* (2015), we adapt the hybrid-scheme by Bennedsen *et al.* (forthcoming) 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 organization 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.

*Corresponding author. Email: a.jacquier@imperial.ac.uk

2. Rough volatility and the rough Bergomi model

Comte and Renault (1996) 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 $(1/2, 1)$. Originally introduced by Alòs *et al.* (2007) and Fukasawa (2011), and recently promoted further by Gatheral *et al.* (2014), stochastic volatility models with Hurst index smaller than $1/2$ are able to produce 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 classical Markovian stochastic volatility models. Our work focuses on the pricing measure $\mathbb{Q}$ and we assume through this paper that the model presented by Gatheral *et al.* (2014) under $\mathbb{P}$ is a reasonable model. Finally, and most importantly, we follow the recent paper by Bayer *et al.* (2015), in order to extend the RFSV model to pricing schemes under the measure $\mathbb{Q}$. More precisely, Bayer *et al.* (2015) 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 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

$$\mathbb{E}(\mathcal{V}_t\mathcal{V}_s) = s^{2H}\int_0^1 \left(\frac{t}{s}-u\right)^{H_-}(1-u)^{H_-}\mathrm{d}u,$$

for any $s, t \in [0, 1]$.

We shall also introduce, for any $0 \leq t_1 \leq t_2 \leq t$, the notation

$$\mathcal{V}_t^{[t_1,t_2]} := \int_{t_1}^{t_2} (t-u)^{H_-}\mathrm{d}Z_u, \tag{2.3}$$

and write $\mathcal{V}_t^{t_2} = \mathcal{V}_t^{[0,t_2]}$ whenever the lower bound of the integral is null. 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, $\mathcal{V}^T$ is a centred Gaussian process with variance

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

and covariance structure

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

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

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

and ${}_2F_1$ is the hypergeometric function (Abramowitz and Stegun 1972, Chapter 15).

2.1. Hybrid simulation scheme

Bayer *et al.* (2015) used a Cholesky decomposition approach 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 *et al.* (forthcoming) presented a new simulation scheme for Brownian semistationary ($\mathrm{t}\mathcal{BSS}$) processes. This method, as opposed to Cholesky, is an approximate method. However, in Bennedsen *et al.* (forthcoming) 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 ($\mathrm{t}\mathcal{BSS}$) 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 $\mathrm{t}\mathcal{BSS}(\alpha, W)$ process if furthermore

(i) there exists $\alpha \in \left(-\frac{1}{2}, \frac{1}{2}\right)\backslash\{0\}$ 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‡ 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 Bennedsen *et al.* (forthcoming) and recalled in appendix 1, provides an efficient way to simulate $\mathrm{t}\mathcal{BSS}$ processes, and in particular applies to the rough Bergomi model. Indeed, from (2.2), $g(x) \equiv x^{H_-}$ and $\sigma(\cdot) \equiv 1$ as in definition 2.1; since

†For a continuous semimartingale Y starting from zero, the Doléans-Dade stochastic exponential (Doléans-Dade 1970) is defined as $\mathcal{E}(Y)_t := \exp(Y_t - \frac{1}{2}[Y, Y]_{0,t})$. Here, $\mathrm{Y} = \mathcal{V}$ is a zero-mean Gaussian process, with infinite quadratic variation, and this definition does not make sense. However, the Wick exponential $\mathcal{E}(Y)_t := \exp(Y_t - \frac{1}{2}\mathbb{E}(Y_t^2))$ does, and we conveniently keep this notation without further details, following Bayer *et al.* (2015, Remark 1.1).

‡A measurable function $L : (0, 1] \to [0, \infty)$ is slowly varying Bingham *et al.* (1989) at 0 if for any $t > 0$, $\lim_{x\downarrow 0} L(tx)/L(x) = 1$.

$H_- \in (-\frac{1}{2}, 0)$, then $L_g \equiv 1$, and $\mathcal{V}$ satisfies definition 2.1(i). Definition 2.1(ii) trivially holds, so that the Volterra process $\mathcal{V}$ in (2.2) is a $t\mathcal{BSS}(H_-, Z)$ process.

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 $n_T := \lfloor nT \rfloor$)

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

where the coefficients $\{b_i^*\}$ are defined in (A1). 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 Bennedsen *et al.* (forthcoming).

3. Rough Bergomi and VIX

Bayer *et al.* (2015) 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 Bayer *et al.* (2015).

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

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

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

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

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

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

PROPOSITION 3.1 *The VIX dynamics are given by*

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

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

$$\begin{aligned} \mathrm{VIX}_T^2 &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left(V_s | \mathcal{F}_T\right) \mathrm{d}s \\ &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\Big[\xi_0(t) \mathcal{E}\left(2\nu C_H \mathcal{V}_t\right) | \mathcal{F}_T \Big] \mathrm{d}t \\ &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\Bigg[\xi_0(t) \eta_T(t) \\ &\qquad \times \exp\left(2\nu C_H \mathcal{V}_t^{[T,t]} - \frac{\nu^2 C_H^2 t^{2H}}{H} \right) \Bigg| \mathcal{F}_T \Bigg] \mathrm{d}t, \end{aligned}$$

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

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

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

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. Since $\mathbb{E}\left(V_t | \mathcal{F}_T\right) = \xi_T(t)$ by (3.1), the equality

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

yields that the following representation for the forward variance curve ξ_T in the rough Bergomi model:

$$\xi_T(t) = \xi_0(t) \eta_T(t) \exp\left(\frac{\nu^2 C_H^2}{H} \left[(t-T)^{2H} - t^{2H} \right] \right), \quad \text{for any } t \geq T. \tag{3.3}$$

Bayer *et al.* (2015) did not derive such a representation, and their approach for pricing VIX derivatives relies on an approximation which avoids the computations developed in this section. The identity (3.3) allows for a better understanding of the process ξ_T, and for an innovative approach to price VIX derivatives.

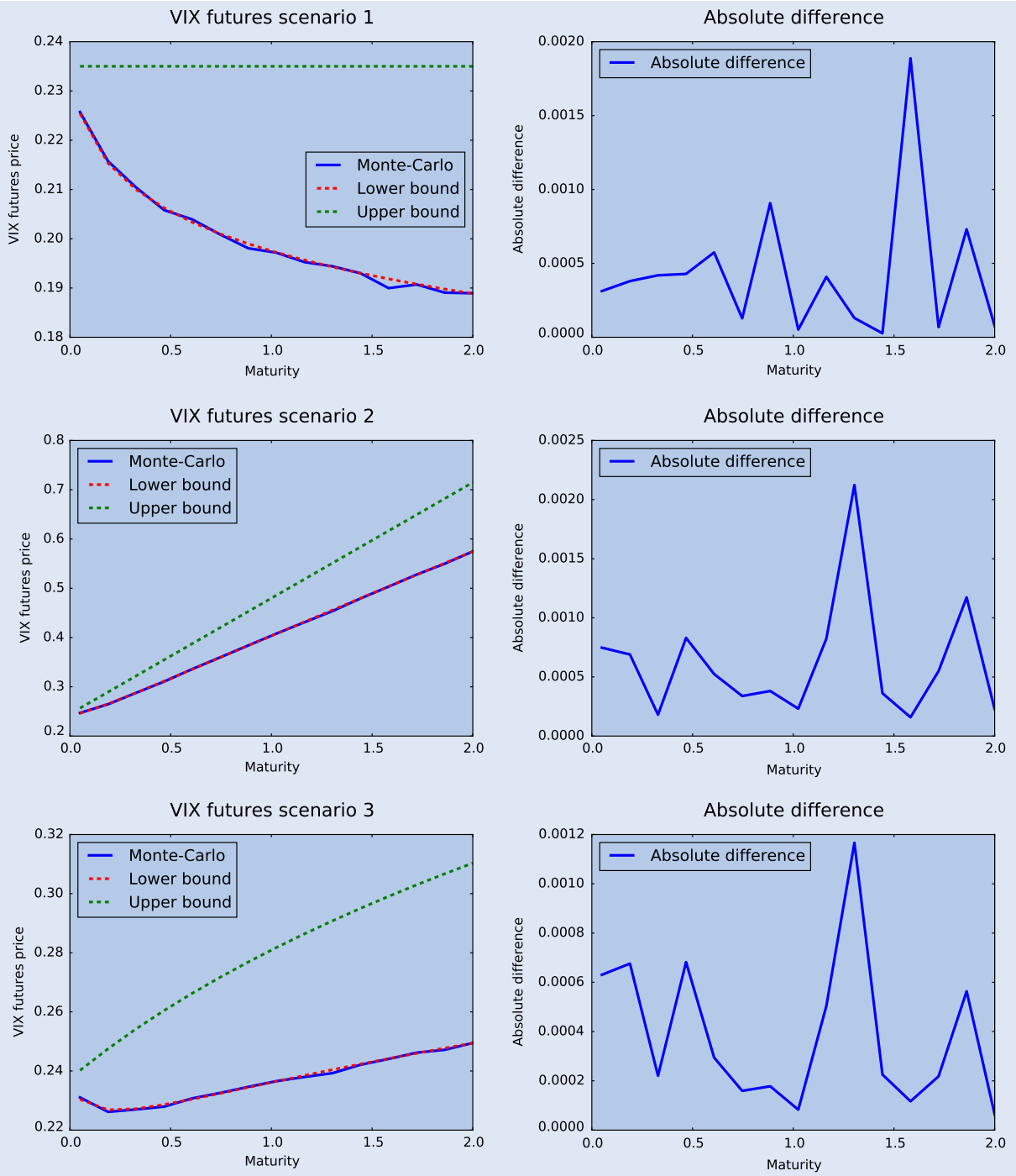

Figure 1. Bounds vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

3.2. *Upper and lower bounds for VIX futures*

THEOREM 3.2 *The following bounds hold for VIX futures:*

$$\frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t \leq \mathfrak{V}_T \leq \left\{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s\right\}^{1/2}. \tag{3.4}$$

Proof. The conditional Jensen's inequality gives

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

Furthermore, since ξ_0 is $\mathcal{F}_0$-adapted, Fubini's theorem along with the martingale property of ξ_T yield the upper bound $\mathfrak{V}_T = \mathbb{E}\left(\mathrm{VIX}_T|\mathcal{F}_0\right) \leq \sqrt{\Delta^{-1}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s}$. To obtain a lower

bound we use the representation in (3.3), and Cauchy–Schwarz's inequality, and Fubini's theorem, so that

$$\begin{aligned}
\mathfrak{V}_T &= \mathbb{E}\left(\mathrm{VIX}_T|\mathcal{F}_0\right)\\
&= \mathbb{E}\left[\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2C_H^2}{H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t}\,\middle|\,\mathcal{F}_0\right]\\
&\geq \mathbb{E}\left[\frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)\eta_T(t)}\exp\left(\frac{\nu^2C_H^2}{2H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t\,\middle|\,\mathcal{F}_0\right]\\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\mathbb{E}\left(\sqrt{\eta_T(t)}\right)\exp\left(\frac{\nu^2C_H^2}{2H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t\\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2C_H^2}{4H}\left[t^{2H}-(t-T)^{2H}\right]\right)\\
&\quad\times\exp\left(\frac{\nu^2C_H^2}{2H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t\\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2C_H^2}{4H}\left[(t-T)^{2H}-t^{2H}\right]\right)\mathrm{d}t,
\end{aligned}$$

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

We perform a numerical experiment to check the tightness of the bounds obtained in proposition 3.2). For this analysis we consider three qualitative scenarios for the initial forward variance curve:

$$\begin{aligned}
\text{Scenario 1}: &\ \xi_0(t) = 0.234^2;\\
\text{Scenario 2}: &\ \xi_0(t) = 0.234^2(1+t)^2;\\
\text{Scenario 3}: &\ \xi_0(t) = 0.234^2\sqrt{1+t}.
\end{aligned}\tag{3.5}$$

Figure 1 suggests that the lower bound given in proposition 3.2) is surprisingly tight for very different shapes 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

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

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

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

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

$$\phi(T)^2-\psi(T)^2 = \Delta^2\left(\frac{f''(T)}{6}-\frac{f'(T)^2}{16f(T)}\right)+\mathcal{O}(\Delta^3).\tag{3.6}$$

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.

Remark 3.3 Recall that $f(t) = \sqrt{\xi_0(t)}\exp(\frac{\nu^2C_H^2}{4H}[(t-T)^{2H}-t^{2H}]) > 0$ has a financial interpretation and cannot be arbitrarily small. Therefore, the term $\frac{f'(T)^2}{f(T)}$ in (3.6) does not explode.

The use of the lower bound as a proxy for the price allows one, in principle, to invert the price and calibrate ξ_0 consistent with the VIX as a functional of H, ν and the current term structure of VIX futures. This approach is left for future research.

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 summarized as follows:

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

(1) Simulate the Volterra process $(\mathcal{V}_t)_{t\in[0,T]}$ using the hybrid scheme in appendix 1, 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.

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

(3) fix a grid $\mathfrak{T} = \{\tau_j\}_{j=0,\dots,N}$ 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:

$$\begin{aligned}
\widetilde{\mathcal{V}}_{\tau_0}^T &:= \mathcal{V}_T^T \qquad\text{and}\\
\widetilde{\mathcal{V}}_{\tau_j}^T &:= \sum_{i=1}^{n_T}\frac{Z_{t_i}-Z_{t_{i-1}}}{(\tau_j-t_{i-1})^{-H_-}}, \quad \text{for } j=1,\dots,N;
\end{aligned}$$

(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 $Q_{T,\tau_j}^2 := \xi_0(\tau_j)\exp\left(2\nu C_H\widetilde{\mathcal{V}}_{\tau_j}^T\right)\exp(\frac{\nu^2C_H^2}{H}((\tau_j-T)^{2H}-\tau_j^{2H}))$.

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.

Remark 3.6 In order to analyse the convergence of the scheme, consider equidistant grids $\mathcal{T}$ and $\mathfrak{T}$ with increment sizes $h_{\mathcal{T}}$ and $h_{\mathfrak{T}}$, respectively. In the first three steps, the Hybrid and forward Euler schemes yield an error of order $\mathcal{O}(h_{\mathcal{T}})$, hence we have $\mathcal{V}^T = \widetilde{\mathcal{V}}^T + \mathcal{O}(h_{\mathcal{T}})$. In addition, the trapezoidal rule (Atkinson 1989, Chapter 5) in step 4 yields an error of order $\mathcal{O}(h_{\mathfrak{T}}^2)$. Therefore, the total error of convergence is of order $\mathcal{O}(h_{\mathcal{T}})$.

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 $n_T = 8$ discretization 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 $\text{corr}(\mathcal{V}^T_{t_1}, \mathcal{V}^T_{t_1+\varepsilon}) < \text{corr}(\mathcal{V}^T_{t_2}, \mathcal{V}^T_{t_2+\varepsilon})$ holds for all $T < t_1 < t_2$, as well as the following:

PROPOSITION 3.7 *The limit* $\lim_{\varepsilon \downarrow 0} corr\left(\mathcal{V}^T_t, \mathcal{V}^T_{t+\varepsilon}\right) = 1$ *holds for any* $t \in [T, T+\Delta]$.

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

$$
\begin{aligned}
&E\left[\left(\mathcal{V}^T_{t+\varepsilon} - \mathcal{V}^T_t\right)^2\right] \\
&= \int_0^T [(t+\varepsilon-u)^{H_-} - (t-u)^{H_-}]^2 \mathrm{d}u \\
&= \int_0^T \Big[(t+\varepsilon-u)^{2H_-} - 2[(t+\varepsilon-u)(t-u)]^{H_-} \\
&\quad + (t-u)^{2H_-}\Big] \mathrm{d}u.
\end{aligned}
$$

By monotone convergence, the sum of the first two terms converges to the third one, and the result follows. □

In light of proposition 3.7, 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 $\rho_i := \text{corr}(\mathcal{V}^T_{t_{8+i}}, \mathcal{V}^T_{t_{8+i+1}})$, 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 3.8 (VIX simulation (truncated Cholesky)) Fix a grid $\mathfrak{T} = \{\tau_j\}_{j=1,\ldots,N}$ on $[T, T+\Delta]$,

(i) compute the covariance matrix of $(\mathcal{V}^T_{\tau_j})_{j=1,\ldots,8}$ using (2.5) and simulate the process using the Cholesky method;
(ii) generate $\{\mathcal{V}^T_{\tau_j}\}_{j=1,\ldots,N}$ by correlating and rescaling using (2.5):

$$
\begin{aligned}
\mathcal{V}^T_{\tau_j} = \sqrt{\mathbb{V}(\mathcal{V}^T_{\tau_j})} \Bigg(& \frac{\rho(\mathcal{V}^T_{\tau_{j-1}}, \mathcal{V}^T_{\tau_j})\mathcal{V}^T_{\tau_{j-1}}}{\sqrt{\mathbb{V}(\mathcal{V}^T_{\tau_{j-1}})}} \\
& + \sqrt{1 - \rho(\mathcal{V}^T_{\tau_{j-1}}, \mathcal{V}^T_{\tau_j})^2}\mathcal{N}(0,1)\Bigg), \\
& \text{for } j = 9, \ldots, N;
\end{aligned}
$$

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

3.3.3. Singular value decomposition approach. As the numerical considerations in the previous section indicated, all Gaussian simulation schemes based on the computation of the determinant (including Cholesky) are not feasible to simulate the process $\mathcal{V}^T$. This is due to the fact that the driving Brownian motion in the stochastic integral representation of $\left(\mathcal{V}^T_t\right)_{0 \le t \le T}$ is always the same for all t. In addition, since $\mathcal{V}^T$ and its increments are non-stationary, many 'exact' methods (such as circulant embedding, Fourier methods or wavelets) are not applicable. This, however, can be bypassed using a Singular Value Decomposition (SVD) in order to compute the square root of the covariance matrix.

Algorithm 3.9 (VIX simulation using SVD) Fix a grid $\mathfrak{T} = \{\tau_j\}_{j=1,\ldots,N}$ on $[T, T+\Delta]$,

(i) compute the covariance matrix of $(\mathcal{V}^T_{\tau_j})_{j=1,\ldots,N}$ using (2.5);
(ii) perform the SVD decomposition of the covariance matrix obtaining $U\Lambda U^*$, where U is a unitary matrix and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ a diagonal matrix of positive singular values;
(iii) draw a sample of a standard Gaussian $Z \sim \mathcal{N}(0, I_N)$ and generate

$$\mathcal{V}^T \sim U\Lambda^{1/2}U^*Z,$$

where $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$;
(iv) compute the VIX by numerical integration as in algorithm 3.4(4).

Remark 3.10 The complexity of the SVD decomposition (Golub and Van Loan 1996, Algorithm 8.6.2) is $\mathcal{O}(N^3)$.

Remark 3.11 Recall that the simulation of $\mathcal{V}^T$ using SVD is exact, since U being unitary implies $U\Lambda^{1/2}U^*Z \sim \mathcal{N}(0, C)$ where $C = U\Lambda^{1/2}U^*U\Lambda^{1/2}U^* = U\Lambda U^*$. However, the numerical integration in step 4 produces an overall error of order $\mathcal{O}(h_{\mathfrak{T}}^2)$ where $h_{\mathfrak{T}} = \max_{1<j<N-1}(\tau_{j+1} - \tau_j)$.

3.3.4. 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 Bayer *et al.* (2015) and Bennedsen *et al.* (forthcoming):

$$
\begin{aligned}
&\xi_0 = 0.235^2, \qquad H = 0.07, \\
&\nu = 1.9\frac{C_H\sqrt{2H}}{2} \approx 1.2287, \qquad \kappa = 2.
\end{aligned} \tag{3.7}
$$

We perform 10^6 simulations for all three methods, with 300 equidistant points for $\mathcal{T}$ and, in the case of the HS+FE, 500 equidistant points for $\mathfrak{T}$. Figure 2 suggests that all methods agree qualitatively and converge to a similar output. Figure

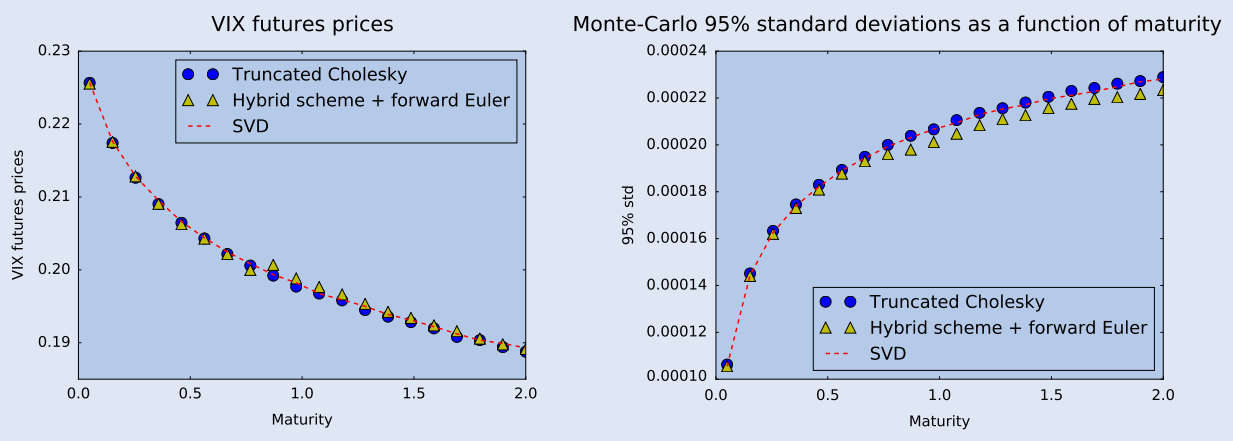

Figure 2. Monte-Carlo benchmarking for all three methods with 10^6 simulations and the corresponding 95% confidence level standard deviations.

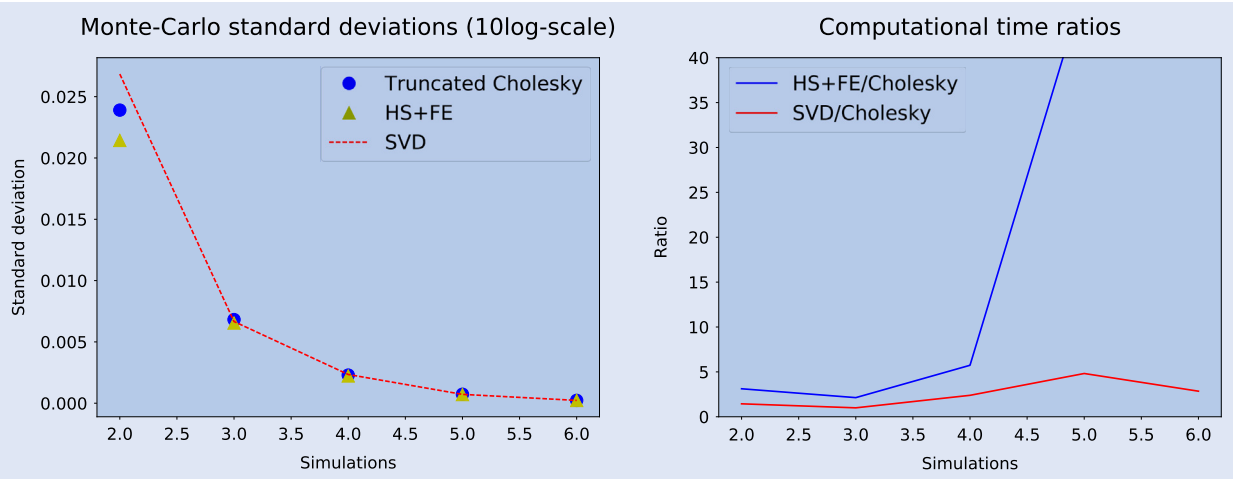

Figure 3. Monte-Carlo standard deviations and computational time ratios for a fixed maturity ($T = 2$ years) for all methods using efficient parallel computing.

2 also indicates that the Monte-Carlo standard deviation increases in T for all schemes. Figure 3 shows that a large number of simulations is needed in all cases to obtain reliable output. On the other hand, figure 3 shows that the HS+FE method is much slower than the other two methods. Also, the truncated Cholesky method is 2.5 times faster than the SVD, which is consistent with the computational complexities discussed in the previous section. In light of this analysis, all three methods seem to approximate the prices similarly well. The computational time, however, is still too slow for reasonable calibration, and we shall only use the truncated Cholesky approach as a benchmark for the approximations in the next sections.

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 $\Delta\text{VIX}_T^2 = \int_T^{T+\Delta} \xi_T(t)\mathrm{d}t$. In Bayer *et al.* (2015), assumed that $\log(\Delta\text{VIX}_T^2)$ follows a Gaussian distribution, and computed directly its first and second moments, which hence fully characterize the distribution of ΔVIX_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 ΔVIX_T^2 is. In a different context (geometric Brownian motion and Asian options), Dufresne (2004) 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 (Dufresne 2004), 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 3.12 *The following holds:*

$$\begin{aligned}\sigma^2 &:= \mathbb{V}(\log(\Delta\text{VIX}_T^2)) \\ &= -2\log\mathbb{E}(\Delta\text{VIX}_T^2) + \log\mathbb{E}\left((\Delta\text{VIX}_T^2)^2\right),\end{aligned}$$

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

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

$$\mathbb{E}\left[(\Delta \mathrm{VIX}_T^2)^2\right] = \int_{[T,T+\Delta]^2} \xi_0(u)\xi_0(t) \times \exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H}\right]\right\} \times \mathrm{e}^{\overline{\Theta}_{u,t}}\mathrm{d}u\mathrm{d}t. \tag{3.8}$$

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

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

Remark 3.13 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 3.14 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 $x \leq 1$ (details on the radius of convergence of hypergeometric functions can be found in Abramowitz and Stegun (1972, p. 556)), even though the integral representation (3.8) 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 3.12 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 (3.3), we get

$$\mathbb{E}\left(\xi_T(u)\xi_T(t)\right) = \xi_0(u)\xi_0(t) \times \exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} - u^{2H} + (t-T)^{2H} - t^{2H}\right]\right\}\mathbb{E}\left(\mathrm{e}^{\vartheta_{u,t}}\right), \tag{3.9}$$

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

$$\begin{aligned}\mathbb{V}(\vartheta_{u,t}) &= 4\nu^2 C_H^2\left(\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + 2\int_0^T (u-s)^{H_-}(t-s)^{H_-}\mathrm{d}s\right)\\ &= 4\nu^2 C_H^2\left\{\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + \frac{2(u-t)^{H_-}}{H_+}\left[t^{H_+}\mathrm{F}\left(\frac{-t}{u-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{u-t}\right)\right]\right\}.\end{aligned}$$

Then,

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

and the second moment follows from the immediate computation

$$\begin{aligned}\mathbb{E}\left((\Delta \mathrm{VIX}_T)^2\right) &= \int_T^{T+\Delta}\int_T^{T+\Delta} \mathbb{E}\left[\xi_T(u)\xi_T(t)\right]\mathrm{d}u\mathrm{d}t\\ &= \int_T^{T+\Delta}\int_T^{T+\Delta} \xi_0(u)\xi_0(t) \times \exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H}\right]\right\} \times \mathrm{e}^{\frac{1}{2}\mathbb{V}(\vartheta_{u,t})}\mathrm{d}u\mathrm{d}t,\end{aligned}$$

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 consider the following approximation:

APPROXIMATION 3.15 $\log(\Delta \mathrm{VIX}_T^2)$ *is approximately Gaussian with mean* μ *and* σ^2.

In fact, this is almost the same as in Bayer *et al.* (2015); however, they did not compute the variance exactly as in proposition 3.12, and instead considered the lognormal approximation

APPROXIMATION 3.16 (BAYER- - FRIZ- - GATHERAL) $\log(\Delta \mathrm{VIX}_T^2)$ *is approximately Gaussian with mean* $\tilde{\mu}$ *and variance* $\tilde{\sigma}^2$ *given by*

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

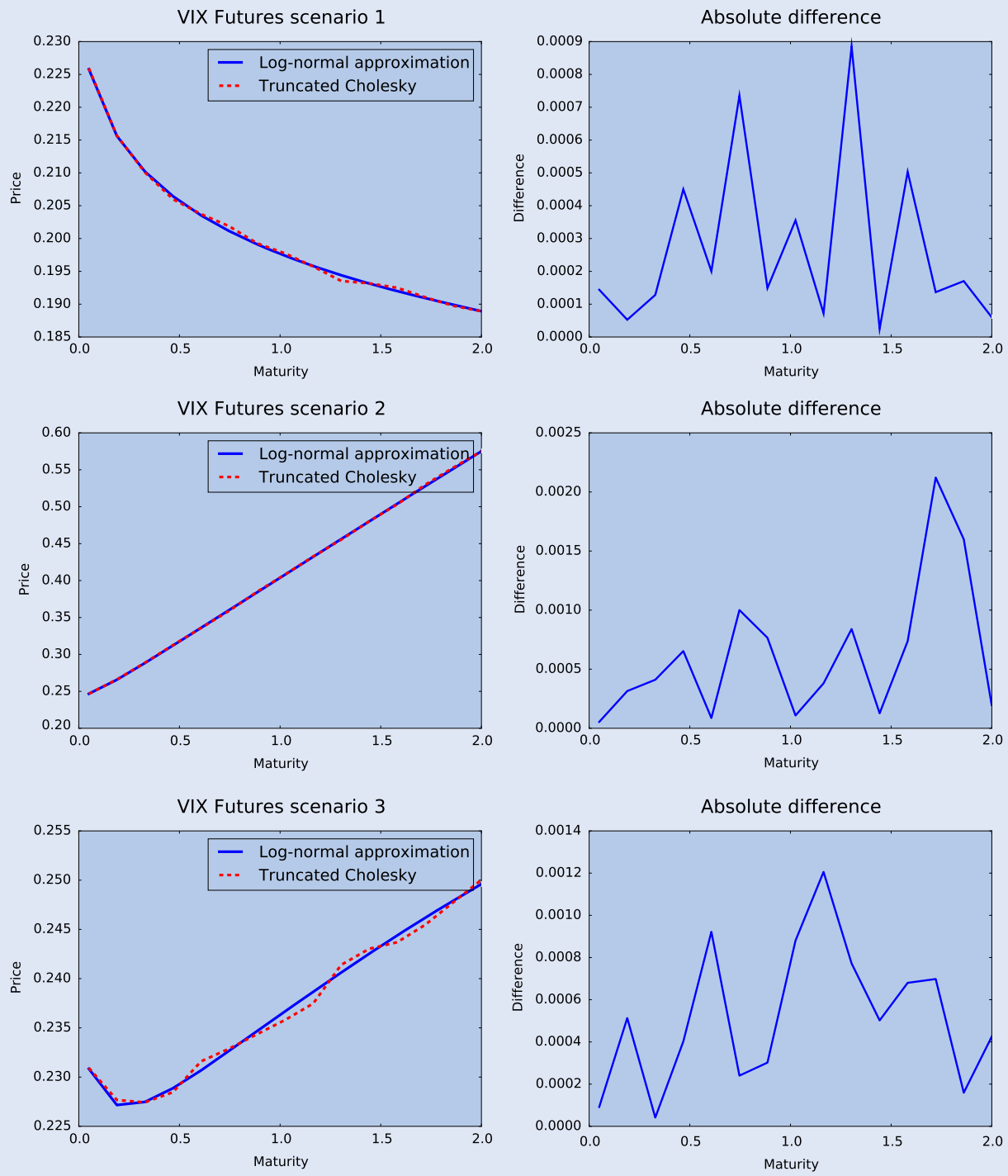

Figure 4. Log-normal approximations vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

LEMMA 3.17 *A VIX future is worth*

$$\mathfrak{V}_T = \begin{cases} \sqrt{\dfrac{1}{\Delta}\displaystyle\int_T^{T+\Delta} \xi_0(t)\mathrm{d}t}\, \exp\left(-\dfrac{\sigma^2}{8}\right), \\ \qquad \textit{under approximation 3.15}, \\ \sqrt{\dfrac{1}{\Delta}\displaystyle\int_T^{T+\Delta} \xi_0(t)\mathrm{d}t}\, \exp\left(-\dfrac{\tilde{\sigma}^2}{8}\right), \\ \qquad \textit{under approximation 3.16}. \end{cases}$$

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

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

and the second case follows analogously. □

As opposed to the simulation schemes, the log-normal approximation does depend on the qualitative properties 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 approximation 3.15 to obtain a closed-form Black-Scholestype formulae for European options on the VIX.

LEMMA 3.18 *For $0 \le t \le T$, 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 approximation 3.15, a European Call option on a VIX future maturing at T is worth*

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

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 3.12.

Proof. Under approximation 3.15, the lemma follows directly from the following trivial computations:

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

□

3.4.1. Numerical tests of the log-normal approximation on VIX futures. We perform a numerical analysis of the approximation by pricing VIX futures using the parameters in (3.7). We consider 10^6 simulations and the three qualitative scenarios introduced in (3.5) for the forward variance curve. Figure 4 suggests that the approximation is accurate 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 5 shows that the log-normal approximation is accurate for at-the-money 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 (3.7). Figures 6 and 7 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 6 and 7. 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 6 our approximation was 40 times slower than Bayer, Friz and Gatheral's.

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 *et al.* (2015). Even if our approximation seems to be more accurate the computational cost is much larger and the difference 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 optimization and calibration is extremely useful.

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

$$\mathcal{L}^{\mathfrak{F}}(\nu, H) := \sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)^2, \tag{3.10}$$

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

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

which is the closed-form expression of the variance in approximation 3.16. The gradient of the objective function is an important source of information in many optimization algorithms. To compute it, we differentiate the objective function in (3.10) 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

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

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

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

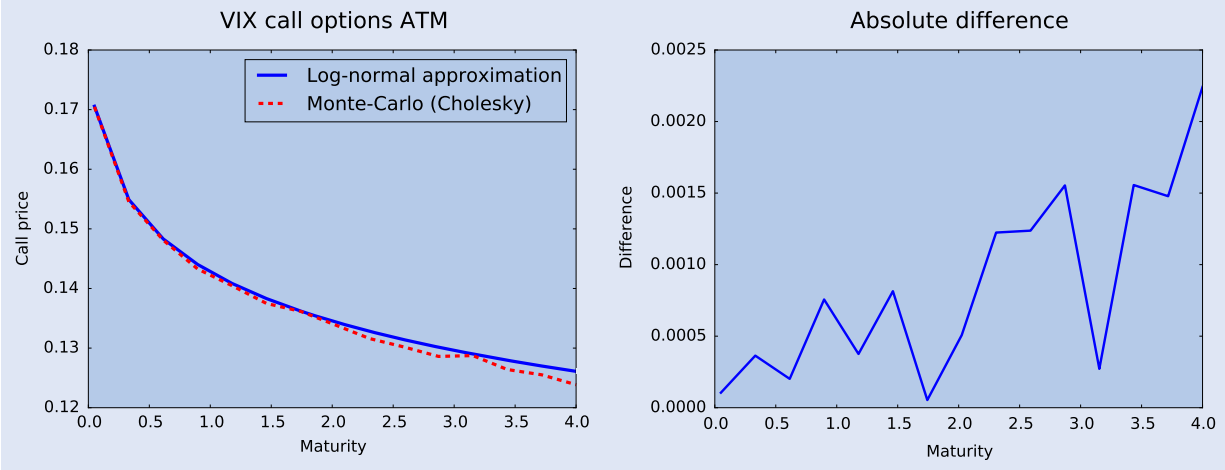

Figure 5. Log-normal approximation vs. Monte Carlo (Truncated Cholesky) with 10^6 simulations

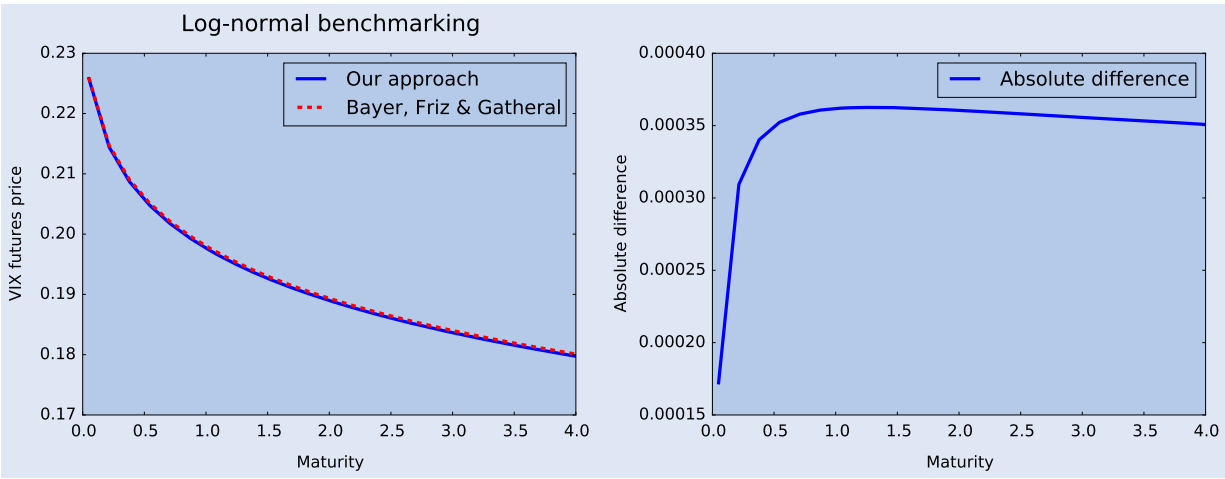

Figure 6. VIX futures prices following approximations 3.15 and 3.16.

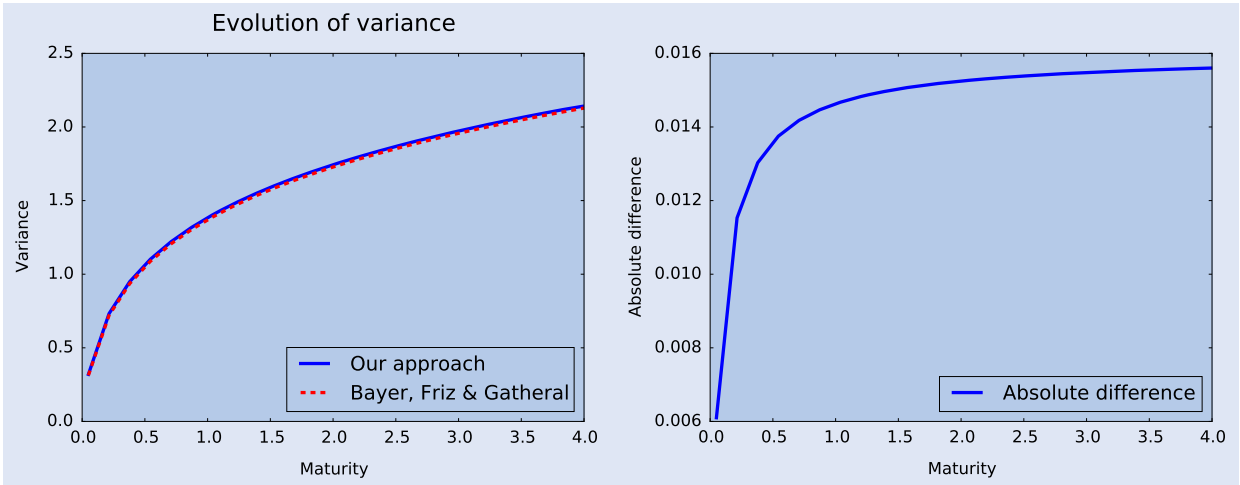

Figure 7. Comparison of the variance following approximation 3.15 (σ^2) and approximation 3.16 ($\tilde{\sigma}^2$).

$$
\begin{aligned}
&+\frac{8\nu^2C_H^2}{\Delta^2H_+^2}\int_0^{T_i}\left[(T_i-s+\Delta)^{H+\frac{1}{2}}-(T_i-s)^{H_+}\right]\\
&\times\left[(T_i-s+\Delta)^{H_+}\log(T_i-s+\Delta)-(T_i-s)^{H_+}\log(T_i-s)\right]\mathrm{d}s\\
&=\frac{\tilde{\sigma}_i^2(K_H-2)}{H_+}\\
&+\frac{8\nu^2C_H^2}{\Delta^2H_+^2}\int_0^{T_i}\left[(T_i-s+\Delta)^{H_+}-(T_i-s)^{H_+}\right]\\
&\times\left[(T_i-s+\Delta)^{H_+}\log(T_i-s+\Delta)-(T_i-s)^{H_+}\log(T_i-s)\right]\mathrm{d}s,
\end{aligned}
$$

where

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

and $K_H = \frac{H_+}{H}\{1 + H\psi(2-H_+) - H(\psi(H_+)+\psi(1-2H_-))\}$, 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 (Bergomi 2005) and the rough Bergomi models (Bayer *et al.* 2015), 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 (1998), 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 parameterization for the implied volatility surface. The first approach involves a discretization 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. We follow here the latter approach, and consider a parametric form for the total implied variance. Gatheral and Jacquier (2014) originally proposed such a parameterization (called SSVI), and studied absence of arbitrage and calibration to S&P options data. We use here the eSSVI parameterization (Hendriks and Martini 2017) for the implied volatility surface, a refinement of SSVI parameterization, relaxing the constant correlation assumption:

$$
\begin{aligned}
\sigma_{\mathrm{BS}}^2(t,k)t &= w(t,k)\\
&:=\frac{\theta_t}{2}\Bigg\{1+\rho(\theta_t)\varphi(\theta_t)k\\
&\quad+\sqrt{(\varphi(\theta_t)k+\rho(\theta_t))^2+1-\rho(\theta_t)^2}\Bigg\},
\end{aligned}
\tag{3.11}
$$

for all log-(forward) moneyness $k \in \mathbb{R}$, where $(\theta_t)_{t\geq 0}$ is the observed at-the-money variance curve, and where the shape function 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:

$$
\rho(\theta) = (A-C)\mathrm{e}^{-B\theta}+C, \qquad \text{for } (A,C)\in(-1,1)^2,\ B\geq 0, \tag{3.12}
$$

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 (2014) 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 (3.12), and care must be taken in order to preclude arbitrage. Concretely, following Hendriks and Martini (2017), the restriction

$$
|\theta\partial_\theta(\rho(\theta)) + \rho(\theta)\gamma| \leq \gamma, \tag{3.13}
$$

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 parameterization, it is sufficient (Gatheral and Jacquier 2014, 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 parameterization, variance swaps can be computed in closed form, as proved in Hendriks and Martini (2017), based on earlier works by Gatheral (2006) and Fukasawa (2011):

PROPOSITION 3.19 *The fair strike (in total variance) of a variance swap in the eSSVI model reads*

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

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

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

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 3.20 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 parameterization (3.11) on the SPX implied volatility surface on the 4th of December 2015. Figure 8 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 algorithm reads as follows:

Algorithm 3.21 (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 3.19;

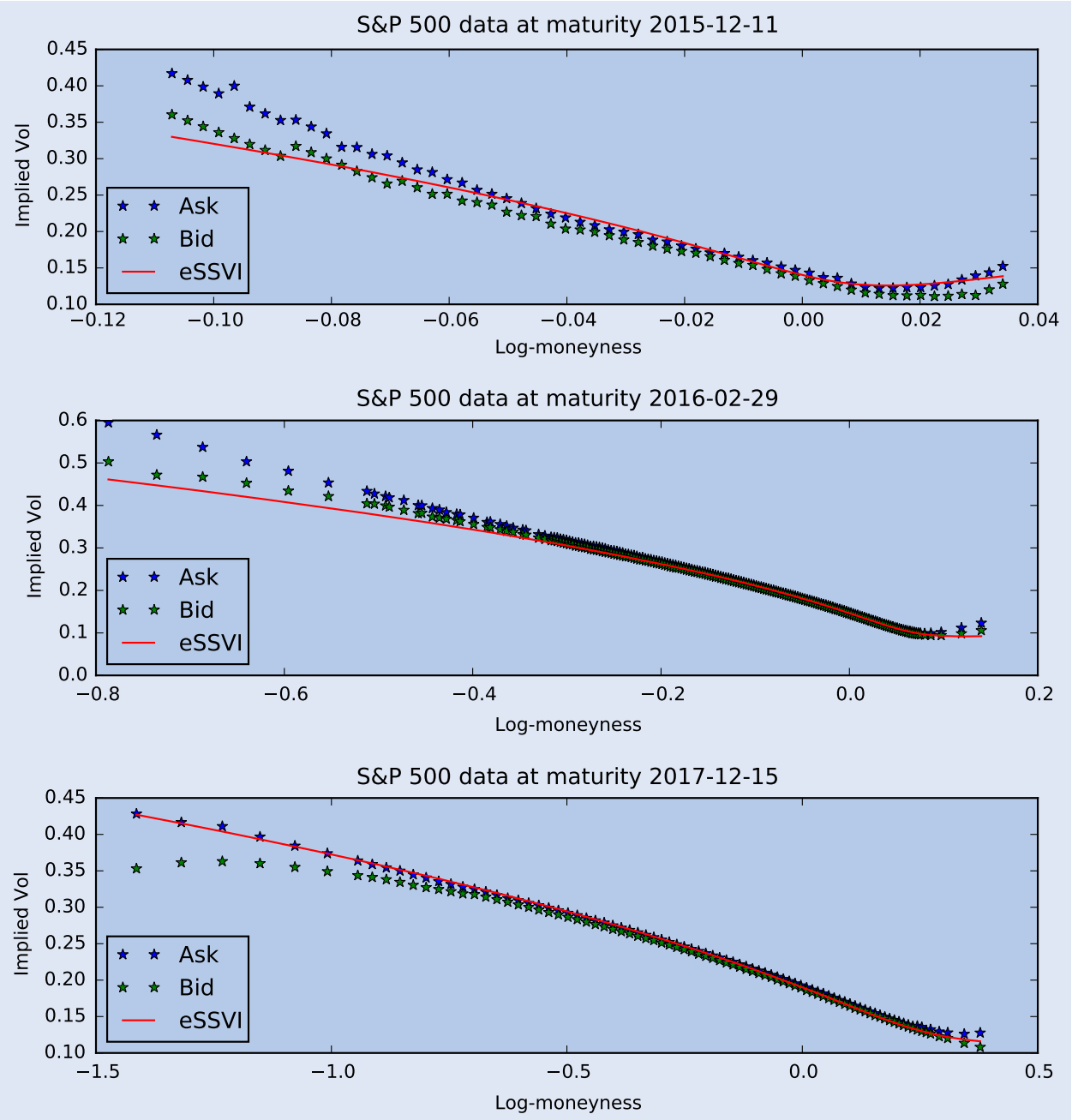

Figure 8. eSSVI calibration results on 4 December 2015 using traded SPX options.

(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) minimize (over ν, H) the objective function in (3.10).

Figures 9–11 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.

Remark 3.22 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.

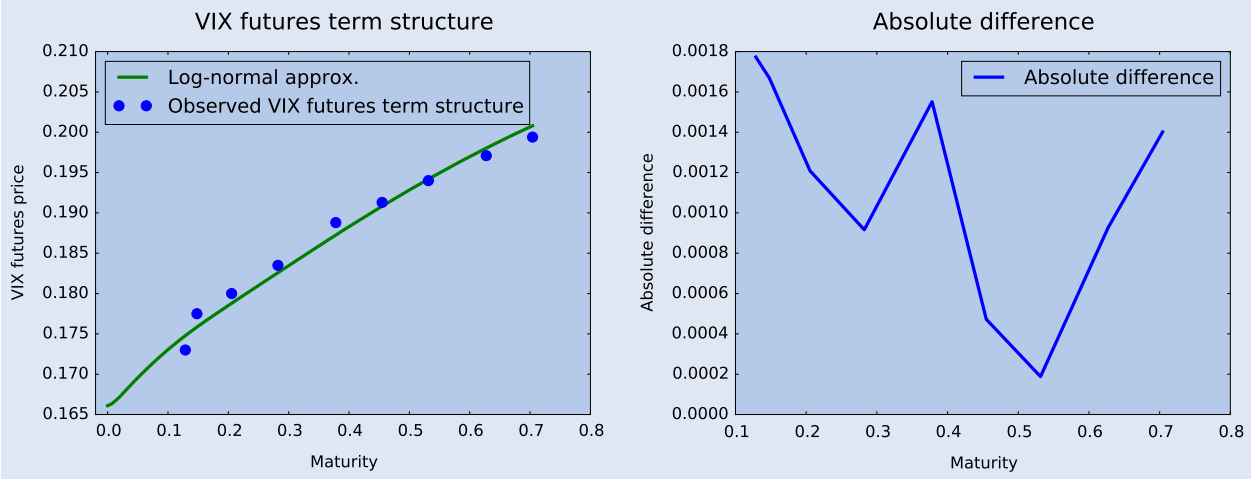

Figure 9. VIX futures calibration on 4 December 2015. Optimal parameters: $(H, \nu) = (0.09237, 1.004)$.

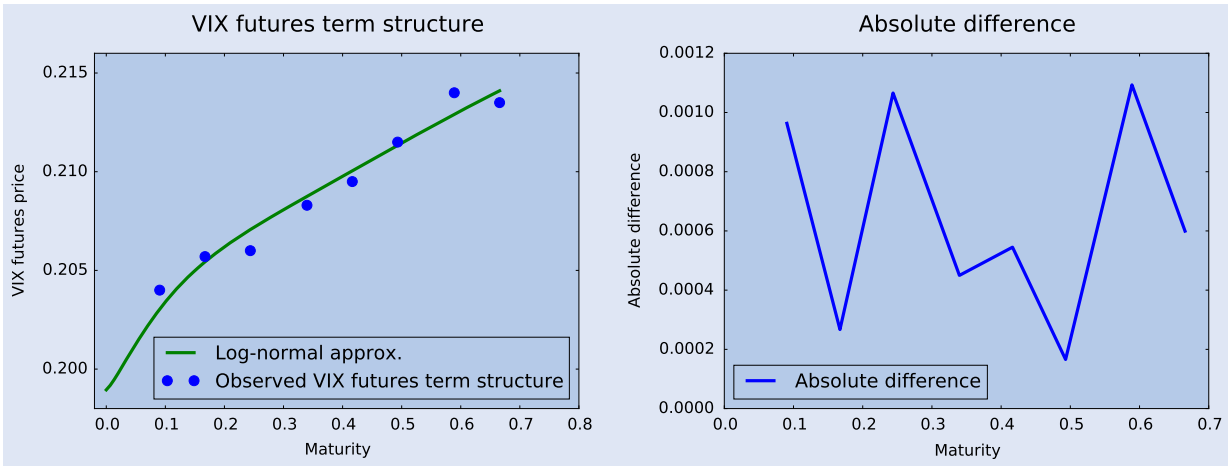

Figure 10. VIX futures calibration on 22 February 2016. Optimal parameters: $(H, \nu) = (0.10093, 1.00282)$.

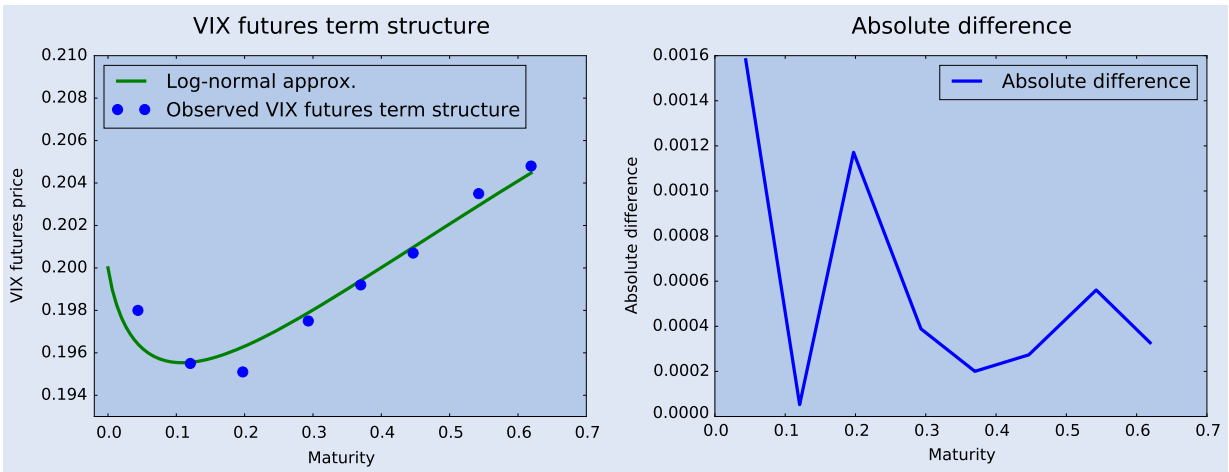

Figure 11. VIX futures calibration on 4 January 2016. Optimal parameters: $(H, \nu) = (0.0509, 1.2937)$.

Remark 3.23 Our calibration involves two different data-sets, SPX options and VIX futures: the Vanilla quotes are extracted from the CBOE delayed option quotes page† and the VIX futures from the CBOE VIX futures historical data page.‡ 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 $\mathcal{T} := \{t_i\}_{i=0,\dots,n_T}$, 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 $t \in \mathcal{T}$;
(iii) extract the path of the Brownian motion Z driving $\mathcal{V}$:

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

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 $W_{t_i} - W_{t_{i-1}} = \rho\overline{Z}_{i-1} + \sqrt{1-\rho^2}\overline{Z}^{\perp}_{i-1}$;
(v) simulate $S_{t_i} = \exp(X_{t_i})$ using a forward Euler scheme:

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

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

†CBOE delayed option quotes: http://www.cboe.com/delayedquote/quotetable.aspx.
‡CBOE VIX futures historical data: http://cfe.cboe.com/data/historicaldata.aspx.

4.2. Calibration of SPX options via VIX futures

We first follow the calibration algorithm 3.21 to obtain H and ξ_0, and we then aim at minimizing, over (ν, ρ), the objective function

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

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, $(\mathrm{C}^{\mathrm{obs}}_{i,j})_{i,j}$ is the set observed Call prices in the time grid $T_1 < \dots < T_N$ and strike grid $K^{(1)} < \dots < K^{(L)}$. In order to optimize 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:

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:

$$V_t^{(u)} = \xi_0(t)\mathcal{E}\left(2\nu C_H \mathcal{V}_t^{(u)}\right), \quad u = 1, \dots, M,$$
$$W^{(u)} = \rho Z^{(u)} + \sqrt{1-\rho^2}Z^{\perp(u)}, \quad u = 1, \dots, M,$$
$$S^{(u)}_{t+\Delta} = S^{(u)}_t + S^{(u)}_t\sqrt{v_t^{(u)}}\left(W^{(u)}_{t+\Delta} - W^{(u)}_t\right), \quad u = 1, \dots, M;$$

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

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

(v) minimize 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 12. 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 parameter ν obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility

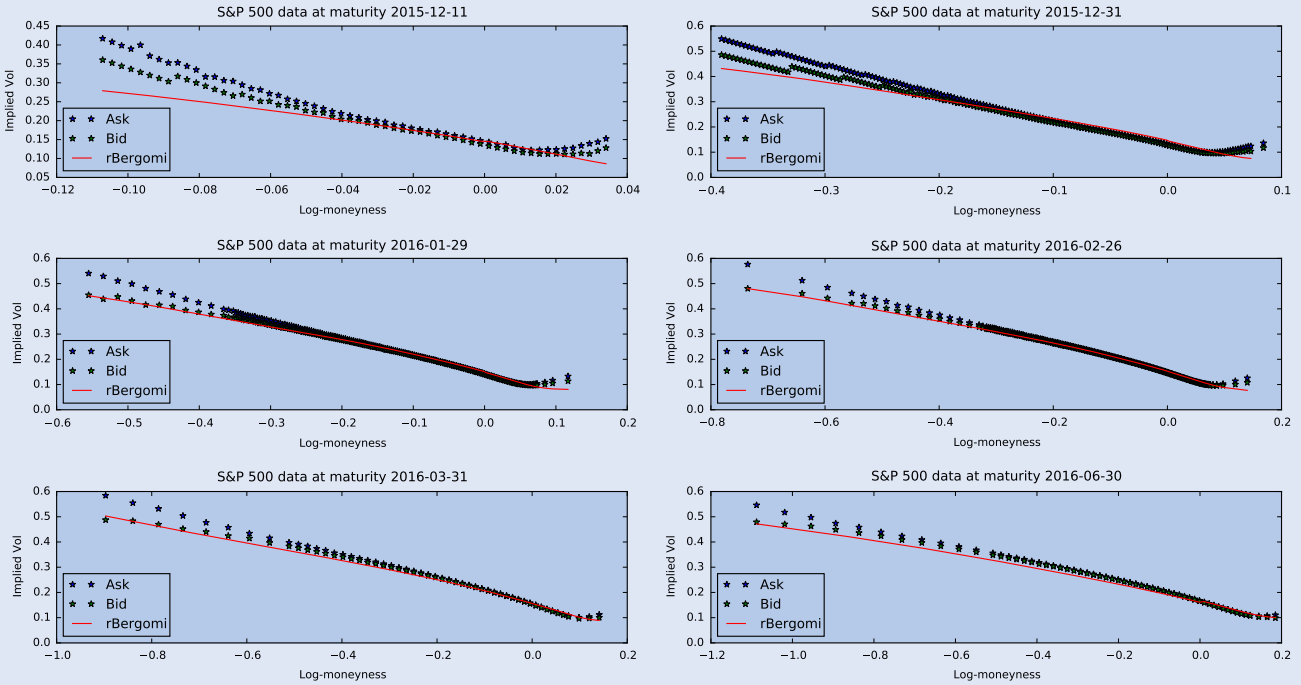

Figure 12. Calibration of SPX smiles on 4 December 2015. Calibrated parameters: $(\nu, \rho) = (1.19, -0.999)$.

of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?). Nevertheless, we emphasize the importance of an accurate ξ_0 curve to improve the fit to SPX and to provide an efficient joint calibration. Note in passing that the calibration drives the correlation parameter to the lower bound of its admissible values.

5. Conclusion

Following the path set by Bayer *et al.* (2015), we developed here a relatively fast algorithm to calibrate VIX futures, 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.

Acknowledgements

The authors would like to thank Christian Bayer, Jim Gatheral, Mikko Pakkanen and Mathieu Rosenbaum for useful discussions. AJ acknowledges financial support from the EPSRC First Grant EP/M008436/1, and AM is grateful to the Centre for Doctoral Training in Financial Computing & Analytics for financial support. The numerical implementations have been carried out on the collaborative platform Zanadu (www.zanadu.io).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by EPSRC [grant number EP/M008436/1].

References

Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions*, 1972 (Dover: New York).

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

Atkinson, L., *An Introduction to Numerical Analysis*, 2nd ed., 1989 (Wiley: Hoboken, NJ).

Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. *Quant. Finance*, 2015, **16**(6), 1–18.

Bennedsen, M., Lunde, A. and Pakkanen, M.S., Hybrid scheme for Brownian semistationary processes. *Finance Stoch.*, forthcoming.

Bergomi, L., Smile dynamics II. *Risk.*, 2005, 67–73.

Bingham, N.H., Goldie, C.M. and Teugels, J.L., *Regular Variation*, 1989 (Cambridge University Press: Cambridge).

Carr, P. and Madan, D., *Towards a Theory of Volatility Trading*, 1998 (Risk Publications). pp. 417-427.

Carr, P. and Madan, D., Joint modeling of VIX and SPX options at a single and common maturity with risk management applications. *IIE Trans.*, 2014, **46**(11), 1125–1131.

Comte, F. and Renault, E., Long memory continuous time models. *J. Econom.*, 1996, **73**(1), 101–149.

Doléans-Dade, C., Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrscheinlichkeit. Verw. Geb.*, 1970, **16**, 181–194.

Dufresne, D., The log-normal approximation in financial and other computations. *Adv. App. Prob*, 2004, **36**, 747–773.
Fukasawa, M., Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance Stoch.*, 2011, **15**(4), 635–654.
Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, 2006 (Wiley: Hoboken, NJ).
Gatheral, J., *Consistent Modelling of SPX and VIX Options*, 2008 (Presentation, Bachelier Congress: London).
Gatheral, J. and Jacquier, A., Arbitrage-free SVI volatility surfaces. *Quant. Finance*, 2014, **14**(1), 59–71.
Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough, 2014. arXiv: 1410.3394.
Golub, G.H. and Van Loan, C.F., *Matrix Computations*, 3rd ed., 1996 (Johns Hopkins University Press: Baltimore, MD).
Hagan, P.S., Kumar, D., Lesniewski, A. and Woodward, D.E., Managing smile risk. *Wilmott Mag.*, 2002, 84–108.
Hendriks, S. and Martini, C., The extended SSVI volatility surface, 2017. ssrn:2971502.
Heston, S., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**(2), 327–342.
Kokholm, T. and Stisen, M., Joint pricing of VIX and SPX options with stochastic volatility and jump models. *J. Risk Finance*, 2015, **16**(1), 27–48.

Appendix 1. The hybrid scheme

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

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

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

with L_g introduced in definition 2.1, and where

$$\overline{W}_i := \int_{t_i}^{t_{i+1}} \mathrm{d}W_s, \quad \overline{W}_{i,k} := \int_{t_i}^{t_{i+1}} (t_{i+k} - s)^\alpha \, \mathrm{d}W_s,$$

$$b_k^* = \left(\frac{k^{\alpha+1} - (k-1)^{\alpha+1}}{\alpha+1}\right)^{1/\alpha}, \text{ for } k \geq \kappa + 1. \tag{A1}$$

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

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

$$\mathbb{E}\left(\overline{W}_{i,k}\overline{W}_j\right) = 0, \quad \text{for } k \neq j,$$

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

$$\mathbb{V}\left(\overline{W}_{i,k}\right) = \frac{k^{2\alpha+1} - (k-1)^{2\alpha+1}}{n^{2\alpha+1}2\alpha + 1}, \quad \mathbb{V}\left(\overline{W}_i\right) = \frac{1}{n}.$$