Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

VIX pricing in the rBergomi model under a regime switching change of measure

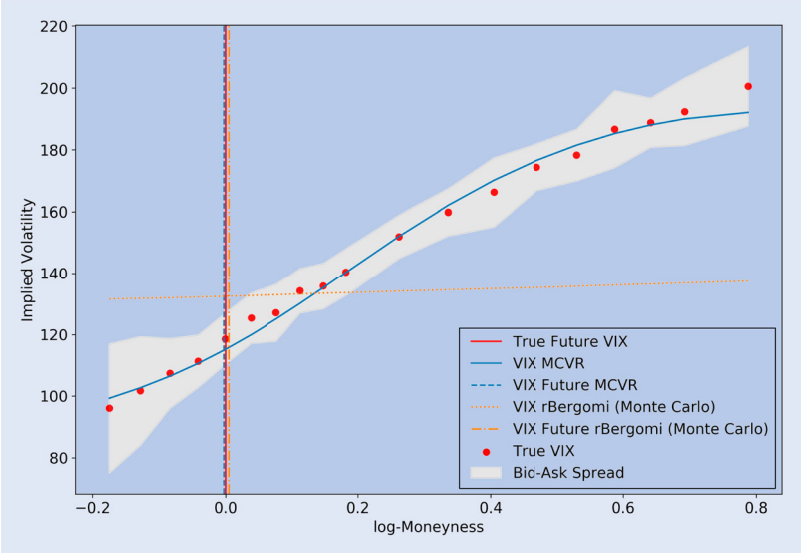
Henrique Guerreiro & João Guerra

To cite this article: Henrique Guerreiro & João Guerra (2023) VIX pricing in the rBergomi model under a regime switching change of measure, Quantitative Finance, 23:5, 721-738, DOI: 10.1080/14697688.2023.2178321

To link to this article: https://doi.org/10.1080/14697688.2023.2178321

Published online: 07 Mar 2023.

Quantitative Finance, 2023
Vol. 23, No. 5, 721–738, https://doi.org/10.1080/14697688.2023.2178321



VIX pricing in the rBergomi model under a regime switching change of measure

HENRIQUE GUERREIRO * and JOÃO GUERRA

ISEG—School of Economics and Management, REM—Research in Economics and Mathematics, Universidade de Lisboa, CEMAPRE, Rua do Quelhas 6, Lisboa 1200-781, Portugal

(*Received 10 March 2022; accepted 20 January 2023; published online 8 March 2023*)

A generalised rough Bergomi model with regime switching stochastic change of measure yields upward sloping VIX smiles

1. Introduction

Finding a mathematical model that reproduces the key features of observed market smiles has been a longstanding problem in mathematical finance. To this end, Gatheral *et al.* (2018) have introduced rough volatility models, where the log-variance behaves similarly to a fractional Brownian Motion (fBm). The rBergomi model, introduced in Bayer *et al.* (2016), is a rough volatility model that is able to adjust very well to the smiles of the SP500 with a small number of parameters. Moreover, it produces a power-law decaying *at the money skew*, a feature that is not shared by many conventional stochastic volatility models. The rBergomi model is first obtained by modeling the log-variance as a truncated Brownian semi-stationary process (TBSS), which is motivated by the empirical finding that increments of log-volatility behave similarly to those of fBm. For more details about TBSS processes, see Barndorff-Nielsen and Schmiegel (2009) and Bennedsen *et al.* (2017). Afterwards, a deterministic change of measure is applied, preserving analytic tractability. For evidence for rough volatility see Alòs *et al.* (2007), El Euch *et al.* (2018), Fukasawa (2020) and Livieri *et al.* (2018).

Unfortunately, by virtue of the deterministic change of measure, the rBergomi model produces flat smiles for the VIX index. This feature is very inconsistent with the market, where the VIX smile is upward sloping. In order to circumvent this problem, multiple solutions have been proposed. One possibility is to propose a stochastic volatility of the TBSS (which acts as a stochastic vol-of-vol). This approach was first proposed in Horvath *et al.* (2020), where analytic tractability is ensured by assuming the stochastic vol-of-vol is Markovian and independent of the volatility, together with some extra assumptions. This approach is further explored in Guerreiro and Guerra (2023), where a Least Squares Monte Carlo method is proposed to keep numerical tractability whilst dropping the independence assumption. For a discussion about capturing the VIX skew, see Alos *et al.* (2018).

One of the challenges of proposing more complex rough volatility models is to preserve analytic tractability. Due to non-Markovianity, classical techniques involving PDE's are not available, and costly Monte Carlo simulations are the only viable alternative. Recently, there has been an attempt to solve this problem by considering rough volatility models of the

*Corresponding author. Email: hguerreiro@iseg.ulisboa.pt

affine type, from which the most well known is the rough Heston model proposed in El Euch *et al.* (2019). Remarkably, the classical Riccati equation appearing in the characteristic function of the Heston model is replaced by a fractional Riccati equation in the rough Heston model. For details, see El Euch and Rosenbaum (2019). We may also name the fractional Ornstein–Uhlenbeck process, where the diffusion term is non-random. The original RFSV model of Gatheral *et al.* (2018) consisted of modeling the log-volatility as a process of the fractional Ornstein–Uhlenbeck type. It has also been considered in Wang *et al.* (2021).

The general theory of affine Volterra processes is extensively studied in Abi Jaber *et al.* (2019b), where it is shown that the conditional moment generating function (cMGF) can be written in terms of a Riccati-Volterra equation, thereby extending previous results concerning classical affine diffusions (see, for instance, Filipović 2005). The body of literature concerning affine processes, namely in finance, is vast. See, per example, Abi Jaber (2020), Abi Jaber (2021), Abi Jaber *et al.* (2019a), Comte *et al.* (2012), Wang (2008), Ackermann *et al.* (2020) and references therein.

In this paper, we propose a fractional Ornstein–Uhlenbeck stochastic change of measure for the rBergomi model that produces upward sloping VIX smiles whilst maintaining analytic tractability. The change of measure is obtained by explicitly solving the corresponding fractional affine Volterra equation. This approach has the advantage of giving us a description of market dynamics both under the physical measure $\mathbb{P}$ and the pricing measure $\mathbb{Q}$.

The paper is organized as follows. Section 2 provides the notation we will use in the rest of the paper. In Section 3, we introduce the (generalized) rBergomi model, both under the physical measure and under a general stochastic change of measure. Afterwards, in Section 4, we discuss the fractional Ornstein–Uhlenbeck process and derive the analytical formulae needed for efficient VIX pricing. Next, in Section 5, we propose the stochastic change of measure for the generalized rBergomi model and obtain a semi-closed formula for the forward variance curve. In Section 6, we apply the control variate trick inspired by Horvath *et al.* (2020) and obtain an approximation for the VIX method which significantly reduces the computing time. In Section 7, we further reduce computing times by applying a Monte Carlo variance reduction technique through importance sampling of the continuous time Markov chain. Then, in Section 9, we discuss model calibration and display the results. The comparison of the various numerical methods considered throughout the paper can be found in Section 10. Finally, in Section 11, we summarize our conclusions and mention further research problems.

2. Notation

The notation L^p always means $L^p(\mathbb{R}^d)$ for $d \in \mathbb{N}_1$, where $\mathbb{N}_1$ denotes the positive natural numbers. If we wish to talk about L^p with respect to a specific measurable set $A \subset \mathbb{R}^d$, we write $L^p(A)$. The same convention applies to L^p_{loc}, the space of functions that are in $L^p(K)$ for all compact sets K.

We work under a filtered probability space with filtration $(\mathcal{F}_t)_{t\geq 0}$. We shall denote the physical (or real world) measure by $\mathbb{P}$ and the pricing (or risk neutral) measure by $\mathbb{Q}$. Expected values and Brownian motions are defined with respect to the pricing measure, except when indicated with a $\mathbb{P}$ in superscript.

The forward variance curve is denoted by

$$\xi_t(u) = \mathbb{E}\left[v_u \mid \mathcal{F}_t\right], \tag{1}$$

where v is the variance process.

We denote the convolution operator by $*$, meaning

$$(f * g)(t) = \int f(s)g(t-s)\,\mathrm{d}s. \tag{2}$$

If f, g have support in $\mathbb{R}^+$ it becomes

$$(f * g)(t) = \int_0^t f(s)g(t-s)\,\mathrm{d}s. \tag{3}$$

For $K \in L^2([0,T])$ and a continuous semi-martingale $\mathrm{d}M = b\,\mathrm{d}s + \sigma dB$, with b, σ locally bounded and adapted, and B a standard Brownian motion (sBm), we may also define the convolution for $t \in [0,T]$ as

$$(K * \mathrm{d}M)(t) = \int_0^t K(t-s)b_s\,\mathrm{d}s + \int_0^t K(t-s)\sigma_s\,\mathrm{d}B_s. \tag{4}$$

We use the notation $\mathcal{E}$ for the stochastic exponential:

$$\mathcal{E}(X)_t = \exp\left(X_t - \frac{1}{2}\langle X\rangle_t\right), \tag{5}$$

where $\langle\cdot\rangle$ denotes the quadratic variation.

3. The rBergomi model

The rBergomi model was introduced in Bayer *et al.* (2016). Motivated by empirical data concerning the log-increments of volatility, the authors propose the following model under the physical measure $\mathbb{P}$:

$$v_u = A_0(u)\exp\left(2\sqrt{\gamma}\int_0^u K_\alpha(u-s)\,\mathrm{d}W_s^{\mathbb{P}}\right), \tag{6}$$

where K_α is the fractional kernel defined by

$$K_\alpha(u-s) = (u-s)^{\alpha-1}, \tag{7}$$

for $\alpha \in (1/2, 1)$, and A_0 is a deterministic function. We assume zero interest rates for simplicity. Thus, the dynamics of the price process S_t are

$$\mathrm{d}S_t = S_t(\zeta_t\,\mathrm{d}t + \sqrt{v_t}\,\mathrm{d}B_t^{\mathbb{P}}), \tag{8}$$

where $B^{\mathbb{P}}$ is a sBm correlated with $W^{\mathbb{P}}$ and defined as

$$B^{\mathbb{P}} = \rho W^{\mathbb{P}} + \bar{\rho}\bar{W}^{\mathbb{P}}, \tag{9}$$

for $\rho \in (-1,1)$, $\bar{\rho} = \sqrt{1-\rho^2}$ and $\bar{W}^{\mathbb{P}}$ a sBm independent of $W^{\mathbb{P}}$. In order to price options on a fixed time horizon $T > 0$,

we need to apply a change of measure to the pricing measure $\mathbb{Q}$. There are essentially two components to the change of measure, applied to the independent Brownian motions $W^{\mathbb{P}}$ and $\bar{W}^{\mathbb{P}}$. Thus, a general change of measure for W is characterized by a suitable adapted process λ such that

$$\mathrm{d}W_t^{\mathbb{P}} = \mathrm{d}W_t + \lambda_t \,\mathrm{d}t. \tag{10}$$

Moreover, the change of measure makes S a $\mathbb{Q}$-martingale so that

$$\mathrm{d}B_t = \mathrm{d}B_t^{\mathbb{P}} + \frac{\zeta_t}{\sqrt{v_t}}\,\mathrm{d}t. \tag{11}$$

Thus, we obtain the (extended) rBergomi model, where the variance is given by

$$v_u = A_0(u)\exp\left(2\sqrt{\gamma}\int_0^u K_\alpha(u-s)\,\mathrm{d}W_s + 2\sqrt{\gamma}\int_0^u \lambda_s K_\alpha(u-s)\,\mathrm{d}s\right), \tag{12}$$

and the price satisfies

$$\mathrm{d}S_t = S_t\sqrt{v_t}\,\mathrm{d}B_t. \tag{13}$$

When λ is deterministic, it gets absorbed into the deterministic function A_0 and gives rise to the original rBergomi model of Bayer *et al.* (2016). In this case, the variance is log-normal. This means that the VIX (see Chicago Board Options Exchange 2019), which is given by

$$VIX_t = \sqrt{\frac{1}{\Delta}\int_t^{t+\Delta} \xi_t(u)\,\mathrm{d}u}, \tag{14}$$

will be also approximately log-normal and lead to flat VIX smiles.

The goal of this paper is to propose a stochastic change of measure λ for the rBergomi model which produces upward sloping VIX smiles and at the same time provides an efficient semi-analytic Monte Carlo method to price VIX options.

To this end, we will first need to make an excursus to the theory of affine Volterra processes, and in particular the fractional Ornstein–Uhlenbeck process.

4. Fractional Ornstein–Uhlenbeck process

4.1. Homogeneous case

The d-dimensional fractional Ornstein–Uhlenbeck process is the solution to the fractional SDE

$$X_t = X_0 + \int_0^t K(t-s)\theta(\mu - X_s)\,\mathrm{d}s + \int_0^t K(t-s)\sigma\,\mathrm{d}Z_s, \tag{15}$$

where Z is a d-dimensional sBm, $\mu \in \mathbb{R}^d$ and $\sigma, \theta \in \mathbb{R}^{d\times d}$. The kernel K is required to satisfy some regularity and integrability conditions, which in our setting are specified in Assumption A1. The above process is a particular case of the more general class of affine Volterra processes, discussed in Abi Jaber *et al.* (2019b). Although not Markovian, this class of processes still possesses an exponentially-affine cMGF, as it can be seen in Abi Jaber *et al.* (2019b, Theorem 4.3). Moreover, we have an explicit expression for the solution.

4.2. Inhomogeneous fractional Ornstein–Uhlenbeck

The inhomogeneous fractional Ornstein–Uhlenbeck process is defined similarly to the above, but the parameters may depend on time:

$$X_t = X_0 + \int_0^t K(t-s)\theta(s)(\mu(s) - X_s)\,\mathrm{d}s + \int_0^t K(t-s)\sigma(s)\,\mathrm{d}Z_s. \tag{16}$$

The class of inhomogeneous affine processes was recently studied in Ackermann *et al.* (2020). By Ackermann *et al.* (2020, Theorem 2.1), we also have an exponentially-affine formula for the cMGF in inhomogeneous affine Volterra models with continuous coefficients, under some mild assumptions.

4.3. Time-dependent mean-reversion

Now we specialize in the case when only μ is time dependent. The idea is to later (in Section 5) work under a stochastic μ to remove the log-normality of the variance. Then, after conditioning, we will be able to use the analysis performed for deterministic time-dependent μ to keep the model analytically tractable.

A time-dependent μ is in general not continuous, and leads to a non-homogenous equation. Nevertheless, we may still use Abi Jaber *et al.* (2019b, Lemma 2.5) to build an explicit solution.

In order to ensure $K * \mathrm{d}Z$ has a continuous version, we assume each component of the kernel satisfies (Abi Jaber *et al.* 2019b, condition (2.5)).

Assumption A1 *Assume $K \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^{d\times d})$ and for each component H of K there is $\gamma \in (0,2]$ such that $\int_0^h H(t)^2\,\mathrm{d}t = O(h^\gamma), h \downarrow 0$, and $\int_0^T (H(t+h) - H(t))^2\,\mathrm{d}t = O(h^\gamma), h \downarrow 0$ for every $T < \infty$.*

We also introduce general assumptions on the function μ, that in particular allow for piecewise constant functions.

Assumption A2 *Assume $\mu \in L^q_{loc}(\mathbb{R}^+, \mathbb{R}^d)$, where q is such that $1/p + 1/q = 1$, and $K \in L^p_{loc}(\mathbb{R}^+, \mathbb{R}^{d\times d})$.*

Remark 4.1 If a kernel K satisfies Assumption A1, we know $K \in L^2_{loc}$, but it may be that $K \in L^p_{loc}$ for $p > 2$, which will yield a weaker condition on μ (since q will be required to be lower). In the case of the fractional kernel K_α, it is known it satisfies A1. Moreover, we know $K_\alpha \in L^p_{loc}$ for any $p < 1/(1-\alpha)$ and in particular for $p = 2$ (since $\alpha > 1/2$). Thus, at worst, choosing $q = 2$ always works.

Thus, we have the following theorem.

THEOREM 4.1 *Let $K : \mathbb{R}^+ \to \mathbb{R}^{d\times d}$, for some $d \in \mathbb{N}_1$. Suppose K satisfies Assumption* A1 *and μ satisfies Assumption* A2. *Let $T > 0$. Let $\theta, \sigma \in \mathbb{R}^{d\times d}$. Let $F : [0,T] \to \mathbb{R}^d$ be a continuous function and Z be a sBm. Denote also by R_θ the resolvent of second kind (see Definition* A.1*) of $K\theta$ and define $E_\theta = K - R_\theta * K$. Define also, for $0 < c < d$, the deterministic function*

$$H_{c,d}(u) = \int_c^d E_\theta(u-s)\theta\mu(s)\,\mathrm{d}s \tag{17}$$

and the Volterra process

$$Y^\sigma_{c,d}(u) = \int_c^d E_\theta(u-s)\sigma\,\mathrm{d}Z_s. \tag{18}$$

Then the affine Volterra equation

$$X_t = F(t) + \int_0^t K(t-s)\theta(\mu(s) - X_s)\,\mathrm{d}s + \int_0^t K(t-s)\sigma\,\mathrm{d}Z_s \tag{19}$$

has a unique continuous strong solution on $[0,T]$. Moreover, the solution is given explicitly by

$$X_u = g(u) + H_u + Y^\sigma_u, \tag{20}$$

where

$$g(u) = F(u) - \int_0^u R_\theta(u-s)F(s)\,\mathrm{d}s, \tag{21}$$

$H_u := H_{0,u}(u)$ is a deterministic function and $Y^\sigma_u := Y^\sigma_{0,u}(u)$ is a (fractional) Gaussian Volterra process.

Proof Let $b = \theta\mu$. Consider the function

$$\tilde{F}(t) = F(t) + (K * b)(t). \tag{22}$$

Since the deterministic function b satisfies $b \in L^q_{loc}$ and $K \in L^p_{loc}$, it follows by Lemma A.2 that $K * b$ is continuous. Since F is assumed to be continuous, it follows that $\tilde{F}$ is continuous. Note that (19) can be written as

$$X = \tilde{F} + (-K\theta) * X + K * (\sigma\,\mathrm{d}Z). \tag{23}$$

Provided X is continuous, by Abi Jaber *et al.* (2019b, Lemma 2.5), X solves the above if and only if

$$X = \tilde{F} - R_\theta * \tilde{F} + E_\theta * (\sigma\,\mathrm{d}Z). \tag{24}$$

Now notice that

$$\begin{aligned}\tilde{F} - R_\theta * \tilde{F} + E_\theta * (\sigma\,\mathrm{d}Z) &= F - R_\theta * F + (K * b)\\ &\quad - R_\theta * (K * b) + E_\theta * (\sigma\,\mathrm{d}Z)\\ &= g + (K - R_\theta * K) * b\\ &\quad + E_\theta * (\sigma\,\mathrm{d}Z)\\ &= g + E_\theta * b + E_\theta * (\sigma\,\mathrm{d}Z),\end{aligned}$$

where we used the associativity of the convolution operator for deterministic functions in the second equality and the definition of E_θ in the third equality. Thus, we only have to check that X admits a continuous version. Indeed, by the properties of the resolvent (see Remark A.1), since $K \in L^2_{loc}$, then also $R_\theta \in L^2_{loc}$. Since each component of K satisfies Assumption A1, it follows by Lemma A.3 that each component of $R_\theta * K$ also satisfies Assumption A1. Again, by Assumption A.3, we conclude that each component of E_θ also satisfies Assumption A1. We may now apply (Abi Jaber *et al.* 2019b, Lemma 2.4) and conclude that $E_\theta * (\sigma\,\mathrm{d}Z)$ admits a continuous version. Finally, since $\tilde{F}$ is continuous and $R_\theta \in L^p_{loc} \subset L^1_{loc}$, by Lemma A.2 we also conclude that $\tilde{F} * R_\theta$ is continuous. Thus, X admits a continuous version. ■

We now turn to one dimensional case $d = 1$. Using the above theorem it is easy to obtain the exponentially affine formula for the cMGF.

PROPOSITION 4.2 *Let $w \in \mathbb{R}$ and X, Y^σ, g be as in* (20)*, with $d = 1$. To lighten notation, write $Y = Y^{\sigma=1}$. Then*

$$\begin{aligned}&\mathbb{E}\left[\exp\left(wX_u\right) \mid \mathcal{F}_t\right]\\ &\quad = \exp\left[w(g(u) + H_u + \sigma Y_{0,t}(u)) + w^2 e_t(u)\right]\end{aligned} \tag{25}$$

where

$$e_t(u,\sigma) = \frac{1}{2}\sigma^2\int_t^u E^2_\theta(u-s)\,\mathrm{d}s. \tag{26}$$

Proof It is clear that

$$\mathbb{E}\left[\exp(wX_u) \mid \mathcal{F}_t\right] = \exp\left[w(g(u) + H_u)\right]\mathbb{E}\left[\exp(wY_u) \mid \mathcal{F}_t\right]. \tag{27}$$

Then

$$\begin{aligned}&\mathbb{E}\left[\exp(wY_u) \mid \mathcal{F}_t\right]\\ &\quad = \mathbb{E}\left[\exp(wY_{0,t}(u) + wY_{t,u}(u) \mid \mathcal{F}_t\right]\\ &\quad = \exp(wY_{0,t}(u))\mathbb{E}\left[\exp(wY_{t,u}(u)) \mid \mathcal{F}_t\right]\\ &\quad = \exp(wY_{0,t}(u))\mathbb{E}\left[\exp(wY_{t,u}(u))\right]\\ &\quad = \exp\left(wY_{0,t}(u) + \frac{1}{2}w^2\sigma^2\int_t^u E^2_\theta(u-s)\,\mathrm{d}s\right),\end{aligned}$$

where we used the fact that $Y_{0,t}(u)$ is $\mathcal{F}_t$-measurable and the fact that $\sigma Y_{t,u}(u)$ is Gaussian and independent of $\mathcal{F}_t$ with zero mean, and variance given by $2e_t(u,\sigma)$. ■

If K is the fractional kernel K_α, it is actually possible to express R_θ and E_θ in terms of the Mittag–Leffler function, as done in Lemma A.4. Moreover, if we assume μ is piecewise constant and F is constant, we get explicit expressions for H and g.

PROPOSITION 4.3 *Let $a < b$. Suppose μ is piecewise constant so that it can be written as*

$$\mu(s) = \sum_{k=0}^n \mu_k \mathbb{1}_{[t_k,t_{k+1})}(s), \tag{28}$$

where $a = t_0 < t_1 < \cdots < t_n < t_{n+1} = b$. Then

$$H_{a,b}(u) = \sum_{k=0}^n \mu_k I_k,$$

$$I_k := \int_{t_k}^{t_{k+1}} \theta E_\theta(u-s)\,\mathrm{d}s. \tag{29}$$

Moreover, if K is the fractional kernel K_α,

$$I_k = E_{\alpha,1}(-\theta\Gamma(\alpha)(u-t_{k+1})^\alpha) - E_{\alpha,1}(-\theta\Gamma(\alpha)(u-t_k)^\alpha). \tag{30}$$

Also, if $F \equiv x_0 \in \mathbb{R}$ we have

$$g(u) = x_0 E_{\alpha,1}\left(-\theta\Gamma(\alpha)u^\alpha\right). \tag{31}$$

Proof The equality (29) immediately follows from the fact that μ is piecewise constant together with the definition of H. Both remaining facts are an easy consequence of the elementary integral equalities involving R_θ and E_θ, stated in Lemmas A.4 and A.6. ■

5. Change of measure via regime switching fractional Ornstein–Uhlenbeck

We now turn to the case where μ follows a continuous time Markov chain (CTMC) with m states, independent of the driving Brownian motions. Since μ is independent of the Brownian motions, we can first simulate μ, and then for each μ simulate X given μ. We now make this rigorous.

For a fixed function f, let $X^{\{f(s)|0\le s\le T\}}$ be the unique continuous strong solution to (19) (with f taking the role of μ in (19)). Then X follows (20). Now, noting that μ is independent of Z, we set

$$X = X^{\{\mu_s|0\le s\le T\}}, \tag{32}$$

which amounts to replacing the deterministic function $\mu(s)$ with the stochastic process μ_s in (20).

PROPOSITION 5.1 *Let θ and σ be constant and μ be independent of Z, Markovian with respect to the filtration $(\mathcal{F}_t)_{t\ge0}$ and such that it satisfies Assumption* A2 *almost surely. Let $0 < t \le u \le T = t + \Delta$. Let F be a continuous function and g be as in* (21)*. Let X follow* (32) *as above.*

Then

$$\mathbb{E}\left[\exp(wX_u) \mid \mathcal{F}_t\right] = \exp\Big(w(g(u) + H_{0,u}(u) + \sigma Y_{0,t}(u)) + w^2 e_t(u,\sigma)\Big)\, G(w, u-t, \mu_t), \tag{33}$$

where G is the deterministic function

$$G(w,\tau,z) = \mathbb{E}\left[\exp\left(w\int_0^\tau \theta E_\theta(\tau-s)\mu_s\,\mathrm{d}s\right)\middle|\,\mu_0 = z\right]. \tag{34}$$

Proof By the tower property we have

$$\mathbb{E}\left[\exp(wX_u) \mid \mathcal{F}_t\right] = \mathbb{E}\left[\left[\exp(wX_u) \mid \mathcal{F}_t \vee \mathcal{G}_T\right] \mid \mathcal{F}_t\right], \tag{35}$$

where $\mathcal{G}_T$ is the sigma algebra generated by $(\mu_s)_{0\le s\le T}$. Since μ is independent of Z, we may apply (25):

$$\mathbb{E}\left[\exp(wX_u) \mid \mathcal{F}_t\right] = \mathbb{E}\Big[\exp\Big(w(g(u) + H_{0,u}(u) + \sigma Y_{0,t}(u)) + w^2 e_t(u,\sigma)\Big) \mid \mathcal{F}_t\Big]. \tag{36}$$

Note that $H_u = H_{0,u}(u)$ is no longer deterministic. By taking out the measurable terms and rearranging

$$\exp\Big(w(g(u) + H_{0,t}(u) + \sigma Y_{0,t}(u)) + w^2 e_t(u,\sigma)\Big) \times \mathbb{E}\left[\exp(wH_{t,u}(u)) \mid \mathcal{F}_t\right]. \tag{37}$$

Finally, the fact that μ is a CTMC, together with a change of variables implies that

$$\mathbb{E}\left[\exp(wH_{t,u}(u)) \mid \mathcal{F}_t\right] = \mathbb{E}\left[\exp(wH_{t,u}(u)) \mid \mu_t\right] = G(w, u-t, \mu_t). \tag{38}$$

■

We may now introduce a stochastic change of measure based on the fractional Ornstein–Uhlenbeck process.

DEFINITION 5.1 *Let $Z^{\mathbb{P}}, \bar{Z}^{\mathbb{P}}$ and $\bar{W}^{\mathbb{P}}$ be three independent Brownian motions and $\eta \in [0,1)$, with $\bar{\eta} = \sqrt{1-\eta^2}$. Define $W^{\mathbb{P}}$ as*

$$W_s^{\mathbb{P}} = \eta Z_s^{\mathbb{P}} + \bar{\eta}\bar{Z}_s^{\mathbb{P}},$$

which is a Brownian motion independent of $\bar{W}^{\mathbb{P}}$. Then let v, S and B be defined as in the rBergomi model of Section 3 so that the variance follows (12)*. Now let μ follow a CTMC, independent of all the Brownian motions. Let X be as in Proposition* 5.1*, where $\sigma = \eta$, $K = K_\alpha$ is the fractional kernel, and the driving sBm is $Z = Z^{\mathbb{P}}$. We define the fractional Ornstein–Uhlenbeck regime switching change of measure by*

$$\lambda_s(\omega) = \theta\,(\mu_s(\omega) - X_s(\omega)). \tag{39}$$

REMARK 5.1 One could think that letting $\eta \in (-1,1)$ would make the model more general. But note that if we replace Z by $-Z$ and η by $-\eta$ we obtain the exact same paths for v and S. Thus, we only need to consider the case $\eta \in [0,1)$.

REMARK 5.2 The change of measure λ is no longer deterministic, which means it is not obvious that W is a $\mathbb{Q}$ Brownian motion. To this end, we apply Girsanov's theorem. For details, we refer to Appendix 2.

REMARK 5.3 Note that, since v follows (12) and X solves (19), we may write

$$\log(v_u/A_0(u)) = 2\sqrt{\gamma}\,(X_u - F(u) + \bar{\eta}M_u), \tag{40}$$

where M is the Riemann–Liouville fBm

$$M_u = (K_\alpha * \mathrm{d}\bar{Z})(u) = \int_0^u K_\alpha(u-s)\,\mathrm{d}\bar{Z}_s \tag{41}$$

and $\bar{Z}$ is defined by

$$\bar{Z} = \frac{1}{\bar{\eta}}\,(W - \eta Z). \tag{42}$$

so that

$$\eta Z + \bar{\eta}\bar{Z} = W. \tag{43}$$

Now it is easy to derive a semi-closed expression for the forward variance curve, which can be used to obtain the VIX via a standard quadrature method.

PROPOSITION 5.2 *Suppose the assumptions of Proposition* 5.1 *are satisfied. Let X and μ be as in Definition* 5.1 *with $w = 2\sqrt{\gamma}$. Then we have the following formula for the forward variance curve*

$$\xi_t(u) = \xi_0(u)\frac{G(w, u-t, \mu_t)}{G(w, u, \mu_0)} \exp\left[w\Lambda(0,t,u) + w^2\lambda(0,t,u)\right], \tag{44}$$

where

$$\Lambda(0,t,u) = H_{0,t}(u) + \eta Y_{0,t}(u) + \bar{\eta} M_{0,t}(u), \tag{45}$$

$$\lambda(0,t,u) = e_t(u,\eta) - e_0(u,\eta) + m_t(u,\eta) - m_0(u,\eta), \tag{46}$$

$$M_{a,b}(u) = \int_a^b K_\alpha(u-s)\,\mathrm{d}\bar{Z}_s, \tag{47}$$

and

$$\begin{aligned} m_t(u,\eta) &= \log \mathbb{E}\left[\exp(\bar{\eta} M_{t,u}(u)) \mid \mathcal{F}_t\right] \\ &= \frac{1}{2}(1-\eta^2)\frac{(u-t)^{2H}}{2H}. \end{aligned} \tag{48}$$

Proof Note that, by definition

$$\xi_t(u) = A_0(u)\mathbb{E}\left[v_u/A_0(u) \mid \mathcal{F}_t\right]. \tag{49}$$

Then, by (40),

$$\xi_t(u) = A_0(u)\mathbb{E}\left[\exp\left(w(X_u - F(u) + \bar{\eta} M_u)\right) \mid \mathcal{F}_t\right]. \tag{50}$$

Now, using (20), observe we can write X as

$$X_u = g(u) + H_{0,t}(u) + \eta Y_{0,t}(u) + H_{t,u}(u) + \eta Y_{t,u}(u). \tag{51}$$

Since Z, μ and $\bar{Z}$ are all independent, it follows that $H_{t,u}(u), Y_{t,u}(u)$ and $M_{t,u}(u)$ are all conditionally independent given $\mathcal{F}_t$. Using also the fact that $M_{0,t}(u), Y_{0,t}(u)$ and $H_{0,t}(u)$ are all $\mathcal{F}_t$-measurable, it follows that X and M are conditionally independent given $\mathcal{F}_t$. Hence

$$\begin{aligned} \xi_t(u) = A_0(u)\,\mathrm{e}^{-wF(u)}&\mathbb{E}\left[\exp(wX_u) \mid \mathcal{F}_t\right] \\ &\times \mathbb{E}\left[\exp(w\bar{\eta} M_u) \mid \mathcal{F}_t\right]. \end{aligned} \tag{52}$$

Now apply Proposition 5.1 together with formula (48) and obtain

$$\begin{aligned} \xi_t(u) = A_0(u)G(w, u-t, \mu_t) \exp\big[w(g(u) - F(u) + \Lambda(0,t,u)) \\ + w^2(e_t(u,\eta) + m_t(u,\eta))\big]. \end{aligned} \tag{53}$$

Finally, if we apply (44) with $t = 0$ we have

$$\begin{aligned} \xi_0(u) = A_0(u)G(w, u, \mu_0) \exp\big[w(g(u) - F(u)) \\ + w^2\left(e_0(u,\eta) + m_0(u,\eta)\right)\big] \end{aligned} \tag{54}$$

and hence

$$\begin{aligned} A_0(u) = \frac{\xi_0(u)}{G(0, u, \mu_0)} \exp\big[-w(g(u) - F(u)) \\ - w^2\left(e_0(u,\eta) + m_0(u,\eta)\right)\big]. \end{aligned} \tag{55}$$

Plugging this equation into (44) yields the result. ■

REMARK 5.4 The forward variance appearing in (44) can either be extracted from the market (see Bayer *et al.* 2016), or we can make the simplifying assumption of a constant (or piecewise constant) forward variance curve. The constant(s) can be then considered as model parameter(s).

REMARK 5.5 We can express the forward variance curve in our model in terms of the classical rBergomi forward variance curve ξ^{RB}. In the rBergomi model, for a given initial curve ξ_0 and $v = 2\sqrt{\gamma}$, the forward variance curve will be given by

$$\begin{aligned} \log \xi_t^{RB}(u \mid v) := \log \xi_0(u) + vM_{0,t}(u) \\ + v^2(m_t(u,0) - m_0(u,0)). \end{aligned} \tag{56}$$

For our model, we can express the forward variance curve in terms of the rBergomi forward variance curve as follows

$$\log \xi_t(u) = \log \xi_t^{RB}(u \mid \bar{\eta} w) + \Xi, \tag{57}$$

where

$$\begin{aligned} \Xi = \log \frac{G(w, u-t, b_t)}{G(w, u, b_0)} + w(H_{0,t}(u) + \eta Y_{0,t}(u)) \\ + w^2(e_t(u) - e_0(u)). \end{aligned} \tag{58}$$

We can verify that, as must be the case, should we parameterize our model so that the change of measure is deterministic, i.e. $\eta = 0$ and μ deterministic, then we would recover the rBergomi model. Indeed, since $\eta = 0$, we have $e_t \equiv e_0 \equiv Y_{0,t} \equiv 0$. Moreover, since H is deterministic,

$$\begin{aligned} \log \frac{G(w, u-t, \mu_t)}{G(w, u, \mu_0)} &= \log \mathbb{E}\left[\exp(wH_{t,u}(u)) \mid \mathcal{F}_t\right] \\ &\quad - \log \mathbb{E}\left[\exp(wH_{0,u}(u))\right] \\ &= -wH_{0,t}(u). \end{aligned}$$

This implies $\Xi = 0$. Finally, since $\eta = 0$, we have $\bar{\eta} = 1$ and thus

$$\log \xi_t(u) = \log \xi_t^{RB}(u \mid w), \tag{59}$$

as expected.

REMARK 5.6 If μ were deterministic, the change of measure would amount to changing the kernel of the Volterra process, and the variance would still be log-normal, leading to an approximately flat VIX smile.

6. Control variate

As done in Horvath *et al.* (2020), we may use the fact that $\xi_t(u)$ is conditionally log-normal to approximate the VIX and use the approximation as a control variate. Namely, we approximate the integral over a family of conditionally log-normal random variables by the exponential of the integral of their logarithms.

By Proposition 5.2, we know that, conditional on $\mu(s)_{0\le s\le t}$, $\log \xi_t(u)$ is a Gaussian process. This allows us to conclude that

$$N_t := \frac{1}{\Delta}\int_t^{t+\Delta} \log \xi_t(u)\,\mathrm{d}u \tag{60}$$

is also Gaussian. Let us denote its mean by μ_N and its variance by σ_N^2. As a consequence of Proposition 5.2,

$$\log \xi_t(u) = \bar{m}(u) + w(\eta Y_{0,t}(u) + \bar{\eta} M_{0,t}(u)), \tag{61}$$

where

$$\begin{aligned}\bar{m}(u) &= \log \xi_0(u) + \log G(w, u-t, \mu_t)\\ &\quad - \log G(w,u,\mu_0) + w^2\lambda(0,t,u) + wH_{0,t}(u).\end{aligned} \tag{62}$$

Since Y and M do not depend on μ and have zero mean, the conditional mean of $\log \xi_t(u)$ is given by $\bar{m}(u)$. Hence, by Fubini's theorem, the conditional mean of N_t is

$$\mu_N = \frac{1}{\Delta}\int_t^{t+\Delta} \bar{m}(u)\,\mathrm{d}u. \tag{63}$$

Since Y, M are independent given μ, the conditional variance of N satisfies

$$\sigma_N^2 = \frac{w^2}{\Delta^2}\left[\eta^2\sigma_Y^2 + (1-\eta^2)\sigma_M^2\right], \tag{64}$$

where

$$\sigma_Y^2 = \mathrm{var}\left[\int_t^{t+\Delta}\left(\int_0^t E_\theta(u-s)\,\mathrm{d}Z_s\right)\mathrm{d}u\right] \tag{65}$$

and

$$\sigma_M^2 = \mathrm{var}\left[\int_t^{t+\Delta}\left(\int_0^t K_\alpha(u-s)\,\mathrm{d}\bar{Z}_s\right)\mathrm{d}u\right]. \tag{66}$$

By the stochastic Fubini theorem it follows that

$$\sigma_Y^2 = \int_0^t\left(\int_t^{t+\Delta} E_\theta(u-s)\,\mathrm{d}u\right)^2 \mathrm{d}s \tag{67}$$

and

$$\sigma_M^2 = \int_0^t\left(\int_t^{t+\Delta} K_\alpha(u-s)\,\mathrm{d}u\right)^2 \mathrm{d}s. \tag{68}$$

In particular, σ_N^2 does not depend on μ.

We may use Lemma A.6 to make the expression for σ_Y^2 more explicit:

$$\sigma_Y^2 = \frac{1}{\theta^2}\int_0^t\left[E_{\alpha,1}(-c(t-s)^\alpha) - E_{\alpha,1}(-c(t-s+\Delta)^\alpha)\right]^2\mathrm{d}s, \tag{69}$$

where $c = \theta\Gamma(\alpha)$. Likewise,

$$\begin{aligned}\sigma_M^2 &= \int_0^t\left(\int_t^{t+\Delta} K_\alpha(u-s)\,\mathrm{d}u\right)^2\mathrm{d}s\\ &= \frac{1}{\alpha^2}\left[\int_0^t (t+\Delta-s)^{2\alpha} + (t-s)^{2\alpha}\right.\\ &\quad \left. + 2(t+\Delta-s)^\alpha(t-s)^\alpha\,\mathrm{d}s\right]\\ &= \frac{1}{\alpha^2}\left[\frac{1}{2\alpha+1}\left((t+\Delta)^{2\alpha+1} - \Delta^{2\alpha+1} + t^{2\alpha+1}\right)\right.\\ &\quad \left. - 2\int_0^t (x^2 + x\Delta)^\alpha\,\mathrm{d}x\right].\end{aligned}$$

Note that, since $\alpha \in (1/2, 1)$, the computation of the above integrals does not involve singularities.

Using the above, we may obtain the cMGF:

$$\mathbb{E}\left[\exp(zN_t) \mid \mu\right] = \exp\left(z\mu_N + \frac{1}{2}z^2\sigma_N^2\right). \tag{70}$$

Thus, the logarithm of the cMGF is given by

$$\begin{aligned}&\log \mathbb{E}\left[\exp(zN_t) \mid \mu_t\right]\\ &= \frac{z}{\Delta}\left(\int_t^{t+\Delta} \log\xi_0(u)\log G(w,u,\mu_0) + w^2\lambda(0,t,u)\,\mathrm{d}u\right)\\ &\quad + \frac{1}{2}z^2\sigma_N^2 + \log\mathbb{E}\left[\exp\left(\frac{z}{\Delta}\int_t^{t+\Delta}\log G(w,u-t,\mu_t)\right.\right.\\ &\quad \left.\left. + wH_{0,t}(u)\,\mathrm{d}u\right)\right].\end{aligned}$$

If we apply the approximation

$$VIX_t^2 = \frac{1}{\Delta}\int_t^{t+\Delta}\xi_t(u)\,\mathrm{d}u \approx \exp\left(N_t\right), \tag{71}$$

we obtain an approximation of the VIX Future

$$\mathbb{E}\left[VIX_t\right] \approx \mathbb{E}\left[\exp\left(\frac{1}{2}N_t\right)\right]. \tag{72}$$

For a call option on the VIX, we have the approximation

$$\mathbb{E}\left[\left[(\exp(N_t/2) - K)^+ \mid \mu\right]\right] = \left[BS(\mu_N/2, \sigma_N^2/4)\right], \tag{73}$$

where BS is defined by the Black–Scholes formula:

$$BS(\mu,\sigma^2) = \mathcal{N}(d_+)F - \mathcal{N}(d_-)K, \tag{74}$$

where

$$d_\pm = \frac{1}{\sigma}\left(\log\frac{F}{K} \pm \frac{1}{2}\sigma^2\right) \tag{75}$$

and

$$F = \exp\left(\mu + \frac{1}{2}\sigma^2\right). \tag{76}$$

REMARK 6.1 In the rBergomi model, μ is deterministic and thus we obtain a price given (approximately) by the Black-Scholes formula, leading to a flat VIX smile. We can thus see the VIX option price in our model as (approximately) a weighted average of the rBergomi prices, where the weights are determined by the parameters of the CTMC μ.

7. Variance reduction via importance sampling

The distribution of dwelling times for the more extreme values of the CTMC can be deeply unbalanced: only a small percentage of generated paths will contain a significant dwelling time for the more extreme states. For out-of-the-money (OTM) options, both for the SP500 and the VIX, these paths have a significant contribution on the mean, since they will generate non-zero payouts for call options. Thus, it is natural to employ a Monte Carlo Variance Reduction technique (MCVR) via importance sampling.

In general, if we are interested in estimating $\mathbb{E}[f(X)]$ for a measurable function f and a random variable X distributed according to a density p, the simple Monte Carlo approach is to compute

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i), \quad X_i \sim p \; \forall i = 1, \dots, n. \tag{77}$$

Should we consider an alternative density h with the intent of reducing the variance of the estimator, we may sample $Y \sim h$ and weight the samples by the likelihood ratio p/h:

$$\frac{1}{n}\sum_{i=1}^{n} f(Y_i)\frac{p(Y_i)}{h(Y_i)}, \quad Y_i \sim h \; \forall i = 1, \dots, n, \tag{78}$$

producing an unbiased estimator for $\mathbb{E}[f(X)]$. If $p = h$ we of course recover (77). Thus, in order to perform the MCVR, we need access to the densities p and h.

Let us consider our case of the CTMC with m states starting at state s_0. Let $\{q_i\}_{i=1}^m$ be the jump intensities associated with each state and $p_{i,j}$ the probability of switching from state i to state j. Each path of the CTMC with k jumps is identified by the sequence of states it attains $s = (s_0, s_1, \dots, s_k)$ and dwelling times $t = (t_0, \dots, t_{k-1})$. By successive conditioning and using the Markov property it is easy to see that the density is given by

$$p(s,t) = \mathrm{e}^{-q_{s_k} t_k} \prod_{i=0}^{k-1} p_{s_i, s_{i+1}} q_{s_i} \mathrm{e}^{-q_{s_i} t_i} \mathbb{1}_{A_k}, \tag{79}$$

where

$$t_k = T - \sum_{i=0}^{k-1} t_i \tag{80}$$

and

$$A_k = \left\{ (t_0, t_1, \dots, t_{k-1}) \in \mathbb{R}^k : \sum_{i=0}^{k-1} t_i < T \right\}. \tag{81}$$

In order to reduce the variance of the estimator, we may simply use a uniform distribution for the dwelling times, which is given by the inverse of the volume of the region under a $(k-1)$-simplex:

$$\frac{1}{h(\mathbf{s}, \mathbf{t})} = \int_{A_k} 1 = \frac{T^k}{k!}. \tag{82}$$

Thus, for a function of interest $f = f(\mathbf{s}, \mathbf{t})$, our MCVR estimator is given by

$$\sum_{k=0}^{+\infty} \frac{T^k}{k!} \sum_{s \in J(k)} \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{s}, \mathbf{t}_i) p(\mathbf{s}, \mathbf{t}_i), \tag{83}$$

where $J(k)$ is the set of all possible state sequences starting at s_0 with k jumps, and $\mathbf{t} \sim \mathrm{Uniform}(A_k)$, i.e. it is uniformly distributed under the $(k-1)$-simplex. Such a sample can be obtained by sampling from the k-dimensional flat Dirichlet distribution (thus obtaining a uniform sample on the k-simplex) and removing the last coordinate.

Next, there are two adjustments that still have to be made to (83). First, we do not consider the number of jumps up to infinity, but rather truncate the series by an adequate maximum number of jumps K_M, which is chosen based on the model parameters. Secondly, using the same number of samples for every number of jumps is sub-optimal since it dramatically increases the computational cost. For this reason, we applied a stratified sample approach, where the sample size for each number of jumps n_k is in general proportional to its probability mass, but at the same time each number of jumps is required to have a minimum sample size. This allows us to substantially reduce the variance with only a small increase of computational cost. Thus, the MCVR estimator is given by

$$\sum_{k=0}^{K_M} \frac{T^k}{k!} \sum_{s \in J(k)} \frac{1}{n_k} \sum_{i=1}^{n_k} f(\mathbf{s}, \mathbf{t}_i) p(\mathbf{s}, \mathbf{t}_i). \tag{84}$$

Before we dwell into the application of the MCVR to pricing VIX and SP500 options, we note that it can be used to compute (34) for each τ, which is needed for both the VIX and the SP500 options valuation, with

$$f_\tau(\mathbf{s}, \mathbf{t}) = \exp\left(w \int_0^\tau \theta E_\theta(\tau - u) \mu(\mathbf{s}, \mathbf{t})(u) \, \mathrm{d}u \right). \tag{85}$$

7.1. VIX

Let us consider the approximations for the VIX future and option given by (72) and (73), respectively. Note that the randomness in these approximations depends only on the CTMC μ. The MCVR estimator for the VIX future is then given by (84) with

$$f(\mathbf{s}, \mathbf{t}) = \exp\left(\frac{1}{2}\mu_N(\mathbf{s}, \mathbf{t}) + \frac{1}{8}\sigma_N^2 \right), \tag{86}$$

with μ_N and σ_N^2 as in Section 6. Note that μ_N is a deterministic function of the path of the CTMC and σ_N^2 is deterministic.

In a similar way, the MCVR estimator for the VIX option is built using

$$f(\mathbf{s}, \mathbf{t}) = BS(\mu_N(\mathbf{s}, \mathbf{t})/2, \sigma_N^2/4). \tag{87}$$

7.2. SP500

To price SP500 options, the payout function will also depend on the Brownian paths. But even in this scenario, we may apply the MCVR. We start by considering the Brownian increments of the sBm's $Z, \bar{Z}$ and B, which we will denote by $\mathbf{W}$. Note that $\mathbf{W}$ is independent of $(\mathbf{s}, \mathbf{t})$. In the MCVR, we do not consider an alternative density for $\mathbf{W}$ but only for $(\mathbf{s}, \mathbf{t})$. By independence we obtain the following MCVR estimator of $\mathbb{E}[f(\mathbf{s}, \mathbf{t}, \mathbf{W})]$

$$\sum_{k=0}^{K_M} \frac{T^k}{k!} \sum_{s \in J(k)} \frac{1}{n_k} \sum_{i=1}^{n_k} f(\mathbf{s}, \mathbf{t}_i, \mathbf{W}_i) p(\mathbf{s}, \mathbf{t}_i). \tag{88}$$

In the case of the SP500 call option with maturity t and strike K, the function f is simply given by

$$f(\mathbf{s}, \mathbf{t}, \mathbf{W}) = (S_t(\mathbf{s}, \mathbf{t}, \mathbf{W}) - K)^+, \tag{89}$$

where we use the fact that the Brownian increments, together with the path of the CTMC, totally determine the (discretization of the) underlying price path.

The results in Section 10 show that the generated smiles are, on average, very similar, but the variance (and hence the computational cost) reduction is very significant for the VIX when applying MCVR.

8. Numerical implementation

All numerical simulations are performed on a Linux machine (Ubuntu 20.4), with an AMD Ryzen 7 3800X CPU with 16 threads. The Python language was used, resorting to *numpy* whenever possible for C-like speed. We also made use of the *numba* package, which allows for just in time compilation of code. This is especially useful for code that involves long/nested loops over nested lists of varying length, such as the ones that commonly appear in code related to CTMC's. These kinds of routines are not easily vectorized in *numpy* and using *numba* allows for much faster execution time without writing convoluted *numpy* code.

8.1. Pricing SP500 options

To price SP500 options, we use the MCVR method of Section 7.2. Once the variance process is computed, the price process (13) is easily obtained using a standard Riemann sum approximation.

In order to simulate the variance, we use (40). To simulate X, we use (20). The process M is a Riemann-Liouville fBm, so we may use the hybrid scheme of Bennedsen *et al.* (2017).

By Lemma A.4, $E_\theta(t) = t^{\alpha-1}\psi(t)$, where ψ is a continuous function expressed in terms of the Mittag-Leffler function. Thus, the process Y can also be computed with the hybrid scheme, setting $L_g = \psi$ in the notation of Bennedsen *et al.* (2017, Remark 3.1). For the computation of the Mittag–Leffler function, we used the library provided by Hinsen (2017). The function g is known exactly. The process H can be approximated using Proposition 4.3.

Using the *numba* package significantly improved the simplicity and performance of the simulation of H, due to the fact that the number of jumps is potentially different for each path.

Finally, the initial curve A_0 can be obtained from ξ_0 via (55).

8.2. Pricing VIX options

8.2.1. Simple Monte Carlo. Pricing VIX options via simple Monte Carlo amounts to simulating the forward variance curve, for which we resort to (44).

The terms $Y_{0,t}(u)$ and $M_{0,t}(u)$ are Volterra Gaussian processes with non-singular kernels, and thus can be approximated using the standard Riemann sum method. The term $H_{0,t}(u)$ can be approximated using Proposition 4.3. The function m_t poses no difficulty, since it has a closed expression. For the function e_t, which will involve an integral with a weakly singular integrand, we may use the procedure of Appendix 3.

Finally, we need to compute $G(w, u-t, b_t^{(i)})$, for each simulation i and $u \in [t, t+\Delta]$. Fortunately, μ_t can only attain a finite number of values. Thus, we may easily perform a simple Monte Carlo simulation for each of the m possible values of μ_t. That is, we may create M paths of μ conditioned on $\mu_0 = z_k$ for each possible state $z_1, z_2, \ldots, z_m$ and then approximate

$$G(w, u-t, z_k) \approx \frac{1}{M} \sum_{j=1}^{M} \exp\left(w \int_0^{u-t} \theta E_\theta(u-t-s) \mu^{(j)}(s)\, \mathrm{d}s \right).$$

Again, we use Proposition 4.3 to compute the integral.

Remark 8.1 The pricing of VIX options only requires the Brownian increments of $Z, \bar{Z}$ and B, but does not require simulating the processes H, Y and M. Recall these processes are computationally more expensive to simulate than $H_{0,t}(u)$, $Y_{0,t}(u)$ and $M_{0,t}(u)$ because of the presence of singularities in the kernels, which requires the hybrid scheme in the case of Y and M.

8.2.2. Control variate and variance reduction. We may use the approximations of Section 6 as a control variate to reduce the variance of the Monte Carlo estimator, or we may see it as an approximation and drop the simple Monte Carlo method altogether. In practice, it is much more efficient to use it as an approximation since the correlation between the control variate and the Monte Carlo estimator is very high. Finally, the variance reduction techniques described in Section 7.1 let us further reduce the variance of the estimator.

9. Calibration

We calibrate our model to the VIX and SP500 smiles separately and then perform a joint calibration. Due to much faster computing times, we used the MCVR estimators of Section 7 for both the SP500 and VIX Smiles. We found the maximum number of jumps $K_M = 4$ to be adequate for the

purposes of this paper. Moreover, to illustrate the contribution of our model, we performed a calibration for the VIX smile using a deterministic change of measure (i.e. the original rBergomi model). As expected, the produced smile is flat and not upward sloping like the one observed in the market.

We used the *LMFIT* library for the calibration routines (see Newville *et al.* 2014). We performed some exploratory tests and found the *Levenberg-Marquardt* method and the *Levenberg-Marquardt* algorithm formulated as a trust-region type algorithm to be good choices for the calibration routine.

The parameter space is simply a multidimensional rectangle. We fixed $\rho = -0.95$, $x_0 = 0$ and $s_0 = 0$ in all calibrations, in order to simplify the calibration procedure.

To make the interpretation easier, we used the parameterization with the Hurst index $H = \alpha - 1/2$. The forward variance curve was assumed constant in the case of the VIX and SP500 individual calibrations and piecewise constant in the case of joint calibration. This way, at each calibration step, we can adjust its value in $(T, T + \Delta)$ to the VIX Future.

The loss function measures the relative error between the implied volatilities produced by the model and those observed in the market.

We tested the joint calibration with both $m = 3$ and $m = 2$ states for the CTMC. Since $m = 3$ requires 6 more parameters than $m = 2$, the calibration displays a higher complexity and the model is less parsimonious. Moreover, the quality of the fits did not substantively differ. Perhaps a calibration method more tailored to this complex problem is needed. Devising such robust and efficient calibration procedure, one that substantially improves the quality of the fits for the joint calibration problem, is an endeavor that we leave for further research.

The bid and ask option prices extracted from Yahoo finance and date from 19 January 2021. The implied volatilities were computed from the average of the bid and ask prices. The time to maturity is 22 days.

9.1. Calibration to SP500 options

Although the original rBergomi model is a particular case of our fractional Ornstein–Uhlenbeck model, and for this reason our model inherits the ability to fit SP500 smiles, it is still interesting to study the calibration to the SP500 under the general fractional Ornstein–Uhlenbeck model we propose. First, because potentially multiple minima exist, and thus there can be parameterizations of the model that are not the rBergomi model. Secondly, because the rBergomi model corresponds to choices of parameters in the boundary of the parameter space (zero intensity for jumps, zero mean reversion value or setting all states of CTMC to the same value), making it hard for the calibration procedure to converge to the rBergomi model. Testing the ability of the calibration procedure and our proposed model to adjust to SP500 smiles adds confidence in both the reliability of the calibration procedure and the flexibility of the model.

In table 1, we present the calibration results and parameter space for the SP500 calibration. As expected, the model did not converge to the rBergomi model, which lies in the boundary of the parameter region space. Nevertheless, the Hurst parameter was lower than 0.1, which is consistent with the fact that the roughness of the volatility is important in reproducing the large skew of the SP500 smile. In figure 1, we observe the calibrated smile provides a good fit to the observed market smile, staying mostly inside the bid-ask spreads.

Table 1. SP500 calibration–calibrated values.

Parameter	Min	Max	Calibrated
H	0.05	0.3	0.0876
ρ	*fixed*	*fixed*	−0.95
η	0.0	0.99	0.0686
θ	0.1	10.0	2.4865
γ	0.01	0.5	0.3377
μ_1	0.0	7.5	0.1069
μ_2	0.0	30.0	9.06
q_1	0.0	5.0	0.1728
q_2	0.0	75.0	32.3703
$\xi_0(0)$	0.0001	0.25	0.0562
x_0	*fixed*	*fixed*	0.0

9.2. Calibration to VIX options and future

In figure 2, we can see that the fitted smile is always within the bid-ask spread and the future is calibrated almost exactly. In order to better show the contribution of our stochastic change of measure, we calibrated the VIX smile under a deterministic change of measure, which is essentially the original rBergomi model. As expected, the VIX smiles are flat.

We can see the calibrated values in table 2. We observe the Hurst parameter H is much higher than when calibrating for the SP500 smile. In our experiments, we observed that it is possible to produce good fits also with lower values of H. Nevertheless, we noticed increasing H slightly increases the quality of the fit. Essentially, the CTMC provides the model enough freedom to produce similar slopes and levels of the volatility smile for different values of the Hurst parameter. This can be explained by the fact that the CTMC is what produces the upward sloping feature in the VIX smile. Indeed, if μ were deterministic, then the smile would be flat. It is important to note that the increase in the quality of the fit is relatively small when considering larger values of H. Since larger values of H produce unrealistic SP500 smiles, we decided to keep the calibration interval for H in the $[0, 0.3]$ range.

9.3. Joint calibration to both SP500 and VIX options

The SP500-VIX Joint calibration has been called *the holy grail of volatility modeling* (see Gatheral *et al.* 2020). One of the main challenges in rBergomi-like models is that the large SP500 skew observed in the market will require for the vol-of-vol to be very large, which would produce VIX implied volatilities of a much higher magnitude than the ones we observe. The extra flexibility provided by our regime switching change of measure allows the model to reproduce the large SP500 skew whilst keeping VIX implied volatilities close to the ones observed in the market.

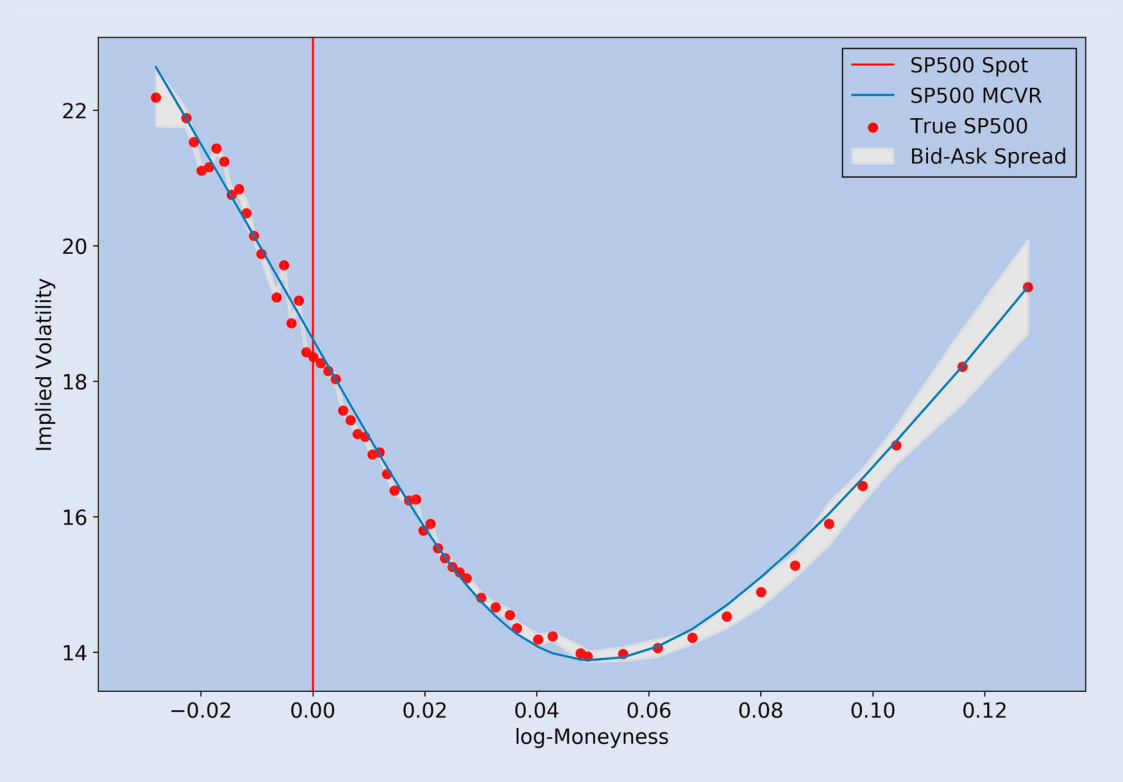


Figure 1. Calibrated SP500 smile.

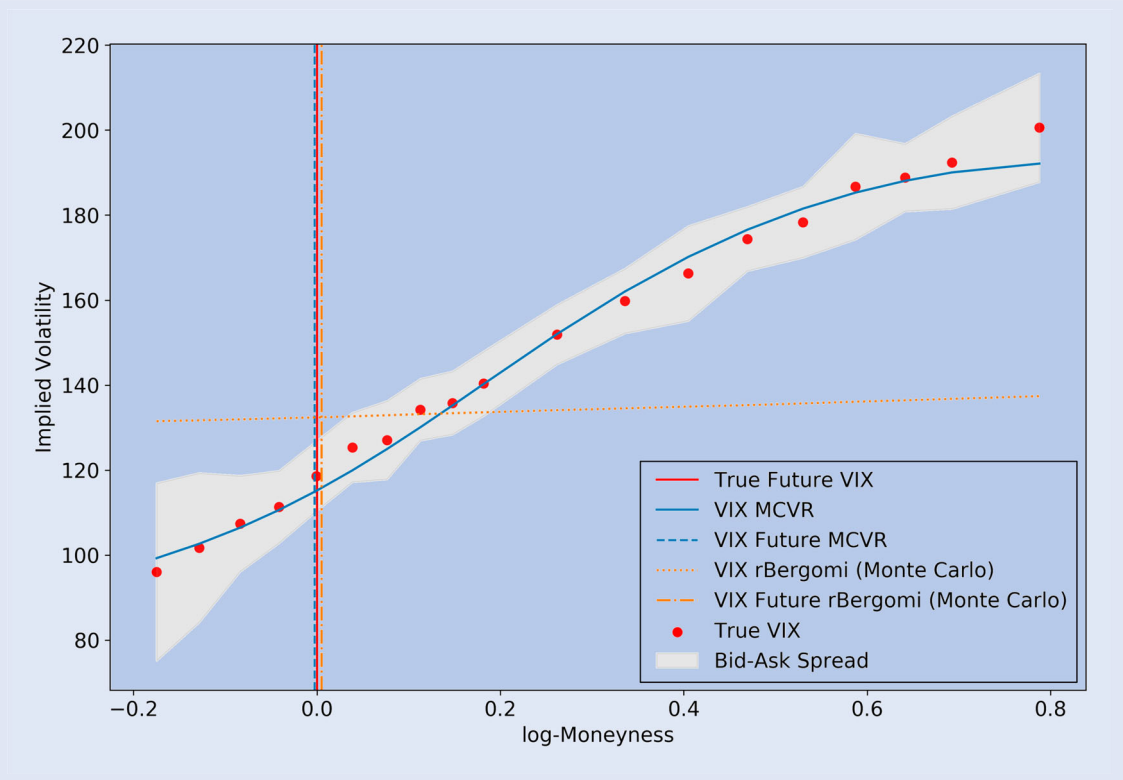


Figure 2. Calibrated VIX smile.

Table 2. VIX Calibration–calibrated values.

Parameter	Min	Max	Calibrated
H	0.05	0.3	0.3
ρ	*fixed*	*fixed*	-0.95
η	0.0	0.99	0.7753
θ	0.1	10.0	3.0297
γ	0.01	0.5	0.2739
μ_1	0.0	7.5	1.233
μ_2	0.0	30.0	8.3383
q_1	0.0	5.0	2.2913
q_2	0.0	75.0	25.295
x_0	*fixed*	*fixed*	0.0
$\xi_0(T)$	NA	NA	0.0714

Table 3. Joint calibration–calibrated values.

Parameter	Min	Max	Calibrated
H	0.05	0.3	0.1073
ρ	*fixed*	*fixed*	-0.95
η	0.0	0.99	0.9892
θ	0.1	10.0	2.85
γ	0.01	0.5	0.4989
μ_1	0.0	7.5	7.1608
μ_2	0.0	30.0	12.1098
q_1	0.0	5.0	2.0501
q_2	0.0	75.0	42.3636
$\xi_0(0)$	0.0001	0.25	0.0489
x_0	*fixed*	*fixed*	0.0
$\xi_0(T)$	NA	NA	0.0748

The joint calibration problem is a lot harder to calibrate also because a move in one direction of the parameter space might improve the SP500 fit but deteriorate the VIX fit and vice-versa. Indeed, we observed the calibration procedure displayed significantly more dependence on the initial condition when compared to the VIX or SP500 individual calibrations.

The calibrated values can be found in table 3. We can observe that the value for the Hurst parameter is much lower than the one that was obtained when calibrating only for the VIX. As explained in Section 9.1, a lower Hurst parameter allows the model to reproduce the large SP500 skew. Moreover, although exhibiting different values for each calibration, we consistently see the behavior of the CTMC is similar. The initial state μ_1 is of lower magnitude than μ_2 and stable (since q_1 is small). The other state μ_2 is a crisis-like state which has larger magnitude than μ_1 but is very unstable (since q_2 is large).

In figure 3, we see that the joint calibration is not as good as individual SP500 or VIX calibrations. Nevertheless, it is still a qualitatively good fit, especially considering bid-ask spreads. Indeed, it captures the overall shape and level of both SP500 and VIX smiles. This fit can potentially be improved in future works. First, by devising a calibration routine that is less reliant on the initial condition and is more robust. Secondly, such routine could perhaps allow for an increase in the number of states of the CTMC to produce better fits.

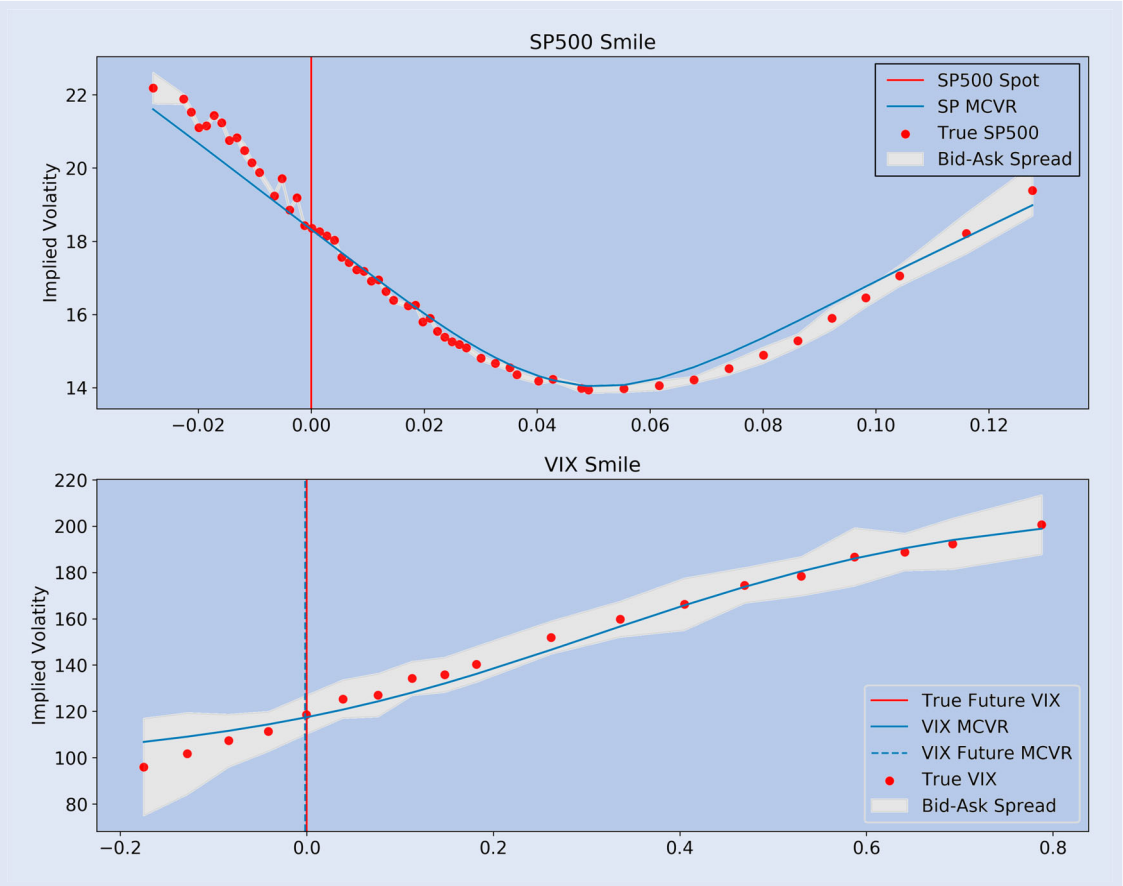


Figure 3. Calibrated joint SP500-VIX smile

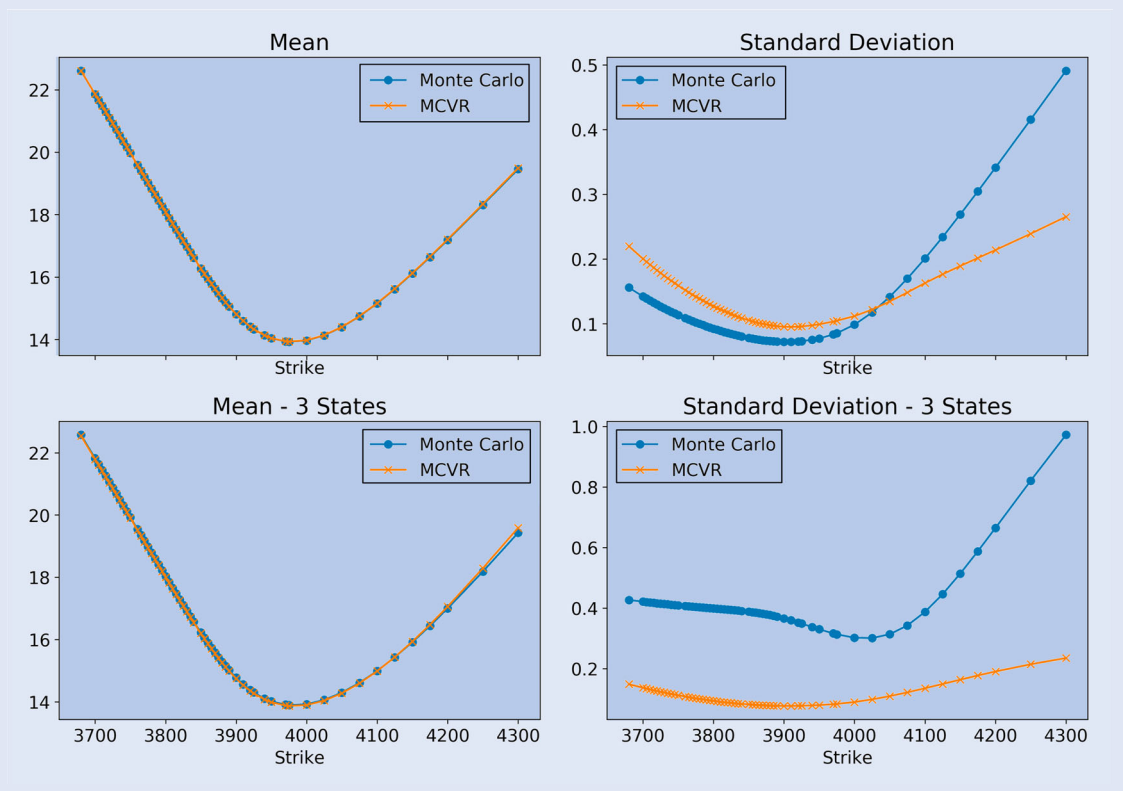


Figure 4. Statistics comparing the various numerical methods—SP500.

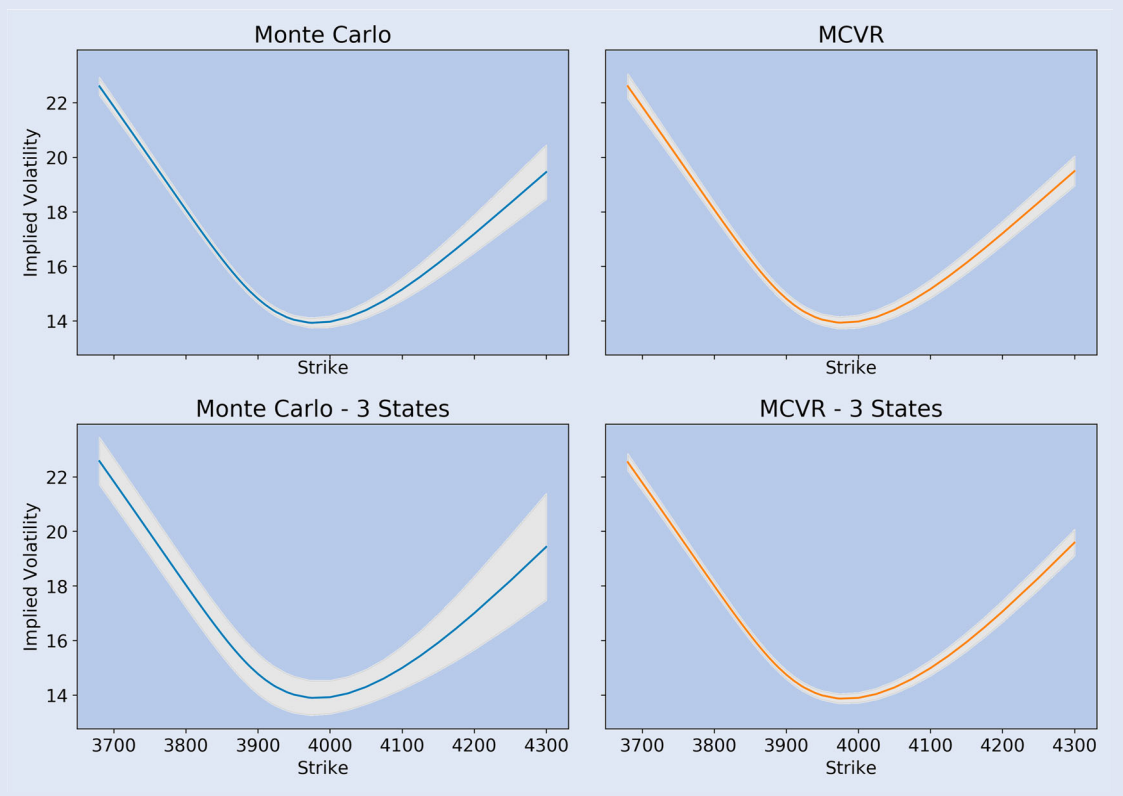


Figure 5. Error bars for various numerical methods (2 Standard Deviations)—SP500.

10. Performance of numerical methods

In order to evaluate the performance of the MCVR, we generate multiple smiles (1000 for the SP500 and 2000 for the VIX) with approximately the same computational budget, using the Monte Carlo Variance Reduction and Simple Monte Carlo methods. In the case of the VIX, we also present the comparison with the Control Variate approximation method. We

compute the mean smile across the entire sample, in order to validate that the methods are equivalent in the sense that they produce approximately the same results. We also compute the standard deviation, to compare the performance of the methods. We used the calibrated values found in Section 9.

10.1. SP500

We also note that the impact of the MCVR depends on the chosen model parameters. Indeed, if the CTMC is more complex, the impact of the variance reduction techniques can be larger. To exemplify this, we also generated smiles using $m = 3$. In figure 4 we can see that the smiles generated by MCVR and simple Monte Carlo are indeed very close. We also observe that the MCVR has a much lesser impact on performance than in the case of VIX smiles. This is due to the fact that the SP500 smile depends on the generated Brownian increments, which are not affected by the variance reduction method. Moreover, we see that it is more effective for OTM options than for ITM options, as expected. This is especially evident in the case of $m = 2$. In the case $m = 3$, with a more complex CTMC, the MCVR has a greater impact in variance reduction. Indeed, this is true even for ITM options. This behavior is also displayed in figure 5.

10.2. VIX

In the case of the VIX, we see that the Control Variate method substantially decreases the variance compared to the simple Monte Carlo, and the MCVR further decreases the variance of the Control Variate method (figures 6 and 7).

11. Conclusion and further research

In this paper, we proposed a generalization of the rBergomi model, where the change of measure is stochastic and written in terms of a regime-switching fractional Ornstein–Uhlenbeck process. From the semi-closed expression for the solution of the fractional SDE, we obtain a semi-closed form for the conditional moment generating function, which in turn lets us obtain a semi-closed expression for the VIX.

The fact that the forward variance curve is conditionally log-normal allows for an approximation of the VIX, whose

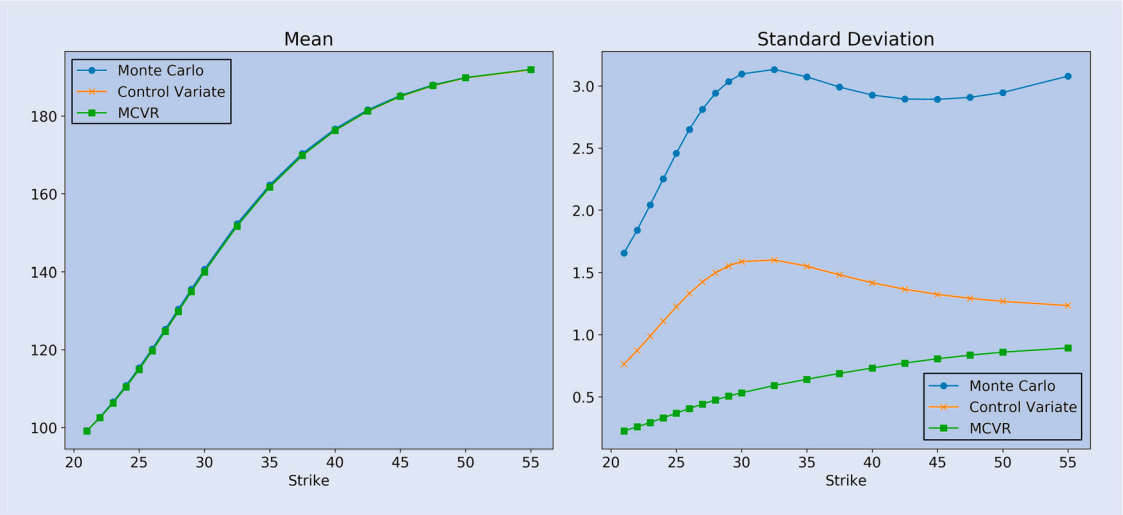


Figure 6. Statistics comparing the various numerical methods - VIX.

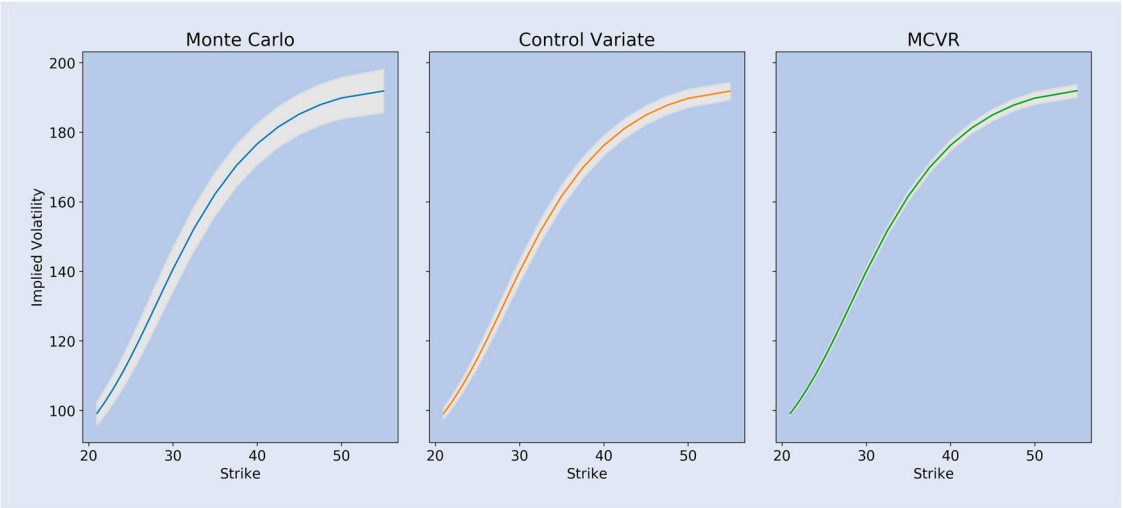


Figure 7. Error bars for various numerical methods (2 Standard Deviations)—VIX.

options and future prices are much more efficient to compute. Applying an importance sampling technique to the continuous time Markov chain, allows to further reduce the variance, which leads to fast pricing of VIX options and futures.

The explicit stochastic change of measure we propose allows us to reproduce the upward sloping VIX smiles observed in the market, which the original rBergomi could not do, since it produced flat smiles. The flexibility of the model is also displayed in the fact it can produce good fits to SP500 smiles without falling back into the original rBergomi model. For this reason, the model is even able to provide reasonable fits to the SP500-VIX joint calibration. Thus, the fractional Ornstein–Uhlenbeck model we propose has the ability to describe market dynamics both under the physical measure $\mathbb{P}$ and the pricing measure $\mathbb{Q}$.

We conclude with a few topics for further research. First, to devise a more robust calibration routine that is less reliant on the initial condition for the joint calibration problem. Secondly, this could be accompanied by a more complex CTMC that would be able to improve the fits. Finally, it remains an open question if a different choice for the dynamics of the long term mean of the fractional Ornstein–Uhlenbeck process would be able to provide better fits to observed smiles whilst keeping pricing and calibration computing times at a reasonable level.

Acknowledgments

We would like to express our gratitude to the Department of Mathematics of ISEG, Universidade de Lisboa and to CEMAPRE for the use of their computational resources in our numerical simulations. Finally, we also wish to thank the anonymous referees for helpful comments and suggestions.

Data availability

The data that supports the findings of this study was retrieved from Yahoo Finance in 19 January 2021.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding

Supported by FCT Grant SFRH/BD/147161/2019 and Partially supported by the project CEMAPRE/REM-UiDB/05069/2020—financed by FCT/MCTES through national funds.

ORCID

Henrique Guerreiro http://orcid.org/0000-0002-3519-7371
João Guerra http://orcid.org/0000-0002-5307-7170

References

Abi Jaber, E., The characteristic function of Gaussian stochastic volatility models: An analytic expression. arXiv preprint arXiv:2009.10972, 2020.

Abi Jaber, E., Weak existence and uniqueness for affine stochastic Volterra equations with L^1-kernels. *Bernoulli*, 2021, **27**(3), 1583–1615.

Abi Jaber, E., Cuchiero, C., Larsson, M. and Pulido, S., A weak solution theory for stochastic Volterra equations of convolution type. arXiv preprint arXiv:1909.01166, 2019a.

Abi Jaber, E., Larsson, M. and Pulido, S., Affine Volterra processes. *Ann. Appl. Probab.*, 2019b, **29**(5), 3155–3200.

Abi Jaber, E., Miller, E. and Pham, H., Markowitz portfolio selection for multivariate affine and quadratic Volterra models. *SIAM J. Financ. Math.*, 2021, **12**(1), 369–409.

Ackermann, J., Kruse, T. and Overbeck, L., Inhomogeneous affine Volterra processes. arXiv preprint arXiv:2012.10966, 2020.

Alos, E., García-Lorite, D. and Muguruza, A., On smile properties of volatility derivatives and exotic products: Understanding the VIX skew. arXiv preprint arXiv:1808.03610, 2018.

Alòs, E., León, J.A. 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.

Barndorff-Nielsen, O.E. and Schmiegel, J., Brownian semistationary processes and volatility/intermittency. In *Advanced Financial Modelling*, edited by H. Albrecher, W. J. Runggaldier, and W. Schachermayer, pp. 1–26, 2009 (De Gruyter).

Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. *Quant. Finance*, 2016, **16**(6), 887–904.

Bennedsen, M., Lunde, A. and Pakkanen, M.S., Hybrid scheme for Brownian semistationary processes. *Finance Stoch.*, 2017, **21**(4), 931–965.

Chicago Board Options Exchange, VIX: CBOE volatility index, 2019.

Comte, F., Coutin, L. and Renault, É., Affine fractional stochastic volatility models. *Ann. Finance*, 2012, **8**, 337–378.

El Euch, O. and Rosenbaum, M., The characteristic function of rough Heston models. *Math. Finance*, 2019, **29**(1), 3–38.

El Euch, O., Fukasawa, M. and Rosenbaum, M., The microstructural foundations of leverage effect and rough volatility. *Finance Stoch.*, 2018, **22**(2), 241–280.

El Euch, O., Gatheral, J. and Rosenbaum, M., Roughening Heston. *Risk Manage. Anal. Financ. Inst. eJ.*, 2019, **00**, 84–89. Available at SSRN: https://ssrn.com/abstract=3116887

Filipović, D., Time-inhomogeneous affine processes. *Stoch. Process. Their. Appl.*, 2005, **115**, 639–659.

Fukasawa, M., Volatility has to be rough. *Quant. Finance*, 2020, **21**, 1–8.

Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough. *Quant. Finance*, 2018, **18**(6), 933–949.

Gatheral, J., Jusselin, P. and Rosenbaum, M., The quadratic rough Heston model and the joint S&P500/VIX smile calibration problem. arXiv preprint arXiv:2001.01789, 2020.

Gripenberg, G., Londen, S.O. and Staffans, O., *Volterra Integral and Functional Equations*. Encyclopedia of Mathematics and its Applications, 1990 (Cambridge University Press).

Guerreiro, H. and Guerra, J., Least squares Monte Carlo methods in stochastic Volterra rough volatility models. *J. Comput. Finance*, 2023. doi:10.21314/JCF.2022.027.

Hinsen, K., The Mittag-Leffler function in Python. 2017. https://github.com/khinsen/mittag-leffler.

Horvath, B., Jacquier, A. and Tankov, P., Volatility options in rough volatility models. *SIAM J. Financ. Math.*, 2020, **11**, 437–469.

Livieri, G., Mouti, S., Pallavicini, A. and Rosenbaum, M., Rough volatility: Evidence from option prices. *IISE Trans.*, 2018, **50**, 767–776.

Newville, M., Stensitzki, T., Allen, D.B. and Ingargiola, A., LMFIT: Non-linear least-square minimization and curve-fitting for Python, 2014. doi:10.5281/zenodo.598352.

Rogosin, S., Gorenflo, R., Kilbas, A.A. and Mainardi, F., *Mittag-Leffler Function, Related Topics and Applications*, 2014.
Wang, Z., Existence and uniqueness of solutions to stochastic Volterra equations with singular kernels and non-Lipschitz coefficients. *Stat. Probab. Lett.*, 2008, **78**, 1062–1071.
Wang, X., Xiao, W. and Yu, J., Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process. *J. Econom.*, 2023, **232**(2), 389–415. doi:10.1016/j.jeconom.2021.08.001

Appendices

Appendix 1. Useful convolution results

In this section, we present some well-known convolution results.

The following two results are adapted from Gripenberg *et al.* (1990, Theorem 2.2.2) and Gripenberg *et al.* (1990, Corollary 2.2.3).

LEMMA A.1 *Let* $1 \le p \le \infty$, *and* $f \in L^p, g \in L^q$, *where* $1/p + 1/q = 1$. *Then* $f * g$ *is continuous.*

LEMMA A.2 *Let* $1 \le p \le \infty$, *and* $f \in L^p_{loc}, g \in L^q_{loc}$, *where* $1/p + 1/q = 1$ *and* f, g *have support on* $\mathbb{R}^+$. *Then* $f * g$ *is continuous and has support on* $\mathbb{R}^+$.

DEFINITION A.1 *For a kernel* $K \in L^1_{loc}(\mathbb{R}^+, \mathbb{R}^{d\times d})$, *the resolvent of* K *(also called resolvent of the second kind) is the unique kernel* $R \in L^1_{loc}(\mathbb{R}^+, \mathbb{R}^{d\times d})$ *that solves the resolvent equation*

$$K * R = R * K = K - R. \tag{A1}$$

For more details, see Abi Jaber et al. (2019b).

REMARK A.1 The resolvent inherits various properties of the original kernel K. Indeed, if $K \in L^p_{loc}$, with $1 \le p \le +\infty$, then also $R \in L^p_{loc}$. For more details, see Gripenberg *et al.* (1990, Theorem 3.5).

The following Lemma is adapted from Abi Jaber *et al.* (2019b, Example 2.3) to the multidimensional cases that interest us.

LEMMA A.3 *Let* $K : \mathbb{R}^+ \to \mathbb{R}^{d\times d}$. *Suppose* K *satisfies Assumption* A1. *Then:*

(i) *Let* $G : \mathbb{R}^+ \to \mathbb{R}^{d\times d}$. *If* G *satisfies Assumption* A1, *then* $K + G$ *also satisfies Assumption* A1.
(ii) *Let* $F \in L^2_{loc}(\mathbb{R}^+, \mathbb{R}^{d\times d})$. *Then* $K * F$ *satisfies Assumption* A1 *with the same* γ.

The fact we can state the above lemma in the multidimensional case is a consequence of the fact that each component of $K * F$ is the sum of convolutions of components of K and F.

LEMMA A.4 *Let* K *be the fractional kernel* $K_\alpha(x) = x^{\alpha-1}$, *where* $\alpha \in (1/2, 1)$. *Let* $\theta \in \mathbb{R} \setminus \{0\}$. *Let* R_θ *be the resolvent of* $K_\alpha\theta$, *and* $E_\theta = K_\alpha - R_\theta * K$. *Then*

$$R_\theta(t) = \theta t^{\alpha-1}\psi(t) \tag{A2}$$

and

$$E_\theta(t) = t^{\alpha-1}\psi(t), \tag{A3}$$

where ψ *is the continuous function defined on* $\mathbb{R}^+_0$ *by*

$$\psi(t) = \Gamma(\alpha)E_{\alpha,\alpha}(-\theta\Gamma(\alpha)t^\alpha) \tag{A4}$$

and E *denotes the Mittag–Leffler function*

$$E_{\alpha,\beta}(z) = \sum_{n=0}^{+\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}. \tag{A5}$$

LEMMA A.5 *Let* $\alpha \in (1/2, 1]$. *Then*

$$\int_0^u t^{\alpha-1}E_{\alpha,\alpha}(-t^\alpha)\,\mathrm{d}t = 1 - E_{\alpha,1}(-u^\alpha). \tag{A6}$$

Proof This an easy consequence of Rogosin *et al.* (2014, 4.11.5), which states that, for $\alpha > 0$,

$$t^{\alpha-1}E_{\alpha,\alpha}(-t^\alpha) = -\frac{\mathrm{d}}{\mathrm{d}t}E_{\alpha,1}(-t^\alpha). \tag{A7}$$

■

LEMMA A.6 *Let* R_θ *be as in Lemma* A.4, *with* $\theta > 0$. *Then*

$$\int_0^u R_\theta(t)\,\mathrm{d}t = 1 - E_{\alpha,1}\left(-\theta\Gamma(\alpha)u^\alpha\right). \tag{A8}$$

Proof This is a trivial consequence of Lemma A.5, together with a change of variables $x = [\theta\Gamma(\alpha)]^{1/\alpha}t$. ■

Appendix 2. Girsanov change of measure

The Brownian motions involved in our model are written in terms of the three-dimensional Brownian motion

$$\mathbf{B}^{\mathbb{P}} = (\bar{W}^{\mathbb{P}}, Z^{\mathbb{P}}, \bar{Z}^{\mathbb{P}}). \tag{A9}$$

For this Brownian motion, a general Girsanov change of measure is given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = M_T, \tag{A10}$$

where

$$M_t = \mathcal{E}\left(\int \alpha \cdot \mathrm{d}\mathbf{B}^{\mathbb{P}}\right)_t = \exp\left(\int_0^t \alpha_s \cdot \mathrm{d}\mathbf{B}^{\mathbb{P}}_s - \frac{1}{2}\int_0^t \|\alpha_s\|^2\,\mathrm{d}s\right), \tag{A11}$$

and

$$\alpha_s = (a_s, b_s, c_s), \tag{A12}$$

for adapted processes a, b, c.

Intuitively, we only need to apply the change of measure to two sources of randomness: the price and volatility drivers. Thus, the Brownian motion Z may be left unchanged and we may set $b_t \equiv 0$. The above change of measure implies that

$$a_t = \frac{1}{\bar{\rho}}(u_t + \rho\lambda_t), \tag{A13}$$

and

$$c_t = -\frac{1}{\bar{\eta}}\lambda_t, \tag{A14}$$

where ϑ is the market price of risk $\vartheta_t = \zeta_t/\sqrt{v_t} > 0$ and λ_t is the change of measure for the volatility component, as in (10). In the case of our fractional Ornstein–Uhlenbeck regime switching change of measure, the process λ_t is given by (39). To ensure $\mathbb{P} \sim \mathbb{Q}$, we still have to verify that

$$\mathbb{E}^{\mathbb{P}}[M_T] = 1. \tag{A15}$$

To this end, we adapt the proof from the ones found in Abi Jaber *et al.* (2019b, Lemma 7.3) and Abi Jaber *et al.* (2021, Appendix C).

Consider the stopping times

$$\tau_n = \inf\{t > 0 \mid \max_{i=1,2} U_t^{(i)} > n\} \wedge T, \tag{A16}$$

where

$$U_t^1 = \int_0^t \vartheta_s^2 \, \mathrm{d}s \tag{A17}$$

and

$$U_t^2 = \int_0^t X_s^2 \, \mathrm{d}s. \tag{A18}$$

Consider now the processes α_n defined by

$$\alpha_s^n = \mathbb{1}_{\{s \le \tau_n\}} \alpha_s \tag{A19}$$

and the corresponding sequence of measures $\mathbb{Q}^n$

$$\frac{\mathrm{d}\mathbb{Q}^n}{\mathrm{d}\mathbb{P}} = M_T^{\tau_n} = M_{\tau_n}, \tag{A20}$$

where the process M^{τ_n} is defined by

$$M_t^{\tau_n} = \mathcal{E}\left(\int \alpha^n \cdot \mathrm{d}\mathbf{B}^{\mathbb{P}}\right)_t. \tag{A21}$$

We have not specified the dynamics of the market price of risk ϑ, since it is outside the scope of this paper, but we will assume it can be controlled in the same way as X in the following sense:

Assumption A3 *There exists a constant C_2', which does not depend on n, such that for all $n \ge 1$,*

$$\sup_{0 \le t \le T} \mathbb{Q}^n\left[\vartheta_t^2\right] \le C_2'. \tag{A22}$$

For more details concerning this assumption and some sufficient conditions to satisfy it, see Remark A.3.

We can now verify Novikov's condition for α^n. Indeed

$$\begin{aligned}\|\alpha_t^n\|^2 &= a_t^2 + c_t^2 \\ &= \frac{1}{1-\rho^2}(\vartheta_t + \rho\lambda_t)^2 + \frac{1}{1-\eta^2}\lambda_t^2 \\ &\le \frac{2}{1-\rho^2}(\vartheta_t^2 + \rho^2\lambda_t^2) + \frac{1}{1-\eta^2}\lambda_t^2,\end{aligned}$$

where we used the useful inequality $(a+b)^2 \le 2(a^2+b^2)$ for $a, b \in \mathbb{R}$. Recall that $\lambda_t = \theta(\mu_t - X_t)$, where μ follows a CTMC and hence is bounded by a certain number $C_\mu \in \mathbb{R}$. Thus

$$\lambda_s^2 \le 2\theta^2(C_\mu^2 + X_t^2). \tag{A23}$$

Putting it all together, it follows that

$$\|\alpha_t^n\|^2 \le \mathbb{1}_{\{s \le \tau_n\}}\left(\beta_0 + \beta_1 \vartheta_t^2 + \beta_2 X_t^2\right), \tag{A24}$$

where the constants $\beta_0, \beta_1, \beta_2$ are given by

$$\begin{aligned}\beta_0 &= 2\theta^2 C_\mu^2\left(\frac{2\rho^2}{1-\rho^2} + \frac{1}{1-\eta^2}\right), \\ \beta_1 &= \frac{2}{1-\rho^2}, \\ \beta_2 &= \frac{1}{C_\mu^2}\beta_0.\end{aligned} \tag{A25}$$

Hence

$$\int_0^T \|\alpha_s^n\|^2 \, \mathrm{d}s \le \beta_0 T + \beta_1 U_{\tau_n}^1 + \beta_2 U_{\tau_n}^2. \tag{A26}$$

By construction, $U_{\tau_n}^i \le n$ a.s. for $i = 1, 2$. Thus, Novikov's condition is verified and therefore the process M^{τ_n} is a true martingale. Hence

$$\begin{aligned}1 = M_0^{\tau_n} &= \mathbb{E}\left[M_T^{\tau_n}\right] = \mathbb{E}\left[M_{\tau_n}\right] \\ &= \mathbb{E}\left[M_{\tau_n}\mathbb{1}_{\{\tau_n < T\}}\right] + \mathbb{E}\left[M_T \mathbb{1}_{\{\tau_n = T\}}\right].\end{aligned} \tag{A27}$$

Since M is a supermartingale, $M_T \in L^1$. Moreover, since X is continuous, $U_T < \infty$ a.s. and thus $\tau_n \to T$ a.s. when $n \to \infty$. By the dominated convergence theorem it follows that

$$\mathbb{E}\left[M_T \mathbb{1}_{\{\tau_n = T\}}\right] \to \mathbb{E}\left[M_T\right]. \tag{A28}$$

We are left to show that

$$\mathbb{E}\left[M_{\tau_n}\mathbb{1}_{\{\tau_n < T\}}\right] \to 0. \tag{A29}$$

By construction, we have that

$$\begin{aligned}\mathbb{E}\left[M_{\tau_n}\mathbb{1}_{\{\tau_n < T\}}\right] &= \mathbb{Q}^n\left[\tau_n < T\right] \\ &\le \sum_{i=1}^2 \mathbb{Q}^n\left[U_T^i > n\right] \\ &\le \frac{1}{n}\sum_{i=1}^2 \mathbb{Q}^n\left[|U_T^{(i)}|\right],\end{aligned}$$

where we used the Markov inequality in the last step.

For more general processes, we could proceed here as in Abi Jaber *et al.* (2019b, Lemma 7.3), but since we know our process X exactly we will simplify the argument. Indeed, by (20), we know that

$$|X_t| \le f(t) + \sigma|Y_t|. \tag{A30}$$

where f is the continuous function

$$f(t) = |g(t)| + C_\mu \int_0^t E_\theta(s) \, \mathrm{d}s. \tag{A31}$$

Since Y is Gaussian and its variance is a continuous function, described in (48), it follows that there exists a constant $C_2 \in \mathbb{R}$ such that

$$\sup_{0 \le t \le T} \mathbb{E}\left[|X_t|^2\right] \le C_2. \tag{A32}$$

By the Fubini-Tonelli theorem

$$\mathbb{E}\left[|U_T^{(2)}|\right] = \mathbb{E}\left[\int_0^T X_s^2 \, \mathrm{d}s\right] \le T \sup_{0 \le s \le T} \mathbb{E}\left[X_s^2\right] \le TC_2. \tag{A33}$$

The sBm Z is also a $\mathbb{Q}^n$-sBm since $b_t \equiv 0$ in the change of measure. Thus, the above constant does not depend on n. Finally, note that Assumption A3 guarantees that

$$\mathbb{E}\left[|U_T^{(1)}|\right] \le TC_2'. \tag{A34}$$

Remark A.2 The fact that the Brownian motion Z is not affected by the change of measure is by no means essential to the argument, as it can be seen in the proof of Abi Jaber *et al.* (2019b, Lemma 7.3) and also in the argument for ϑ in Remark A.3 below.

Remark A.3 The Assumption A3 may look *ad hoc*, but note that it will be satisfied if ϑ itself satisfies an equation similar to that of X. Indeed, assume

$$\vartheta = \vartheta_0 + K * (q \, \mathrm{d}t + \sigma \mathrm{d}\mathbf{B}^{\mathbb{P}}). \tag{A35}$$

The above is equivalent to

$$\vartheta = \vartheta_0 + K * (\tilde{q} \, \mathrm{d}t + \sigma \, \mathrm{d}\mathbf{B}^{\mathbb{Q}^n}), \tag{A36}$$

where

$$\tilde{q}(t, x, \omega) = q(t, x, \omega) + \sigma \mathbb{1}_{\{t \le \tau_n\}}(\omega). \tag{A37}$$

Provided q satisfies a uniform linear growth condition on x, so will $\tilde{q}$. Moreover, since α^n satisfies Novikov's condition, we are guaranteed that $\mathbf{B}^{\mathbb{Q}^n}$ is a $\mathbb{Q}_n$-sBm. The existence of the constant C_2' will then be guaranteed again by Abi Jaber *et al.* (2019b, Lemma 3.1, Remark 3.2).

Note also that the assumption will also be satisfied if ϑ can be written as a function of such process (which can even be X) and that function itself satisfies the linear growth condition.

Appendix 3. Weakly singular kernels

When dealing with fractional processes, it is common to encounter integrals with weakly singular kernels of the form

$$\int_0^u s^x (u-s)^y \varphi(s)\,\mathrm{d}s, \tag{A38}$$

where $-1 < x, y < 0$ and φ is a continuous function on $[0, u]$. To approximate such integrals, we use an idea similar to Bennedsen *et al.* (2017) and write them as

$$\int_0^u s^x (u-s)^y \varphi(s)\,\mathrm{d}s = I_1 + I_2 + I_3, \tag{A39}$$

where

$$\begin{aligned} I_1 &= \int_0^{\varepsilon} s^x (u-s)^y \varphi(s)\,\mathrm{d}s, \\ I_2 &= \int_{\varepsilon}^{u-\varepsilon} s^x (u-s)^y \varphi(s)\,\mathrm{d}s, \\ I_3 &= \int_{u-\varepsilon}^{u} s^x (u-s)^y \varphi(s)\,\mathrm{d}s, \end{aligned} \tag{A40}$$

for a small $\varepsilon > 0$, which is usually taken to be the grid time step. The integral I_2 can be approximated by a standard quadrature method since the integrand does not contain any singularities. For integrals I_1 and I_3, we approximate the non-singular part by a constant and integrate the singular part analytically:

$$\begin{aligned} I_1 &\approx \frac{\phi(0)u^y + \phi(\varepsilon)(u-\varepsilon)^y}{2} \frac{\varepsilon^{x+1}}{x+1}, \\ I_3 &\approx \frac{\phi(u-\varepsilon)(u-\varepsilon)^x + \phi(u)u^x}{2} \frac{\varepsilon^{y+1}}{y+1}. \end{aligned} \tag{A41}$$