FUNCTIONAL CENTRAL LIMIT THEOREMS FOR ROUGH VOLATILITY

BLANKA HORVATH, ANTOINE JACQUIER, AITOR MUGURUZA, AND ANDREAS SØJMARK

Abstract. The non-Markovian nature of rough volatility processes makes Monte Carlo methods challenging and it is in fact a major challenge to develop fast and accurate simulation algorithms. We provide an efficient one for stochastic Volterra processes, based on an extension of Donsker's approximation of Brownian motion to the fractional Brownian case with arbitrary Hurst exponent $H \in (0,1)$. Some of the most relevant consequences of this 'rough Donsker (rDonsker) Theorem' are functional weak convergence results in Skorokhod space for discrete approximations of a large class of rough stochastic volatility models. This justifies the validity of simple and easy-to-implement Monte-Carlo methods, for which we provide detailed numerical recipes. We test these against the current benchmark Hybrid scheme [14] and find remarkable agreement (for a large range of values of H). This rDonsker Theorem further provides a weak convergence proof for the Hybrid scheme itself, and allows to construct binomial trees for rough volatility models, the first available scheme (in the rough volatility context) for early exercise options such as American or Bermudan options.

Contents

Date: November 17, 2023.

2010 *Mathematics Subject Classification.* 60F17, 60F05, 60G15, 60G22, 91G20, 91G60, 91B25.

Key words and phrases. functional limit theorems, fractional Brownian motion, rough volatility, binomial trees.

The authors would like to thank Christian Bayer, Peter Friz, Masaaki Fukasawa, Paul Gassiat, Jim Gatheral, Mikko Pakkanen and Mathieu Rosenbaum for useful discussions. BH gratefully acknowledges financial support from the SNSF Early Postdoc.Mobility grant 165248, and AM is grateful to the Centre for Doctoral Training in Financial Computing & Analytics for financial support. Part of this work was carried out while AJ was Visiting Professor in Baruch College, CUNY, and AJ further acknowledges financial support from the EPSRC/T032146 grant. The numerical implementations have been carried out on the collaborative Zanadu platform, and the code is fully available at GitHub:RoughFCLT .

Introduction

Fractional Brownian motion has a long and famous history in probability, stochastic analysis and their applications to diverse fields [50, 51, 57, 65]. Recently, it has experienced a new renaissance in the form of fractional volatility models in mathematical finance. These were first introduced by Comte and Renault [22], and later studied theoretically by Djehiche and Eddahbi [26], Alòs, León and Vives [3] and Fukasawa [37], and given financial motivation and data consistency by Gatheral, Jaisson and Rosenbaum [41] and Bayer, Friz and Gatheral [11]. Since then, a vast literature has pushed the analysis in many directions [10, 12, 15, 31, 34, 38, 43, 44, 48, 54, 70], leading to theoretical and practical challenges to understand and implement these models. One of the main issues, at least from a practical point of view, is on the numerical side: absence of Markovianity rules out PDE-based schemes, and simulation is the only possibility. However, classical simulation methods for fractional Brownian motion (based on Cholesky decomposition or circulant matrices) are notoriously slow, and faster techniques are needed. The state of the art, so far, is the recent hybrid scheme developed by Bennedsen, Pakkanen and Lunde [14], and its turbocharged version [67]. We rise here to this challenge, and propose an alternative tree-based approach, mathematically rooted in an extension of Donsker's theorem to rough volatility.

Donsker [28] (and later Lamperti [61]) proved a functional central limit for Brownian motion, thereby providing a theoretical justification of its random walk approximation. Many extensions have been studied in the literature, and we refer the interested reader to [30] for an overview. In the fractional case, Sottinen [81] and Nieminen [72] constructed–following Donsker's ideas of using iid sequences of random variables– an approximating sequence converging to the fractional Brownian motion, with Hurst parameter $H > 1/2$. In order to deal with the non-Markovian behaviour of fractional Brownian motion, Taqqu [83] considered sequences of non-iid random variables, again with the restriction $H > 1/2$. Unfortunately, neither methodologies seem to carry over to the 'rough' case $H < 1/2$, mainly because of the topologies involved. The recent development of rough paths theory [36, 64] provided an appropriate framework to extend Donsker's results to processes with sample paths of Hölder regularity strictly smaller than 1/2. For $H \in (1/3, 1/2)$, Bardina, Nourdin, Rovira and Tindel [6] used rough paths to show that functional central limit theorems (in the spirit of Donsker) apply. This in particular suggests that the natural topology at work for rough fractional Brownian motion is the topology induced by the Hölder norm of the sample paths. Indeed, switching the topology from the Skorokhod one used by Donsker to the (stronger) Hölder topology is the right setting for rough central limit theorems, as we outline in this paper. Recent results [13, 73, 74] provide convergence for (geometric) fractional Brownian motions with general $H \in (0, 1)$ using Wick calculus, assuming that the approximating sequences are Bernoulli random variables.

We extend this (Theorem 1.10) to a universal functional central limit theorem, involving general (discrete or continuous) random variables as approximating sequences, only requiring finiteness of moments.

We consider a general class of continuous processes with any Hölder regularity, including fractional Brownian motion with $H \in (0,1)$, truncated Brownian semi-stationary processes, Gaussian Volterra processes, as well as rough volatility models recently proposed in the financial literature. The fundamental novelty here is an approximating sequence capable of simultaneously keeping track of the approximated rough volatility process (fractional Brownian motion, Brownian semistationary process, or any continuous path functional thereof) and of the underlying Brownian motion. This is crucial in order to take into account the correlation of the two processes, the so-called leverage effect in financial modelling. While approximations of two-dimensional (correlated) semimartingales are well understood in the standard case, the rough case is so far an open problem. Our analysis easily generalises beyond Brownian drivers to more general semimartingales, emphasising that the subtle, yet essential, difficulties lie in the passage from the semimartingale setup to the rough case. This is the first Monte-Carlo method available in the literature, specifically tailored to two-dimensional rough systems, based on an approximating sequence for which we prove a Donsker-Lamperti-type functional central limit theorem (FCLT). This further provides a pathwise justification of the hybrid scheme by Bennedsen, Lunde and Pakkanen [14], and to develop tree-based schemes, opening the doors to pricing early-exercise options such as American options. In Section 1, we present the class of models we are considering and state our main results. The proof of the main theorem is developed in Section 2 in several steps. We reserve Section 3 to applications of the main result, namely weak convergence of the hybrid scheme, binomial trees as well as numerical examples. We present simple numerical recipes, providing a pedestrian alternative to the advanced hybrid schemes in [14, 67], and develop a simple Monte-Carlo with low implementation complexity, for which we provide comparison charts against [14] in terms of accuracy and against [67] in terms of speed. Reminders on Riemann-Liouville operators and additional technical proofs are postponed to the appendix.

Notations: On the interval $\mathbb{I} := [0,1]$, $\mathcal{C}(\mathbb{I})$ and $\mathcal{C}^\alpha(\mathbb{I})$ denote the spaces of continuous and α-Hölder continuous functions on $\mathbb{I}$ with Hölder regularity $\alpha \in (0,1)$; $\mathcal{C}^1(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R} : f' \text{ exists and is continuous on } \mathbb{I}\}$ and $\mathcal{C}^1_+(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R}_+ : f' \text{ exists and is continuous on } \mathbb{I}\}$. Both definitions imply bounded first order derivatives on $\mathbb{I}$. We use $C, \widetilde{C}, \widehat{C}, C_1, C_2, \overline{C}, \underline{C}$ as strictly positive real constants which may change from line to line, the exact values of which do not matter.

1. Weak convergence of rough volatility models

Donsker's invariance principle [28] (also termed 'functional central limit theorem') ensures the weak convergence of an approximating sequence to a Brownian motion in the Skorokhod space. As opposed to the central limit theorem, Donsker's theorem is a pathwise statement which ensures that convergence takes place for all times. This result is particularly important for Monte-Carlo methods, which aim to approximate pathwise functionals of a given process (an essential requirement to price path-dependent financial securities for example). We prove here a version of Donsker's result, not only in the Skorokhod topology, but also in the stronger Hölder topology, for a general class of continuous stochastic processes.

1.1. Hölder spaces and fractional operators.

For $\beta \in (0,1]$, the β-Hölder space $\mathcal{C}^\beta(\mathbb{I})$, with the norm

$$\|f\|_\beta := |f|_\beta + \|f\|_\infty = \sup_{\substack{t,s\in\mathbb{I}\\ t\neq s}} \frac{|f(t)-f(s)|}{|t-s|^\beta} + \max_{t\in\mathbb{I}} |f(t)|,$$

is a non-separable Banach space [58, Chapter 3]. In the spirit of Riemann-Liouville fractional operators recalled in Appendix A, we introduce Generalised Fractional Operators (GFO). For $\lambda \in (0,1)$, define the intervals

$$\mathfrak{R}^\lambda := (-\lambda, 1-\lambda), \qquad \mathfrak{R}^\lambda_+ := \mathfrak{R}^\lambda \cap (0,1), \qquad \mathfrak{R}^\lambda_- := \mathfrak{R}^\lambda \cap (-1,0),$$

and the space $\mathcal{L}^\alpha := \{g \in \mathcal{C}^2((0,1]) : \left|\frac{g(u)}{u^\alpha}\right|, \left|\frac{g'(u)}{u^{\alpha-1}}\right| \text{ and } \left|\frac{g''(u)}{u^{\alpha-2}}\right| \text{ bounded}\}$, for $\alpha \in \mathfrak{R}^\lambda$.

Definition 1.1. For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, the GFO associated to $g \in \mathcal{L}^\alpha$ is defined on $\mathcal{C}^\lambda(\mathbb{I})$ as

$$(\mathcal{G}^\alpha f)(t) := \begin{cases} \displaystyle\int_0^t (f(s)-f(0)) \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s, & \text{if } \alpha \in (0, 1-\lambda), \\ \displaystyle\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (f(s)-f(0)) g(t-s)\mathrm{d}s, & \text{if } \alpha \in (-\lambda, 0). \end{cases} \tag{1.1}$$

We shall further use the notation $G(t) := \int_0^t g(u)\mathrm{d}u$, for any $t \in \mathbb{I}$. Of particular interest in mathematical finance are the following kernels:

$$\begin{array}{lll} \text{Riemann-Liouville:} & g(u) = u^\alpha, & \text{for } \alpha \in (-1,1); \\ \text{Gamma fractional:} & g(u) = u^\alpha \mathrm{e}^{\beta u}, & \text{for } \alpha \in (-1,1), \beta < 0; \\ \text{Power-law:} & g(u) = u^\alpha (1+u)^{\beta-\alpha}, & \text{for } \alpha \in (-1,1), \beta < -1. \end{array} \tag{1.2}$$

The next result generalises the classical mapping properties of Riemann-Liouville fractional operators first proved by Hardy and Littlewood [46], and will be of fundamental importance in the rest of our analysis.

Proposition 1.2. *For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, the operator $\mathcal{G}^\alpha$ is continuous from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$.*

The proof can be found in Appendix C. We note that the result is analogous to the classical Schauder estimates, phrased in terms of convolution with a suitable regularising kernel, as e.g. treated in [35, Theorem 14.17] and [17, Theorem 2.13 and Lemma 2.9], but in settings that are slightly different from ours.

We develop here an approximation scheme for the following system, generalising the concept of rough volatility introduced in [3, 37, 41] in the context of mathematical finance, where the process X represents the dynamics of the logarithm of a stock price process:

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2}V_t\mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t, \quad X_0 = 0, \\ V_t &= \Phi\left(\mathcal{G}^\alpha Y\right)(t), \end{aligned} \tag{1.3}$$

with $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, and Y the (strong) solution to the stochastic differential equation

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + a(Y_t)\mathrm{d}W_t, \quad Y_0 \in \mathcal{D}_Y, \tag{1.4}$$

where $\mathcal{D}_Y$ denotes the state space of Y, usually $\mathbb{R}$ or $\mathbb{R}_+$. The two Brownian motions B and W, defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{I}}, \mathbb{P})$, are correlated by the parameter $\rho \in [-1,1]$, and we let the operator Φ be such that, for all $\gamma \in (0,1)$, we have $\Phi : \mathcal{C}^\gamma(\mathbb{I}) \to \mathcal{C}^\gamma_+(\mathbb{I})$ with Φ continuous from $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$ to itself. This in particular implies that whenever $Y \in C^\lambda(\mathbb{I})$ then $V \in C^{\alpha+\lambda}_+(\mathbb{I})$, i.e., V is non-negative and belongs to $C^{\alpha+\lambda}(\mathbb{I})$. As an example, one can consider a so-called Nemyckij operator $\Phi(f) := \phi \circ f$, given by composition with some $\phi : \mathbb{R} \to \mathbb{R}_+$, in which case Drábek [29] has shown that the operator Φ is continuous from $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$ to $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$, for all $\gamma \in (0,1)$, if and only if $\phi \in \mathcal{C}^1(\mathbb{R})$. It remains to formulate a precise definition for $\mathcal{G}^\alpha W$ (Proposition 1.4) and for $\mathcal{G}^\alpha Y$ (Corollary 2.8) to fully specify the system (1.3) and clarify the existence of solutions.

Assumption 1.3. There exist $C_b, C_a > 0$ such that, for all $y \in \mathcal{D}_Y$,

$$|b(y)| \le C_b(1+|y|) \qquad \text{and} \qquad |a(y)| \le C_a(1+|y|),$$

where a and b are continuous functions such that there is a unique strong solution to (1.4).

Existence of solutions to (1.4) along with Assumption 1.3 need to be checked on a case by case basis beyond standard Lipschitz and linear growth conditions (we provide a showcase of models in Examples 1.5-1.8 below satisfying existence and pathwise uniqueness conditions). General conditions for stochastic invariance can be found in [4, 5, 24] for diffusions, and in [1] for affine Volterra processes. Not only is the solution to (1.4) continuous, but $(\frac{1}{2}-\varepsilon)$-Hölder continuous for any $\varepsilon \in (0, \frac{1}{2})$ as a consequence of the Kolmogorov-Čentsov theorem [20]. Existence and precise meaning of $\mathcal{G}^{\alpha}Y$ is delicate, and is treated below.

1.2. **Examples.** Before constructing our approximation scheme, let us discuss a few examples of processes within our framework. As a first useful application, these generalised fractional operators render a (continuous) mapping between a standard Brownian motion and its fractional counterpart:

Proposition 1.4. *For any $\alpha \in \mathfrak{R}^{1/2}$, the equality $(\mathcal{G}^{\alpha}W)(t) = \int_0^t g(t-s)\mathrm{d}W_s$ holds almost surely for all $t \in \mathbb{I}$.*

Proof. Since the paths of Brownian motion are $(\frac{1}{2}-\varepsilon)$-Hölder continuous for any $\varepsilon \in (0, \frac{1}{2})$, existence (and continuity) of $\mathcal{G}^{\alpha}W$ is guaranteed for all $\alpha \in \mathfrak{R}^{1/2}$. When $\alpha \in \mathfrak{R}^{1/2}_{+}$, the kernel is smooth and square integrable, so that Itô's product rule yields (since $g(0)=0$)

$$\begin{aligned}(\mathcal{G}^{\alpha}W)(t) &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)(W(s)-W(0))\mathrm{d}s = g(t)(W(0)-W(0)) - g(0)\left(W(t)-W(0)\right) + \int_0^t g(t-s)\mathrm{d}W_s,\\ &= \int_0^t g(t-s)\mathrm{d}W_s,\end{aligned}$$

and the claim holds. For $\alpha \in \mathfrak{R}^{1/2}_{-}$, and any $\varepsilon > 0$, introduce the operator

$$\left(\mathcal{G}^{1+\alpha}_{\varepsilon} f\right)(t) := \int_0^{t-\varepsilon} g(t-s)(f(s)-f(0))\mathrm{d}s, \qquad \text{for all } t \in \mathbb{I},$$

which satisfies $\frac{\mathrm{d}}{\mathrm{d}t}\lim_{\varepsilon\downarrow 0}\left(\mathcal{G}^{1+\alpha}_{\varepsilon} f\right)(t) = \left(\mathcal{G}^{\alpha} f\right)(t)$ pointwise. Now, for any $t \in \mathbb{I}$, almost surely,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathcal{G}^{1+\alpha}_{\varepsilon} W\right)(t) &= g(\varepsilon)\left(W(t-\varepsilon)-W(0)\right) - g(t)\left(W(0)-W(0)\right) + \int_0^{t-\varepsilon}\frac{\mathrm{d}}{\mathrm{d}t}g(t-s)W(s)\mathrm{d}s\\ &= g(\varepsilon)W(0) + \int_0^{t-\varepsilon} g(t-s)\mathrm{d}W_s. \end{aligned} \tag{1.5}$$

Then, as ε tends to zero, the right-hand side of (1.5) tends to $\int_0^t g(t-s)\mathrm{d}W_s$, and furthermore, the convergence is uniform. On the other hand, the equalities

$$\begin{aligned}(\mathcal{G}^{1+\alpha}_0 W)(t) - (\mathcal{G}^{1+\alpha}_0 W)(0) &= \lim_{\varepsilon\downarrow 0}\left[(\mathcal{G}^{1+\alpha}_{\varepsilon} W)(t) - (\mathcal{G}^{1+\alpha}_{\varepsilon} W)(0)\right] = \lim_{\varepsilon\downarrow 0}\int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{G}^{1+\alpha}_{\varepsilon} W\right)(s)\mathrm{d}s\\ &= \int_0^t \lim_{\varepsilon\downarrow 0}\left(\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{G}^{1+\alpha}_{\varepsilon} W\right)(s)\mathrm{d}s = \int_0^t \left(\int_0^s g(s-u)\mathrm{d}W_u\right)\mathrm{d}s,\end{aligned}$$

hold since convergence is uniform on compacts, and the fundamental theorem of calculus concludes the proof. □

Modulo a constant multiplicative factor C_α, the (left) fractional Riemann-Liouville operator (Appendix A) is identical to the GFO in (1.2), so that the Riemann-Liouville (or Type-II) fractional Brownian motion can be written as $C_\alpha \mathcal{G}^\alpha W$. Proposition 1.2 then implies that the Riemann-Liouville operator is continuous from $\mathcal{C}^{1/2}(\mathbb{I})$ to $\mathcal{C}^{1/2+\alpha}(\mathbb{I})$ for $\alpha \in \mathfrak{R}^{1/2}$. Each kernel in (1.2) gives rise to processes proposed by Barndorff-Nielsen and Schmiegel [8] for turbulence and financial modelling.

Example 1.5. The rough Bergomi model introduced by Bayer, Friz and Gatheral [11] reads

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha \mathrm{d}W_s\right),$$

with $V_0, \nu, \xi_0(\cdot) > 0$, $\alpha \in \mathfrak{R}^{1/2}$ and $\mathcal{E}(\cdot)$ is the Wick stochastic exponential. This corresponds exactly to (1.3) with $g(u) \equiv u^\alpha$, $Y = W$ and

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left\{-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha}\mathrm{d}s\right\}.$$

Example 1.6. A truncated Brownian semistationary ($\mathcal{TBSS}$) process is defined as $\int_0^t g(t-s)\sigma(s)\mathrm{d}W_s$, for $t \in \mathbb{I}$, where σ is $(\mathcal{F}_t)_{t\in\mathbb{I}}$-predictable with locally bounded trajectories and finite second moments, and $g : \mathbb{I}\setminus\{0\} \to \mathbb{I}$ is Borel measurable and square integrable. If $\sigma \in \mathcal{C}^1(\mathbb{I})$, this class falls within the GFO framework.

Example 1.7. Bennedsen, Lunde and Pakkanen [15] considered adding a Gamma kernel to the volatility process, which yields the Truncated Brownian semi-stationary (Bergomi-type) model:

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha \mathrm{e}^{-\beta(t-s)}\mathrm{d}W_s\right),$$

with $\beta > 0$, $\alpha \in \mathfrak{R}^{1/2}$. This corresponds to (1.3) with $Y = W$, Gamma fractional kernel $g(u) \equiv u^\alpha \mathrm{e}^{-\beta u}$ in (1.2),

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left\{-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha}\mathrm{e}^{-2\beta(t-s)}\mathrm{d}s\right\}.$$

Example 1.8. The rough Heston model introduced by Guennoun, Jacquier, Roome and Shi [43] reads

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \kappa(\theta - Y_s)\mathrm{d}t + \int_0^t \xi\sqrt{Y_s}\mathrm{d}W_s, \\ V_t &= \eta + \int_0^t (t-s)^\alpha \mathrm{d}Y_s, \end{aligned}$$

with $Y_0, \kappa, \xi, \theta > 0$, $2\kappa\theta > \xi^2$ and $\eta > 0$, $\alpha \in \mathfrak{R}^{1/2}$. This corresponds exactly to (1.3) with $g(u) \equiv u^\alpha$, $\Phi(\varphi)(t) := \eta + \varphi(t)$, and the coefficients of (1.4) read $b(y) \equiv \kappa(\theta - y)$ and $a(y) \equiv \xi\sqrt{y}$. This model is markedly different from the rough Heston introduced by El Euch and Rosenbaum [31] (for which the characteristic function is known in semi-closed form). Unfortunately, this second version is out of the scope of our invariance principle.

1.3. **The approximation scheme.** We now move on to the core of the project, namely an approximation scheme for the system (1.3)-(1.4). The basic ingredient to construct approximating sequences will be suitable families of iid random variables which satisfy the following assumption:

Assumption 1.9. The family $(\xi_i)_{i\geq 1}$ forms an iid sequence of centered random variables with finite moments of all orders and $\mathbb{E}[\xi_1^2] = \sigma^2 > 0$.

Given $(\zeta_i)_{i\geq 1}$ satisfying Assumption 1.9, Lamperti's [61] generalisation of Donsker's [28] invariance principle tells us that a Brownian motion W can be approximated weakly in Hölder space (for details, see Theorem 2.1) by processes of the form

$$W_n(t) := \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor}\zeta_k + \frac{nt-\lfloor nt\rfloor}{\sigma\sqrt{n}}\zeta_{\lfloor nt\rfloor+1}, \tag{1.6}$$

defined pathwise for any $\omega \in \Omega$, $n \geq 1$, and $t \in \mathbb{I}$. As we explain in Section 2.2, a similar construction holds to weakly approximate the process Y from (1.4) in Hölder space:

$$Y_n(t) := Y_n(0) + \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor} b\left(Y_n^{k-1}\right) + \frac{nt-\lfloor nt\rfloor}{n} b\left(Y_n^{\lfloor nt\rfloor}\right) + \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor} a\left(Y_n^{k-1}\right)\zeta_k + \frac{nt-\lfloor nt\rfloor}{\sigma\sqrt{n}} a\left(Y_n^{\lfloor nt\rfloor}\right)\zeta_{\lfloor nt\rfloor+1}, \tag{1.7}$$

where $Y_n^k := Y_n(t_k)$ and $\mathcal{T}_n := \{t_k = \frac{k}{n}\}_{k=0,\dots,n}$. Here the ζ_i's correspond to the innovations of the Brownian motion W in (1.4). Similarly, we shall use ξ_i when referring to the innovations of the Brownian B from (1.3) which enter into the approximations of the log stock price in (1.8) below. Throughout the paper, we assume that the innovations $\{\xi_i\}_{i=1}^{\lfloor nt\rfloor}$ and $\{\zeta_i\}_{i=1}^{\lfloor nt\rfloor}$ come from two sequences $(\xi_i)_{i\geq 1}$ and $(\zeta_i)_{i\geq 1}$ satisfying Assumption 1.9 such that $((\xi_i, \zeta_i))_{i\geq 1}$ is i.i.d. with $\mathrm{corr}(\xi_i, \zeta_i) = \rho$ for all $i \geq 1$. Naturally, the approximations in (1.7) and in (1.8) below should be understood pathwise, but we omit the ω-dependence in the notations for clarity.

Regarding the approximation scheme for the process X, given by (1.3), we follow a typical route in weak convergence analysis [18, 33] and establish convergence in the Skorokhod space $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Here $\mathcal{D}(\mathbb{I}) = \mathcal{D}(\mathbb{I}, \mathbb{R})$ denotes the space of $\mathbb{R}$-valued càdlàg processes on $\mathbb{I}$ and $d_{\mathcal{D}}$ denotes a metric inducing the Skorokhod topology. To approximate X in this space, we shall then consider the following process:

$$X_n(t) \quad := \quad -\frac{1}{2n}\sum_{k=1}^{\lfloor nt\rfloor}\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_{k-1}) + \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt\rfloor}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_{k-1})}\xi_k. \tag{1.8}$$

Analogously to (1.7), one could view these as continuous processes via linear interpolation, but we note that the interpolating term would decay to zero by Chebyshev's inequality. The following result, proved in Section 2.4, confirms the functional convergence of the approximating sequence $(X_n)_{n\geq 1}$.

Theorem 1.10. *The sequence $(X_n)_{n\geq 1}$ converges weakly to X in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$.*

The construction of the proof allows to extend the convergence to the case where Y is a d-dimensional diffusion without additional work. The proof of the theorem requires a certain number of steps: we start with the convergence of the approximations (Y_n), in some Hölder space, which we then translate into convergence of the sequence $(\Phi(\mathcal{G}^\alpha Y_n))$, by suitable continuity properties of the operations $\mathcal{G}^\alpha$ and Φ, before finally deducing also the convergence of the corresponding stochastic integrals for the approximations of (1.3). These steps are carried out in Sections 2.2, 2.3, and 2.4 below.

2. Functional Central limit theorems for a family of Hölder continuous processes

2.1. **Weak convergence of Brownian motion in Hölder spaces.** Donsker's classical convergence result was proven under the Skorokhod topology. We concentrate here on convergence in the Hölder topology, due to Lamperti [62]. The standard convergence result for Brownian motion can be stated as follows:

Theorem 2.1. *For $\lambda < \frac{1}{2}$, the sequence (W_n) in (1.6) converges weakly to a Brownian motion in $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$.*

The proof relies on finite-dimensional convergence and tightness of the approximating sequence. Not surprisingly, the tightness criterion [18] in the Skorokhod space $\mathcal{D}(\mathbb{I})$ and in a Hölder space are different. In fact, the tightness criterion in Hölder space is strictly related to Kolmogorov-Čentsov's continuity [20]. Note in passing that the approximating sequence (1.6) is piecewise differentiable in time for each $n \geq 1$ even though its limit is obviously not. The proof of Theorem 2.1 follows from Theorem 2.3 under Assumption 1.9.

Theorem 2.2 (Sufficient conditions for weak convergence in Hölder spaces (Račkauskas-Suquet [77]))**.** *Let $Z \in \mathcal{C}^\lambda(\mathbb{I})$ and $(Z_n)_{n\geq 1}$ an approximating sequence in the sense that, for any sequence $(\tau_k)_k$ in $\mathbb{I}$, $(Z_n(\tau_k))_k$ converges in distribution to $(Z(\tau_k))_k$ as n tends to infinity. Assume further that we have*

$$\mathbb{E}\left[|Z_n(t) - Z_n(s)|^\gamma\right] \leq C|t-s|^{1+\beta} \tag{2.1}$$

for all $n \geq 1$, $t, s \in \mathbb{I}$, for some $C, \gamma, \beta > 0$ with $\frac{\beta}{\gamma} \leq \lambda$. Then $(Z_n)_{n\geq 1}$ converges weakly to Z in $\mathcal{C}^\mu(\mathbb{I})$ for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

The proof of this theorem relies on results of Račkauskas and Suquet [77], who prove the convergence in the Hölder space $C_0^\lambda(\mathbb{I})$ endowed with the norm $\|f\|_\lambda^0 := |f|_\lambda + |f(0)|$, for all functions that satisfy

$$\lim_{\delta\downarrow 0} \sup_{\substack{0<t-s<\delta \\ t,s\in\mathbb{I}}} \frac{|f(t)-f(s)|}{(t-s)^\gamma} = 0.$$

From here the proof of Theorem 2.2 is a straightforward consequence, since $\left(C_0^\lambda(\mathbb{I}), \|\cdot\|_\lambda^0\right)$ is a separable closed subspace of $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ (see [45, 77] for details), and one can then use the simple tightness criterion introduced above to conclude. Moreover, as the identity map from $C_0^\lambda(\mathbb{I})$ into $\mathcal{C}^\lambda(\mathbb{I})$ is continuous, weak convergence in the former implies weak convergence in the latter. To conclude our review of weak convergence in Hölder spaces, the following theorem, due to Račkauskas and Suquet [77] provides necessary and sufficient conditions ensuring convergence in Hölder space:

Theorem 2.3 (Račkauskas-Suquet [77])**.** *For any $\lambda \in (0, \frac{1}{2})$, the sequence $(W_n)_{n\geq 1}$ in (1.6) converges weakly to a Brownian motion in $\mathcal{C}^\lambda(\mathbb{I})$ if and only if $\mathbb{E}[\xi_1] = 0$ and $\lim_{t\uparrow\infty} t^{\frac{1}{1-2\lambda}}\mathbb{P}(|\xi_1| \geq t) = 0$.*

Assumption 1.9 ensures the conditions in Theorem 2.3. The following statement allows us to apply Theorem 2.2 on $\mathbb{I}$ and extend the Hölder convergence result via linear interpolation to a continuous sequence.

Theorem 2.4. *Let $Z \in \mathcal{C}^\lambda(\mathbb{I})$ and $(Z_n)_{n\geq 1}$ an approximation sequence such that finite-dimensional convergence holds, i.e. $Z_n(t)$ converges in distribution to $Z(t)$ for $t \in \mathbb{I}$ as n tends to infinity. Moreover, if*

$$\mathbb{E}\left[|Z_n(t_i) - Z_n(t_j)|^\gamma\right] \leq C\left|t_i - t_j\right|^{1+\beta}, \tag{2.2}$$

for any $t_i, t_j \in \mathcal{T}_n$ and some $\beta, \gamma, C > 0$ with $\frac{\beta}{\gamma} \leq \lambda$ and $\gamma \geq 1+\beta$, then the linear interpolating sequence

$$\overline{Z}_n(t) := Z_n\left(\frac{\lfloor nt\rfloor}{n}\right) + (nt - \lfloor nt\rfloor)\left(Z_n\left(\frac{\lfloor nt\rfloor + 1}{n}\right) - Z_n\left(\frac{\lfloor nt\rfloor}{n}\right)\right)$$

satisfies (2.1). In particular, $\overline{Z}_n$ then converges weakly to Z in $\mathcal{C}^\mu(\mathbb{I})$ for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

Proof. For any $t, s \in \mathbb{I}$, we can write, letting $Z_n^k := Z_n(t_k)$ and $\overline{Z}_n^k := \overline{Z}_n(t_k)$,

$$\begin{aligned}
\mathbb{E}\left[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma\right] &= \mathbb{E}\left[\left|Z_n^{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)\left(Z_n^{\lfloor nt \rfloor+1} - Z_n^{\lfloor nt \rfloor}\right) - Z_n^{\lfloor ns \rfloor} - (ns - \lfloor ns \rfloor)\left(Z_n^{\lfloor ns \rfloor+1} - Z_n^{\lfloor ns \rfloor}\right)\right|^\gamma\right] \\
&\leq 3^{\gamma-1}\mathbb{E}\left[\left|Z_n^{\lfloor nt \rfloor} - Z_n^{\lfloor ns \rfloor}\right|^\gamma + (nt - \lfloor nt \rfloor)^\gamma \left|Z_n^{\lfloor nt \rfloor+1} - Z_n^{\lfloor nt \rfloor}\right|^\gamma + (ns - \lfloor ns \rfloor)^\gamma \left|Z_n^{\lfloor ns \rfloor+1} - Z_n^{\lfloor ns \rfloor}\right|^\gamma\right] \\
&\leq C\left(\left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n}\right)^{1+\beta} + \frac{(nt - \lfloor nt \rfloor)^\gamma}{n^{1+\beta}} + \frac{(ns - \lfloor ns \rfloor)^\gamma}{n^{1+\beta}}\right) \leq C(t-s)^{1+\beta},
\end{aligned}$$

where we used (2.2) and the fact that $\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \leq 2(t-s)$, $nt - \lfloor nt \rfloor \leq 1$ for $t \geq 0$ and $\frac{1}{n} \leq (t-s)$.

Finally, it is left to prove the case $\frac{1}{n} > (t-s)$. There are two possible scenarios here:

- If $\lfloor nt \rfloor = \lfloor ns \rfloor$, then, using $\gamma \geq 1+\beta$, we have

$$\mathbb{E}[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma] = \mathbb{E}\left[\left|(nt - ns)\left(Z_n^{\lfloor nt \rfloor+1} - Z_n^{\lfloor nt \rfloor}\right)\right|^\gamma\right] \leq \frac{C|t-s|^\gamma}{n^{1+\beta-\gamma}} \leq C(t-s)^{1+\beta}$$

- If $\lfloor nt \rfloor \neq \lfloor ns \rfloor$, then either $\lfloor nt \rfloor + 1 = \lfloor ns \rfloor$ or $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$. Without loss of generality consider the second case. Then

$$\begin{aligned}
\mathbb{E}\left[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma\right] &= \mathbb{E}\left[\left|\overline{Z}_n(t) - Z_n^{\lfloor nt \rfloor} + Z_n^{\lfloor nt \rfloor} - \overline{Z}_n(s)\right|^\gamma\right] \leq 2^{\gamma-1}\mathbb{E}\left[\left|\overline{Z}_n(t) - Z_n^{\lfloor nt \rfloor}\right|^\gamma + \left|Z_n^{\lfloor nt \rfloor} - \overline{Z}_n(s)\right|^\gamma\right] \\
&\leq C\left((t-s)^{1+\beta} + \mathbb{E}\left[\left|(\lfloor nt \rfloor - ns)\left(Z_n^{\lfloor nt \rfloor} - Z_n^{\lfloor nt \rfloor-1}\right)\right|^\gamma\right]\right),
\end{aligned}$$

  and the result follows as before since $t - \frac{\lfloor nt \rfloor}{n} < |t-s|$ and $|s - \frac{\lfloor nt \rfloor}{n}| \leq |t-s|$.

□

2.2. **Weak convergence of Itô diffusions in Hölder spaces.** The first important step in our analysis is to extend Donsker-Lamperti's weak convergence from Brownian motion to the Itô diffusion Y in (1.4).

Theorem 2.5. *The sequence* $(Y_n)_{n\geq 1}$ *in* (1.7) *converges weakly to* Y *in* (1.4) *in* $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ *for all* $\lambda < \frac{1}{2}$, *Proof.* Finite-dimensional convergence is a classical result by Kushner [60], so only tightness needs to be checked. In particular, using Theorem 2.4 we need only consider the partition $\mathcal{T}_n$. Thus, we get

$$\begin{aligned}
\mathbb{E}\left[|Y_n^j - Y_n^i|^{2p}\right] &= \mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n} b\left(Y_n^{k-1}\right) + \frac{1}{\sigma\sqrt{n}} a\left(Y_n^{k-1}\right)\zeta_k\right|^{2p}\right] \\
&\leq 2^{2p-1}\left\{\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n} b\left(Y_n^{k-1}\right)\right|^{2p}\right] + \mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{a\left(Y_n^{k-1}\right)\zeta_k}{\sigma\sqrt{n}}\right|^{2p}\right]\right\} \\
&\leq 2^{2p-1}\left\{\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n} b\left(Y_n^{k-1}\right)\right|^{2p}\right] + C(p)\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{a\left(Y_n^{k-1}\right)^2\zeta_k^2}{\sigma^2 n}\right|^{p}\right]\right\} \\
&\leq 2^{2p-1}\left\{\frac{(j-i)^{2p-1}}{n^{2p}}\sum_{k=i+1}^{j}\mathbb{E}\left[\left|b\left(Y_n^{k-1}\right)\right|^{2p}\right] + \frac{(j-i)^{p-1}}{n^p}C(p)\frac{\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\sum_{k=i+1}^{j}\mathbb{E}\left[a\left(Y_n^{k-1}\right)^{2p}\right]\right\} \\
&\leq 2^{2p-1}\frac{(j-i)^{p-1}}{n^p}\sum_{k=i+1}^{j}\left(C_b^{2p}\mathbb{E}\left[(1+|Y_n^{k-1}|)^{2p}\right] + C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\mathbb{E}\left[\left(1+|Y_n^{k-1}|\right)^{2p}\right]\right) \\
&\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\frac{(j-i)^{p-1}}{n^p}\left\{(j-i) + \sum_{k=i+1}^{j}\left(\mathbb{E}\left[|Y_n^{k-1}|)^{2p}\right]\right)\right\} \\
&\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp\left(\sum_{k=i+1}^{j} 2^{2p}\frac{(j-i)^{p-1}}{n^p}\right)(t_j-t_i)^p \\
&\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp\left(2^{2p}\right)(t_j-t_i)^p := \mathfrak{C}(p)(t_j-t_i)^p,
\end{aligned}$$

where we have used the discrete version of the BDG inequality [9, Theorem 6.3] in the martingale term $\sum_{k=i+1}^{j}\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k-1}\right)\zeta_k$ with $C(p) := 6^p(p-1)^{p-1}$. Indeed, for the discrete-time martingale process $(x_n^{i,j})_u := \sum_{k=1}^{u}\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k+i-1}\right)\zeta_{i+k}$ for $u \in \{1,...,j-i\}$, we have $|(x_n^{i,j})_{j-i}| \leq \max_{u\in\{1,..,j-i\}}|(x_n^{i,j})_u|$ and the BDG inequality clearly also applies to $|x_{j-i}^{i,j}|$. We also used independence of ζ_k and Y_{k-1} and the linear growth of $b(\cdot)$ and $a(\cdot)$ from Assumption 1.3, Hölder inequality and the discrete version of Gronwall's lemma [21] in the last step. Since $\mathbb{E}[\zeta_k^{2p}]$ is bounded by Assumption 1.9 and the constant $\mathfrak{C}(p)$ only depends on p, but not on n, then the tightness criterion (2.2) of Theorem 2.4 holds for $p > 1$ with $\gamma = 2p$ and $\beta = p - 1$. $\square$

Corollary 2.6. *Let $(Y_n)_{n\geq 1}$ be defined as in Theorem 2.5 with innovations $(\zeta_i)_{i\geq 1}$, and suppose $(B^n)_{n\geq 1}$ is defined by the Donsker approximations* (1.6)*, for some innovations $(\xi_i)_{i\geq 1}$ satisfying Assumption 1.9 such that $((\zeta_i,\xi_i))_{i\geq 1}$ is iid with $\operatorname{corr}(\zeta_i,\xi_i) = \rho$, for all $i \geq 1$. Then there is joint weak convergence of (B_n, Y_n) to (B, Y) in $\left(\mathcal{C}^\lambda(\mathbb{I},\mathbb{R}^2), \|\cdot\|_\lambda\right)$, for all $\lambda < \frac{1}{2}$, for a standard Brownian motion B such that $[B,W]_t = \rho t$, for $t \in \mathbb{I}$, where W is the standard Brownian motion driving the dynamics of the weak limit Y in* (1.4)*.*

Proof. Take $(\zeta_i^\perp)_{i\geq 1}$ to satisfy Assumption 1.9 and be independent of the innovations $(\zeta_i)_{i\geq 1}$ defining $(Y^n)_{n\geq 1}$. Then set $\xi_i := \rho\zeta_i + \sqrt{1-\rho^2}\zeta_i^\perp$, for $i \geq 1$, and let B_n be defined in terms of $(\xi_i)_{i\geq 1}$. This yields the same finite-dimensional distributions of (B_n, Y_n) as for the general $(\xi_i)_{i\geq 1}$ in the statement of the corollary. Consider now the drift vector $\mathbf{b}(y) = (0, b(y))$ and the 2×2 diffusion matrix $\mathbf{a}(y)$ with rows $(\rho, \sqrt{1-\rho^2})$ and $(0, a(y))$. Then Kushner [60] applies directly to give finite-dimensional convergence with the desired limit. Finally, tightness of (B_n, Y_n) follows analogously to the proof of Theorem 2.5. Hence the claim follows. $\square$

2.3. **Invariance principle for rough processes.** We have set the ground to extend our results to processes that are not necessarily $(1/2-\varepsilon)$-Hölder continuous, Markovian nor semimartingales. More precisely, we are interested in α-Hölder continuous paths with $\alpha \in (0,1)$, such as Riemann-Liouville fractional Brownian motion or some $\mathcal{TBSS}$ processes. A key tool is the Continuous Mapping Theorem, first proved by Mann and Wald [66], which establishes the preservation of weak convergence under continuous operators.

Theorem 2.7 (Continuous Mapping Theorem). *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be two normed spaces and assume that $g : \mathcal{X} \to \mathcal{Y}$ is a continuous operator. If the sequence of random variables $(Z_n)_{n\geq 1}$ converges weakly to Z in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, then $(g(Z_n))_{n\geq 1}$ also converges weakly to $g(Z)$ in $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$.*

Many authors have exploited the combination of Theorems 2.1 and 2.7 to prove weak convergence [76, Chapter IV]. This path avoids the lengthy computations of tightness and finite-dimensional convergence in classical proofs [18]. In fact, Hamadouche [45] already realised that Riemann-Liouville fractional operators are continuous, hence Theorem 2.7 holds under mapping by Hölder continuous functions. In contrast, the novelty here is to consider the family of GFO applied to Brownian motion together with the extension of Brownian motion to Itô diffusions. In fact, minimal changes to the proof of Proposition 1.4 yield the following:

Corollary 2.8. *If Y solves* (1.4), *then $(\mathcal{G}^\alpha Y)(t) = \int_0^t g(t-s)\mathrm{d}Y_s$ almost surely for all $t \in \mathbb{I}$ and $\alpha \in \mathfrak{R}^{\frac{1}{2}}$.*

The analogue of Theorem 2.5 for $\mathcal{G}^\alpha Y$ holds as follows:

Theorem 2.9 (Generalised rough Donsker). *For (Y_n) in* (1.7), *Y its weak limit in $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ for $\lambda < \frac{1}{2}$, then the representation*

$$(\mathcal{G}^\alpha Y_n)(t) = \sum_{i=1}^{\lfloor nt \rfloor} n\big[G(t-t_{i-1}) - G(t-t_i)\big]\left(Y_n^i - Y_n^{i-1}\right) + n\,G(t-t_{\lfloor nt\rfloor})\big(Y_n(t) - Y_n^{\lfloor nt\rfloor}\big), \quad t \in \mathbb{I}, \tag{2.3}$$

holds. Furthermore this sequence $(\mathcal{G}^\alpha Y_n)_{n\geq 1}$ converges weakly to $\mathcal{G}^\alpha Y$ in $\left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$ for any $\alpha \in \mathfrak{R}^\lambda$.

Proof. Recall that the sequence (1.7) is piecewise differentiable in time. For $\alpha \in \mathfrak{R}^\lambda_+$, note that $g(0) = 0$ and therefore by integration by parts [84, Section 2.4] (where Y_n is piecewise differentiable), for $n \geq 1$ and $t \in \mathbb{I}$,

$$\begin{aligned}
(\mathcal{G}^\alpha Y_n)(t) &= \int_0^t g'(t-s)(Y_n(s) - Y_n(0))\mathrm{d}s = \int_0^t g(t-s)\frac{\mathrm{d}(Y_n(s) - Y_n(0))}{\mathrm{d}s}\mathrm{d}s \\
&= \frac{1}{\sigma\sqrt{n}}\left[\sum_{i=1}^{\lfloor nt\rfloor} n\int_{t_{i-1}}^{t_i} g(t-s)a\left(Y_n^{i-1}\right)\zeta_i \mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^t g(t-s)a\left(Y_n^{\lfloor nt\rfloor}\right)\zeta_{\lfloor nt\rfloor+1}\mathrm{d}s\right] \\
&+ \frac{1}{n}\left[n\sum_{i=1}^{\lfloor nt\rfloor}\int_{t_{i-1}}^{t_i} g(t-s)b\left(Y_n^{i-1}\right)\mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^t g(t-s)b\left(Y_n^{\lfloor nt\rfloor}\right)\mathrm{d}s\right] \\
&= \sum_{i=1}^{\lfloor nt\rfloor} n\left[G(t-t_{i-1}) - G(t-t_i)\right]\left(Y_n^i - Y_n^{i-1}\right) + n(G(t-t_{\lfloor nt\rfloor}) - G(0))\big(Y_n(t) - Y_n^{\lfloor nt\rfloor}\big),
\end{aligned}$$

and (2.3) follows since $G(0) = 0$ in the last line. When $\alpha \in \mathfrak{R}^\lambda_-$, using $G(0) = 0$, we similarly get

$$\begin{aligned}
\int_0^t g(t-s)(Y_n(s)-Y_n(0))\mathrm{d}s &= \int_0^t G(t-s)\frac{\mathrm{d}(Y_n(s)-Y_n(0))}{\mathrm{d}s}\mathrm{d}s \\
&= \frac{1}{\sigma\sqrt{n}}\left[\sum_{i=1}^{\lfloor nt\rfloor} n\int_{t_{i-1}}^{t_i} G(t-s)a\left(Y_n^{i-1}\right)\zeta_i\mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^t G(t-s)a\left(Y_n^{\lfloor nt\rfloor}\right)\zeta_{\lfloor nt\rfloor+1}\mathrm{d}s\right] \\
&+ \frac{1}{n}\left[n\sum_{i=1}^{\lfloor nt\rfloor}\int_{t_{i-1}}^{t_i} G(t-s)b\left(Y_n^{i-1}\right)\mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^t G(t-s)b\left(Y_n^{\lfloor nt\rfloor}\right)\mathrm{d}s\right] \\
&= n\left\{\sum_{i=1}^{\lfloor nt\rfloor}\left[\frac{b\left(Y_n^{i-1}\right)}{n} + \frac{a\left(Y_n^{i-1}\right)}{\sigma\sqrt{n}}\zeta_i\right]\int_{t_{i-1}}^{t_i} G(t-s)\mathrm{d}s + \left[\frac{b\left(Y_n^{\lfloor nt\rfloor}\right)}{n} + \frac{a\left(Y_n^{\lfloor nt\rfloor}\right)}{\sigma\sqrt{n}}\zeta_{\lfloor nt\rfloor+1}\right]\int_{t_{\lfloor nt\rfloor}}^t G(t-s)\mathrm{d}s\right\} \\
&= n\left\{\sum_{i=1}^{\lfloor nt\rfloor}\left(Y_n^i - Y_n^{i-1}\right)\int_{t_{i-1}}^{t_i} G(t-s)\mathrm{d}s + \left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right)\int_{t_{\lfloor nt\rfloor}}^t G(t-s)\mathrm{d}s\right\},
\end{aligned}$$

and from there it follows readily that

$$\begin{aligned}
(\mathcal{G}^\alpha Y_n)(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t g(t-s)(Y_n(s)-Y_n(0))\mathrm{d}s \\
&= \sum_{i=1}^{\lfloor nt\rfloor} n\left[G(t-t_{i-1}) - G(t-t_i)\right]\left(Y_n^i - Y_n^{i-1}\right) + nG(t-t_{\lfloor nt\rfloor})\big(Y_n(t) - Y_n^{\lfloor nt\rfloor}\big),
\end{aligned}$$

as desired (when $t = \frac{k}{n}$ the difference quotients pick up an extra term, but this vanishes in the limit). Finally, the claimed convergence follows analogously to that in Theorem 2.5 by continuous mapping, along with the fact that $\mathcal{G}^\alpha$ is a continuous operator from $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $\left(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha}\right)$ for all $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$. □

Notice here that the mean value theorem implies

$$(\mathcal{G}^\alpha Y_n)(t) = \sum_{i=1}^{\lfloor nt\rfloor} g\big(t_i^*\big)\big(Y_n^i - Y_n^{i-1}\big) + g\big(t^*_{\lfloor nt\rfloor+1}\big)\big(Y_n(t) - Y_n^{\lfloor nt\rfloor}\big), \tag{2.4}$$

where $t_i^* \in [t-t_i, t-t_{i-1}]$ and $t^*_{\lfloor nt\rfloor+1} \in [0, t-t_{\lfloor nt\rfloor}]$ and we use that $G(0) = 0$. This expression is closer to the usual left-point forward Euler approximation. For numerical purposes, (2.4) is much more efficient, since the integral G required in (2.3) is not necessarily available in closed form. Nevertheless, not any arbitrary choice of t_i^* gives the desired convergence from the above argument. We shall present a suitable candidate for optimal t_i^* in Section 3.3.1, which guarantees weak convergence in Hölder sense.

As could be expected, the Hurst parameter influences the speed of convergence of the scheme. We leave a formal proof to further study, but the following argument provides some intuition about the correct normalising factor: Given $g \in \mathcal{L}^\alpha$, we can write $g(u) = u^\alpha L(u)$, where L is a bounded function on $\mathbb{I}$. At time $t = t_i$, take $t_k^* = t_i - t_k + \frac{\varepsilon}{n}$ for $\varepsilon \in [0,1]$. For $\alpha \in \mathfrak{R}^\lambda_-$, since $g \in \mathcal{L}^\alpha$, we can rewrite the approximation (2.4) as

$$(\mathcal{G}^\alpha Y_n)(t_i) = \frac{1}{n^{1/2+\alpha}}\sum_{k=1}^{i}(i-k+\varepsilon)^\alpha L(t_k^*)\left(Y_n^k - Y_n^{k-1}\right)\sqrt{n}, \quad \text{for } i = 0, \ldots, n.$$

Here, $(i-k+\varepsilon)^\alpha \leq \varepsilon^\alpha$ is bounded in $n \geq 1$ as long as $\varepsilon \in (0,1]$, so the normalisation factor is of order $n^{-\alpha-1/2}$. When $\alpha \in \mathfrak{R}_+^\lambda$, the approximation (2.4) instead reads as

$$(\mathcal{G}^\alpha Y_n)(t_i) = \frac{1}{\sqrt{n}} \sum_{k=1}^{i} \left(t_i - t_k + \tfrac{\varepsilon}{n}\right)^\alpha L(t_k^*) \left(Y_n^k - Y_n^{k-1}\right) \sqrt{n}, \quad \text{for } i = 0, \ldots, n,$$

in which case $(t_i - t_k + \frac{\varepsilon}{n})^\alpha \leq t_i^\alpha$ is bounded in $n \geq 1$, and hence the normalisation factor is of order $n^{-1/2}$. This intuition is consistent with the result by Neuenkirch and Shalaiko [71], who found the strong rate of convergence of the Euler scheme to be of order $\mathcal{O}(n^{-H})$ for $H < \frac{1}{2}$ for fractional Ornstein-Uhlenbeck. So far, our results hold for α-Hölder continuous functions; however, for practical purposes, it is often necessary to constrain the volatility process $(V_t)_{t\in\mathbb{I}}$ to remain strictly positive at all times. The stochastic integral $\mathcal{G}^\alpha Y$ need not be so in general. However, a simple transformation (e.g. exponential) can easily overcome this fact. The remaining question is whether the α-Hölder continuity is preserved after this composition.

Proposition 2.10. *Let $(Y_n)_{n\geq 1}$ be the approximating sequence* (1.7) *in $\mathcal{C}^\lambda(\mathbb{I})$ for $\lambda < 1/2$. Then $(\Phi(\mathcal{G}^\alpha Y_n))$ converges weakly to $\Phi(\mathcal{G}^\alpha Y)$ in $\left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$ for all $\alpha \in \mathfrak{R}^\lambda$.*

Proof. Theorem 2.9 gives that $\mathcal{G}^\alpha Y_n$ converges weakly to $\mathcal{G}^\alpha Y$ in $\left(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha}\right)$. By our assumptions, Φ is continuous from $\left(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha}\right)$ to $\left(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha}\right)$. The proposition thus follows from the continuous mapping theorem. The diagram below summarises the steps with $\lambda < 1/2$. The double arrows show weak convergence, and we indicate next to them the topology in which it takes place.

$$\begin{tikzcd}
\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right) \arrow[r, bend left, "\mathcal{G}^\alpha"] & \left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right) \arrow[r, bend left, "\Phi"] & \left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right) \\
Y_n \arrow[r, "\mathcal{G}^\alpha"] \arrow[d, Rightarrow, "\|\cdot\|_\lambda"'] & \mathcal{G}^\alpha(Y_n) \arrow[r, "\Phi"] \arrow[d, Rightarrow, "\|\cdot\|_{\alpha+\lambda}"'] & \Phi(\mathcal{G}^\alpha Y_n) \arrow[d, Rightarrow, "\|\cdot\|_{\alpha+\lambda}"'] \\
Y \arrow[r, "\mathcal{G}^\alpha"] & \mathcal{G}^\alpha Y \arrow[r, "\Phi"] & \Phi(\mathcal{G}^\alpha Y)
\end{tikzcd}$$

□

2.4. **Extending the weak convergence to the Skorokhod space and proof of Theorem 1.10.** The Skorokhod space of càdlàg processes equipped with the Skorokhod topology has been widely used to prove weak convergence [18, 33]. The Skorokhod space of càdlàg processes equipped with the Skorokhod norm, which we denote $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$, markedly simplifies when we only consider continuous processes (as is the case of our framework with Hölder continuous processes). Billingsley [18, Chapter 3, Section 12] proved that the identity $(\mathcal{D}(\mathbb{I}) \cap \mathcal{C}(\mathbb{I}), d_\mathcal{D}) = (\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$ always holds. This seemingly simple statement allows us to reduce proofs of weak convergence of continuous processes in the Skorokhod topology to that in the supremum norm, usually much simpler. We start with the following straightforward observation:

Lemma 2.11. *For any $\lambda \in (0,1)$, the identity map is continuous from $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$.*

Proof. Since the identity map is linear, it suffices to check that it is bounded. For this observe that $\|f\|_\lambda = |f|_\lambda + \sup_{t\in\mathbb{I}}|f(t)| = |f|_\lambda + \|f\|_\infty > \|f\|_\infty$, where $|f|_\lambda > 0$, which concludes the proof since the Skorokhod norm in the space of continuous functions is equivalent to the supremum norm. $\square$

Applying the Continuous Mapping Theorem twice, first with the Generalised fractional operator (Theorem 2.9), then with the identity map, yields the following result directly:

Theorem 2.12. *For any $\alpha \in \mathfrak{R}^{1/2}$, the sequence $(\Phi(\mathcal{G}^\alpha Y_n))$ converges weakly to $\Phi(\mathcal{G}^\alpha Y)$ in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Moreover, the sequence is tight in $(\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$.*

The final step in the proof of our main theorem is to extend the functional weak convergence to the log-stock price X. For this, we will rely on the weak convergence theory for stochastic integrals due to Jakubowski, Memin and Pagès [52] and further developed by Kurtz and Protter [59]. Throughout, we write $H \bullet N := \int_0^\cdot H(s)\mathrm{d}N(s)$ and we use the notation H^- for the process $H^-(t) := H(t-)$ obtained by taking left limits.

Theorem 2.13 (Kurtz and Protter [59])**.** *For each $n \geq 1$, let $N_n = M_n + A_n$ be an $(\mathcal{F}_t^n)$-semimartingale and let H_n be an $(\mathcal{F}_t^n)$-adapted càdlàg process on $\mathbb{I}$. Suppose that, for all $\gamma > 0$, there are $(\mathcal{F}_t^n)$-stopping times (τ_n^γ) such that $\sup_{n\geq 1}\mathbb{P}(\tau_n^\gamma \leq \gamma) \leq 1/\gamma$ and $\sup_{n\geq 1}\mathbb{E}[[M_n]_{\tau_n^\gamma\wedge 1} + T_{\tau_n^\gamma\wedge 1}(A_n)] < \infty$, where T_t denotes the total variation on $[0,t]$. If (H_n, N_n) converges weakly to (H,N) in $(\mathcal{D}(\mathbb{I},\mathbb{R}^2), d_{\mathcal{D}})$, then N is a semimartingale in the filtration generated by (H,N) and $(H_n, N_n, H_n^- \bullet N_n)$ converges weakly to $(H, N, H^- \bullet N)$ in $(\mathcal{D}(\mathbb{I},\mathbb{R}^3), d_{\mathcal{D}})$.*

The above amounts to a restatement of [59, Theorem 2.2] in the special case $\delta = \infty$ (in their notations) and restricted to real-valued processes on $\mathbb{I}$. With this, we can now give the proof of Theorem 1.10, which asserts the functional weak convergence of the approximations X_n from (1.8) to the desired log-price X from (1.3).

Proof of Theorem 1.10. We begin by considering, for all $n \geq 1$, the particular approximations

$$M_n(t) := \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\xi_i,$$

$t \in \mathbb{I}$, of the driving Brownian motion B in the dynamics of X. Here the ξ_i satisfy Assumption 1.9 and so does the ζ_i in the construction of Y_n from (1.7). While each pair ξ_i and ζ_i are correlated, they form an iid sequence $\{(\zeta_i,\xi_i)\}_{i\geq 1}$ across the pairs. In particular, it is straightforward to see that each M_n is a martingale on $\mathbb{I}$ for the filtration $(\mathcal{F}_t^n)$ defined by $\mathcal{F}_t^n := \sigma(\zeta_i, \xi_i : i = 1, \dots, \lfloor nt\rfloor)$. Moreover, we have

$$\mathbb{E}\big[[M_n]_t\big] = \lfloor nt\rfloor \frac{1}{\sigma^2 n}\mathbb{E}[\xi_1^2] = \frac{\lfloor nt\rfloor}{n} \leq 1,$$

for $t \in \mathbb{I}$, for all $n \geq 1$. Consequently, we can simply take $\tau_n^\gamma := +\infty$, for all $\gamma > 0$ and $n \geq 1$, to satisfy the required control on the integrators $N_n := M_n$ in Theorem 2.13. By [33, Chapter 7, Theorem 1.4], the M_n converge weakly to a Brownian motion B in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Now fix $\alpha \in \mathfrak{R}^{\frac{1}{2}}$ and define a sequence of càdlàg processes H_n on $\mathbb{I}$, for all $n \geq 1$, by setting $H_n(1) := \Phi(\mathcal{G}^\alpha Y_n)(1)$ and $H_n(t) := \Phi(\mathcal{G}^\alpha Y_n)(t_{k-1})$ for $t \in [t_{k-1}, t_k)$, for each $k = 1, \dots, n$. In view of Theorem 2.12, the Arzela–Ascoli characterisation of tightness [18, Theorem 8.2] for the space $(\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$ allows us to conclude that the H_n converge weakly to $H := \Phi(\mathcal{G}^\alpha Y)$ in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Furthermore, recalling the definition of Y_n in (1.7), each H_n is adapted to the filtration $(\mathcal{F}_t^n)$ introduced above. By Corollary 2.6, we readily deduce that there is joint weak convergence of (Y_n, H_n, M_n) to (Y, H, B) on $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}}) \times (\mathcal{D}(\mathbb{I}), d_{\mathcal{D}}) \times (\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$, where Y satisfies (1.4) for a Brownian motion W with $[W,B]_t = \rho t$, for all $t \in \mathbb{I}$. As noted in [59], the Skorokhod topology on $\mathcal{D}(\mathbb{I},\mathbb{R}^2)$ is stronger than the product topology on

$\mathcal{D}(\mathbb{I}) \times \mathcal{D}(\mathbb{I})$, but here it automatically follows that we have weak convergence of the pairs (H_n, M_n) to (H, B) in $\big(\mathcal{D}(\mathbb{I}, \mathbb{R}^2), d_{\mathcal{D}}\big)$, by standard properties of the Skorokhod topology (e.g. [33, Chapter 3, Theorem 10.2]), since the limiting pair (H, B) is continuous. Consequently, we are in a position to apply Theorem 2.13. To this end, observe that

$$(H_n^- \bullet M_n)(t) = \sum_{k=1}^{\lfloor nt \rfloor} H_n(t_k-)(M_n(t_k) - M_n(t_k-)) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \Phi(\mathcal{G}^\alpha Y_n)(t_{k-1})\xi_k,$$

which is precisely the second term on the right-hand side of (1.8). Therefore, Theorem 2.13 gives that the stochastic integral $H \bullet M = \sqrt{\Phi(\mathcal{G}^\alpha Y)} \bullet B$ is the weak limit in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$ of the second term on the right-hand side of (1.8). For the first term on the right-hand side of (1.8), we have $-\frac{1}{2}\int_0^\cdot H_n(s)\mathrm{d}s$ converging weakly to $-\frac{1}{2}\int_0^\cdot H(s)\mathrm{d}s$, by the continuous mapping theorem, as the integral is a continuous operator from $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$ to itself. Since there is weak convergence of the pairs $(H_n, H_n^- \bullet M_n)$ to $(H, H \bullet B)$ in $(\mathcal{D}(\mathbb{I}, \mathbb{R}^2), d_{\mathcal{D}})$, the sum of the two terms on the right-hand side of (1.8) are then also weakly convergent in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Recalling that the limit Y satisfies (1.4) for a Brownian motion W such that W and B are correlated with parameter ρ, we hence conclude that X converges in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$ to the desired limit. □

3. Applications

3.1. **Weak convergence of the Hybrid scheme.** The Hybrid scheme (and its turbocharged version [67]) introduced by Bennedsen, Lunde and Pakkanen [14] is the current state-of-the-art to simulate $\mathcal{TBSS}$ processes. However, only convergence in mean-square-error was proved, but not (functional) weak convergence, which would justify the use of the scheme for path-dependent options. Unless otherwise stated, we shall denote by $\mathcal{T}_n := \{t_k = \frac{k}{n}\}_{k=0,\dots,n}$ the uniform grid on $\mathbb{I}$. We show that the Hölder convergence also holds for the case $g(x) = x^\alpha$:

Proposition 3.1. *The Hybrid scheme sequence* $(\widetilde{\mathcal{G}}^\alpha W_n)$ *defined as*

$$\widetilde{\mathcal{G}}^\alpha W_n(t) := \widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt\rfloor}{n}\right) + (nt - \lfloor nt\rfloor)\left(\widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt\rfloor + 1}{n}\right) - \widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt\rfloor}{n}\right)\right), \tag{3.1}$$

for $t \in \mathbb{I}$*, where*

$$\widetilde{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{(i-\kappa)\vee 0} \int_{t_{k-1}}^{t_k} \sqrt{n}(t_i - s)^\alpha \mathrm{d}s\xi_k + \int_{0\vee t_{i-\kappa}}^{t_i} (t_i - s)^\alpha \mathrm{d}W_s, \quad i = 0, \dots, n, \quad \kappa \geq 1. \tag{3.2}$$

with $\xi_k := \int_{t_{k-1}}^{t_k} \mathrm{d}W_s \sim \mathcal{N}(0, 1/n)$*, converges to* $\mathcal{G}^\alpha W$ *in* $\big(\mathcal{C}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\big)$ *for* $\alpha \in \mathfrak{R}^{1/2}$ *and* $\kappa = 1$.

Proof. Finite-dimensional convergence follows trivially as the target process is centered Gaussian, thus convergence of the covariance matrix ensures finite-dimensional convergence. To prove convergence we only need to to show that the approximating sequence is tight, by verifying the criteria from Theorem 2.4 as follows:

$$\mathbb{E}\left[\left|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)\right|^{2p}\right] \leq C|t_i - t_j|^{2p\alpha+p},$$

for all $t_i, t_j \in \mathcal{T}_n$, for $p \geq 1$ and some constant $C \geq 0$. Without loss of generality assume $t_j < t_i$ and take $\kappa = 1$. Define

$$\widetilde{\sigma}^2 := \mathbb{E}\left[\left|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)\right|^2\right].$$

We note that

$$\mathbb{E}\left[n\left(\int_{t_{j-1}}^{t_j}(t_i-s)^\alpha \mathrm{d}s\int_{t_{j-1}}^{t_j}\mathrm{d}W_s-\int_{t_{j-1}}^{t_j}(t_j-s)^\alpha \mathrm{d}W_s\right)^2\right]\leq\int_{t_{j-1}}^{t_j}\left((t_i-s)^\alpha-(t_j-s)^\alpha\right)^2\mathrm{d}s,$$

where we have used Chebyshev's integral inequality. Therefore,

$$\begin{aligned}
\widetilde{\sigma}^2 &\leq \sum_{k=1}^{j}\left(\int_{t_{k-1}}^{t_k}(t_i-s)^\alpha-(t_j-s)^\alpha \mathrm{d}s\right)^2+\int_{t_{j-1}}^{t_j}\left((t_i-s)^\alpha-(t_j-s)^\alpha\right)^2\mathrm{d}s+\int_{t_j}^{t_i}(t_i-s)^{2\alpha}\mathrm{d}s\\
&\leq \sum_{k=1}^{j-1}\left(\int_{t_{k-1}}^{t_k}(t_i-s)^\alpha-(t_j-s)^\alpha \mathrm{d}s\right)^2+\int_{t_{j-1}}^{t_j}(t_i-t_j)^{2\alpha}\mathrm{d}s+\frac{1}{2\alpha+1}(t_i-t_j)^{2\alpha+1}\\
&\leq \sum_{k=1}^{j-1}\left(\int_{t_{k-1}}^{t_k}(t_i-t_j)^\alpha\,\mathrm{d}s\right)^2+\frac{1}{n}(t_i-t_j)^{2\alpha}\mathrm{d}s\frac{1}{2\alpha+1}(t_i-t_j)^{2\alpha+1}\\
&\leq \frac{1}{n}\left(t_i-t_j\right)^{2\alpha}+\frac{1}{n}(t_i-t_j)^{2\alpha}\mathrm{d}s\frac{1}{2\alpha+1}(t_i-t_j)^{2\alpha+1}\leq C(t_i-t_j)^{2\alpha+1}
\end{aligned}$$

where we have used the power inequality $|x|^p-|y|^p\leq|x-y|^p$ for $p\leq 1$. Thus, by standard moment properties of Gaussian random variables [19, Theorem 2.1] we obtain

$$\mathbb{E}\left[\left|\widetilde{\mathcal{G}}^\alpha W_n(t_i)-\widetilde{\mathcal{G}}^\alpha W_n(t_j)\right|^{2p}\right]\leq\widetilde{C}\widetilde{\sigma}^{2p}\leq\widetilde{C}(t_i-t_j)^{2p\alpha+p},$$

which gives the desired result. □

We further note that Proposition 3.1 and Theorem 1.10 ensure the weak convergence of the log-stock price for the Hybrid scheme as well.

Remark 3.2. Proposition 3.1 may easily be extended to a d-dimensional Brownian motion W (for example for multifactor volatility models), also providing a weak convergence result for the d-dimensional version of the Hybrid scheme recently developed by Heinrich, Pakkanen and Veraart [47].

3.2. **Application to fractional binomial trees.** We consider a binomial setting for the Riemann-Liouville fractional Brownian motion $\mathcal{G}^{H-1/2}W$ with $g(u)\equiv u^{H-1/2}$, $H\in(0,1)$, for which Theorem 2.9 provides a weakly converging sequence. On the partition $\mathcal{T}_n$, with a Bernoulli sequence $\{\zeta_i\}_{i=1}^n$ satisfying $\mathbb{P}(\zeta_i=1)=\mathbb{P}(\zeta_i=-1)=\frac{1}{2}$ for all i (justified by Theorem 1.10), the approximating sequence reads

$$(\mathcal{G}^{H-1/2}W_n)(t_i):=\frac{1}{\sqrt{n}}\sum_{k=1}^{i}\left(t_i-t_{k-1}\right)^{H-1/2}\zeta_k,\quad\text{for } i=0,\dots,n.$$

Figure 1 shows a fractional binomial tree structure for $H=0.75$ and $H=0.1$. Despite being symmetric, such trees cannot be recombining due to the (non-Markovian) path-dependent nature of the process. It might be possible, in principle, to modify the tree at each step to make it recombining, following the procedure developed in [2] for Markovian stochastic volatility models. It is not so straightforward though, and requires a dedicated thorough analysis which we leave for future research.

3.3. **Monte-Carlo.** Theorem 1.10 introduces the theoretical foundations of Monte-Carlo methods (in particular for path-dependent options) for rough volatility models. In this section we give a general and easy-to-understand recipe to implement the class of rough volatility models (1.3). For the numerical recipe to be as general as possible, we shall consider the general time partition $\mathcal{T}:=\{t_i=\frac{iT}{n}\}_{i=0,\dots,n}$ on $[0,T]$ with $T>0$.

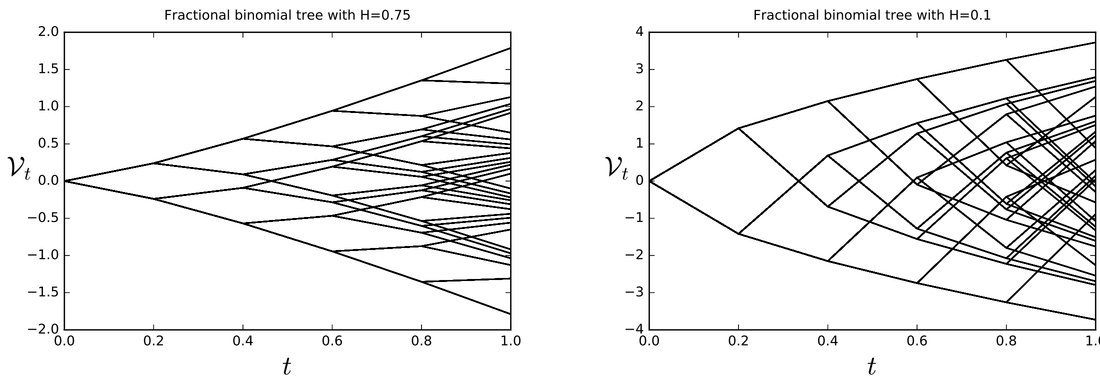

FIGURE 1. Binomial tree for the Riemann-Liouville fractional Brownian motion with $n = 5$ discretisation points for $H = 0.75$ (left) and $H = 0.1$ (right).

Algorithm 3.3 (Simulation of rough volatility models)**.**

(1) Simulate two $\mathcal{N}(0,1)$ matrices $\{\xi_{j,i}\}_{\substack{j=1,\dots,M\\ i=1,\dots,n}}$ and $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M\\ i=1,\dots,n}}$ with $\mathrm{corr}(\xi_{j,i},\zeta_{j,i}) = \rho$;

(2) simulate M paths of Y_n via[1]

$$Y_n^j(t_i) = \frac{T}{n}\sum_{k=1}^{i} b(Y_n^j(t_{k-1})) + \frac{T}{\sqrt{n}}\sum_{k=1}^{i} a\left(Y_n^j(t_{k-1})\right)\zeta_{j,k}, \quad i = 1,\dots,n \text{ and } j = 1,\dots,M,$$

and also compute

$$\Delta Y_n^j(t_i) := Y_n^j(t_i) - Y_n^j(t_{i-1}), \quad i = 1,\dots,n \text{ and } j = 1,\dots,M,$$

(3) Simulate M paths of the fractional driving process $((\mathcal{G}^\alpha Y_n)(t))_{t\in\mathcal{T}}$ using

$$(\mathcal{G}^\alpha Y_n)^j\,(t_i) := \sum_{k=1}^{i} g(t_{i-k+1})\Delta Y_n^j(t_k) = \sum_{k=1}^{i} g(t_k)\Delta Y_n^j(t_{i-k+1}), \quad i = 1,\dots,n \text{ and } j = 1,\dots,M.$$

The complexity of this step is in general of order $\mathcal{O}(n^2)$ (see Appendix B for details). However, this step is easily implemented using discrete convolution with complexity $\mathcal{O}(n\log n)$ (see Algorithm B.4 in Appendix B for details in the implementation). With the vectors $\mathfrak{g} := (g(t_i))_{i=1,\dots,n}$ and $\Delta Y_n^j := (\Delta Y_n^j(t_i))_{i=1,\dots,n}$ for $j = 1,\dots,M$, we can write $(\mathcal{G}^\alpha Y_n)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \Delta Y_n^j)$, for $j = 1,\dots,M$, where $*$ represents the discrete convolution operator.

(4) Use the forward Euler scheme to simulate the log-stock process, for all $i = 1,\dots,n$, $j = 1,\dots,M$, as

$$X^j(t_i) = X^j(t_{i-1}) - \frac{1}{2}\frac{T}{n}\Phi\left(\mathcal{G}^\alpha Y_n\right)^j(t_{i-1}) + \sqrt{\frac{T}{n}}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)^j(t_{i-1})}\,\xi_{j,i}.$$

Remark 3.4.

- When $Y = W$, we may skip step (2) and replace $\Delta Y_n^j(t_i)$ by $\sqrt{T/n}\zeta_{i,j}$ on step (3).
- Step (3) may be replaced by the Hybrid scheme algorithm [14] only when $Y = W$.

Antithetic variates in Algorithm 3.3 are easy to implement as it suffices to consider the uncorrelated random vectors $\zeta_j := (\zeta_{j,1},\zeta_{j,2},\dots,\zeta_{j,n})$ and $\xi_j := (\xi_{j,1},\xi_{j,2},\dots,\xi_{j,n})$, for $j = 1,\dots,M$. Then $(\rho\xi_j + \overline{\rho}\zeta_j, \xi_j)$, $(\rho\xi_j -$

[1]Here, $Y_n^j(t_i)$ denotes the j-th path Y_n evaluated at the time point t_i, which is different from the notation Y_n^j in the theoretical framework above, but should not create any confusion.

$\overline{\rho}\zeta_j, \xi_j)$, $(-\rho\xi_j - \overline{\rho}\zeta_j, -\xi_j)$ and $(-\rho\xi_j + \overline{\rho}\zeta_j, -\xi_j)$, for $j = 1, \dots, M$, constitute the antithetic variates, which significantly improves the performance of the Algorithm 3.3 by reducing memory requirements, reducing variance and accelerating execution by exploiting symmetry of the antithetic random variables.

3.3.1. *Enhancing performance.* A standard practice in Monte-Carlo simulation is to match moments of the approximating sequence with the target process. In particular, when the process is Gaussian, matching first and second moments suffices. We only illustrate this approximation for Brownian motion: the left-point approximation (2.4) (with $Y = W$) may be modified to match moments as

$$(\mathcal{G}^\alpha W)(t_i) \approx \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{i} g\left(t_k^*\right)\zeta_k, \qquad \text{for } i = 0, \dots, n, \tag{3.3}$$

where t_k^* is chosen optimally. Since the kernel $g(\cdot)$ is deterministic, there is no confusion with the Stratonovich stochastic integral, and the resulting approximation will always converge to the Itô integral. The first two moments of $\mathcal{G}^\alpha W$ read

$$\mathbb{E}\left[\left(\mathcal{G}^\alpha W\right)(t)\right] = 0 \qquad \text{and} \qquad \mathbb{V}\left[\left(\mathcal{G}^\alpha W\right)(t)\right] = \int_0^t g(t-s)^2 \mathrm{d}s.$$

The first moment of the approximating sequence (3.3) is always zero, and the second moment reads

$$\mathbb{V}\left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{j-1} g\left(t_k^*\right)\zeta_k\right) = \frac{1}{n}\sum_{k=1}^{j-1} g\left(t_k^*\right)^2.$$

Equating the theoretical and approximating quantities we obtain $\frac{1}{n} g(t_k^*)^2 \mathrm{d}s = \int_{t_{k-1}}^{t_k} g(t-s)^2 \mathrm{d}s$ for $k = 1, \dots, n$, so that the optimal evaluation point can be computed as

$$g(t_k^*) = \sqrt{n \int_{t_{k-1}}^{t_k} g(t-s)^2 \mathrm{d}s}, \qquad \text{for any } k = 1, \dots, n. \tag{3.4}$$

With the optimal evaluation point the scheme is still a convolution so that Algorithm B.4 in Appendix B can still be used for faster computations. In the Riemann-Liouville fractional Brownian motion case, $g(u) = u^{H-1/2}$, and the optimal point can be computed in closed form as

$$t_k^* = \left(\frac{n}{2H}\left[\left(t - t_{k-1}\right)^{2H} - \left(t - t_k\right)^{2H}\right]\right)^{1/(2H-1)}, \qquad \text{for each } k = 1, \dots, n.$$

Proposition 3.5. *The moment matching sequence* $(\widehat{\mathcal{G}}^\alpha W_n)$ *defined as*

$$\widehat{\mathcal{G}}^\alpha W_n(t) := \widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(\widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - \widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right), t \in \mathbb{I}, \tag{3.5}$$

where

$$\widehat{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{i} \sqrt{n \int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s}\,\xi_k. \tag{3.6}$$

with (ξ_k) *an iid family of centered sub-Gaussian random variables with* $\mathbb{E}[\xi_k^2] = \frac{1}{n}$ *(namely* $\mathbb{P}(|\xi_k| > x) \le C\mathrm{e}^{-vx^2}$ *for all* $x > 0$ *and some* $C, v > 0$*). Then, convergence to* $\mathcal{G}^\alpha W$ *holds in* $\left(\mathcal{C}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ *for* $\alpha \in \mathfrak{R}^{1/2}$.

Proof. Finite-dimensional convergence follows from the Central Limit Theorem as the target process is centered Gaussian, thus convergence of the covariance matrix ensures finite-dimensional convergence. It then suffices to prove that the approximating sequence is tight in the desired space, which, in view of Theorem 2.4, can be deduced by establishing the control

$$\mathbb{E}\left[\left|\widetilde{\mathcal{G}}^{\alpha}W_n(t_i) - \widetilde{\mathcal{G}}^{\alpha}W_n(t_j)\right|^{2p}\right] \leq C|t_i - t_j|^{2p\alpha+p},$$

for all $t_i, t_j \in \mathcal{T}_n$, for $p \geq 1$ and some constant $C \geq 0$. We have

$$\widetilde{\sigma}^2 := \mathbb{E}\left[\left(\widehat{\mathcal{G}}^{\alpha}W_n(t_i) - \widehat{\mathcal{G}}^{\alpha}W_n(t_j)\right)^2\right] = \mathbb{E}\left[\left(\widehat{\mathcal{G}}^{\alpha}W_n(t_i)\right)^2\right] + \mathbb{E}\left[\left(\widehat{\mathcal{G}}^{\alpha}W_n(t_j)\right)^2\right] - 2\mathbb{E}\left[\widehat{\mathcal{G}}^{\alpha}W_n(t_i)\widehat{\mathcal{G}}^{\alpha}W_n(t_j)\right].$$

We note that

$$\mathbb{E}\left[\left(\widehat{\mathcal{G}}^{\alpha}W_n(t_i)\right)^2\right] = \sum_{k=1}^{i}\left(\sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2\mathrm{d}s}\right)^2 = \int_0^{t_i} g(t_i - s)^2\mathrm{d}s = \mathbb{E}\left[\left(\mathcal{G}^{\alpha}W(t_i)\right)^2\right],$$

and by Cauchy-Schwarz we also have

$$\begin{aligned}\mathbb{E}\left[\widehat{\mathcal{G}}^{\alpha}W_n(t_i)\widehat{\mathcal{G}}^{\alpha}W_n(t_j)\right] &= \sum_{k=1}^{t_i\wedge t_j}\sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2\mathrm{d}s\int_{t_{k-1}}^{t_k} g(t_j - s)^2\mathrm{d}s} \\ &\geq \int_0^{t_i\wedge t_j} g(t_i - s)g(t_j - s)\mathrm{d}s = \mathbb{E}\left[\mathcal{G}^{\alpha}W(t_i)\mathcal{G}^{\alpha}W(t_j)\right].\end{aligned}$$

We then obtain $\widetilde{\sigma}^2 \leq \mathbb{E}\left[\left|\mathcal{G}^{\alpha}W(t_i) - \mathcal{G}^{\alpha}W(t_j)\right|^2\right]$. Finally, $\widetilde{\mathcal{G}}^{\alpha}W_n(t_i) - \widetilde{\mathcal{G}}^{\alpha}W(t_j)$ is sub-Gaussian as a linear combination of sub-Gaussian random variables, and the Gaussian moment inequality [19, Theorem 2.1] with the variance estimate $\widetilde{\sigma}^2$ yields

$$\mathbb{E}\left[\left|\widehat{\mathcal{G}}^{\alpha}W_n(t_i) - \widehat{\mathcal{G}}^{\alpha}W(t_j)\right|^{2p}\right] \leq \mathbb{E}\left[\left|\mathcal{G}^{\alpha}W(t_i) - \mathcal{G}^{\alpha}W(t_j)\right|^{2p}\right] \leq \widetilde{C}|t_i - t_j|^{2p\alpha+p}.$$

□

3.3.2. *Reducing variance.* As Bayer, Friz and Gatheral [11] and Bennedsen, Lunde and Pakkanen [14] pointed out, a major drawback in simulating rough volatility models is the very high variance of the estimators, so that a large number of simulations are needed to produce a decent price estimate. Nevertheless, the rDonsker scheme admits a very simple conditional expectation technique which reduces both memory requirements and variance while also admitting antithetic variates. This approach is best suited for calibrating European type options. We consider $\mathcal{F}_t^B = \sigma(B_s : s \leq t)$ and $\mathcal{F}_t^W = \sigma(W_s : s \leq t)$ the natural filtrations generated by the Brownian motions B and W. In particular the conditional variance process $V_t|\mathcal{F}_t^W$ is deterministic. As discussed by Romano and Touzi [78], and recently adapted to the rBergomi case by McCrickerd and Pakkanen [67], we can decompose the stock price process as

$$\mathrm{e}^{X_t} = \mathcal{E}\left(\rho\int_0^t \sqrt{\Phi\left(\mathcal{G}^{\alpha}Y\right)(s)}\mathrm{d}W_s\right)\mathcal{E}\left(\sqrt{1-\rho^2}\int_0^t \sqrt{\Phi\left(\mathcal{G}^{\alpha}Y\right)(s)}\mathrm{d}W_s^{\perp}\right) =: \mathrm{e}^{X_t^{||}}\mathrm{e}^{X_t^{\perp}},$$

and notice that

$$X_t|(\mathcal{F}_t^W \wedge \mathcal{F}_0^B) \sim \mathcal{N}\left(X_t^{||} - (1-\rho^2)\int_0^t \Phi\left(\mathcal{G}^{\alpha}Y\right)(s)\mathrm{d}s, (1-\rho^2)\int_0^t \Phi\left(\mathcal{G}^{\alpha}Y\right)(s)\mathrm{d}s\right).$$

Thus $\exp(X_t)$ becomes log-normal and the Black-Scholes closed-form formulae are valid here (European, Barrier options, maximum, etc.). The advantage of this approach is that the orthogonal Brownian motion $W^{\perp}$ is completely unnecessary for the simulation, hence the generation of random numbers is reduced to a half, yielding proportional memory saving. Not only this, but this simple trick also reduces the variance of the

Monte-Carlo estimate, hence fewer simulations are needed to obtain the same precision. We present a simple algorithm to implement the rDonsker with conditional expectation and assuming that $Y = W$.

Algorithm 3.6 (Simulation of rough volatility models with Brownian drivers)**.** Consider the equidistant grid $\mathcal{T}$.

(1) Draw a random matrix $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M/2\\ i=1,\dots,n}}$ with unit variance, and create antithetic variates $\{-\zeta_{j,i}\}_{\substack{j=1,\dots,M/2\\ i=1,\dots,n}}$;

(2) Simulate M paths of the fractional driving process $\mathcal{G}^\alpha W$ using discrete convolution (see Algorithm B.4 in Appendix B for details in the implementation):
$$(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \zeta_j), \quad j = 1, \dots, M,$$
and store in memory $(1-\rho^2)\int_0^T (\mathcal{G}^\alpha W)^j(s)\mathrm{d}s \approx (1-\rho^2)\frac{T}{n}\sum_{k=0}^{n-1}(\mathcal{G}^\alpha W)^j(t_k) =: \Sigma^j$ for each $j = 1, \dots, M$;

(3) use the forward Euler scheme to simulate the log-stock process, for each $i = 1, \dots, n$, $j = 1, \dots, M$, as
$$X^j(t_i) = X^j(t_{i-1}) - \frac{\rho^2}{2}\frac{T}{n}\Phi\left(\mathcal{G}^\alpha W\right)^j(t_{i-1}) + \rho\sqrt{\frac{T}{n}}\sqrt{\Phi\left(\mathcal{G}^\alpha W\right)^j(t_{i-1})}\zeta_{j,i};$$

(4) Finally, we may compute any option using the Black-Scholes formula. For instance a Call option with strike K and maturity $T \in \mathbb{I}$ would be given by $C^j(K) = \exp(X^j(T))\mathcal{N}(d_1^j) - K\mathcal{N}(d_2^j)$ for $j = 1, \dots, M$, where $\Sigma^j = Var(X^j(T))$, $d_1^j := \frac{1}{\sqrt{\Sigma^j}}(X^j(T) - \log(K) + \frac{1}{2}\Sigma^j)$ and $d_2^j = d_1^j - \sqrt{\Sigma^j}$. Thus, the output of the model would be $C(K) = \frac{1}{M}\sum_{k=1}^M C^j(K)$.

The algorithm is easily adapted to general diffusions Y as drivers of the volatility (Algorithm 3.3(2)). Algorithm 3.3 is obviously faster than 3.6, especially when using control variates. Nevertheless, with the same number of paths, Algorithm 3.6 remarkably reduces the Monte-Carlo variance, meaning in turn that fewer simulations are needed, making it very competitive for calibration.

3.4. **Numerical example: Rough Bergomi model.** Figures 2-5 perform a numerical analysis of the Monte Carlo convergence as a function of n. We observe that the lower the H, the larger n needs to be to achieve convergence. However, we also observe that for the Cholesky, rDonsker (naive and moment match) and Hybrid schemes and $H \geq 0.1$, with $N = 252$ we already achieve a precision of order 10^{-4}, which is equivalent to a basis point in financial terms. For $H < 0.1$ we might require n larger than 252, if precision is required beyond 10^{-4}. We also observe in Figure 5 that the naive rDonsker approximation converges extremely slow for small H. Additionally, Figures 6-11 measure the price estimations compared to the Cholesky method which is taken as benchmark. The Hybrid scheme tends to be closer to this benchmark especially for $H < 0.1$. When $H \geq 0.1$ for both the Hybrid scheme and rDonsker moment-match we observe an error less than 10^{-4} for $n \geq 252$. It is noteworthy to mention that the naive rDonsker scheme has substantially worse convergence (at least an order of magnitude) than the other methods. We note that the black lines in all figures represent the 99% Monte Carlo standard deviations, hence errors below that threshold should be interpreted as noise.

3.5. **Speed benchmark against Markovian stochastic volatility models.** In this section we benchmark the speed of the rDonsker scheme against the Hybrid scheme and a classical Markovian stochastic volatility model using 10^5 simulations and averaging the speeds over 10 trials. For the former we simulate the rBergomi model [11], whereas for the latter we use the classical Bergomi [16] model using a forward Euler scheme in both volatility and stock price. All three schemes are implemented in `Cython` to make the comparison fair, and to obtain speeds comparable to `C++`. Figure 12 shows that rDonsker is about twice slower than the Markovian

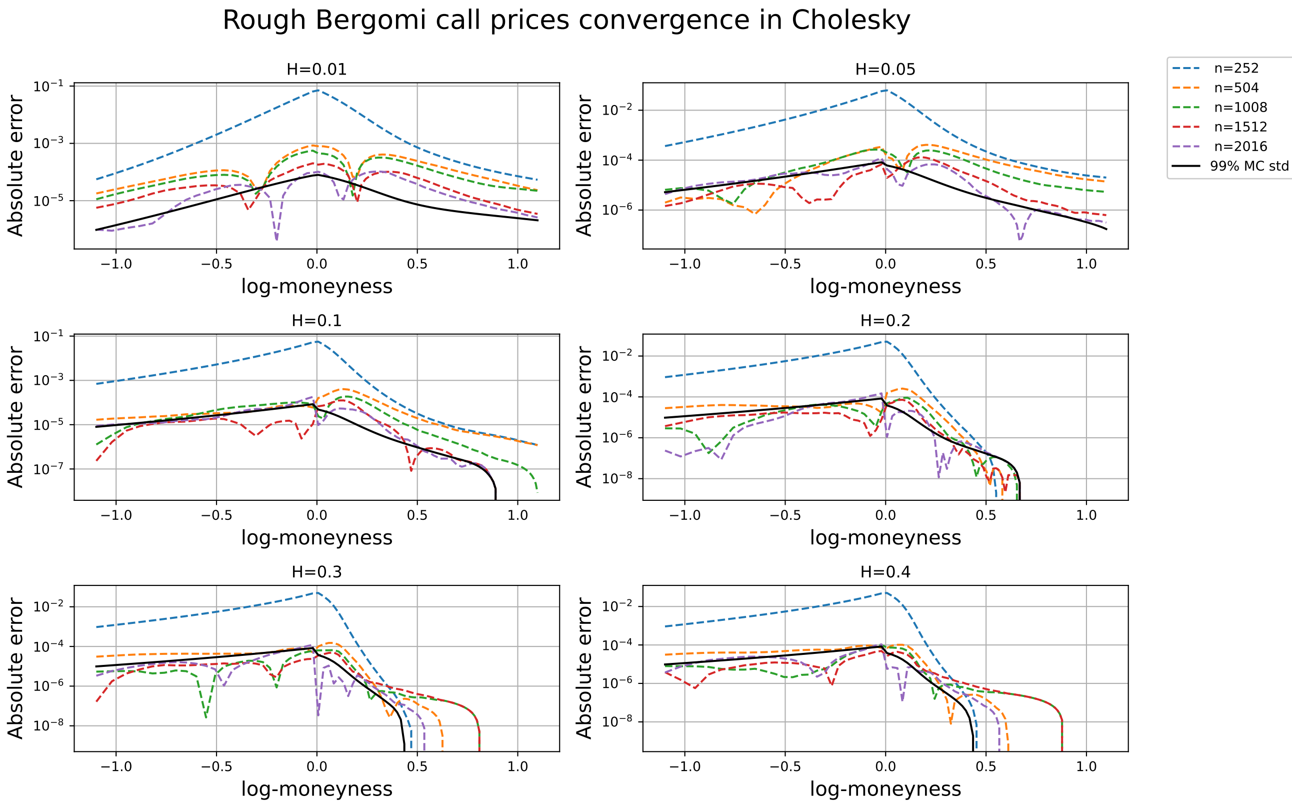

FIGURE 2. Rough Bergomi Call option price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

case whereas the Hybrid scheme is approximately 2.5 times slower, which is expected from the complexities of both schemes. However, it is remarkable that the $\mathcal{O}(n \log n)$ complexity of the FFT stays almost constant with the grid size n and the computational time grows almost linearly as in the Markovian case. We presume that this is the case since $n \ll 10000$ is relatively small. Figure 12 also shows that rough volatility models can be implemented very efficiently and are not particularly slower than classical stochastic volatility models.

3.6. **Implementation guidelines and conclusion.** The numerical analysis above suggests some guidelines to implement rough volatility models driven by $\mathcal{TBSS}$ processes of the form $\mathcal{G}^{H-1/2}Y$, for some Itô diffusion Y:

$H > 0.1$	$H \in [0.05, 0.1]$	$H < 0.05$
rDonsker	choice depends on error sensitivity	Hybrid scheme

Regarding empirical estimates, Gatheral, Jaisson and Rosenbaum [41] suggest that $H \approx 0.15$. Bennedsen, Lunde and Pakkanen [15] give an exhaustive analysis of more than 2000 equities for which $H \in [0.05, 0.2]$. On the pricing side, Bayer, Friz and Gatheral [11] and Jacquier, Martini and Muguruza [53] found that calibration

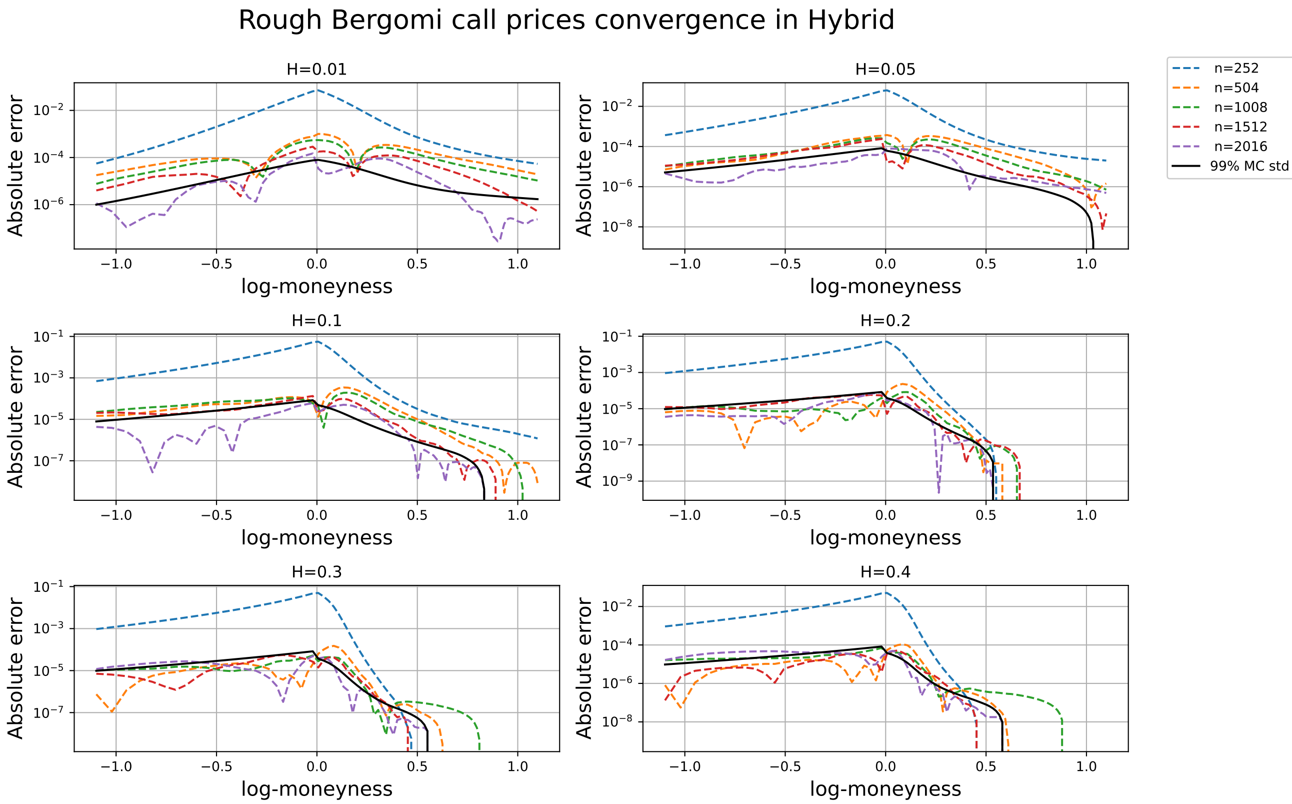

FIGURE 3. Rough Bergomi Call option price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

routines yield $H \in [0.05, 0.10]$. Finally, Livieri, Mouti, Pallavicini and Rosenbaum [63] found evidence in options data that $H \approx 0.3$. Despite the diverse ranges found so far, there is a common agreement that $H < 1/2$.

Remark 3.7. The rough Heston model presented by Guennoun, Jacquier, Roome and Shi [43] is out of the scope of the Hybrid scheme. Moreover, any process of the form $\mathcal{G}^\alpha Y$, for some Itô diffusion Y under Assumptions 1.3 is, in general, out of the scope of the Hybrid scheme. This only leaves the choice of using the rDonsker scheme, for which reasonable accuracy is obtained at least for Hölder regularities greater than 0.05.

3.7. **Bushy trees and binomial markets.** Binomial trees have attracted a lot of attention from both academics and practitioners, as their apparent simplicity provides easy intuition about the dynamics of a given asset. Not only this, but they are by construction arbitrage free and allow to price path-dependent options, together with their hedging strategy. In particular, early exercise options, in particular Bermudan or American options, are usually priced using trees, as opposed to Monte-Carlo methods. The convergence stated in Theorem 1.10 lays the theoretical foundations to construct fractional binomial trees (note that Bernoulli random variables satisfy the conditions of the theorem). Figure 1 already showed binomial trees for fractional Brownian motion, but we ultimately need trees describing the dynamics of the stock price.

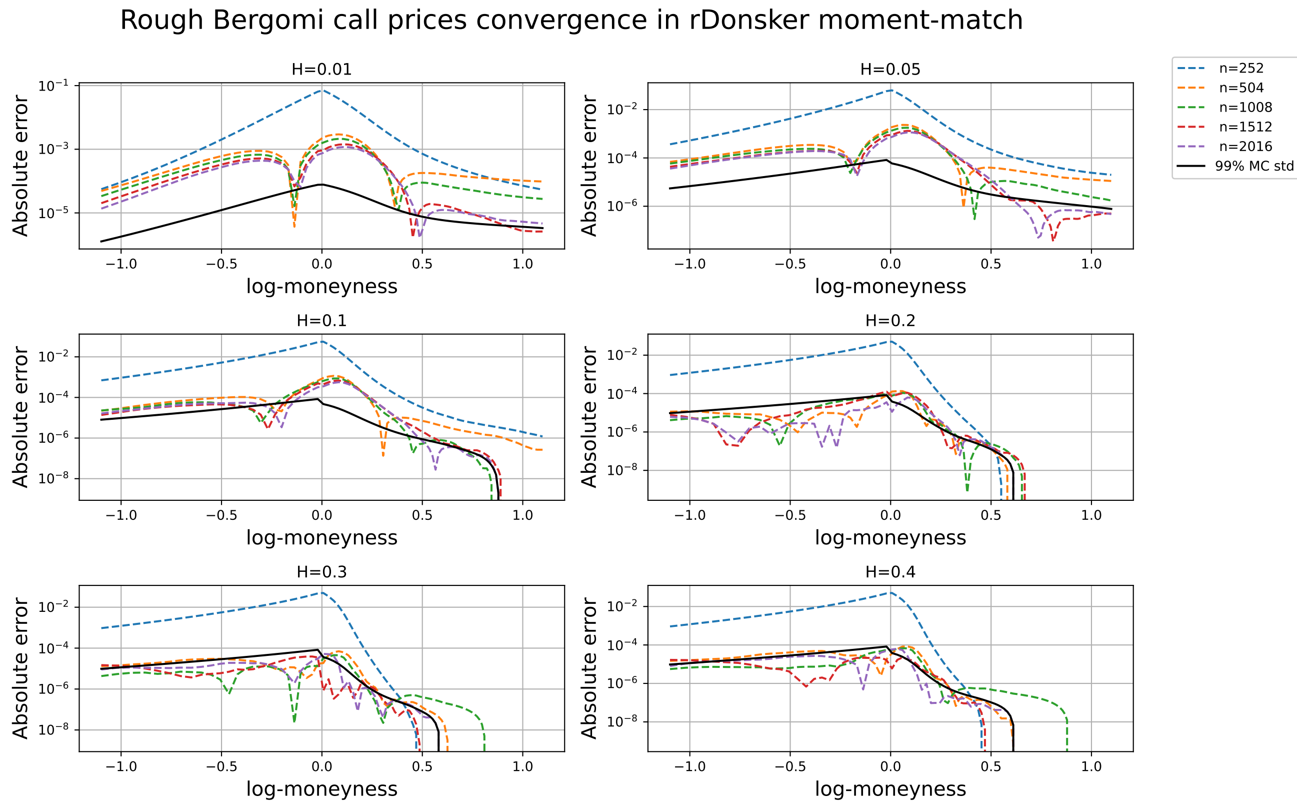

Figure 4. Rough Bergomi Call option price convergence using rDonsker with moment-matching and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, with n the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

3.7.1. *A binary market.* We invoke Theorem 1.10 with the independent sequences $\{\zeta_i\}_{i=1}^n$, $\{\zeta_i^\perp\}_{i=1}^n$ such that $\mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i^\perp = 1) = \mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i^\perp = -1) = \frac{1}{2}$ for all i. We further define, on $\mathcal{T}$, for any $i = 1, \dots, n$,

$$B_n(t_i) = \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \left(\rho\zeta_k + \overline{\rho}\zeta_k^\perp\right),$$

$$Y_n(t_i) = \frac{T}{n} \sum_{k=1}^{i} b\left(Y_n(t_{k-1})\right) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sigma\left(Y_n(t_{k-1})\right)\zeta_k,$$

the approximating sequences to B and Y in (1.3). The approximation for X is then given by

$$X_n(t_i) = X_n(t_{i-1}) - \frac{1}{2}\frac{T}{n} \sum_{k=1}^{i} \Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k)}\left(\rho\zeta_k + \overline{\rho}\zeta_k^\perp\right).$$

In order to construct the tree we have to consider all possible permutations of the random vectors $\{\zeta_i\}$ and $\{\zeta_i^\perp\}$. Since each random variable only takes two values, this adds up to 4^n possible combinations, hence the 'bushy tree' terminology. When $\rho \in \{-1, 1\}$, the magnitude is reduced to 2^n.

3.8. **American options in rough volatility models.** There is so far no available scheme for American options (or any early-exercise options for that matter) under rough volatility models, but the fractional trees

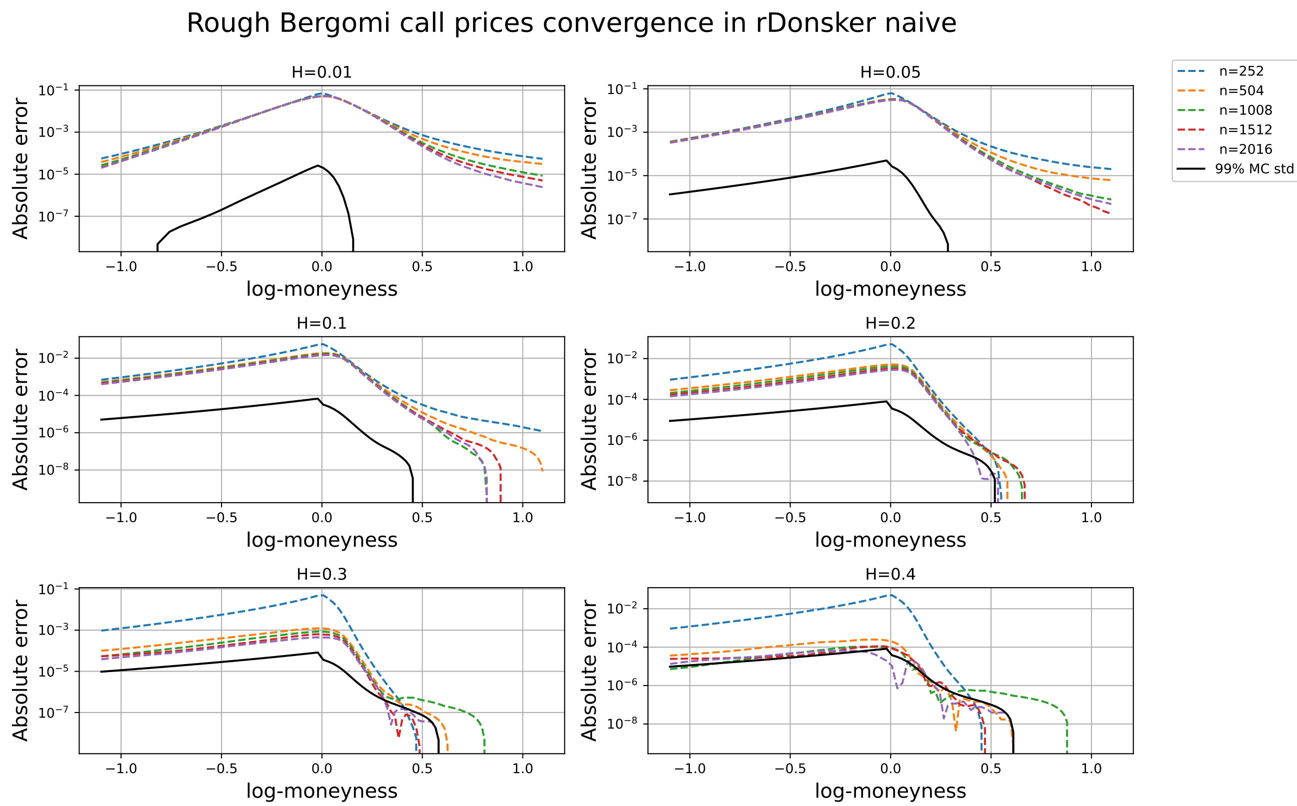

FIGURE 5. Rough Bergomi Call option price convergence using rDonsker method with left point Euler and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

constructed above provide a framework to do so. In the Black-Scholes model, American options can be priced using binomial trees by backward induction. A key ingredient is the Snell envelope [80] and the following representation by El Karoui [32] ($\widetilde{\mathbb{I}}$ denotes the set of stopping times with values in $\mathbb{I}$):

Definition 3.8. Let $(X_t)_{t\in\mathbb{I}}$ be an $(\mathcal{F}_t)_{t\in\mathbb{I}}$ adapted process, and $\tau \in \widetilde{\mathbb{I}}$. The Snell envelope $\mathcal{J}$ of X is defined as $\mathcal{J}(X)(t) := \operatorname{ess\,sup}_{\tau\in\widetilde{\mathbb{I}}} \mathbb{E}(X_\tau|\mathcal{F}_t)$ for all $t \in \mathbb{I}$.

In plain words, the Snell envelope of X is the smallest supermartingale that dominates it. Strictly speaking, it is necessary for X_τ to be uniformly integrable for any $\tau \in \widetilde{\mathbb{I}}$. Following [55], an American option is nothing else than the smallest supermartingale dominating its European counterpart:

Definition 3.9. Let $C_t^e(k,T)$ and $P_t^e(k,T)$ denote European Call and Put prices at time t, with log-strike k and maturity T. Then the American counterparts, $C_t^a(k,T)$ and $P_t^a(k,T)$, are given by

$$C_t^a(k,T) = \mathcal{J}(C^e(k,T))(t) \qquad \text{and} \qquad P_t^a(k,T) = \mathcal{J}(P^e(k,T))(t).$$

Preservation of weak convergence under the Snell envelope map is due to Mulinacci and Pratelli [69], who proved that convergence takes place in the Skorokhod topology only if the Snell envelope is continuous. In our setting, the scheme for American options is fully justified by the following theorem:

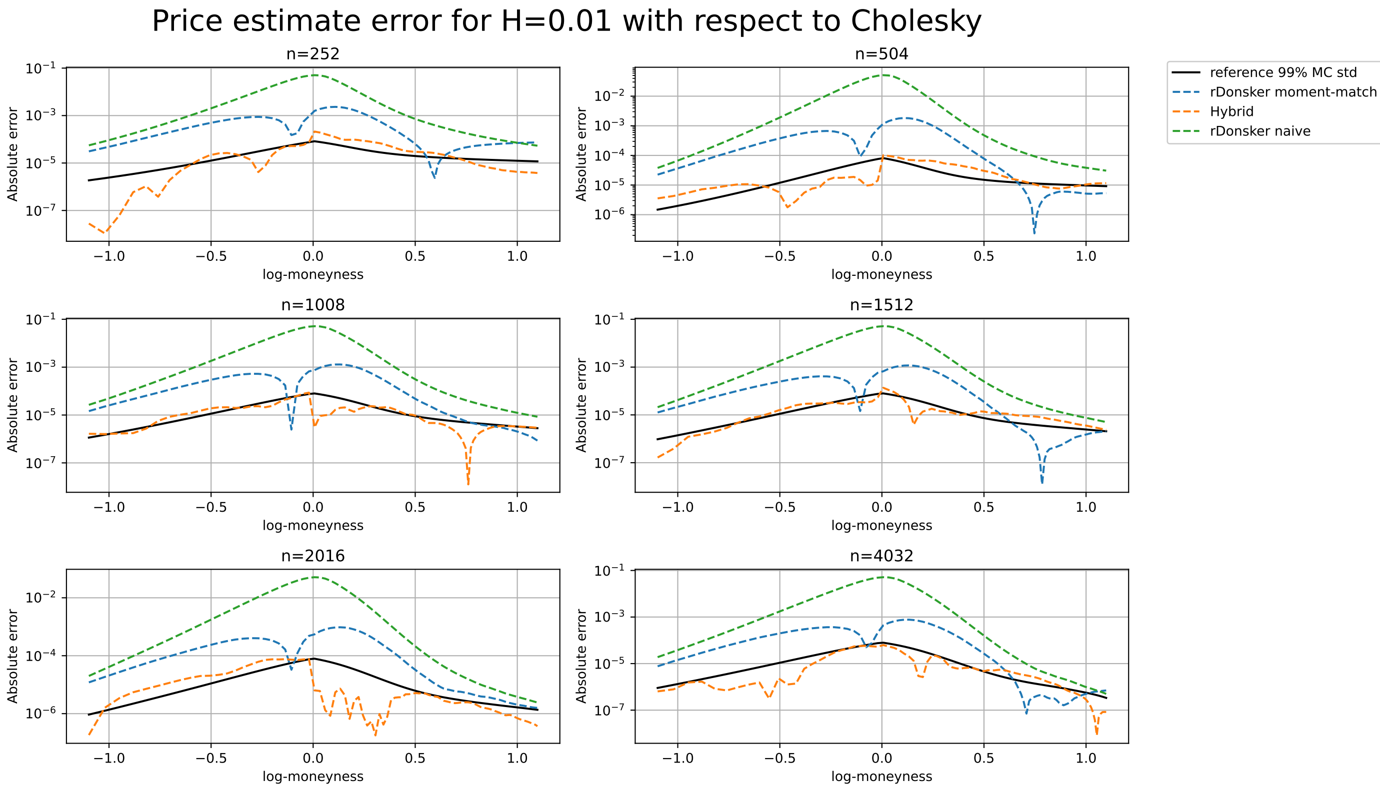

FIGURE 6. Rough Bergomi Call option price comparison with $H = 0.01, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

Theorem 3.10. *For V in* (1.3)*, if* e^X *is a true martingale then* $\left(\mathcal{J}(\mathrm{e}^{X_n})\right)_{n\geq 1}$ *converges weakly to* $\mathcal{J}\left(\mathrm{e}^X\right)$ *in the Skorokhod topology* $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$.

Proof. Since the sequence $(X_n)_{n\geq 1}$ converges weakly to X in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$, for X in (1.3), the theorem follows from the Continuous Mapping Theorem if we can show that $\mathcal{J}$ is continuous. El Karoui proved in [32, Chapter 2.14] that the Snell envelope of an optional process, uniformly integrable for all stopping times $\tau \in \widetilde{\mathbb{I}}$, is continuous. To prove the proposition, we therefore only need to check uniform integrability of the stock price e^X. As $\mathbb{I}$ is a finite time horizon, Doob's optimal stopping theorem for martingales gives $\mathrm{e}^{X_t} = \mathbb{E}[\mathrm{e}^{X_1}|\mathcal{F}_t]$ for all $t \in \mathbb{I}$, thus e^X on $\mathbb{I}$ is uniformly integrable and the result follows. $\square$

Mulinacci and Pratelli [69] also gave explicit conditions for weak convergence to be preserved in the Markovian case. It is trivial to see that the pricing of American options in the rough tree scheme coincides with the classical backward induction procedure. We consider continuously compounded interest rate r and dividend yield d.

Algorithm 3.11 (American options in rough volatility models)**.** On the equidistant grid $\mathcal{T}$,

(1) construct the binomial tree using the explicit construction in Section 3.7.1 and obtain $\{S_t^j\}_{t\in\mathcal{T}, j=1,\dots,4^n}$;
(2) the backward recursion for the American with exercise value $h(\cdot)$ is given by $\widetilde{h}_{t_N} := h(S_{t_N})$ and

$$\widetilde{h}_{t_i} := \mathrm{e}^{(d-r)/n}\mathbb{E}\left[\widetilde{h}_{t_{i+1}}|\mathcal{F}_{t_i}\right] \vee h(S_{t_i}), \qquad \text{for } i = N-1, \dots, 0,$$

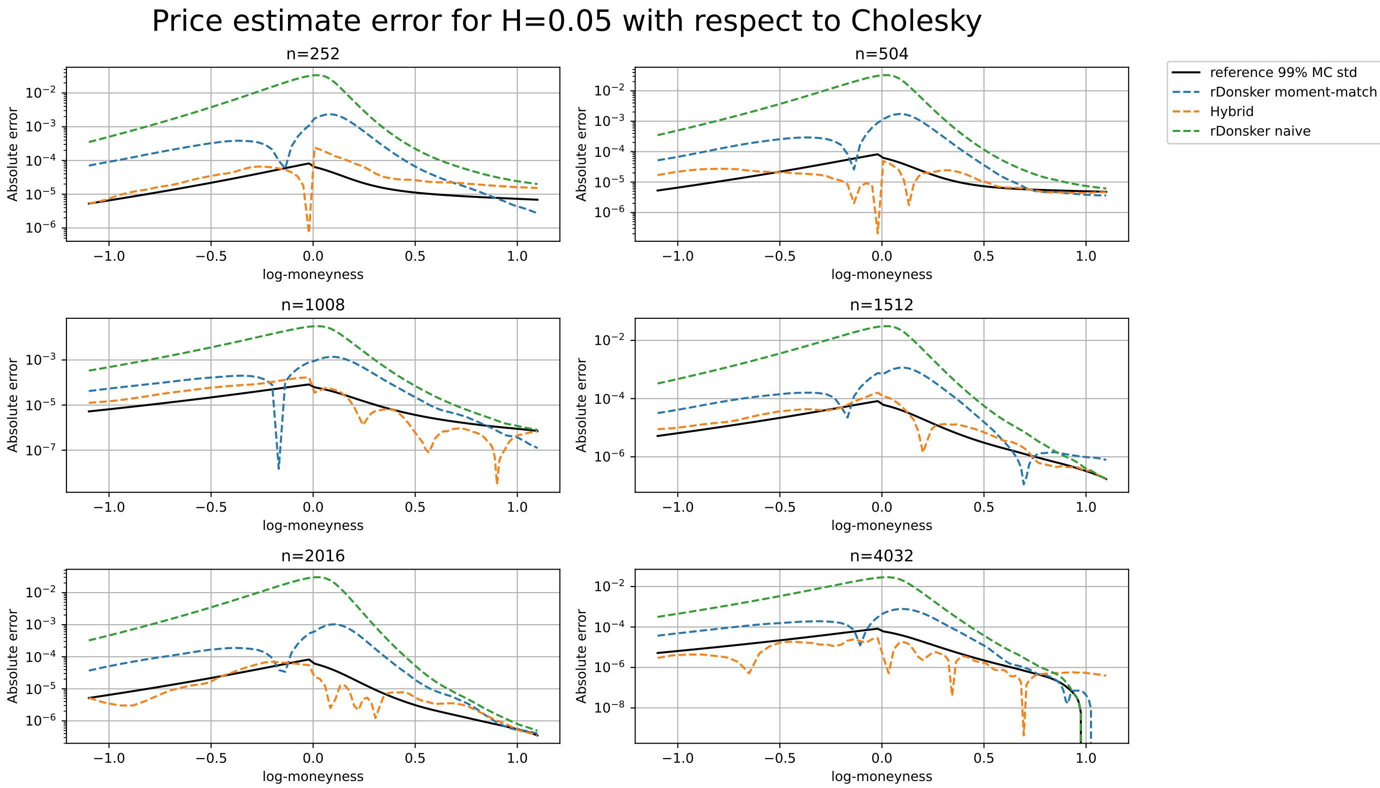

FIGURE 7. Rough Bergomi Call option price comparison with $H = 0.05, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

where $\mathbb{E}[\cdot|\mathcal{F}_{t_i}] = \frac{1}{4}\left(\widetilde{h}^{++}_{t_{i+1}} + \widetilde{h}^{+-}_{t_{i+1}} + \widetilde{h}^{-+}_{t_{i+1}} + \widetilde{h}^{--}_{t_{i+1}}\right)$ and $\widetilde{h}^{\pm\pm}_{t_i}$ represents the outcome $(\zeta_i, \zeta_i^{\perp}) = (\pm 1, \pm 1)$ for the driving binomials, following the construction in Section 3.7.1.

(3) finally, $\widetilde{h}_0$ is the price of the American option at inception of the contract.

The main computational cost of the scheme is the construction of the tree in Step 1. Once the tree is constructed, computing American prices for different options is a fast routine.

3.8.1. *Numerical example: rough Bergomi model.* The rough Bergomi model satisfies the martingale property in Theorem 3.10 (b) for $\rho \leq 0$ (see Gassiat [39]). We construct a rough volatility tree for the rough Bergomi model [11] and check the accuracy of the scheme. Figures 13 and 14 show the fractional trees for different values of H and for $\rho \in \{-1, 1\}$. Both pictures show a markedly different behaviour, but as a common property we observe that as H tends to $1/2$, the tree structure somehow becomes simpler.

3.8.2. *European options.* Figure 15 displays volatility smiles obtained using the tree scheme. Even though the time steps are not sufficient for small H, the fit remarkably improves when $H \geq 0.15$, and always remains inside the 95% confidence interval with respect to the Hybrid scheme. Moreover, the moment-matching approach from Section 3.3.1 shows a superior accuracy when $H \leq 0.1$, but is not sufficiently accurate. In Figure 16 a detailed error analysis corroborates these observations: the relative error is smaller than 3% for $H \geq 0.15$.

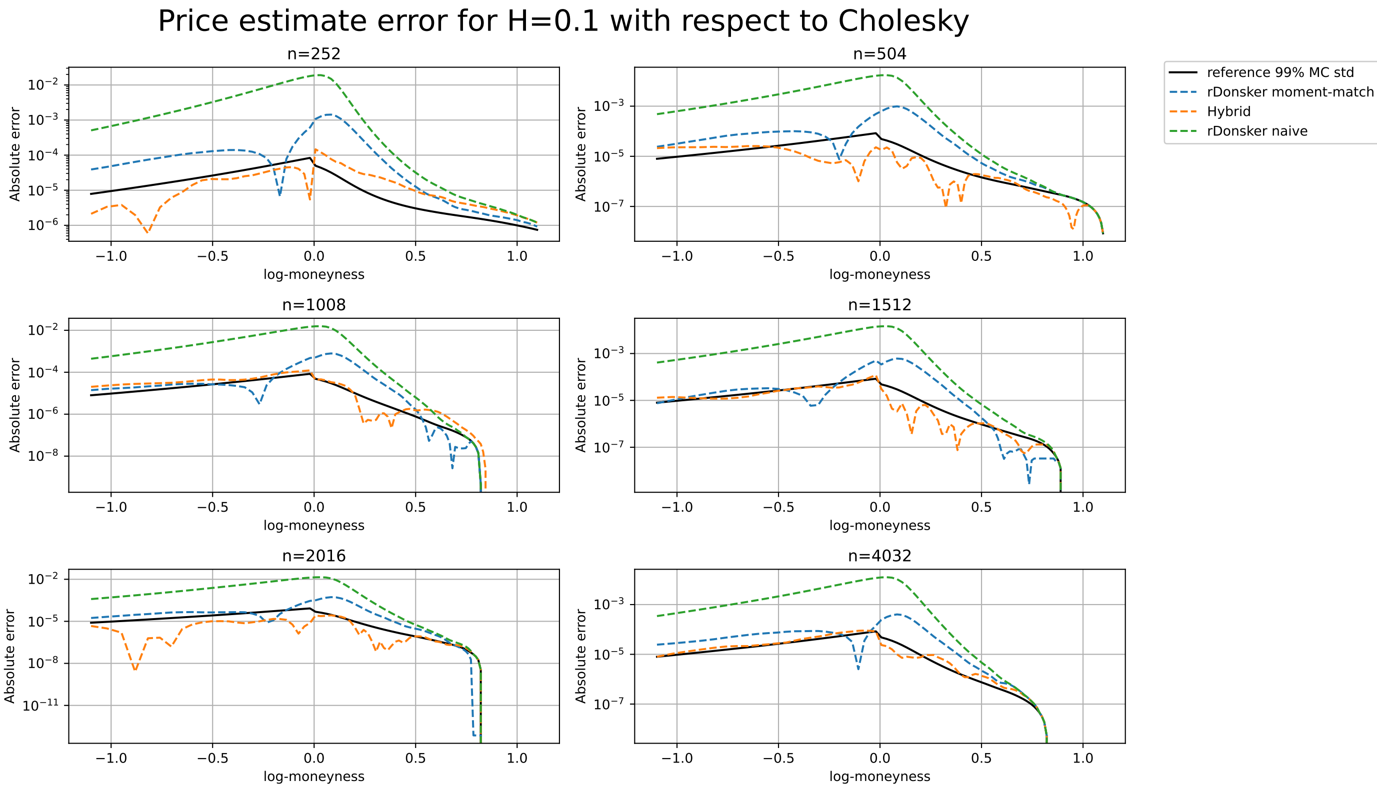

FIGURE 8. Rough Bergomi Call option price comparison with $H = 0.10, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

3.8.3. *American options.* In the context of American options, there is no benchmark to compare our result. However, the accurate results found in the previous section (at least for $H \geq 0.15$) justify the use of trees to price American options. Figure 17 shows the output of American and European Put prices with interest rates equal to $r = 5\%$. Interestingly, the rougher the process (the smaller the H), the larger the difference between in-the-money European and American options.

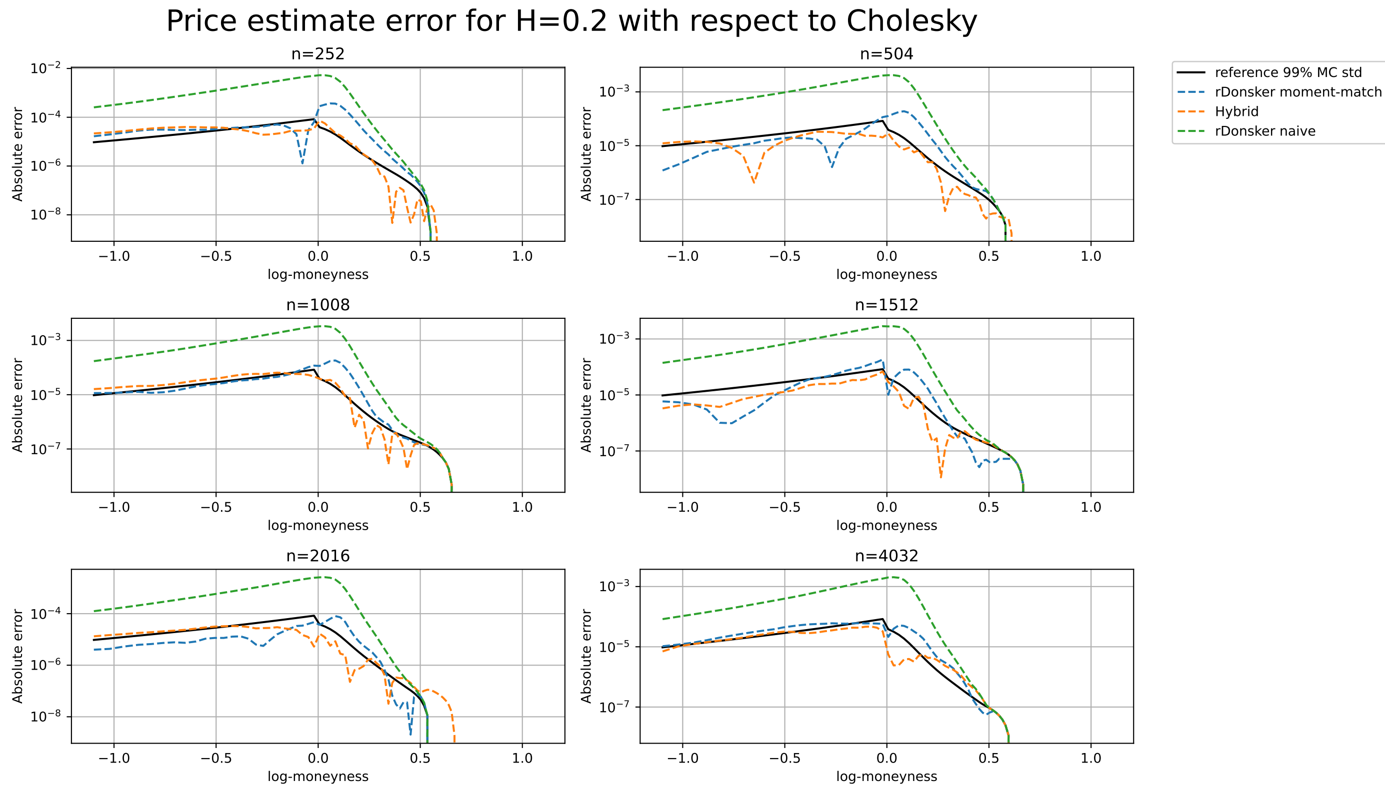

Figure 9. Rough Bergomi Call option price comparison with $H = 0.2, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

Appendix A. Riemann-Liouville operators

We review here fractional operators and their mapping properties. We follow closely the excellent monograph by Samko, Kilbas and Marichev [79], as well as some classical results by Hardy and Littlewood [46]. However, we introduce a modification in their definition, so that the condition $f(0) = 0$ is not necessary as opposed to the original definition in Hardy and Littlewood [46]

A.0.1. *Riemann-Liouville fractional operators.*

Definition A.1. For $\lambda \in (0,1)$, $\alpha \in \mathfrak{R}^\lambda$ the left Riemann-Liouville fractional operator is defined on $\mathcal{C}^\lambda(\mathbb{I})$ as

$$(I^\alpha f)(t) := \begin{cases} \dfrac{1}{\Gamma(\alpha)} \displaystyle\int_0^t \frac{f(s) - f(0)}{(t-s)^{1-\alpha}} \mathrm{d}s, & \text{for } \alpha \in (0, 1-\lambda), \\ \left(\dfrac{\mathrm{d}}{\mathrm{d}t} I^{1+\alpha} f\right)(t) = \dfrac{1}{\Gamma(1+\alpha)} \dfrac{\mathrm{d}}{\mathrm{d}t} \displaystyle\int_0^t (t-s)^\alpha (f(s) - f(0)) \mathrm{d}s, & \text{for } \alpha \in (-\lambda, 0). \end{cases} \tag{A.1}$$

Theorem A.2. *For any $f \in \mathcal{C}^\lambda(\mathbb{I})$, with $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, $I^\alpha f \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. In particular, there exists $C > 0$ such that $|(I^\alpha f)(t)| \leq C t^{\alpha+\lambda}$ for any $t \in \mathbb{I}$.*

Proof. We first consider $\alpha > 0$, then we may easily represent

$$(I^\alpha f)(t) = \frac{1}{\Gamma(\alpha)} \int_0^t \frac{f(u) - f(0)}{(t-u)^{1-\alpha}} \mathrm{d}u.$$

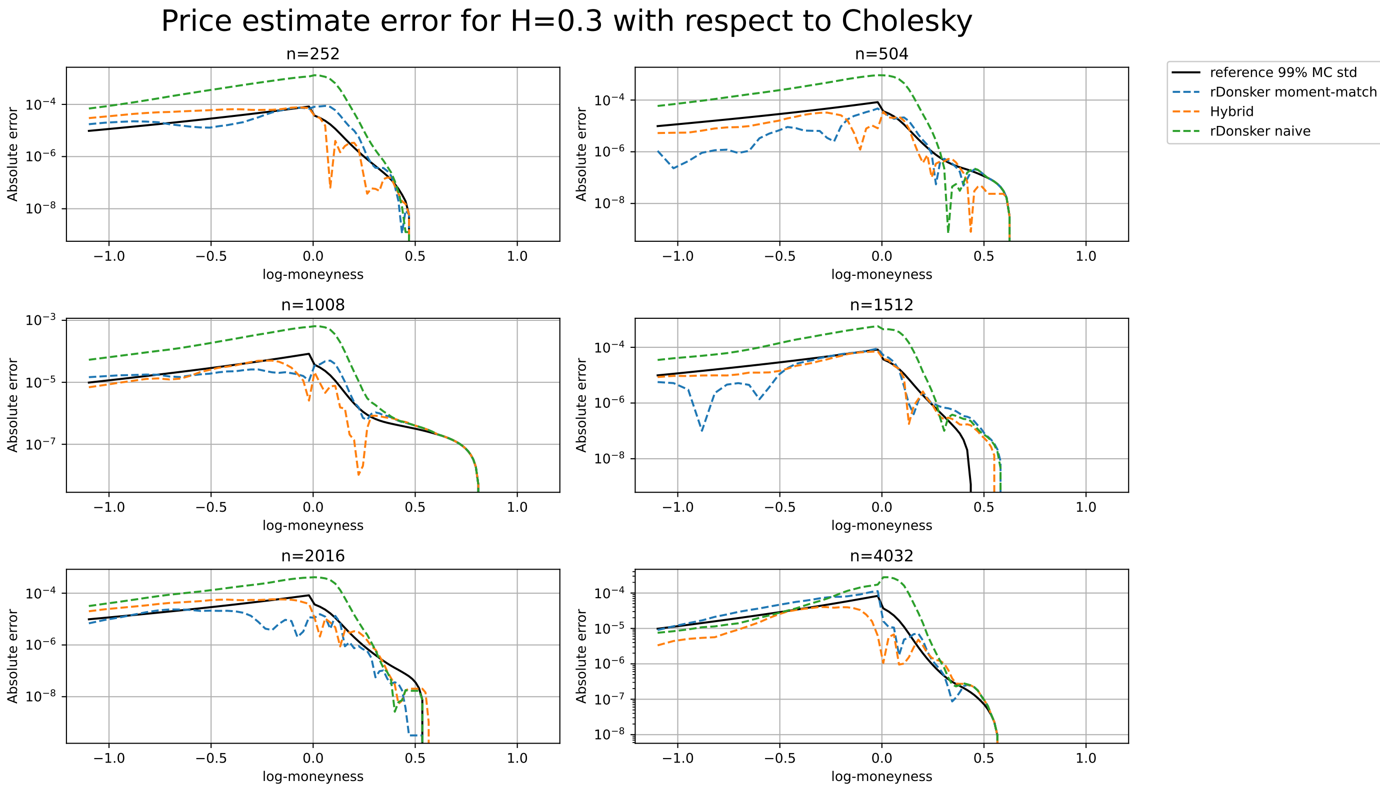

FIGURE 10. Rough Bergomi Call option price comparison with $H = 0.3, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

Since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we obtain $|(I^\alpha f)(t)| \leq \frac{|f|_\lambda}{\Gamma(\alpha)} \int_0^t \frac{u^\lambda \mathrm{d}u}{(t-u)^{1-\alpha}}$, and hence

$$|(I^\alpha f)(t)| \leq \frac{\Gamma(2+\lambda)|f|_\lambda}{(1+\lambda)\Gamma(\alpha+\lambda+1)} t^{\alpha+\lambda},$$

which proves the estimate for $|I^\alpha f|$. Next, we prove that $I^\alpha f \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. For this, introduce $\phi(t) := f(t) - f(0)$ and consider $t, t+h \in \mathbb{I}$ with $h > 0$,

$$\begin{aligned}(I^\alpha f)(t+h) - (I^\alpha f)(t) &= \frac{1}{\Gamma(\alpha)} \left(\int_{-h}^t \frac{\phi(t-u)}{(u+h)^{1-\alpha}} \mathrm{d}u - \int_0^t \frac{\phi(t-u)}{u^{1-\alpha}} \mathrm{d}u \right) \\ &= \frac{\phi(t)}{\Gamma(1+\alpha)} \left[(t+h)^\alpha - t^\alpha \right] + \frac{1}{\Gamma(\alpha)} \left(\int_{-h}^0 \frac{\phi(t-u) - \phi(t)}{(u+h)^{1-\alpha}} \mathrm{d}u \right) \\ &+ \frac{1}{\Gamma(\alpha)} \left(\int_0^t \left[(u+h)^{\alpha-1} - u^{\alpha-1} \right] \left[\phi(t-u) - \phi(t) \right] \mathrm{d}u \right) =: J_1 + J_2 + J_3. \end{aligned} \tag{A.2}$$

We first consider J_1. If $h > t$, then

$$|J_1| \leq \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^\lambda \left[(t+h)^\alpha - t^\alpha \right] \leq C h^{\lambda+\alpha}.$$

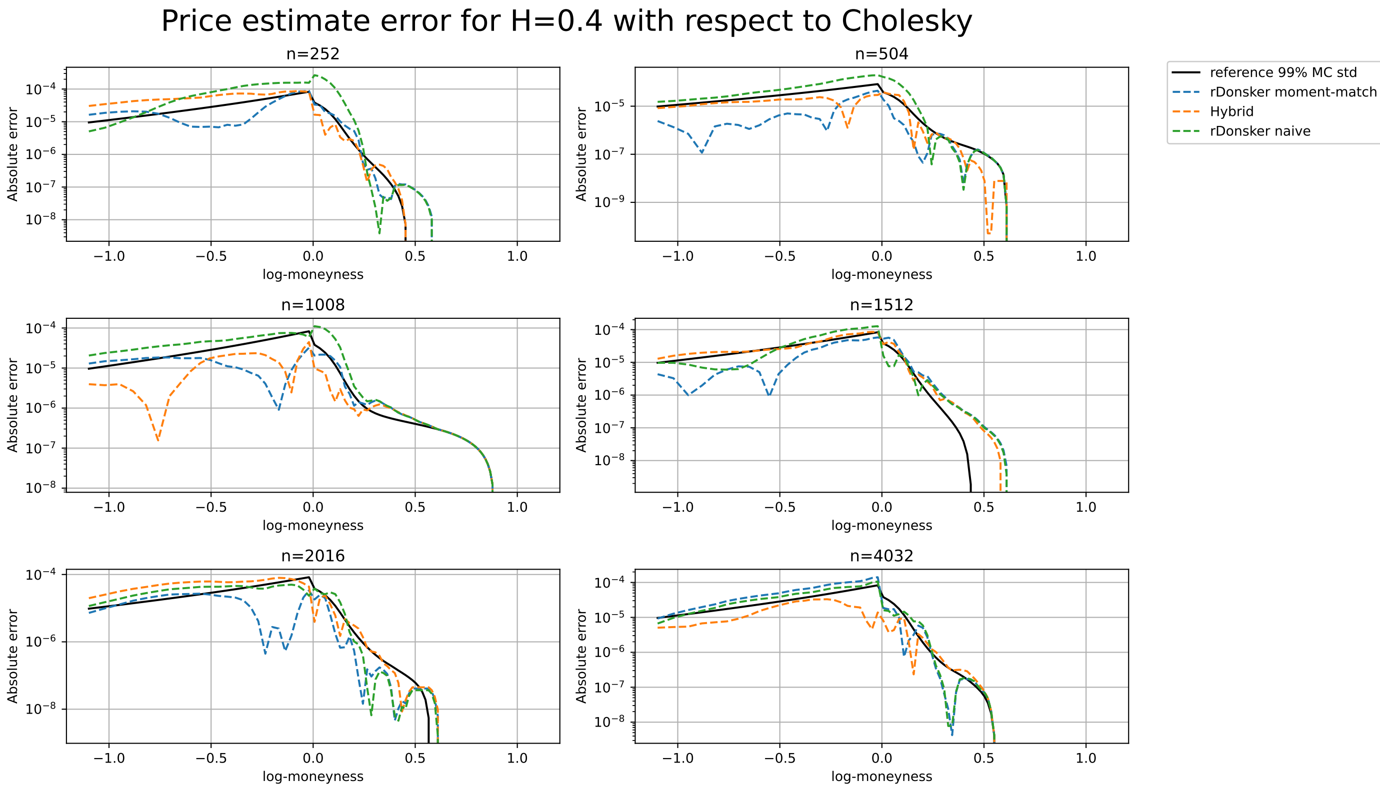

FIGURE 11. Rough Bergomi Call option price comparison with $H = 0.4, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

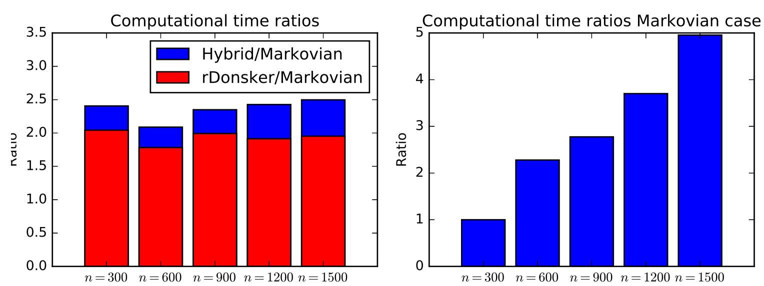

FIGURE 12. Computational time benchmark using Hybrid scheme, rDonsker and Markovian (forward Euler) for different grid sizes n.

On the other hand, when $0 < h < t$, since $(1+u)^\alpha - 1 \le \alpha u$ for $u > 0$, then

$$|J_1| \le \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^{\lambda+\alpha} \left| \left(1 + \frac{h}{t}\right)^\alpha - 1 \right| \le Cht^{\lambda+\alpha-1} \le Ch^{\lambda+\alpha}.$$

For J_2, since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we can write

$$|J_2| \le \frac{|f|_\lambda}{\Gamma(\alpha)} \int_{-h}^0 \frac{|u|^\lambda}{(u+h)^{1-\alpha}} \le Ch^{\lambda+\alpha}.$$

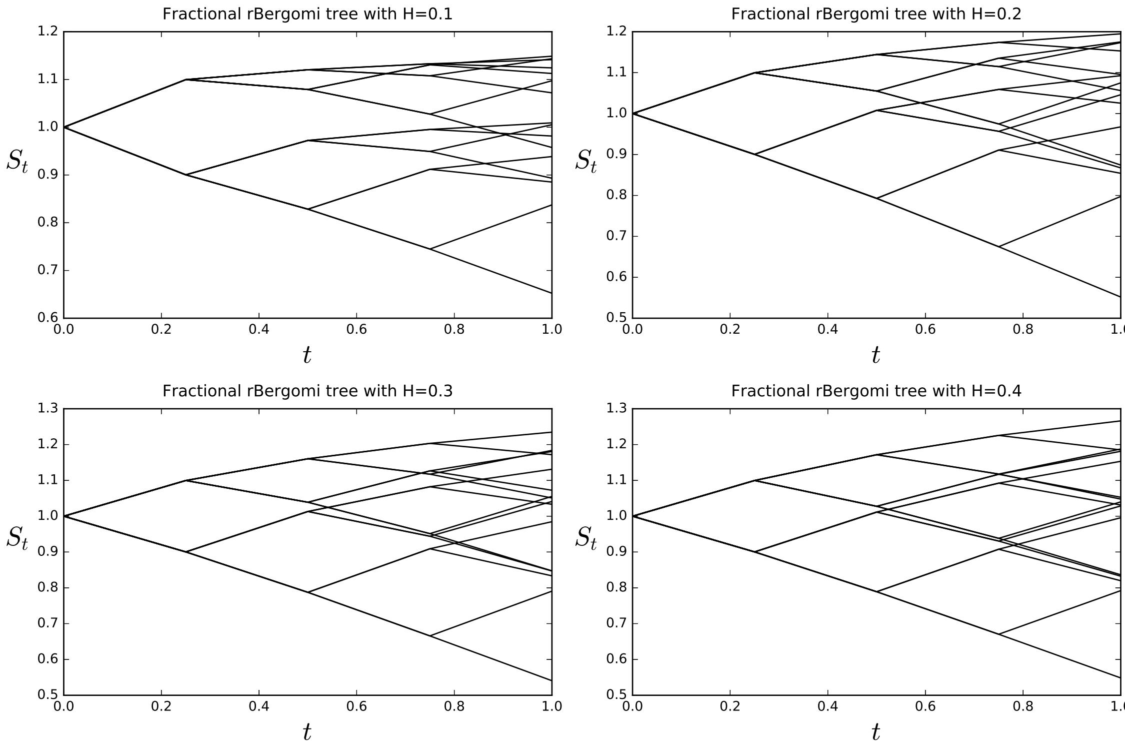

FIGURE 13. rBergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 5 time steps.

Finally,

$$|J_3| \leq \frac{|f|_\lambda}{\Gamma(\alpha)} \int_0^t u^\lambda [u^{\alpha-1} - (u+h)^{\alpha-1}] \mathrm{d}u = \frac{|f|_\lambda}{\Gamma(\alpha)} h^{\lambda+\alpha} \int_0^{t/h} u^\lambda [u^{\alpha-1} - (u+1)^{\alpha-1}] \mathrm{d}u.$$

Hence, if $t \leq h$, then $|J_3| \leq Ch^{\lambda+\alpha}$. Likewise, if $t > h$ and $\lambda + \alpha < 1$, then $|J_3| \leq Ch^{\lambda+\alpha}$ since

$$\left|u^{\alpha-1} - (u+1)^{\alpha-1}\right| = u^{\alpha-1}\left[1 - \left(1 + \frac{1}{u}\right)^{\alpha-1}\right] \leq Cu^{\alpha-2}.$$

Thus, we have shown that $I^\alpha f$ satisfies the $(\lambda+\alpha)$-Hölder condition and belongs to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$ in the case $\alpha > 0$. The conclusion for $\alpha < 0$ follows by taking $g(u) := u^\alpha$ in the proof of Proposition 1.2 in Appendix C. □

Corollary A.3. *For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, I^α is a continuous operator from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$.*

Proof. It is clear that I^α is a linear operator. From Theorem A.2, $\|I^\alpha f\|_{\alpha+\lambda} \leq C_1 \|f\|_\lambda \|(\cdot)^{\alpha+\lambda}\|_{\lambda+\alpha} \leq C\|f\|_\lambda$, since $|f|_\lambda \leq \|f\|_\lambda$. Therefore I^α is also bounded and hence continuous. □

APPENDIX B. DISCRETE CONVOLUTION

Definition B.1. For $\mathrm{a}, \mathrm{b} \in \mathbb{R}^n$, the discrete convolution operator $* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$(\mathrm{a} * \mathrm{b})_i := \sum_{m=0}^{i} \mathrm{a}_m \mathrm{b}_{i-m}, \quad i = 0, \dots, n-1.$$

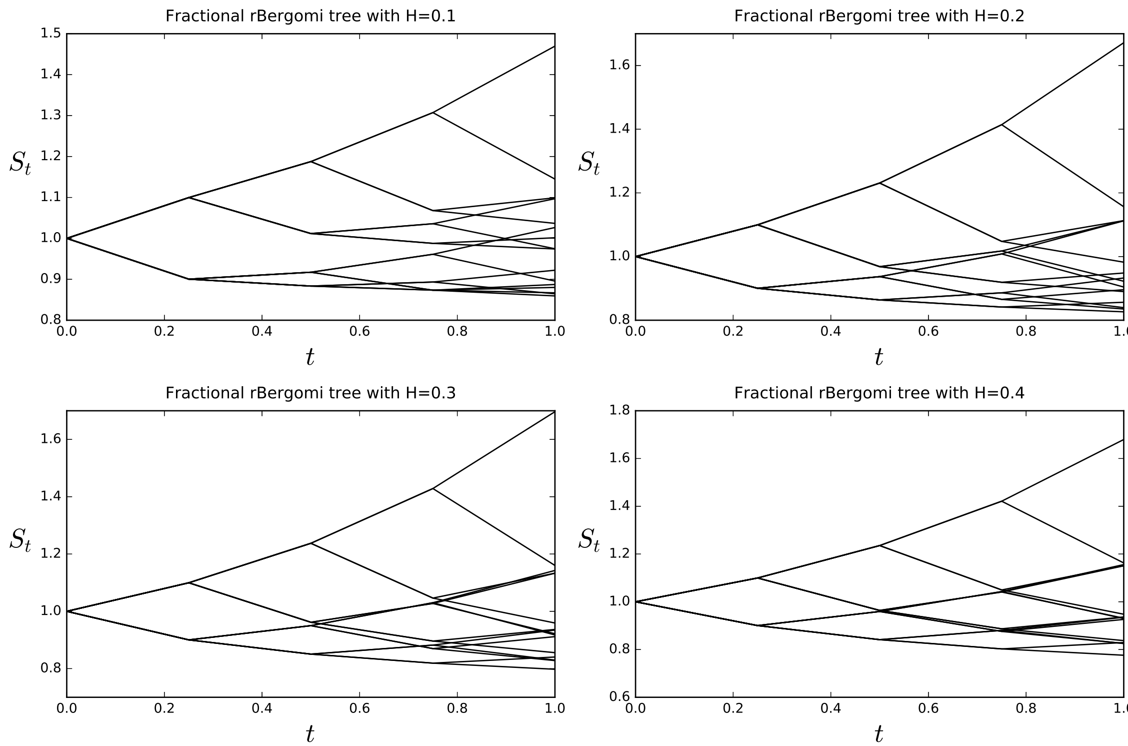

FIGURE 14. rBergomi trees with $(\nu, \rho, \xi_0) = (1, 1, 0.04)$ and five time steps.

When simulating $\mathcal{G}^\alpha W$ on the uniform partition $\mathcal{T}$, the scheme reads

$$(\mathcal{G}^\alpha W)^j(t_i) = \sum_{k=1}^{i} g(t_i - t_{k-1})\xi_k = \sum_{k=1}^{i} g(t_k)\zeta_{j,k-i+1}, \qquad \text{for } i = 1, \dots, n,$$

which has the form of the discrete convolution in Definition B.1. Rewritten in matrix form,

$$\begin{pmatrix} g(t_1) & 0 & \cdots & 0 \\ g(t_2) & g(t_1) & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ g(t_n) & g(t_{n-1}) & \cdots & g(t_1) \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{pmatrix},$$

it is clear that this operator yields a complexity of order $\mathcal{O}(n^2)$, which can be improved drastically.

Definition B.2. The Discrete Fourier Transform (DFT) of a sequence $\mathrm{c} := (c_0, c_1, ..., c_{n-1}) \in \mathbb{C}^n$ is given by

$$\widehat{f}(\mathrm{c})[j] := \sum_{k=0}^{n-1} c_k \exp\left(-\frac{2\mathtt{i}\pi jk}{n}\right), \qquad \text{for } j = 0, \dots, n-1,$$

and the Inverse DFT of c is given by

$$f(\mathrm{c})[k] := \frac{1}{n}\sum_{j=0}^{n-1} c_j \exp\left(\frac{2\mathtt{i}\pi jk}{n}\right), \qquad \text{for } k = 0, \dots, n-1.$$

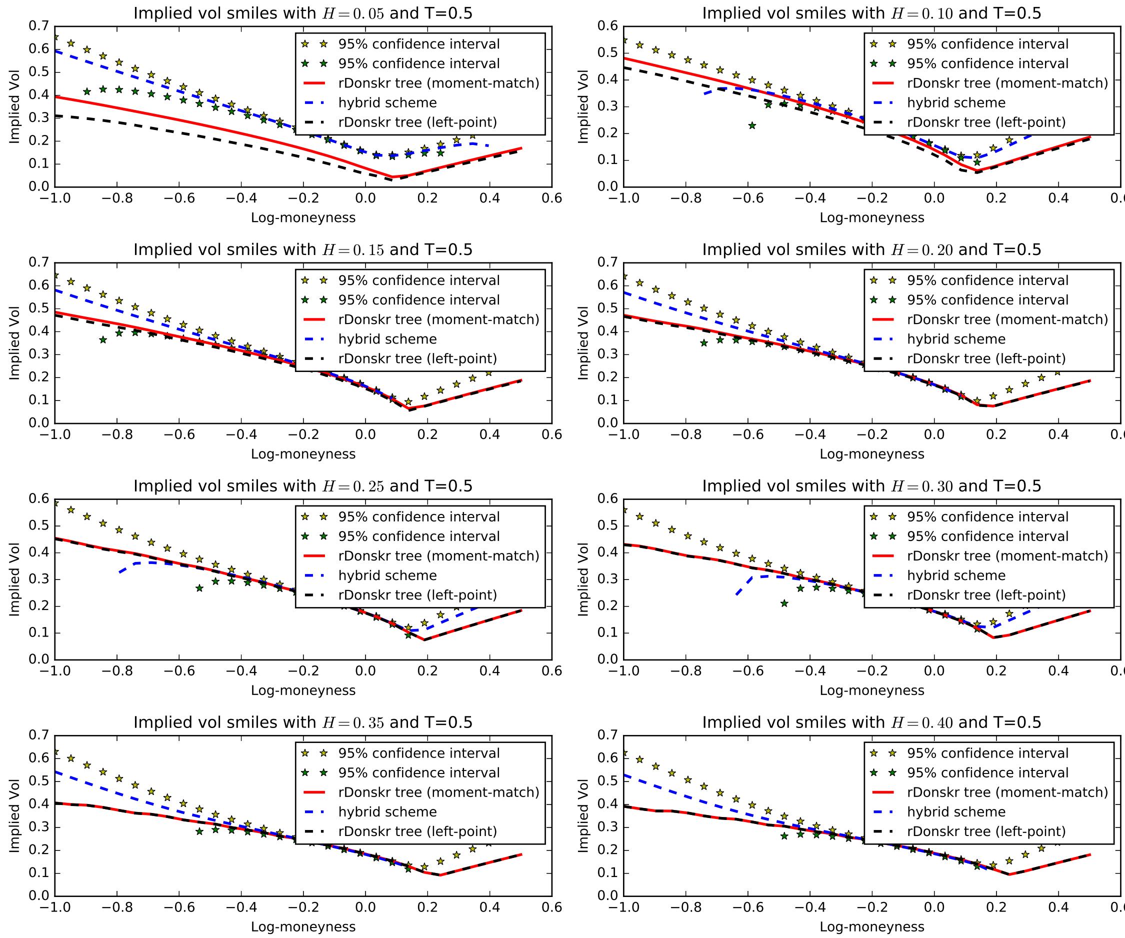

FIGURE 15. rBergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$, 24 time steps.

In general, both transforms require a computational effort of order $\mathcal{O}(n^2)$, but the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey [23] exploits the symmetry and periodicity of complex exponentials of the DFT and reduces the complexity of both transforms to $\mathcal{O}(n \log n)$.

Theorem B.3. *For* $\mathrm{a}, \mathrm{b} \in \mathbb{R}^n$*, the identity* $(\mathrm{a} * \mathrm{b}) = f\big(\widehat{f}(\mathrm{a}) \bullet \widehat{f}(\mathrm{b})\big)$ *holds, with* $\bullet$ *the pointwise multiplication.*

This in particular implies that the complexity of the discrete convolution is reduced to $\mathcal{O}(n \log n)$ by FFT.

Algorithm B.4 (FFT Discrete convolution for $\mathcal{B}$)**.** On the equidistant grid $\mathcal{T}$,

(1) draw a random matrix $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M \\ i=1,\dots,n}}$ such that $\mathbb{V}(\zeta_{j,i}) = 1$;
(2) define the vectors $\mathfrak{g} := (g(t_i))_{i=1,\dots,n}$ and $\zeta_j := (\zeta_{j,i})_{i=1,\dots,n}$, for $j = 1, \dots, M$;
(3) using FFT, compute $\varphi_j := \widehat{f}(\mathfrak{g}) \cdot \widehat{f}(\zeta_j)$, for $j = 1, \dots, M$;
(4) simulate M paths of $(\mathcal{G}^\alpha W)$ using FFT, as $(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}} f(\varphi_j)$ for $j = 1, \dots, M$.

In Step 2 we may replace the evaluation points $\mathfrak{g}$ by any optimal evaluation point $\{g\,(t_i^*)\}_{i=1}^n$ as in (3.4). Many packages offer a direct implementation of the discrete convolution such as `numpy.convolve` in `Python`. The user then only needs to pass the arguments $\mathfrak{g}$ and ξ_j, and Steps 3 and 4 are computed automatically (using

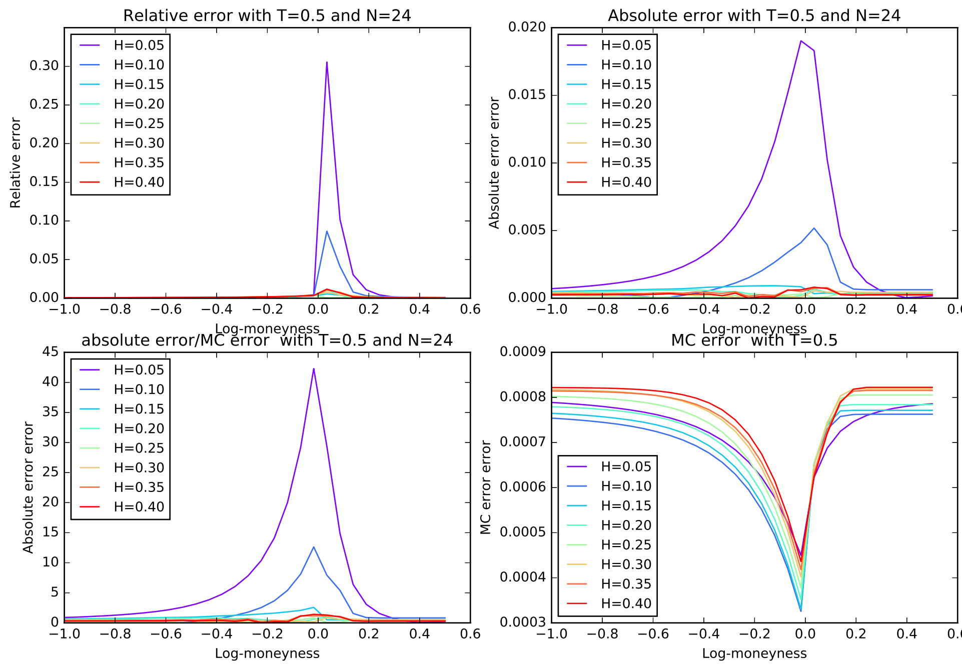

Figure 16. Error analysis for the rDonsker moment-match tree for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps.

efficient FFT techniques) by the function. Although the FFT step is the heaviest computation on the simulation of rough volatility models, the actual time grid $\mathcal{T}$ is not specially large ($n \ll 1000$). Hence, the fastest FFT for very large n is not essential, as the implementation is run on smaller time grids. In this aspect we find that `numpy.convolve` is a very competitive implementation.

Appendix C. Proof of Proposition 1.2

In this section, we present a proof of Proposition 1.2. We first consider the case $\alpha \in (-\lambda, 0)$ with $0 < \lambda \leq 1$. Fix $f \in \mathcal{C}^\lambda(\mathbb{I})$ and $g \in \mathcal{L}^\alpha$. As our first step, we derive a useful representation akin to [79, Equation (13.1)], but for the operator $\mathcal{G}^\alpha$, which amounts to $\mathcal{G}^\alpha f(0) = 0$ and

$$\mathcal{G}^\alpha f(t) = (f(t) - f(0))g(t) - \int_0^t (f(t) - f(s)) \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s, \tag{C.1}$$

for $t \in (0, 1]$ (note that $g(0)$ need not be defined, but our assumptions guarantee $\mathcal{G}^\alpha f(0) = 0$). To show that (C.1) holds, we look at the difference quotients for the definition of $\mathcal{G}^\alpha f$ in (1.1). For any $t \in [0, 1)$ and

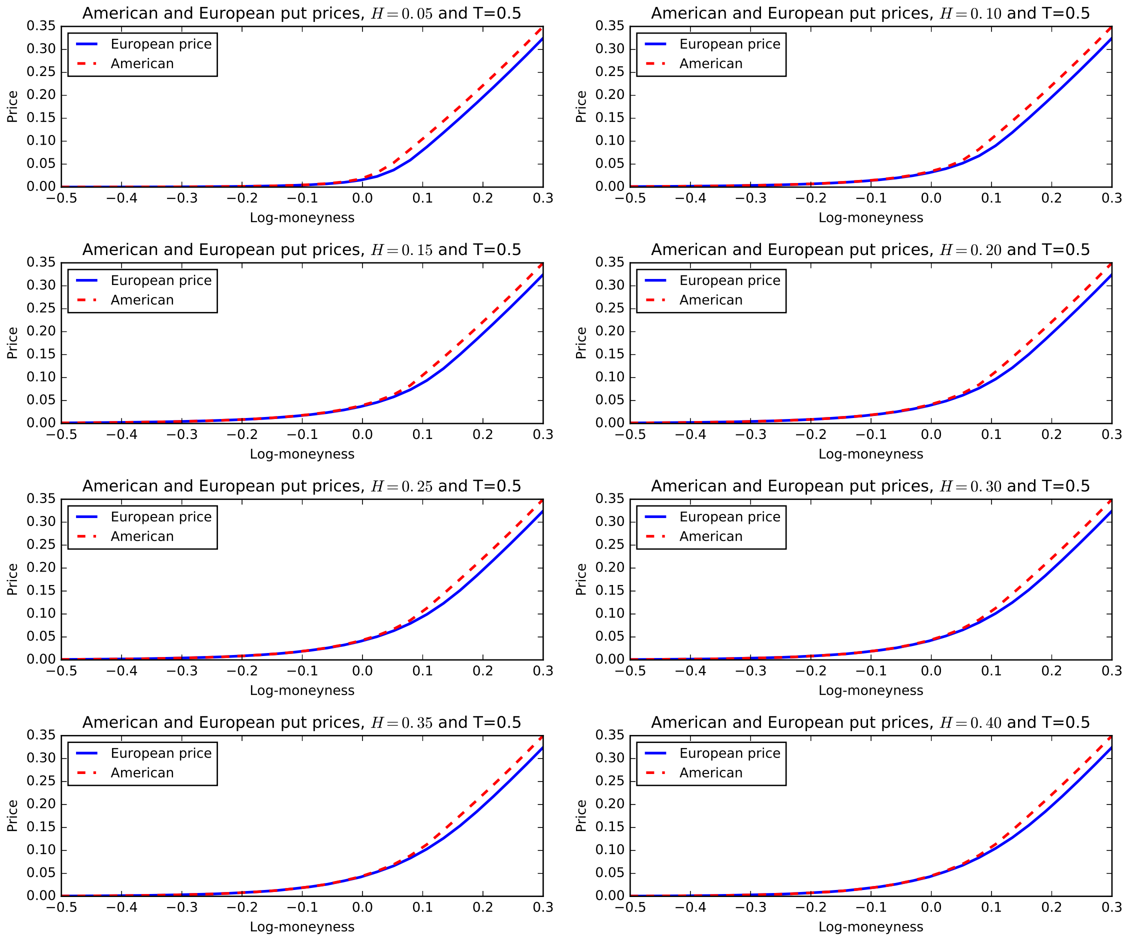

FIGURE 17. American and European Put prices in the rough Bergomi model for different values of H and $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 26 time steps.

any small enough $h > 0$, a bit of rewriting leads to the equality

$$
\begin{aligned}
\text{(C.2)} \qquad & \int_0^{t+h} (f(s) - f(0)) g(t+h-s) \mathrm{d}s - \int_0^t (f(s) - f(0)) g(t-s) \mathrm{d}s \\
&= \int_0^t (f(s) - f(t)) \big(g(t+h-s) - g(t-s) \big) \mathrm{d}s \\
&\quad + (f(t) - f(0)) \left(\int_0^{t+h} g(t+h-s) \mathrm{d}s - \int_0^t g(t-s) \mathrm{d}s \right) + \int_t^{t+h} (f(s) - f(t)) g(t+h-s) \mathrm{d}s.
\end{aligned}
$$

In the second term on the right-hand side of (C.2), a change of variables gives

$$
(f(t) - f(0)) \left(\int_0^{t+h} g(t+h-s) \mathrm{d}s - \int_0^t g(t-s) \mathrm{d}s \right) = (f(t) - f(0)) \int_{-h}^0 g(t-r) \mathrm{d}r.
$$

Looking at the third term on the right-hand side of (C.2), our assumptions yield

$$
\left| \int_t^{t+h} (f(s) - f(t)) g(t+h-s) \mathrm{d}s \right| \le \int_t^{t+h} C_1 h^\lambda C_2 h^\alpha \mathrm{d}s \le C_0 h^{1+\lambda+\alpha} = o(h),
$$

as h tends to zero, since $\lambda+\alpha\in(0,1)$. As for the first term, we have

$$\left|(f(t)-f(s))\frac{g(t+h-s)-g(t-s)}{h}\right| \le \frac{C_1|t-s|^\lambda}{h}\int_{t-s}^{t-s+h} g'(r)\mathrm{d}r \le C_1(t-s)^\lambda C_2|t-s|^{\alpha-1} = C_0(t-s)^{\lambda+\alpha-1},$$

for all $s\in(0,t)$ and $h>0$, where $\lambda+\alpha-1\in(-1,0)$, so the right-hand side is in $L^1([0,t])$ and hence we can apply the dominated convergence theorem. Specifically, dividing by h in (C.2) and sending h to zero, we obtain (C.1) in the limit, by using dominated convergence on the first term, Lebesgue's differentiation on the second term, and noting that the third term vanishes.

Having established (C.1), we can now use it to obtain the desired Hölder estimates. We begin with the first term on the right-hand side of (C.1). Let $\phi(t) := (f(t)-f(0))g(t)$, for $t\in\mathbb{I}$, where we note that $|(f(t)-f(0))g(t)| \le Ct^{\lambda+\alpha}$ with $\lambda+\alpha\in(0,1)$, so $\phi(0)=0$ is well defined. Rewriting, and using the assumptions on f and g, we get, for every $t\in\mathbb{I}$ and $h\in(0,1-t]$,

$$\begin{aligned}|\phi(t+h)-\phi(t)| &\le |f(t)-f(0)|\int_t^{t+h}|g'(r)|\mathrm{d}r + |g(t+h)||f(t+h)-f(t)| \\ &\le C|f|_\lambda t^{\lambda+\alpha}(t+h)^\alpha\big((t+h)^{-\alpha}-t^{-\alpha}\big) + C|f|_\lambda h^{\lambda+\alpha} \le C'|f|_\lambda h^{\lambda+\alpha},\end{aligned}\tag{C.3}$$

where the last inequality follows by elementary considerations, as in the arguments on [68, Chapter 1, Page 15]. The case $h\in[-t,0]$ is analogous.

For the second term on the right-hand side of (C.1), we can follow a procedure similar to the proof of [79, Lemma 13.1]. Defining

$$\varphi(t) := \int_0^t (f(t)-f(s))\frac{\mathrm{d}}{\mathrm{d}t}g(t-s)\mathrm{d}s = \int_0^t (f(t)-f(t-r))\frac{\mathrm{d}}{\mathrm{d}r}g(r)\mathrm{d}r,$$

and rewriting things, for any $t\in\mathbb{I}$ and $h\in[-t,1-t]$, we arrive at

$$\begin{aligned}\varphi(t+h)-\varphi(t) &= \int_0^t (f(t)-f(t-r))\big(g'(r+h)-g'(r)\big)\mathrm{d}r - \int_0^t (f(t)-f(t-r))g'(r+h)\mathrm{d}r \\ &\qquad + \int_{-h}^t (f(t+h)-f(t-u))g'(u+h)\mathrm{d}u \\ &= \int_0^t (f(t)-f(t-r))\big(g'(r+h)-g'(r)\big)\mathrm{d}r - \int_0^t (f(t+h)-f(t))g'(r+h)\mathrm{d}r \\ &\qquad + \int_{-h}^0 (f(t+h)-f(t-u))g'(u+h)\mathrm{d}u := I_1+I_2+I_3.\end{aligned}\tag{C.4}$$

Without loss of generality, we assume $h>0$. For the first integral, a change of variables gives

$$\begin{aligned}|I_1| &\le |f|_\lambda\int_0^t r^\lambda\int_r^{r+h}|g''(u)|\mathrm{d}u\mathrm{d}r \le C|f|_\lambda\int_0^t r^\lambda\big(r^{\alpha-1}-(r+h)^{\alpha-1}\big)\mathrm{d}r \\ &= C|f|_\lambda h^{\lambda+\alpha-1}\int_0^t \big(\frac{r}{h}\big)^\lambda\Big(\big(\frac{r}{h}\big)^{\alpha-1}-\big(\frac{r}{h}+1\big)^{\alpha-1}\Big)\mathrm{d}r = C|f|_\lambda h^{\lambda+\alpha}\int_0^{t/h} u^{\alpha+\lambda-1}\Big(1-\big(1+\frac{1}{u}\big)^{\alpha-1}\Big)\mathrm{d}u \\ &\le C|f|_\lambda h^{\lambda+\alpha}\Big(\int_0^1 u^{\lambda+\alpha-1}\mathrm{d}u + (1-\alpha)\int_1^\infty u^{\lambda+\alpha-2}\mathrm{d}u\Big),\end{aligned}$$

where $\lambda+\alpha\in(0,1)$, so the final two terms on the right-hand side are finite. In the final line, we have used that the mapping $y\mapsto -(1+y)^{\alpha-1}+(\alpha-1)y$ is concave with a maximum value of -1 at $y=0$.

As regards the two remaining integrals I_2 and I_3, we see immediately that

$$|I_2| \le C|f|_\lambda h^\lambda \int_0^\infty (r+h)^{\alpha-1}\mathrm{d}r = \frac{C}{|\alpha|}|f|_\lambda h^{\lambda+\alpha}, \quad \text{and}$$
$$|I_3| \le C|f|_\lambda \int_{-h}^0 (u+h)^{\lambda+\alpha-1}\mathrm{d}u = \frac{C}{|\lambda+\alpha|}|f|_\lambda h^{\lambda+\alpha}.$$

By linearity, the desired continuity of the operator $\mathcal{G}^\alpha : \mathcal{C}^\lambda(\mathbb{I}) \to \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$, for $\alpha \in (-\lambda, 0)$, now follows from (C.1), (C.3), and the three above estimates for (C.4).

It remains to consider $\alpha \in (0, 1-\lambda)$. As before, recall $0 < \lambda \le 1$, and fix $f \in \mathcal{C}^\lambda(\mathbb{I})$ along with $g \in \mathcal{L}^\alpha$. Unlike above, $s \mapsto \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)$ is now integrable on $\mathbb{I}$ which makes things go through more easily: in particular, we can work directly with the definition of $\mathcal{G}^\alpha f$ in (1.1), applying arguments analogous to (C.4). The case $g(u) = u^\alpha$ is already covered by the proof of Theorem A.2. For a general $g \in \mathcal{G}^\alpha$, we can retrace those same steps, except that, in (A.2) and the subsequent estimates for J_1, J_2, and J_3, we must now invoke our control on g and its derivatives (similarly to how we did it above for (C.4) and the subsequent estimates of I_1, I_2, and I_3). This completes the proof of Proposition 1.2.

References

[1] E. Abi Jaber, M. Larsson and S. Pulido. Affine Volterra processes. *Annals of Applied Probability*, 29(5): 3155-3200, 2019.

[2] E. Akyildirim, Y. Dolinsky and H.M. Soner. Approximating stochastic volatility by recombinant trees. *Annals of Applied Probability* 24(5): 2176-2205, 2014.

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

[4] J.P. Aubin and H. Doss. Characterization of stochastic viability of any non-smooth set involving its generalized contingent curvature. *Stochastic Analysis and Applications*, 21(5): 955-981, 2003.

[5] M. Bardi and R. Jensen. A geometric characterization of viable sets for controlled degenerate diffusions. *Set-Valued Analysis*, 10(2-3): 129-141, 2002.

[6] X. Bardina, I. Nourdin, C. Rovira and S. Tindel. Weak approximation of a fractional SDE. *Stochastic Processes and their Applications*, 120(1): 39-65, 2010.

[7] O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64(2): 253-280, 2002.

[8] O. E. Barndorff-Nielsen and J. Schmiegel. Ambit processes: with applications to turbulence and tumour growth. In F. E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal and T. Zhang (Eds.), 2007. *Stochastic Analysis and Applications, volume 2 of Abel Symp.*: 93-124, Springer, Berlin.

[9] M. Beiglböck and P. Siorpaes. Pathwise versions of the Burkholder-Davis-Gundy inequality. *Bernoulli* 21(1): 360-373, 2015.

[10] C. Bayer, P. Friz, P. Gassiat, J. Martin and B. Stemper. A regularity structure for rough volatility. *Mathematical Finance*, 30(3): 782-832, 2020.

[11] C. Bayer, P. Friz and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6): 1-18, 2016.

[12] C. Bayer, P. Friz, A. Gulisashvili, B. Horvath and B. Stemper. Short-time near the money skew in rough fractional stochastic volatility models. *Quantitative Finance*, 19(5): 779-798, 2019.

[13] C. Bender and P. Parczewski. On the connection between discrete and continuous Wick calculus with an application to the fractional Black-Scholes model. *Stochastic Processes, Filtering, Control and Their Applications*, Cohen et al. (Eds.), World Scientific Press, 2012.

[14] M. Bennedsen, A. Lunde and M.S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4): 931-965, 2017.

[15] M. Bennedsen, A. Lunde and M.S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. Forthcoming in *Journal of Financial Econometrics*.

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

[17] L. Broux, F. Caravenna and L. Zambotti. Hairer's multilevel Schauder estimates without regularity structures. *arXiv: 2301.07517*, 2023.

[18] P. Billingsley. Convergence of probability measures. *New York, Wiley*, 1968.

[19] S. Boucheron, G. Lugosi and P. Massart. Concentration inequalities: a nonasymptotic theory of independence. *OUP*, 2013.

[20] N.N. Čentsov. Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the 'heuristic' approach to the Kolmogorov-Smirnov tests. *Theory of Probability & its Applications*, 1(1): 140-144, 1956.

[21] D.S. Clark. Short proof of a discrete gronwall inequality. *Discrete Applied Mathematics*, 16(3): 279-281, 1987.

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

[23] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19: 297-301, 1965.

[24] G. Da Prato and H. Frankowska. Stochastic viability of convex sets. *Journal of Mathematical Analysis and Applications*, 333(1): 151-163, 2007.

[25] C.R. Dietrich and G.N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4): 1088-1107, 1997.

[26] M. Djehiche and M. Eddahbi. Hedging options in market models modulated by the fractional Brownian motion. *Stochastic Analysis and Applications*, 19(5): 753-770, 2001.

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

[28] M.D. Donsker. An invariance principle for certain probability limit theorems. *Memoirs of the AMS*, 6, 1951.

[29] P. Drábek. Continuity of Nemyckij's operator in Hölder spaces. *Comm. Math. Univ. Carolinae*, 16(1): 37-57, 1975.

[30] R.M. Dudley. Uniform central limit theorems. CUP, 1999.

[31] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. Math. Finance, 29 (1): 3-38, 2019.

[32] N. El Karoui. Les aspects probabilistes du contrôle stochastique. *Springer Lecture Notes in Mathematics*, 876: 73-238, 1979.

[33] S. N. Ethier and T. G. Kurtz. Markov processes: Characterization and convergence. *Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons*, 1986.

[34] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal Fin. Math.*, 8: 114-145, 2017.

[35] P.K. Friz and M. Hairer. A course on rough paths. *Springer*, 2021.

[36] P.K. Friz and N. Victoir. Multidimensional stochastic processes as rough paths: theory and applications. CUP, 2010.

[37] M. Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15: 635-654, 2011.

[38] M. Fukasawa, T. Takabatake and R. Westphal. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics (Is volatility rough?) Forthcoming in *Mathematical Finance*.

[39] P. Gassiat. On the martingale property in the rough Bergomi model. Electron. Commun. Probab. 24(33), 2019.

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

[41] J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6): 933-949, 2018.

[42] I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & its Applications*, 5(3): 285-301, 1960.

[43] H. Guennoun, A. Jacquier, P. Roome and F. Shi. Asymptotic behaviour of the fractional Heston model. *SIAM Journal on Financial Mathematics*, 9(3), 1017-1045, 2018.

[44] A. Gulisashvili. Large deviation principle for Volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, 9(3): 1102-1136, 2018.

[45] D. Hamadouche. Invariance principles in Hölder spaces. *Portugal. Math.*, 57: 127-151, 2000.

[46] G.H. Hardy and J.E. Littlewood. Some properties of fractional integrals I. *Mathematische Zeitschrift*, 27: 565-606, 1928.

[47] C. Heinrich, M. Pakkanen and A.E.D. Veraart. Hybrid simulation scheme for volatility modulated moving average fields. Mathematics and Computers in Simulation, 166: 224-244, 2019.

[48] B. Horvath, A. Jacquier and C. Lacombe. Asymptotic behaviour of randomised fractional volatility models. Journal of Applied Probability, 56(2): 496-523, 2019.

[49] Y. Hu, D. Nualart and J. Song. Fractional martingales and characterisation of the fractional Brownian motion. *Annals of Probability*, 37(6): 2404-2430, 2009.

[50] H.E. Hurst. Long-term storage capacity of reservoirs. *Trans. American Society of Civil Engineers*, 116(1): 770-799, 1956.

[51] H.E. Hurst, R.P. Black and Y.M. Simaika. Long-term storage: an experimental study. *London: Constable,* 1965.

[52] A. Jakubowski, J. Memin and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace $D1$ de Skorokhod. *PTRF*, 81(1): 111-137, 1989.

[53] A. Jacquier, C. Martini and A. Muguruza. On VIX futures in the rough Bergomi model. *Quant. Finance*, 18(1): 45-61, 2018.

[54] A. Jacquier, M. Pakkanen and H. Stone. Pathwise large deviations for the rough Bergomi model. *Journal of Applied Probability*, 55(4): 1078-1092, 2018.

[55] I. Karatzas. On the pricing of American options. *Applied Mathematics and Optimization*, 17(1): 37-60, 1988.

[56] I. Karatzas and S.E. Shreve. Brownian motion and stochastic calculus. Springer-Verlag, 1997.

[57] A. Kolmogorov. Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C.R.(Doklady) Acad. URSS (N.S)*, 26: 115-118, 1940.

[58] N.V. Krylov. Lectures on elliptic and parabolic equations in Hölder spaces. *Graduate Studies in Mathematics*, 12, 1996.

[59] T.G. Kurtz and P.E. Protter. Weak convergence of stochastic integrals and stochastic differential equations. *CIME School in Probability. Lecture Notes in Mathematics*, 1627: 1-41, 1991.

[60] J. Kushner. On the weak convergence of interpolated Markov chains to a diffusion. *Annals of Probability*, 2(1): 40-50, 1974.

[61] J. Lamperti. Semi-stable stochastic processes. *Transactions of the AMS*, 104: 62-78, 1962.

[62] J. Lamperti. On convergence of stochastic processes. *Transactions of the AMS*, 104: 430-435, 1962.

[63] G. Livieri, S. Mouti, A. Pallavicini and M. Rosenbaum. Rough volatility: evidence from option prices. *IISE Transactions*, 50(9), 2018.

[64] T.J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2): 215-310, 1998.

[65] B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4): 422-437, 1968.

[66] H.B. Mann and A. Wald. On stochastic limit and order relationships. *Annals of Mathematical Statistics*, 14(3): 217-226, 1943.

[67] R. McCrickerd and M.S. Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11): 1877-1886, 2018.

[68] N.I. Muskhelishvili. Singular integral equations. Boundary problems of functions theory and their applications to mathematical physics. Revised translation from the Russian, edited by J. R. M. Radok. Reprinted. Wolters-Noordhoff Publishing, Groningen, 1972. pp. xii+7–447.

[69] S. Mulinacci and M. Pratelli. Functional convergence of Snell envelopes: applications to American options approximations. *Finance and Stochastics*, 2: 311-327, 1998.

[70] E. Neuman and M. Rosenbaum. Fractional Brownian motion with zero Hurst parameter: a rough volatility viewpoint. *Electronic Communications in Probability*, 23(61): 1-12, 2018.

[71] A. Neuenkirch and T. Shalaiko. The order barrier for strong approximation of rough volatility models. arXiv:1606.03854, 2016.

[72] A. Nieminen. Fractional Brownian motion and martingale-differences. *Statistics & Probability Letters*, 70(1): 1-10, 2004.

[73] P. Parczewski. Donsker-type theorems for correlated geometric fractional Brownian motions and related processes. *Electronic Communications in Probability*, 22(55): 1-13, 2017.

[74] P. Parczewski. A fractional Donsker theorem. *Stochastic Analysis and Applications* 32(2): 2014

[75] J. Picard. Representation formulae for the fractional Brownian motion. *Séminaire de Probabilités*, 43: 3-70, 2011.

[76] D. Pollard. Convergence of stochastic processes. Springer-Verlag, 1984.

[77] A. Račkauskas and C. Suquet. Necessary and sufficient condition for the functional central limit theorem in Hölder spaces. *Journal of Theoretical Probability*, 17(1): 221-243, 2004.

[78] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4): 399-412, 1997.

[79] S.G. Samko, A.A. Kilbas and O.I. Marichev. Fractional integrals and derivatives: theory and applications. Gordon and Breach, Yverdon, 1993.

[80] J.L. Snell. Applications of martingale system theorems. *Transactions of the AMS*, 73: 293-312, 1952.

[81] T. Sottinen. Fractional Brownian motion, random walks and binary market models. *Finance Stoch.*, 5(3): 343-355, 2001.

[82] L.C.G. Rogers and D. Williams. Diffusions, Markov processes and martingales, Volume 2, Itô Calculus. CUP, 2000.

[83] M. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31(4): 287-302, 1975.

[84] R. Weinstock. Calculus of variations: with applications to Physics and Engineering. Dover Publications, 1974.

[85] D. Williams. Probability with martingales. CUP, 1991.

[86] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1): 155-167, 1971.

Department of Mathematics, University of Oxford
Email address: blanka.horvath@maths.ox.ac.uk

Department of Mathematics, Imperial College London, and the Alan Turing Institute
Email address: a.jacquier@imperial.ac.uk

Kaiju Capital Management and Department of Mathematics, Imperial College London
Email address: aitor.muguruza-gonzalez15@imperial.ac.uk

Department of Statistics, London School of Economics
Email address: a.sojmark@lse.ac.uk