FUNCTIONAL CENTRAL LIMIT THEOREMS FOR ROUGH VOLATILITY

BLANKA HORVATH, ANTOINE JACQUIER, AND AITOR MUGURUZA

Abstract. We extend Donsker's approximation of Brownian motion to fractional Brownian motion with Hurst exponent $H \in (0,1)$ and to Volterra-like processes. Some of the most relevant consequences of our 'rough Donsker (rDonsker) Theorem' are convergence results for discrete approximations of a large class of rough 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 [11] 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.

Introduction

Fractional Brownian motion has a long and famous history in probability, stochastic analysis and their applications to diverse fields [42, 43, 50, 58]. 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 [17], and later studied theoretically by Djehiche and Eddahbi [20], Alòs, León and Vives [3] and Fukasawa [31], and given financial motivation and data consistency by Gatheral, Jaisson and Rosenbaum [35] and Bayer, Friz and Gatheral [8]. Since then, a vast literature has pushed the analysis in many directions [7, 9, 12, 26, 28, 35, 36, 40, 47, 65], 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 any 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 too slow, and faster techniques are needed. The state of the art, so far, is the recent hybrid scheme developed by Bennedsen, Pakkanen and Lunde [11], and its turbocharged version [60]. 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 [22] (and later Lamperti [54]) proved a functional central limit for Brownian motion, thereby providing a theoretical justification of random walk approximation to it. Many extensions have been studied in the literature, and we refer the interested reader to [24] for an overview. In the fractional case, Sottinen [74]

Date: November 29, 2017.

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

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

The authors would like to thank Christian Bayer, Peter Friz, 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. The numerical implementations have been carried out on the collaborative platform Zanadu (www.zanadu.io).

and Nieminen [64] constructed–following Donsker's ideas of using iid sequences of random variables– an approximation 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 [76] 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 considered for the convergence. The recent development of rough paths theory [29, 30, 57] provided an appropriate framework to extend Donsker's results to processed with sample paths of Hölder regularity strictly smaller than $1/2$. For Hurst parameters $H \in (1/3, 1/2)$, Bardina, Nourdin, Rovira and Tindel [4] 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 motions is the topology induced by the Hölder norms 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 [10, 66, 67] 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 here a general class of continuous processes with any Hölder regularity, including fractional Brownian motion with $H \in (0, 1)$, but also truncated Brownian semi-stationary processes, Gaussian Volterra process, as well as some rough volatility models recently proposed in the literature. The fundamental novelty of our approach is that we create an approximating sequence capable of simultaneously keeping track of the approximation of the 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 extension to the rough case is so far an open problem. Our analysis easily generalises beyond Brownian drivers to more general semimartingales, emphasising that the subtle, but 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 functional a Donsker-Lamperti-type functional central limit theorem (FCLT). This further allows us (i) to provide a pathwise justification of the hybrid scheme by Bennedsen, Lunde and Pakkanen [11], and (ii) to develop tree-based simulation 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, with which we provide a pedestrian alternative to the advanced hybrid schemes of [11, 60]. That is, we present a simple Monte-Carlo pricer with the particular appeal of a low implementation complexity and provide comparison charts against [11] in terms of accuracy and against [60] in terms of speed, with respect to the Hurst parameter H. Reminders on Riemann-Liouville operators and additional technical proofs are postponed to the appendix.

Notations: We consider the unit interval $\mathbb{I} := [0,1]$. We denote by $\mathcal{C}(\mathbb{I})$ and $\mathcal{C}^\alpha(\mathbb{I})$ the space of continuous and α-Hölder continuous functions on $\mathbb{I}$ with Hölder regularity $\alpha \in (0,1)$. In addition, we shall denote by $\mathcal{C}^1(\mathbb{I})$ and $\mathcal{C}^1_b(\mathbb{I})$ the space of continuously differentiable functions and bounded continuously differentiable functions. We shall use C, C_1, C_2 are 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 [22] (also termed 'functional central limit theorem') ensures the weak convergence of an approximating sequence to a Brownian motion in the Skorohod 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 (essential requirement in order to price path-dependent financial securities for example). We shall 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 $\mathcal{C}^\beta(\mathbb{I})$, equipped 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 [51, Chapter 3]. Following the spirit of Riemann-Liouville fractional operators recalled in Appendix A, we introduce the class of Generalised Fractional Operators (GFO). For any $\alpha \in (-1,1)$, we introduce the space $\mathcal{L}^\alpha := \left\{u \mapsto u^\alpha L(u) : L \in \mathcal{C}^1_b(\mathbb{I})\right\}$, as well as the following subset of $\mathbb{R}^2$:

$$\mathfrak{R} := \Big\{(\alpha,\lambda) \in (-1,1) \times (0,1) \text{ such that } \alpha + \lambda \in (0,1)\Big\}.$$

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

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

We shall further use the notation $G(t) := \int_0^t g(u)\mathrm{d}u$, for any $t \in \mathbb{I}$. The following kernels and operators are well-known examples of Generalised Fractional Operators:

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

The following theorem generalises the classical mapping properties of Riemann-Liouville fractional operators first proved by Hardy and Littlewood [38], and will be of fundamental importance in the rest of our analysis. To ease the flow of the paper, we postpone its proof to Appendix C.1.

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

In this paper we develop an approximation scheme for the following system, generalising the concept of rough volatility introduced in [3, 31, 33] in the context of mathematical finance, where the X process represents the dynamics of the logarithm of a stock price process:

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2}V_t\mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t, \quad && X_0 = 0, \\ V_t &= \Phi\left(\mathcal{G}^\alpha Y\right)(t), && V_0 > 0, \end{aligned} \tag{1.3}$$

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

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

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 $\rho \in [-1,1]$. The functional Φ is continuous on $\mathcal{C}(\mathbb{I})$, and for any $\varphi \in \mathcal{C}(\mathbb{I})$, $\Phi(\varphi)$ is continuously differentiable and integrable. This is enough to ensure that the first stochastic differential equation is well defined. It remains to formulate the precise definition for $\mathcal{G}^\alpha Y$ (Propsition 1.4) to fully specify the system (1.3) and clarify the existence of solutions. Existence and (strong) uniquess of a solution to the second SDE in (1.4) is guaranteed by the following standard assumption [78]:

Assumption 1.3. There exist $C_b, C_a > 0$ and an increasing function $\varrho : (0,\infty) \to (0,\infty)$ with $\lim_{\epsilon\downarrow 0} \int_\epsilon^1 \frac{\mathrm{d}x}{\varrho(x)} = \infty$. such that

$$|b(x) - b(y)| \le C_b|x-y| \qquad \text{and} \qquad |a(x)-a(y)| \le C_a\sqrt{\varrho(|x-y|)}.$$

Not only is the solution to (1.4) continuous, but $\frac{1}{2}$-Hölder continuous as a consequence of Kolmogorov-Čentsov's theorem [16]. Existence and precise meaning of the term $\mathcal{G}^\alpha Y$ is more delicate, and is treated further 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. *The equality* $(\mathcal{G}^\alpha W)(t) = \int_0^t g(t-s)\mathrm{d}W_s$ *holds almost surely for all* $t \in \mathbb{I}$.

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-Riouville fractional Brownian motion (or Type-II fractional Brownian motion) can be written as $C_\alpha\mathcal{G}^\alpha W$. Furthermore, Theorem 2.5 yields that the Riemann-Liouville operator is continuous from $\mathcal{C}^{1/2}(\mathbb{I})$ to $\mathcal{C}^{1/2+\alpha}(\mathbb{I})$. Each kernel in (1.2) gives rise to such processes, that have been proposed in turbulence modelling and in mathematical finance by Barndorff-Nielsen and Schmiegel [6].

Example 1.5. The first example is the rough Bergomi model introduced by Bayer, Friz and Gatheral [8], where

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

with $V_0, \nu, \xi_0(\cdot) > 0$, $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ 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 $\sigma \in \mathcal{C}_b^1(\mathbb{I})$, this class falls within the GFO framework.

Example 1.7. Bennedsen, Lunde and Pakkanen [12] considered adding a Gamma kernel to the volatility process, which yields the Truncated Brownian semi-stationary (Bergomi-type) model:

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

with $V_0 > 0 > 0$ and $\lambda < 0$. This corresponds exactly to (1.3) with $g(u) \equiv u^\alpha \mathrm{e}^{-\lambda u}$, $Y = W$ and

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

Example 1.8. The rough Heston model introduced by Guennoun, Jacquier, Roome and Shi [35] 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 $\eta > 0$, $\alpha \in (-\frac{1}{2}, 0)$. 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 [26] (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 our project, namely an approximation scheme for the system (1.3). The basic ingredient to construct approximating sequences is a family of iid random variables, which satisfies 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 $\mathbb{E}(\xi_1^2) = \sigma^2 > 0$.

Following Donsker [22] and Lamperti [54], we first define, for any $\omega \in \Omega$, $n \geq 1$, $t \in \mathbb{I}$, the approximating sequence for the driving Brownian motion B as

$$B_n(t) := \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt \rfloor}\xi_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}}\xi_{\lfloor nt \rfloor + 1}. \tag{1.5}$$

As will be explained later, a similar construction holds to approximate the process Y:

$$Y_n(t) := \frac{1}{n}\sum_{k=1}^{\lfloor nt \rfloor} b\left(Y_n^{k-1}\right) + \frac{nt - \lfloor nt \rfloor}{n} b\left(Y_n^{\lfloor nt \rfloor}\right) + \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt \rfloor} a\left(Y_n^{k-1}\right)\xi_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} a\left(Y_n^{\lfloor nt \rfloor}\right)\xi_{\lfloor nt \rfloor + 1}, \tag{1.6}$$

where $Y_n^k := Y_n(t_k)$, where $t_k := \frac{k}{n}$, from which we naturally deduce an approximating scheme (up to the interpolating term which decays to zero by Chebyshev's inequality) for X as

$$X_n(t) := -\frac{1}{2n}\sum_{k=1}^{\lfloor nt \rfloor}\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k) + \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{\lfloor nt \rfloor}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k)}\left(B_n^{k+1} - B_n^k\right) \tag{1.7}$$

All the approximations above, as well as all the convergence statements below should be understood pathwise, but we omit the ω dependence in the notations for clarity. The main result of the paper is a convergence statement about the approximating sequence $(X_n)_{n\geq 1}$. In usual weak convergence analysis [14], convergence is stated in the Skorohod space $(\mathcal{D}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$ of càdlàg processes equipped with the Skorohod topology. While Theorem 1.10 proves it, it further provides convergence in the stronger Hölder space $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ for $\lambda < \frac{1}{2}$, albeit with additional restrictions.

Theorem 1.10. *The sequence $(X_n)_{n\geq 1}$ converges weakly to X in $(\mathcal{D}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$. Furthermore, for $\lambda < \frac{1}{2}$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, convergence in the Hölder space $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ holds if either of the following two conditions holds:*
(i) all the ξ_i are distributed as $\mathcal{N}(0,1)$ and $\mathbb{E}[\mathrm{e}^{\Phi(\mathcal{G}^\alpha Y)}]$ is finite;
(ii) all the ξ_i are bounded almost surely and $\mathbb{E}[\mathrm{e}^{\Phi(\mathcal{G}^\alpha Y_n)}]$ is finite for each n.

In (ii), the moment condition on the sequence (Y_n) is difficult to check. However, it clearly holds as soon as the iid sequence (ζ_i), approximating the stochastic driver of Y, is bounded. 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 approximation (Y_n) in some Hölder space, which we then translate, first into convergence of the stochastic integral in (1.3), then, by continuity of the mapping Φ into convergence of the sequence $(\Phi(\mathcal{G}^\alpha Y_n))$. All these ingredients are detailed in Section 2 below. Once this is achieved, the proof of the theorem itself is relatively straightforward, as illustrated in Section 2.5.

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 Skorohod topology. We concentrate here on convergence in the Hölder topology, due to Lamperti [55]. The standard convergence result for Brownian motion can be stated as follows:

Theorem 2.1. *For $\alpha < \frac{1}{2}$, the sequence (B_n) in (1.5) converges weakly to a Brownian motion in $(\mathcal{C}^\alpha(\mathbb{I}), \|\cdot\|_\alpha)$.*

The proof relies on finite-dimensional convergence and tightness of the approximating sequence. Not surprisingly, the tightness criterion [14] in the Skorohod space $\mathcal{D}(\mathbb{I})$ and in a Hölder space setting are very different. In fact, the tightness criterion in Hölder spaces is strictly related to Kolmogorov-Čentsov's continuity theorem [16]. Note, in passing, that the approximating sequence (1.5) is differentiable (in time) for each $n \geq 1$, even though its limit is obviously not.

Theorem 2.2 (Sufficient conditions for weak convergence in Hölder spaces)**.** *Let $Z \in \mathcal{C}^\lambda(\mathbb{I})$ and $(Z_n)_{n\geq 1}$ its corresponding approximating sequence in the sense that for any $t_1 \leq \ldots \leq t_k$ in $\mathbb{I}$, $(Z_n(t_1), \ldots, Z_n(t_k))$ converges in distribution to $(Z(t_1), \ldots, Z(t_k))$ as n tends to infinity. Assume further that the tightness criterion*

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

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

As pointed out by Račkauskas and Suquet in [70], strictly speaking the convergence takes place 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

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

Then $\left(C_0^\lambda(\mathbb{I}), \|\cdot\|_\lambda^0\right)$ becomes a separable closed subspace of $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ (see [70, 37] for details), and one can then use the simple tightness criteria introduced in Theorem 2.2. 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 later. To conclude our review of weak convergence in Hölder spaces, the following theorem, due to Račkauskas and Suquet [70] provides necessary and sufficient conditions ensuring convergence in Hölder space:

Theorem 2.3 (Račkauskas-Suquet [70]). *Let $\alpha \in (0, \frac{1}{2})$ and $p(\alpha) := \frac{1}{1-2\alpha}$. The sequence $(B_n)_{n\geq 1}$ in (1.5) converges (pathwise) weakly to a Brownian motion in $\mathcal{C}^\alpha(\mathbb{I})$ if and only if $\mathbb{E}(\xi_1) = 0$ and $\lim_{t\uparrow\infty} t^{p(\alpha)}\mathbb{P}(|\xi_1| \geq t) = 0$.*

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 2.4. *The sequence $(Y_n)_{n\geq 1}$ converges weakly to Y in (1.4) in $(\mathcal{C}^\alpha(\mathbb{I}), \|\cdot\|_\alpha)$ for all $\alpha < \frac{1}{2}$.*

Proof. Finite-dimensional convergence is a classical result by Kushner [53], so only tightness needs to be checked. Using $Y_n^i := Y_n\left(\frac{i}{n}\right)$ as above, and without loss of generality assume $Y_n^0 = 0$ and $b(Y_n^0) = 0$, so that

$$\mathbb{E}\left(\left|Y_n^1\right|^{2p}\right) = \mathbb{E}\left(\left|\frac{b\left(Y_n^0\right)}{n} + \frac{a\left(Y_n^0\right)}{\sigma\sqrt{n}}\xi_1\right|^{2p}\right) \leq \frac{C}{n^p}\mathbb{E}\left(|\xi_1|^{2p}\right).$$

Assumption 1.3 yields

$$\begin{aligned}
\mathbb{E}\left(\left|Y_n^2\right|^{2p}\right) &= \left(\left|\frac{1}{n}\sum_{k=1}^{2} b\left(Y_n^{k-1}\right) + \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{2} a\left(Y_n^{k-1}\right)\xi_k\right|^{2p}\right) \\
&\leq \left\{\mathbb{E}\left[\left|Y_n^1\right|\right] + \frac{1}{\sqrt{n}}\mathbb{E}\left[\left(\left|\frac{C_b Y_n^1}{\sqrt{n}}\right| + \frac{\left|b(Y_n^0)\right|}{\sqrt{n}} + \frac{C_a}{\sigma}\sqrt{\rho\left(|Y_n^1|\right)}\xi_2 + \left|a(Y_n^0)\xi_2\right|\right)\right]\right\}^{2p} \leq \frac{C}{n^p}\mathbb{E}\left(\left(|\xi_1| + |\xi_2|\right)^{2p}\right).
\end{aligned}$$

By induction we find $\mathbb{E}\left(\left|Y_n^i - Y_0\right|^{2p}\right) \leq \frac{C}{n^p}\mathbb{E}\left[\left(\sum_{k=1}^{i}|\xi_i|\right)^{2p}\right]$, which implies the tightness criterion (2.1) for $p > 1$ for $\alpha = 2p$ and $\beta = p - 1$. □

2.3. **Invariance principle for rough processes.** We have set the ground to extend our results to processes that are not necessarily 1/2-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 [59], which establishes the preservation of weak convergence under continuous operators.

Theorem 2.5 (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 2.5 in order to prove weak convergence [69, Chapter IV]. This path avoids the lengthy computations of tightness and finite-dimensional convergence in classical proofs [14]. In fact, Hamadouche [37] already realised that Riemann-Liouville fractional operators are continuous, hence Theorem 2.5 holds under mapping by Hölder continuous functions. In contrast, the novelty of our approach is to consider, on the one hand the family of GFO applied to a Brownian motion, and on the other hand the extension of Brownian motion to Itô diffusions. In fact, minimal changes to the proof in Proposition 1.4 yield the following:

Corollary 2.6. *If Y is the solution to (1.4), then $(\mathcal{G}^\alpha Y)(t) = \int_0^t g(t-s)\mathrm{d}Y_s$ almost surely for all $t \in \mathbb{I}$.*

The analogue of Theorem 2.4 for Y follows by continuous mapping along with the fact that $\mathcal{G}^\alpha$ is a continuous operator from $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\alpha\right)$ to $\left(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_\alpha\right)$ for all $\lambda \in (0, 1)$ such that $(\alpha, \lambda) \in \mathfrak{R}$.

Theorem 2.7 (Generalised rough Donsker)**.** *For (Y_n) in (1.6), Y its weak limit in $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ for $\lambda<\frac{1}{2}$,*

$$(\mathcal{G}^\alpha Y_n)(t) = \sum_{i=1}^{\lfloor nt\rfloor} n\left[G\left(t-\frac{i-1}{n}\right) - G\left(t-\frac{i}{n}\right)\right]\left(Y_n^i - Y_n^{i-1}\right) + nG\left(t-\frac{\lfloor nt\rfloor}{n}\right)\left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right) \tag{2.2}$$

converges weakly to $\mathcal{G}^\alpha Y$ in $\left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$ for any $\alpha \in (-\frac{1}{2},\frac{1}{2})$.

Proof. We apply directly the definition (1.1) of the GFO to the sequence (1.6), recalling that the latter is differentiable in time. For $\alpha>0$, integration by parts yields, for any $n\geq 1$ and $t\in\mathbb{I}$,

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

since $G(0)=g(0)=0$. When $\alpha<0$, similar steps imply

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

when $\frac{\lfloor nt\rfloor}{n}=t$, $G(0)=0$, and the expression is well defined. □

We may omit the interpolation term in Donker's linear interpolating sequence (1.5), and Lamperti's proof [55] still holds with the sequence $B_n(t) := \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\xi_i$, as the rightmost term in (1.5) converges to zero by Chebychev inequality. This statement also holds for the sequence (1.6), so that, $\mathcal{G}^\alpha Y_n$ in (2.2) reduces to

$$(\mathcal{G}^\alpha Y_n)(t) = \sum_{k=1}^{\lfloor nt\rfloor} g(t-t_{k-1})\left(Y_n^k - Y_n^{k-1}\right) = \sum_{k=1}^{\lfloor nt\rfloor-1}\left[g(t-t_{k-1}) - g(t-t_k)\right]Y_n^{k-1} \tag{2.3}$$

which coincides with the usual left-point forward Euler approximation. For numerical purposes, (2.3) is much more efficient, since the integral G required in (2.2) is not necessarily available in closed form. The speed of convergence of the rDonsker scheme is not of order $\mathcal{O}\left(n^{-1/2}\right)$ as one might assume. In fact the Hurst parameter (in particular $\alpha\in(-\frac{1}{2},0)$) influences the speed of convergence adversely, as the following proposition shows.

Proposition 2.8. *The speed of convergence of the rDonsker scheme is of order $\mathcal{O}\left(n^{-\alpha-1/2}\right)$ when $\alpha\in(-\frac{1}{2},0]$ and $\mathcal{O}\left(n^{-1/2}\right)$ when $\alpha\in(0,\frac{1}{2})$.*

Proof. Let $\alpha\in(-\frac{1}{2},0]$. Since $g\in\mathcal{L}^\alpha$, the approximation (2.3) reads, for any $n\geq1$,

$$(\mathcal{G}^\alpha Y_n)(t_i) = \frac{1}{n^{1/2-\alpha}\sigma}\sum_{k=1}^{i}\left(nt_i-(k+1)T\right)^\alpha L(t_i-t_{k-1})\left(Y_n^k - Y_n^{k-1}\right)\sigma\sqrt{n}, \quad \text{for } i=0,\dots,n.$$

Here, $(nt_i - T(k-1))^\alpha \leq (t_i)^\alpha$ is bounded for any $n \geq 1$ so that the claim follows directly. When $\alpha \in (0, \frac{1}{2})$ we may rewrite (2.3) as

$$\left(\mathcal{G}^\alpha Y_n\right)(t_i) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{i} \left(t_i - t_{k-1}\right)^\alpha L\left(t_i - t_{k-1}\right)\left(Y_n^k - Y_n^{k-1}\right)\sigma\sqrt{n}, \quad \text{for } i = 0, \ldots, n.$$

In this case, $\left(t_i - t_{k-1}\right)^\alpha \leq (t_i)^\alpha$ is also bounded for any $n \geq 1$, and the proposition follows. □

So far, our results hold for a class of α-Hölder continuous functions. It is often necessary, at least for practical reasoning purposes, 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 to know whether the α-Hölder continuity is preserved after this composition.

Proposition 2.9. *Let $(Y_n)_{n\geq 1}$ be the approximating sequence* (1.6)*. Then $\left(\Phi\left(\mathcal{G}^\alpha Y_n\right)\right)$ converges weakly to $\Phi\left(\mathcal{G}^\alpha Y\right)$ in $\left(\mathcal{C}^{\alpha+1/2}(\mathbb{I}), \|\cdot\|_{\alpha+1/2}\right)$ for any $\alpha \in (-\frac{1}{2}, \frac{1}{2})$.*

Proof. Drábek [23] found necessary and sufficient conditions ensuring that Hölder continuity is preserved under composition (which he calls Nemyckij operators). More precisely, he proved that the composition $f \circ g$ is continuous from $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $\left(\mathcal{C}^\lambda(\mathbb{I}, \|\cdot\|_\lambda\right)$ if and only if f is of class $\mathcal{C}^1$. The proof of the proposition then follows by applying the Continuous Mapping Theorem to Theorem 2.7 along with Drábek's continuity property. The following diagram summarises the steps, where $\lambda < 1/2$. The double arrows indicate weak convergence, and we indicate next to them the topology in which it takes place.

$$\begin{tikzcd}
\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_{1/2}\right) \arrow[r, bend left, "\mathcal{G}^\alpha"] & \left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right) \arrow[r, bend left, "\Phi"] & \left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+1/2}\right) \\
Y_n \arrow[r, "\mathcal{G}^\alpha"] \arrow[d, Rightarrow, "\|\cdot\|_\lambda"'] & \mathcal{G}^\alpha(Y_n) \arrow[r, "\Phi"] \arrow[d, Rightarrow, "\|\cdot\|_{\alpha+\lambda}"'] & \Phi(\mathcal{G}^\alpha Y_n) \arrow[d, Rightarrow, "\|\cdot\|_{\alpha+\lambda}"'] \\
Y \arrow[r, "\mathcal{G}^\alpha"] & \mathcal{G}^\alpha Y \arrow[r, "\Phi"] & \Phi(\mathcal{G}^\alpha Y)
\end{tikzcd}$$

□

2.4. **Convergence of the (log-)stock process in the Hölder topology.** We extend here the convergence to the log-stock process maintaining the Hölder space framework. To start with, the Hölder regularity coefficient of an Itô integral with an integrand having λ-Hölder continuous paths is not at all obvious. The following proposition gives an answer to this question.

Proposition 2.10. *Let W be a standard Brownian motion, and Θ a càdlàg process on the same filtered probability space with finite moments up to order $2p$. Then $\Theta \bullet W \in \mathcal{C}^\lambda(\mathbb{I})$ for all $\lambda < \frac{1}{2}\left(1 - \frac{1}{p}\right)$.*

Proof. For this we will use Kolmogorov-Čentsov's continuity theorem [16].

$$\mathbb{E}\left[\left(\int_0^t \Theta(u)\mathrm{d}W_u - \int_0^s \Theta(u)\mathrm{d}W_u\right)^{2p}\right] = \mathbb{E}\left[\left(\int_s^t \Theta(u)\mathrm{d}W_u\right)^{2p}\right] = \mathbb{E}\left[\left(\int_s^t \Theta(u)^2\mathrm{d}u\right)^{p}\right]$$
$$\le C(t-s)^{p-1}\left(\int_s^t \mathbb{E}\left[\Theta(u)^{2p}\right]\mathrm{d}u\right) \le C(t-s)^p$$

by Itô's isometry and Hölder's inequality along with the finite moments of Θ. Thus, by Kolmogorov's continuity criterion the stochastic integral $\Theta \bullet W$ has continuous paths with Hölder regularity $\frac{1-1/p}{2}$ for all $p \ge 1$. □

The finiteness of all moments might be too restrictive for some applications, and in fact this will be relaxed in Section 2.5 at the cost of switching to the Skorohod topology. Nevertheless, in the Hölder setting, once Proposition 2.10 applies, it suffices to prove continuity of the Itô map between the corresponding Hölder spaces.

Proposition 2.11. *The Itô map $\Theta \mapsto \Theta \bullet W$ is continuous from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^\upsilon(\mathbb{I})$ for all $\lambda \in (0,1)$, $\upsilon < \frac{1}{2}$.*

Proof. Let $f \in \mathcal{C}^\lambda(\mathbb{I})$ and $W \in \mathcal{C}^\upsilon(\mathbb{I})$. Since the Itô map is linear, it suffices to check boundedness.

$$\left\|\int_0^t f(s)\mathrm{d}W_s\right\|_\Upsilon \le \left\|\int_0^t ||f||_\lambda \mathrm{d}W_s\right\|_\upsilon \le ||f||_\lambda \|W_t\|_\Upsilon \le ||f||_\lambda \left\|Ct^{1/2}\right\|_\upsilon \le CT^{1/2}||f||_\lambda,$$

where we have used the Hölder continuity of W, and the proposition follows. □

Finally, we present the main convergence result.

Theorem 2.12. *Let $\Phi(\mathcal{G}^\alpha Y_n)$ as in (2.3) with $\xi_i \sim \mathcal{N}(0,1)$, and weak limit $\Phi(\mathcal{G}^\alpha Y)$ in $\left(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$, for $\lambda < \frac{1}{2}$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. If $\mathbb{E}[\mathrm{e}^{\Phi(\mathcal{G}^\alpha Y)}] < \infty$, then the sequence defined by*

$$-\frac{1}{2n}\sum_{i=1}^{\lfloor nt\rfloor} \Phi\left(\mathcal{G}^\alpha Y_n\right)(t_i) + \frac{\rho}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_i)}\xi_i + \frac{\overline{\rho}}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_i)}\zeta_i$$

where (ζ_i) is an iid family of $\mathcal{N}(0,1)$ random variables, converges weakly in $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to

$$-\frac{1}{2}\int_0^t \Phi\left(\mathcal{G}^\alpha Y\right)(s)\mathrm{d}s + \int_0^t \sqrt{\Phi\left(\mathcal{G}^\alpha Y\right)(s)}\left(\rho \mathrm{d}W_s + \overline{\rho}\mathrm{d}W_s^\perp\right).$$

Proof. The proof follows by repeatedly applying the continuous mapping theorem after Proposition 2.9. For the deterministic integral part one can easily prove that the integral mapping is continuous from $\mathcal{C}^{\alpha+\lambda}$ to $\mathcal{C}^\lambda$ using a similar argument to Proposition 2.11. Then we get

$$\int_0^t \Phi\left(\mathcal{G}^\alpha Y_n\right)(s)\,\mathrm{d}s = \sum_{i=1}^{\lfloor nt\rfloor}\int_{\frac{i}{n}}^{\frac{i+1}{n}} \Phi\left(\mathcal{G}^\alpha Y_n\right)\left(\frac{i}{n}\right)\mathrm{d}s = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\Phi\left(\mathcal{G}^\alpha Y_n\right)\left(\frac{i}{n}\right).$$

For the stochastic integral by definition we have that $\Phi\left(\mathcal{G}^\alpha Y\right) \in L^1$ is well defined and the finiteness of all moments allows us to apply Proposition 2.10. Then using the continuity of the Itô map we obtain the following approximating sequence weakly convergent in $\mathcal{C}^{1/2}(\mathbb{I})$:

$$\int_0^t \sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(s)}\mathrm{d}Z_s = \sum_{i=1}^{\lfloor nt\rfloor}\int_{\frac{i}{n}}^{\frac{i+1}{n}}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_i)}\mathrm{d}Z_s = \sum_{i=1}^{\lfloor nt\rfloor}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_i)}\left(Z(t_{i+1}) - Z(t_i)\right).$$

Then, the problem reduces to being able to simulate the increments of Z exactly, taking into account that $\mathrm{corr}(Z,W) = \rho$ must also hold. Since the increments of Z are Gaussian we may easily construct this explicitly

$$Z(t_{i+1}) - Z(t_i) = \frac{1}{\sqrt{n}}\left(\rho\xi_i + \sqrt{1-\rho^2}\zeta_i\right)$$

where the independence of the iid $\mathcal{N}(0,1)$ sequences (ξ_i) and (ζ_i) is crucial for this to be exact. □

We used here the approximation (2.3), instead of (2.2), essentially for computational reasons. It is of course possible to use the latter, at the cost of increasing complexity of the approximating sequence due to the interpolating term involving double integrals, in general not available in closed form. Proposition 2.11 allows to maintain the Hölder space framework but only if the family (ξ_i) is restricted to be Gaussian, which is in any case sufficient for Monte-Carlo simulations. Nevertheless, the following proposition relaxes this condition.

Theorem 2.13. *Let the sequences $(\Phi(\mathcal{G}^\alpha Y_n), W_n)$ defined by 2.3 and 1.5 converge weakly to $(\mathcal{G}^\alpha Y, W))$ in the joint Hölder topology $\mathcal{C}^{\alpha+\lambda} \times \mathcal{C}^\lambda(\mathbb{I})$ for $\lambda < \frac{1}{2}$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. Assume further $\mathbb{E}[\mathrm{e}^{\Phi(\mathcal{G}^\alpha Y_n)}] < \infty$ and that the iid family (ξ_i) in Assumption 1.9 is bounded. Then the sequence of stochastic integrals $(\Phi(\mathcal{G}^\alpha Y_n) \bullet W_n)$ also converges to $\Phi(\mathcal{G}^\alpha Y) \bullet W$ in $\mathcal{C}^\lambda(\mathbb{I})$.*

Proof. We will make use of Theorem 2.2. Finite dimensional convergence follows from Jakubowski, Memin and Pagès [45], since the approximating sequence 1.5 with bounded random variables satisfies the Uniform Tightness (see [45] for details) criterion. Then it remains to prove tightness of the approximating sequence,

$$\begin{aligned}
\mathbb{E}\left[\left\{\sum_{j=ns}^{nt} \Phi\left(\mathcal{G}^\alpha Y_n\right)(t_j)\left(W_n(t_{j+1}) - W_n(t_j)\right)\right\}^{2p}\right] &\leq \frac{C}{n^{2p}} \mathbb{E}\left[\left(\sum_{j=ns}^{nt} \Phi\left(\mathcal{G}^\alpha Y_n\right)(t_j)^{2p}\right)\right] \\
&\leq \frac{C}{n^{2p}} \sum_{j=ns}^{nt} \mathbb{E}\left[\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_j)^{2p}\right] \leq \frac{C}{n^{2p}},
\end{aligned}$$

where we have made use of the boundedness of ξ, Jensen's inequality and the finiteness of all moments of $\Phi\left(\mathcal{G}^\alpha Y_n\right)$. The inequality then gives the desired convergence result in $\mathcal{C}^\lambda(\mathbb{I})$ □

As opposed to Theorem 2.12 (where the driving random variables are forced to be Gaussian), Theorem 2.13 allows to use any family of bounded random variables as approximating sequences of $W^\perp$ and any family random variables ensuring the moment condition $\mathbb{E}[\mathrm{e}^{\Phi(\mathcal{G}^\alpha Y_n)}] < \infty$. The gap between these two sets of conditions, that neither theorem covers, but this will be discussed in Section 2.5.

2.5. **Extending the weak convergence to the Skorohod space and proof of Theorem 1.10.** The Skorohod space of càdlàg processes equipped with the Skorohod topology has been widely used to prove weak convergence [14]. The Skorohod space of càdlàg processes equipped with the Skorohod norm, which we denote $(\mathcal{D}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$, markedly simplifies when we only consider continuous processes (as is the case of our framework with Hölder continuous processes). Billingsley [14, Chapter 3 Section 12] proved that the identity $(\mathcal{D}(\mathbb{I}) \cap \mathcal{C}(\mathbb{I}), \|\cdot\|_{\mathcal{D}}) \equiv (\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$ always holds. This seemingly simple statement allows us to reduce proofs of weak convergence of continuous processes in the Skorohod topology to that in the supremum norm, usually much simpler. We start with the following straightforward observation:

Lemma 2.14. *The identity map is continuous from $\left(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $(\mathcal{D}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$ for all $\lambda \in (0,1)$.*

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 Skorohod norm in the space of continuous functions is equivalent to the supremum norm. □

Applying the Continuous Mapping Theorem twice, first with the Generalised fractional operator (Theorem 2.7), then with the identity map, yields the following result directly:

Theorem 2.15. *The sequence* $(\Phi(\mathcal{G}^\alpha Y_n))$ *converges weakly to* $\Phi\left(\mathcal{G}^\alpha Y\right)$ *in* $(\mathcal{D}(\mathbb{I}), ||\cdot||_{\mathcal{D}})$ *for any* $\alpha \in (-\frac{1}{2}, \frac{1}{2})$.

The final step in the proof of our main theorem, is to extend weak convergence to the log-stock price. For this, the following result on weak convergence of stochastic integrals $X \bullet Y := \int X \mathrm{d}Y$ due to Jakubowski, Memin and Pagès [45], and later generalised to SDEs by Kurtz and Protter [52] is the key ingredient.

Theorem 2.16. *Let* $(W_n)_{n\geq 1}$ *be as in* (1.5)*,* N *a càdlàg process on* $\mathbb{I}$*, and* $(N_n)_{n\geq 1}$ *an approximating sequence such that* (N_n, W_n) *converges weakly in* $\left(\mathcal{D}(\mathbb{I}^2), \|\cdot\|_{\mathcal{D}}\right)$ *to* (N, W)*. Then, there exists a filtration* $\mathcal{H}$ *under which* W *is an* $\mathcal{H}$*-continuous martingale and* $(N_n, W_n, N_n \bullet W_n)_{n\geq 1}$ *converges weakly to* $(N, W, N \bullet W)$.

As noted in [52], the Skorohod topology in $\mathcal{D}(\mathbb{I}^2)$ is stronger than in $\mathcal{D}(\mathbb{I}) \times \mathcal{D}(\mathbb{I})$. In order to use this result, we first need to have the joint convergence of the two correlated driving Brownian motions W and Z. Let $(W_n)_{n\geq 1}$ and $(W_n^\perp)_{n\geq 1}$ be two sequences as in (1.5), with weak limits W and $W^\perp$, and let $\overline{\rho} := \sqrt{1-\rho^2}$. Donsker's invariance implies that $(W_n, W_n^\perp)_{n\geq 1}$ converges weakly to $(W, W^\perp)$ in $\left(\mathcal{C}^\alpha(\mathbb{I}^2), \|\cdot\|_\alpha\right)$, and hence by the Continuous Mapping Theorem with $f(x,y) := \left(x, \rho x + \sqrt{1-\rho^2} y\right)$, the sequence $(W_n, \rho W_n + \overline{\rho} W_n^\perp)_{n\geq 1}$ converges weakly to $(W, \rho W + \overline{\rho} W^\perp)$ in $\left(\mathcal{C}^\alpha(\mathbb{I}^2), \|\cdot\|_\alpha\right)$ for all $\alpha < \frac{1}{2}$. Finally, the first term on the right-hand side of (1.7) converges weakly to $-\frac{1}{2}\int_0^T \Phi\left(\mathcal{G}^\alpha Y\right)(s)\mathrm{d}s$ by the Continuous Mapping Theorem, as the integral is a continuous operator from $(\mathcal{D}(\mathbb{I}), ||\cdot||_{\mathcal{D}})$ to itself. Since the couple (Y_n, W_n) converges weakly to (Y, W) in $\left(\mathcal{D}(\mathbb{I}^2), ||\cdot||_{\mathcal{D}}\right)$, Theorem 2.16 implies that the second term on the right-hand side of (1.7) converges weakly to $\sqrt{\Phi(\mathcal{G}^\alpha Y)} \bullet W$, and Theorem 1.10 follows.

3. Applications

3.1. **Weak convergence of the Hybrid scheme.** The Hybrid scheme (and its turbocharged version [60]) introduced by Bennedsen, Lunde and Pakkanen [11] is the current state-of-the-art to simulate $\mathcal{TBSS}$ processes. However, only convergence in the mean-square-error sense was proved, but not weak convergence, which would justify the use of the scheme for path-dependent options. Unless otherwise stated, we shall denote by $\mathcal{I} := \{t_i = \frac{i}{n}\}_{i=0,\dots,n}$ the uniform grid on $\mathbb{I}$. The framework developed above provides such a convergence result:

Proposition 3.1. *The sequence* $(\widetilde{\mathcal{G}}^\alpha W_n)$ *in the Hybrid scheme (defined below in* (3.1)*) converges to* $\mathcal{G}^\alpha W$ *in* $\left(\mathcal{C}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ *for* $\alpha \in (-1/2, 1/2)$.

Proof. The Hybrid scheme in [11] with $\kappa \geq 1$ can be written

$$\widetilde{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{(i-\kappa)\vee 0} g(t_i - t_{k-1})\xi_k + \int_{0\vee t_{i-\kappa}}^{t_i} g(t_i - s)\mathrm{d}W_s, \quad i = 0, \dots, n, \tag{3.1}$$

with $\xi_k := \int_{t_{k-1}}^{t_k} \mathrm{d}W_s \sim \mathcal{N}(0, 1/n)$ Gaussian, hence satisfying the conditions in Theorem 2.1. Comparing (2.3) with (3.1), weak convergence in the former implies weak convergence in the latter, since the error of the Hybrid scheme is smaller. The result then follows by Theorem 2.7 □

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 [39].

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 $g(u) \equiv u^{H-1/2}$ for $H \in (0,1)$, for which Theorem 2.7 provides a weakly converging sequence. On the partition $\mathcal{I}$, with Bernoulli random variables $\{\xi_i\}_{i=1}^n$ satisfying $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = \frac{1}{2}$ for all i (justified by Theorem 2.13), the approximating sequence reads

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

Figures 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.

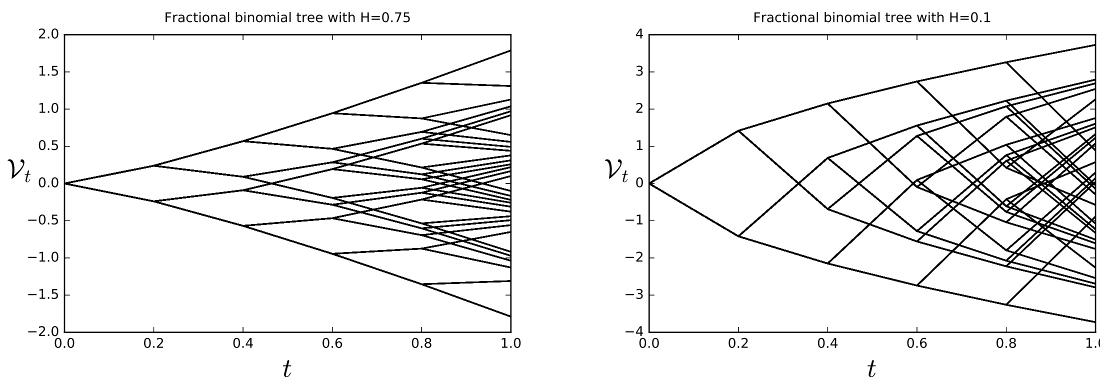

FIGURE 1. Binomial tree for the Riemann-Liouville fractional Brownian motion with $n = 5$ discretisation points for $H = 0.75$ (left) and $H = 0.1$ (right).

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 $T > 0$.

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]

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

and also compute

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

[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.

(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

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

Remark 3.4.

- When $Y=W$, we may skip step (2) and replace $\Delta Y_n^j(t_i)$ by $\sqrt{T/n}\xi_{i,j}$ on step (3).
- Step (3) may be replaced by the Hybrid scheme algorithm [11] only when $Y=W$.

Antithetic variates in Algorithm 3.3 are easy to implement as it suffices to consider the uncorrelated random vectors $\xi_j := (\xi_{j,1},\xi_{j,2},\dots,\xi_{j,n})$ and $\zeta_j := (\zeta_{j,1},\zeta_{j,2},\dots,\zeta_{j,n})$, for $j=1,\dots,M$. Then $(\rho\zeta_j+\overline{\rho}\xi_j,\zeta_j)$, $(\rho\zeta_j-\overline{\rho}\xi_j,\zeta_j)$, $(-\rho\zeta_j-\overline{\rho}\xi_j,-\zeta_j)$ and $(-\rho\zeta_j+\overline{\rho}\xi_j,-\zeta_j)$, 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 (2.3) (with $Y=W$) may be modified to match moments as

$$(\mathcal{G}^\alpha W)(t_i) \approx \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{i} g\left(t_k^*\right)\xi_k, \qquad \text{for } i=0,\dots,n, \tag{3.2}$$

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 (3.2) is always zero, and the second moment reads

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

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

$$g(t_k^*) = \sqrt{n\int_{t_{k-1}}^{t_k} g(t-s)^2\mathrm{d}s}, \qquad \text{for any } k=1,\dots,n. \tag{3.3}$$

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

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

This optimal evaluation point framework is also valid for the Hybrid scheme [11]. The authors originally proposed an optimal evaluation point minimising the mean square error. Nevertheless, we have seen in Proposition 3.1 that the scheme converges weakly already with a left-point approximation, hence the user is free to choose the optimal evaluation point based on criteria different from the mean square error.

3.3.2. *Reducing Variance.* As Bayer, Friz and Gatheral [8] and Bennedsen, Lunde and Pakkanen [11] 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 $\mathcal{F}_t^B = \sigma(B_s : s \le t)$ and $\mathcal{F}_t^W = \sigma(W_s : s \le t)$ 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 [71], and recently adapted to the rBergomi case by McCrickerd and Pakkanen [60], we can decompose the stock price process as

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

and notice that

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

Thus $\exp(X_t)$ becomes log-normal and the Black-Scholes closed-form formulae are valid here (European, Barrier options, maximum,...). The advantage of this approach is that the orthogonal Brownian motion $B^\perp$ 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 also this simple trick 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 3.5 (Simulation of rough volatility models with Brownian drivers)**.** Consider the equidistant grid $\mathcal{T}$.

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

(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):

$$(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \xi_j), \quad j = 1,\dots,M,$$

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

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

(4) Finally, we have $X^j(T) \sim \mathcal{N}(X_T^j - \Sigma^j, \Sigma^j)$ for $j = 1, \ldots, M$; we may compute any option using the Black-Scholes formula. For instance a Call option with strike K would be given by $C^j(K) = \exp(X_T^j)\mathcal{N}(d_1^j) - K\mathcal{N}(d_2^j)$ for $j = 1, \ldots, M$, where $d_1^j := \frac{1}{\sqrt{\Sigma^j}}(X_T^j - \log(K) + \frac{1}{2}\Sigma^j)$ 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 the case of general diffusions Y as drivers of the volatility (see Algorithm 3.3 step (2)). Algorithm 3.3 is obviously faster than 3.5, especially when using control variates. Nevertheless, with the same number of paths, Algorithm 3.5 remarkably reduces the Monte-Carlo variance, meaning in turn that fewer simulations are needed, making it very competitive for calibration.

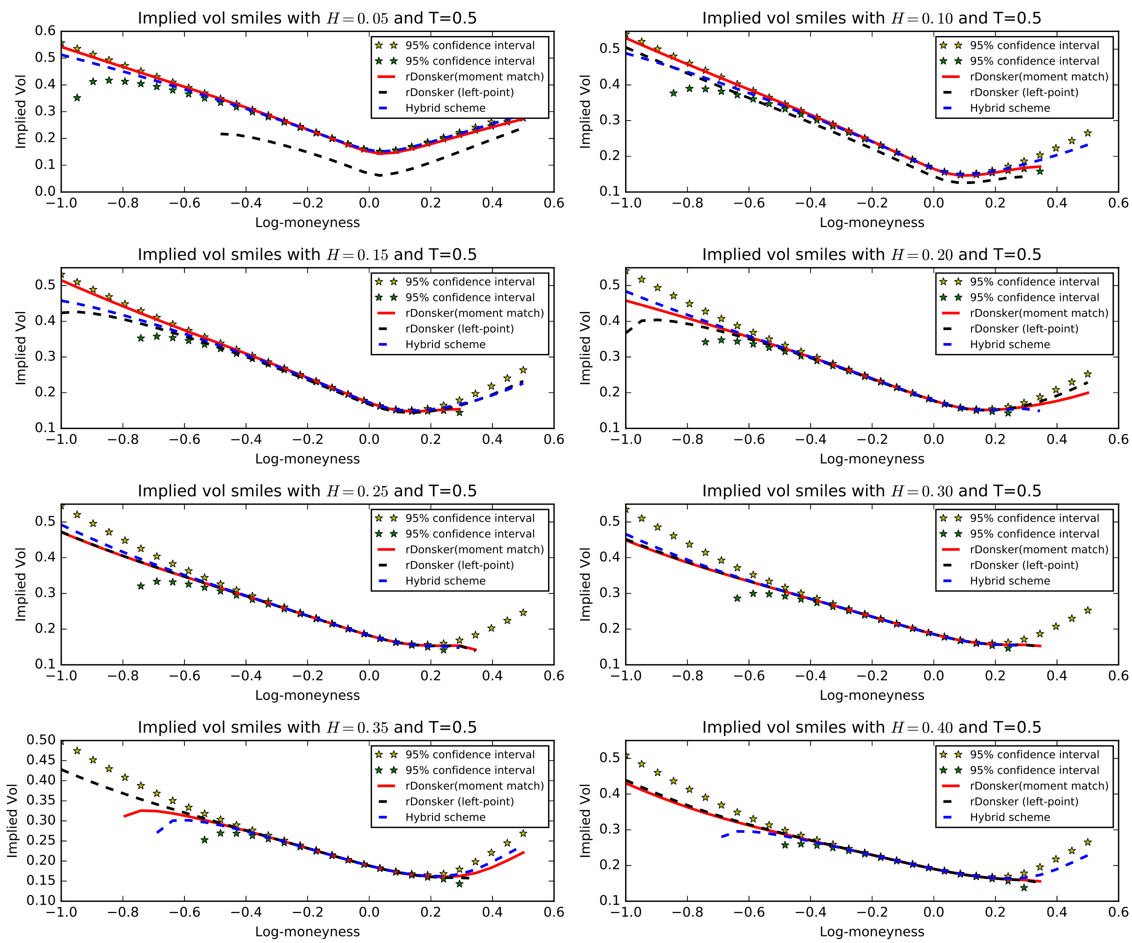

FIGURE 2. Implied volatilities of rDonsker with left-point and variance matching, and in the Hybrid scheme with $5 \cdot 10^5$ simulations. Conditional expectation and antithetic variates where used in both methods.

3.4. **Numerical example: Rough Bergomi model.** Figure 2 shows implied volatilities for different values of H, when using left-point and moment matching optimal evaluation point in the rDonsker scheme and also the Hybrid scheme. In Figures 3 and 4, we give a exhaustive comparison analysis of the errors when using a left-point evaluation and moment matching optimal evaluation. It is obvious from Figure 3 that as H tends to

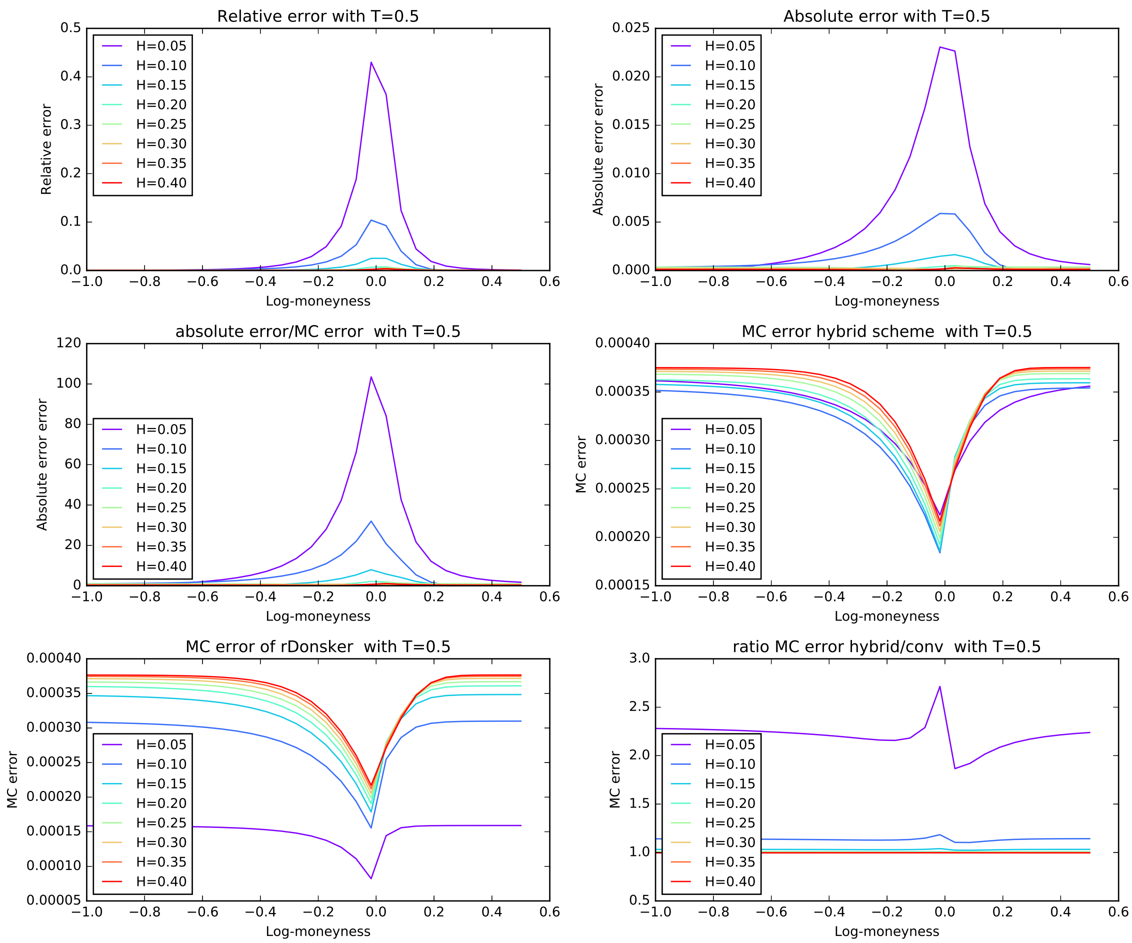

FIGURE 3. Monte-Carlo errors between left-point rDonsker and the Hybrid scheme with $5 \cdot 10^5$ simulations. Conditional expectation and antithetic variates where used in both methods.

zero, the left-point rDonsker converges too slowly to the required output as opposed to the Hybrid scheme, which was show in [11] to converge to the output regardless of H and the discretisation grid. This phenomenon is not surprising, since we already discussed that the rate of convergence of the rDonsker scheme is of order $\mathcal{O}(n^{-H}$. Nevertheless, for $H > 0.15$ there is no significant difference between both schemes. In particular, we notice that the biggest error for the rDonsker scheme happens when the options is around-the-money. Now, in Figure 4 we observe how the optimal evaluation point improves substantially the performance of the rDonsker scheme. The relative error and absolute errors are reduced by a factor of 10 when $H = 0.05$ is very small. This maintains the relative error below 4% for $H \geq 0.05$. Specifically, it is worth noticing that from Figure 3 to Figure 4 the behaviour of the Monte-Carlo error when $H = 0.05$ dramatically changes when using the optimal evaluation point, becoming more similar to the Hybrid scheme.

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 ones we simulate the rBergomi model [8], whereas for the later we use the classical Bergomi [13] model using a forward Euler scheme

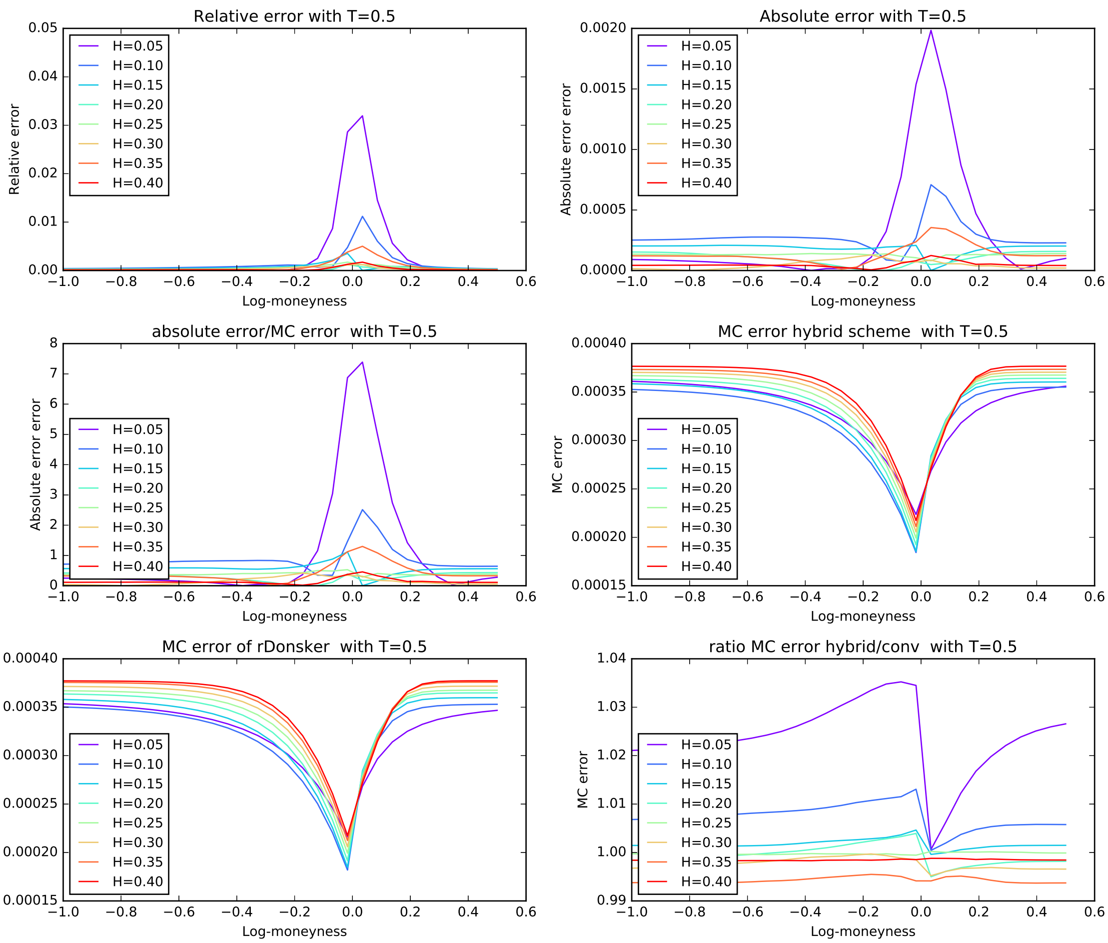

Figure 4. Monte-Carlo errors between rDonsker using moment matching evaluation and the Hybrid scheme with $5 \cdot 10^5$ simulations. Conditional expectation and antithetic variates where used in both methods.

in both volatility and stock price. All three schemes are implemented in Cython to make the comparisson fair and to obtain C++ like speeds. Figure 5 shows that rDonsker is approximately 2 times slower than the Markovian case whereas the Hybrid scheme is approximately 2.5 times slower. This is of course 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 $n << 10000$ is relatively small. Figure 5 also proves 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.** Based on the numerical analysis above, we suggest the following 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:

Regarding empirical estimates, Gatheral, Jaisson and Rosenbaum [35] suggest that $H \approx 0.15$. Bennedsen, Lunde and Pakkanen [12] give an exhaustive analysis of more than 2000 equities for which $H \in [0.05, 0.2]$. On

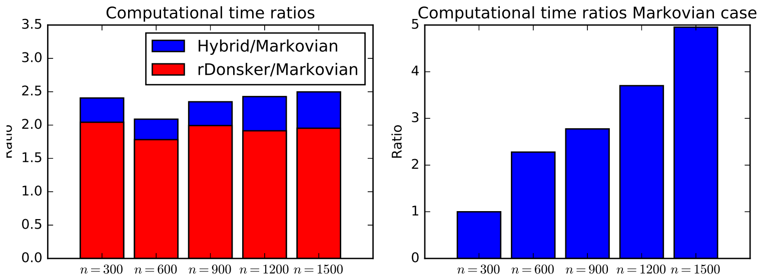

FIGURE 5. Computational time benchmark using Hybrid scheme, rDonsker and Markovian(forward Euler) for different grid sizes n.

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

the pricing side, Bayer, Friz and Gatheral [8] and Jacquier, Martini and Muguruza [46] found that calibration routines yield $H \in [0.05, 0.10]$. Finally, Livieri, Mouti, Pallavicini and Rosenbaum [56] 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 3.6. The rough Heston model presented by Guennoun, Jacquier, Roome and Shi [35] 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 provide 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.

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

$$
\begin{aligned}
B_n(t_i) &= \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \left(\rho\xi_k + \overline{\rho}\zeta_k\right), \\
Y_n(t_i) &= \frac{T}{n} \sum_{k=1}^{i} b\left(Y_n(t_{k-1})\right) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sigma\left(Y_n(t_{k-1})\right)\xi_k,
\end{aligned}
$$

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

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

In order to construct the tree we have to consider all possible permutations of the random vectors $\{\xi_i\}$ and $\{\zeta_i\}$. 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 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 [73] and the following representation by El Karoui [27] ($\widetilde{\mathbb{I}}$ denotes the set of stopping times with values in $\mathbb{I}$):

Definition 3.7. 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 $\mathscr{J}$ of X is defined as $\mathscr{J}(X)(t) := \operatorname{ess\,sup}_{\tau\in\widetilde{\mathbb{I}}} \mathbb{E}(X_\tau|\mathcal{F}_t)$ 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 [48], an American option is nothing else than the smallest supermartingale dominating its European counterpart:

Definition 3.8. Let $C^e_t(k, T)$ and $P^e_t(k, T)$ denote European Call and Put prices at time t, with log-strike k and maturity T. Then the American counterparts, $C^a_t(k, T)$ and $P^a_t(k, T)$, are given by

$$C^a_t(k, T) = \mathscr{J}(C^e(k, T))(t) \qquad \text{and} \qquad P^a_t(k, T) = \mathscr{J}(P^e(k, T))(t).$$

The most general result regarding the preservation of weak convergence under the Snell envelope map is due to Mulinacci and Pratelli [62], who proved that convergence takes place in the Skorohod topology only if the Snell envelope is continuous. In our setting, the scheme for American options is fully justified by the following theorem:

Theorem 3.9. *For V in (1.3), if $\mathbb{E}\left\{\exp\left(\int_0^t V_s \mathrm{d}s\right)\right\}$ is finite, then $(\mathscr{J}(X_n))_{n\geq1}$ converges weakly to $\mathscr{J}(X)$ in the Skorokhod topology.*

Proof. Since the sequence $(X_n)_{n\geq1}$ converges weakly to X in $(\mathcal{D}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$, for X in (1.3), the theorem follows using the Continuous Mapping Theorem if we can show that $\mathscr{J}$ is continuous. El Karoui proved [27] proved 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. Using the de la Vallée-Poussin theorem, for any $t \in \mathbb{I}$,

$$\mathbb{E}\left(\mathrm{e}^{2X_t}\right) = \mathbb{E}\left[\exp\left(-\int_0^t V_s \mathrm{d}s + 2\int_0^t \sqrt{V_s}\mathrm{d}B_s\right)\right] = \mathbb{E}\left[\mathcal{E}\left(2\int_0^t \sqrt{V_s}\mathrm{d}B_s\right)\exp\left(3\int_0^t V_s \mathrm{d}s\right)\right].$$

With $\mathcal{F}^W_t := \sigma(W_s : s \leq t)$ the filtration generated by W (or equivalently by V), the tower property yields

$$\mathbb{E}\left(\mathrm{e}^{2X_t}\right) = \mathbb{E}\left[\mathbb{E}\left(\mathrm{e}^{2X_t}|\mathcal{F}^W_t\right)\right] = \mathbb{E}\left[\exp\left(3\int_0^t V_s \mathrm{d}s\right)\right],$$

by the martingale property of $\mathcal{E}\left(2\int_0^t \sqrt{V_s}\mathrm{d}B_s\right)|\mathcal{F}^W_t$. Therefore $\exp(X)$ is a uniformly integrable martingale, and so is $\exp(X_{\tau\wedge\cdot})$ by Doob's optimal stopping theorem, and the proposition follows. □

Corollary 3.10. *Theorem 3.9 also holds under the stronger condition $\mathbb{E}\left(\mathrm{e}^{V_t}\right) < \infty$ for all $t \in \mathbb{I}$.*

Proof. Jensen's inequality implies that

$$\mathbb{E}\left[\exp\left(\int_0^t V_s \mathrm{d}s\right)\right] \leq \mathbb{E}\left(C_t \int_0^t \mathrm{e}^{V_s}\mathrm{d}s\right) = C_t \int_0^t \mathbb{E}\left(\mathrm{e}^{V_s}\right)\mathrm{d}s$$

for some constant $C_t > 0$, and the right-hand side is finite if $\mathbb{E}\left(\mathrm{e}^{V_s}\right)$ is and the proposition follows. □

Mulinacci and Pratelli [62] also gave explicit conditions for the 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 rates and dividend yields, denote by r and 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

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

where $\mathbb{E}(\cdot|\mathcal{F}_{t_i}) = \frac{1}{2}\left(\widetilde{h}^{\mathrm{up}}_{t_{i+1}} + \widetilde{h}^{\mathrm{down}}_{t_{i+1}}\right)$ with $\widetilde{h}^{\mathrm{up}}$ and $\widetilde{h}^{\mathrm{down}}$ being the adjacent nodes;
(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.* We construct a rough volatility tree for the rough Bergomi model [8] and check the accuracy of the scheme. Figures 6 and 7 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 8 displays implied 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 9 a detailed error analysis corroborates these observations: the relative error is smaller than 3% for $H \geq 0.15$.

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 10 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.

Appendix A. Riemann-Riouville operators

We review here fractional operators and their mapping properties. We follow closely the excellent monograph by Samko, Kilbas and Marichev [72], as well as some classical results by Hardy and Littlewood [38].

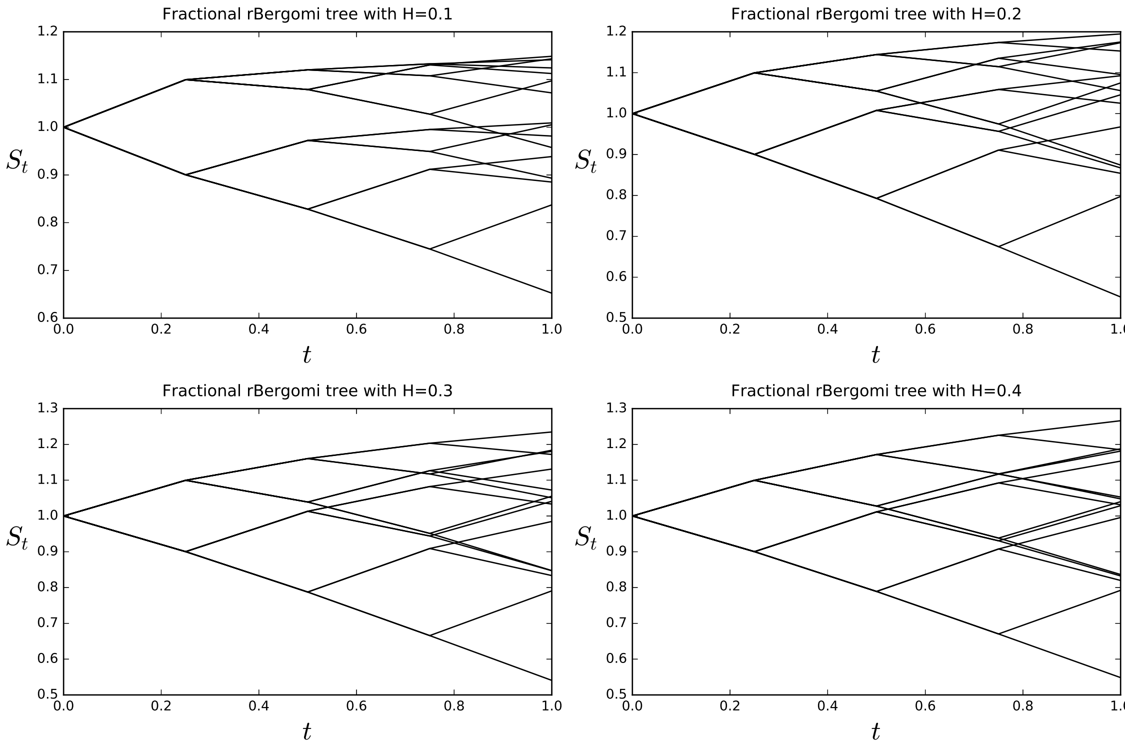

FIGURE 6. rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 5 time steps.

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

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

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

Theorem A.2. *For any $f \in \mathcal{C}^\lambda(\mathbb{I})$, with $\lambda \in (0,1)$ and $\alpha \in (0,1)$, the identity*

$$(I^\alpha f)(t) = \frac{f(0)}{\Gamma(1+\alpha)} t^\alpha + \psi(t),$$

holds for all $t \in \mathbb{I}$, for some $\psi \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$ satisfying $|\psi(t)| \le C t^{\lambda+\alpha}$ on $\mathbb{I}$ for some $C > 0$.

Proof. We may easily represent

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

with $\psi(t) := \dfrac{1}{\Gamma(\alpha)} \displaystyle\int_0^t \frac{f(u)-f(0)}{(t-u)^{1-\alpha}} \mathrm{d}u$. Since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we obtain $|\psi(t)| \le \dfrac{|f|_\lambda}{\Gamma(\alpha)} \displaystyle\int_0^t \frac{u^\lambda}{(t-u)^{1-\alpha}} \mathrm{d}u$, and hence

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

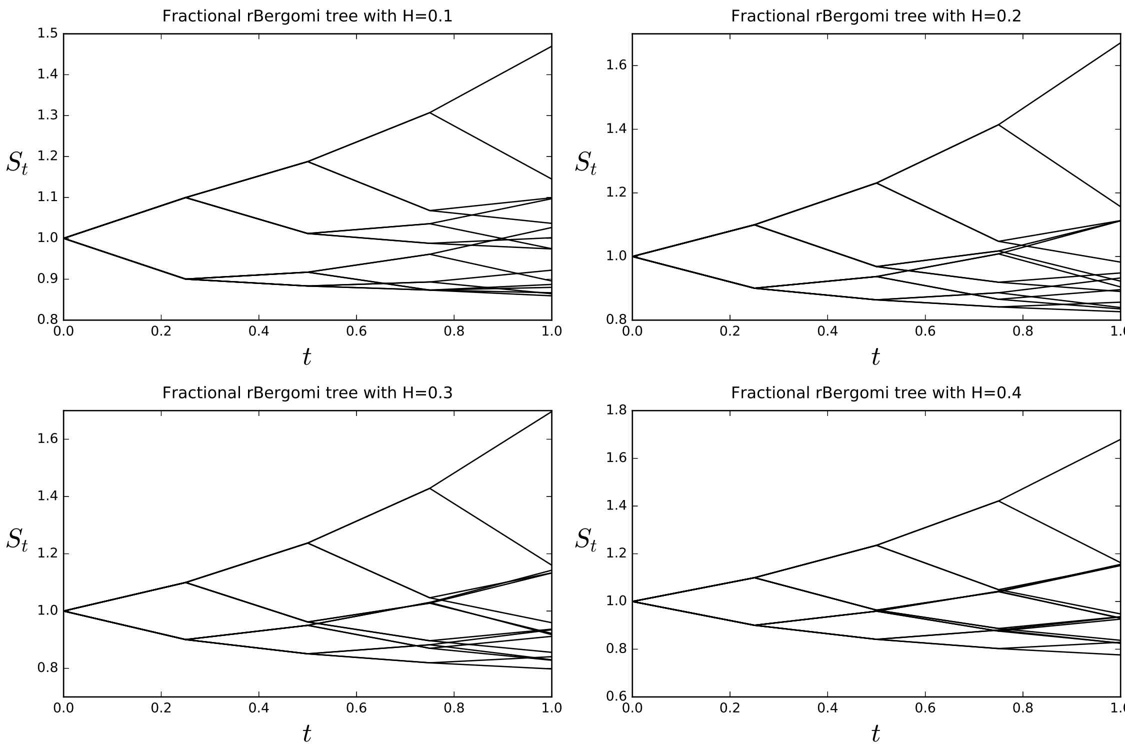

FIGURE 7. rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, 1, 0.04)$ with 5 time steps.

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

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

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

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

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

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

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

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

Finally,

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

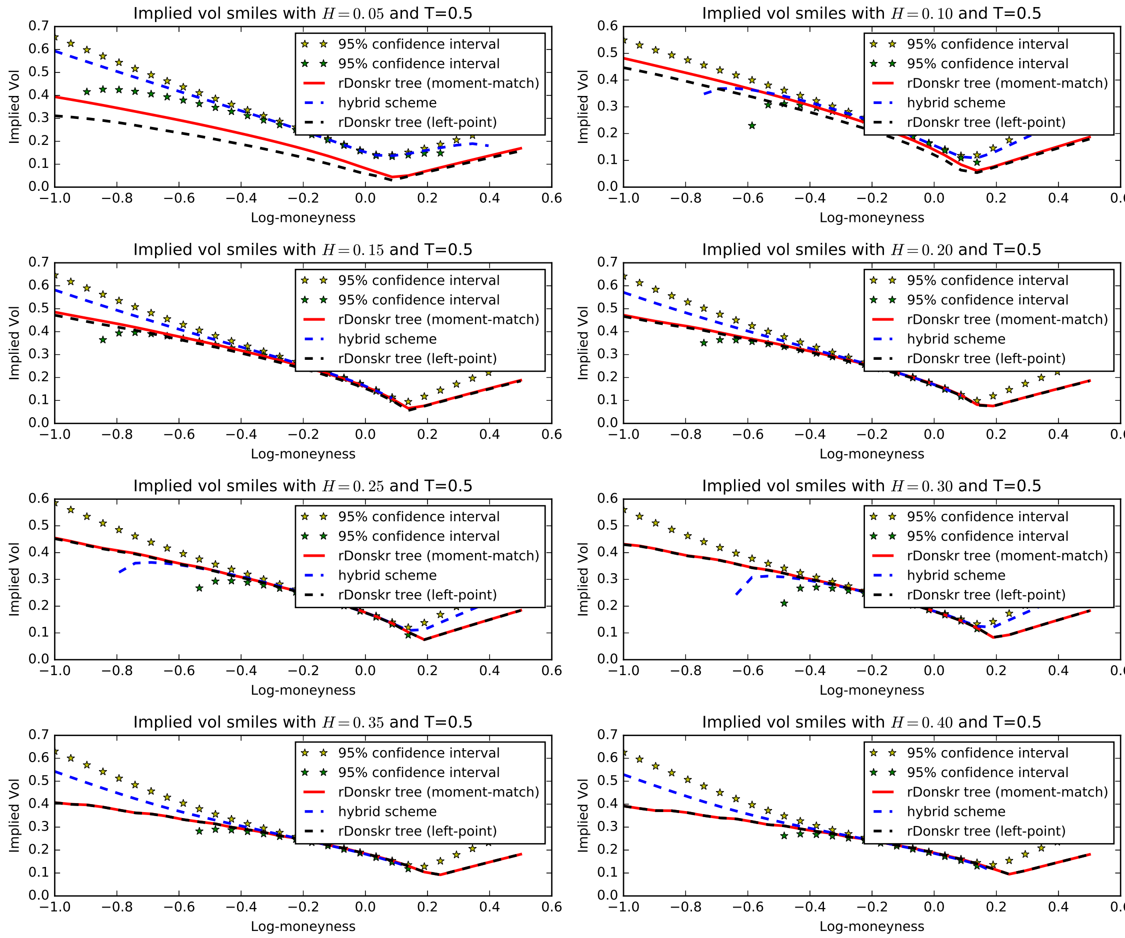

FIGURE 8. rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps.

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

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

Thus ψ satisfies the $(\lambda+\alpha)$-Hölder condition and belongs to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. □

Corollary A.3. *For any $\alpha, \lambda \in (0,1)$ such that $\lambda + \alpha \le 1$, 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. Using the estimate in Theorem A.2 we have

$$\|I^\alpha f\|_{\alpha+\lambda} \le \frac{f(0)}{\Gamma(1+\alpha)}\|(\cdot)^\alpha\|_{\lambda+\alpha} + \|\psi\|_{\lambda+\alpha} \le C_1\|f\|_\lambda\|(\cdot)^\alpha\|_{\lambda+\alpha} + C_2\|f\|_\lambda\|(\cdot)^{\alpha+\lambda}\|_{\lambda+\alpha} \le C\|f\|_\lambda,$$

since $|f|_\lambda \le \|f\|_\lambda$, $f(0) \le \|f\|_\lambda$. Therefore I^α is also bounded and hence continuous. □

Theorem A.4. *For any $0 < -\alpha < \lambda \le 1$, let $f \in \mathcal{C}^\lambda(\mathbb{I})$. Then $I^\alpha f$ exists, $I^{-\alpha}I^\alpha f = f$ and, for all $t \in \mathbb{I}$,*

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

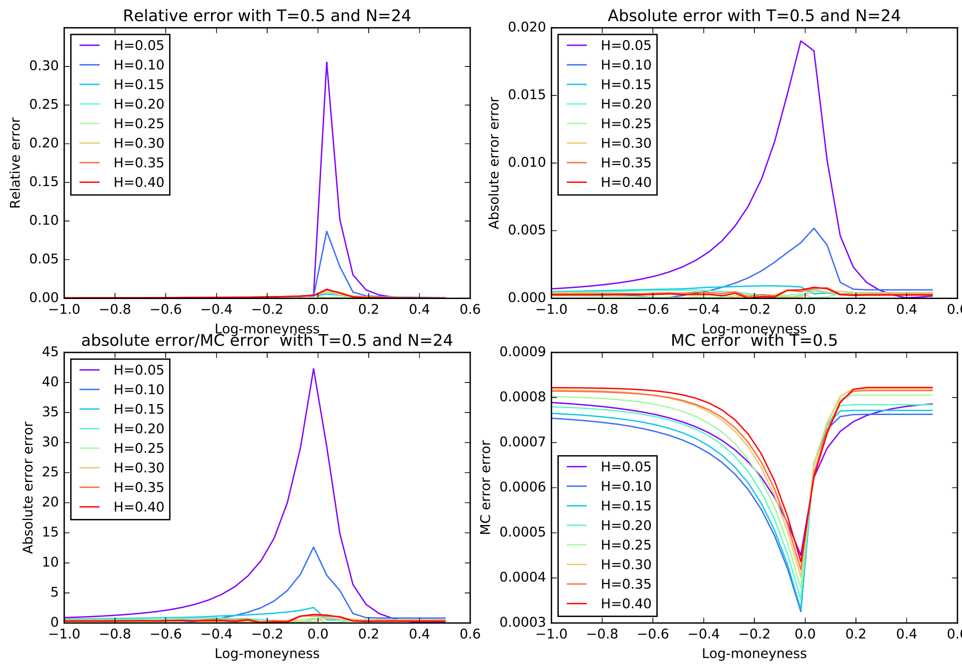

FIGURE 9. error analysis for the rDonsker moment-match tree for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps.

Proof. For $f \in \mathcal{C}^\lambda(\mathbb{I})$, define, for any $\varepsilon > 0$ and $t \in \mathbb{I}$,

$$(I_\varepsilon^{1+\alpha} f)(t) := \frac{1}{\Gamma(\alpha+1)} \int_0^{t-\varepsilon} (t-u)^\alpha f(u) \mathrm{d}u,$$

and note that $I_0^{1+\alpha} = I^{1+\alpha}$. Then, we have

$$\begin{aligned} \Gamma(1+\alpha) \left(\frac{\mathrm{d}}{\mathrm{d}t} I_\varepsilon^{1+\alpha} f \right)(t) &= \varepsilon^\alpha f(t-\varepsilon) + \alpha \int_0^{t-\varepsilon} (t-u)^{\alpha-1} f(u) \mathrm{d}u \\ &= -\alpha \int_0^{t-\varepsilon} (t-u)^{\alpha-1} (f(t) - f(u)) \mathrm{d}u - \varepsilon^\alpha (f(t) - f(t-\varepsilon)). \end{aligned} \tag{A.2}$$

where Hölder continuity implies that $f(t) - f(u) \leq C(t-u)^\lambda$, so that the integral in (A.2) is well defined. Then, as ε tends to zero, the right-hand side of (A.2) tends uniformly to

$$\psi(t) = -\alpha \int_0^t (t-u)^{\alpha-1} (f(t) - f(u)) \mathrm{d}u.$$

Now, for $t \in \mathbb{I}$,

$$\begin{aligned} (I^{1+\alpha} f)(t) - (I^{1+\alpha} f)(0) &= \lim_{\varepsilon \downarrow 0} \left\{ (I_\varepsilon^{1+\alpha} f)(t) - (I_\varepsilon^{1+\alpha} f)(0) \right\} = \lim_{\varepsilon \downarrow 0} \int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}u} I_\varepsilon^{1+\alpha} f \right)(u) \mathrm{d}u \\ &= \int_0^t \lim_{\varepsilon \downarrow 0} \left(\frac{\mathrm{d}}{\mathrm{d}u} I_\varepsilon^{1+\alpha} f \right)(u) \mathrm{d}u = \frac{1}{\Gamma(1+\alpha)} \int_0^t \psi(u) \mathrm{d}u, \end{aligned}$$

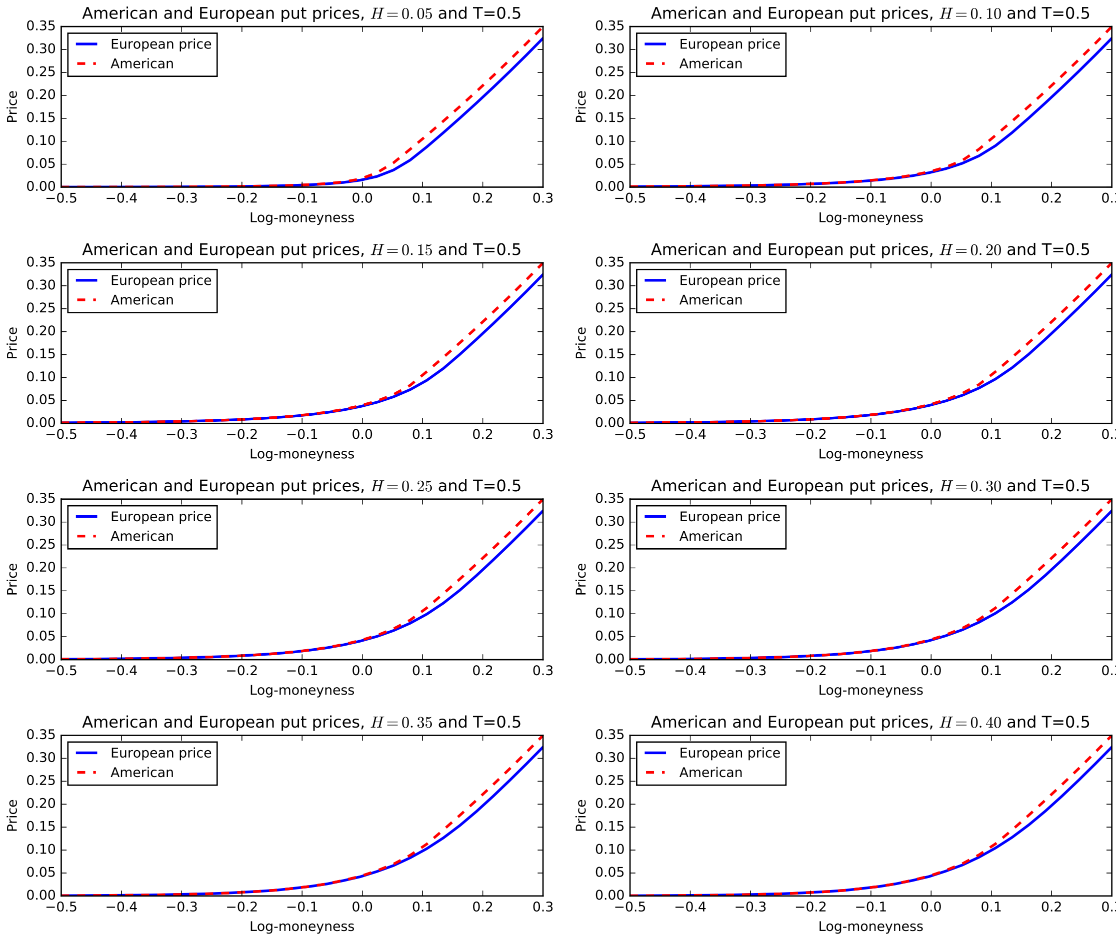

FIGURE 10. rough Bergomi American and European Put prices for different values of H and $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 26 time steps.

where the exchange of limit and integral holds since the convergence is uniform and the interval compact. Therefore, $\Gamma(\alpha+1)(I^{1+\alpha}f)$ is the integral of ψ and, by the Fundamental Theorem of Calculus,

$$\psi(t) = \Gamma(\alpha+1)\left(\frac{\mathrm{d}}{\mathrm{d}t}I^{1+\alpha}f\right)(t) = \Gamma(\alpha+1)(I^{\alpha}f)(t).$$

Therefore it exists and, similarly to Theorem A.2, $\psi \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. Finally, since, for $0<\beta<1$, the equality

$$(I^{\beta}I^{1-\beta}f)(t) = (I^{1-\beta}I^{\beta}f)(t) = (I^{1}f)(t) = \int_0^t f(u)\mathrm{d}u$$

holds for all $t \in \mathbb{I}$, we conclude that

$$\Big(\Gamma(1+\alpha)I^{1+\alpha}f - I^{-\alpha}\psi\Big)(t) = \int_0^t \Gamma(1+\alpha)f(u) - (I^{-\alpha}\psi)(u))(t-u)^{\alpha}\mathrm{d}u = 0,$$

and hence, by continuity of both f and $I^{-\alpha}\psi$, $f = I^{-\alpha}I^{\alpha}f$. $\square$

Appendix B. Discrete convolution

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

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

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

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

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

$$\begin{pmatrix} g(t_1) & 0 & \cdots & 0 \\ g(t_2) & g(t_1) & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ g(t_n) & g(t_{n-1}) & \cdots & g(t_1) \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix},$$

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

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

and the Inverse DFT of c is given by

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

In general, both transforms require a computational effort of order $\mathcal{O}(n^2)$, but the the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey [18] exploit 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* $(\mathrm{a} * \mathrm{b}) = f\left(\widehat{f}(\mathrm{a}) \bullet \widehat{f}(\mathrm{b})\right)$ *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 $\{\xi_{j,i}\}_{\substack{j=1,\ldots,M \\ i=1,\ldots,n}}$ such that $\mathbb{V}(\xi_{j,i}) = 1$;
(2) define the vectors $\mathfrak{g} := (g(t_i))_{i=1,\ldots,n}$ and $\xi_j := (\xi_{j,i})_{i=1,\ldots,n}$, for $j = 1, \ldots, M$;
(3) using FFT, compute $\varphi_j := \widehat{f}(\mathfrak{g}) \cdot \widehat{f}(\xi_j)$, for $j = 1, \ldots, M$;
(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 $j = 1, \ldots, M$.

In Step 2 we may replace the evaluation points $\mathfrak{g}$ by any optimal evaluation point $\{g(t_i^*)\}_{i=1}^n$ as in (3.3). Many numerical packages offer a direct implementation of the discrete convolution such as the numpy.convolve function in the NumPy library of Python. The user then only needs to pass the arguments $\mathfrak{g}$ and ξ_j to this function and Steps 3 and 4 are computed automatically (using efficient FFT techniques) by the package. 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, i.e. $n << 1000$. Hence, it is not important to have the fastest possible FFT for very large n, it is much more important for the implementation to be fast on small time grids. In this aspect we find that numpy.convolve is a very competitive implementation.

Appendix C. Additional Proofs

C.1. **Proof of Proposition 1.2.** Since $g \in \mathcal{L}^\alpha$, there exists $C > 0$ such that $|g(u)| \le Cu^\alpha$; hence, for $t \in \mathbb{I}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t |f(s)g(t-s)|\mathrm{d}s \le C\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t |f(s)(t-s)^\alpha|\mathrm{d}s.$$

Therefore, for $f \in \mathcal{C}^\lambda(\mathbb{I})$, the inequalities involving the Riemann-Liouville fractional operator (Appendix A)

$$(\mathcal{G}^\alpha f)(t) \le C(I^\alpha f)(t) \le C\|f\|_\lambda \tag{C.1}$$

hold for $\alpha \le 0$ and all $t \in \mathbb{I}$. Since Riemann-Liouville operators are continuous (Appendix A), continuity of the GFO follows directly from (C.1) along with linearity. To prove that $\mathcal{G}^\alpha$ belongs to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$, we may invoke (C.1) and easily adapt Theorem A.4. Similarly, when $\alpha \ge 0$, for any $u \in \mathbb{I}$, $g'(u) = u^\alpha L'(u) + u^{\alpha-1}L(u) \le C_1 + C_2 u^{\alpha-1}$, and the λ-Hölder continuity of f yields, for any $t \in \mathbb{I}$,

$$\int_0^t \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)f(s)\mathrm{d}s \le \frac{C_1}{\lambda+1}t^{\lambda+1} + C_2\int_0^t (t-s)^{\alpha-1}f(s)\mathrm{d}s \le C_1 + C_2\int_0^t (t-s)^{\alpha-1}f(s)\mathrm{d}s.$$

Since the time horizon $\mathbb{I}$ is compact, the first constant does not affect continuity or mapping properties of the GFO. The second term is bounded by the Riemann-Liouville integral operator (Appendix A), hence continuity and mapping properties follow as before by straightforward modification of Theorem A.2.

C.2. **Proof of Proposition 1.4.** Since the paths of Brownian motion are 1/2-Hölder continuous, existence (and continuity) of $\mathcal{G}^\alpha W$ is guaranteed for all $\alpha \in (-1/2, 1/2)$. When $\alpha \in [0, 1/2)$, the kernel is smooth and square integrable, so that Itô's product rule yields (since $g(0) = 0$)

$$(\mathcal{G}^\alpha W)(t) = \int_0^t \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)W(s)\mathrm{d}s = \int_0^t g(t-s)\mathrm{d}W_s,$$

and the claim holds. For $\alpha \in (-1/2, 0)$, and any $\varepsilon > 0$, introduce the operator

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

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

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}_\varepsilon^{1+\alpha} W\right)(t) = g(\varepsilon)W(t-\varepsilon) + \int_0^{t-\varepsilon} \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)W(s)\mathrm{d}s = \int_0^{t-\varepsilon} g(t-s)\mathrm{d}W_s. \tag{C.2}$$

Then, as ε tends to zero, the right-hand side of (C.2) tends to $\int_0^t g(t-s)\mathrm{d}W_s$, and furthermore, the convergence is uniform. On the other hand, the equalities

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

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

References

[1] M. Abramowitz and I.A. Stegun. Handbook of mathematical functions. *National Bureau of Standards*, 1965.

[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] 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.

[5] 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.

[6] 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.

[7] C. Bayer, P. Friz, P. Gassiat, J. Martin and B. Stemper. A regularity structure for rough volatility. arXiv:1710.07481, 2017.

[8] C. Bayer, P. Friz and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6): 1-18, 2015.

[9] C. Bayer, P. Friz, A. Gulisashvili, B. Horvath and B. Stemper. Short-time near the money skew in rough fractional stochastic volatility models. arXiv:1703.05132, 2017.

[10] 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.

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

[12] M. Bennedsen, A. Lunde and M.S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. arXiv:1610.00332, 2017.

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

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

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

[16] 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.

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

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

[19] 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.

[20] 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.

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

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

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

[24] R.M. Dudley. Uniform central limit theorems. Cambridge University Press, 1999.

[25] P. Embrechts and M. Maejima. An introduction to the theory of self-similar stochastic processes. *International Journal of Modern Physics B*, 14(12-13), 2000.

[26] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. To appear in *Mathematical Finance*.

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

[28] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8: 114-145, 2017.

[29] P.K. Friz and M. Hairer. A course on rough paths, with an introduction to regularity structures. *Springer*, 2014.

[30] P.K. Friz and N. Victoir. Multidimensional stochastic processes as rough paths: theory and applications. *Cambridge Studies in Advanced Mathematics ed. Cambridge University Press*, 2010.

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

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

[33] J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is rough. To appear in *Quantitative Finance*.

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

[35] H. Guennoun, A. Jacquier, P. Roome and F. Shi. Asymptotics of the fractional Heston model. arXiv:1411.7653, 2017.

[36] A. Gulisashvili. Large deviation principle for Volterra type fractional stochastic volatility models. arXiv:1710.10711, 2017.

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

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

[39] C. Heinrich, M. Pakkanen and A.E.D. Veraart. Hybrid simulation scheme for volatility modulated moving average fields. arXiv:1709.01310, 2017.

[40] B. Horvath, A. Jacquier and C. Lacombe. Asymptotic behaviour of randomised fractional volatility models. arXiv:1708.01121, 2017.

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

[42] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of American Society of Civil Engineers. 116: 770*, 1951.

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

[44] K. Itô. *Preceedings of the Imperial Academy*, 20: 54-55, 1944.

[45] 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.

[46] A. Jacquier, C. Martini and A. Muguruza. On VIX futures in the rough Bergomi model. Forthcoming in *Quant, Finance*, 2017.

[47] A. Jacquier, M. Pakkanen and H. Stone. Pathwise large deviations for the Rough Bergomi model. arXiv:1706.05291, 2017.

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

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

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

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

[52] 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, 1995.

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

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

[55] J. Lamperti. On convergence of stochastic processes. *Transactions of the AMS*, 104: 430-435, 1965.

[56] G. Livieri, S. Mouti, A. Pallavicini, M. Rosenbaum. Rough volatility: evidence from option prices. arXiv:1702.02777, 2017.

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

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

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

[60] R. McCrickerd and M.S. Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. arXiv:1708.02563, 2017.

[61] R.E. Megginson. An introduction to Banach spaces. *Graduate Texts in Mathematics*, 108, Springer-Verlag, 1998.

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

[63] L. Mytnik and E. Neuman. Sample path properties of Volterra processes. *Communications on Stochastic Analysis*, 6(3): 359-377, 2012.

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

[65] E. Neuman and M. Rosenbaum. Fractional Brownian motion with zero Hurst parameter: a rough volatility viewpoint. arXiv:1711.00427, 2017.

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

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

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

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

[70] 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.

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

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

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

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

[75] L.C.G. Rogers and D. Williams. Diffusions, Markov processes and martingales: Volume 2, Itô Calculus. Cambridge Mathematical Library, 2000.

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

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

[78] 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, Imperial College London
E-mail address: b.horvath@imperial.ac.uk

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

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