HAL open science

From rough to multifractal volatility: The log S-fBM model

Peng Wu, Jean-François Muzy, Emmanuel Bacry

To cite this version:

Peng Wu, Jean-François Muzy, Emmanuel Bacry. From rough to multifractal volatility: The log S-fBM model. Physica A: Statistical Mechanics and its Applications, 2022, 604, pp.127919. 10.1016/j.physa.2022.127919 . hal-03852446

HAL Id: hal-03852446

https://hal.science/hal-03852446v1

Submitted on 15 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:2201.09516v2 [q-fin.ST] 17 Jul 2022

From Rough to Multifractal volatility: the log S-fBM model

Peng Wu[1,*], Jean-François Muzy[2], and Emmanuel Bacry[1]

[1]Ceremade, CNRS-UMR 7534, Université Paris-Dauphine PSL
Place du Maréchal de Lattre de Tassigny, 75016 Paris, France
[2]SPE CNRS-UMR 6134, Université de Corse
BP 52, 20250 Corte, France
[*]Corresponding author, peng.wu@dauphine.eu
Emails: [2]muzy@univ-corse.fr, [1]bacry@ceremade.dauphine.fr

July 19, 2022

Abstract

We introduce a family of random measures $M_{H,T}(\,\mathrm{d}t)$, namely log S-fBM, such that, for $H > 0$, $M_{H,T}(\,\mathrm{d}t) = e^{\omega_{H,T}(t)}\,\mathrm{d}t$ where $\omega_{H,T}(t)$ is a Gaussian process that can be considered as a stationary version of an H-fractional Brownian motion. Moreover, when $H \to 0$, one has $M_{H,T}(\,\mathrm{d}t) \to \widetilde{M}_T(\,\mathrm{d}t)$ (in the weak sense) where $\widetilde{M}_T(\,\mathrm{d}t)$ is the celebrated log-normal multifractal random measure (MRM). Thus, this model allows us to consider, within the same framework, the two popular classes of multifractal ($H = 0$) and rough volatility ($0 < H < 1/2$) models. The main properties of the log S-fBM are discussed and their estimation issues are addressed. We notably show that the direct estimation of H from the scaling properties of $\ln(M_{H,T}([t, t+\tau]))$, at fixed τ, can lead to strongly over-estimating the value of H. We propose a better GMM estimation method which is shown to be valid in the high-frequency asymptotic regime. When applied to a large set of empirical volatility data, we observe that stock indices have values around $H = 0.1$ while individual stocks are characterized by values of H that can be very close to 0 and thus well described by a MRM. We also bring evidence that unlike the log-volatility variance ν^2 whose estimation appears to be poorly reliable (though used widely in the rough volatility literature), the estimation of the so-called "intermittency coefficient" λ^2, which is the product of ν^2 and the Hurst exponent H, appears to be far more reliable leading to values that seem to be universal for respectively all individual stocks and all stock indices.

Keywords— Rough volatility, Multifractal volatility, fractional Brownian motion, GMM estimation, Intermittency coefficient

1 Introduction

During the past few years, new insights on stochastic volatility models have been obtained after the observation by Gatheral et al. [16] (see also [8, 9]), that the logarithm of the realized volatility is rough, i.e., is less regular than a standard Brownian motion. Rough volatility models have become very popular not only because they allow one to account for main empirical realized volatility properties but also because, when they are considered in asset price models, they provide a very good fit of option prices and notably their ATM skew power-law behavior close to maturity [20, 6, 11]. The first empirical evidence reported in [16] suggests that the logarithm of the asset price stochastic variance can be represented by a fractional Brownian motion (fBM) of Hurst exponent H close to

$H \simeq 0.1 < 1/2$. More recent studies based either on quasi-likelihood approach [14] or GMM-approach [10], consistently suggest the H is even closer to $H = 0$, i.e., $H \lesssim 0.05$ for a large panel of equity data. In that respect, it is natural to consider the limit $H \to 0$ in the rough process driving the volatility logarithm. Even if one cannot plug $H = 0$ in the power-law expression of the fractional Brownian motion covariance, formally, it corresponds to a logarithmic behavior.

Such a logarithmic behavior is precisely the one that characterises the so-called continuous random cascade models introduced two decades ago by Bacry et al. [26, 1]. Indeed, in 2000, these authors proposed the "Multifractal Random Walk" (MRW) as a model for asset prices in order to account for their multifractal properties, i.e., the fact observed by various authors (see e.g. [18, 17, 21]) that asset return empirical moments obey non-trivial scaling properties. The MRW model relies on a multifractal stochastic volatility model, namely the "Multifractal Random Measure" (MRM) model [26, 1], in which the log-volatility is provided by a log-correlated Gaussian field. Such a class of processes, also referred to as Gaussian multiplicative chaos, has been at the heart of many studies in a large variety of applications [29]. Gaussian multiplicative chaos and the associated log-normal random cascades have been extended to any infinitely divisible distribution by Bacry and Muzy in [24, 4].

Recovering a multifractal volatility model as the limit $H \to 0$ of a rough volatility model or, from a more general perspective, defining a meaningful limit $H \to 0$ of a fractional Brownian motion and one of its variants has been the subject of various recent studies. In [15], the authors build an $H{=}0$ - fBM by considering a regularisation from the harmonizable representation of fBM's while in [27, 19] a $H = 0$ limiting process is obtained using a peculiar normalisation and centering of the fBM. In [13] (see also [12]), the authors consider the limit $H \to 0$ of the exponential of a rescaled Riemann-Liouville fBM and its relationship with Gaussian multiplicative chaos. Finally, in [7], Bayer et al. propose a new class of rough models that consists in modulating the Riemann-Liouville fBM power-law kernel by a logarithmic factor. The so-obtained "super-rough" stochastic volatility remains well-defined as a continuous process when $H = 0$.

In this paper, our goal is to add a contribution to this problem by introducing a new version of rough volatility models based on the so-called "stationary" fBM (S-fBM). S-fBM is a variant of fBM whose covariance function is exactly the one obtained when considering the small-time approximation of the correlation of the fractional Ornstein-Uhlenbeck process considered in [16]. We prove that when $H \to 0$, one recovers the exact self-similar multifractal measure defined in [26, 1]. Our construction is based on the same approach proposed in [24, 4] where the log-volatility is obtained from the integration of a 2D Gaussian white noise over a triangular domain in a time-scale plane. It turns out such an approach corresponds to the same method defined by Takenaka to build the fractional Brownian motion [30]. Our model therefore provides a unified framework to consider both rough and multifractal stochastic volatility models. Beyond defining the main statistical properties of the model, we aim at estimating its parameters on a large panel of market data. For that purpose, we extend the GMM method proposed in [3] that is based on a "small intermittency" expansion of the moments of the measure logarithms.

The paper is organized as follows: in section 2, after recalling the basic notions underlying usual rough volatility models and the definition of the multifractal random measure (MRM), we introduce the log S-fBM random measure $M_{H,T}(\,\mathrm{d}t)$ as the exponential of the S-fBM random process which is nothing but a "stationary" version of the fractional Brownian motion of Hurst parameter $0 < H < 1$. We show that one recovers the celebrated Mandelbrot-Van Ness fBM when T, the correlation parameter of our model, tends to infinity. In this section, we also show that the log S-fBM converges, when $H \to 0$, towards a Multifractal Random Measure, consequently leading to a unified framework for rough volatility models ($M_{H,T}(\,\mathrm{d}t)$, for $H \neq 0$) and multifractal volatility models (by extension, $M_{H=0,T}(\,\mathrm{d}t)$). In section 3, we establish, within this unified framework, analytical expressions for the second-order moments of respectively $M_{H,T}([t, t+\Delta])$ and its logarithm, while in section 4 we define two GMM parameter estimation methods based on these expressions. Our approach is illustrated by various numerical examples. Application to empirical data, namely the daily volatility of many individual stocks as well as market indices is provided in section 5. Section 6 summaries our findings while technical material and mathematical proofs are provided in Appendices.

2 The log Stationary fractional Brownian Motion (log S-fBM) stochastic volatility model

2.1 Multifractal and rough volatility models

Before introducing our new model of stochastic volatility measure (log S-fBM), let us briefly walk through the two popular former classes of stochastic volatility models it is notably designed to unify, namely the Rough Fractional Stochastic Volatility (RFSV) model and the Multifractal Random Walk (MRW) or Multifractal Random Measure (MRM) models.

The MRM/MRW models The MRW was firstly introduced in 2001 by Bacry et al. [26, 1] as a model for log-prices $X(t)$ that has exact (log-normal) multifractal properties, i.e., such that the moment of price returns $\delta_\tau X(t) = X(t+\tau) - X(t)$ obeys exact scaling properties:

$$\mathbb{E}[|\delta_\tau X(t)|^q] = \mathbb{E}[|X(t+\tau) - X(t)|^q] \sim C_q \tau^{\zeta(\frac{q}{2})} \tag{1}$$

where the multifractal scaling spectrum $\zeta(q)$ is a non-linear (namely parabolic) concave function that only depends on a single positive parameter λ^2 (which quantifies the level of non linearity of $\zeta(q)$) and such that $\zeta(1) = 1$. Let us point out that the parameter λ^2 is generally referred to as the *intermittency coefficient* since it governs the degree of multifractality of the model, i.e., the range of the Hölder exponents that characterise the paths $X(t)$. It consequently controls the degree of appearance of volatility bursts. When $\lambda = 0$, the model is said to be *monofractal*, X then simply corresponds to a Brownian motion which is almost everywhere of Hölder regularity $H = 1/2$.

The MRW model involves a log-normal stochastic volatility, that is a multifractal random measure (MRM) $\widetilde{M}_T$, obtained as the weak limit

$$\widetilde{M}_{\ell,T}(\,\mathrm{d}t) \xrightarrow[\ell\to 0]{w} \widetilde{M}_T(\,\mathrm{d}t), \tag{2}$$

where $\widetilde{M}_{\ell,T}(\,\mathrm{d}t)$ is defined by

$$\widetilde{M}_{\ell,T}(\,\mathrm{d}t) = e^{\omega_{\ell,T}(t)}\,\mathrm{d}t,$$

where $\xrightarrow{w}$ stands for the weak convergence and the process $\omega_{\ell,T}(t)$ is Gaussian and stationary with a logarithmic covariance vanishing for lags greater than T (see Eq. (75)). Let us point out that multifractality of the limit process is obtained in the Gaussian multiplicative chaos context [29] which implies that, at the same time ℓ goes to 0, the mean (resp. variance) of $\omega_{\ell,T}$ has to go to $-\infty$ (resp. $+\infty$). Thus though the stochastic measure $e^{\omega_{\ell,T}(t)}\,\mathrm{d}t$ has a weak limit, the Gaussian process $\omega_{\ell,T}(t)$ does not have a limit. We refer the reader to the beginning of Appendix A.3 for detailed construction of the log-normal MRM.

Since such a logarithmic decreasing covariance can be interpreted using random multiplicative cascades as the limit case where the scale ratio goes to 1, one often refers to such a model as "continuous cascade" [3] models. In [24, 4], MRM measures have been extended from log-normal statistics to any log-infinitely divisible law so that they obey the exact scaling law:

$$\mathbb{E}[|\delta_\tau \widetilde{M}_T(t)|^q] = \mathbb{E}[|\widetilde{M}_T(t+\tau) - \widetilde{M}_T(t)|^q] \sim C_q \tau^{\zeta(q)}, \tag{3}$$

where ζ_q is the cumulant generating index of the infinitely divisible law (let us point out that it is parabolic only in the Gaussian case).

The MRM process has been used in various works since 2001 for volatility modeling. Not only it has stationary increments but it reproduces most stylized facts of volatility (including scale invariance and self-similarity properties). Moreover, it also benefits from a concise geometric construction, which allows one to easily obtain the auto-covariance function in the desired form.

The original RFSV model. In 2018, Gatheral et al. [16] introduced a new (but related) class of models called "rough" fractional stochastic volatility (RFSV) models. Instead of focusing on the scaling properties of price increments, Gatheral et al. examined the regularity properties of the log-volatility and observed (as the case for a multifractal model) that volatility appears to be far less regular than a Brownian motion. RFSV model quickly became a popular model. Within the RSFV framework, the volatility measure $V_{H,T}([t, t+\tau])$ of some given interval $[t, t+\tau]$ is supposed to be provided by a density measure $v(t)$ corresponding to a log-normal stationary process:

$$V_{H,T}([t,t+\tau]) = \int_t^{t+\tau} v(s)\,\mathrm{d}s = \int_t^{t+\tau} e^{o_{H,T}(s)}\,\mathrm{d}s, \tag{4}$$

where $o_{H,T}(t)$ is a fractional Ornstein-Uhlenbeck (fOU) process that satisfies, for some $0 < H < 1/2$, the equation

$$\mathrm{d}o_{H,T}(t) = \nu\,\mathrm{d}B_t^H - \alpha(o_H(t) - m)\,\mathrm{d}t, \tag{5}$$

where B_t^H is a fractional Brownian motion with Hurst parameter H. The parameter ν^2 (resp. m) is the variance (resp. mean) of $o_{H,T}(t)$ and $\alpha = \frac{1}{T}$, where T represents a characteristic correlation time that accounts for the typical mean reversion length of the process. Indeed, Gatheral et al. show that, for $\tau > 0$ small enough, the covariance function of $o_{H,T}(t)$ can be approximated as:

$$\mathrm{Cov}[o_{H,T}(t), o_{H,T}(t+\tau)] \simeq \frac{\nu^2}{2}\Big(T^{2H}\Gamma(2H+1) - \tau^{2H}\Big), \tag{6}$$

where Γ represents the Gamma function. In [16], it is also shown that when $T \to \infty$, $o_{H,T}(t)$ behaves locally as a fractional Brownian motion B_t^H in the sense that, $\forall t_0$:

$$\mathbb{E}[\sup_{t\in[0,t_0]} |o_{H,T}(t) - o_{H,T}(0) - \nu B_t^H|] \to 0. \tag{7}$$

This result can be of practical importance for application in finance since empirically it appears that T is very large and consequently νB_t^H can be used as a volatility model instead of the associated fOU process $o_{H,T}(t)$ as long as $t \ll T$.

Let us point out that, since the original work [16], many other versions of RFSV models have been introduced in the literature, each of them serving some specific purposes (making some explicit computations or estimations simpler) while keeping the main feature of the original RFSV model, i.e., the "roughness" of the volatility modelled using a fBM-like process. In the next section, we will introduce a new version that will enable us to unify in the same framework an RFSV model and the MRM framework.

2.2 The log S-fBM random measure : a common framework for RFSV and MRM models

In this section, we build the main model of this paper. This model allows us to define a common framework for RFSV and MRM models. It is built in three steps. First we introduce a stationary version of a fractional Brownian motion, namely the S-fBM process $\{\omega_{H,T}(t)\}_t$ for $H > 0$. Then using this S-fBM process, we define the log S-fBM stochastic measure $M_{H,T}$ ($H > 0$) which can be seen as a new version of a rough volatility model (RFSV). Finally, we prove that this process converges when H goes to 0 to a measure that we will refer to as $M_{0,T}$, which is shown to be an MRM.

Step 1/3 : Defining the S-fBM process $\{\omega_{H,T}(t)\}_t$ for $H > 0$

The S-fBM process $\{\omega_{H,T}(t)\}_t$ is a stationary Gaussian process and can thus be defined by its mean and its covariance function. In Appendix A.1, we provide the details of its construction by following the one proposed by Bacry & Muzy ([24, 4]) for building log-infinitely divisible Multifractal Random Measures (MRM). Let us point out that such a construction can also be related to the original approach proposed by Takenaka to build correlated fields (see [30] and Appendix A.2). Thus, following the construction detailed in Appendix A.1, the S-fBM process $\{\omega_{H,T}(t)\}_t$ is defined for $H > 0$ as a stationary Gaussian process whose covariance function is:

$$C_\omega(\tau) = \mathrm{Cov}[\omega_{H,T}(t), \omega_{H,T}(t+\tau)] = \begin{cases} \frac{\nu^2}{2}[T^{2H} - \tau^{2H}], & \text{when} |\tau| < T \\ 0, & \text{when} |\tau| \geq T \end{cases} \tag{8}$$

The parameter H is analog to the Hurst parameter of the fBM process since it controls the "roughness" of the model. The variance parameter ν controls the average amplitude of the process and the constant T is a large time scale that corresponds to the correlation scale. Let us point out that the approximated covariance provided by Eq. (6) of the fOU process $o_{H,T}$, involved in the construction of the RFSV model, holds exactly for the S-fBM $\omega_{H,T}$ (up to a rescaling of T) for lags smaller than T. Both the S-fBM process and the fOU could then be regarded as stationary versions of a fBM process but, unlike the fOU process, the correlation function of S-fBM exactly vanishes

for lags greater than T, i.e., the S-fBM values at different timestamps are independent when the distance between timestamps is large enough (i.e., greater than T).

It is noteworthy that, when $T \to \infty$, one recovers the original Takenaka construction of the fBM [30] by proving that $\omega_{H,T}(t) - \omega_{H,T}(0) \to B_H(t)$. More precisely, in Appendix A.2 we show that, when $T \to \infty$, the analog of Eq. (7) holds for $\omega_{H,T}$:

Proposition 1. *There exists $B_H(t)$ a fractional Brownian motion of Hurst index H and unit variance at $t = 1$ such that, $\forall t_0 > 0$, one has:*

$$\lim_{T\to\infty} \mathbb{E}[\sup_{t\in[0,t_0]} |\omega_{H,T}(t) - \omega_{H,T}(0) - \nu B_H(t)|] = 0 \ . \tag{9}$$

A direct result from the similarity in auto-covariance function is that S-fBM has the same scaling property as RFSV. According to Appendix A.1, for $\frac{\tau}{2} < T$,

$$\delta_\tau \omega_{H,T}(t) = \omega_{H,T}(t+\tau) - \omega_{H,T}(t) \sim \mathcal{N}(0, \nu^2 \tau^{2H}). \tag{10}$$

It leads to the following scaling property of generalized moments, $\forall q > 0$:

$$\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q] = \nu^q \frac{2^{q/2}\Gamma(\frac{q+1}{2})}{\sqrt{\pi}} \tau^{qH} = C_q \tau^{qH} \ . \tag{11}$$

This means that $\log(\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q])$ is linear against $\log(\tau)$ with slope qH. From Kolmogorov continuity theorem it results that the paths of $\omega_{H,T}$ are continuous functions. More precisely, $\omega_{H,T}(t)$ is $\alpha-$Hölder continuous for all regularity exponents $\alpha < H$. We especially point out that the calibration of the Hurst parameter H of the log-volatility process in [16] is based on the obtained scaling behavior (11).

Step 2/3 : Defining the log S-fBM stochastic measure $\{M_{H,T}(t)\}_t$ for $H > 0$
The log S-fBM stochastic measure is then defined as:

$$M_{H,T}(\,\mathrm{d}t) = e^{\omega_{H,T}(t)}\,\mathrm{d}t. \tag{12}$$

Then for any interval I, one has:

$$M_{H,T}(I) = \int_I e^{\omega_{H,T}(t)}\,\mathrm{d}t \ . \tag{13}$$

Under this setting, we retrieve the so-called stationarity of volatility process, i.e.:

$$\mathbb{E}[M_{H,T}(I)] = \sigma^2 |I| \tag{14}$$

with

$$\sigma^2 = e^{m+\frac{\nu^2}{2}},$$

where m (resp. ν^2) is the mean (resp. variance) of $\omega_{H,T}$. The quantity σ^2 can be regarded as the variance of the price fluctuations on a unit-time interval.

Step 3/3 : Convergence of $\{M_{H,T}(t)\}_t$ towards an MRM when H goes to 0
As shown in Appendix A.3, the MRM measure can be recovered from the log S-fBM by taking the limit $H \to 0$. More precisely, the following proposition holds true:

Proposition 2. *Let $M_{H,T}(t) = M_{H,T}([0,t])$ be the log S-fbm process defined by* (13) *and define the intermittency coefficient,*

$$\lambda^2 = H(1-2H)\nu^2 \ . \tag{15}$$

Considering both λ^2 and the variance of the price fluctuations

$$\sigma^2 = e^{m+\nu^2/2}$$

are fixed, then, when $H \to 0$ (and consequently, $\nu^2 \to +\infty$ and $m \to -\infty$), one has

$$M_{H,T}(\,\mathrm{d}t) \xrightarrow{w} \widetilde{M}_T(\,\mathrm{d}t) \tag{16}$$

where $\xrightarrow{w}$ stands for the weak convergence and $\widetilde{M}_T$ is a log-normal MRM (as defined by (2)*) with the intermittency coefficient λ^2 and integral scale T.*

The proof is provided in Appendix A.3. This result indicates that the MRM can be considered as a limit case of a log S-fBM and therefore could be regarded as an "extremely rough" case. Let us remark that very much like the scale parameter ℓ involved in the regular construction of the MRM, the parameter H in the context of Proposion 2 can be considered as regularization parameter. However, such a regularization through $H > 0$ impacts observations at all (time) scales since it explicitly breaks scale-invariance that is not the case of regularization through ℓ as defined in Appendix A.1.

Conclusion and notations for the remaining of the paper
For the sake of simplicity, in the following, the MRM $\widetilde{M}_T(\,\mathrm{d}t)$ will be referred to as $M_{0,T}(\,\mathrm{d}t)$. Thus, we can consider that we have built a class of models $M_{H,T}(\,\mathrm{d}t)$, which correspond for $H > 0$ to an RFSV model and for $H = 0$ to an MRM model.

3 Second order properties of $M_{H,T}([0,t])$ and its logarithm

In section 4, we will consider the problem of estimating the parameters of the S-fBM, namely H, ν^2 (or equivalently λ^2) and T through the expression of various "statistical moments" of the process. Among these moments, the correlation function of $M_{H,T}$ or of $Z_{H,T} = \ln M_{H,T}$ are particularly interesting since, as emphasized below, they can be approximated by simple analytical expressions.

Let us first remark that in [16], Gatheral et al. proposed to estimate the roughness exponent H of the RFSV model (equivalently $H > 0$ in the log-SfBM model) by considering the scaling of the increments of $\omega_{H,T}$ as in Eq. (11). However, since $\omega_{H,T}(t)$ cannot be directly observable, they consider as a proxy of $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$, the observable moments:

$$m(q, H, \tau, \Delta) = \mathbb{E}\Big(|\ln M_{H,T,\Delta}(t+\tau) - \ln M_{H,T,\Delta}(t)|^q\Big), \tag{17}$$

where $M_{H,T,\Delta}(t)$ is the so-called integrated variance over an interval of size Δ:

$$M_{H,T,\Delta}(t) = \sigma^2 \int_t^{t+\Delta} e^{\omega_{H,T}(s)}\,\mathrm{d}s\ . \tag{18}$$

Thus, the exponent H is measured from the scaling behavior in τ of this proxy of $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ using the Eq. (11).

However, as emphasized below (see Section 4.1), the estimation of H based on Eq. (11) can be highly biased. In order to obtain an unbiased estimation of H in the framework of an RFSV model, Ref. [10] introduces a totally different framework. The authors provide a GMM method that is based on the correlation function of $M_{H,T,\Delta}$:

$$C_M(\Delta, \tau) = \mathbb{E}[M_{H,T,\Delta}(t) M_{H,T,\Delta}(t+\tau)]. \tag{19}$$

More precisely, they show that under peculiar conditions, its asymptotic behavior when $\tau \gg \Delta$ can be obtained and then a GMM formula can be derived. Within the framework of various RFSV models (namely the one involving an fBM or its Riemann-Liouville variant), the authors advocate the use of this GMM method and show that it provides reliable estimates for both the roughness parameter H and the variance parameter ν^2.

Following this latter path, in this work, we aim at defining a GMM method for the log-SfBM framework, that works for both $H > 0$ (the RFSV case) and $H = 0$ (the MRM case). We thus need to establish exact or good approximations of correlation function $C_M(\Delta, \tau)$ of $M_{H,T,\Delta}$. This is the purpose of the next section (Section 3.1).

Moreover, as we will see, the process $\ln M_{H,T,\Delta}(t)$ is, in some sense, close to be a Gaussian process, consequently it is also natural to operate the GMM not on the process $M_{H,T,\Delta}$ itself but on its logarithm $\ln M_{H,T,\Delta}(t)$. We therefore also need to establish exact or good approximations of the correlation function of $\ln M_{H,T,\Delta}$, which is defined by :

$$C_{\ln M}(\Delta, \tau) = \mathrm{Cov}[\ln M_{H,T,\Delta}(t), \ln M_{H,T,\Delta}(t+\tau)]\ . \tag{20}$$

This is the purpose of Section 3.2.

3.1 Integrated variance correlation function

In Appendix A.4 we prove the following Proposition that gives an explicit analytic formula for $C_M(\Delta,\tau)$:

Proposition 3. *For any $\tau \leq T$, one has*

$$C_M(\Delta,\tau) = K_1\big(F(\tau+\Delta) + F(\tau-\Delta) - 2F(\tau)\big) \tag{21}$$

with

$$F(z) = \begin{cases} zK_2^{-\frac{1}{2H}}\gamma(\frac{1}{2H}, K_2 z^{2H}) - K_2^{-\frac{1}{H}}\gamma(\frac{1}{H}, K_2 z^{2H}), & \text{when } H>0 \\ \frac{z^{2-\lambda^2}}{(2-\lambda^2)(1-\lambda^2)}, & \text{when } H=0 \end{cases} \tag{22}$$

where $\gamma(a,z)$ stands for the (lower) incomplete Gamma function,

$$\gamma(a,z) = \int_0^z t^{a-1}e^{-t}\,\mathrm{d}t,$$

and where we have denoted

$$\begin{aligned} K_1 &= \frac{\sigma^4 e^{K_2 T^{2H}}}{2H}, \quad \text{if } H>0 \\ &= \sigma^4 T^{\lambda^2}, \quad \text{if } H=0 \\ K_2 &= \frac{\nu^2}{2} = \frac{\lambda^2}{2H(1-2H)}. \end{aligned}$$

Let us notice that when $\tau > T$, since $M_{H,T,\Delta}(t)$ and $M_{H,T,\Delta}(t+\tau)$ are independent, one has:

$$C_M(\Delta,\tau) = \mathbb{E}[M(\Delta,\tau)]^2 = \sigma^4\Delta^2 .$$

Moreover, using the equality:

$$\gamma(s,z) = s^{-1}z^s e^{-z}U(1,s+1,z),$$

where $U(1,s,z)$ is the Kummer's confluent hypergeometric function, the function $F(z)$ can be simply rewritten as:

$$F(z) = \sigma^4 z^2 e^{C_\omega(z)}\left(U(1,1+\frac{1}{2H},K_2z^{2H}) - \frac{1}{2}U(1,1+\frac{1}{H},K_2z^{2H})\right), \tag{23}$$

where $C_\omega(z)$ is the covariance of $\omega_{H,T}$ provided by Eq. (8). Since, when $|b|\to\infty$, $U(1,b,z)\simeq 1+\frac{z}{b}$, and finally, when $H \ll 1$, the following approximation for $F(z)$ holds:

$$F(z) \simeq \sigma^4 z^2 e^{C_\omega(z)}\left(\frac{1}{2} + \frac{3H}{2}z^{2H}\right). \tag{24}$$

3.2 Small λ^2 approximation of the logarithm integrated variance moments

In this section, our goal is to obtain analytical expressions for the moments of $\ln M([t,t+\Delta])$ instead of $M([t,t+\Delta])$. In [3] a GMM method to estimate the parameters of the MRM $\widetilde{M}_T$ has been proposed relying on the expression of such logarithmic moments that were obtained within a *small intermittency*, i.e. $\lambda^2 \ll 1$, asymptotic behavior. In fact, it is straightforward to check that all proofs and results established in [3] in the limit $\lambda^2 \to 0$ for the log-normal MRM measure $\widetilde{M}_T$ remain valid for $M_{H,T,\Delta}$ for $H>0$, i.e. in the log S-fBM framework introduced in this paper. Indeed, in particular by simply checking that all conditions required for MRM also hold for the log S-fBM measure $M_{H,T,\Delta}$, a direct consequence of Proposition 13 in [3] is the following result:

Proposition 4. *Let $t_1,\dots,t_n$ be n arbitrary times. The generalized moments of the logarithm of $\Delta^{-1}M_{H,T,\Delta}(t)$ admit the following Taylor series expansion around $\lambda^2=0$:*

$$\mathbb{E}\Big[\ln\Big(\frac{M_{H,T,\Delta}(t_1)}{\Delta}\Big)\cdots\ln\Big(\frac{M_{H,T,\Delta}(t_n)}{\Delta}\Big)\Big] = \lambda^n\Delta^{-n}\mathbb{E}\Big[\Omega_{H,T,\Delta}(t_1)\cdots\Omega_{H,T,\Delta}(t_n)\Big] + o(\lambda^n), \tag{25}$$

where $\Omega_{H,T,\Delta}(t)$ is the Gaussian process defined by

$$\Omega_{H,T,\Delta}(t) = \frac{1}{\lambda}\int_t^{t+\Delta}\Big(\omega_{H,T}(u) - \mathbb{E}(\omega_{H,T}(u))\Big)\,\mathrm{d}u. \tag{26}$$

Within this approximation, one can directly compute $C_{\ln M}(\Delta,\tau)$ the correlation function of $\ln M_{H,T,\Delta}(t)$ as defined in (20). From the definition of $\Omega_{H,T,\Delta}$ and the expression (8) for the covariance of $\omega_{H,T}(t)$, it results:

Proposition 5. *To the first order in $\lambda^2 \ll 1$, the covariance function of $\ln M_{H,T,\Delta}$, reads:*

$$\begin{aligned} C_{\ln M}(\Delta,\tau) &= \frac{\lambda^2}{2H(1-2H)}\Delta^{-2}\int_0^{\Delta} \mathrm{d}u \int_{\tau}^{\tau+\Delta}\left(T^{2H}-|u-v|^{2H}\right)\ \mathrm{d}v \qquad (27)\\ &= \frac{\lambda^2}{2H(1-2H)}\left(T^{2H}-\frac{(\tau+\Delta)^{2H+2}+|\tau-\Delta|^{2H+2}-2\tau^{2H+2}}{\Delta^2(2H+1)(2H+2)}\right)+o(\lambda^2). \qquad (28)\end{aligned}$$

Let us first start with two direct consequences of these propositions

- When $\Delta \to 0$ one has $C_{\ln M}(\Delta,\tau) \simeq \frac{\lambda^2}{2H(1-2H)}(T^{2H}-\tau^{2H})$ which is nothing but the covariance of $\omega_{H,T}(t)$.
- When $H \to 0$, one recovers the expression in Proposition 10 of [3] in the MRM case.
- Proposition 4 leads to approximating $\ln M_{H,T,\Delta}(t)$ by a Gaussian process.

Using this last consequence, Proposition 4 can also be used to obtain an approximation to the first order in λ^2 of the moments defined in (17). Indeed, if one supposes that $\ln M_{H,T,\Delta}(t+\tau) - \ln M_{H,T,\Delta}(t)$ is a Gaussian random variable of variance $V(H,\tau,\Delta)$, then,

$$m(q,H,\tau,\Delta) \overset{\lambda}{=} \pi^{-1/2}2^{\frac{q}{2}}\Gamma\left(\frac{q+1}{2}\right)\Big(V(H,\tau,\Delta)\Big)^{\frac{q}{2}} \qquad (29)$$

in which $\overset{\lambda}{=}$ indicates that equality holds in the first order of λ^2.

From expression (28), one has when $\tau < T$,

$$\begin{aligned} V(H,\tau,\Delta) &= 2\mathrm{Var}(\ln M_{H,T,\Delta}(t)) - 2C_{\ln M}(\Delta,\tau)\\ &\overset{\lambda}{=} \frac{\lambda^2}{H(1-2H)}\left(\frac{(\tau+\Delta)^{2H+2}+|\tau-\Delta|^{2H+2}-2\tau^{2H+2}}{\Delta^2(2H+1)(2H+2)} - \frac{2\Delta^{2H+2}}{\Delta^2(2H+1)(2H+2)}\right)\\ &\overset{\lambda}{=} \lambda^2\tau^{2H}g_H(\frac{\Delta}{\tau})\end{aligned}$$

with

$$g_H(z) = \frac{|1+z|^{2H+2}+|1-z|^{2H+2}-2|z|^{2H+2}-2}{z^2H(1-2H)(2H+1)(2H+2)}\ . \qquad (30)$$

The final expression for the moments of the increments of the measure logarithm $\ln M_{H,T,\Delta}$ reads, in the first oder in λ^2,

$$m(q,H,\tau,\Delta) \overset{\lambda}{=} 2^{\frac{q}{2}}\pi^{-1/2}\Gamma\left(\frac{q+1}{2}\right)\lambda^q\tau^{qH}\left[g_H\left(\frac{\Delta}{\tau}\right)\right]^{q/2}. \qquad (31)$$

Let us remark that we have the following asymptotic relation:

$$g_H(z) \sim \begin{cases} \frac{1}{H(1-2H)}+O(z^{2H}) & \text{when } z\to 0\\ z^{-2}\Big((1+z)^2\ln(1+z)+(1-z)^2\ln(1-z)-2z^2\ln(z)\Big)+O(H) & \text{when } H\to 0\end{cases} \qquad (32)$$

and when $H > 0$, one recovers that when $\Delta \to 0$ one has, to the first oder in λ^2,

$$m(q,H,\tau,\Delta) \overset{\lambda}{=} C\tau^{qH}+O((\Delta/\tau)^{2H}), \qquad (33)$$

which is the expression used to estimate H in [16] where $\ln M_{H,T,\Delta}(t)$ corresponds to the logarithm of the (daily) realized volatility.

4 Estimation

This section is devoted to the estimation of H in the framework of log-SfBM. We first show (in Section 4.1) that if H is measured from the scaling behavior of $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ against τ using Eq. (11) (as advocated in [16]), the estimation of H can be highly biased.

In order to obtain an unbiased estimation of H, we consider two GMM based estimators in Sections 4.2 and 4.3. The first one is based on the use of the moments of the log-SfBM process itself mainly relying on the explicit covariance formula in Eq. (21). The second one is based on the use of moments of the logarithm of the log-SfBM process and involves the explicit covariance provided by Eq. (28).

We show that both estimators are expected to be reliable even in the "high-frequency regime" when data are only available over an interval that is smaller than the overall correlation scale T, i.e. in a regime when one does not expect any ergodic hypothesis to hold.

4.1 Bias of the moment scaling method proposed in Ref. [16]

In [16], the parameter H is estimated from the scaling behavior of $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ against τ as described in Eq. (11). More precisely, the unobservable quantity $\mathbb{E}[|\delta_\tau \omega_{H,T}(t)|^q]$ is substituted by its observable proxy $m(q, H, \tau, \Delta)$ as defined in Eq. (17), whose explicit form is worked out in Eq. (31). Then, a linear regression of $\ln m(q, H, \tau, \Delta)$ against $\ln(\tau)$ is performed in order to estimate H.

We can show see that this approach can lead to a significantly biased estimation of H. Indeed, by taking logarithm on both sides of Eq. (31), one has (using again the notation $\overset{\lambda}{=}$ for equality up to the first order of λ^2)

$$\ln(m(q, H, \tau, \Delta)) \overset{\lambda}{=} C(q, \nu) + qH \ln(\tau) + \frac{q}{2} \ln\left(g_H(\frac{\Delta}{\tau})\right). \tag{34}$$

where the expression of $g_H(z)$ is provided in Eq. (30). Since the term $\ln(g_H(\frac{\Delta}{\tau}))$ also depends on $\ln(\tau)$, assuming, on a given range of τ, that

$$\ln(g_H(\frac{\Delta}{\tau})) \simeq B_H \ln(\tau/\Delta) + C, \tag{35}$$

the measured slope $\hat{H}$ in the relation $\ln m_q(H, \tau, \Delta)$ against $\ln(\tau)$ is biased as:

$$\hat{H} = H + \frac{B_H}{2}, \tag{36}$$

and the bias depends on both the considered range of τ and the value of H.

Let us illustrate this phenomenon on some numerical simulations. For that purpose, let us consider an arbitrary value $\Delta = 1$ and $\tau \in [1, 500]$. For the specific value $H = 0.002$, Fig. 1 plots $\ln(g_H(\frac{\Delta}{\tau}))$ as a function of $\ln(\frac{\tau}{\Delta}))$. We note that the behavior is, to a first approximation, assumed to be linear in the range when τ is sufficiently large in front of Δ. A linear regression leads to a slope value of $B \simeq 0.160$ which dominates the (highly biased) estimation of H : $\hat{H} \simeq 0.08$. Using the same procedure, we checked that for different values of H in the range $[0.0, 0.15]$, one systematically overestimates H with a bias that decreases from 0.08 to 0.03. It is noteworthy that the same kind of bias analysis has been considered by the author of Ref. [16] themselves (see their Appendix C).

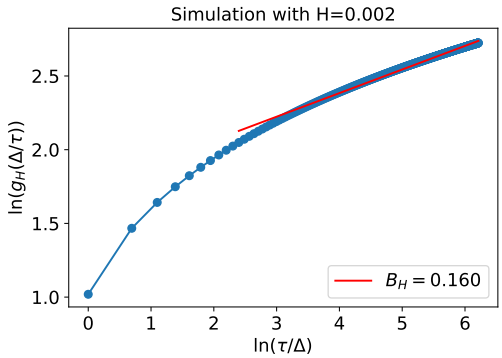

Figure 1: Estimation of the bias term B (Eq. (35)) involved in the estimation of H using linear regression on Eq. (34). Following Eq. (35), $\ln(g_H(\Delta/\tau))$ is displayed against $\ln(\tau/\Delta)$ (where $\Delta = 1$ and τ varies from 1 to 500). B_H is estimated by linear regression on this curve over the range $\tau \in [10, 500]$

4.2 Low versus high frequency regime for GMM estimations

As already explained, our purpose is to build two GMM estimators based on the second order moments of the log-SfBM process $M_{H,T}$ or its logarithm. More precisely, we will consider respectively $C_M(\Delta, \tau)$ the correlation function of $M_{H,T,\Delta}$ (using the explicit covariance formula (21)) and $C_{\ln M}(\Delta, \tau)$, the covariance function of $\ln M_{H,T,\Delta}$ (using the explicit covariance formula (28)). If L denotes the overall size of the interval where the empirical data are available at scale Δ, one can measure $M_{H,T,\Delta}(k\Delta)$ (or equivalently $\ln M_{H,T,\Delta}(k\Delta)$) for $k = 1 \dots N$ where $N = \frac{L}{\Delta}$ and the estimators of previous correlation functions read:

$$\widehat{C_M}(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} M_{H,T,\Delta}(j\Delta) M_{H,T,\Delta}((j+k)\Delta) \tag{37}$$

$$\widehat{C_{\ln M}}(\Delta, k\Delta) = N^{-1} \sum_{j=1}^{N-k} \Big(\ln M_{H,T,\Delta}(j\Delta) - \widehat{\mu}_\Delta \Big) \Big(\ln M_{H,T,\Delta}((j+k)\Delta) - \widehat{\mu}_\Delta \Big) \tag{38}$$

$$\widehat{\mu}_\Delta = \frac{1}{N} \sum_{k=1}^{N} \ln M_{H,T,\Delta}(k\Delta) \tag{39}$$

In general, GMM methods rely on some ergodic hypothesis that ensures the convergence of previous empirical means towards the expected values. As advocated in [3] or in [10], these approaches allow one to build efficient parameter estimator in the limit $N = \frac{L}{\Delta} \to \infty$, which, when Δ is kept fixed, corresponds to $L \to \infty$. When $L \gg T$ (recall that T is the correlation length of $M_{H,T,\Delta}(t)$), this ergodicity assumption can be proven to hold. We refer to such a situation as the "low-frequency regime".

However, as first remarked in [3], one can alternatively consider the asymptotic regime $N \to \infty$ when $\Delta \to 0$, while $L = \mathcal{O}(T)$ is fixed. This is the "high-frequency regime". Thus, whereas the low-frequency regime corresponds to $\Delta < T \ll L$, the second one corresponds to $\Delta \ll L = \mathcal{O}(T)$.

Let us point out that, as emphasized in [3] and motivated by the empirical results reported in [25] (see also Sec. 5 below), in many practical situations and notably for financial time series, the high-frequency regime appears to fit more precisely the empirical conditions. Notably, it appears that the correlation scale T of the realized volatility always seems to be larger than the observation size L. For instance, in Fig. 6(b) of [25], the authors plotted the logarithm of Dow-Jones realized daily volatility from 1928 to 2011 and observed deviations far from the "mean value" that are lasting for decades. The same kind of observation can be done in Fig. 8(a) below. In [25], it is also observed that the estimated correlation scale increases linearly with the observation size L from a few days to several years in agreement with the hypothesis that the true correlation scale is extremely large. In such a situation, assuming that the low-frequency regime $L \gg T$ is reachable and consequently that the ergodic hypothesis holds, is clearly unrealistic.

These remarks call for developing GMM estimations in the high-frequency regime $\Delta \ll L \leq T$. Let us first start by noticing that from the expression of the covariance of $\omega_{H,T}$ (Eq. (8)), for $t \in I$ (where I is any interval such that $|I| < L$), one has

$$\{\omega_{H,T}(t)\}_{t\in I} \stackrel{\mathcal{L}}{=} \{\Omega + \omega_{H,L}(t)\}_{t\in I} \tag{40}$$

where $\stackrel{\mathcal{L}}{=}$ means an equality of all finite dimensional distributions and Ω is a Gaussian random variable independent of $\omega_{H,L}$ and of variance $\frac{\nu^2}{2}(T^{2H} - L^{2H})$. It thus results that we have, in any interval I of size $|I| < L$,

$$\{M_{H,T,\Delta}\}_{t\in I} \stackrel{\mathcal{L}}{=} \{e^{\Omega} M_{H,L,\Delta}(t)\}_{t\in I} \text{ and} \tag{41}$$

$$\{\ln M_{H,T,\Delta}\}_{t\in I} \stackrel{\mathcal{L}}{=} \{\Omega + \ln M_{H,L,\Delta}(t)\}_{t\in I} \,. \tag{42}$$

As already discussed in [3] for the case of the MRM measure $M_{H=0,L,\Delta}$, these properties show that one cannot measure the parameters T and σ^2 over an interval of size $L < T$ since by redefining σ^2 as $\sigma^2 e^{\Omega}$, one can always assume that $T = L$. It can also be seen on expression (22) that, when $\tau < T$, the large correlation scale T can be absorbed in a redefinition of the variance parameter σ^2.

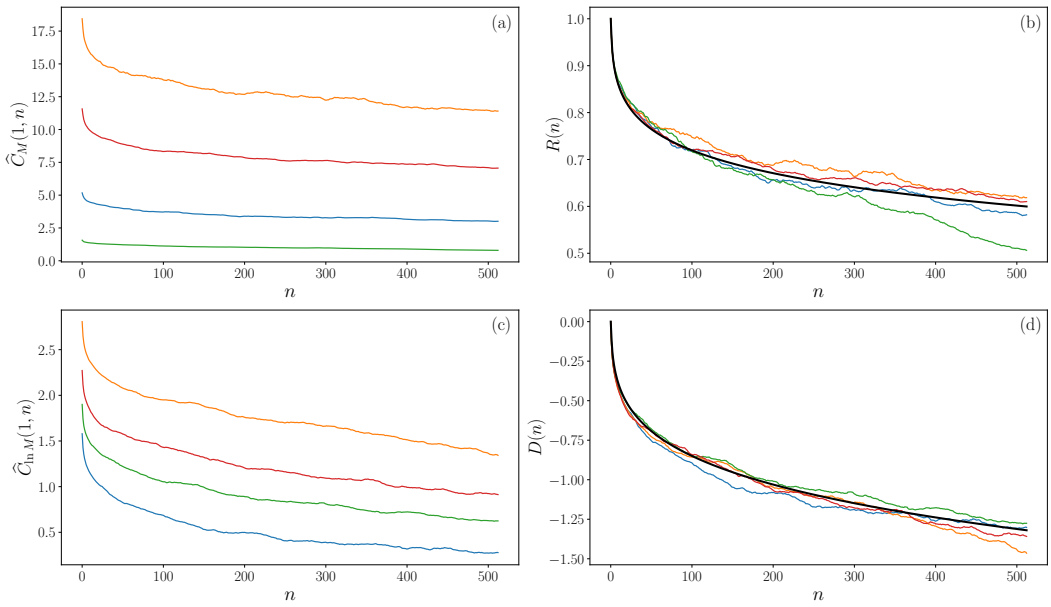

Figure 2: Estimation of the correlation functions of $M_{H,T,\Delta}$ and $\ln M_{H,T,\Delta}$ over an interval of size L such that $\Delta = 1 \ll L < T$. (a) $\widehat{C_M}(\Delta = 1, n)$ as estimated from 4 independent realisations of $M_{H=0.1,T=2^{17},\Delta=1}$ ($\lambda^2 = 0.03$) over an interval of size $L = 2^{14}$. Each estimation appears to be multiplied by an arbitrary random factor. (b) the ratio $R(n) = \frac{\widehat{C}_M(1,n)}{\widehat{C}_M(1,0)}$ vs. n. With such normalisation, all curves are superimposed and appear to be well fitted by the analytical expression $\widetilde{R}(n)$ represented by the bold black curve. (c) $\widehat{C}_{\ln M}(\Delta = 1, n)$ as estimated from 4 independent realisations of $M_{H=0.1,T=2^{17},\Delta=1}$ ($\lambda^2 = 0.08$) over an interval of size $L = 2^{14}$. Each estimation appears to be shifted by a random term. (d) the difference $D(n) = \widehat{C}_{\ln M}(1,n) - \widehat{C}_{\ln M}(1,0)$ vs. n. When shifting all curves in such a way, they are superimposed and appear to be well fitted by the analytical expression $\widetilde{D}(n)$ represented by the bold black curve.

If one seeks to consider correlation function based GMM estimators in the high frequency regime, one thus needs to study the behavior of respectively the estimators, $\widehat{C}_{\ln M}(\Delta, n\Delta)$, $\widehat{C}_M(\Delta, n\Delta)$ in

the limit $\Delta \to 0$. A rigorous study of this problem is beyond the scope of the present paper, but we can refer to Theorem 10 of [3] where the authors proved that, in the multifractal case ($H = 0$), the behavior of $\widehat{C}_{\ln M}(\Delta, n\Delta)$ can be used to build an asymptotically unbiased and consistent estimator of λ^2 in the high frequency regime. In the present paper, we just give a sketch of proof that one can build moments functions with vanishing fluctuations in the limit $\Delta \to 0$.

First, let us notice that, without loss of generality, one can always perform an overall change of scale, $\Delta \to 1$, $L \to \frac{L}{\Delta}$, $T \to \frac{T}{\Delta}$. This amounts to assume that $\Delta = 1$ while the limit $\Delta \to 0$ becomes $L, T \to \infty$ and $L = CT$ with $C = \mathcal{O}(1)$.

Then, Appendix A.5 provides an heuristic proof of the following result:

Proposition 6. *Suppose that $H < \frac{1}{2}$. Then, for any $n_{\max} < \infty$, $C \leq 1$, when $L = CT \to \infty$, then, to the first order in λ^2, one has for all $n \leq n_{\max}$:*

$$D(n) \stackrel{def}{=} \widehat{C}_{\ln M}(1, n) - \widehat{C}_{\ln M}(1, 0) \xrightarrow{P} \widetilde{D}_{\ln M}(n) \tag{43}$$

where $\xrightarrow{P}$ means that the convergence holds in "in probability" and where

$$\widetilde{D}_{\ln M}(n) = -\nu^2 \frac{|n+1|^{2H+2} + |n-1|^{2H+2} - 2n^{2H+2}}{2(2H+1)(2H+2)}. \tag{44}$$

Moreover, numerical experiments (see Fig. 2 below) also suggest that an equivalent result holds for $\widehat{C}_M$, i.e.,

$$R(n) = \frac{\widehat{C}_M(1, n)}{\widehat{C}_M(1, 0)} \xrightarrow{P} \widetilde{R}_M(n) \tag{45}$$

with

$$\widetilde{R}_M(n) = F(n+1) + F(n-1) - 2F(n) \tag{46}$$

where $F(z)$ is defined in Eq. (22).

The consequence of Eqs. (43), (45) is that, for L large enough, there exist two positive random variables K_1 and K_2 such that, in the first order of λ^2):

$$\widehat{C}_{\ln M}(1, n) \simeq K_1 + \widetilde{D}_{\ln M}(n), \tag{47}$$

$$\widehat{C}_M(1, n) \simeq K_2 \widetilde{R}_M(n). \tag{48}$$

Numerical illustrations of these relations are given in Figs. 2 and 3.

In Fig. 2, we have displayed the estimated correlation functions $\widehat{C_M}(\Delta, n\Delta)$ and $\widehat{C}_{\ln M}(\Delta, n\Delta)$ for 2 sets of 4 realisations of $M_{H,T,\Delta}(t)$ over an interval of size $L = 2^{14}$ with $\Delta = 1$, $H = 0.1$, $T = 2^{17}$ and $\lambda^2 = 0.03$ (for $\widehat{C_M}$) or $\lambda^2 = 0.08$ (for $\widehat{C}_{\ln M}$). One clearly sees in Fig. 2(a) that each estimate $\widehat{C_M}$ seems to differ from the other one by a significant geometric random factor while estimates of $\widehat{C}_{\ln M}$ appear to be randomly shifted in Fig. 2(c). In order to check these assertions, we have plotted respectively the ratios $R(n)$ and the differences $D(n)$ in Figs 2(b) and 2(d). As expected, all the curves appear to collapse to a single curve that is well described by analytical expressions obtained from respectively Eq. (44) and (46) (represented by bold curves).

The asymptotic convergence of Proposition 6 is illustrated in Fig. 3 where we have plotted $D(n)$ as defined in Eq. (43) as obtained from random samples of $M_{H,T,\Delta}$ with $\Delta = 1$, $H = 0.1$, $\lambda^2 = 0.08$, $T = 2L$ and $L = 2^{12}, 2^{14}, 2^{16}, 2^{18}$. All the curves are shifted by an arbitrary small constant for clarity purpose. As predicted by Eq. (43), one sees that, as L increases, the empirical curves become less and less noisy and increasingly close to the analytical expectation (44) (black curve).

4.3 Defining two GMM estimators for H and λ^2

We are now ready for defining two GMM estimators for H based respectively on the moments of $M_{H,T,\Delta}$ or its logarithm in the high-frequency limit. By using the previously established expressions (44) and (46) for the empirical correlation function $\widehat{C}_{\ln M}(1, n)$ and $\widehat{C}_M(1, n)$, one can devise two GMM methods along the same line as the methods proposed respectively in [3] and [10].

As explained in the previous section (Section 4.2), in the high-frequency regime, estimations of T or σ^2 are unreachable. Thus, hereafter, we consider exclusively the problem of estimating

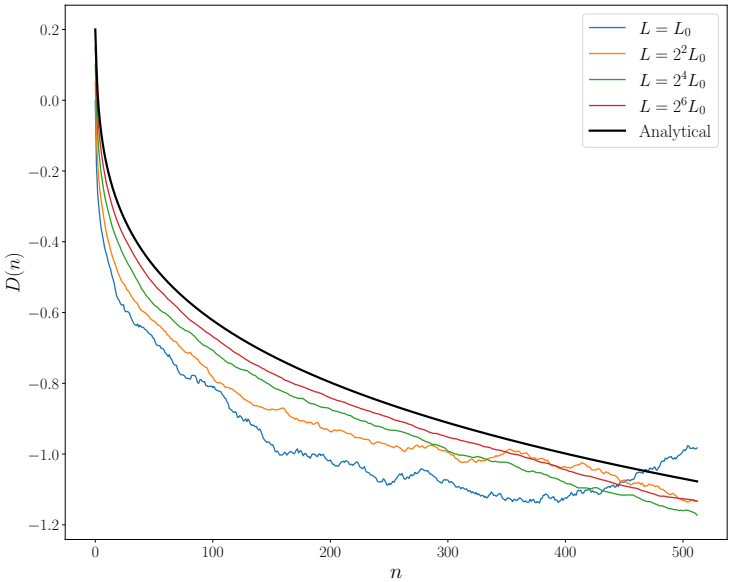

Figure 3: Estimation of the correlation function difference $D(n)$ as defined in Proposition 6. $D(n)$ from observations over intervals of increasing size $L = L_0$, $L = 4L_0$, $L = 16L_0$ and $L = 64L_0$ with $L_0 = 2^{12}$. The remaining parameters are $\Delta = 1$, $H = 0.1$, $\lambda^2 = 0.08$ and $T = 2L$. One sees that as $L \to \infty$, the empirical fluctuations become smaller and smaller and the empirical estimations appear to converge towards to theoretical expression (44) (black bold curve). The curves have been shifted by an increasing constant for the sake of clarity.

the values of the parameters H and λ^2 (or alternatively ν^2) using one of the following two sets of moments:

$$\begin{aligned} \mathrm{GMM}_M : \quad \mathcal{M}_1 &= \left(\widehat{C_M}(1, j_1) - \widetilde{C}_M(1, j_1), \ldots, \widehat{C_M}(1, j_Q) - \widetilde{C}_M(1, j_Q)\right) , \\ \mathrm{GMM}_{\ln M} : \quad \mathcal{M}_2 &= \left(\widehat{C_{\ln M}}(1, j_1) - \widetilde{C}_{\ln M}(1, j_1), \ldots, \widehat{C_{\ln M}}(1, j_Q) - \widetilde{C}_{\ln M}(1, j_Q)\right) , \end{aligned}$$

where Q is the number of moments, $j_1, j_2, \ldots, j_Q$ different time indices, $\widehat{C_M}$ and $\widehat{C}_{\ln M}$ are the empirical estimators of respectively C_M and $C_{\ln M}$ and $\widetilde{C}_M, \widetilde{C}_{\ln M}$ the following analytical expressions:

$$\widetilde{C}_{\ln M}(1, n) = K_1 + \widetilde{D}_{\ln M}(n) + V_1 \delta_n \tag{49}$$

$$\widetilde{C}_M(1, n) = K_2 \widetilde{R}_M(n) \tag{50}$$

where K_1, K_2 and V_1 are 3 random positive constants and δ_n stands for the Kronecker function. Notice that the term $V_1 \delta_n$ allows one to account for the eventual presence of a white noise (of variance V_1 superimposed to $\ln M_{L,T,1}(t)$ as described in ref. [10].

4.4 Numerical illustrations and empirical performances of the GMM methods

In order to verify our approach and compare the performances of GMM_M and $\mathrm{GMM}_{\ln M}$, we have carried out various numerical experiments. However, since historical volatility is not directly observable in financial markets, in order to consider a more realistic scenario, we decided to run the experiments directly on a price model. We consider that a "price" X_t is modelled by a Brownian motion whose variance is a log S-fBM measure $\mathrm{d}M$, i.e.,

$$dX_t = e^{\omega_{H,T}(t)} \, \mathrm{d}B_t = \frac{M_{H,T}(\mathrm{d}t)}{\mathrm{d}t} \, \mathrm{d}B_t \ , \tag{51}$$

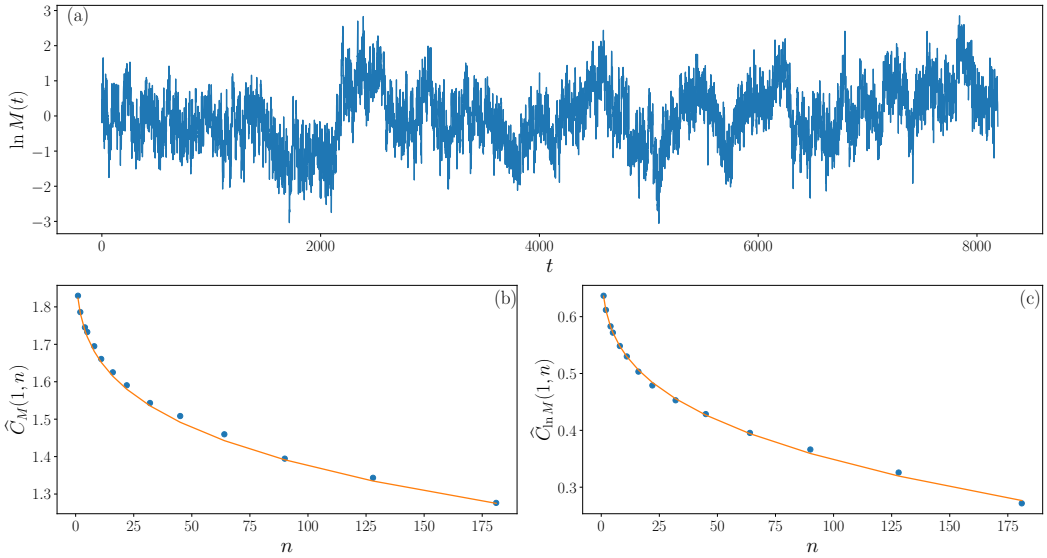

Figure 4: GMM estimations of H and λ^2 (a) A sample of length $L = 2^{14}$ of the "log-volatility", $\ln \widehat{M}_{H,T,\Delta}(t)$, with $n = 32$, $\Delta = 1$, $H = 0.08$, $\lambda^2 = 0.1$ and $T = 2^{17}$. (b) Best GMM fit of $\widehat{C}_M(\tau)$ as a function of τ. (c) Best GMM fit of $\widehat{C}_{\ln M}(\tau)$ as a function of τ.

where $M_{H,T}$ is the log-fBM defined in (12) while B_t is a Brownian motion independent of $M_{H,T}$. Let us notice that, when $H = 0$, X_t is precisely the MRW process introduced in [26, 1]. Let us also remark that, in many respects (e.g., by rewriting $\omega_{H,T}$ as a stochastic integral of the form $\int_{-\infty}^{t} g(t-s)dW_s$), the model (51) can be seen as a peculiar non skewed variant a the rough Bergomi model introduced in [6].

Alternatively, an equivalent definition of X_t can be obtained using a time-warp of the Brownian motion:

$$X_t = B(M_{H,T}(t)) \ . \tag{52}$$

Within this framework, $M_{H,T}(t)$ is called the (stochastic) volatility of X_t. If one does not observe directly $M_{H,T}(\,\mathrm{d}t)$ but only the process X_t, as emphasized notably in [5], a proxy of the integrated volatility over an interval of size Δ is provided by an estimation of the quadratic variation of X:

$$\widehat{M}_{H,T,\Delta}(t) = \sum_{i=1}^{n} \left(X_{t+\frac{i\Delta}{n}} - X_{t+\frac{(i-1)\Delta}{n}} \right)^2 \ . \tag{53}$$

As shown in [5] (see also [10]), as $n \to \infty$, under mild conditions, $\widehat{M}_{H,T,\Delta} \to M_{H,T,\Delta}$ while even for moderate n, $\widehat{M}_{H,T,\Delta}(t)$ and $\ln \widehat{M}_{H,T,\Delta}(t)$ provide excellent approximations of the integrated volatility and its logarithm. For the purpose of this paper, we have checked that $n = 32$ is sufficient to disregard any significant difference between $\widehat{M}$ and M.

We simulated independent samples of S-fBM processes and the associated processes X_t with $H = 0.02$, $H = 0.08$ and $H = 0.15$, with 2 different values of λ^2, namely 0.02 and 0.1. We chose $T = 2^{17}$, $L = 2^{14}$ and fixed arbitrary $\sigma^2 = 1$.

For all these parameters we run both GMM_M and $\mathrm{GMM}_{\ln M}$ estimators with $Q = 19$ and $\{\tau_k\}_{k=0,\dots,18} = \left\lfloor \sqrt{2^k} \right\rfloor$. Our GMM implementations closely follow the one detailed in [10] and notably the error covariance is estimated using the Newey-West HAC type estimator with a lag $L^{1/3}$ and the initialisation is performed using the scaling estimator provided in [16]. We used the *L-BFGS-B* minimisation algorithm as provided by *scipy.optimize* library in Python but we find similar results using alternative methods.

In Fig. 4, are displayed a fit of respectively $\widehat{C_M}(1,n)$ and $\widehat{C}_{\ln M}(1,n)$ using expressions (44) and (46) with the estimated GMM parameters for a sample of length $L = 16384$ with $H = 0.08$ and $\lambda^2 = 0.02$ or $\lambda^2 = 0.1$. Our estimation results are summarised in Table 1 where we reported the obtained mean values and standard deviation of estimated H and λ^2 for each set of parameters.

$\lambda^2 = 0.02$	$H = 0$	$H = 0.02$	$H = 0.08$	$H = 0.15$
$\widehat{H}$ (GMM$_M$)	0.010 (0.01)	0.007 (0.015)	0.077 (0.033)	0.146 (0.05)
$\widehat{H}$ (GMM$_{\ln M}$)	0.010 (0.01)	0.018 (0.015)	0.082 (0.02)	0.153 (0.02)
$\widehat{\lambda}^2$ (GMM$_M$)	0.010 (0.01)	0.010 (0.01)	0.018 (0.006)	0.021 (0.005)
$\widehat{\lambda}^2$ (GMM$_{\ln M}$)	0.019 (0.001)	0.020 (0.001)	0.019 (0.002)	0.020 (0.002)
$\lambda^2 = 0.1$	$H = 0$	$H = 0.02$	$H = 0.08$	$H = 0.15$
$\widehat{H}$ (GMM$_M$)	0.010 (0.02)	0.018 (0.02)	0.11 (0.22)	0.16 (0.26)
$\widehat{H}$ (GMM$_{\ln M}$)	0.010 (0.01)	0.02 (0.01)	0.078 (0.02)	0.16 (0.02)
$\widehat{\lambda}^2$ (GMM$_M$)	0.08 (0.03)	0.08 (0.02)	0.09 (0.045)	0.08 (0.07)
$\widehat{\lambda}^2$ (GMM$_{\ln M}$)	0.095 (0.001)	0.10 (0.005)	0.10 (0.008)	0.10 (0.008)

Table 1: Summary of GMM$_M$ & GMM$_{\ln M}$ estimation performances. For each parameter set, we report the mean values and standard deviations as obtained from estimations realized on 50 independent samples of length $L = 2^{14}$ of log S-fBM stochastic volatility model.

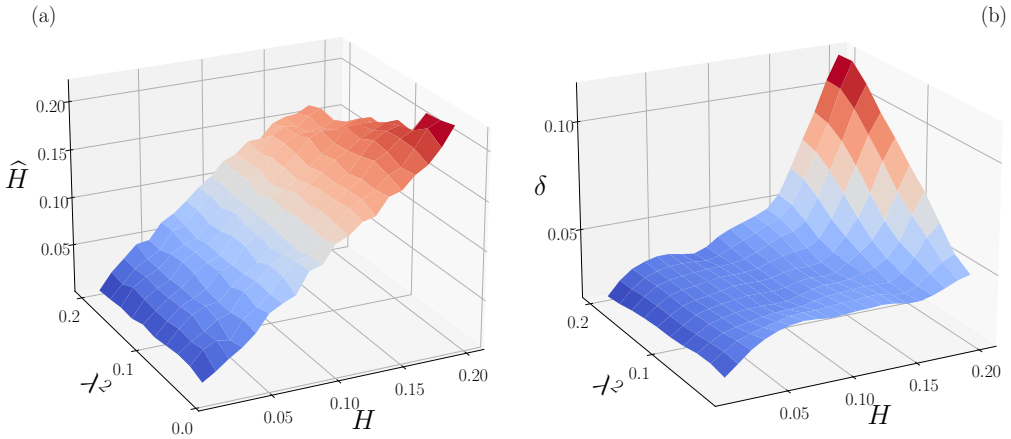

Figure 5: Surface plot of the Hurst parameter H estimation error as a function of the parameters H and λ^2. For each pair of parameters values, the considered log S-fBM stochastic volatility series are of length $L = 4096$, with $T = 2^{17}$ and the averages $E(.)$ are computed over 50 independent draws. (a) $E(\widehat{H})$, the average GMM$_{\ln M}$ estimation $\widehat{H}$ is plotted as as function of H and λ^2. On observes a significant negative bias for largest values of H and λ^2 ($H \geq 0.15$ and $\lambda^2 \geq 0.1$). (b) The error $\delta = \sqrt{E[(H - \widehat{H})^2]}$ as a function of H and λ^2. On can see that the error remains small (i.e. less than 0.03) as long as $H < 0.15$ and $\lambda^2 < 0.1$.

We clearly see that the GMM$_{\ln M}$ method relies on logarithms of integrated volatilities outperforms the GMM$_M$ method built on integrated volatilities. This latter approach appears to have significantly larger bias and variance errors notably for very small H values. GMM$_{\ln M}$ method provides more reliable estimates and in particular one sees that the error on λ^2 is very small for all sets of parameters. In order to better illustrate the variations of δ, the estimation error on $\widehat{H}$, as a function of the S-fBm parameters, we have displayed it as a surface plot in Fig. 5 for a size $L = 2^{12}$ comparable to ones of the empirical time series considered in section 5. One can see that the error increases when both H increases and λ^2 increase but remains rather small in the domain

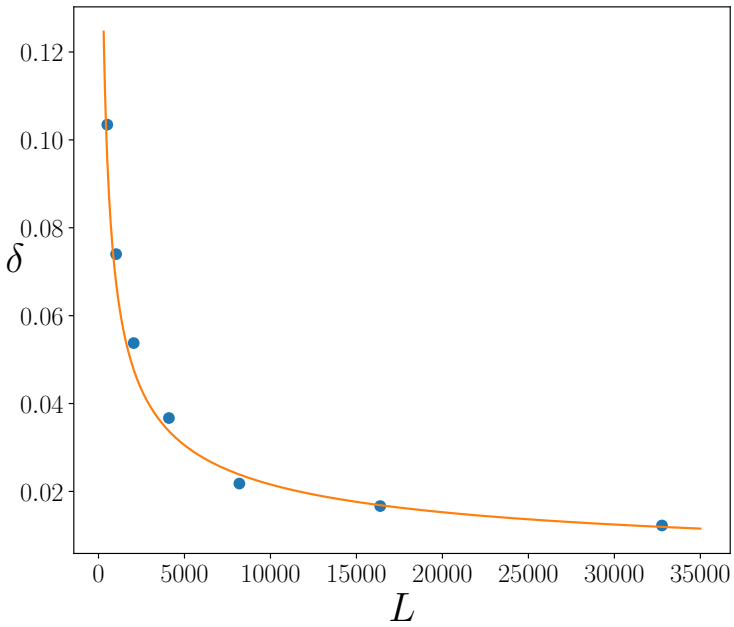

Figure 6: $\mathrm{GMM}_{\ln M}$ Hurst exponent estimation error δ as a function of the sample size L. For each size $L \in [2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}, 2^{15}]$, we have generated 200 realizations of log S-fBM stochastic volatility series of length L with $H = 0.1$, $\lambda^2 = 0.03$ and $T = 4L$. Empirical error are represented by symbols ($\bullet$) while the solid line represents the curve $\delta = KL^{-1/2}$.

$0 \leq H \leq 0.15$ and $0 < \lambda^2 \leq 0.1$. As far as the estimation error of λ^2 is concerned, we have observed similar results though with a relative error smaller than 0.1 over the whole domain.

Regarding the scaling of the error with respect to the sample size L, by considering, besides $L = 2^{12}$ as in Fig. 5 or $L = 2^{14}$ as in Table 1), various sample lengths ($L = 2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}, 2^{15}$) with $T = 4L$, we checked that, as predicted by Prop. 6, the estimation errors vanishes when L increases. As illustrated in Fig. 6, empirically, it appears that, even in the high-frequency regime, the error behaves as $L^{-1/2}$.

Finally, let us emphasise that the reported estimations were obtained by estimating H and the variance parameter ν^2 from which λ^2 is estimated using Eq. (15). We checked that estimating directly λ^2 instead of deriving it from ν^2, provides the same results. However we observed that the errors on ν^2 are much larger than the errors on λ^2. More precisely, it appears that, for a fixed λ^2, the measured bias is strongly related to $\widehat{H}$ as precisely expected from :

$$\widehat{\nu}^2 \simeq \frac{\lambda^2}{\widehat{H}(1-2\widehat{H})}. \tag{54}$$

This is illustrated in Fig. 7 in which two experiments where run with $H = 0.02$. For the first one we chose $\lambda^2 = 0.02$ and for the other one we chose $\lambda^2 = 0.1$.

For each sample, we have reported $\widehat{\nu}^2$ as a function of $\frac{1}{\widehat{H}(1-2\widehat{H})}$ estimated by $\mathrm{GMM}_{\ln M}$ method. One can easily see that in each case ($\lambda^2 = 0.02$ or $\lambda^2 = 0.1$), one observes a very large dispersion on $\widehat{\nu}^2$ (whose expected values should be respectively $\nu^2 = 1.04$ and $\nu^2 = 5.2$) that, however, strikingly appears to be proportional to $\frac{1}{\widehat{H}(1-2\widehat{H})}$ (which, when H is very small, has a large dispersion). As shown by the linear fits predicted by Eq (54) (continuous line in Fig. 7), the proportionality constant is precisely the value of the intermittency coefficient λ^2 for which the estimation is quite accurate. These observations suggest that while λ^2 can be estimated with a very small error, this is not at all the case of ν^2, when $H \ll 1$. The intermittency coefficient λ^2 appears to be a much more reliable

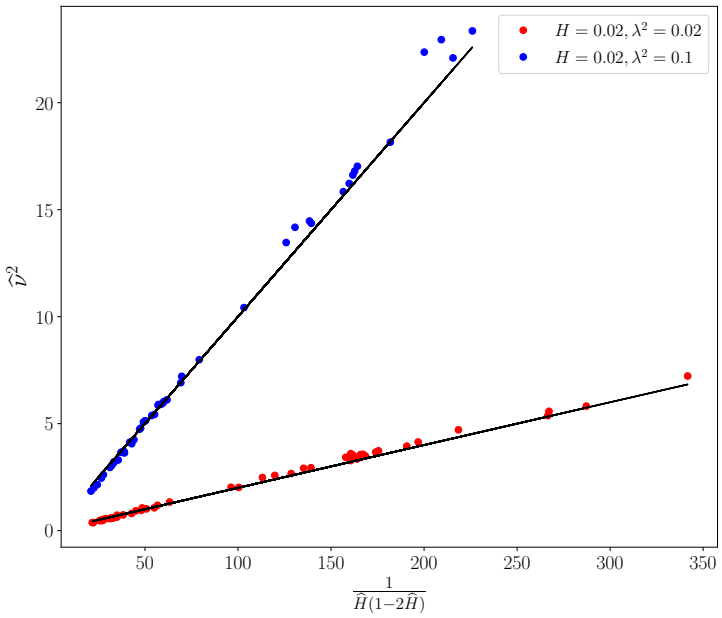

Figure 7: $\mathrm{GMM}_{\ln M}$ estimation of ν^2 as a function of $\frac{1}{\widehat{H}(1-2\widehat{H})}$. Each point corresponds to one estimation on a sample of length $L = 2^{14}$. The parameters are $H = 0.02$ and $\lambda^2 = 0.02$ (red symbols) or $\lambda^2 = 0.1$ (blue symbols). The straight lines represent the fit provided by Eq. (54).

quantity than the variance ν^2 of the log volatility. This can be easily explained by the fact that, in order for the S-fBM measure to converge when $H \to 0$ (towards the MRM $\tilde{M}$), one has to choose a variance proportional to $1/H$. Therefore, in the moment estimation method, in order to match the empirical covariance values when the estimated H is very small, the parameter ν^2 must scale as H^{-1}.

5 Application to realized volatility of asset returns

In this section we consider the application of the estimator of the former section to characterise the roughness exponent H and the intermittency coefficient λ^2 of realized volatility associated with various assets. Section 4 suggests that the GMM estimator $\mathrm{GMM}_{\ln M}$ outperforms the other candidate GMM_M. This is why we exclusively consider the $\mathrm{GMM}_{\ln M}$ applied to various empirical daily volatility data. Our study is based on 2 datasets containing respectively stock market indices and individual stock prices:

Oxford-Man Institute of Quantitative Finance Realized Library (OM)

The Oxford-Man Institute's Realized Library[1], contains historical records of various estimators of daily realized volatility of several stock indices. This dataset is widely used in various empirical studies and in particular, it was used as a benchmark database in many former studies on rough volatility (see e.g. [16, 10]). So we apply $\mathrm{GMM}_{\ln M}$ estimator to analyse the daily volatility time series associated with 24 major stock market indices considered in [10]. Following this latter work, in the following, we only report obtained results when using bipower variation volatility estimator but we have checked that the same results are obtained when using realized variance estimators at scale 5 min or 10 min. Two $\mathrm{GMM}_{\ln M}$ estimations are illustrated in Fig. 8 : one on CAC40 data and one on S&P500 data. The corresponding daily historical volatilities (using bipower-variation

[1] http://realized.oxford-man.ox.ac.uk/data

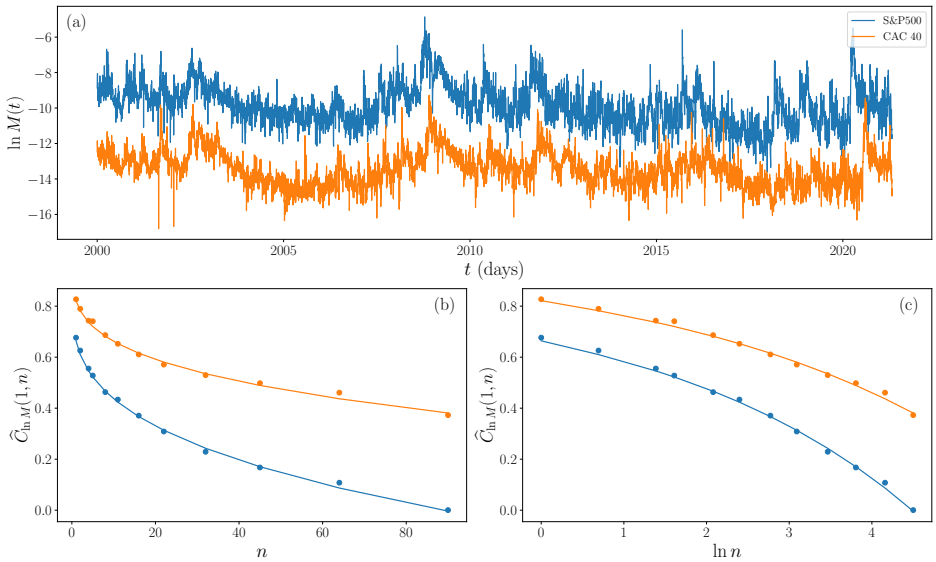

Figure 8: $\mathrm{GMM}_{\ln M}$ estimation of daily volatility of S&P500 and CAC40 indices from Oxford-Man dataset (a) log-realized bipower-variation from January 2000 to March 2021. (b) (resp. (c)) : the dots represent the estimations $\widehat{C}_{\ln M}(\tau,\Delta)$ of the corresponding correlation function $\widehat{C}_{\ln M}(\tau,\Delta)$ for each index as a function of τ (resp. $\ln\tau$). The plain lines correspond to the $\mathrm{GMM}_{\ln M}$ fits. The so-obtained estimated values of H are respectively $H \simeq 0.14$ (for S&P) and $H \simeq 0.13$ (for CAC40). CAC40 curves in Figures (a),(b) and (c) have been arbitrary shifted for the sake of clarity.

estimator) are illustrated in Fig. 8(a). We observe that over the 20-years period, the volatilities of S&P 500 and CAC40 are strongly correlated. One can also notice that some of the correlated departures from the mean value are lasting several years. This observation seriously questions any ergodic hypothesis that would result from short-term correlations as assumed in many papers (see, e.g., [16, 10]). Figs 8(b) (resp. (c)) displays the corresponding estimated correlation functions $\hat{C}_{\ln M}(\Delta,\tau)$ as a function of τ (resp. $\ln\tau$) and their $\mathrm{GMM}_{\ln M}$ fits. The so-obtained estimations for H are $H \simeq 0.14$ (for S&P) and $H \simeq 0.13$ (for CAC40). For both indices, we estimate $\lambda^2 \simeq 0.05$.

Yahoo Finance database (YF)

We collected historical daily open, high, low, and close price time-series of 296 individual stocks and also of a set of 24 stock indices from Yahoo Finance[2]. Stocks were taken from either the S&P 500 index (historical data from 1985-01-01 to 2021-12-31) or the CAC 40 index ((historical data from 2000-01-01 to 2021-12-31) while the indices were chosen as being those we considered in the OM database (over the period 200-01-01 to 2021-12-31). For each asset, we constructed a proxy of the daily volatility using Garman-Klass (GK) estimator described in [22]. This allows us, for any individual stocks or any index, to perform a $\mathrm{GMM}_{\ln M}$ estimation of H and λ^2 from the estimated log-volatility time series. As illustrated in Figs 10 and 11 below, on stock indices our results using Oxford-Man realized volatility and GK volatility estimation from YF data provide results that are fully consistent. In Fig. 9, following the exact same structure as Fig. 8, we illustrated the estimation procedure with the examples of Apple and Bank of America realized volatility. Again, we observe that volatility fluctuations seem to be long-term correlated. For the selected two stocks, the estimated values of H are respectively 0.01 and 0.02.

The $\mathrm{GMM}_{\ln M}$ estimations on all the 24 stock indices from both OM and YF databases and the 296 individual stocks from the YF database are summarized in Figures 10 and 11. In Fig. 10, we have reported the normalized histograms of the estimation $\widehat{H}$ for the Hurst exponents of the stock indices and the individual stocks of the two datasets. We can observe that the two distributions are

[2]http://YahooFinance.com

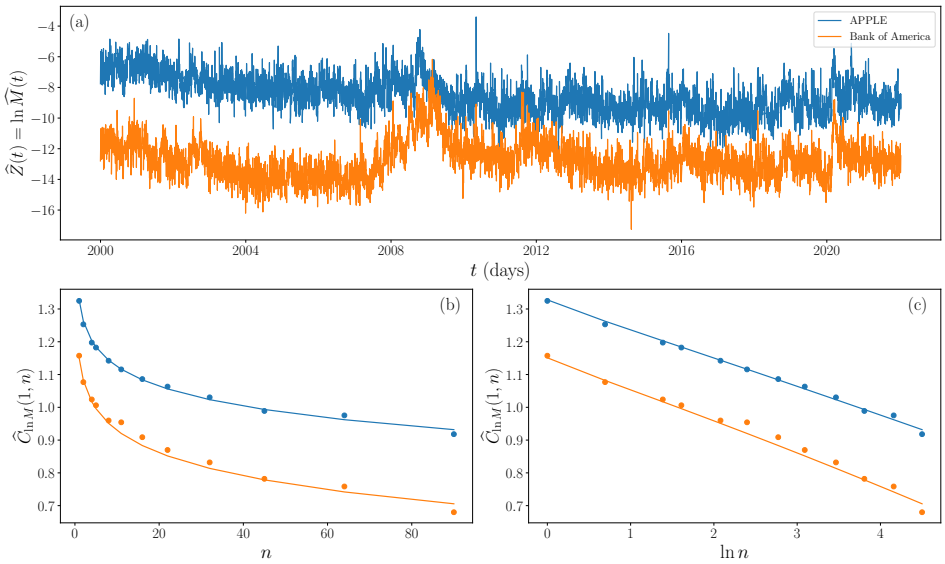

Figure 9: $\mathrm{GMM}_{\ln M}$ estimation of daily volatility of Apple and Bank of America from Yahoo Finance dataset. (a) log-realized Garman-Klass estimation of volatility [22] from January 2000 to December 2021. (b) (resp. (c)) : the dots represent the estimations $\widehat{C}_{\ln M}(\tau, \Delta)$ of the corresponding correlation function $\widehat{C}_{\ln M}(\Delta, \tau)$ for each index as a function of τ (resp. $\ln \tau$. The plain lines correspond to the $\mathrm{GMM}_{\ln M}$ fits. The so-obtained estimated values of H are respectively $H \simeq 0.01$ (for Apple) and $H \simeq 0.022$ (for Bank of Am.). Apple curves in Figures (a),(b) and (c) have been arbitrary shifted for the sake of clarity.

quite different: while the Hurst exponents of the stock market indices are spread around $H \simeq 0.13$ with a rather large dispersion (corresponding to a rms of 0.03), the distribution of H values of individual stocks is mainly peaked around a very small average value $H \simeq 0.01$ (with a rms of 0.015). It therefore clearly appears that the log-volatility of stock indices is much more regular than the log-volatility of individual stocks which turns out to be well described by a multifractal model characterized by $H = 0$. Moreover, in agreement with the findings of [16] (and in contrast with the results reported in [10]), Stock indices are confirmed to be well described by a "rough volatility" model with a typical value of the Hurst exponent close to $H = 0.15$. Notice that that estimation from either OM bipower variation realized volatility (orange bars) or YF Garman Klass realized volatility (green bars) provides a similar result[3].

As far as the intermittency coefficient λ^2 is concerned, we reported in Fig. 11(a) the estimated values $\hat{\lambda}^2$ for the 296 individual stocks (blue bullets) and the 24 stock indices (orange and green triangles for respectively OM and YF data) as a function of the logarithm of the estimated exponent $\hat{H}$. First, we can remark in both figures that OM and YF index data provide close estimations and therefore lead to the same conclusions. We can see that all the points are distributed around the value $\lambda^2 \simeq 0.07$ for stocks and $\lambda^2 \simeq 0.05$ for indices. In contrast, if one estimates the variance parameter ν^2, one observes a very large dispersion of its values. Actually, as it can be checked in Fig. 11(b), the data closely follow the curve $\nu^2 = \frac{0.07}{H(1-2H)}$ as represented by the solid line. Whether $\widehat{H}$ varies because H itself is varying or because of estimation errors, it appears that ν^2 is related to $\widehat{H}$ through the relationship (54). This suggests that the intermittency coefficient λ^2 is more likely to be the pertinent parameter to account for volatility fluctuations. Moreover, this latter quantity appears to be "almost universal" with a value $\lambda^2 \simeq 0.07$ for stocks and 0.05 for indices. Let us remark that, because of data dispersion due to estimation errors, one can not exclude in Fig. 11(a), a direct relationship between λ^2 and H, since one can observe that λ^2 slightly decreases as the Hurst exponent increases. A possible explanation could be that a linear combination of multifractal

[3]To be more precise, empirically the find a correlation coefficient of 0.7 between the two series of Hurst index estimates (i.e., from OM and YF data) over the 24 indices.

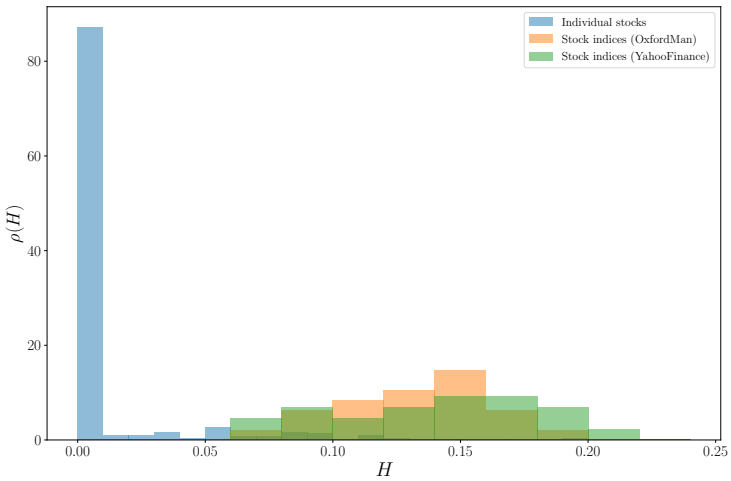

Figure 10: Estimation of the probability density distribution of Hurst exponent estimation $\hat{H}$ for the 296 individual stocks (blue bars) of the YF database and for the 24 stock market indices of the OM database (orange bars) or YF database (green bars).

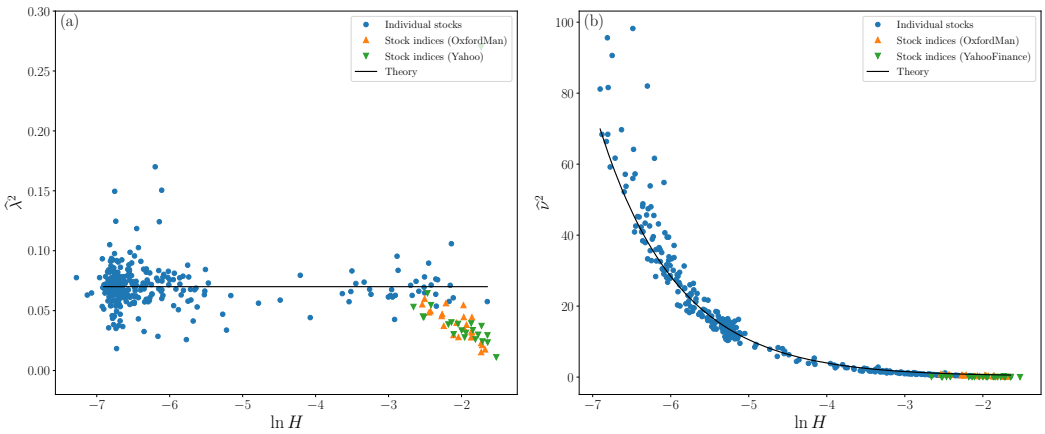

Figure 11: Estimation of the intermittency and variance parameters as a function of the estimated Hurst exponent. (a) Estimated intermittency coefficient $\widehat{\lambda^2}$ as a function of the logarithm of the estimated Hurst exponent H. The solid line represents the constant value $\lambda^2 = 0.07$ corresponding to the best fit of individual stock data. (b) Estimated variance coefficient $\widehat{\nu^2}$ as a function of the logarithm of the estimated Hurst exponent H. The solid line represents the log S-fBM expression (54). In (a) and (b) blue dots represent the individual stock data from YF database, orange up-pointing triangles represent index data from OM database while down-pointing green triangles correspond index data from YF database.

processes with some particular correlations in their increments or in their volatilities (the individual stock prices) may appear as a rough volatility process with an intermittency that depends on H (some market index). This question will be considered in future work.

6 Conclusion

We have introduced the log S-fBM, a class of log-normal "rough" random measures $M_{H,T}(\mathrm{d}t)$ that converge, when $H \to 0$, to the log-normal multifractal random measure. This model allows us to consider, within the same framework, the two popular classes of multifractal ($H = 0$) and rough volatility ($0 < H < 1/2$) models. Besides the roughness exponent H, the model involves 3 supplementary parameters: $\sigma^2 = \frac{\mathbb{E}(M_{H,T}(\mathrm{d}t))}{\mathrm{d}t}$ that provides the mean value of $M_{H,T}([t,t'])$, the

intermittency coefficient λ^2 which is related to the variance of $\ln M_{H,T}$ and the correlation length T (also referred to as the "integral scale" in the multifractal literature) above which the process values are independent. The second-order properties are studied and notably, we have computed the correlation function of $\ln M_{H,T}$ to the first order in λ^2. By studying the self-similarity properties of $M_{H,T}(\,\mathrm{d}t)$ when one changes the correlation length T, it appears that one cannot estimate T and σ^2 in the "high-frequency" estimation regime, i.e., if one observes, at a small scale $\Delta \ll T$, a single sample of $M_{H,T}$ over an interval of length $L = \mathcal{O}(T)$.

We design two efficient GMM estimation methods, GMM_M and $\mathrm{GMM}_{\ln M}$ based on the expressions of respectively $M_{H,T}$ and $\ln M_{H,T}$ correlation functions. We provide theoretical arguments and numerical evidence showing that very much like the method introduced in [3], $\mathrm{GMM}_{\ln M}$ provides an efficient estimation of H and λ^2 even in the high-frequency asymptotic regime.

We illustrate on various numerical examples that, when $H < 1/2$, the most pertinent parameter for accounting for volatility fluctuations is not, as it is always used in the rough volatility literature [16, 10], the variance parameter $\nu^2 = \frac{\lambda^2}{H(1-2H)}$, but the intermittency parameter λ^2. Indeed the estimation of the variance parameter is shown to have large fluctuations and to strongly depend on the estimation error on H.

Finally, when calibrating the log S-fBM model on a large set of empirical daily volatility data, we observe that stock market indices have values around $H = 0.1$ (close to a rough volatility behavior) whereas individual stocks are characterized by values of H that can be very close to 0 (close to a multifractal volatility behavior). The conclusions of section 4 concerning the Hurst estimation errors for the typical sample sizes we considered ($L \simeq 5.10^3$ corresponding to 20 years of daily data) and the agreement of the estimated values from two very different volatility estimators associated with YF and OM databases, make it unlikely that such an observation is caused by a statistical bias. Furthermore, nothing guarantees that a linear combination (indices) of multifractal processes (individual stocks) appears to be itself multifractal. It may simply appear as a "rougher" volatility process with eventually a smaller intermittency coefficient. This specific question will be addressed in future work. Finally, we pointed out that the estimations of the intermittency coefficient λ^2 are much more robust than the ones of the variance parameter ν^2. Its value seems to be quite universal and spread around $\lambda^2 = 0.07$ for stocks and $\lambda^2 = 0.05$ for stock market indices in agreement with the values formerly reported for the multifractal model [3].

In a future work, we will consider the issue of defining a faithful model for asset and option prices within the log S-fBm framework. To that end, the problem of introducing a specific skewness in our model will be considered along the same line as in Ref. [2]

A Appendix

A.1 Construction of the S-fBM process

In this Appendix, we explain in every details how the S-fBM process $\omega_{H,T}(t)$ is defined. It depends on three parameters :

- the (Hurst) parameter $H \in]0, 1[$,
- the decorrelation time scale $T > 0$
- and the coefficient $\lambda^2 > 0$ which is linked to the variance parameter ν^2 by

$$\nu^2 = \frac{\lambda^2}{H(1-2H)}.$$

  This parameter will be referred to as the intermittency parameter since it controls the intensity of intermittent "bursts" observed in $M_{H,T}$ and it is the name given to that quantity in the framework of MRM.

Construction of the S-fBM process $\omega_{H,T}(t)$
In the upper half-plane $(t, h) \in \mathcal{S} = \mathbb{R} \times \mathbb{R}^\star$, we first consider the area $C_{\ell,T}(t^*)$ illustrated in Fig. 12 which is defined as:

$$C_{\ell,T}(t^*) = \{(t, s)|h > \ell, |t - t^*| < \frac{1}{2}\min(h, T)\}. \tag{55}$$

For $\ell = 0$, we will use the notation $C_T(t^*) = C_{0,T}(t^*)$.

We then consider in $\mathcal{S}$ a non homogeneous Gaussian white noise $dG_H(t,h)$ of variance:

$$\mathrm{d}p_H(t,h) = \mathbb{E}\left(\mathrm{d}G_H(t,h)^2\right) = \lambda^2 h^{2H-2}\,\mathrm{d}h\,\mathrm{d}t. \tag{56}$$

We will see below that H is the analog of the Hurst parameter of the fBM process.

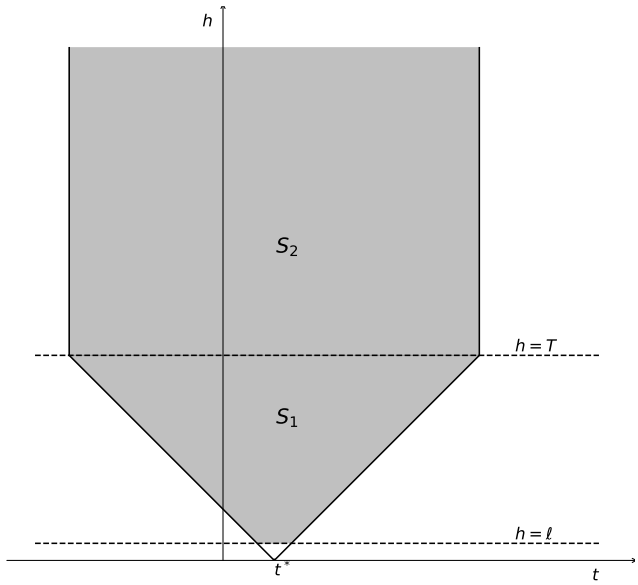

Figure 12: Definition of time-scale domain $C_{\ell,T}(t)$

We then define the Gaussian process $\omega_{H,T}(t)$ as:

$$\omega_{H,T}(t) = \mu_{H,T}(t) + \int_{C_T(t)} \mathrm{d}G_H, \tag{57}$$

where $\mu_{H,T}(t)$ is a normalising constant such that

$$\mathbb{E}[e^{\omega_{H,T}(t)}] = 1. \tag{58}$$

Covariance function of the S-fBM process $\omega_{H,T}(t)$
As a Gaussian process, the S-fBM is mainly characterised by its covariance function. This covariance can be directly calculated as the variance of integral of the random measure $\mathrm{d}G_H(t,h)$ on the overlapping area of $C_T(t_1)$ and $C_T(t_2)$ displayed in Fig. 13:

$$\mathrm{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \mathrm{Var}\left(\int_{C_{\ell,T}(t_1)\cap C_{\ell,T}(t_2)} \mathrm{d}G_H\right). \tag{59}$$

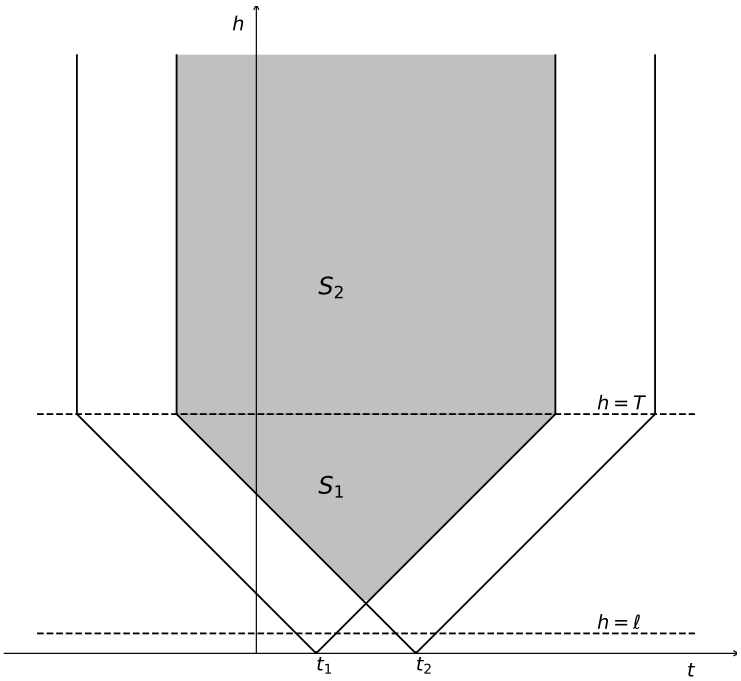

Figure 13: The overlapping area

Let us assume, without loss of generality, that $t_2 > t_1$ and denote $\tau = t_2 - t_1$. When $\tau > T$, $C_T(t_1) \cap C_T(t_2) = \emptyset$ and thus $\mathrm{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = 0$. For $\tau < T$, we have, using the notations of Fig. 13,

$$\begin{aligned} \mathrm{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) &= \int_{S_1 \cup S_2} \mathrm{d}p_H(h,t) \\ &= \int_{S_1} \mathrm{d}p_H(t,h) + \int_{S_2} \mathrm{d}p_H(h,t). \end{aligned} \tag{60}$$

Using (56), we have for the first term,

$$\begin{aligned} \int_{S_1} dp_H(h,t) &= \lambda^2 \int_{t_2-t_1}^{T} h^{2H-2}\,\mathrm{d}h \int_{t_2-h/2}^{t_1+h/2} \mathrm{d}t \\ &= \lambda^2 \int_{\tau}^{T} h^{2H-2}(h-\tau) \\ &= \frac{\lambda^2}{2H}\left(T^{2H} - \tau^{2H}\right) - \frac{\lambda^2}{2H-1}\tau\left(T^{2H-1} - \tau^{2H-1}\right). \end{aligned} \tag{61}$$

For the second term,

$$\begin{aligned} \int_{S_2} p(h,t)\,\mathrm{d}h\,\mathrm{d}t &= \lambda^2 (T-\tau) \int_{T}^{\infty} h^{2H-2}\,\mathrm{d}h \\ &= -\frac{\lambda^2}{2H-1}(T-\tau)T^{2H-1}. \end{aligned} \tag{62}$$

By composing the results above,

$$\mathrm{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \frac{\lambda^2}{2H(1-2H)}[T^{2H} - \tau^{2H}]. \tag{63}$$

Similarly, if we consider a strictly positive ℓ and $\tau < \ell$, direct calculation shows:

$$\mathrm{Cov}(\omega_{H,T}(t_1), \omega_{H,T}(t_2)) = \lambda^2 \left[\frac{1}{2H(1-2H)}(T^{2H} - \ell^{2H}) + \frac{\ell^{2H}}{1-2H}(1 - \frac{\tau}{\ell})\right]. \tag{64}$$

A.2 Takenaka fBM and proof of Proposition 1

Let us denote by $C(t_0)$ the full cone obtained by considering $T \to \infty$ in $C_T(t)$:

$$C(t_0) = \{(t,h)|h > 0, |t - t^*| < \frac{h}{2}\} \tag{65}$$

and consider the domain:

$$D(t) = C(t) \,\Delta\, C(0) = D_+(t) \cup D_-(t), \tag{66}$$

where Δ stands for the symmetric difference between two sets and $D_+(t)$, $D_-(t)$ are the two disjoint sets:

$$D_+(t) = C(t) - C(0) \;\text{ and }\; D_-(t) = C(0) - C(t).$$

Along the same line as definition (57), let us define the Gaussian processes:

$$\omega_\pm(t) = \int_{D_\pm(t)} dG_H \;\text{ and }\; B_H(t) = \omega_+(t) + \omega_-(t)\ . \tag{67}$$

Notice that $\omega_\pm(0) = 0$ and therefore $B_H(0) = 0$. It is easy to show that, after a little algebra that, for $0 \le t_1 \le t_2$:

$$\begin{aligned}
\mathbb{E}\left[\omega_-(t_1)\,\omega_-(t_2)\right] &= \frac{\lambda^2}{2H(1-2H)} t_1^{2H}, \\
\mathbb{E}\left[\omega_+(t_1)\,\omega_+(t_2)\right] &= \frac{\lambda^2}{2H(1-2H)} \left(t_2^{2H} - |t_2 - t_1|^{2H}\right), \\
\mathbb{E}\left[\omega_\pm(t_1)\,\omega_\mp(t_2)\right] &= 0\ .
\end{aligned}$$

It directly results that:

$$\gamma(t_1,t_2) = \mathbb{E}\left[B_H(t_1)\,B_H(t_2)\right] = \frac{\nu^2}{2}\left(t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}\right) \tag{68}$$

with $\nu^2 = \frac{\lambda^2}{H(1-2H)}$. Since $B_H(0) = 0$, we see that $B_H(t)$ is nothing but a fractional Brownian motion. This construction corresponds to the 1D version of Takenaka fractional Brownian fields as discussed in [30].

In order to prove Proposition 1, let us first work out

$$R_T(t,s) = \mathbb{E}\left[(\omega_{H,T}(t) - \omega_{H,T}(0))B_H(s)\right]\ .$$

This amounts to compute the "areas" of the intersections of $D_\pm(s)$ with $C_T(t)$ and $C_T(0)$ respectively. After a little algebra, one obtains, for any $0 \le t, s \le T$:

$$\begin{aligned}
R_T(t,s) &= \gamma(t,s) + \gamma'(t,s) \text{ with} \\
\gamma'(t,s) &= B\left(\left(\frac{T}{2}\right)^{2H} + \left(|t-s| + \frac{T}{2}\right)^{2H} - \left(t + \frac{T}{2}\right)^{2H} - \left(s + \frac{T}{2}\right)^{2H}\right)
\end{aligned}$$

where B is a positive constant depending on H and λ^2. Similarly, if

$$S_T(t,s) = \mathbb{E}\left[(\omega_{H,T}(t) - \omega_{H,T}(0))(\omega_{H,T}(s) - \omega_{H,T}(0))\right]\ ,$$

one has if $0 \le s, t \le T$:

$$S_T(t,s) = \gamma(t,s)\ . \tag{69}$$

Let

$$Z_{H,T}(t) = \omega_{H,T}(t) - \omega_{H,T}(0) - B_H(t)$$

and

$$d(t,s) = \left[\mathbb{E}\left(Z_{H,T}(t) - Z_{H,T}(s)\right)^2\right]^{1/2}\ .$$

By expanding the square $\left(Z_{H,T}(t) - Z_{H,T}(s)\right)^2$ one directly obtains:

$$\begin{aligned}
d(t,s)^2 &= 2\gamma(t,t) - 2R_T(t,t) + 2\gamma(s,s) - 2R_T(s,s) - 4\gamma(t,s) + 4R_T(t,s) \\
&= -2\gamma'(t,t) - 2\gamma'(s,s) + 4\gamma'(t,s) \\
&= 4B\left(\left(|t-s| + \frac{T}{2}\right)^{2H} - \left(\frac{T}{2}\right)^{2H}\right).
\end{aligned}$$

Therefore, $\forall t_0$, $t, s < t_0$, we have when $T \to \infty$:

$$d(t,s) = CT^{H-\frac{1}{2}}|t-s|^{1/2} + o(T^{H-\frac{1}{2}}) \ . \tag{70}$$

On can consider $d(t,s)$ as a metric and define $N(t_0,\varepsilon)$ as the number of boxes of $d-$radius ε need to cover the set $[0,t_0]$. Let $D = \sup_{t,s\in[0,t_0]} d(t,s)$. Then, according to Dudley inequality [23], there exists a positive universal constant K such that:

$$\mathbb{E}(\sup_{t\in[0,t_0]} |Z_H(t)|) \leq K \int_0^D \sqrt{\log N(t_0,\varepsilon)}\, \mathrm{d}\varepsilon. \tag{71}$$

From Eq. (70), one has $D \sim CT^{H-\frac{1}{2}} t_0^{\frac{1}{2}}$. Moreover, one has

$$N(t_0,\varepsilon) \simeq 1 + \left\lfloor \frac{t_0 T^{2H-1} C^2}{\varepsilon^2} \right\rfloor = 1 + \left\lfloor \frac{D^2}{\varepsilon^2} \right\rfloor,$$

where $\lfloor x \rfloor$ stands for the largest integer not greater than x. We then have when ε is small with respect to D:

$$\log N(t_0,\varepsilon) \simeq 2\ln(D) - 2\ln(\varepsilon).$$

Thus

$$\int_0^D \sqrt{\log N(t_0,\varepsilon)} d\varepsilon \sim \int_0^D \sqrt{2\ln(D) - 2\ln(\varepsilon))} d\varepsilon = \sqrt{\frac{\pi}{2}} D,$$

since for $H < 1/2$, $D \to 0$ when $T \to \infty$. Proposition 1 follows directly from inequality (71).

A.3 The case $H \to 0$: Convergence towards the MRM log-normal measure

We now examine the case $H = 0$ in the geometric construction above. The definition of $C_{\ell,T}$ remains unchanged and we consider the Gaussian random noise when $H = 0$, $dG_0(t,h)$ of variance:

$$\mathrm{d}p_0(t,h) = \lambda^2 h^{-2}\, \mathrm{d}h\, \mathrm{d}t \tag{72}$$

Then we define a random process $\omega_{\ell,T}$ as previously

$$\omega_{\ell,T}(t) = \mu_{\ell,T} + \int_{C_{\ell,T}(t)} \mathrm{d}G_0. \tag{73}$$

As proven in [4], provided $\mu_{\ell,T}$ is chosen such that $\mathbb{E}e^{\omega_{\ell,T}(t)} = 1$, when $\ell \to 0$ we have

$$\widetilde{M}_{\ell,T}(\,\mathrm{d}t) = e^{\omega_{\ell,T}(t)}\, \mathrm{d}t \xrightarrow[\ell\to 0]{w} \widetilde{M}_T(dt), \tag{74}$$

where $\xrightarrow{w}$ stands for the weak convergence and where $\widetilde{M}_T(dt)$ is the so-called log-normal "Multifractal Random Measure" (MRW), a non trivial singular continuous random measure with exact multifractal properties [24, 4, 3].
In [24, 4], it is also shown that the covariance of $\omega_{\ell,T}$ reads (for $\tau \geq 0$):

$$\mathrm{Cov}(\omega_{\ell,T}(t), \omega_{\ell,T}(t+\tau)) = \begin{cases} \lambda^2 \ln\left(\frac{T}{\tau}\right) & \text{if } \ell \leq \tau \leq T \\ \lambda^2 \left(\ln\left(\frac{T}{\ell}\right) + 1 - \frac{\tau}{\ell}\right) & \text{if } \tau \leq \ell \\ 0, & \text{otherwise} \end{cases} \tag{75}$$

We can remark that this expression of the covariance of $\omega_{\ell,T}$ in the range $\tau \geq \ell$, can be recovered from Eq. (63), (64) when $H \to 0$.

Let us show a strong mean square convergence of S-fBM to MRM when $H \to 0$ as claimed in Proposition 1.
Since $\widetilde{M}(t) = \widetilde{M}([0,t])$ is regular enough, in order to establish the weak convergence we just have to prove that $\forall t$,

$$\lim_{H\to 0} \mathbb{E}[(M_{H,T}[0,t] - \widetilde{M}_T[0,t])^2] = \lim_{H\to 0} \lim_{\ell\to 0} \mathbb{E}[(M_{H,T}[0,t] - \widetilde{M}_{\ell,T}[0,t])^2] = 0. \tag{76}$$

Before starting, it is useful to calculate the covariance between $\omega_{H,T} = \omega_{\ell=0,H,T}$ and $\omega_{\ell,T} = \omega_{\ell,H=0,T}$. Following similar computation as in Appendix A.1,

$$\mathrm{Cov}(\omega_{H,T}(t_1),\omega_{\ell,T}(t_2)) = \begin{cases} \frac{\lambda^2}{H(1-H)}(T^H-\tau^H) & \text{if } \ell \le \tau \le T \\ \lambda^2\Big(\frac{1}{H(1-H)}(T^H-\ell^H)+\frac{\ell^H}{1-H}(1-\frac{\tau}{\ell})\Big) & \text{if } \tau < \ell \\ 0 & \text{otherwise} \end{cases} \tag{77}$$

By expanding the square in Eq. (76), we have:

$$\mathbb{E}[(M_{H,T}[0,t]-\widetilde{M}_{\ell,T}[0,t])^2] = \int_0^t\int_0^t \Big(e^{\mathrm{Cov}[\omega_{H,T}(u),\omega_{H,T}(v)]} + e^{\mathrm{Cov}[\omega_{\ell,T}(u),\omega_{\ell,T}(v)]} - 2e^{\mathrm{Cov}[\omega_{H,T}(u),\omega_{\ell,T}(v)]}\Big)\,\mathrm{d}u\,\mathrm{d}v. \tag{78}$$

Since, for a symmetric function f, one has:

$$\int_0^t\int_0^t f(u-v)\ \mathrm{d}u\,\mathrm{d}v = 2\int_0^t (t-z)f(z)\,\mathrm{d}z.$$

Then the previous expression becomes:

$$\mathbb{E}[(M_{H,T}[0,t]-\widetilde{M}_{\ell,T}[0,t])^2] = 2\int_0^t \mathrm{d}z(t-z)\Big(e^{\mathrm{Cov}[\omega_{H,T}(0),\omega_{H,T}(z)]} + e^{\mathrm{Cov}[\omega_{\ell,T}(0),\omega_{\ell,T}(z)]} - 2e^{\mathrm{Cov}[\omega_{H,T}(0),\omega_{\ell,T}(z)]}\Big). \tag{79}$$

Let us split this integral as a sum of two integrals, I_1 and I_2 according to whether one considers the integration domains $z > \ell$ and $z \le \ell$ respectively. In the first case, by replacing the covariance by their expressions, one has:

$$I_1 = \int_\ell^t (t-z)\left(e^{\frac{\lambda^2}{2H(1-2H)}(T^{2H}-z^{2H})} + \left(\frac{T}{z}\right)^{\lambda^2} - 2e^{\frac{\lambda^2}{H(1-H)}(T^H-z^H)}\right)\mathrm{d}z. \tag{80}$$

Since $\lambda^2 < 1$, one can the safely take $\ell \to 0$ in the lower integral bound and then, thanks to dominated convergence theorem, observe that I_1 converges to 0 when $H \to 0$ since the expression inside the integral vanishes in this limit. The second integral, when $z \le \ell$ is:

$$I_2 = \int_0^\ell (t-z)\left(e^{\frac{\lambda^2}{2H(1-2H)}(T^{2H}-\tau^{2H})} + e^{\lambda^2(\frac{T}{\ell}+1-\frac{z}{\ell})} - 2e^{\lambda^2\left(\frac{T^H-\ell^H}{H(1-H)}+\frac{\ell^H}{1-H}(1-\frac{z}{\ell})\right)}\right)\mathrm{d}z. \tag{81}$$

For $0 \le z \le \ell$, the first and last terms inside the integral can be bounded by a constant that does not depend on ℓ while the second term can be bounded by $C\ell^{-\lambda^2}$. Therefore we can see that, if $\lambda^2 < 1$, $I_2 \to 0$ when $\ell \to 0$. This concludes the proof.

A.4 Proof of Eqs. (21) and (22)

Let us compute the analytical expression of $C_M(\tau,\Delta) = \mathbb{E}[M_{H,T,\Delta}(t)M_{H,T,\Delta}(t+\tau)]$ and establish expressions (21) and (22). For that purpose, let us first remark that from the definition (18) of $M_{H,T,\Delta}(t)$ and from the expression (8) of the covariance of $\omega_{H,T}(t)$, we have (when $\tau < T$):

$$C_M(\tau,\Delta) = K_1\int_0^\Delta \mathrm{d}u\int_\tau^{\tau+\Delta}\mathrm{d}v\ e^{-K_2|u-v|^{2H}} \tag{82}$$

with $K_1 = e^{K_2T^{2H}}$ and $K_2 = \frac{\lambda^2}{2H(1-2H)}$.
Moreover, let us prove that, if $f(z)$ is a symmetric function, then

$$\int_0^\Delta \mathrm{d}u\int_\tau^{\tau+\Delta}\mathrm{d}v f(u-v) = \int_0^\Delta \mathrm{d}z(\Delta-z)\Big(f(z+\tau)+f(z-\tau)\Big). \tag{83}$$

Indeed, as shown in [28], we have, when $\tau = 0$:

$$\int_0^\Delta \mathrm{d}u\int_0^\Delta \mathrm{d}v f(u-v) = 2\int_0^\Delta \mathrm{d}z(\Delta-z)f(z).$$

In the l.h.s. of (83), let us set $v' = v - \tau$ and use respectively symmetry argument and previous expression to obtain

$$\begin{aligned}
l.h.s. &= \int_0^\Delta \mathrm{d}u \int_0^\Delta \mathrm{d}v f(\tau + u - v) \\
&= \int_0^\Delta \mathrm{d}u \int_0^u \mathrm{d}v f(\tau + |u - v|) + \int_0^\Delta \mathrm{d}u \int_u^\Delta \mathrm{d}v f(\tau - |u - v|) \\
&= \frac{1}{2}\Big(\int_0^\Delta \mathrm{d}u \int_0^\Delta \mathrm{d}v f(\tau + |u - v|) + \int_0^\Delta \mathrm{d}u \int_0^\Delta \mathrm{d}v f(\tau - |u - v|)\Big) \\
&= \int_0^\Delta \mathrm{d}z (\Delta - z)\Big(f(z + \tau) + f(z - \tau)\Big).
\end{aligned}$$

By using (83) in (82), we have:

$$\begin{aligned}
C_M(\tau, \Delta) &= K_1 \int_0^\Delta \mathrm{d}z\ (\Delta - z)\ \Big(e^{-K_2|\tau+z|^{2H}} e^{-K_2|\tau-z|^{2H}}\Big) \\
&= K_1 \int_\tau^{\tau+\Delta} \mathrm{d}z\ (\Delta + \tau - z) e^{-K_2|z|^{2H}} + K_1 \int_{\tau-\Delta}^{\tau} \mathrm{d}z\ , (\Delta - \tau + z) e^{-K_2|z|^{2H}} \\
&= F(\tau + \Delta) + F(\tau - \Delta) - 2F(\tau),
\end{aligned}$$

where we have denoted

$$F(x) = K_1\Big(x \int_0^x \mathrm{d}z\ e^{-K_2|z|^{2H}} - \int_0^x \mathrm{d}z\ z e^{-K_2|z|^{2H}}\Big)\ .$$

If one considers the lower-incomplete Gamma function $\gamma(a, z)$,

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t}\, \mathrm{d}t$$

and makes the change of variable $t = K_2|z|^{2H}$ in previous integrals, one obtains the following exact expression for $F(x)$:

$$F(x) = \frac{K_1}{2H}\Big(\frac{x}{K_2^{\frac{1}{2H}}}\gamma(\frac{1}{2H}, K_2 x^{2H}) - \frac{1}{K_2^{\frac{1}{H}}}\gamma(\frac{1}{H}, K_2 x^{2H})\Big),$$

which corresponds to Eq. (22). When $H = 0$, i.e. for $\widetilde{M}_T$, one can show that the former expression reduces to:

$$F(x) = \frac{T^{\lambda^2}}{(2 - \lambda^2)(1 - \lambda^2)} x^{2-\lambda^2}.$$

A.5 Proof of proposition 6

In this section we provide a proof of Proposition 6 based on small intermittency approximation of Proposition 4. Let $\Delta = 1$ and $N = \frac{L}{\Delta} = L$ be the number of samples $M_{H,T,1}$ in the interval $[0, L]$. We will suppose that Ł $\to \infty$ with $T = CL$, so that we are in the high frequency regime. Let us consider the empirical mean:

$$\widehat{\mu}_N = \frac{1}{N}\sum_{k=1}^{N} \ln M_{H,T,\Delta=1}(k) \tag{84}$$

and define the "centered" random variable:

$$Z(k) = \ln M_{H,T,1}(k) - \widehat{\mu}_N\ . \tag{85}$$

If $C_Z(k) = \mathrm{Cov}\,[Z(j), Z(j + k)]$, one has obviously:

$$C_Z(k) = C_{\ln M}(1, k) - \mathrm{Var}\,[\widehat{\mu}_N]\ . \tag{86}$$

One can use Proposition 4 to compute, to the first order in λ^2, all terms in Eq. (86). Indeed, the expression of $C_{\ln M}(k)$ is provided by Proposition 5 (Eq. (28)) and order to compute $\mathrm{Var}\,[\widehat{\mu}_{\Delta,N}]$, one can use Prop. 4 to show that, to the first order in λ^2,

$$\mathrm{Var}\,[\widehat{\mu}_N] = \mathrm{Var}\,[\ln M_{H,T,\Delta=L}]$$

and therefore, from expression (28), one has:

$$\mathrm{Var}\left[\widehat{\mu}_{\Delta,N}\right] = \frac{\lambda^2}{2H(1-2H)}\Big(T^{2H} - \frac{L^{2H}}{(2H+1)(H+1)}\Big)\,. \tag{87}$$

It thus results that:

$$C_Z(k) = \frac{\lambda^2}{2H(1-2H)}\left(\frac{L^{2H}}{(1+2H)(1+H)} - \frac{|k+1|^{2H+2}+|k-1|^{2H+2}-2|k|^{2H+2}}{(2H+1)(2H+2)}\right). \tag{88}$$

Let us consider the empirical covariance:

$$\widehat{C}_{\ln M}(1,k) = N^{-1}\sum_{k=1}^{N-k}\Big(\ln M_{H,T,1}(j) - \widehat{\mu}_N\Big)\Big(\ln M_{H,T,1}((j+k)) - \widehat{\mu}_N\Big), \tag{89}$$

Since one has $\mathbb{E}\left[D(n)\right] = \mathbb{E}\left[\widehat{C}_{\ln M}(1,n) - \widehat{C}_{\ln M}(1,0)\right] = C_Z(k) - C_Z(0) = \widetilde{D}(n)$ (as defined in Eq. (44)), in order to prove Eq. (43), it is sufficient to show that

$$\lim_{N\to\infty}\mathrm{Var}\left[D(n)\right] = 0$$

To that end, remark that, from the definition of $D(n)$,

$$\begin{aligned}\mathrm{Var}\left[D(n)\right] &= N^{-2}\sum_{i=1}^{N-n}\sum_{j=1}^{N-n}\mathrm{Cov}\left[Z(i)Y(i), Z(j)Y(j)\right] && (90)\\ &= N^{-2}\sum_{i=1}^{N-n}\mathrm{Var}\left[Z(i)Y(i)\right] + 2N^{-2}\sum_{i=1}^{N-n}\sum_{j=i+1}^{N-n}\mathrm{Cov}\left[Z(i)Y(i), Z(j)Y(j)\right] && (91)\end{aligned}$$

where we have denoted

$$Y(i) = Z(i+n) - Z(i).$$

From proposition 4, because $\Omega_{T,H,\Delta}(t)$ is a Gaussian process, we have, to the first order in λ^2,

$$\begin{aligned}\mathrm{Cov}\left[Z(i)Y(i), Z(j)Y(j)\right] &= \mathrm{Cov}\left(Z(i),Z(j)\right)\mathrm{Cov}\left(Y(i),Y(j)\right) + \mathrm{Cov}\left(Z(i),Y(j)\right)\mathrm{Cov}\left(Z(j),Y(j)\right)\\ &= C_Z(j-i)\Big(2C_Z(j-i) - C_Z(i-j+n) - C_Z(j-i-n)\Big)\\ &+ \Big(C_Z(j+n-i) - C_Z(j-i)\Big)\Big(C_Z(j-i-n) - C_Z(j-i)\Big)\,.\end{aligned}$$

Thereby, from the expression (88) of $C_Z(k)$, after a little algebra, one can show that there exists a constant C such that

$$\mathrm{Cov}\left[Z(i)Y(i), Z(j)Y(j)\right] \le CN^{2H}(1+|i-j|)^{2H-2}\,.$$

Then, Eq. (91) gives:

$$\mathrm{Var}\left[D(n)\right] \le CN^{2H-1} + 2CN^{2H-2}\sum_{i=1}^{N}\int_i^N x^{2H-2}dx \le C'N^{2H-1} \tag{92}$$

and thus, if $H < \frac{1}{2}$,

$$\lim_{N\to\infty}\mathrm{Var}\left[D(n)\right] = 0.$$

This concludes the proof of Eq. (43).

Acknowledgement

This research is partially supported by the Agence Nationale de la Recherche as part of the "Investissements d'avenir" program (reference ANR-19-P3IA-0001; PRAIRIE 3IA Institute).

References

[1] E. Bacry, J. Delour, and J. F. Muzy. Multifractal random walk. *Phys. Rev. E*, 64:026103, Jul 2001.

[2] E. Bacry, L. Duvernet, and J.F. Muzy. Continuous-time skewed multifractal processes as a model for financial returns. *Journal of Applied Probability*, 49(2):482 – 502, 2012.

[3] E. Bacry, A. Kozhemyak, and J. F. Muzy. Log-normal continuous cascade model of asset returns: aggregation properties and estimation. *Quantitative Finance*, 13(5):795–818, 2013.

[4] E. Bacry and J.F. Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3):449–475, 2003.

[5] O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J.R. Statist. Soc. B*, 64, 2002.

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

[7] C. Bayer, F.A. Harang, and P. Pigato. Log-modulated rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 12(3):1257–1284, 2021.

[8] M. Bennedsen. Semiparametric estimation and inference on the fractal index of gaussian and conditionally gaussian time series data. *Econometric Reviews*, 39(9):875–903, Feb 2020.

[9] M. Bennedsen, A. Lunde, and M.S. Pakkanen. Decoupling the Short- and Long-Term Behavior of Stochastic Volatility. *Journal of Financial Econometrics*, Jan 2021.

[10] A.E. Bolko, K. Christensen, M.S. Pakkanen, and B. Veliyev. Roughness in spot variance? A gmm approach for estimation of fractional log-normal stochastic volatility models using realized measures. *arXiv:2010.04610*, 2020.

[11] Masaaki F. Volatility has to be rough. *Quantitative Finance*, 21(1):1–8, 2021.

[12] M. Forde, M. Fukasawa, S. Gerhold, and B. Smith. The rough bergomi model as $h \to 0$ skew flattening/blow up and non-gaussian rough volatility, 2020.

[13] M. Forde and B. Smith. The riemann-liouville field as a $h \to 0$ limit- sub, critical and super critical gmc, decompositions and explicit spectral expansions, 2020.

[14] M. Fukasawa, T. Takabatake, and R. Westphal. Is volatility rough ? *arXiv:1905.04852v2*, 2019.

[15] Y. V. Fyodorov, B. A. Khoruzhenko, and N. J. Simm. Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian Unitary Ensemble. *The Annals of Probability*, 44(4):2980 – 3031, 2016.

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

[17] S. Ghashghaie, W. Breymann, J. Peinke, and P. Talkner. Turbulence and financial markets. In S. Gavrilakis, L. Machiels, and P. A. Monkewitz, editors, *Advances in Turbulence VI*, pages 167–170, Dordrecht, 1996. Springer Netherlands.

[18] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381:767—770, 1996.

[19] P. Hager and E. Neumann. The multiplicative chaos of $h = 0$ fractional brownian fields. Technical report, 2021.

[20] G. Livieri, S. Mouti, A. Pallavicini, and M. Rosenbaum. Rough volatility: Evidence from option prices. *IISE Transactions*, 50(9):767–776, 2018.

[21] B. B. Mandelbrot, A. J. Fisher, and L. E. Calvet. A multifractal model of asset returns. *Cowles Foundation Discussion Paper No. 1164, Sauder School of Business Working Paper, Available at SSRN*, Sep 1997.

[22] Garman M.B and Klass M.J. On the estimation of security price volatility from historical data. *The Journal of Business*, 53:67–78, 1980.

[23] Y. Mishura and M. Zili. Gaussian processes. In Y. Mishura and M. Zili, editors, *Stochastic Analysis of Mixed Fractional Gaussian Processes*, pages 1–29. Elsevier, 2018.

[24] J.F. Muzy and E. Bacry. Multifractal stationary random measures and multifractal random walks with log infinitely divisible scaling laws. *Physical Review E*, 66(5), Nov 2002.

[25] J.F. Muzy, R. Baïle, and E. Bacry. Random cascade model in the limit of infinite integral scale as the exponential of a nonstationary 1/f noise: Application to volatility fluctuations in stock markets. *Physical Review E*, 87(4), 2013.

[26] J.F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *Eur. Phys. J. B*, 17(3):537–548, 2000.

[27] E. Neuman and M. Rosenbaum. Fractional brownian motion with zero hurst parameter: a rough volatility viewpoint. *Electronic Communications in Probability*, 23, 2018.

[28] M. Rambaldi, E. Bacry, and J.F. Muzy. Disentangling and quantifying market participant volatility contributions. *Quantitative Finance*, 19(10):1613–1625, 2019.

[29] R. Rhodes and V. Vargas. Gaussian multiplicative chaos and applications: A review. *Probability Surveys*, 11:315–392, 05 2013.

[30] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance: Stochastic Modeling.* 1994.