Quantitative Finance

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

Detecting rough volatility: a filtering approach

Camilla Damian & Rüdiger Frey

To cite this article: Camilla Damian & Rüdiger Frey (2024) Detecting rough volatility: a filtering approach, Quantitative Finance, 24:10, 1493-1508, DOI: 10.1080/14697688.2024.2399284

To link to this article: https://doi.org/10.1080/14697688.2024.2399284

Published online: 25 Sep 2024.

Quantitative Finance, 2024
Vol. 24, No. 10, 1493–1508, https://doi.org/10.1080/14697688.2024.2399284

Detecting rough volatility: a filtering approach

CAMILLA DAMIAN*† and RÜDIGER FREY‡

†Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
‡Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Vienna, Austria

(*Received 15 September 2023; accepted 17 August 2024*)

In this paper, we focus on filtering and parameter estimation in stochastic volatility models when observations arise from high-frequency data. We are particularly interested in rough volatility models where spot volatility is driven by fractional Brownian motion with Hurst index $H < \frac{1}{2}$. Since volatility is not directly observable, we rely on particle filtering techniques for statistical inference regarding the current level of volatility and the parameters governing its dynamics. In order to obtain numerically efficient and recursive algorithms, we use the fact that a fractional Brownian motion can be represented through a superposition of Markovian semimartingales (Ornstein-Uhlenbeck processes). We analyze the performance of our approach on simulated data and we compare it to similar studies in the literature. The paper concludes with an empirical case study, where we apply our methodology to high-frequency data of a liquid stock.

Keywords: High-frequency data; Rough volatility; Fractional Brownian motion; Stochastic filtering; Nested particle filter

1. Introduction

A stylized fact of empirical finance states that volatility fluctuates randomly and should therefore be modeled as a stochastic process. In classical stochastic volatility models, this process is given by the solution of a stochastic differential equation (SDE) driven by a Brownian motion. More recently, the influential paper (Gatheral *et al.* 2018) advocates so-called *rough volatility models*, where the log-volatility is driven by a fractional Brownian motion with Hurst index $H < \frac{1}{2}$. In these models the paths of the volatility process are Hölder-continuous for every exponent $h < H < \frac{1}{2}$ and hence less regular (rougher) than solutions of SDEs driven by Brownian motion. In order to justify the use of rough volatility models, Gatheral *et al.* (2018) employ proxy data for historical volatility to estimate the smoothness of the volatility process with a regression approach. They find that the increments of log-volatility of various assets enjoy a monofractal scaling property; together with the well-established fact that the distribution of these increments is approximately Gaussian, this implies that log-volatility behaves essentially as a fractional Brownian motion. In particular, given their estimates of the smoothness parameter, these authors claim that log-volatility can be modeled as a fractional Brownian motion with Hurst exponent $H \approx 0.1$.

These findings were challenged by Fukasawa *et al.* (2019) and more recently by Cont and Das (2022) and Rogers (2023), who claim that the roughness of volatility might be spurious and due to observation errors. These papers stress that historical volatility is not directly observable, so that its level needs to be estimated from the observable time series of asset prices. Hence, the volatility data entering into the regression approach of Gatheral *et al.* (2018) (the so-called *realized volatility* data) are noisy estimates of the true spot volatility, where the estimation error is attributed to *microstructure noise*. Fukasawa *et al.* (2019) use simulation experiments to show that, in the presence of microstructure noise, the regression method of Gatheral *et al.* (2018) leads to an estimate of $\hat{H} \approx 0.1$, *regardless* of the true value of H used in the simulation.† In a similar vein, Cont and Das (2022) study a nonparametric roughness estimator. Their numerical experiments show that, in absence of microstructure noise, this estimator differentiates well between rough and non-rough volatility. However, if volatility is estimated from simulated asset price trajectories, this nonparametric estimator always points to a rough behavior, even if spot volatility follows a classical stochastic volatility model driven by a Brownian motion.

*Corresponding author. Email: c.damian@vu.nl

† In particular, they consider the case where the true log-volatility process is a standard Ornstein Uhlenbeck process corresponding to $H = \frac{1}{2}$.

Identifying the 'true' cause for the apparent roughness in realized volatility thus constitutes a challenging statistical problem. Fukasawa *et al.* (2019) tackle this problem in a setting where the microstructure noise is modeled as an independent and identically distributed series of error terms with decaying variance, independent of the underlying volatility process. They propose a modified Whittle estimator for H (essentially a quasi-maximum likelihood estimator) and study its asymptotic properties. When applying their estimator to real data, Fukasawa *et al.* (2019) find that volatility is indeed rough: remarkably, their estimates for H for various stock indices are not only smaller than 0.5, but even smaller than those of Gatheral *et al.* (2018). In a similar spirit, Bennedsen *et al.* (2021) study a nonlinear version of the regression approach of Gatheral *et al.* (2018), which is robust with respect to microstructure noise. Applying this method to the S&P500 E-mini futures contract, they obtain estimates of the order of $\hat{H} \approx 0.15$ for the roughness parameter of spot volatility.

In this paper, we adopt a Bayesian perspective and propose to estimate the current level and the roughness parameter of the spot volatility process via stochastic filtering techniques. We consider a model that is suitable for ultra high-frequency data and assume that the asset price process is piecewise constant and jumps when new orders arrive to the market. We model these times as jump times of a doubly-stochastic Poisson process with a stochastic intensity. We show that for frequent small jumps—the typical case for high frequency asset price data—this model is close to a diffusion model for volatility where the instantaneous variance is proportional to the intensity of the jump model. In this way our analysis can be understood as a contribution to the literature on rough volatility models. To mimic a real trading setting, we assume that the analyst observes only the price trajectory, so that the intensity—or equivalently its logarithm—plays the role of an unobservable signal process. To account for roughness, we model the logarithmic intensity as a fractional Brownian motion (fBM) with a Hurst index $H < \frac{1}{2}$. Hence the the signal process is neither a semimartingale nor a Markov process, posing considerable challenges for filtering and parameter estimation. In fact, even in the flexible context of particle filtering, the non-Markovianity of the signal implies that sampling a particle at a given time instance requires calculations involving the entire trajectory up to that point. Clearly, the practical consequence is a great increase in computational cost and in memory allocation requirements over time. We overcome this issue by employing a suitable approximation of the underlying non-Markovian volatility process. The starting point is the work of Carmona and Coutin (1998) and Carmona *et al.* (2000), who show that an fBM with $H < \frac{1}{2}$ can be approximated by a finite superposition of Ornstein-Uhlenbeck (OU) processes with different mean-reversion speeds but driven by the same Brownian motion. Using these results, we obtain a finite-dimensional Markovian approximation to the log-intensity process, which allows us to use a standard recursive particle filter to retrieve the signal. Moreover, we can make inference about the Hurst index by adapting the *nested particle filter algorithm* of Crisan and Miguez (2018) to our specific framework.

Simulation experiments indicate that this approach yields satisfactory results, both in terms of filtering and of parameter estimation, while still fully taking into account the intrinsic unobservability of the volatility process. Next, in the spirit of Cont and Das (2022) and Rogers (2023), we apply our methodology to synthetic data generated in a setting where spot volatility is driven by a Brownian diffusion. We find that our filtering approach seems to be less sensitive to the issue of market microstructure noise than the nonparametric roughness estimator analyzed in Cont and Das (2022). This shows that it might indeed be possible to detect rough volatility via stochastic filtering. We proceed with an analysis of the so-called OU-OU model proposed in Rogers (2023). Finally, we carry out a small empirical case study where we apply our methodology to high-frequency data for the mid-price of Apple shares and obtain an estimate for the Hurst parameter H which is substantially smaller than $\frac{1}{2}$, suggesting that volatility paths indeed exhibit roughness.

The remainder of this paper is as follows. Section 2 introduces the model. In section 3 we discuss the approximation of fractional Brownian motion by a superposition of Ornstein Uhlenbeck processes, which is a prerequisite for the filtering algorithms in section 4. The numerical experiments are discussed in section 5. We present an application of the proposed methodology to the 'alternative' models of Cont and Das (2022) and Rogers (2023) in section 6 and an empirical case study using Apple tick-by-tick data in section 7. Section 8 concludes and illustrates possible extensions.

2. Setup

Fix some horizon date T. We work on some underlying filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}$ satisfies the usual conditions and where all processes are $\mathbb{F}$-adapted. In this paper we consider a model for ultra high frequency or event data. As illustrated by figure 1, on very fine time scales an asset price process is piecewise constant and jumps at time points where new orders arrive to the market. We therefore assume that S follows a marked point process of the form

$$S_t = S_0 + \sum_{i=1}^{D_t} \nu_i, \tag{1}$$

where $D = (D_t)_{t \ge 0}$ is a doubly stochastic Poisson process with intensity driven by a stochastic process X, see Brémaud (1981), so that there is randomness in market activity. Furthermore, $\{\nu_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d variables with variance $\sigma^2 > 0$ and, for simplicity, with mean zero, and D and $\{\nu_i\}_{i \in \mathbb{N}}$ are independent. Summarizing, S is a doubly stochastic compound Poisson process with jump intensity $\lambda(X_t)$ and jump size distribution equal to the law of ν_1. Note that (1) is an arithmetic model, so that S_t can in principle become negative. Since we are interested in a relatively short time frame of the order of one or several days, this issue can be neglected.

We assume that the intensity of D is of the form

$$\lambda_t := \lambda(X_t) = b \cdot \exp(X_t), \tag{2}$$

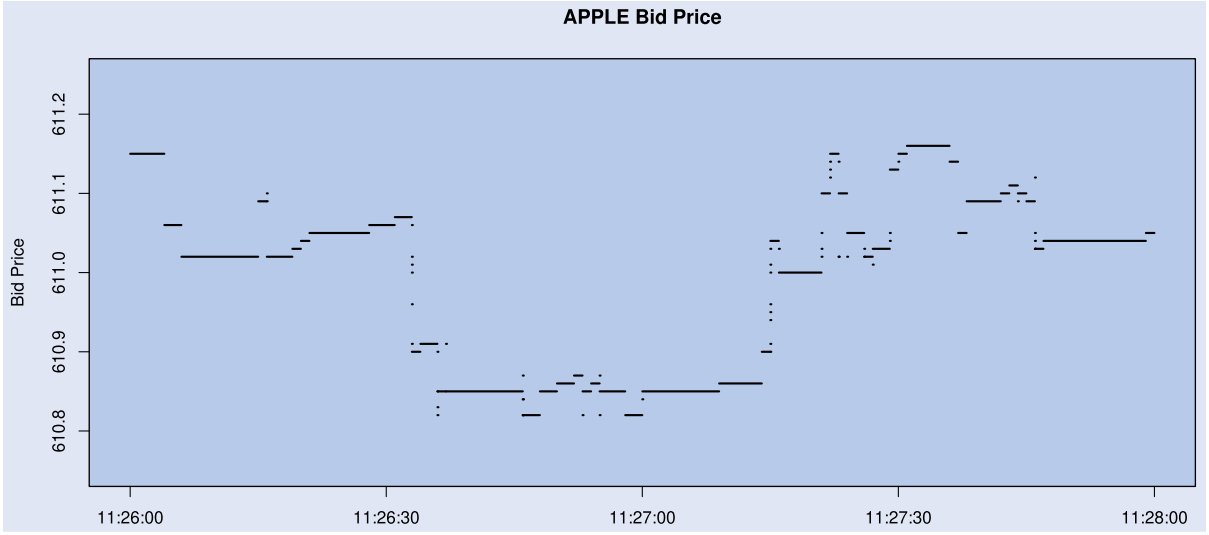

Figure 1. Two minutes of Apple bid price movements on July 9th, 2012.

where b is a positive constant. We consider the case where X is given by a fractional Brownian motion (fBM) or a Riemann-Liouville fractional Brownian motion (RLfBM) with Hurst index $H < \frac{1}{2}$ (see below for a precise definition).

Next we link our model to stochastic volatility models with stock price dynamics driven by Brownian motion. For typical event data the price jumps $\{v_i\}_{i\in\mathbb{N}}$ and the jump variance σ^2 are small, as prices move only by few ticks at a time. Moreover, there are frequent trades and hence many jumps, so that estimates for the constant b in equation (2) will be large. The following proposition shows that the model (1) is therefore close (in the sense of weak convergence) to an arithmetic stochastic volatility model with spot variance $\sigma_t^2 = \sigma^2\lambda(X_t)$, see in particular Hypothesis (iv) of the proposition.

PROPOSITION 2.1 *Consider a sequence of stock price models models* $S_t^\ell = S_0 + \sum_{i=1}^{D_t^\ell} v_i^\ell$, $\ell \in \mathbb{N}$, *where for each* ℓ

(i) D^ℓ *is a doubly-stochastic Poisson process with intensity* $\lambda^\ell(X_t)$ *for a fixed fBM* X;
(ii) $\{v_i\}_{i\in\mathbb{N}}$ *are i.i.d. with zero mean and variance* $(\sigma^\ell)^2$, *independent of* D^ℓ;
(iii) $(v_i^\ell)^2 \le \bar{c}_\ell$ *for a sequence* $\bar{c}_\ell \to 0$;
(iv) *There is some* $\sigma > 0$ *such that* $(\sigma^\ell)^2\lambda^\ell(x) \to \sigma^2\lambda(x)$ *uniformly on compacts.*

Then the sequence $(S^\ell, X)_{\ell\in\mathbb{N}}$ *converges in distribution to the process* (S, X), *where* S *has dynamics*

$$\mathrm{d}S_t = \sigma\sqrt{\lambda(X_t)}\,\mathrm{d}W_t \tag{3}$$

for a standard Brownian motion W *independent of* X.

Proof Since X is not a diffusion process we cannot apply directly a convergence criterion for diffusions such as Ethier and Kurtz (1986, Theorem 7.4.1). To address this issue we therefore first condition on $\mathcal{F}_T^X = \sigma(X_s, 0 \le s \le T)$, so that the path $(x_s)_{s=0}^T$ of X is known and D^ℓ is a Poisson process with time-dependent intensity $\tilde{\lambda}^\ell(t) := \lambda^\ell(x_t)$. Thus, consider the filtration $\tilde{\mathbb{F}} = \{\mathcal{F}_s^S \vee \mathcal{F}_T^X\}_{0\le s\le T}$, and note that, for each ℓ, S^ℓ is a $\tilde{\mathbb{F}}$ martingale. Moreover, it holds that

$$d(S_t^\ell)^2 = 2S_{t^-}^\ell\,\mathrm{d}S_t^\ell + d[S^\ell]_t,$$

where $[S^\ell]_t = \sum_{i=1}^{D_t^\ell}(v_i^\ell)^2$ is the pathwise quadratic variation of S^ℓ. It follows that $[S^\ell]_t - \int_0^t(\sigma^\ell)^2\lambda^\ell(X_s)\,\mathrm{d}s$ is a $\tilde{\mathbb{F}}$ martingale, so that $(S_t^\ell)^2 - \int_0^t(\sigma^\ell)^2\lambda^\ell(X_s)\,\mathrm{d}s$ is a $\tilde{\mathbb{F}}$ martingale as well. By *(iv)*, one has that

$$\sup_{t\le T}\left|\int_0^t(\sigma^\ell)^2\lambda^\ell(X_s)\,\mathrm{d}s - \int_0^t\sigma^2\lambda(X_s)\,\mathrm{d}s\right| \xrightarrow{P} 0,$$

where $\xrightarrow{P}$ denotes convergence in probability. Moreover,

$$\sup_{t\le T}\left(S_t^\ell - S_{t^-}^\ell\right)^2 = \sup_{i\le D_T^\ell}(v_i^\ell)^2 \le \bar{c}_\ell,$$

which converges to zero by (iii). By Theorem 7.4.1 in Ethier and Kurtz (1986) the law of S^ℓ given $\mathcal{F}_T^X$ therefore converges to the law of the time inhomogeneous diffusion process $\tilde{S}$ with $\mathrm{d}\tilde{S}_t = \sigma\sqrt{\tilde{\lambda}(t)}\,\mathrm{d}W_t$. Note that the theorem is formulated for homogeneous diffusions (time-independent drift and dispersion). In order to apply the result to inhomogeneous diffusions we consider the pair (t, S^ℓ) respectively $(t, \tilde{S})$, that is we add time as an additional state variable.† The law of $\tilde{S}$ is the conditional law of the stochastic volatility model (3) given $\mathcal{F}_T^X$. Hence we get that for every bounded and continuous function $F\colon \mathcal{C}([0,T],\mathbb{R}) \to \mathbb{R}$

$$\lim_{n\to\infty}\mathbb{E}\big(F(D^\ell) \mid \mathcal{F}_T^X\big) = \mathbb{E}\big(F(S) \mid \mathcal{F}_T^X\big),$$

and the result follows by integrating out the conditioning on $\mathcal{F}_T^X$. ■

In our setting the process X is not directly observable. Hence, this paper is concerned with statistical inference for

† Alternatively, we might apply a convergence criterium for semimartingales such as Theorem IX.3.21 in Jacod and Shiryaev (2003).

the current level of the process X and for the Hurst parameter H using the observed path of the asset price process S or, equivalently, the observed path of D as input.† Denote by $\mathbb{F}^D = (\mathcal{F}_t^D)_{0\le t\le T}$ the filtration generated by the point process D, so that $\mathcal{F}_t^D$ models the available observation at time t for making inference about X and H. We are interested in two related problems, filtering and parameter estimation. In the filtering task we compute an approximation to the conditional distribution of X_t given $\mathcal{F}_t^D$ for *known* parameters. In the parameter estimation task we adopt a Bayesian perspective and consider H as a random variable endowed with a prior distribution; particle filtering methods are then used to approximate the joint posterior distribution of H and X_t given $\mathcal{F}_t^D$.

In the case where X follows a Markovian diffusion, the filtering problem was studied previously by Frey and Runggaldier (2001) and Cvitanić *et al.* (2006) via standard stochastic filtering methods. Since fractional Brownian motion is neither Markovian nor a semimartingale, in our setup these methods cannot be applied in a straightforward way.‡ To address this issue, we use results from Carmona and Coutin (1998) and Carmona *et al.* (2000) to approximate X by a finite superposition of Markov processes. Details are given in the next section.

Remark 2.2 Note that the non-observability of X is intrinsic to stochastic volatility models and not due to the fact that we are modeling the stock price as a marked point process. In fact, if one works with a discretely sampled stochastic volatility model of the form $\mathrm{d}S_t = \sigma\sqrt{\lambda(X_t)}\,\mathrm{d}W_t$, the process X cannot be observed perfectly due to estimation errors and microstructure noise. These estimation errors vanish only if one samples S on a very fine time scale. However, as illustrated by figure 1, on these time scales a diffusion does not provide an accurate description of stock price trajectories.

3. Approximation by a Markovian model

In this section, we discuss the approximation of X by a finite superposition of Markov processes.

3.1. Representation of (RL)fBM

As a starting point, one considers the Mandelbrot-Van Ness representation of a fractional Brownian motion W^H, which is given by

$$W_t^H = c_H \int_0^t (t-s)^{H-\frac{1}{2}}\,\mathrm{d}B_s + c_H \int_{-\infty}^0 \left((t-s)^{H-\frac{1}{2}} - (-s)\right)^{H-\frac{1}{2}}\,\mathrm{d}B_s, \tag{4}$$

where B is a two-sided Brownian motion and c_H a constant depending on H. In particular, the choice

$$c_H = \left(\frac{\pi H(2H-1)}{\Gamma(2-2H)\Gamma(H+\frac{1}{2})^2\sin(\pi(H-\frac{1}{2}))}\right)^{\frac{1}{2}}, \tag{5}$$

$\Gamma(\cdot)$ the gamma function, ensures that the autocovariance function of W^H is given by

$$\mathbb{E}\{W_t^H W_s^H\} = \frac{1}{2}\left\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right\}. \tag{6}$$

Similarly, we introduce the so-called Riemann-Liouville fractional Brownian motion V^H as

$$V_t^H = c_H \int_0^t (t-s)^{H-\frac{1}{2}}\,\mathrm{d}B_s. \tag{7}$$

Now, if $H < \frac{1}{2}$, it has been shown in Carmona and Coutin (1998) and Carmona *et al.* (2000) that expressing $u \mapsto u^{H-\frac{1}{2}}$ as a Laplace transform,

$$u^{H-\frac{1}{2}} = \frac{1}{\Gamma(\frac{1}{2}-H)}\int_0^\infty e^{-xu}x^{-H-\frac{1}{2}}\,\mathrm{d}x,$$

and applying the stochastic Fubini theorem results in an infinite dimensional Markovian representation for W^H and V^H. In fact, we can express (4) as

$$W_t^H = \int_0^\infty \int_0^t e^{-x(t-s)}\,\mathrm{d}B_s\mu(\mathrm{d}x) + \int_0^\infty \left(e^{-xt}-1\right) \times \int_{-\infty}^0 e^{sx}\,\mathrm{d}B_s\,\mu(\mathrm{d}x),$$

where

$$\mu(\mathrm{d}x) = c_H\frac{x^{-H-\frac{1}{2}}}{\Gamma(\frac{1}{2}-H)}\,\mathrm{d}x. \tag{8}$$

Define for $x>0$ the process Z^x with $Z_t^x = \int_0^t e^{-x(t-s)}\,\mathrm{d}B_s$ and note that Z^x is an OU process with $Z_0^x = 0$, dispersion equal to one and mean-reversion speed equal to x. Consider the infinite dimensional Markov process $(\mathbf{Z}_t)_{0\le t\le T}$ with $\mathbf{Z}_t = (Z_t^x, x>0)$ and let $Q_0^x := \int_{-\infty}^0 e^{xs}\,\mathrm{d}B_s$. Then, for $H<\frac{1}{2}$, we can write W_t^H as function of $\mathbf{Z}_t$ and of the family of random variables $(Q_0^x, x>0)$ as follows

$$W_t^H = \int_0^\infty Z_t^x\,\mu(\mathrm{d}x) + \int_0^\infty \left(e^{-xt}-1\right)Q_0^x\,\mu(\mathrm{d}x). \tag{9}$$

Similarly, we obtain an infinite dimensional Markovian representation for RLfBM of the form

$$V_t^H = \int_0^\infty Z_t^x\mu(\mathrm{d}x). \tag{10}$$

† Observing S or D is equivalent in the context of our model since the price jumps $\{\nu_i\}_{i\in\mathbb{N}}$ are independent of X.

‡ Some theoretical results regarding nonlinear filtering for fractional Brownian motion (both in the state and in the observation process) have been obtained for example by Decreusefond and Üstünel (1998) and Coutin and Decreusefond (1999), but they are restricted to the case $H \ge \frac{1}{2}$ and thus not appropriate for the *rough* volatility setting considered here.

3.2. *Finite-dimensional approximation*

The representation (9) (respectively (10)) can be exploited to obtain an approximation of W^H (respectively V^H) by a finite sum of Ornstein-Uhlenbeck processes which are all driven by the same Brownian motion. The key idea is to approximate the measure μ in (8) by a finite sum of Dirac measures; that is, for some $J \in \mathbb{N}$, $\mu \approx \sum_{j=1}^{J} c_j \delta_{\kappa_j}$ for positive coefficients $(c_j)_{j=1,\dots,J}$ and positive mean-reversion speeds $(\kappa_j)_{j=1,\dots,J}$. More precisely, following Carmona and Coutin (1998) and Carmona *et al.* (2000), we perform the following approximation of W^H and V^H, which we refer to as spatial discretization. Given $H \in (0, \frac{1}{2})$, fix $J \in \mathbb{N}$ and consider a compact subset $[\xi_0, \xi_J]$ of $(0, \infty)$. Split this interval into subintervals by auxiliary terms $0 < \xi_0 < \xi_1 < \cdots < \xi_J < \infty$ and compute, for $j = 1, \dots, J$,

$$c_j = \int_{\xi_{j-1}}^{\xi_j} \mu(\mathrm{d}x) \quad \text{and} \quad \kappa_j = \frac{1}{c_j} \int_{\xi_{j-1}}^{\xi_j} x\mu(\mathrm{d}x).$$

With a slight abuse of notation we write Z^j and Q_0^j in place of Z^{κ_j} resp. $Q_0^{\kappa_j}$. Then, an approximation of the fBM defined by (4) is given by

$$W_t^H \approx W_t^{(H,J)} := \sum_{j=1}^{J} c_j \left(Z_t^j + \left(e^{-\kappa_j t} - 1 \right) Q_0^j \right) = \sum_{j=1}^{J} c_j Y_t^j, \tag{11}$$

where $Y_t^j = Z_t^j + (e^{-\kappa_j t} - 1) Q_0^j$. Note that $(Q_0^j)_{j=1}^J$ can be easily simulated, since $(Q_0^x)_{x>0}$ is a centered Gaussian process with covariance function $\mathrm{cov}(Q_0^x, Q_0^y) = \frac{1}{x+y}$. Similarly, we consider the following approximation of the RLfBM:

$$V_t^H \approx V_t^{(H,J)} := \sum_{j=1}^{J} c_j Z_t^j. \tag{12}$$

3.3. *Choices for spatial discretization*

It is shown in Carmona and Coutin (1998) and Carmona *et al.* (2000) that the approximations (11) and (12) converge in the sense that $\sup_{0 \le t \le T} |W_t^H - W_t^{(H,J)}|$ converges to zero in L^2 for every fixed T as the partition $\{\xi_0, \xi_1, \dots, \xi_J\}$ converges to the identity. Further results on convergence and accuracy of this approximation are given in Harms (2019) and in Coutin and Pontier (2007).

For the filtering applications in this paper we need an approximation involving only a moderate number J of processes, since J determines the dimension of the state space for the particle filter. A standard choice for spatial discretization is to use a geometric partition $(\xi_j)_{j=0}^J$ with $\xi_j = r \cdot \xi_{j-1}$ for $j = 1, \dots, J$ for a ratio $r \in (1, 2)$, as this choice ensures that each of the intervals $[\xi_{j-1}, \xi_j]$ has the same mass under the measure μ. The quality of the approximation obviously depends on the chosen J and r, where the approximation quality increases for large J and r close to one. On the other hand, for particle filtering we need to keep the dimension of the approximation manageable. Therefore, in this paper we fix the number J of OU processes and a compact set $[\xi_0, \xi_J]$, which is split into subintervals using a geometric partition with ratio $r = (\frac{\xi_J}{\xi_0})^{\frac{1}{J}}$. This section motivates our choices for the dimension of the approximation J, as well as for ξ_0 and ξ_J.

For fixed H, the choice of J depends on the fineness of the time scale on which one wants to approximate W^H. Fix a time horizon $T > 0$ and consider the partition $0 = t_0 < t_1 < \cdots < t_{TN} = T$ of $[0, T]$ with a constant step size $\Delta = \frac{1}{N}$, where N is the sampling frequency (the number of discrete-time steps in one continuous-time unit). Then, for smaller Δ, we need to increase J to maintain a satisfactory approximation accuracy. This is studied in detail in Coutin and Pontier (2007). Thus, given the sampling frequency N and a parameter $H < \frac{1}{2}$, we choose the number of OU processes as

$$\begin{aligned} J := J(N, H) &= \lfloor 2 \cdot N^{\zeta} \cdot \log(N) \rfloor \vee 16, \\ \zeta &= \log(1.5 - H). \end{aligned} \tag{13}$$

Note that in the parameter estimation task it is helpful to work with the same value J for all sampled values of H, as this allows for an efficient implementation of the nested particle filter in section 4.3. In that case, we take the value of J that corresponds to the midpoint $H = 0.25$, that is

$$\begin{aligned} J := J(N) &= \lfloor 2 \cdot N^{\zeta} \cdot \log(N) \rfloor \vee 16, \\ \zeta &= \log(1.5 - 0.25) \approx 0.223. \end{aligned} \tag{14}$$

Next, we turn to the choice of ξ_0 and ξ_J. For H close to 0.5, the mixing measure μ puts a large mass on small values of x so that the approximation quality is improved by taking a small value of ξ_0 in that case; for H close to zero, on the other hand, it is preferable to choose a comparably large value of ξ_J. With $\alpha = H + \frac{1}{2}$, we can, for example, choose the values $\xi_0 = J^{-2\alpha}$ and $\xi_J = J^{4-2\alpha}$, so that $(\frac{\xi_J}{\xi_0})^{\frac{1}{J}} = J^{\frac{4}{J}}$. Numerical experiments have shown that these choices perform well for modest $J > 16$. We report the results of some of these experiments in Appendix.

4. Nested particle filter

When discussing statistical inference for the model described in section 2, we distinguish the *filtering* problem and the *parameter estimation* problem. To deal with parameter estimation we adapt the nested particle filtering (NPF) algorithm of Crisan and Miguez (2018) to track the posterior distribution of the unknown parameters in our model, as well as the joint posterior distribution of parameter and state variables, in a recursive fashion. More specifically, Crisan and Miguez (2018) introduce a nested structure, employing two layers of particle filters: an 'outer' filter, which approximates the posterior of the unknown parameter vector $\boldsymbol{\Theta}$ given the observations, and a set of 'inner' filters, each corresponding to a sample generated in the outer layer and yielding an approximation of the posterior measure of the state conditional on both the observations and the given sample of $\boldsymbol{\Theta}$.

A key feature of the algorithm is that, at each recursive step, existing parameter particles are first *jittered* (that is, subjected to a small random perturbation) to restore diversity in the

sample, which might have been greatly reduced due to a previous resampling step. The inner bootstrap filtering step is then performed by propagating existing state particles using the jittered parameter particles, which is crucial to make the procedure recursive. Note that this assumes that the optimal filter for the model of interest must be continuous with respect to the parameter; that is, small changes in the parameter should lead only to small changes in the posterior of the state given the observations. For an in-depth theoretical discussion of the methodology, including assumptions and convergence results, we refer the reader to Crisan and Miguez (2018), particularly section 5.

In this section, we briefly describe the building blocks of this methodology, tailored to our setup, starting in section 4.1 with the approximate discrete-time model. We then describe the standard bootstrap filtering algorithm for the state in section 4.2. A set of these bootstrap filters, each conditional on a given parameter sample, are then batched to build the inner layer of the NPF algorithm, which we describe in detail in section 4.3.

4.1. Discrete-time modelling framework

First, in order to fit our setup within the framework of Crisan and Miguez (2018), we must consider a time-discretized version of our model. To this, we fix a time horizon $T > 0$ and consider the partition $0 = t_0 < t_1 < \cdots < t_{TN} = T$ of $[0, T]$ with a (for simplicity assumed constant) step size $\Delta = \frac{1}{N}$, where N is the number of discrete time steps in one continuous-time unit. We only consider changes in the unobservable state process at these discrete time points, which implies that the intensity has a piecewise-constant form. In particular, we assume that

$$\lambda_u = b \cdot \exp\left(X_{t_{n-1}}\right) \tag{15}$$

for all $u \in [t_{n-1}, t_n)$, $n = 1, \cdots, TN$. As long as Δ is small enough, this assumption is not too restrictive as the choice (15) still allows to mimic true market behavior reasonably well. Recall that here and in what follows X denotes a fBM resp. a RLfBM and, when filtering, it is approximated by (11) resp. by (12).

Then, the approximate discrete-time model is characterized by random sequences indexed by n, with $n = 1, \cdots, TN$. In particular, X_n denotes the value, assumed constant, of the state process within the continuous time interval $[t_{n-1}, t_n)$. Similarly, λ_n denotes the corresponding intensity and Z_n^j (resp. Y_n^j) the jth process used in approximating X_n. For what concerns observations, we consider the sequence given by the increments $I_n = D_{t_n} - D_{t_{n-1}}$, $n = 1, \ldots, TN$. Note that the piecewise-constant form of the intensity implies that I_n is Poisson distributed; in particular we have

$$I_n | \lambda_n \sim \text{Pois}(\lambda_n \Delta). \tag{16}$$

Finally, we adopt the usual convention that lower-case letters denote variable realizations (this holds true for sampled particles as well). In particular, we record the values of an observed realization of the counting process D on a grid of step size $\Delta = \frac{1}{N}$ to then obtain the observation sequence of its increments, denoted by $(i_n)_{1 \le n \le TN}$.

4.1.1. Updating scheme. Note that the particle filtering algorithm will require us to update the value of J processes building the approximations (11) and (12) at discretized time instances. Since each of the Ornstein-Uhlenbeck processes building the approximation of RLfBM given by (12) starts at and reverts to zero and has unit variance, the dynamics of each $Z^j, j = 1, \ldots, J$ are given by

$$\mathrm{d}Z_t^j = -\kappa_j Z_t^j \,\mathrm{d}t + \mathrm{d}B_t.$$

As explained in Gillespie (1996), an Euler-Maruyama discretization of these dynamics is accurate only for a suitably small time-discretization step Δ. In particular, Δ should be much smaller than the reciprocal of the mean reversion speed, which is clearly problematic in our context, given the high values of some of the κ_j's in the approximation (12). However, from the explicit solution of the OU SDE we can express z_n^j given the previous realization z_{n-1}^j as

$$z_n^j = e^{-\kappa_j \Delta} z_{n-1}^j + \sqrt{\frac{1 - e^{-2\kappa_j \Delta}}{2\kappa_j}} v, \quad n = 1, \ldots, TN, \tag{17}$$

where v is sampled from a standard normal distribution and is identical for all j, as all OU processes in (12) are driven by the same Brownian motion.

For what concerns the approximation of fBM given by (11), note that (17) above allows us to recursively express y_n^j, conditional on the previous realization y_{n-1}^j and on q_0^j, as

$$\begin{aligned} y_n^j = z_n^j + (e^{-\kappa_j n\Delta} - 1) q_0^j &= e^{-\kappa_j \Delta} z_{n-1}^j + \sqrt{\frac{1 - e^{-2\kappa_j \Delta}}{2\kappa_j}} v \\ &\quad + (e^{-\kappa_j n\Delta} - 1) q_0^j \\ &= e^{-\kappa_j \Delta} y_{n-1}^j + \sqrt{\frac{1 - e^{-2\kappa_j \Delta}}{2\kappa_j}} v + (e^{-\kappa_j \Delta} - 1) q_0^j, \end{aligned} \tag{18}$$

where v is sampled from a standard normal distribution and is the same for all j's. Note that the simultaneous propagation of all J random variables z_n^j respectively y_n^j by means of a single standard normal random variable makes for an efficient implementation of the filtering algorithms in sections 4.2 and 4.3.

4.2. 'Inner' (Bootstrap) filter

To fix notation and describe how the approximation of (RL)fBM can be exploited in the given context, in this section we review the standard particle filter (so-called *bootstrap* filter); this represents the 'inner' filter of the two nested layers of particle filters described in Crisan and Miguez (2018) and it yields, at each time point n, an approximation π_n of the posterior measure for the state, conditional on the observations up to the time point n and the given (known or sampled) parameters.

In Algorithm 1 below we give a schematic representation of the bootstrap filtering algorithm, tailored to our setup,

Algorithm 1: Bootstrap Filter Algorithm for the fBM Case

Input: number of continuous time units T, number of discrete-time steps per unit N, recorded observations $(i_n)_{1\le n\le TN}$, Hurst index $H \in (0, \frac{1}{2})$, constant $b > 0$, number of state particles M

Initialization ($n = 0$)

- Compute approximation dimension $J = J(H, N)$ `// See (13)`
- Compute coefficients $\boldsymbol{c}$ and mean-reversion speeds $\boldsymbol{\kappa}$ `// See sections 3.2 and 3.3`
- **for** $m = 1$ **to** M
  - Sample $\boldsymbol{y}_0^{(m)}$ from a suitable prior `// E.g., ` $\boldsymbol{y}_0^{(m)} \sim \mathcal{N}_J(\boldsymbol{0}, \sigma^y \boldsymbol{I})$`, where` σ^y `is small`
  - Generate $\boldsymbol{q}_0^{(m)}$ `//` $\boldsymbol{q}_0^{(m)} \sim \mathcal{N}_J(\boldsymbol{0}, \boldsymbol{\Gamma})$`, where` $\boldsymbol{\Gamma} = (\gamma_{i,j})_{1\le i,j\le J}$ `and` $\gamma_{i,j} = (\kappa_i + \kappa_j)^{-1}$
- **end**

Recursive step ($n \ge 1$)

- **Propagate**
  - **for** $m = 1$ **to** M
    - Generate $v^{(m)} \sim \mathcal{N}(0, 1)$
    - Draw $\bar{\boldsymbol{y}}_n^{(m)}$ conditional on $\boldsymbol{y}_{n-1}^{(m)}$ and $\boldsymbol{q}_0^{(m)}$ `// See updating scheme (18)`
    - Set $\bar{\boldsymbol{q}}_0^{(m)} = \boldsymbol{q}_0^{(m)}$
  - **end**
- **Compute normalized likelihood weights**
  - **for** $m = 1$ **to** M
    - Compute $\bar{x}_n^{(m)} = \boldsymbol{c}^\top \bar{\boldsymbol{y}}_n^{(m)}$ `// See state approximation (11)`
    - Let $\lambda_n^{(m)} = b\Delta \exp(\bar{x}_n^{(m)})$ and $a_n^{(m)} = (\lambda_n^{(m)})^{i_n} \exp(-\lambda_n^{(m)})$ `// See (16)`
    - Compute $\bar{w}_n^{(m)} = \frac{a_n^{(m)}}{\sum_{m=1}^M a_n^{(m)}}$
  - **end**
- **Resample with replacement**
  - **for** $m = 1$ **to** M
    - Let $(\boldsymbol{y}_n^{(m)}, \boldsymbol{q}_0^{(m)}) = (\bar{\boldsymbol{y}}_n^{(q)}, \bar{\boldsymbol{q}}_0^{(q)})$ with probability $\bar{w}_n^{(q)}$, $q \in \{1, \dots, M\}$
  - **end**

Output: approximate state posterior distribution given observations and parameters

for the case where the state process is a fractional Brownian motion.† Here, we consider both the constant b and the Hurst index $H < \frac{1}{2}$ to be known; then, using (13) we compute the number $J = J(H, N)$ of mean-reverting processes needed for the approximation. We use the following notation. We write $\boldsymbol{c} = (c_1, \dots, c_J)^\top$ for the vector of coefficients and $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_J)^\top$ for the vector of mean reversion speeds used for the approximation (11). Moreover, we let $\boldsymbol{Y}_n := (Y_n^1, \dots, Y_n^J)^\top$ and $\boldsymbol{Q}_0 := (Q_0^1, \dots, Q_0^J)^\top$ denote state vectors and we thus use the notation $\boldsymbol{y}_n^{(m)}$ and $\boldsymbol{q}_0^{(m)}$ for vectors corresponding to the mth state particle.

4.3. Nested particle filter and parameter estimation

Denote by $\boldsymbol{\Theta}$ be the vector of model parameters to be estimated. In this paper, we aim at estimating $\boldsymbol{\Theta} = H$, the Hurst index of the underlying fractional Brownian motion. The estimation of b will not be addressed in the specific context of this model, as it can typically be chosen to match the average number of price movements observed in one (continuous-time) unit.

The nested particle filter for the approximation, at each discrete time point n, of the posterior distribution of the parameter is described in Algorithm 2 below. Besides the time step size $\Delta = \frac{1}{N}$ and the recorded observations $(i_n)_{1\le n\le TN}$, the required inputs are the value of constant b and the number of particles K and M. Here, we consider the Hurst index $H < \frac{1}{2}$ to be unknown and, using (14), we compute the number $J = J(N)$ of OU processes needed for the approximation *independently* of the value of the sampled parameter particle. Note that in the following we keep the vector notation explained in section 4.2 and include a superscript denoting the parameter particle index; for instance, $\boldsymbol{y}_n^{(k,m)}$ corresponds to the m^{th} state particle in the inner filter given the k^{th} parameter particle in the outer filter.

5. Numerical results

In this section, we present the results of a simulation study aimed at testing the accuracy of the presented algorithms in the context of our model. Throughout this section, we fix a time horizon $T = 1$ and a small time step of size $\Delta = \frac{1}{480}$; that is, we consider one-minute intervals in a trading day of 8 hours. Except when analyzing the impact of the 'informativeness' of the observation process, we set $b = 8000$ as the average amount of price changes in a given day—a conservative choice for a liquid stock.

† The algorithm for the RLfBM case can be retrieved by considering each sampled Q_0 to be equal to zero.

Algorithm 2: Nested Particle Filter Algorithm for the fBM Case

Input: number of continuous time units T, number of discrete-time steps per unit N, recorded observations $(i_n)_{1\le n\le TN}$, constant $b > 0$, jittering hyperparameters, number of parameter particles K, number of state particles M

Initialization ($n = 0$)

Compute approximation dimension $J = J(N)$ // See (14)

for $k = 1$ **to** K

Sample θ_0^k from a suitable prior distribution // E.g., $\theta_0^k \sim \mathcal{U}_{(0,\frac{1}{2})}$

for $m = 1$ **to** M

Sample $\boldsymbol{y}_0^{(k,m)}$ from a suitable prior // E.g., $\boldsymbol{y}_0^{(m)} \sim \mathcal{N}_J(\boldsymbol{0}, \sigma^y \boldsymbol{I})$, where σ^y is small

Generate $\boldsymbol{q}_0^{(k,m)}$ // $\boldsymbol{q}_0^{(k,m)} \sim \mathcal{N}_J(\boldsymbol{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} = (\gamma_{i,j})_{1\le i,j\le J}$ and $\gamma_{i,j} = (\kappa_i + \kappa_j)^{-1}$

end

end

Recursive step ($n \ge 1$)

for $k = 1$ **to** K

Draw $\bar{\theta}_n^k$ from $\kappa_K^{\theta_{n-1}^k}(d\theta)$ // See Crisan and Miguez (2018) for choices of jittering kernel κ_K

Given $\bar{\theta}_n^k$, compute coefficients $\boldsymbol{c}^k$ and speeds $\boldsymbol{\kappa}^k$ // See sections 3.2 and 3.3

for $m = 1$ **to** M

Inner filter

Propagate

Generate $v^{(k,m)} \sim \mathcal{N}(0,1)$

Draw $\bar{\boldsymbol{y}}_n^{(k,m)}$ conditional on $\boldsymbol{y}_{n-1}^{(k,m)}$, $\boldsymbol{q}_0^{(k,m)}$ and $\boldsymbol{\kappa}^k$ // See updating scheme (18)

Set $\bar{\boldsymbol{q}}_0^{(k,m)} = \boldsymbol{q}_0^{(k,m)}$

Compute normalized likelihood weights

Compute $\bar{x}_n^{(k,m)} = (\boldsymbol{c}^k)^\top \bar{\boldsymbol{y}}_n^{(k,m)}$ // See state approximation (11)

Let $\lambda_n^{(k,m)} = b\Delta \exp(\bar{x}_n^{(k,m)})$ and $a_n^{(k,m)} = (\lambda_n^{(k,m)})^{i_n} \exp(-\lambda_n^{(k,m)})$ // See (16)

Compute $\bar{w}_n^{(k,m)} = \frac{a_n^{(k,m)}}{\sum_{m=1}^M a_n^{(k,m)}}$

Resample with replacement

Let $(\tilde{\boldsymbol{y}}_n^{(k,m)}, \tilde{\boldsymbol{q}}_0^{(k,m)}) = (\bar{\boldsymbol{y}}_n^{(k,q)}, \bar{\boldsymbol{q}}_0^{(k,q)})$ with probability $\bar{w}_n^{(k,q)}$, $q \in \{1,\dots,M\}$

end

Compute normalized likelihood weights

Compute $\bar{w}_n^k = \frac{\sum_{m=1}^M a_n^{(k,m)}}{\sum_{k=1}^K \sum_{m=1}^M a_n^{(k,m)}}$

Resample with replacement

Let $(\theta_n^k, \{(\boldsymbol{y}_n^{(k,m)}, \boldsymbol{q}_0^{(k,m)})\}_{1\le m\le M}) = (\bar{\theta}_n^l, \{(\tilde{\boldsymbol{y}}_n^{(k,m)}, \tilde{\boldsymbol{q}}_0^{(k,m)})\}_{1\le m\le M})$ with probability $\bar{w}_n^l$, $l \in \{1,\dots,K\}$.

end

Output: approximate parameter posterior distribution $\hat{\mu}_n = \frac{1}{K}\sum_{k=1}^K \delta_{\theta_n^k}$ for each $n = 1,\dots,TN$, where δ_θ denotes the Dirac measure at θ

5.1. *Bootstrap filter for known H*

Here we consider the Hurst index H to be known and thus we implement the bootstrap filter of section 4.2 with $M = 300$ particles. We consider two cases for the true Hurst index, $H = 0.1$ and $H = 0.4$; given the total number of time steps $N = 480$, the number of OU processes building the approximation (12) is given by equation (13) as $J(480, 0.1) = 98$ and $J(480, 0.4) = 22$, respectively. The prior distribution for the state (that is, for the vector of mean-reverting processes approximating X) is a centered J-dimensional normal distribution with covariance matrix given by $0.1 \cdot \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the $J \times J$ identity matrix.

Figure 2 shows simulated trajectories of RLfBM(right) resp. fBM (left) with Hurst index $H = 0.1$ (upper panel) resp. $H = 0.4$ (lower panel) as black solid lines, as well as the corresponding filtered estimates as dotted-dashed lines (blue for the $H = 0.1$ case and coral for the $H = 0.4$ case).† We observe that the filtered trajectory tracks the true one reasonably well in both cases.

5.2. *Nested particle filter for unknown H*

Here we consider the Hurst index H to be unknown and thus we estimate it accordingly by implementing the nested particle filter of section 4.3 with $K \cdot M = 300^2$ particles. We consider two cases for the true Hurst index, $H = 0.1$ and $H = 0.4$; however, unlike in the filtering experiment, we keep the number of OU processes building the approximation (12)

† If figure 2 is printed in black and white, it might be hard to distinguish the true trajectories from the filtered ones. In that case, we recommend to look at the figure in color on screen.

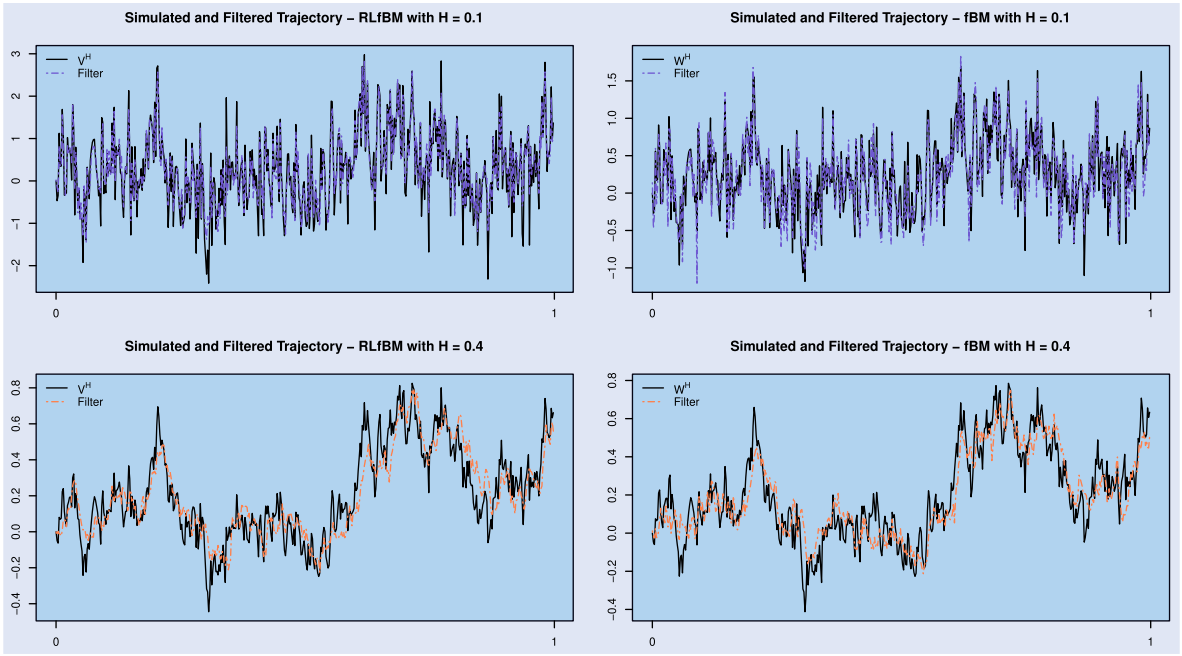

Figure 2. Upper panel: simulated (black solid line) and filtered (blue dotted-dashed line) trajectory of RLfBM (right) resp. fBM (left) with $H = 0.1$. Lower panel: simulated (black solid line) and filtered (coral dotted-dashed line) trajectory of RLfBM (right) resp. fBM (left) with $H = 0.4$.

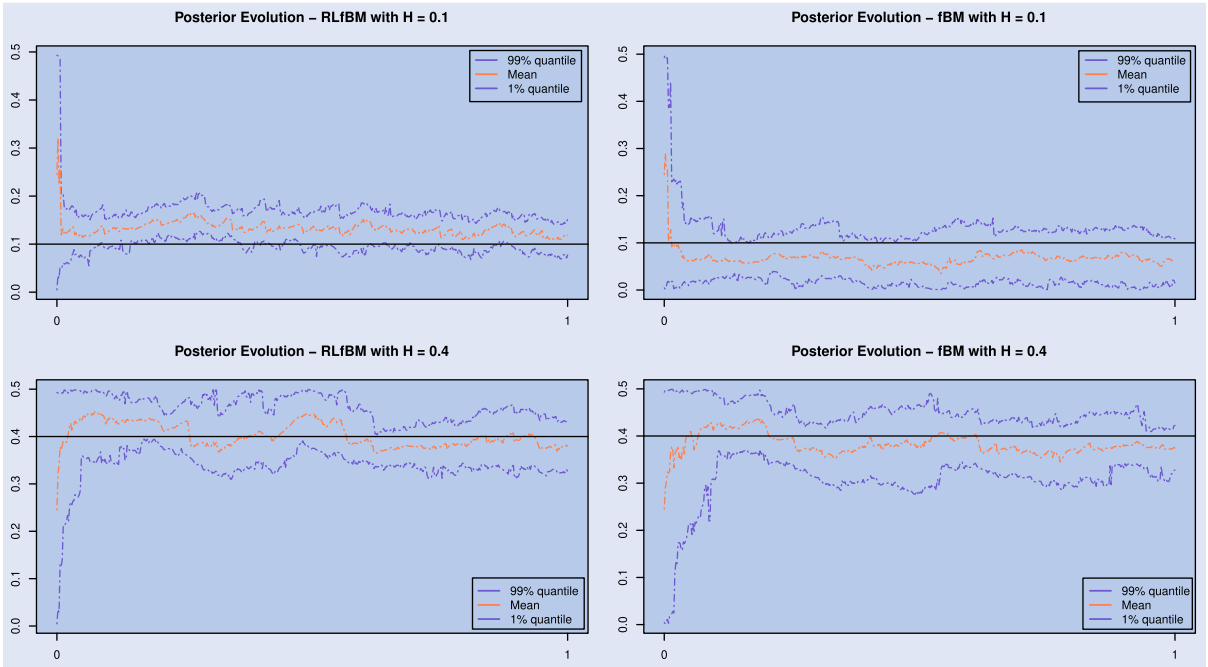

Figure 3. Mean (coral dotted line), 1%- resp. 99%- quantiles (blue dotted lines) and true value (solid black line) of the estimated posterior distribution for H. Upper panel: $H = 0.1$, lower panel: $H = 0.4$. Right: RLfBM case, left: fBM case.

equal in both cases. This number depends only on the number of time steps per continuous-time unit $N = 480$, and it is given by equation (14) as $J(480) = 48$. The prior distribution for the state, that is for the vector $\boldsymbol{Y}_0$ or $\boldsymbol{Z}_0$, is a centered J-dimensional normal distribution with covariance matrix given by $0.1 \cdot \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the $J \times J$ identity matrix. The prior for the parameter $\boldsymbol{\Theta} = H$ is uniform between 0 and $\frac{1}{2}$. For what concerns the jittering kernel, we jitter each parameter particle using a Gaussian kernel κ_K^θ with mean θ (the value of the parameter particle to be jittered) and variance K^{-2}.

Figure 3 shows the mean (coral dotted line), the 1%- and 99%- quantiles (blue dotted lines) of the estimated posterior distribution μ_n, $n = 1, \ldots, TN$, of the unknown parameter H (the true value of H is depicted with a solid black line). The upper panel corresponds to $H = 0.1$, the lower one to $H = 0.4$. Moreover, the plots on the left correspond to the RLfBM case, while those on the right to the fBM case. Overall, the behavior of parameter estimates over time is quite satisfactory as the estimation problem at hand is relatively complex, particularly when considering a conservative average number of trades and a short time horizon.

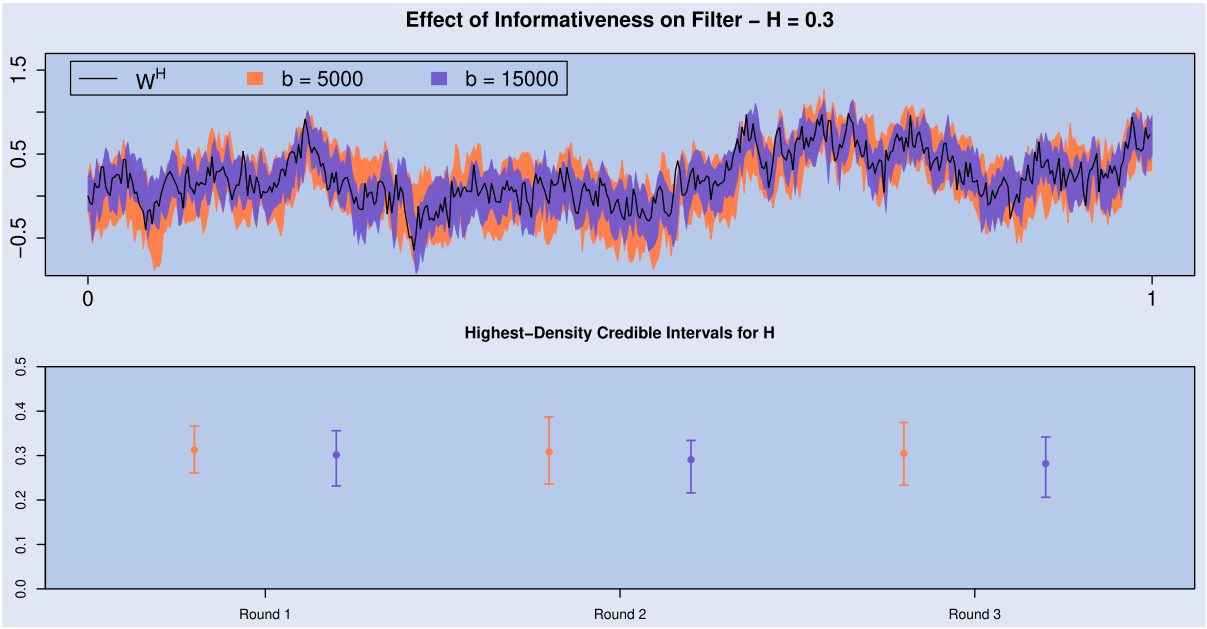

Figure 4. True and filtered state trajectory of W^H with $H = 0.3$ (upper panel) and 99% highest-density credible intervals for H at final time (lower panel) in the cases $b = 5000$ (coral) resp. $b = 15\,000$ (blue).

5.3. *Sensitivity with respect to b*

Next we study the influence of the parameter b on the accuracy of the filtering algorithms. We assume that the underlying signal process is an fBM with true Hurst index $H = 0.3$ and we analyze two cases, one in which the observation process is less informative (i.e. $b = 5000$) and one in which it is more informative (i.e. $b = 15\,000$). First, we consider the Hurst index to be known and thus we implement the bootstrap filter of section 4.2 with $M = 300$ particles and—as specified by equation (13)—we use $J(0.3, 480) = 38$ OU processes to build the approximation (12). Then, we consider H to be unknown and we estimate it by implementing the nested particle filter of section 4.3 with $K \cdot M = 300^2$ particles, where the dimension of the approximation is given by equation (14) as $J(480) = 48$. The choices for the state resp. parameter priors, as well as the specifications for the jittering kernel, are the same as in section 5.2. In particular, the prior for H is chosen to be uniform on $(0, \frac{1}{2})$ and thus not informative.

The upper panel of figure 4 shows the unobserved trajectory of the fBM W^H with Hurst index $H = 0.3$ (solid black line) and the area comprised between the 1%- and 99%-quantile of the posterior distribution of the state process resulting from one run of bootstrap filtering (coral corresponds to the less informative case and blue to the more informative one). The lower panel of figure 4 shows 99% highest-density (that is, narrowest) *credible* intervals for the Hurst index at final time $T = 1$, with the posterior mean depicted as a filled dot.† In particular, we plot the credible intervals corresponding to the less informative case in coral and those corresponding to a more informative case in blue. Moreover, we plot one such credible interval for each of 3 independent runs of the NPF algorithm, using the same seed for each round in both informativeness specifications for a fairer comparison. We observe that a higher value of b improves the filter accuracy; however, results are still satisfactory in the less informative case.

6. Experiments in the context of non-rough stochastic volatility models

In this section, in the spirit of Cont and Das (2022) and Rogers (2023) we apply our methodology to synthetic data generated in a setting where spot volatility is driven by standard Brownian diffusions and assess whether the resulting estimates for H are roughly consistent with such a setup, rather than mistakenly reflecting a spurious strong roughness effect. This amounts to modifying the continuous-time model of section 2, discretizing it in the same fashion as in section 4.1 and estimating the unobservable volatility trajectory and the Hurst index H via the nested particle filter algorithm of section 4.3.

6.1. *Volatility as modulus of Brownian motion*

Motivated by Cont and Das (2022), here we assume our observation process D to have an intensity of the form (to be compared with (2)):

$$\lambda_t := \lambda(W_t) = b \cdot |W_t|^2, \tag{19}$$

where b is a positive constant and W is a standard Brownian motion, so that $H = \frac{1}{2}$. However, we consider the Hurst index H to be unknown and thus we estimate it accordingly by implementing the nested particle filter of section 4.3 with

† Credible intervals are the Bayesian analog of confidence intervals. In particular, the highest-density 99% credible interval for H is the narrowest interval $[a, b]$ such that the posterior probability $\mathbb{P}(H \in [a, b] \mid \mathcal{F}_T^D) \geq 0.99$. We use the `R` package `bayestestR` to compute these credible intervals, see Makowski *et al.* (2019).

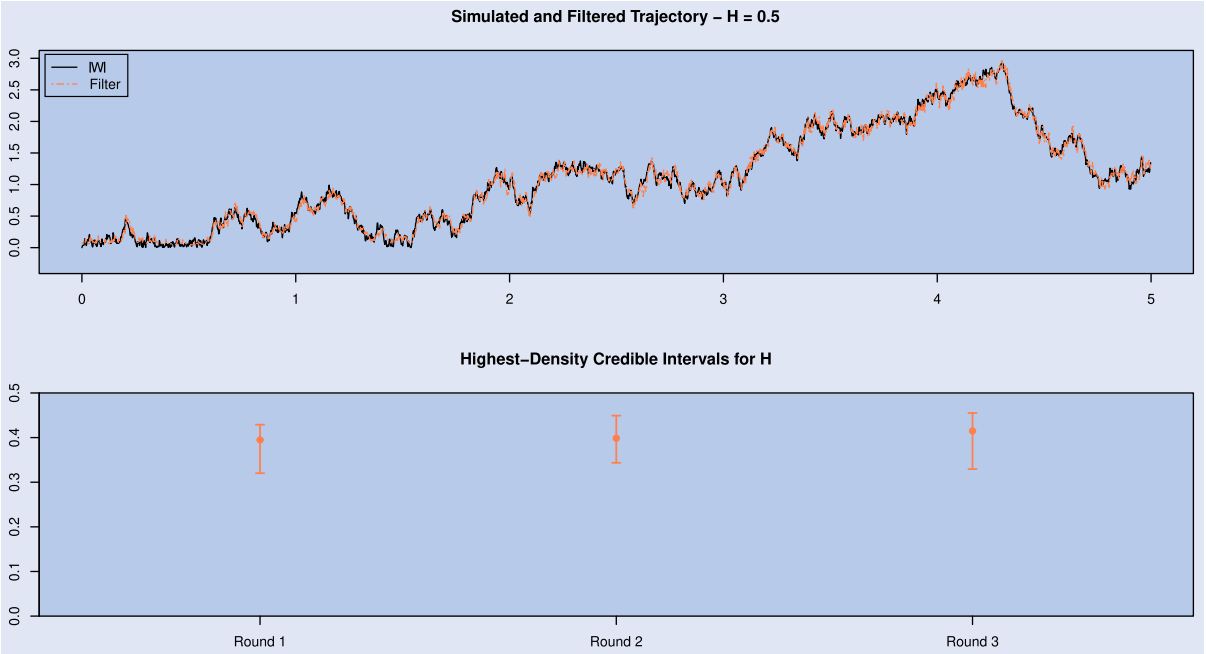

Figure 5. True and filtered state trajectory of $|W|$ and 99% highest-density credible intervals for H at final time (lower panel).

$K \cdot M = 300^2$ particles.† We fix a time step of size $\Delta = \frac{1}{480}$ (that is, we consider one-minute intervals in a trading day of 8 hours) and a time horizon of $T = 5$ days, so that the dimension of the approximation (11) is given by equation (14) as $J(480) = 48$. The choices for the state resp. parameter priors, as well as the specifications for the jittering kernel, are the same as in section 5.2. In particular, the prior for H is chosen to be uniform on $(0, \frac{1}{2})$ and thus not informative.

The upper panel of figure 5 shows a simulated trajectory of $|W|$ (black solid line) and the corresponding filtered estimate (coral dotted-dashed line). The lower panel shows 99% highest-density (that is, narrowest) *credible* intervals for the Hurst index at final time T, one for each of 3 independent runs of the NPF algorithm, with the posterior mean depicted as a filled dot. We stress that this experiment aims at assessing whether the resulting estimate for H reflects a spurious strong roughness effect, attributable for instance to microstructure noise, rather than at estimating H correctly. In fact, the value $H = \frac{1}{2}$ can *never* be estimated correctly without modifying the methodology and the algorithm substantially, as the approximations (11) and (12) are valid for $H < \frac{1}{2}$ and the parameter space of nested particle filtering algorithm of section 4.3 is assumed to be a compact subset of $(0, \frac{1}{2})$. Overall, given these considerations, the results are quite satisfactory, both for the filtered trajectory and for the parameter estimate, since the latter is relatively close to $\frac{1}{2}$ (and not erroneously estimated to be close to zero).

6.2. *OU-OU model for volatility*

As an alternative to rough volatility models, Rogers (2023) proposes the so-called OU-OU model, where spot volatility is modeled as an OU process with high mean-reversion speed and high volatility, reverting to a level given by a second, 'slower' OU process. Rogers (2023) claims that the empirical properties of this model are akin to those of rough volatility models.

To mimic this framework, we assume that the process D has an intensity of the form:

$$\lambda_t := \lambda(X_t) = b \cdot \exp(X_t), \tag{20}$$

where X is given by the so-called OU-OU model:

$$\begin{aligned} \mathrm{d}R_t &= -\beta R_t \,\mathrm{d}t + \sigma_R \,\mathrm{d}W'_t, \\ \mathrm{d}X_t &= \kappa(R_t - X_t)\,\mathrm{d}t + \sigma_X \,\mathrm{d}W_t, \end{aligned} \tag{21}$$

for independent standard Brownian motions W' and W. To simulate model and observations, we choose the same parameters as Rogers (2023), that is, $\sigma_X^2 = 20$, $\sigma_R^2 = 0.625$, $\kappa = 210$ and $\beta = 2.5$. However, when implementing our estimation methodology, we suppose that the unobservable process X is not of its 'true' form (21), but rather a fractional Brownian motion with unknown Hurst index H, which we estimate using the nested particle filter of section 4.3 using $K \cdot M = 300^2$ particles. We fix a time horizon of $T = 1$ day and make two distinct choices for the time step Δ to analyze its impact on the estimation procedure. In particular, we consider $\Delta = \frac{1}{480}$ (corresponding to one-minute intervals in a trading day of 8 hours) and $\Delta = \frac{1}{960}$ (corresponding to thirty-second intervals in a trading day of 8 hours); thus, equation (14) gives $J(480) = 48$ respectively $J(960) = 63$.

The upper panel of figure 6 shows 99% highest-density (that is, narrowest) *credible* intervals for the Hurst index at final time $T = 1$, one for each of 3 independent runs of the NPF algorithm applied to observations generated by the OU-OU model, with the mean depicted as a filled dot. The three (coral) intervals on right-hand side correspond to the coarser time scale case (that is, $\Delta = \frac{1}{480}$), while the three (blue)

† We adapt the filtering algorithms accordingly; in particular, the intensity is to be computed using (19).

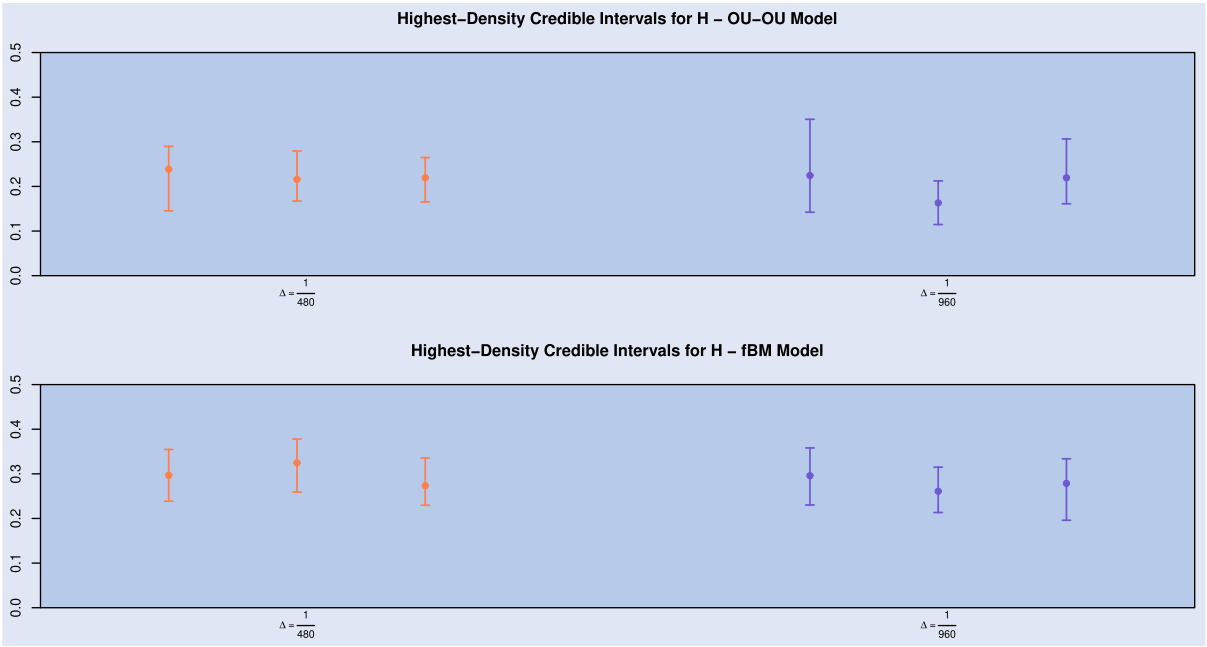

Figure 6. 99% highest-density credible intervals for H at final time for the OU-OU model (upper panel) and 99% highest-density credible intervals for H at final time for the fBM model with $H = 0.3$ (lower panel). In both panels, the (coral) intervals on the left correspond to the time step $\Delta = \frac{1}{480}$, the (blue) intervals on the right to $\Delta = \frac{1}{960}$.

intervals on the left-hand side to the more finely discretized case (that is, $\Delta = \frac{1}{960}$). We can observe that estimates for H obtained in the OU-OU case are consistent with roughness in volatility paths, meaning that the OU-OU model seems to mimic that type of behavior. In particular, the model estimated on a coarser time scale seems to correspond to an fBM with $H \approx 0.2$. However, if we simulate the model on a finer time grid and apply the corresponding algorithm, the estimate for the Hurst index becomes more volatile, which is reflected in a wider credible interval. This does not seem to be the case for data generated by a 'true' rough model: for comparison, we apply the NPF algorithm to observations where the signal is a fBM with Hurst index $H = 0.3$, again for the two different time scales. The corresponding 99% highest-density credible intervals are shown in the lower panel of figure 6. We can observe that, when synthetic data are generated from the correct model, estimates are far less sensitive with respect to the fineness of the time scale considered.

7. Application to real data

In this section, we discuss an empirical case study where we apply our methodology to real data from LOBSTER† More specifically, we consider event data for the mid-price of Apple from July 9th, 2012, to July 13th, 2012, and we sum all movements in mid-price occurring in a given time interval to obtain our observation sequence $(i_n)_{1 \le n \le TN}$. Note that, since our continuous-time unit is days, we have $T = 5$ and, since Apple is traded on the NASDAQ stock exchange which is open six and a half hours per day, we have $N = 390$ for the one-minute intervals case, $N = 780$ for the thirty-second intervals case and $N = 2340$ for the ten-second intervals case. We estimate the value of b to be the average number of per-day price movements over this week of trading; that is, we set $b = 78\,291$.

As H is unknown, we apply the NPF algorithm of section 4.3 to each observation series. As in section 5.2 we use a total of $K \cdot M = 300^2$ particles and choose the prior for the state to be a centered J-dimensional normal distribution with covariance matrix given by $0.1 \cdot \boldsymbol{I}$, while the prior for H is uniform on $(0, \frac{1}{2})$. The variance of the jittering kernel is again set to K^{-2}. In our data application we only consider the fBM case and we approximate the signal process using $J(390) = 45$ (see equation (14)) mean-reverting processes for the one-minute time scale, $J(780) = 58$ for the thirty-second time scale and $J(2340) = 87$ for the ten-second time scale.

The results are depicted in figures 7 and 8. Figure 7 shows the filtered estimate for the signal trajectory (blue line) when the observation series consists of per-minute price movements. The filtering results display U-shape behavior, in agreement with stylized facts about intra-day trading activity. Figure 8 shows 99% highest-density (that is, narrowest) *credible* intervals for the Hurst index at final time $T = 5$, one for each of 3 independent runs of the NPF algorithm, with the mean depicted as a filled dot. In particular, the (coral) intervals on the left correspond to the ten-second time scale, the (blue) intervals in the middle to the thirty-second time scale and the (purple) intervals on the left to the one-minute time scale. Parameter estimation results are qualitatively consistent with the results of Fukasawa *et al.* (2019) or Bennedsen *et al.* (2021), as the mean estimates for H are smaller than 0.1 in all three time-discretization scenarios. However, the credible intervals are slightly wider on a coarser time grid: this is not really surprising, as we use a piecewise constant discrete approximation to a continuous-time model which is best

† For details on the LOBSTER database see https://lobsterdata.com/info/WhatIsLOBSTER.php.

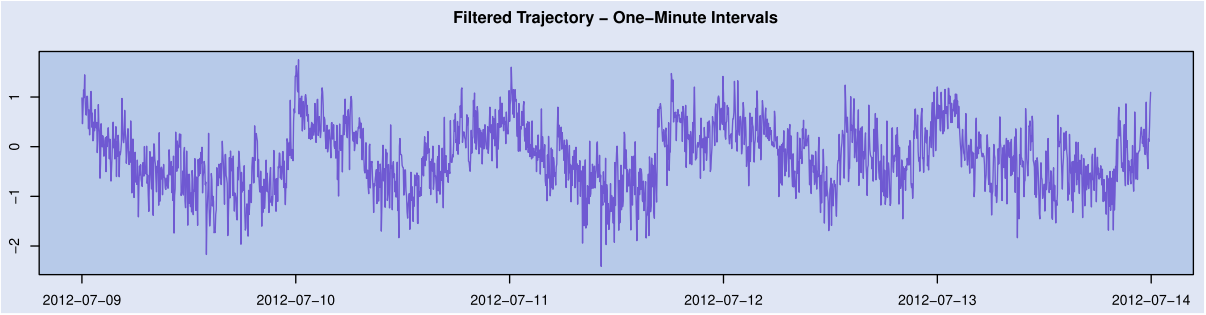

Figure 7. Filtered state trajectory for the real data application when obersations are recorded on a one-minute grid. Data source: LOBSTER.

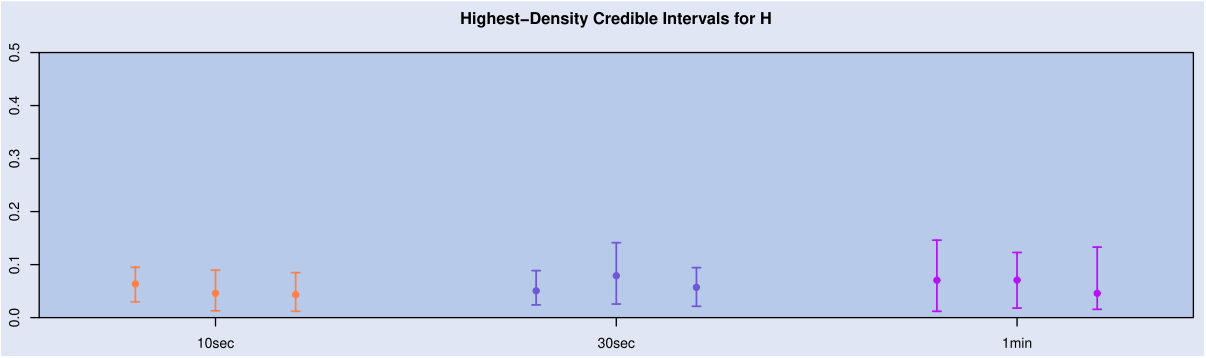

Figure 8. 99% highest-density credible intervals for H at final time for the real data application. Observations are recorded on a ten-second (left), thirty-second (middle) and one-minute (right) grid. Data source: LOBSTER.

suited for fine time scales. In any case, more work is needed to truly assess this *rough* behavior; in particular, the analysis should be repeated for different assets and time points; this is left for future research. Furthermore, we caution against an interpretation of the form 'volatility has to be modeled by a rough process'. In fact, the choice of a model in quantitative finance always involves a trade-off between realism and tractability, and for many tasks the improved tractability of standard Markovian volatility models may well outweigh the potentially greater realism of a rough volatility model.

8. Conclusions and outlook

In this paper, we discussed filtering and parameter estimation in a rough volatility model for high-frequency data. We considered a time-discretized version of a continuous-time framework where observations are given by a trajectory of a Cox process whose intensity is driven by a 'rough' process, such as (Riemann-Liouville) fractional Brownian motion with Hurst parameter $H < \frac{1}{2}$. Basing ourselves on a representation of such signal as a superposition of Ornstein Uhlenbeck processes (as introduced by Carmona and Coutin (1998)), we detailed how it can be filtered using standard particle filtering techniques. Moreover, we explained how to estimate H using the *nested particle filtering* algorithm of Crisan and Miguez (2018). We ran a comprehensive simulation study to test the accuracy of the algorithms and their sensitivity with respect to specific modeling choices. We found that the results are satisfactory; in particular, we were able to estimate the Hurst index with reasonable accuracy, while still taking the unobservability of volatility fully into account. Moreover, the parameter estimation methodology seems to be able to adequately distinguish between 'true' rough dynamics in spot volatility and spurious roughness effects arising from microstructure noise, as documented e.g. in Cont and Das (2022). Finally, we presented an empirical case study using real high-frequency data, from which we could conclude that there is some evidence of roughness in volatility in real markets. However, further analysis is needed to fully corroborate this finding.

In this paper, we focused exclusively on the short-term behavior of volatility and our modeling framework does not allow to capture long memory, a volatility feature which is has been investigated in several empirical studies (see, for instance, Andersen *et al.* 2001) and it is often considered a stylized fact. As explained in Gneiting and Schlather (2004), models based on processes with the self-similarity property cannot account for both roughness and long-run dependence. Moreover, fractional Brownian motion is not stationary, which represents a problem when one seeks to study the long-term properties of volatility. Therefore, a possible extension of the present work is to find an alternative model that allows for rough paths and for a polynomial decay of the autocorrelation function $\rho(s)$ for $s \to \infty$ (that is, for long memory). This is very much in the spirit of Bennedsen *et al.* (2021), who advocate the use of the so-called Brownian semistationary processes (see Barndorff-Nielsen and Schmiegel 2009).

Acknowledgments

We are thankful to Katia Colaneri, Wolfgang Runggaldier and two anonymous referees for useful comments and remarks.

Data availability statement

The data used in this article are not publicly available, we obtained them from the data provider LOBSTER (see https://lobsterdata.com/info/WhatIsLOBSTER.php).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The first author was in part supported by the Austrian Science Fund (FWF, grant ZK 35, grant Y 1235 and grant TAI 517-G).

References

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H., The distribution of realized stock return volatility. *J. Financ. Econ.*, 2001, **61**(1), 43–76.

Barndorff-Nielsen, O. and Schmiegel, J., Brownian semistationary processes and volatility/intermittency. *Adv. Financ. Modell.*, 2009, **8**, 1–26.

Bennedsen, M., Lunde, A. and Pakkanen, M., Decoupling the short- and long-term behavior of stochastic volatility. *J. Financ. Econom.*, 2021, **20**(5), 961–1006.

Brémaud, P., *Point Processes and Queues: Martingale Dynamics*, 1981 (Springer: New York).

Carmona, P. and Coutin, L., Fractional Brownian motion and the Markov property. *Electron. Commun. Probab.*, 1998, **3**, 95–107.

Carmona, P., Coutin, L. and Montseny, G., Approximation of some Gaussian processes. *Stat. Inf. Stoch. Process.*, 2000, **3**(1-2), 161–171.

Cont, R. and Das, P., Rough volatility: Fact or artefact? arXiv preprint arXiv:2203.13820, 2022.

Coutin, L. and Decreusefond, L., Abstract nonlinear filtering theory in the presence of fractional Brownian motion. *Ann. Appl. Probab.*, 1999, **9**(4), 1058–1090.

Coutin, L. and Pontier, M., Approximation of the fractional Brownian sheet via Ornstein-Uhlenbeck sheet. *ESAIM Probab. Stat.*, 2007, **11**, 115–146.

Crisan, D. and Miguez, J., Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *Bernoulli*, 2018, **24**(4A), 3039–3086.

Cvitanić, J., Liptser, R. and Rozovskii, B., A filtering approach to tracking volatility from prices observed at random times. *Ann. Appl. Probab.*, 2006, **00**, 1633–1652.

Decreusefond, L. and Üstünel, A.S., Fractional Brownian motion: Theory and applications. In *ESAIM: Proceedings*, Vol. 5, pp. 75–86, 1998.

Ethier, S. and Kurtz, T., *Markov Processes: Characterization and Convergence*, 1986 (John Wiley & Sons: New York).

Frey, R. and Runggaldier, W., A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *Int. J. Theor. Appl. Finance*, 2001, **4**(02), 199–210.

Fukasawa, M., Takabatake, T. and Westphal, R., Is volatility rough? arXiv preprint arXiv:1905.04852, 2019.

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

Gillespie, D., Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E.*, 1996, **54**(2), 2084.

Gneiting, T. and Schlather, M., Stochastic models that separate fractal dimension and the hurst effect. *SIAM Rev.*, 2004, **46**(2), 269–282.

Harms, P., Strong convergence rates for Markovian representations of fractional Brownian motion. arXiv preprint arXiv:1902.01471, 2019.

Jacod, J. and Shiryaev, A., *Limit Theorems for Stochastic Processes*, 2003 (Springer: Berlin, Heidelberg).

Makowski, D., Ben-Shachar, M.S. and Lüdecke, D., bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *J. Open Source Softw.*, 2019, **4**(40), 1541.

Rogers, L.C.G., Things we think we know. In *Options–45 Years Since the Publication of the Black–Scholes–Merton Model: The Gershon Fintech Center Conference*, pp. 173–184, 2023 (World Scientific).

Appendix. Robustness with respect to the choice of J

In this section, we report the results of some numerical experiments intended to investigate the robustness of the estimation method for varying numbers of approximating processes J.

First, we analyze the approximation described in section 3; that is, we aim at comparing the approximating process $W^{(H,J)}$ (see equation (11)) to the true process W^H in terms of the behavior of the first-lag autocorrelation of increments, as the autocorrelation structure is closely related to the Hurst index H and thus instrumental for the performance of our parameter estimation method. To this, for fixed $T = 1$, $N = 480$ and given J, we first simulate W^H via Choleski and then use the same driving Brownian path (that is, the same random numbers) to construct the approximating process. We repeat this procedure 1000 times to be able to compute the mean squared error (MSE) with respect to the true value of the first-lag autocorrelation of increments of an fBM, which is known theoretically. The MSE computation is done for each $H \in \{0.1, 0.2, 0.3, 0.4\}$ and for varying $J \in \{18, 48, 96\}$. The results (on the log scale) are presented in figure A1, where the horizontal lines correspond to the MSE of the Choleski procedure (which obviously does not vary with J, but has its own error due to process discretization). We find that the approximation method (particularly, our choices for ξ_0 and ξ_J) results in a close approximation of the true autocorrelation structure of increments in terms of MSE even for parsimonious J. The approximation mimics the true autocorrelation structure most closely for the case $H = 0.1$, while it is somewhat less accurate in the case $H = 0.4$. This is not surprising, as in the case $H = 0.4$ we use a smaller number of approximating processes and, most importantly, as this case exhibits a higher MSE also for what concerns the simulation of the 'true' process via Choleski.

Second, we study the robustness of the parameter estimates with respect to J. Since the autocorrelation structure is very relevant for parameter estimation in our setup, the previous results translate to reasonable performance of parameter estimators. Figure A2 shows 99% highest-density (that is, narrowest) *credible* intervals for the Hurst index at final time $T = 1$ (with $N = 480$), with the posterior mean depicted as a filled dot for $H \in \{0.1, 0.2, 0.3, 0.4\}$ and for varying $J \in \{18, 48, 96\}$. We can observe that the choice of J does not dominate the Monte Carlo error but has a relatively small impact on the credible intervals for H. We can conclude that, although theoretical results on pathwise convergence hold only for $J \to \infty$, when filtering it is in fact possible to use a relatively small number of approximating processes to keep the computational cost manageable, while simultaneously achieving reasonable estimation accuracy.

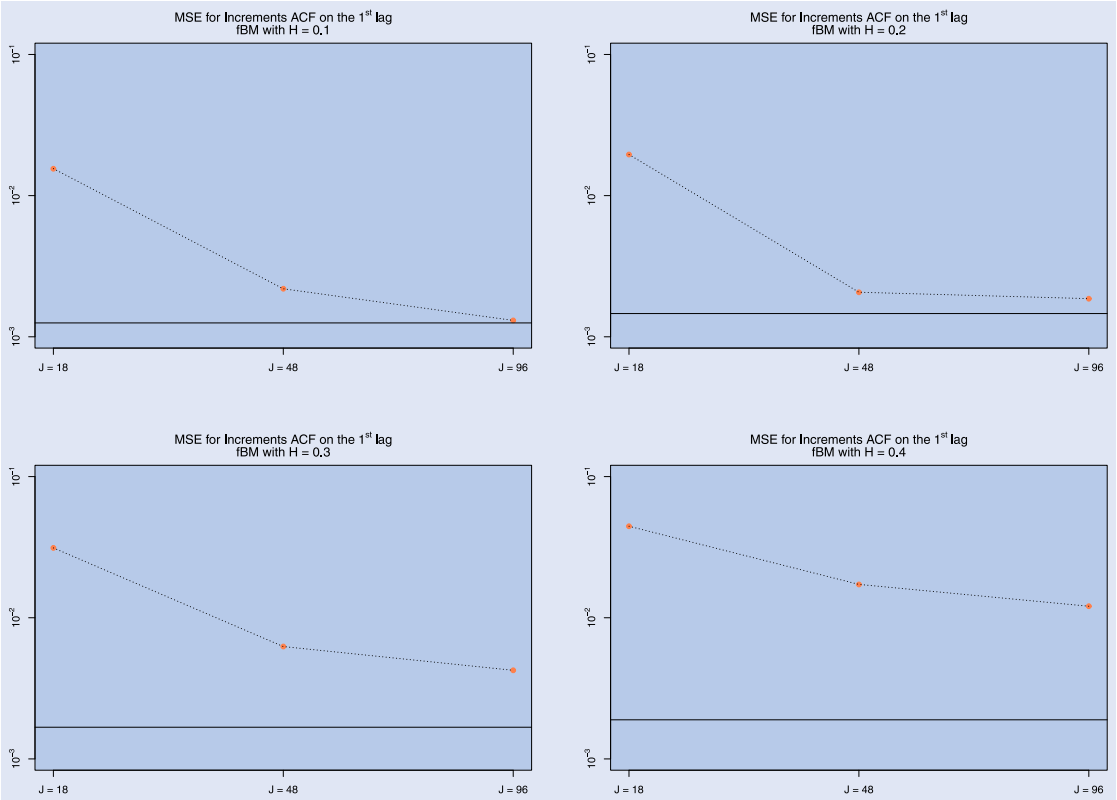

Figure A1. The coral dots and dotted line correspond to the behavior of the MSE (log scale) in first-lag autocorrelation of increments of the approximating process $W^{(H,J)}$ for varying J. The solid horizontal black line depicts the corresponding MSE for the 'true' process simulated with Choleski. Each panel corresponds to a different value of H: starting from the top-left panel and proceeding clockwise, we have the cases $H = 0.1$, $H = 0.2$, $H = 0.3$ and $H = 0.4$.

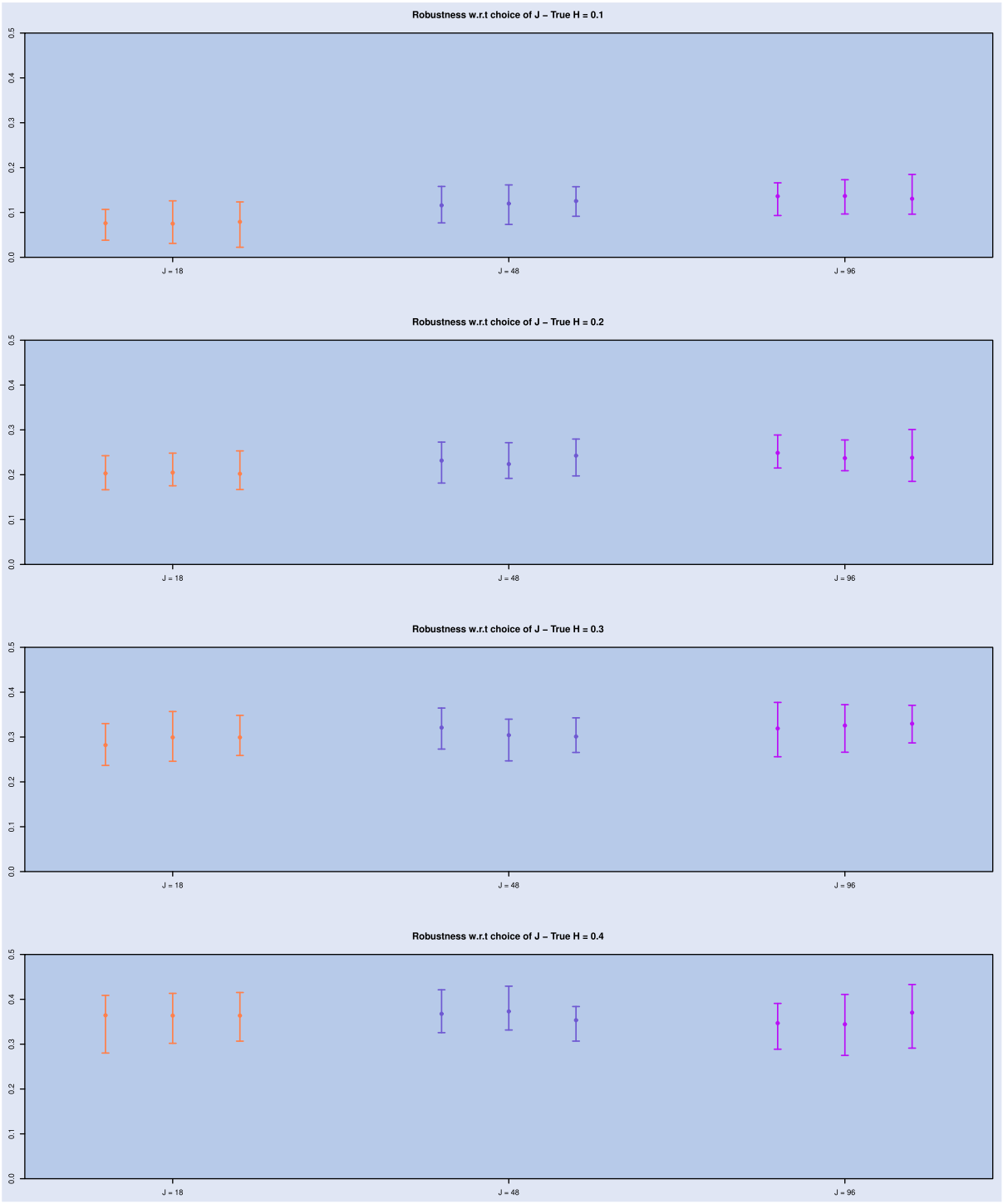

Figure A2. 99% highest-density credible intervals for H at final time. Each panel corresponds to a different value of H: from top to bottom we have $H = 0.1$, $H = 0.2$, $H = 0.3$ and $H = 0.4$. In all panels, the (coral) intervals on the left correspond to $J = 18$, the (blue) intervals in the middle to $J = 48$ and the (purple) intervals on the right to $J = 96$.