

Contents

1	Introduction	1
2	Fractional Brownian motion	3
3	Estimating the roughness of volatility	4
3.1	Instantaneous volatility and realized volatility	5
3.2	Roughness index estimator via p -th variation	7
3.2.1	Estimating the roughness from discrete observations	9
3.2.2	Sample behaviour of $\hat{H}_{L,K}$	9
3.3	Roughness estimator via logarithmic regression	14
3.4	Sequential scale estimator of roughness exponent	18
4	Numerical Experiments	23
4.1	Simple stochastic volatility diffusion model	23
4.2	OU-SV model	29
4.3	A fractional Ornstein-Uhlenbeck model	35
4.4	RFSV model	42
5	Conclusion	44
6	Appendix	46
6.1	Simulation of fractional Brownian motion	46
6.1.1	Proof of Fourier transform in spectral method	48
	References	50

1 Introduction

When modelling prices of financial assets, modelling the volatility process itself plays a crucial role. For the price of an asset S_t price dynamics are often given by a stochastic volatility model driven by a Brownian motion on the form

$$dS_t = \mu_t S_t dt + \sigma_t S_t dB_t \quad (1)$$

where μ_t is a drift term and B_t is a one-dimensional Brownian motion. The term σ_t represents the volatility process, and it is a very important ingredient in the model. However, modelling the volatility process itself is also a complex task. Different approaches to estimate σ_t is used throughout the literature. In the simplest models, the volatility is either constant or a deterministic function of time but in most models the volatility process is modelled as a stochastic process itself [Gatheral et al. \[2018\]](#).

Various authors have proposed various approaches to modelling volatility. The use of fractional Brownian motions and fractional Gaussian noise in volatility modelling was in early applications such as [Bollerslev and Mikkelsen \[1996\]](#) and [Comte and Renault \[1998\]](#) introduced to model long-range dependence in financial time series. A well-known example of such a fractional stochastic volatility model is the one proposed by [Comte and Renault \[1998\]](#) who models the dynamics of the volatility σ_t of an asset as:

$$Y_t = \ln \sigma_t, \quad dY_t = -\gamma Y_t dt + \theta dB_t^H$$

where B^H is a fractional Brownian motion with Hurst exponent H . The long range dependence is modelled by choosing $1 > H > \frac{1}{2}$ [Comte and Renault \[1998\]](#). Fractional Brownian motions can be useful building blocks in stochastic volatility models for several reasons. Firstly, a fractional Brownian motion has the ability to model long range dependence which is measured by the slow decay T^{2H-2} of auto-correlation of functions. Secondly, it can generate trajectories of varying level of roughness (Hölder regularity) which is also governed by the Hurst exponent H . The long-range dependence is manifested over long time scales while roughness is manifested over short time scales. These two properties are therefore very different, and are not necessarily related. However, for a fractional Brownian motion they are linked through self-similarity and are both governed by the Hurst exponent H .

In more recent literature starting with [Gatheral et al. \[2018\]](#), it has been suggested to use fractional stochastic volatility models with $H < \frac{1}{2}$ for modelling volatility. Processes driven by a fractional Brownian motion with $H < \frac{1}{2}$ are referred to as 'rough processes' since these fractional Brownian motion have trajectories rougher than a standard Brownian motion. The reasoning for the choice of $H < \frac{1}{2}$ rely on empirical estimates of the roughness of realized volatility over short intraday time scales. [Gatheral et al. \[2018\]](#) asses the roughness of realized volatility and find that realized volatility exhibit rough behaviour and

concludes that 'volatility is rough'. Since then several rough fractional stochastic volatility models have been suggested such as the rough Heston model [El Euch et al. \[2019\]](#) and the rough Bergomi model [Bayer et al. \[2016\]](#) using a Hurst exponent H near 0.1.

In practice, true spot volatility of the underlying model cannot be observed. Only realized volatility can be observed. Therefore, one cannot necessarily conclude that true spot volatility is rough just because realized volatility exhibit rough behaviour. [Fukasawa et al. \[2022\]](#) and [Cont and Das \[2024\]](#) show that estimation errors when estimating volatility can substantially distort the outcome of roughness estimation and lead to a downward bias in the measurement of roughness such that realized volatility appears to be rough even when instantaneous volatility is not.

To overcome this problem of "illusive roughness" caused by volatility discretization error, roughness estimators taking into account and controlling the measurement error have been introduced by [Han and Schied \[2023\]](#) and [Bolko et al. \[2023\]](#). Both of these estimators take cumulative integrated variance as input.

In this thesis, we extend the analysis from [Cont and Das \[2024\]](#) and investigate further if there is statistical evidence to conclude that 'volatility is rough'. We use the roughness estimators described in both [Cont and Das \[2024\]](#) and [Gatheral et al. \[2018\]](#) to assess the validity of roughness estimates based on realized volatility through detailed simulation studies. Furthermore, we introduce the sequential scale roughness estimator from [Han and Schied \[2023\]](#) to our numerical examples which take cumulative integrated variance as input. We investigate the performance of all three estimators for measuring the roughness of sample paths of stochastic processes based on fractional Brownian motions and other stochastic processes. We then estimate the roughness from realized volatility signals and cumulative realized variance based on high-frequency observations.

Our results are broadly consistent with the findings of [Cont and Das \[2024\]](#). When using our two roughness estimators that take instantaneous or realized volatility as input, we find that even when instantaneous volatility has diffusive dynamics or is smooth (i.e. $H \geq \frac{1}{2}$) realized volatility exhibits rough behaviour corresponding to a Hurst exponent much smaller than 0.5. In all our numerical examples realized volatility is estimated to be rough no matter what the true roughness of the underlying model is. Especially, we show that in some cases diffusive models or even smooth processes are consistent with an apparent Hurst exponent $H \simeq 0.1$ for realized volatility, similar to the values reported in empirical studies in 'rough volatility' literature. This indicates that the claim 'volatility is rough' is far from an established fact.

For the sequential scale estimator, we find that it performs well when integrated variance approximated directly from instantaneous volatility is used as

input. With minor modifications, the estimator also performs reasonably well when using realized variance as input. In particular, for diffusive models, it estimates the true roughness with high accuracy.

However, the sequential scale estimator remains somewhat susceptible to estimation errors in the cumulative integrated variance. Although this estimation error is small, it is sufficient to distort the outcome of the roughness estimation. Specifically, for smooth processes (i.e., $H > \frac{1}{2}$), the roughness estimates derived from realized variance are often near 0.5, suggesting diffusive behaviour. Similarly, rough processes are estimated to be closer to diffusive behaviour (i.e., $H = \frac{1}{2}$) than they actually are.

Thus, while we do not observe 'illusory roughness' in realized data when using the sequential scale estimator, it does not fully mitigate the distortion caused by estimation errors.

2 Fractional Brownian motion

Fractional Brownian motions work well as building blocks for stochastic volatility models. In this section, we will define the fractional Brownian motion, and briefly discuss some of fBm's most important properties. The definitions and properties will mainly follow the theory in [Biagini \[2008\]](#).

A fractional Brownian motion (fBm for short) $(B_t^H)_{t \in \mathbb{R}}$ with Hurst parameter $H \in (0, 1)$ is a centered continuous Gaussian process with covariance function

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}) \quad (2)$$

for $s, t \geq 0$. For $H = \frac{1}{2}$ the fractional Brownian motion reduces to an ordinary Brownian motion. The incremental process of a fractional Brownian motion is called fractional Gaussian noise.

From (2) it can be seen that a fractional Brownian motion has the following properties:

1. $B_0^H = 0$ and $\mathbb{E}[B_t^H] = 0$ for all $t \geq 0$.
2. B^H has homogeneous increments i.e. $B_{t+s}^H - B_s^H$ has the same law of B_t^H for $s, t \geq 0$.
3. B^H is a Gaussian process and $\mathbb{E}[(B_t^H)^2] = t^{2H}$, $t \geq 0$ for all $H \in (0, 1)$.
4. B^H has continuous trajectories.

The fractional Brownian motion can be divided into three very different families corresponding to $0 < H < \frac{1}{2}$, $H = \frac{1}{2}$ and $\frac{1}{2} < H < 1$. Looking at the correlation between two increments, we have that for $H = \frac{1}{2}$ the increments are independent since it is a Brownian motion. For $H \neq \frac{1}{2}$ the increments are not independent.

From (2) it can be seen that the covariance between the increments $B_{t+h}^H - B_t^H$ and $B_{s+h}^H - B_s^H$ with $s+h \leq t$ and $t-s = nh$ is

$$\rho_H(n) = \frac{1}{2} h^{2H} ((n+1)^{2H} + (n-1)^{2H} - 2n^{2H}).$$

In particular, we have for $n = 1$ meaning two increments of the form $B_{t+h}^H - B_t^H$ and $B_{t+2h}^H - B_{t+h}^H$ that the increments are positively correlated for $H > \frac{1}{2}$ and negatively correlated for $H < \frac{1}{2}$.

When $h = 1$ meaning that the incremental process has time step 1 we have that

$$\rho_H(n) = \frac{1}{2} ((n+1)^{2H} + (n-1)^{2H} - 2n^{2H}) \sim H(2H-1)n^{2H-2}$$

as $n \rightarrow \infty$. Thus, we see that $\sum_{n=1}^{\infty} \rho_H(n) = \infty$ for $H > \frac{1}{2}$ and $\sum_{n=1}^{\infty} |\rho_H(n)| < \infty$ for $H < \frac{1}{2}$. The non-summable autocovariance function in the case $H > \frac{1}{2}$ is equivalent to long-range dependence and it indicates slow decay of the covariance function. That is, for $H > \frac{1}{2}$ the fBm has long-range dependence but for $H < \frac{1}{2}$ it does not.

Furthermore, we have that the covariance function of the fBm is homogeneous of order $2H$. That is,

$$\mathbb{E}[B_{at}^H B_{as}^H] = a^{2H} \mathbb{E}[B_t^H B_s^H].$$

That implies that B_{at}^H has the same distribution law as $a^{-H} B_t^H$ for all $H \in (0, 1)$. This is known as statistical self-similarity with Hurst index H .

Furthermore, it can be shown that B_t^H is almost surely Hölder continuous of order α for any $\alpha < H$. Hölder continuity is also known as Hölder regularity or simply as the roughness or smoothness of a sample path. Therefore, the roughness of B_t^H completely follows the Hurst exponent H . Thus, we have that a fractional Brownian motion with Hurst exponent $\frac{1}{2} < H < 1$ has long range dependence in increments and trajectories smoother than a Brownian motion while for $0 < H < \frac{1}{2}$ the fBm has negatively correlated increments (without long range dependence) and trajectories rougher than a Brownian motion [Bisagini \[2008\]](#).

In this thesis, we will be simulating high-frequency data based on fractional Brownian motions when performing numerical experiments. Exact simulation of a fractional Brownian motion is computationally challenging, and we will in this thesis simulate paths from the fBm by using a spectral method. More details on this method is described in [Appendix 6.1](#).

3 Estimating the roughness of volatility

Measuring the roughness of volatility is in practice not an easy task. In practice, we observe only a single price path with discrete observation times, and we need

to determine the roughness of this path sampled at high frequency. In order to design proper estimators it is crucial to be able to determine the roughness of realized volatility, and multiple ways of measuring the roughness exist in literature. In this section we will introduce three different roughness estimators and the theory behind them. Firstly, we will be using the roughness index estimator via p -th variation $\widehat{H}_{L,K}$ which was first introduced by [Cont and Das \[2024\]](#). Secondly, we will describe the roughness estimator via log regression \widehat{H}_R introduced by [Gatheral et al. \[2018\]](#). Thirdly, we will be using the sequential scale estimator $\widehat{\mathcal{R}}_n^s$ originally developed by [Han and Schied \[2023\]](#). The explained theory and rationale behind the estimators will closely follow the theory from the original articles. In [Section 4](#) we will conduct numerical experiments, and we will compare the performance of the three roughness estimators.

3.1 Instantaneous volatility and realized volatility

Before we introduce the details of our roughness estimators we will start by explaining the concepts of instantaneous volatility and realized volatility. As mentioned in [\(1\)](#) price dynamics are often modelled on the form

$$dS_t = \mu_t S_t dt + \sigma_t S_t dB_t.$$

The term σ_t represents the volatility process, and σ_t is also called instantaneous volatility or spot volatility. Thus, instantaneous volatility σ_t is the true volatility of the underlying model. In stochastic volatility models, σ_t is represented as a stochastic process itself often driven by a Brownian motion or a fractional process. Contrary to prices of an asset, instantaneous volatility cannot be directly observed and needs to be estimated from prices.

In a practical situation, the price of the asset S_t at time t is usually observed over a non-uniform time grid of $[0, T]$:

$$\pi^n = \left(0 = t_0^n < t_1^n < \dots < t_{N(\pi^n)}^n = T\right)$$

where n is the sample frequency and $N(\pi^n)$ denotes the number of intervals in the partition π^n . In order to study high-frequency asymptotics of roughness estimators, it is often assumed that

$$|\pi^n| := \sup_{i=1, \dots, N(\pi^n)} |t_i^n - t_{i-1}^n| \xrightarrow{n \rightarrow \infty} 0. \quad (3)$$

That is, the size of the largest interval of π^n will converge to 0 as $n \rightarrow \infty$ where n can be thought of as a sampling frequency i.e. the number of samples per second (or per other unit).

Instantaneous volatility σ_t can then be recovered from integrated variance or quadratic variation of log-prices $\log S$ along this time grid by

$$\sigma_t^2 = \frac{d}{dt} [\log S]_\pi(t)$$

where the quadratic variation of the log-price $[\log S]_\pi(t)$ is defined as

$$[\log S]_\pi(t) = \lim_{n \rightarrow \infty} \sum_{\pi^n \cap [0, t]} \left(\log \frac{S(t_{i+1}^n)}{S(t_i^n)} \right)^2 = \lim_{n \rightarrow \infty} RV_t^2(\pi^n)^2$$

where $RV_t^2(\pi^n)^2$ is realized variance along the sampling grid π^n . Hence, σ_t can be recovered from a high-frequency limit of realized variance where the realized variance along the sampling grid π_n , $RV_t^2(\pi^n)^2$, is defined as

$$RV_t(\pi^n)^2 = \sum_{\pi^n \cap [0, t]} \left(\log \frac{S(t_{i+1}^n)}{S(t_i^n)} \right)^2.$$

We then simply define realized volatility as the square root of realized variance. This leads to the following formal definition:

Definition 1. *Realized volatility of a price process S over the time interval $[t, t + \Delta]$ along the sampling grid π^n is defined as*

$$RV_{t, t+\Delta}(\pi^n) = \sqrt{\sum_{\pi^n \cap [t, t+\Delta]} \left(\log \frac{S(t_{i+1}^n)}{S(t_i^n)} \right)^2}.$$

If the price S_t is a continuous semimartingale then the quadratic variation equals the integrated variance of $\log S$ which is given by

$$\langle \log S \rangle = \int_0^t \sigma_s^2 ds.$$

Furthermore, if S_t follows a stochastic volatility model given by price dynamics as in (1), it is known that the realized variance converges in probability to the integrated variance of $\log S$ as sampling frequency increases [Jacod \[2012\]](#). That is,

$$RV_t^2(\pi^n)^2(\pi^n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \int_0^t \sigma_s^2 ds.$$

From that we immediately obtain that realized volatility converges to

$$RV_{t, t+\Delta}(\pi^n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sqrt{\int_t^{t+\Delta} \sigma_s^2 ds}.$$

Hence, along a single price path observed at high-frequency, we can compute the realized volatility as in [Definition 1](#) and use this as an indicator of volatility

$$RV_{t, t+\Delta}(\pi^n) \simeq \sqrt{\Delta} \sigma_t.$$

More precisely, rescaled realized volatility $\frac{1}{\sqrt{\Delta}}RV_{t,t+\Delta}$ can be used as proxy values $\hat{\sigma}_t$ of instantaneous volatility σ_t . As n increases the estimation becomes more accurate [Cont and Das \[2024\]](#).

Several empirical studies including [Gatheral et al. \[2018\]](#) have attempted to estimate roughness of volatility by estimating the roughness of realized volatility signals (i.e. proxy values $\hat{\sigma}_t$) using high-frequency data. Since carried out on high-frequency data, it is then assumed that the difference between $\hat{\sigma}_t$ and σ_t is relatively small, and that the estimated roughness of realized volatility signals also works as a good estimation of the roughness of the instantaneous volatility process σ_t . However, there is no guarantee that $\hat{\sigma}_t$ and σ_t exhibits same kind of roughness behaviour, and as we will illustrate later on estimation error in proxy values $\hat{\sigma}_t$ can substantially distort the roughness estimation.

3.2 Roughness index estimator via p -th variation

In this section, we will describe the roughness index estimator via normalized p -th variation which was introduced in [Cont and Das \[2024\]](#). We will closely follow the theory and methodology behind the estimator described in [Cont and Das \[2024\]](#).

We are working with partitions of our time interval. Consider a sequence of partitions $\pi = (\pi^n)_{n \geq 1}$ of $[0, T]$ where

$$\pi^n = \left(0 = t_0^n < t_1^n < \dots < t_{N(\pi^n)}^n = T\right)$$

represents observation times at frequency n . We again let $N(\pi^n)$ denotes the number of intervals in the partition π^n . We assume that assumption (3) is fulfilled for all time partitions considered. That is, the size of the largest interval of π^n will converge to 0 as the sampling frequency increases $n \rightarrow \infty$.

We will now define the concept of p -th variation along a sequence of partitions $(\pi^n)_{n \geq 1}$:

Definition 2. (p -th variation along a sequence of partitions)
 $x \in C^0([0, T], \mathbb{R})$ has finite p -th variation along the sequence of partitions $\pi = (\pi^n, n \geq 1)$ if there exists a continuous increasing function $[x]_\pi^{(p)} : [0, T] \rightarrow \mathbb{R}_+$ such that

$$\forall t \in [0, T], \quad \sum_{\{t_j^n, t_{j+1}^n\} \in \pi^n : t_j^n \leq t} |x(t_{j+1}^n) - x(t_j^n)|^{p \rightarrow \infty} [x]_\pi^{(p)}(t). \quad (4)$$

If this property holds, then the convergence in (4) is uniform. We call $[x]_\pi^{(p)}$ the p -th variation of x along the sequence of partitions π . We denote $V_\pi^p([0, T], \mathbb{R})$ the set of all continuous paths with finite p -th variation along π .

The concept of roughness will then be formalized using the notion of variation index and roughness index.

Definition 3. (Variation index)

The variation index of a path x along a partition sequence π is defined as the smallest $p \geq 1$ for which x has finite p -th variation along π :

$$p^\pi(x) = \inf \{p \geq 1 : x \in V_\pi^p([0, T], \mathbb{R})\}.$$

Definition 4. (Roughness index)

The roughness index of a path x (along π) is defined as

$$H^\pi(x) = \frac{1}{p^\pi(x)}.$$

When the underlying sequence of partitions is clear, we will simply denote these indexes as $p(x)$ and $H(x)$.

For a real-valued stochastic process $X : [0, T] \times \Omega \rightarrow \mathbb{R}$ the variation index $p^\pi(X(\cdot, \omega))$ of each sample path $X(\cdot, \omega)$ may in principle be different. However, there are many important classes of stochastic processes which have an almost-sure roughness index. For example we have that along the dyadic partition sequence

$$\pi^n = \left(t_i^n = \frac{iT}{2^n}, i = 0, 1, \dots, 2^n \right), \quad (5)$$

the fractional Brownian motion B_t^H has roughness index $H^\pi(x) = H$ matching the Hurst exponent H and the Hölder regularity [Cont and Das \[2024\]](#). However, existence of variation index is in general not obvious. Further details can be seen in [Han and Schied \[2022\]](#).

When estimating roughness from empirical data based on discrete observations, using the p -th variation directly is difficult since it involves checking convergence to an unknown limit. To this end, [Cont and Das \[2024\]](#) introduce the concept of normalized p -th variation which has better asymptotics:

Definition 5. (Normalized p -th variation along a sequence of partitions)

Let π be a sequence of partitions of $[0, T]$ with mesh $|\pi^n| \rightarrow 0$ and $\pi^n = \{0 = t_1^n < t_2^n < \dots < t_{N(\pi^n)}^n = T\}$. $x \in V_\pi^p([0, T], \mathbb{R})$ is said to have normalized p -th variation along π if there exists a continuous function $w(x, p, \pi) : [0, T] \rightarrow \mathbb{R}$ such that:

$$\forall t \in [0, T], \quad \sum_{\pi^n \cap [0, t]} |x(t_{j+1}^n) - x(t_j^n)|^p \rightarrow w(x, p, \pi)(t).$$

Cont and Das [2024] justify the terminology by a result that shows that for a large class of functions with p -th variation, the normalized p -th variation exists and is linear. Furthermore, it can be shown that the normalized p -th variation is a 'sharp' statistic meaning that if a function has finite p -th variation then for all $q \neq p$ the normalized p -th variation is either zero or infinite. Additionally, it is shown that a fractional Brownian motion has normalized p -th variation along the dyadic partition π^n as in (5) almost-surely.

3.2.1 Estimating the roughness from discrete observations

With these concepts defined we now proceed to explaining how roughness can be estimated from discrete observations. Given discrete observations on a refining partition π^L we define the 'normalized p -th variation statistic' which is the discrete counterpart of the normalized p -th variation as

$$W(L, K, \pi, p, t, X) := \sum_{\pi^K \cap [0, t]} \frac{|X(t_{i+1}^K) - X(t_i^K)|^p}{\sum_{\pi^L \cap [t_i^K, t_{i+1}^K]} |X(t_{j+1}^L) - X(t_j^L)|^p} \times (t_{i+1}^K - t_i^K). \quad (6)$$

This definition involves two frequencies K and $L \gg K$. We can consider L as the sample frequency while K is a block frequency such that π^K is a subpartition of π^L . Thus, we are grouping the sample size L into K many groups where each group contains exactly $\frac{L}{K}$ consecutive points. It can be seen that the normalized p -th variation statistic converges to the normalized p -th variation as L and K increase. That is,

$$\lim_{K \rightarrow \infty} \lim_{L \rightarrow \infty} W(L, K, \pi, p, t, X) = w(x, p, \pi)(t).$$

The variation index estimator $\hat{p}_{L,K}^\pi(X)$ associated with data sampled on π^L can then be obtained by computing $W(L, K, \pi, p, t, X)$ for different values of p such that the following equation can be solved for $p_{L,K}^\pi(X)$

$$W(L, K, \pi, \hat{p}_{L,K}^\pi(X), t, X) = T. \quad (7)$$

The corresponding roughness index estimator is then defined as

$$\hat{H}_{L,K}^\pi(X) = \frac{1}{\hat{p}_{L,K}^\pi(X)}.$$

If the underlying dataset and partition is clear we will also denote this estimator simply as $\hat{H}_{L,K}$. We will refer to $\hat{H}_{L,K}$ as the roughness index estimator via p -th variation.

3.2.2 Sample behaviour of $\hat{H}_{L,K}$

We will now study the behaviour of the roughness estimator based on high-frequency simulated data. We will be simulating from a fractional Brownian

motion B_t^H where the true roughness is known to be H . For the simulated paths we then compute the roughness index estimator via p -th variation $\hat{H}_{L,K}$ and investigate its performance. We will in these examples be using a uniform partition of the time interval $[0, 1]$ with

$$\pi^n = \left(0 < \frac{1}{n} < \frac{2}{n} < \dots < 1 \right).$$

We generate data from four fractional Brownian motions with different Hurst exponents $H = \{0.1, 0.3, 0.5, 0.8\}$. For every simulated path we compute the statistic $W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H)$ for different values of p . Thus, we use π^L as our partition of time and generate $L = 300 \times 300$ points uniformly on the interval $[0, 1]$.

The roughness estimator $\hat{H}_{L,K}$ solves equation (7) with $T = 1$. To visualize this we plot $\log(W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H))$ against $H = 1/p$. The results are presented in Figure 1. The solid black line is the value of $\log(W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H))$ for different values of $1/p$. The blue vertical line represents the estimated roughness index whereas the black dotted vertical line represent the true Hurst parameter of the underlying model. The blue horizontal line represents $W(L, K, \pi, p, t, X) = 1$ which is the equation $\hat{H}_{L,K}$ solves. Thus, the estimated roughness index $\hat{H}_{L,K}$ is found in the crossing between $\log(W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H))$ (solid black line) and $W(L, K, \pi, p, t, X) = 1$ (blue horizontal line).

We now repeat the estimation procedure for 150 independent sample paths using again $W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H)$. In Figure 2 we have generated histograms of the final estimates $\hat{H}_{L,K}$ for these 150 independent paths. In addition to this, we provide the corresponding summary statistics of the estimated roughness index $\hat{H}_{L,K}$ for the 150 paths in Table 1.

Figure 1, 2 and Table 1 all indicate that the estimated roughness index via p -th variation $\hat{H}_{L,K}$ seems to be a fairly accurate roughness estimator when used on a data set with of $L = 300 \times 300$ generated from the fractional Brownian motion. All the roughness estimates $\hat{H}_{L,K}$ are within a distance 0.05 from the true Hurst parameter as seen in Figure 2. The mean and median of the estimated roughness provided in Table 1 are in all four cases very close to the true Hurst parameter. However, for the fBm with Hurst parameter $H = 0.8$ even the upper quartile of $\hat{H}_{L,K}$ is smaller than the true Hurst parameter. This suggest that the estimated roughness index might be slightly biased downwards in the case of $H = 0.8$. The estimate is, however, still quite accurate also for $H = 0.8$, and we conclude that the roughness estimator $\hat{H}_{L,K}$ performs very well in these simulation examples.

We now set $L = 2000 \times 2000$ and $K = 2000$, and simulate data from a fBm with Hurst parameter $H = 0.1$. We compute $\hat{H}_{L,K}$ and as before we generate a histogram of the roughness estimates based on 150 independent paths. The

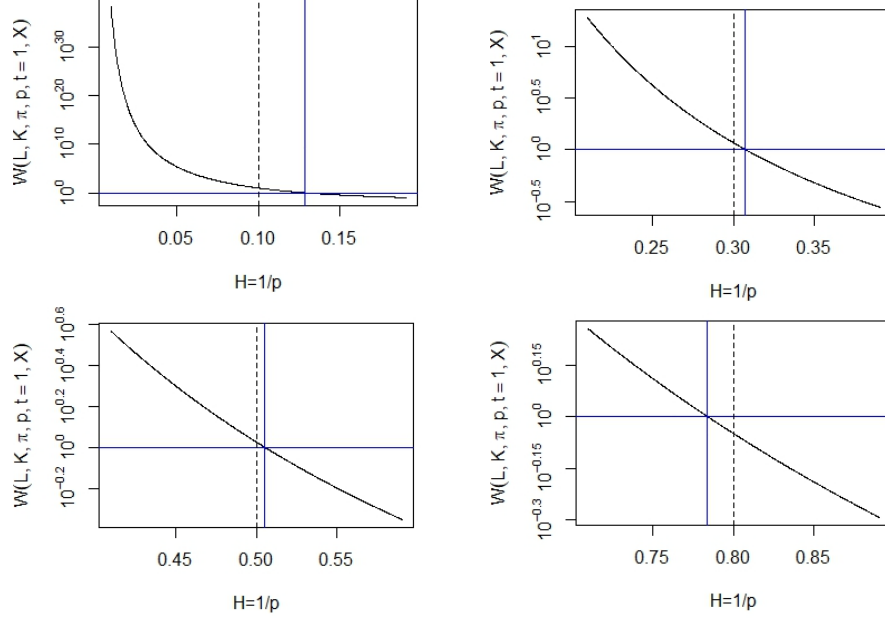


Figure 1: Log-scale plot of the normalized p -th variation statistic for fBm with Hurst parameter $H = \{0.1, 0.3, 0.5, 0.8\}$. The black solid line represents the value of $\log(W(L = 300 \times 300, K = 300, \pi, p, t = 1, X = B^H))$ plotted against $H = 1/p$. The blue vertical line represents $\hat{H}_{L,K}$ using the normalized p -th variation statistic with $L = 300 \times 300$ and $K = 300$. The vertical black dotted line represents the true Hurst parameter.

H	Min.	Lower quartile	Median	Mean	Upper quartile	Max.
0.1	0.0622	0.0944	0.1021	0.1031	0.1149	0.1402
0.3	0.2746	0.2954	0.3012	0.3005	0.3052	0.3177
0.5	0.4743	0.4954	0.5004	0.4997	0.5041	0.5239
0.8	0.7657	0.7797	0.7856	0.7865	0.7937	0.8239

Table 1: Summary of statistics for estimated roughness index $\hat{H}_{L,K}$ from 150 independent simulations of fBm with $L = 300 \times 300$ and $K = 300$.

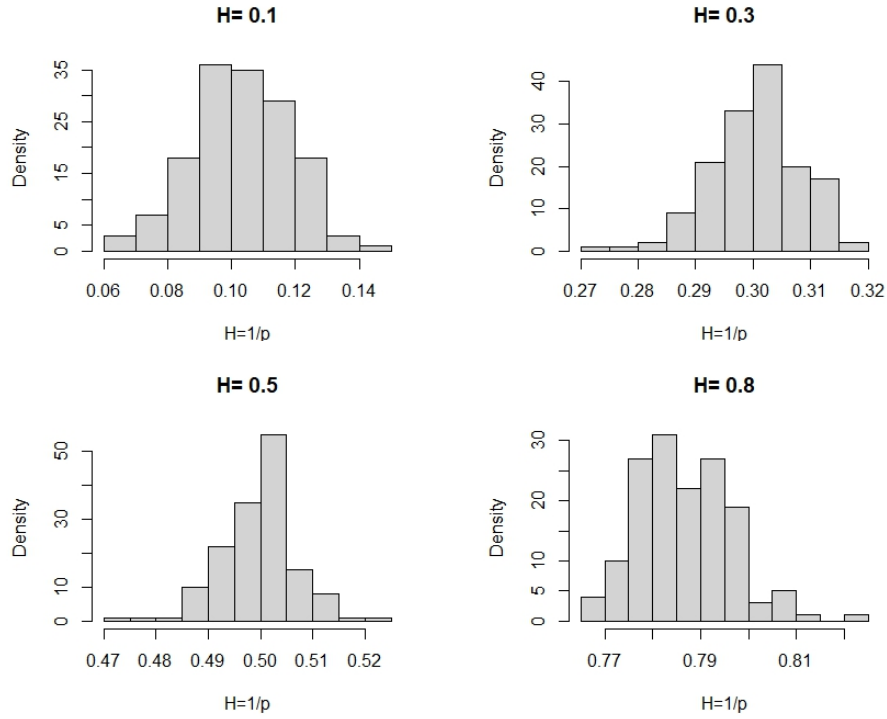


Figure 2: Histogram of estimated roughness index $\hat{H}_{L,K}$ with $L = 300 \times 300$ and $K = 300$ generated from 150 independent simulation of fractional Brownian motions with Hurst parameter $H = \{0.1, 0.3, 0.5, 0.8\}$.

results are presented in Figure 3. Table 2 provides the summary statistics for the estimated roughness index $\hat{H}_{L,K}$ corresponding to the histogram in Figure 3. We observe that the results are similar to those we obtained for $L = 300 \times 300$ and $K = 300$ but that the estimated roughness index $\hat{H}_{L,K}$ seems to be even more accurate. This is not surprising since we know that the normalized p -th variation statistic converges to the normalized p -th variation as L and K increase. However, increasing L and K is also computationally more demanding. The accuracy of $\hat{H}_{L,K}$ obtained when using $L = 300 \times 300$ and $K = 300$ is also satisfactory for many purposes in this thesis.

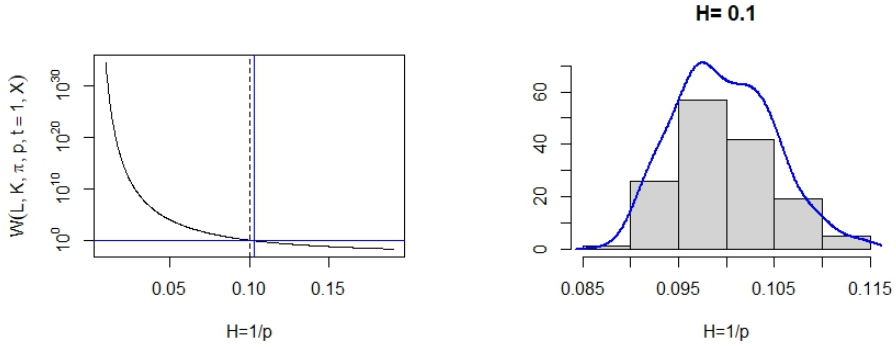


Figure 3: Simulation results for fractional Brownian motion with Hurst parameter $H = 0.1$. **Left:** The value of $\log(W(L = 2000 \times 2000, K = 2000, \pi, p, t = 1, X = B^H))$ plotted against $H = 1/p$ in black. The blue vertical line represents the estimated roughness $\hat{H}_{L,K}$ using $L = 2000 \times 2000$ and $K = 2000$. The vertical black dotted line represents the true Hurst parameter. **Right:** Histogram of estimated roughness index $\hat{H}_{L,K}$ using $L = 2000 \times 2000$ and $K = 2000$ generated by 150 independent simulations of fBm with Hurst parameter $H = 0.1$. The blue line represents a kernel estimator for density.

H	Min.	Lower quartile	Median	Mean	Upper quartile	Max.
0.1	0.0892	0.0963	0.0996	0.0998	0.1031	0.1145

Table 2: Summary of statistics for estimated roughness index $\hat{H}_{L,K}$ from 150 independent simulations of fBm with $H = 0.1$, $L = 2000 \times 2000$ and $K = 2000$.

We further investigate how the choice of $K \ll L$ affects the estimated roughness index $\hat{H}_{L,K}$. In Figure 4, we have plotted the estimated roughness $\hat{H}_{300 \times 300, K}$ for a fractional Brownian motion with Hurst parameter $H = 0.1$ for different values of K with a fixed $L = 300 \times 300$. Note that when $\frac{L}{K}$ is not an integer, the K many groups from the definition of normalized p -th variation statistic (6) do

not contain exactly $\frac{L}{K}$ consecutive points. The code is implemented such that each group will contain either $\lceil \frac{L}{K} \rceil$ or $\lfloor \frac{L}{K} \rfloor$ consecutive points. Figure 4 shows that when K is too low, the estimator $\hat{H}_{300 \times 300, K}$ seems to be underestimating the true Hurst parameter whereas when K is too high the Hurst parameter is overestimated. For $K \approx \sqrt{L}$ the estimated roughness index is quite consistent and close to the true Hurst parameter $H = 0.1$. It is thus natural to use $K = \sqrt{L}$ since $\frac{L}{K}$ will be an integer and the estimator $\hat{H}_{L, K}$ seems to be accurate and consistent in that range.

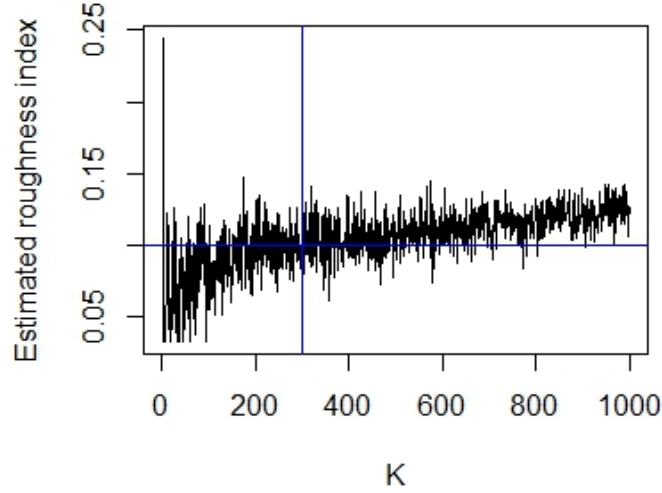


Figure 4: The solid black line represents the estimated roughness index $\hat{H}_{300 \times 300, K}$ plotted against different values of K for a simulation from a fBm with Hurst parameter $H = 0.1$. The blue vertical line represents $K = \sqrt{L} = 300$ whereas the blue horizontal line represents the true Hurst parameter $H = 0.1$.

In summary, these simulation examples show that for realistic sample sizes and frequencies encountered in high-frequency financial data, the roughness estimator via p -th variation $\hat{H}_{L, K}$ is quite accurate and not sensitive to the block size K in the range $K \approx \sqrt{L}$.

3.3 Roughness estimator via logarithmic regression

In this section we will introduce the smoothness estimator via logarithmic regression used in [Gatheral et al. \[2018\]](#). Consider a volatility process on a

time grid $[0, T]$. First we pretend that we have access to discrete observations of the spot volatility process with mesh Δ such that the observations are $\sigma_0, \sigma_\Delta, \dots, \sigma_{k\Delta}, \dots$ for $k \in \{0, \lfloor T/\Delta \rfloor\}$. Here Δ is a natural number and $\Delta \geq 1$. If we set $N = \lfloor T/\Delta \rfloor$, then for $q \geq 0$ we define

$$m(q, \Delta) = \frac{1}{N} \sum_{k=1}^N |\log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta})|^q.$$

Following [Gatheral et al. \[2018\]](#) the main assumption is that for some $s_q > 0$ and $b_q > 0$

$$N^{qs_q} m(q, \Delta) \rightarrow b_q \quad (8)$$

as Δ tends to zero. Under additional technical assumptions, this essentially means that the volatility process belongs to a Besov smoothness space $\mathcal{B}_{q,\infty}^{s_q}$ and does not belong to $\mathcal{B}_{q,\infty}^{s'_q}$ for $s'_q > s_q$. Hence, s_q can be considered as a smoothness parameter. In particular, if $\log(\sigma_t)$ is a fractional Brownian motion with Hurts parameter H , then for any $q \geq 0$ equation (17) holds in probability with $s_q = H$. It can further be shown that the sample paths of the process do indeed belong to $\mathcal{B}_{q,\infty}^{s_q}$. Thus, s_q is a measure of the smoothness (or roughness) of the volatility process [Gatheral et al. \[2018\]](#).

The instantaneous volatility can not be directly observed, and exact computations of $m(q, \Delta)$ is not possible in practice. In order to make use of $m(q, \Delta)$ we must therefore approximate the true spot volatility. The estimated volatility can be computed by the realized volatility or some variation of realized volatility. In the following we will be using the notation $m(q, \Delta)$ with the understanding that we are only approximating the true spot volatility. To estimate the smoothness parameter s_q for each q , we can compute $m(q, \Delta)$ for different values of Δ and regress $\log m(q, \Delta)$ against $\log \Delta$. Note that for a given Δ several $m(q, \Delta)$ can be computed depending on the starting point. For example, if $\Delta = 3$ then the starting point can be $\hat{\sigma}_0$, $\hat{\sigma}_\Delta$ or $\hat{\sigma}_{2\Delta}$. The final measure of $m(q, \Delta)$ is computed as the average of these values of $m(q, \Delta)$ with different starting points.

[Gatheral et al. \[2018\]](#) shows that for a given q , $\log m(q, \Delta)$ values regressed against $\log \Delta$ essentially lie on a straight line. Assuming stationary increments of the log-volatility, this implies that the increments fulfil the scaling property

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = b_q \Delta^{\zeta_q}$$

where $\zeta_q = qs_q > 0$ is the slope of the line associated to q . This result uses that $m(q, \Delta)$ can be seen as the empirical counterpart of $\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q]$. Furthermore, [Gatheral et al. \[2018\]](#) shows that s_q seems to not depend on q , and plotting ζ_q against q shows $\zeta_q \sim qs_q$.

Thus, we can compute an estimate of the smoothness (or roughness) of a volatility process by first regressing $\log m(q, \Delta)$ against $\log \Delta$ for different values of q .

The slope of the straight line is an estimate of ζ_q . Next, we regress ζ_q against q . The slope of the straight line is an estimate of the smoothness parameter s_q . If $\log(\sigma_t)$ is a fractional Brownian motion then $s_q = H$ holds in probability where H is the Hurst parameter.

In Figure 5 we have replicated the method used in [Gatheral et al. \[2018\]](#) to estimate the roughness index of the volatility of S&P500 using the first 3500 days of the daily realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library for S&P. Our results coincide with the results from [Gatheral et al. \[2018\]](#), and the log-regression yields a straight line for all our values of q . Our estimated smoothness of the S&P500 volatility is $H = 0.1421$.

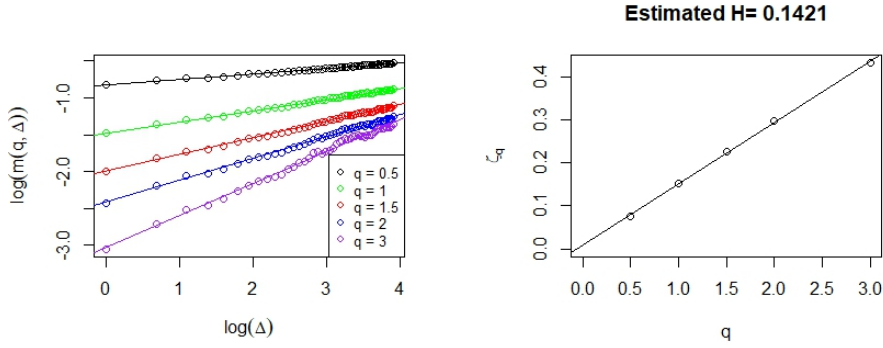


Figure 5: A reproduction of the log-regression method introduced by [Gatheral et al. \[2018\]](#) using the first 3500 days of the daily realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library for S&P. The estimated roughness is $H = 0.1421$.

In Figure 6 we have used the roughness estimator from section 3.1 on the same S&P500 volatility data to estimate the roughness. We have used $W(L = 3500, K = \sqrt{3500}, \pi, p, t = 1, X)$ since the dataset only consist of 3500 days. We have set $t = 1$ since t simply works as a scaling parameter. The conclusions are the same for $t = 1$ as for $t = 3500$. The estimated roughness index is $\hat{H}_{L,K} = 0.1425$. We immediately observe that the estimated roughness index is very close to the estimated smoothness parameter from the log-regression method $s_q = 0.1421$. However, note that $L = 3500$ is much smaller than $L = 300 \times 300 = 90000$ for which we have investigated the accuracy of the roughness estimator via normalized p -th variation statistic. Therefore, the estimated roughness estimator $\hat{H}_{L,K}$ might not be very precise. The results do however indicate that our two different approaches to estimate the roughness of a volatility process coincides.

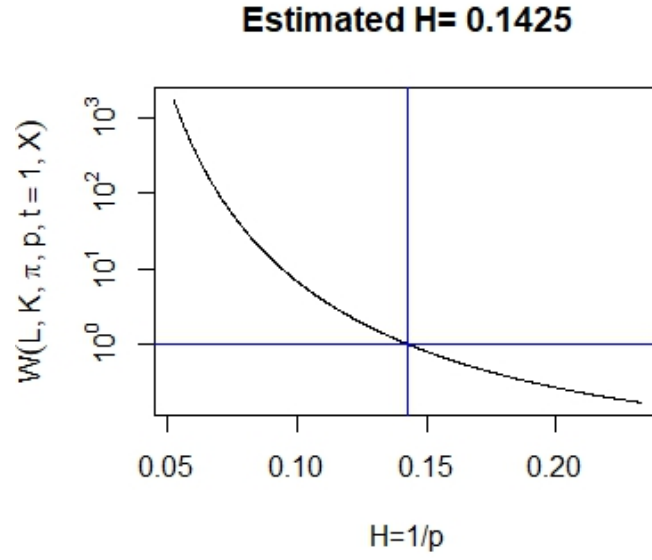


Figure 6: Estimating the roughness of the first 3500 days of S&P500 volatility data from Oxford-Man Institute of Quantitative Finance Realized Library using the roughness estimator from section 3.1 with $L = 3500$ and $K = \sqrt{3500}$. The estimated roughness index is $\hat{H}_{L,K} = 0.1425$.

3.4 Sequential scale estimator of roughness exponent

The third estimator we will introduce, is the sequential scale estimator of roughness first introduced by [Han and Schied \[2023\]](#). This estimator differs from many other roughness estimators by the fact that it is computed directly from discrete observations of realized variance. That is, the input to the estimator needs to be realized variance and not instantaneous variance or volatility. The two previously described estimators are based on having access to discrete observations of instantaneous volatility σ_t . Realized volatility is then used as an approximation of the true spot volatility. This two step approach can be problematic since the estimation error in the realized volatility might substantially distort the outcome of the final roughness estimation. The sequential scale estimator avoids these possible inaccuracies caused by the estimation errors since it is based directly on realized variance.

[Han and Schied \[2023\]](#) are mainly concerned with stochastic volatility models based on fractional Brownian motions, but many aspects of the approach works in a model-free setting. Let $x : [0, 1] \rightarrow \mathbb{R}$ be any continuous function. For $p \geq 1$ the p -th variation of the function x along the n -th dyadic partition is defined as

$$\langle x \rangle_n^{(p)} := \sum_{k=0}^{2^n-1} |x((k+1)2^{-n}) - x(k2^{-n})|^p.$$

The above definition is similar to Definition 1. If there exist a $R \in [0, 1]$ such that

$$\lim_{n \rightarrow \infty} \langle x \rangle_n^{(p)} = \begin{cases} 0 & \text{for } p > \frac{1}{R} \\ \infty & \text{for } p < \frac{1}{R} \end{cases}$$

we refer to R as the roughness exponent of x . The smaller R the rougher is the path x and vice versa. Moreover, if x is a sample path of a fractional Brownian motion, the roughness exponent R is equal to the Hurst parameter H [Han and Schied \[2023\]](#).

Now, since only asset prices and their realized variance can be observed in practice, we can consider this as making discrete observations of

$$y(t) = \int_0^t g(x(s)) ds, \quad 0 \leq t \leq 1, \quad (9)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is sufficiently regular. Here $y(t)$ shall be seen as integrated variance. Therefore, $g(x(t))$ should represent the variance of the underlying model. For instance, if log-volatility is given by a fractional Ornstein-Uhlenbeck process, we will take x as a sample path of a fractional Ornstein-Uhlenbeck process and $g(t) = (e^t)^2 = e^{2t}$.

For the estimator, it is supposed that for a given $n \in \mathbb{N}$ we have access to the discrete observations $\{y(k2^{-n-2}) : k = 0, \dots, 2^{n+2}\}$ of the integrated variance (i.e. the function y in (9)). Based on these data points, the following coefficients are introduced

$$\vartheta_{n,k} := 2^{3n/2+3} \left(y \left(\frac{4k}{2^{n+2}} \right) - 2y \left(\frac{4k+1}{2^{n+2}} \right) + 2y \left(\frac{4k+3}{2^{n+2}} \right) - y \left(\frac{4k+4}{2^{n+2}} \right) \right), \quad (10)$$

for $0 \leq k \leq 2^n - 1$. The estimator for the roughness exponent is then given by

$$\hat{\mathcal{R}}_n(y) := 1 - \frac{1}{n} \log_2 \sqrt{\sum_{k=0}^{2^n-1} \vartheta_{n,k}^2}.$$

Detailed explanation of the rationale behind the estimator can be seen in [Han and Schied \[2023\]](#) where convergence and consistency results are also shown.

The estimator $\hat{\mathcal{R}}_n(y)$ is not scale-invariant, and [Han and Schied \[2023\]](#) suggest to do scale-invariant modifications of $\hat{\mathcal{R}}_n(y)$. Fix $m \in \mathbb{N}$ with $n > m$ and fix $\alpha_0, \dots, \alpha_m \geq 0$ with $\alpha_0 > 0$. The sequential scaling factor λ_n^s is then defined as

$$\lambda_n^s := \arg \min_{\lambda > 0} \sum_{k=n-m}^n \alpha_{n-k} \left(\hat{\mathcal{R}}_k(\lambda y) - \hat{\mathcal{R}}_{k-1}(\lambda y) \right)^2. \quad (11)$$

The sequential scaling estimate $\hat{\mathcal{R}}_n^s(y)$ is then defined as follows

$$\hat{\mathcal{R}}_n^s(y) := \hat{\mathcal{R}}_n(\lambda_n^s y).$$

The idea is that the sequential scaling factor λ_n^s enforces the convergence of $\hat{\mathcal{R}}_n(\lambda_n^s y)$. Thus, $\hat{\mathcal{R}}_n^s(y)$ will in some cases converge faster than $\hat{\mathcal{R}}_n(y)$ and remove bias. [Han and Schied \[2023\]](#) prove that the sequential scale estimator is scale-invariant and has a unique solution for every function $y \in C[0, 1]$.

In practice the roughness exponent is usually estimated from realized variance rather than integrated variance. That is, realized variance is used as an approximation of integrated variance. Let $(m_n)_{n \in \mathbb{N}_0}$ be a fixed increasing sequence, where m_n can be regarded as the number of observed data points used to compute the realized variance over each interval $[k2^{-n}, (k+1)2^{-n}]$. Then we can denote the realized variance used for the sequential scale estimator in the following way

$$\hat{Y}_t^{(n)} := \sum_{k=1}^{\lfloor 2^n m_n t \rfloor} \left(\log S_{\frac{k}{2^n m_n}} - \log S_{\frac{k-1}{2^n m_n}} \right)^2. \quad (12)$$

Thus, the process $\widehat{Y}_t^{(n)}$ is the realized variance calculated from the price process S_t with a mesh size $(2^n m_n)^{-1}$. From $\widehat{Y}_t^{(n)}$ we can define the proxy coefficients $\widetilde{\vartheta}_{n,k}$ as

$$\widetilde{\vartheta}_{n,k} := 2^{3n/2+3} \left(\widehat{Y}_{\frac{4k}{2^{n+2}}}^{(n+2)} - 2\widehat{Y}_{\frac{4k+1}{2^{n+2}}}^{(n+2)} + 2\widehat{Y}_{\frac{4k+3}{2^{n+2}}}^{(n+2)} - \widehat{Y}_{\frac{4k+4}{2^{n+2}}}^{(n+2)} \right).$$

By replacing $\vartheta_{n,k}$ with $\widetilde{\vartheta}_{n,k}$ we can construct an estimator $\widetilde{\mathcal{H}}_n(y)$ that directly estimates the roughness exponent from the realized variance as follows

$$\widetilde{\mathcal{H}}_n(S) := 1 - \frac{1}{n} \log_2 \sqrt{\sum_{k=0}^{2^n-1} \widetilde{\vartheta}_{n,k}^2}$$

where S denotes the price process that $\widetilde{\mathcal{H}}_n$ estimates from. [Han and Schied \[2023\]](#) show that under certain assumptions on the difference between $\widehat{Y}_t^{(n)}$ and $Y_t^{(n)}$, then $\widetilde{\mathcal{H}}_n(S)$ and $\widehat{\mathcal{H}}_n(Y)$ converge to the same limit as n increases provided that m_n grows fast enough. We will by $\widetilde{\mathcal{H}}_n^s(S)$ denote the sequential scale estimator of realized variance which is defined as

$$\widetilde{\mathcal{H}}_n^s(\widehat{Y}_t^{(n)}) := \widetilde{\mathcal{H}}_n(\lambda_n^s \widehat{Y}_t^{(n)})$$

where λ_n^s is computed in the same manner as (11) just using $\widetilde{\mathcal{H}}_n(\widehat{Y}_t^{(n)})$ instead of $\widehat{\mathcal{H}}_n(Y)$.

We will now illustrate the performance of $\widehat{\mathcal{H}}_n^s(y)$ and $\widehat{\mathcal{H}}_n(y)$ in a simple example by replicating a simulation example from [Han and Schied \[2023\]](#). We let x follow a sample path of a fractional Brownian motion B_t^H and set $g(x) = x$. For the computation of $\widehat{\mathcal{H}}_n(y)$ we require observations of y at all values of the time grid $\mathbb{T}_{n+2} := \{k2^{-n-2} : k = 0, 1, \dots, 2^{n+2}\}$. In order to approximate $y(t) = \int_0^t B_s^H ds$ we are simulating the values of B_t^H on a finer time grid \mathbb{T}_N with $N = n + 6$. Then we put

$$Y_{k2^{-n-2}} := 2^{-N} \sum_{j=1}^{2^{N-n-2}k} B_{j2^{-N}}^H, \quad k = 0, 1, \dots, 2^{n+2} \quad (13)$$

which is an approximation of $y(t) = \int_0^t B_s^H ds$ by Riemann sums. In Figure 7 we display the results for the estimators $\widehat{\mathcal{H}}_n$ and $\widehat{\mathcal{H}}_n^s$ for 1000 independent paths of a fractional Brownian motion for different values of n . The left plot of Figure 7 is based on paths from a fractional Brownian motion with Hurst exponent $H = 0.3$ while the right plot is for a fBm with $H = 0.7$. The top plots are from the original estimator $\widehat{\mathcal{H}}_n$ whereas the bottom plots are box plots of the sequential scale estimates $\widehat{\mathcal{H}}_n^s(y)$. We observe from the top plots in Figure 7 that $\widehat{\mathcal{H}}_n$ performs relatively well and seems to converge to the true roughness exponent as n increases. However, the estimator also exhibits a certain bias.

The corresponding results for the sequential scale estimator $\hat{\mathcal{R}}_n^s$ are displayed in the bottom plots of Figure 7. We observe that by passing to the sequential scale estimator we completely remove the bias observed for $\hat{\mathcal{R}}_n$. Overall, $\hat{\mathcal{R}}_n^s$ seems to be a quite accurate estimate especially for $n \geq 15$.

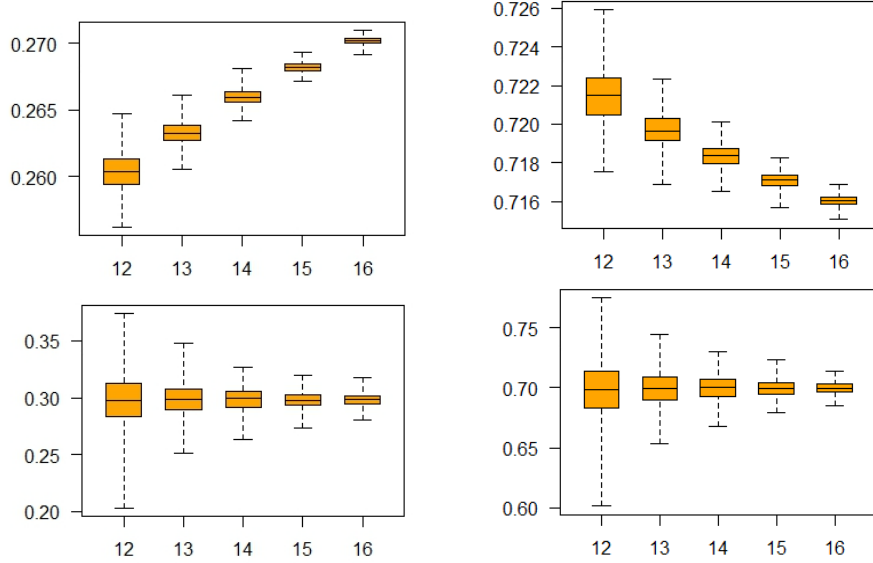


Figure 7: Box plots of the original estimates $\hat{\mathcal{R}}_n(Y)$ (top) and the sequential scale estimates $\hat{\mathcal{R}}_n^s(Y)$ (bottom) for $n = 12, \dots, 16$ based on 1000 independent simulations of a fractional Brownian motion with $H = 0.3$ (left) and $H = 0.7$ (right) and with Y as in (13). The other parameters are chosen to be $m = 3$ and $\alpha_k = 1$ for $k = 0, 1, 2, 3$.

We now make a bit more complicated simulation example. We let log volatility, $\log \sigma_t$ be given by a fractional Ornstein-Uhlenbeck process on the form

$$dX_t = +\rho(\mu - X_t) dt + dB_t^H \quad (14)$$

where B_t^H is a fractional Brownian motion. In this example we use the parameters $\rho = 0.2$, $X_0 = 0$ and $\mu = 2$. As input to the estimator we need discrete observations of

$$\int_0^t \sigma_s^2 ds = \int_0^t e^{2X_s} ds, \quad 0 \leq t \leq 1.$$

To this end, we take again $N = n + 6$ and simulate the values of X_t on the finer time grid $\mathbb{T}_N := \{k2^{-N} : k = 0, 1, \dots, 2^N\}$ by an Euler scheme. We then compute

$$Y_{k2^{-n-2}} := 2^{-N} \sum_{j=1}^{2^{N-n-2}k} \exp(2X_{j2^{-N}}), \quad k = 0, 1, \dots, 2^{n+2} \quad (15)$$

as an approximation of $\int_0^t e^{2X_s} ds$. In Figure 8 we display the results for the estimators $\hat{\mathcal{R}}_n$ and $\hat{\mathcal{R}}_n^s$ for 1000 independent paths of (14) with $n = 12, \dots, 16$. The left plot of Figure 8 is based on paths from a Ornstein-Uhlenbeck process with Hurst exponent $H = 0.3$ while the right plot is for a fOU with $H = 0.7$. The top plots are from the original estimator $\hat{\mathcal{R}}_n(y)$ whereas the bottom plots are box plots of the sequential scale estimates $\hat{\mathcal{R}}_n^s$. We observe from the top plots in Figure 7 that the original estimator $\hat{\mathcal{R}}_n$ performs poorly in this case. However, the sequential scale estimator $\hat{\mathcal{R}}_n^s$ performs almost as well as in the simple case illustrated in Figure 7. This is likely due to the fact that $g(t) = e^{2t}$ used in (15) substantially distort the scale of the underlying process. Since the original estimator is not scale-invariant this leads to poor estimates. However, passing to $\hat{\mathcal{R}}_n^s(y)$ seems to completely remedy this distortion.

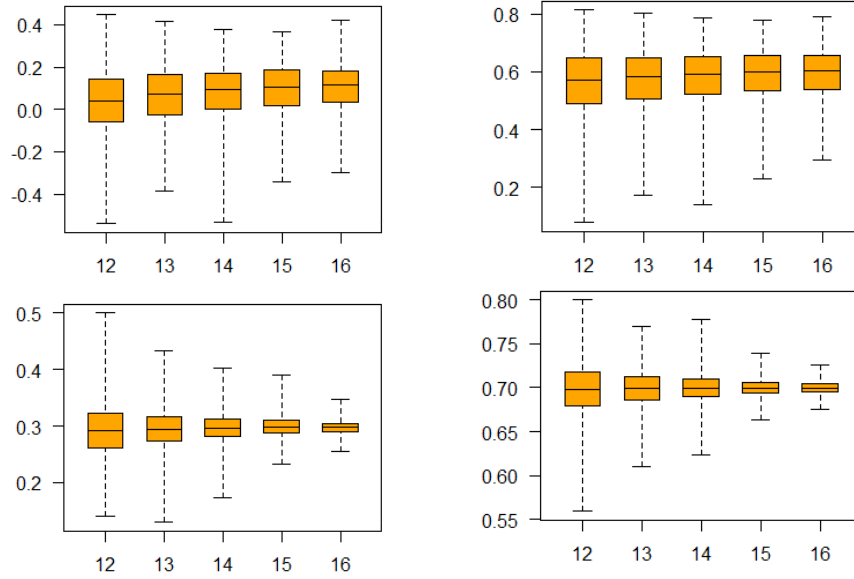


Figure 8: Box plots of the original estimates $\hat{\mathcal{R}}_n(Y)$ (top) and the sequential scale estimates $\hat{\mathcal{R}}_n^s(Y)$ (bottom) for $n = 12, \dots, 16$ based on 1000 independent simulations of a fractional Ornstein-Uhlenbeck process (14) with $H = 0.3$ (left) and $H = 0.7$ (right) and with Y as in (15). The other parameters are chosen to be $m = 3$ and $\alpha_k = 1$ for $k = 0, 1, 2, 3$.

Both of these simulation examples illustrate how the sequential scale estimator $\hat{\mathcal{R}}_n^s$ improves the estimator $\hat{\mathcal{R}}_n$. Overall, $\hat{\mathcal{R}}_n^s$ seems to be performing very well and provide an accurate estimate of the true roughness of the underlying model especially for $n \geq 15$. We will be using $\hat{\mathcal{R}}_n^s$ and $\hat{\mathcal{R}}_n$ with $n \geq 15$ when performing numerical experiments in section 4.

4 Numerical Experiments

In this section we will be performing a number of numerical experiments. We will be estimating the roughness index for stochastic models based directly on the instantaneous volatility σ_t and based on the realized volatility by using price trajectories simulated from the models. By doing this we can investigate how estimation errors impact the estimated roughness index. We will be performing these experiments for stochastic models with various degrees of roughness.

4.1 Simple stochastic volatility diffusion model

First, we will consider the following stochastic volatility model where the volatility is the modulus of a Brownian motion:

$$dS_t = \sigma_t S_t dB_t \quad \text{with} \quad \sigma_t = |W_t|, \quad (16)$$

where S_t is the price of the asset and B_t and W_t are two independent Brownian motions. For the simulation we use $S_0 = 1$ and $T = 1$. The true roughness of this model is simply $H = 0.5$.

We are estimating the realized variance using (??) for 300 consecutive points which corresponds to a 5 minute moving window when the price is updated every second. We estimate the roughness of the volatility process by using the realized volatility and instantaneous volatility as data in our roughness estimators. For every 300 consecutive points we are computing the realized volatility $RV_{t,t+\delta}(\pi^n)$ as in (??) and rescale the volatility to represents the same time interval as the corresponding instantaneous volatility σ_t . The corresponding instantaneous volatility is then the σ_t for the first of these 300 points. In Figure 9, we have made plots for the realized and instantaneous volatility. The left plot of Figure 9 represents the realized volatility of the price process in black and the instantaneous volatility in red. The middle plot of Figure 9 represents the difference between the realized and instantaneous volatility which is the estimation error. The right plot of Figure 9 represents the log difference between the realized and instantaneous volatility. Figure 9 indicates that the estimation error has a complicated dependence structure. There seems to be no obvious pattern in the estimation error and both the estimation error and log estimation error are far from I.I.D.

In Figure 10 we have estimated the roughness index via normalized p -th variation from section 3.1 for realized and instantaneous volatility. In the left plot, we have used the realized volatility as data in the normalized p -th variation statistic and plotted $\log(W(K = 500, L = 500 \times 500, \pi, p, t = 1, X = RV))$ against $H = \frac{1}{p}$. The right graph is a similar plot using instantaneous volatility instead of realized volatility. The obtained roughness estimators are very different for instantaneous and realized volatility. For realized volatility we obtain an estimated roughness of $\hat{H}_{L=500 \times 500, K=500}(RV) = 0.323$ which is much lower than the estimated roughness index for instantaneous volatility $\hat{H}_{L=500 \times 500, K=500}(\sigma) =$

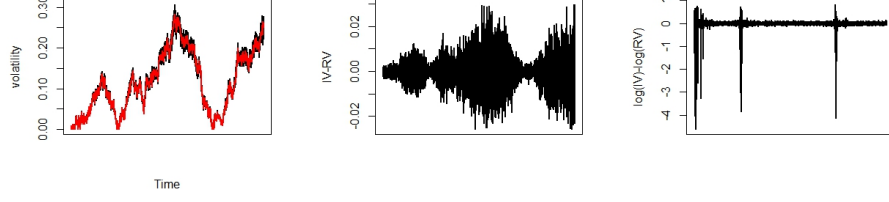


Figure 9: Simulation from model (16). **Left:** The red line represents instantaneous volatility σ_t whereas the black line represents realized volatility RV_t . **Middle:** Corresponding estimation error for the left simulated path. **Right:** Corresponding log estimation error.

0.499. The true roughness of the model is $H = 0.5$. Thus, the roughness estimator is quite accurate when using instantaneous volatility which is in accordance with our results from section 3. However, the estimated roughness when using realized volatility is much rougher than the true roughness. As this is a simulation study we do have any measurement errors, and this rougher behaviour of realized volatility purely comes from the estimation error.

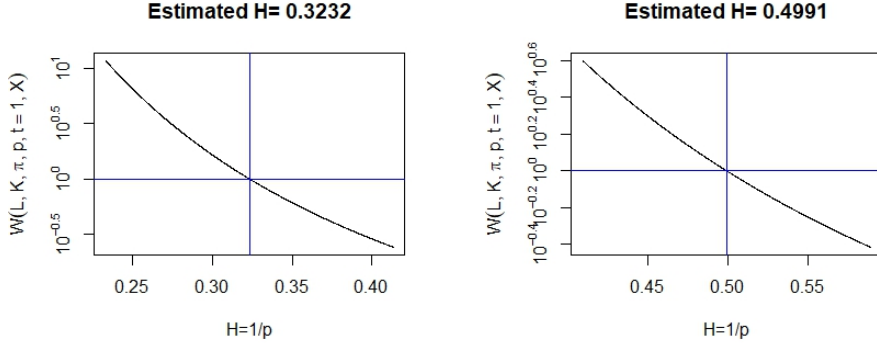


Figure 10: The value of $\log(W(L = 500 \times 500, K = 500, \pi, p, t = 1, X))$ plotted against $H = 1/p$ in black. The blue vertical line represents the estimated roughness $\hat{H}_{L,K}$. **Left:** Estimated roughness index $\hat{H}_{L,K}$ for realized volatility derived from the stochastic volatility model (16). **Right:** Estimated roughness index for instantaneous volatility of the same price path.

In Figure 11 we have plotted the estimated roughness index $\hat{H}_{L=500 \times 500, K}$ against different values of K . The left graph is for realized volatility while the right graph is for instantaneous volatility. The blue lines represents the estimated roughness $\hat{H}_{L,K}$ with $L = 500 \times 500$ and $K = 500$ which are also re-

ported in Figure 10. From the figure, we observe that irrespective of the choice of K for the finite sample dataset of length $L = 500 \times 500$, the realized volatility is significantly rougher than the instantaneous volatility. The estimated roughness of the volatilities are quite consistent as long as K is not too low for both volatilities. However, $\hat{H}_{L=500 \times 500, K}$ seems to be even more consistent for instantaneous volatility than for the realized volatility.

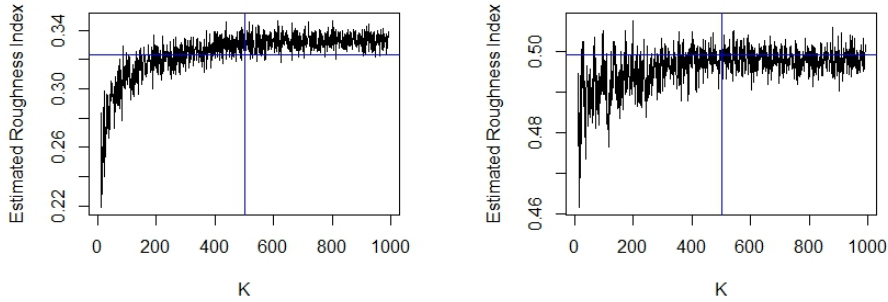


Figure 11: The value of the estimated roughness index $\hat{H}_{L=500 \times 500, K}$ plotted against different values of K . **Left:** Estimated roughness index $\hat{H}_{L=500 \times 500, K}$ for realized volatility derived from the stochastic volatility model (16). The horizontal blue line represents $\hat{H}_{L=500 \times 500, K=500} = 0.323$. **Right:** Estimated roughness index for instantaneous volatility of the same price path. The horizontal blue line represents $\hat{H}_{L=500 \times 500, K=500} = 0.499$.

We now want to use the sequential scale estimator $\hat{\mathcal{R}}_n^s(y)$ to estimate the roughness for the same path that Figure 10 was based on. The sequential scale estimator requires observations of integrated or realized variance $y(t) = \int_0^t g(x(s)) ds$ at all values of the time grid $\mathbb{T}_{n+2} := \{k2^{-n-2} : k = 0, 1, \dots, 2^{n+2}\}$. To accommodate for that we rescale our simulated path from before. We will use $n = 17$ for the roughness estimation. Now, to approximate y properly we set $m_{17} = 2^5 = 32$ which is the number of data points in each interval $[k2^{-n}, (k+1)2^{-n}]$. We then follow the procedure described in section 3.5 and make sure that we have values of the price process on the finer time grid \mathbb{T}_N with $N = n + 7$. Therefore, we need $2^N + 1$ points on the time interval $t \in [0, 1]$. Thus, for these points we have $\Delta t = 2^{-N}$. We rescale the first 2^N points of our simulated Brownian motions B_t and W_t such that they correspond to this new Δt . We then generate the price process as in (16) using these rescaled Brownian motions. In Figure 12 we have plotted the first 100000 data points of the original price process S_t on the left and the corresponding rescaled price process on the right. We observe that the prices processes follow the exact same pattern. The only difference between the graphs is the axes that have been rescaled since we have rescaled the whole

prices process. Therefore, we would expect the same roughness index of the two price processes.

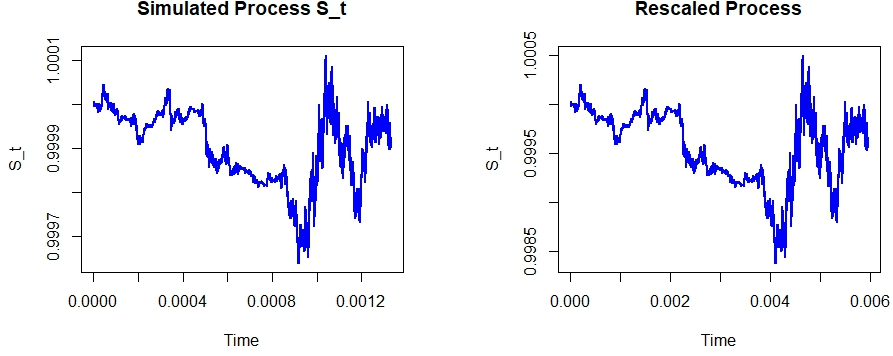


Figure 12: The first 100000 data points of the simulated price process S_t generated from model (16). The left plot is the original simulated path while the right plot is the rescaled price process with fewer points and a different mesh size used for the sequential scale estimator.

We then compute the realized variance as in (12). In order to compare the realized variance/volatility with instantaneous volatility, we also approximate $y(t) = \int_0^t g(x(s)) ds$ directly from σ_t by

$$Y_{k2^{-n-2}} := 2^{-N} \sum_{j=1}^{2^{N-n-2k}} \sigma_{j2^{-N}}^2, \quad k = 0, 1, \dots, 2^{n+2}. \quad (17)$$

In Figure 13 we have plotted these two approximations of integrated variance $y(t)$ and their difference. The left plot of Figure 13 represent the realized variance in black and the approximation Y directly from σ_t as in (17) in red. The middle plot shows the difference between these two approximations of integrated variance, and the right plot shows the difference between the log of the two approximations. Figure 13 indicates that these two ways of computing the integrated variance yield very similar results, and the difference between realized variance and Y is very small. On the left plot it is difficult to distinguish the red and the black line from each other since they fall directly on top of each other. However, there is a small difference as seen on the two plots to the right, and especially the log difference is somewhat big in the very beginning when the integrated variance is very small.

We now proceed to estimating the roughness by using respectively realized variance and spot variance $Y(\sigma_t)$. We denote the sequential scale estimator based directly on realized variance by $\tilde{\mathcal{R}}_n^s(S)$ whereas the roughness estimated from

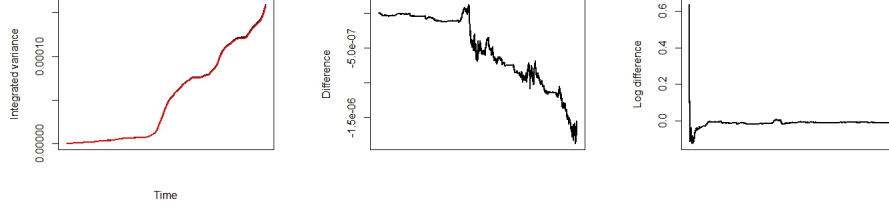


Figure 13: Integrated variance of the original path from model (16) rescaled to fit the sequential scale estimator. **Left:** The red line represents Y computed directly from instantaneous volatility σ_t whereas the black line represents realized variance $\hat{Y}_t^{(n)}$. **Middle:** Corresponding difference between the lines in the left plot. **Right:** Corresponding log difference.

$Y(\sigma_t)$ is denoted by $\hat{\mathcal{R}}_n^s(Y)$. Table 3 provides the roughness estimates obtained from the sequential scale estimator for the rescaled path from model (16). Furthermore, the table also provides the results from the original estimators $\tilde{\mathcal{R}}_n(S)$ and $\hat{\mathcal{R}}_n(Y)$ that are used for the computations of $\tilde{\mathcal{R}}_n^s(S)$ and $\hat{\mathcal{R}}_n^s(Y)$. Even though the difference between realized variance $\hat{Y}_t^{(n)}$ and $Y(\sigma_t)$ is small, the sequential scale estimator yields very different results when used for these two variances as seen in Table 3. The estimated roughness when using spot variance $Y(\sigma_t)$ is very close to the true roughness of the model $H = \frac{1}{2}$. However, when using realized variance the estimated roughness is $\tilde{\mathcal{R}}_n^s(S) = -0.1073$ which is a value that roughness can never take. This indicates that the estimator $\tilde{\mathcal{R}}_n^s(S)$ is very vulnerable to estimation errors in sense of difference between realized variance and integrated variance. This is surprising since the sequential scale estimator has been developed with the focus of avoiding the problems in roughness estimation caused by estimation errors. It is clear that the estimation error is smaller for the sequential scale estimator in the sense that the difference between realized variance $\hat{Y}_t^{(n)}$ and $Y(\sigma_t)$ observed in Figure 13 is much smaller than the difference between realized volatility and instantaneous volatility observed in Figure 9. However, the small difference between $\hat{Y}_t^{(n)}$ and $Y(\sigma_t)$ is enough to substantially distort the outcomes of the roughness estimation. Furthermore, we observe from Table 3 that the original estimator $\tilde{\mathcal{R}}_n(S)$ actually performs better than the sequential scale estimator $\tilde{\mathcal{R}}_n^s(S)$ when used for realized variance. This is surprising since we illustrated in our simulation examples from section 3.5 how passing to the sequential scale estimator could remove bias and make the estimator converge faster. However, both $\tilde{\mathcal{R}}_n^s(S)$ and $\hat{\mathcal{R}}_n^s(S)$ performs very poorly in this example.

In Figure 14 we have repeated the estimation procedure across 150 independent paths drawn from model (16) and computed $\hat{H}_{L,K}$, $\tilde{\mathcal{R}}_n^s$, $\hat{\mathcal{R}}_n$ and $|\tilde{\mathcal{R}}_n^s(S)|$

	Realized variance	Spot variance $Y(\sigma_t)$
Sequential scale estimator	$\tilde{\mathcal{H}}_n^s(S) = -0.4938$	$\hat{\mathcal{H}}_n^s(Y) = 0.5029$
Original scale estimator	$\tilde{\mathcal{H}}_n(S) = 0.1813$	$\hat{\mathcal{H}}_n(Y) = 0.5332$
Absolute scale estimator	$ \tilde{\mathcal{H}}_n^s(S) = 0.4938$	$ \hat{\mathcal{H}}_n^s(Y) = 0.5029$

Table 3: The roughness estimates from the sequential scale roughness estimator with $n = 17$ and $m_{17} = 2^5$ for model (16) by using respectively realized variance and the approximation of integrated variance from instantaneous volatility $Y(\sigma_t)$. The results for the scale-invariant sequential scale estimator are displayed in the top row whereas the corresponding estimates from the original estimator are displayed in the bottom row.

for all paths. The left plot of Figure 14 displays density plots of the estimators when used on realized volatility and realized variance whereas the right plot displays density plots when using instantaneous volatility and $Y(\sigma_t)$ as input to the estimators. For $\hat{H}_{L,K}$ we observe that when using realized volatility the estimator $\hat{H}_{L,K}(RV)$ systematically underestimates the true roughness exponent of $H = \frac{1}{2}$. However, when used on instantaneous volatility $\hat{H}_{L,K}(IV)$ is very accurate and estimates the true roughness of the model very well.

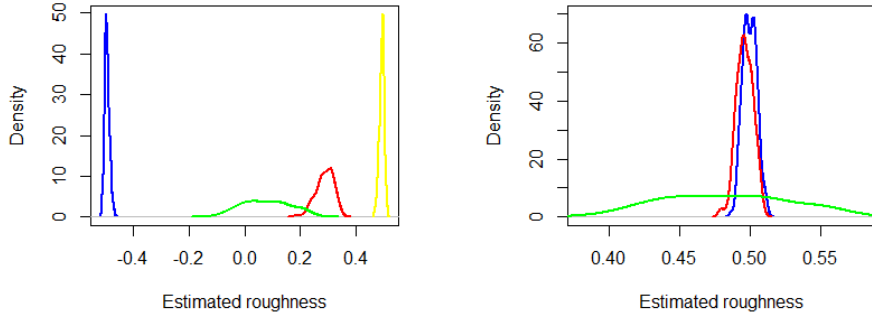


Figure 14: Density plots of the roughness estimators $\hat{H}_{L,K}$ (red), $\tilde{\mathcal{H}}_n^s$ (blue), $\hat{\mathcal{H}}_n$ (green) and $|\tilde{\mathcal{H}}_n^s(S)|$ (yellow) across 150 independent paths drawn from model (16). **Left:** Using realized volatility and realized variance as input. **Right:** Using instantaneous volatility σ_t and $Y(\sigma_t)$ as input.

4.2 OU-SV model

Now, we will simulate from the following OU-SV model:

$$dS_t = S_t \sigma_t dB_t, \quad \sigma_t = \sigma_0 e^{Y_t}, \quad dY_t = -\gamma Y_t dt + \theta dB'_t, \quad (18)$$

where S_t is the price of the asset and B_t and B'_t are two independent Brownian motions. In the simulation, we are using the parameters $\sigma_0 = \gamma = \theta = 1$ and $Y_0 = 0$ and we set $T = 5$ and $S_0 = 1$. The true roughness of this model is $H = 0.5$.

We are using the same procedure as described in section 4.1 to generate our realized and instantaneous volatility. In Figure 15, we have made plots for these volatilities. The left plot of Figure 15 represents the realized volatility of the price process in black and the instantaneous volatility in red. The middle plot of Figure 15 represents the difference between the realized and instantaneous volatility which is the estimation error. The right plot of Figure 15 represents the log difference between the realized and instantaneous volatility. Visually the middle plot suggests that the estimation error has a complicated dependence structure. However, the log estimation error on the right plot of Figure 15 seems to have an I.I.D. Gaussian structure. This is supported by the theory in [Fukasawa et al. \[2022\]](#). However, as shown in Section 4.1 the I.I.D. behaviour of the log estimation error does not in general hold for stochastic diffusive models as assumed in [Fukasawa et al. \[2022\]](#).

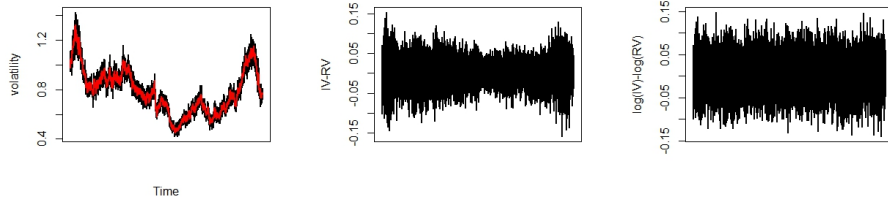


Figure 15: Simulation from the OU-SV model. **Left:** The red line represents instantaneous volatility σ_t whereas the black line represents realized volatility RV_t . **Middle:** Corresponding estimation error for the left simulated path. **Right:** Corresponding log estimation error.

In Figure 16 we have estimated the roughness index via normalized p -th variation for realized and instantaneous volatility. In the left plot, we have used the realized volatility as input to the normalized p -th variation statistic and plotted $\log(W(K = 300, L = 300 \times 300, \pi, p, t = 1, X = RV))$ against $H = \frac{1}{p}$. The right graph is a similar plot using instantaneous volatility instead of realized volatility. The obtained roughness estimators are very different for instantaneous and realized volatility. For realized volatility we obtain an estimated roughness of

$\hat{H}_{L=300 \times 300, K=300}(RV) = 0.155$ which is much lower than the estimated roughness index for instantaneous volatility $\hat{H}_{L=300 \times 300, K=300}(\sigma) = 0.501$. The true roughness of the model is $H = 0.5$. Thus, also for this example the roughness estimator is very accurate when using instantaneous volatility. However, the estimated roughness when using realized volatility is much rougher than the true roughness, and realized variance appear even rougher in this example compared to the previous one. As this is a simulation study this rougher behaviour of realized volatility purely comes from the estimation error, and this illustrates how estimation errors can substantially distort the roughness estimate.

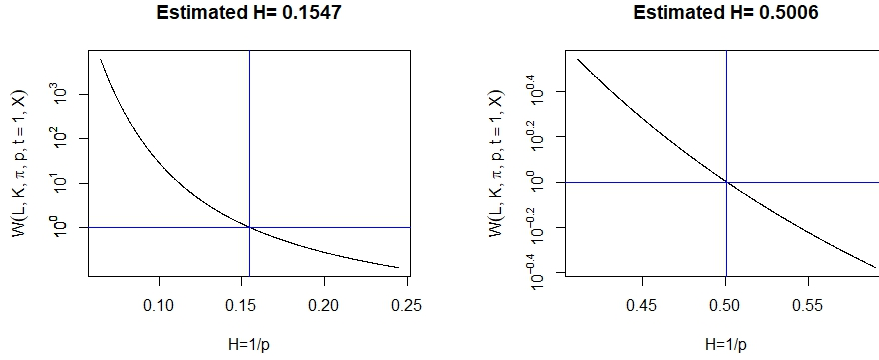


Figure 16: The value of $\log(W(L = 300 \times 300, K = 300, \pi, p, t = 1, X))$ plotted against $H = 1/p$ in black. The blue vertical line represents the estimated roughness $\hat{H}_{L,K}$. **Left:** Estimated roughness index $\hat{H}_{L,K}$ for realized volatility derived from the stochastic volatility model (16). **Right:** Estimated roughness index for instantaneous volatility of the same price path.

Now we use the roughness estimator by logarithmic regression \hat{H}_R for the same instantaneous volatility and realized volatility data. In Figure 17 we display the results for the roughness estimator when using instantaneous volatility. The estimated roughness index is $\hat{H}_R = 0.498$ which is very close to the true roughness of the model $H = \frac{1}{2}$. This indicates that the roughness estimator performs well for instantaneous volatility. In Figure 18 we display the corresponding results for the log regression estimator \hat{H}_R when using realized volatility as input. Here we obtain an estimated roughness of $\hat{H}_R = 0.089$ which is even rougher than the estimate we obtained for the estimator $\hat{H}_{L,K}$ via p -th variation. This rougher behaviour purely comes from the estimation error, and this once again illustrates that using realized volatility for roughness estimation can lead to a dramatic underestimation of the true roughness of the underlying model.

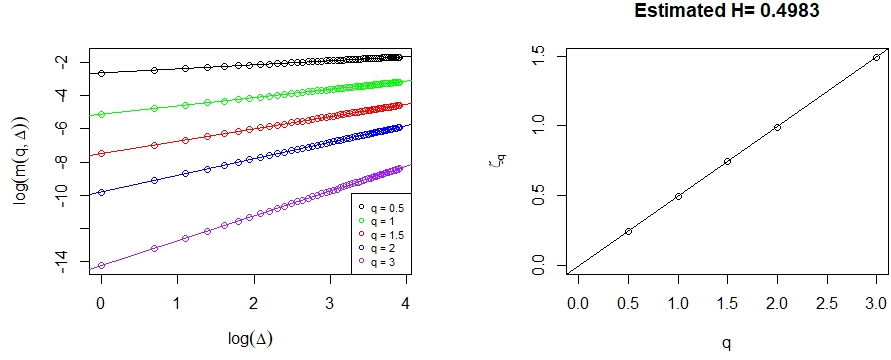


Figure 17: Roughness estimation by log regression using instantaneous volatility simulated from model (18). **Left:** $\log(m(q, \Delta))$ plotted against $\log(\Delta)$. **Right:** Regression coefficients ζ_q plotted against q . The estimated roughness index is $\hat{H}_R = 0.498$.

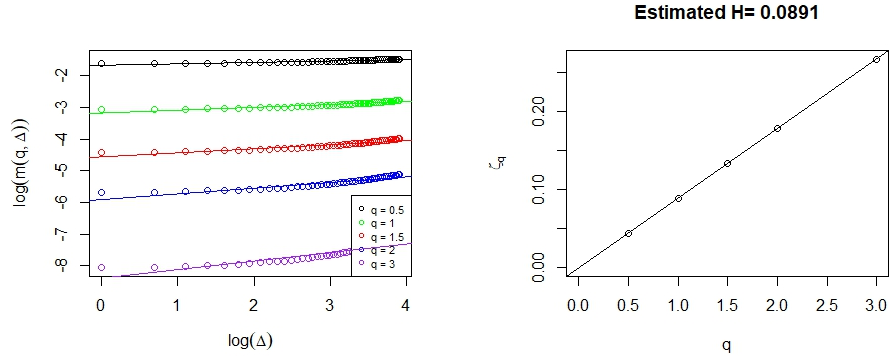


Figure 18: Roughness estimation by log regression using realized volatility for a price path simulated from model (18). **Left:** $\log(m(q, \Delta))$ plotted against $\log(\Delta)$. **Right:** Regression coefficients ζ_q plotted against q . The estimated roughness index is $\hat{H}_R = 0.089$.

We will now use the sequential scale estimator for the same price path. We will be using $n = 17$ and $m_{17} = 2^5$ for the estimation. Furthermore, we use the parameters $m = 4$ and $\alpha_k = 1$ for $k = 0, 1, 2, 3$ for the computation of $\hat{\mathcal{R}}_n^s$. As in section 4.1 we will rescale the price process such that our observed data points fit the required input for the sequential scale estimator. To this end, we need 2^{n+7} data points on the time interval $[0, 1]$. We set $dt = 2^{-n-7}$ and rescale our dB_t and dB'_t to this new dt . We then compute S_t as in (18) using these rescaled dt , dB_t and dB'_t . We compute realized variance $\hat{Y}_t^{(n)}$ as in (12) and $Y(\sigma_t)$ as in (17).

In Figure 19 we have plotted these two approximations of integrated variance $y(t)$ and their difference. The left plot of Figure 13 represent the realized variance in black and the approximation $Y(\sigma_t)$ as in (17) in red. The middle plot shows the difference between these two approximations of integrated variance, and the right plot shows the difference between the log of the two approximations. Just as in the previous example we observe that these two ways of computing the integrated variance yield very similar results, and the difference between realized variance $\hat{Y}_t^{(n)}$ and $Y(\sigma_t)$ is very small. However, the log difference is much bigger in the very beginning when the integrated variance is very small.

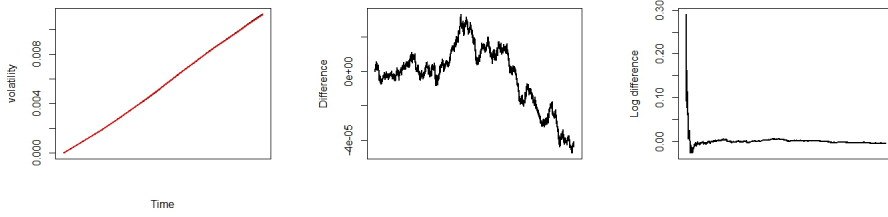


Figure 19: Integrated variance of the original path from model (18) rescaled to fit the sequential scale estimator. **Left:** The red line represents Y computed directly from instantaneous volatility σ_t whereas the black line represents realized variance $\hat{Y}_t^{(n)}$. **Middle:** Corresponding difference between the lines in the left plot. **Right:** Corresponding log difference.

In Table 4 we display the result from the sequential scale roughness estimation. As in the previous example we observe that the sequential scale estimator $\hat{\mathcal{R}}_n^s$ provides an accurate roughness estimation close to the true roughness of the model $H = \frac{1}{2}$ when using $Y(\sigma_t)$ as input. It also significantly improves the original scale estimator $\hat{\mathcal{R}}_n$. However, when using realized variance as input we obtain the roughness estimate $\hat{\mathcal{R}}_n^s(S) = -0.5012$ for the sequential scale estimator which is close to the negative of the true roughness of the model. The small difference between $\hat{Y}_t^{(n)}$ and $Y(\sigma_t)$ observed in Figure 19 completely changes the estimates $\hat{\mathcal{R}}_n^s$ and $\hat{\mathcal{R}}_n$. However, when simply using $|\hat{\mathcal{R}}_n^s|$ we obtain a very

accurate roughness estimate for both realized variance and $Y(\sigma_t)$. These results indicate that $|\hat{\mathcal{R}}_n^s|$ could be a very robust estimator that is unaffected by the estimation error between realized and integrated variance.

	Realized variance	Spot variance $Y(\sigma_t)$
Sequential scale estimator	$\tilde{\mathcal{R}}_n^s(S) = -0.5012$	$\hat{\mathcal{R}}_n^s(Y) = 0.4988$
Original scale estimator	$\tilde{\mathcal{R}}_n(S) = 0.0529$	$\hat{\mathcal{R}}_n(Y) = 0.5407$
Absolute scale estimator	$ \tilde{\mathcal{R}}_n^s(S) = 0.5012$	$ \hat{\mathcal{R}}_n^s(Y) = 0.4988$

Table 4: The roughness estimates from the sequential scale roughness estimator with $n = 17$ and $m_{17} = 2^5$ for model (18) by using respectively realized variance and the approximation of integrated variance from instantaneous volatility $Y(\sigma_t)$.

We now compute the distributions of the estimated roughness index via normalized p -th statistic $\hat{H}_{L,K}$ with $L = 300 \times 300$ and $K = 300$ for realized volatility and instantaneous volatility. We generate the distributions across 2500 independent paths drawn from the OU-SV model (18). The left plot of Figure 21 is the distribution of $\hat{H}_{L,K}$ for realized volatility whereas the right plot is the corresponding distribution for instantaneous volatility as data in the roughness estimator. In addition to this, Table 5 provides the summary statistics for the roughness estimator $\hat{H}_{L,K}$ for realized and instantaneous volatility respectively across the 2500 independent paths. We observe that realized volatility systematically exhibit much rougher behaviour than instantaneous volatility. The roughness estimates based on instantaneous volatility are very close to the roughness of the true model $H = 0.5$. However, the roughness estimates generated from realized volatility are much smaller with a mean of $\hat{H}_{L,K} = 0.1549$ across the 2500 independent simulations.

	Realized volatility	Instantaneous volatility
Min.	0.1192	0.4698
1st quartile	0.1480	0.4938
Median	0.1551	0.4988
Mean	0.1549	0.4987
3rd quartile	0.1621	0.5039
Max.	0.1873	0.5257

Table 5: Summary of statistics for estimated roughness index $\hat{H}_{L,K}$ for realized and instantaneous volatility across 2500 independent simulations the OU-SV model (18) with $L = 300 \times 300$ and $K = 300$.

Note that our results deviates slightly from the results in Cont and Das [2024] for

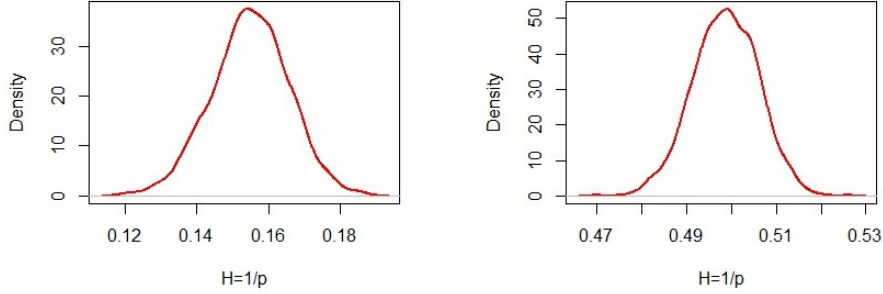


Figure 20: Distribution of the estimated roughness index $\hat{H}_{L,K}$ via normalized p -th variation using $L = 300 \times 300$ and $K = 300$ across 2500 independent simulations of the OU-SV model (18). **Left:** Realized volatility. **Right:** Instantaneous volatility.

a similar model. This might be caused by Cont and Das [2024] using the mean of σ_t across the 300 consecutive data points as their instantaneous volatility which creates a slightly smoother volatility process. Furthermore, Cont and Das [2024] might be using a different T than us which in the OU-SV model will significantly impact the size of the estimation error and thus the estimated roughness index for realized volatility when all other parameters are kept constant. The conclusions are however the same. Realized volatility does systematically exhibit much rougher behaviour than instantaneous volatility and than the true roughness of the model. This rougher behaviour is solely caused by the estimation error, and it indicates that roughness estimated based on realized volatility can be unreliable.

In Figure 21 we have made density plots for the other roughness estimators across 150 independent paths generated from the OU-SV model (18). The left plot of Figure 21 displays density plots of the estimators when used on realized volatility and realized variance whereas the right plot displays density plots when using instantaneous volatility and $Y(\sigma_t)$ as input to the estimators. The density plot of the roughness estimator by log regression \hat{H}_R is represented by the red line. It provides very constant results. For instantaneous volatility it accurately estimates the true roughness of the underlying model whereas it for realized volatility systematically provides a roughness estimate near 0.09 which is much rougher than the true roughness exponent $H = \frac{1}{2}$. Furthermore, we observe that the sequential scale estimator $\hat{\mathcal{R}}_n^s(S)$ systematically estimate the roughness of realized variance near -0.5 . Consequently, we observe that $|\hat{\mathcal{R}}_n^s|$ is very accurate both when used for realized variance $\hat{Y}_t^{(n)}$ and for spot variance $Y(\sigma_t)$, and it is the only estimator in this example which accurately estimate the true roughness of the underlying model based on observations of the actual price process S_t .

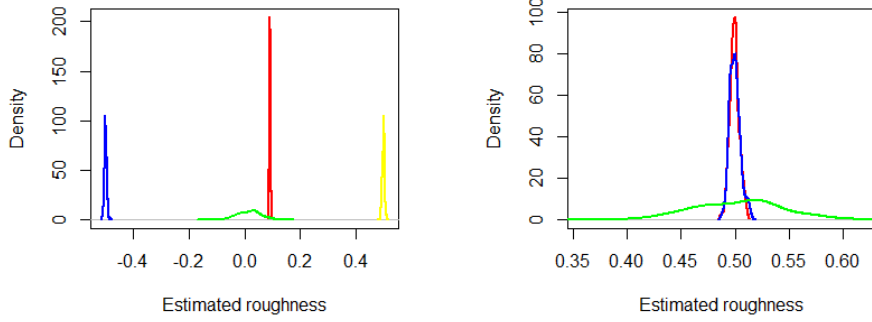


Figure 21: Density plots of the roughness estimators \hat{H}_R (red), $\tilde{\mathcal{R}}_n^s$ (blue), $\hat{\mathcal{R}}_n$ (green) and $|\tilde{\mathcal{R}}_n^s(S)|$ (yellow) across 150 independent paths drawn from the OU-SV model (18). **Left:** Using realized volatility and realized variance as input. **Right:** Using instantaneous volatility σ_t and $Y(\sigma_t)$ as input.

4.3 A fractional Ornstein-Uhlenbeck model

In the two previous models, instantaneous volatility follows a diffusive behaviour similar to a Brownian motion with $H = 0.5$. We now consider a more general case of a stochastic volatility model where the volatility process has a general roughness index $H \in (0, 1)$ generated from a fractional Brownian motion. We investigate the model for different roughness indexes and explore how this affect the estimation error and estimated roughness index. Consider the following process where the volatility is described by a fractional Ornstein-Uhlenbeck model:

$$dS_t = S_t \sigma_t dB_t, \quad \sigma_t = \sigma_0 e^{Y_t}, \quad dY_t = -\gamma Y_t dt + \theta dB_t^H, \quad (19)$$

where S_t is the price process, B_t is a Brownian motion, and B_t^H is a fractional Brownian motions with Hurst exponent H . For the simulation we use the parameters $\sigma_0 = \gamma = \theta = 1$ and $Y_0 = 0$, and we set $T = 5$ and $S_0 = 1$.

We now repeat the procedure described in section 4.1 to compute the estimated roughness index $\hat{H}_{L,K}$ with $L = 300 \times 300$ and $K = 300$ for realized and instantaneous volatility. Furthermore, we compute the smoothness parameter from section 3.5 for both the realized and instantaneous volatility. We compute these roughness estimates across different values of the true Hurst exponent H . In Table 6 we have computed roughness estimates from the fractional Ornstein-Uhlenbeck model (19) with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$.

We have used the same seed in the simulation for all values of H such that the underlying stochastic element of the price processes are the same. Thus, the difference between the price processes solely lies in the Hurst exponent H .

H	Instantaneous vol	Realized vol	IV (log reg)	RV (log reg)
0.1	0.126	0.199	0.102	0.197
0.2	0.220	0.256	0.204	0.250
0.3	0.313	0.261	0.305	0.257
0.4	0.408	0.218	0.407	0.188
0.5	0.504	0.159	0.508	0.095
0.6	0.598	0.102	0.607	0.036
0.7	0.691	0.070	0.703	0.012
0.8	0.778	0.052	0.794	0.004

Table 6: Estimated roughness via p -th variation $\hat{H}_{L,K}$ and via log regression for instantaneous and realized volatility for simulated data from a fractional Ornstein-Uhlenbeck model (19) with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$.

The corresponding paths for price process, realized volatility and instantaneous volatility from model 19 with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ are presented in Figure 22. Visually we observe that the price processes do not differ much as the Hurst exponent changes. However, the paths of realized and instantaneous volatility are very different for different Hurst exponents. For small Hurst exponents instantaneous volatility does visually appear more rough than realized volatility. However, instantaneous volatility becomes smoother and smoother as H increases. That is expected since increasing the Hurst exponent H exactly means that the volatility process becomes smoother. However, realized volatility is not gradually getting smoother as H increases. Instead we observe that for large Hurst exponents, realized volatility appears very rough. This is supported by the estimated roughness indices in Table 6. The estimated roughness based on instantaneous volatility is increasing as H increases both for the roughness estimator via p -th variation and the estimator via logarithmic regression. The estimated roughness based on realized volatility is increasing at first as H increases, but for large Hurst exponents the estimated roughness decreases and the estimated roughness of the volatility process is estimated to be very rough when the true Hurst exponent is $H = 0.8$.

In Figure 23 we have plotted the values from Table 6 in a graph. The blue lines represents roughness estimates for instantaneous volatility while the red lines represents roughness estimates for the corresponding realized volatility. The solid lines are the estimated roughness index via p -th variation $\hat{H}_{L,K}$ with $L = 300 \times 300$ and $K = 300$ while the dashed lines represent roughness estimates using the smoothness parameter computed by logarithmic regression. The x -

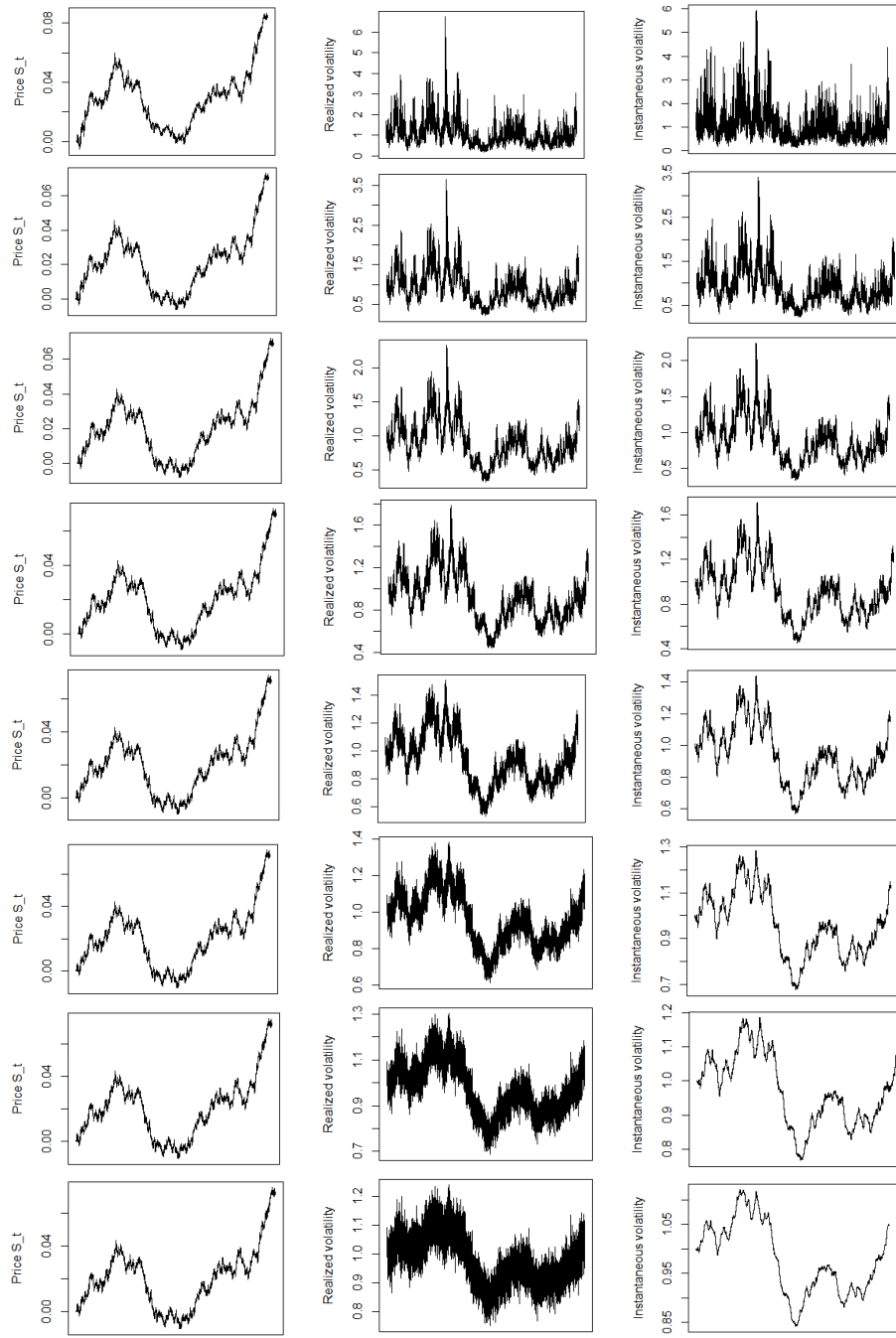


Figure 22: **Left:** Simulated price path $S_t - S_0$ of fractional OU model 19 with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ respectively. **Center:** Realized volatility. **Right:** Instantaneous volatility.

axis represent the true value of H of the simulated fractional OU model. We observe that roughness estimates based on instantaneous volatility give accurate estimates of the true Hurst index H . However, the estimated roughness for realized volatility is quite inaccurate and always stay below 0.3. In particular, when the instantaneous volatility process is generated from a smooth process (i.e. $H > \frac{1}{2}$) the roughness estimate of realized volatility turns out to be a very poor estimate.

It seems that our two roughness estimates coincides and generate similar estimates. Only when the true model is smooth (i.e. $H > 0.5$) the roughness estimate by logarithmic regression for realized volatility seems to systematically be rougher than the corresponding roughness estimate via p -th variation. However, both estimates are very poor in this case.

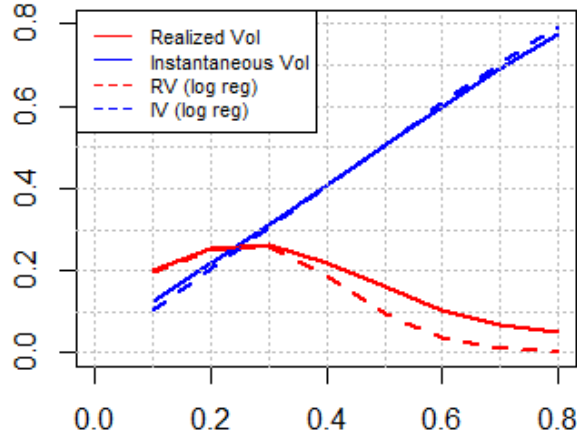


Figure 23: Estimated roughness indexes for realized and instantaneous volatility from a fractional OU model (19) plotted for price paths generated by different values of H . The solid lines are roughness indexes via p -th variation $\hat{H}_{L,K}$ with $L = 300 \times 300$ and $K = 300$ while the dashed lines represents estimated roughness by logarithmic regression.

In Figure 24 we have repeated the above analyses for 100 independent simulated prices processes generated by the fractional Ornstein-Uhlenbeck model (19) but only for the estimated roughness index via p -th variation $\hat{H}_{L,K}$ with $L = 100 \times 100$ and $K = 100$. The figure shows the estimators $\hat{H}_{L,K}(RV)$ and $\hat{H}_{L,K}(\sigma)$

plotted against Hurst exponents $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ for every independent path. The bold black lines represent the mean across 100 independent simulations whereas the dotted lines represent the corresponding 75% confidence interval. If equation (7) have no solution for $H \in (0, 1)$ we set $\hat{H}_{L,K} = 0$.

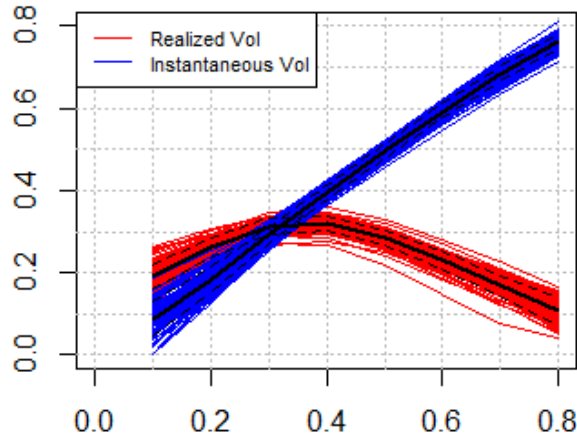


Figure 24: Estimated roughness index via p -th variation for realized and instantaneous volatility plotted against different values of H for 100 independent price paths from a fractional OU model (19). The solid black lines represent the mean across the 100 paths and the dashed black lines represent the corresponding 75% confidence interval.

From Figure 24 we observe that for all 100 independent simulations the roughness estimators exhibit a similar behaviour. Our conclusions about the poor-ness of roughness estimates for realized volatility still hold and was not just a matter of the single price path we studied before. In particular, for the fractional Ornstein-Uhlenbeck model (19) realized volatility always exhibit rough behaviour (i.e. $H < \frac{1}{2}$) no matter what the true roughness of the volatility process is. The estimated roughness based on instantaneous volatility is quite accurate and as this is a simulation study we have no measurement errors so the inaccuracy of the estimators for realized volatility solely comes from the estimation error.

We now introduce the absolute sequential scale estimator $|\hat{\mathcal{R}}_n^s|$ to this example. We take a slightly different approach than in the previous examples and simulate directly the required observations needed for the sequential scale estimator from model (19) with $T = 1$. We use $n = 16$ and $m_n = 2^4$ for the estimation. For the same price path we also compute the roughness index via p -th variation $\hat{H}_{L,K}$ and the roughness estimator via log regression \hat{H}_R by computing instantaneous and realized volatility for every 300 consecutive points. Thus, we use $L = 13980$ and $K = \sqrt{L}$ for the estimator $\hat{H}_{L,K}$. In Table 7 we display the roughness estimates obtained for Hurts exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. We have used the same seed for all H values. The estimates from $|\hat{\mathcal{R}}_n^s|$ are displayed in the two columns furthest to the right in Table 7. IV should for this estimator be understood as integrated variance approximated directly from instantaneous volatility as in (17) whereas RV should be understood as realized variance computed as in (12). For $\hat{H}_{L,K}$ and \hat{H}_R we obtain similar results to the estimates from Table 6 and both of these estimates are inaccurate when used for realized volatility and always stay below 0.3. However, the absolute scale estimator $|\hat{\mathcal{R}}_n^s|$ performs differently. We observe that the estimator is accurate when using spot variance $Y(\sigma_t)$ as input. When using realized variance as input the estimator is not very accurate, but it performs better than the other two roughness estimators, and it does not estimate the process to be rough (i.e. $H < \frac{1}{2}$) when the true underlying model is actually smooth (i.e. $H > \frac{1}{2}$).

H	IV ($\hat{H}_{L,K}$)	RV ($\hat{H}_{L,K}$)	IV (\hat{H}_R)	RV (\hat{H}_R)	IV ($ \hat{\mathcal{R}}_n^s $)	RV ($ \hat{\mathcal{R}}_n^s $)
0.1	0.140	0.226	0.102	0.192	0.067	0.175
0.2	0.219	0.275	0.202	0.246	0.189	0.294
0.3	0.312	0.275	0.300	0.259	0.294	0.445
0.4	0.408	0.216	0.399	0.201	0.396	0.501
0.5	0.503	0.150	0.498	0.109	0.498	0.510
0.6	0.593	0.107	0.596	0.044	0.599	0.510
0.7	0.676	0.077	0.691	0.016	0.699	0.509
0.8	0.749	0.055	0.780	0.005	0.799	0.508

Table 7: Estimated roughness for the three roughness estimators $\hat{H}_{L,K}$, \hat{H}_R and $|\hat{\mathcal{R}}_n^s|$ using instantaneous volatility (IV) and realized volatility/variance (RV) for simulated data from a fractional Ornstein-Uhlenbeck model (19) with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$.

In the left plot of Figure 25 we have plotted the results from Table 7 to visualize the estimates. The right plot of Figure 25 shows roughness estimates from the absolute scale estimator $|\hat{\mathcal{R}}_n^s|$ across 100 independent price paths plotted against the true roughness of the model $H = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. Thus, it is 100 repetitions of the procedure described above. The bold black lines represent the mean across 100 independent simulations whereas the dotted

lines represent a 75% confidence interval.

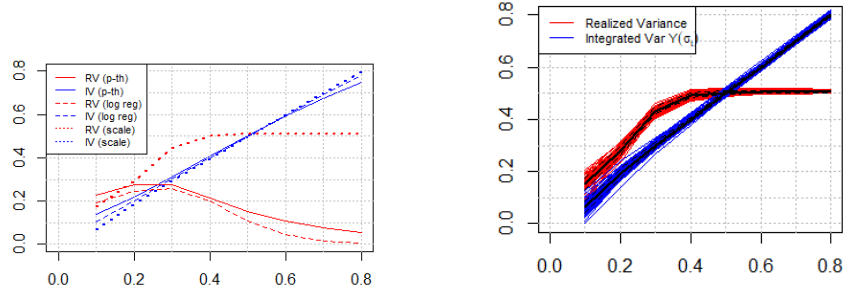


Figure 25: Estimated roughness indexes using instantaneous volatility (IV) and realized volatility/variance (RV) from a fractional OU model (19) with $T = 1$ plotted for price paths generated by different values of H . **Left:** Estimates for all three roughness estimators $\hat{H}_{L,K}$, \hat{H}_R and $|\hat{\mathcal{X}}_n^s|$. **Right:** Absolute scale estimates $|\hat{\mathcal{X}}_n^s|$ for 100 independent price paths. The solid black line represents the mean, and the dashed black lines represent a 75% confidence interval.

From Figure 25 we conclude that the absolute scale estimator systematically exhibits the same kind of behaviour. That is, when used for realized variance the estimator is not very precise especially when the true roughness exponent of the underlying model takes a high value. However, it does remedy some of the inaccuracies observed for the other two estimators, and it does not estimate the path to be rough when the underlying model is not rough. As this is a simulation study, the difference between the blue and red lines are solely caused by estimation errors.

In practice, instantaneous volatility cannot be observed, and roughness estimates are based on realized volatility. If using traditional roughness estimators such as $\hat{H}_{L,K}$ and \hat{H}_R , the above examples illustrate that one cannot necessarily draw the conclusion that (spot) volatility is rough just because realized volatility exhibit rough behaviour with $\hat{H}_{L,K}(RV) < \frac{1}{2}$ even if the estimated roughness is well below $\frac{1}{2}$.

Passing to the absolute sequential scale estimator $|\hat{\mathcal{X}}_n^s|$ which is developed with the ambition of accurately estimating roughness from realized variance does in these examples improve estimation from realized input. However, $|\hat{\mathcal{X}}_n^s|$ is also somewhat vulnerable to estimation errors and does not completely remove inaccuracies caused by estimation errors. Furthermore, $|\hat{\mathcal{X}}_n^s|$ requires observations of realized variance for very specific time points making it less applicable to real

time high frequency data.

4.4 RFSV model

We will now consider the stochastic volatility model suggested by [Gatheral et al. \[2018\]](#). The model is defined in the following way:

$$dS_t = S_t \sigma_t \sqrt{dt} U_n, \quad \sigma_t = e^{Y_t}, \quad dY_t = -\alpha(Y_t - m) dt + \nu dB_t^H, \quad (20)$$

for some $\nu > 0$, $\alpha > 0$, $m \in \mathbb{R}$ and $H < \frac{1}{2}$ where B_t^H is a fractional Brownian motion and U_n are iid standard Gaussian variables. [Gatheral et al. \[2018\]](#) further argues that choosing $\alpha \ll \frac{1}{T}$ is desirable since the log-volatility then locally behaves like a fBm while it is still technically stationary. In addition to this, [Gatheral et al. \[2018\]](#) argues that observed skew term structure in financial data can easier be replicated by choosing $\alpha \ll \frac{1}{T}$.

The specified RFSV model is similar to model (19) from section 4.3 and both models use a fractional Ornstein-Uhlenbeck process to model log-volatility. In this section we use a different approach to compute realized and instantaneous volatility. We follow much of the simulation approach from [Gatheral et al. \[2018\]](#). We set $T = 3601$ and simulate 5000 point per day. For every day we compute the realized volatility of the first 1000 points of the day, and we then rescale this volatility to be a daily volatility estimate. The corresponding instantaneous volatility is computed as the mean of the first 1000 points per day. Following this approach to compute volatility data, we perform many of the same analyses as in section 4.3. When simulating from the model we use the parameters $\nu = 0.2$, $\alpha = 10^{-4}$, $m = -7$ and we set $Y_0 = m$ and $S_0 = 1$.

In Table 8 we have computed roughness estimates from the RSFV model (20) with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

H	Instantaneous vol	Realized vol	IV (log reg)	RV (log reg)
0.1	0.175	0.168	0.146	0.143
0.2	0.241	0.236	0.217	0.214
0.3	0.314	0.310	0.297	0.295
0.4	0.394	0.390	0.386	0.384
0.5	0.475	0.472	0.478	0.476

Table 8: Estimated roughness via p -th variation $\hat{H}_{L,K}$ with $L = 60 \times 60$ and $K = 60$ and via log regression for instantaneous and realized volatility for simulated data from the RSFV model (20) with Hurst exponent $H = \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

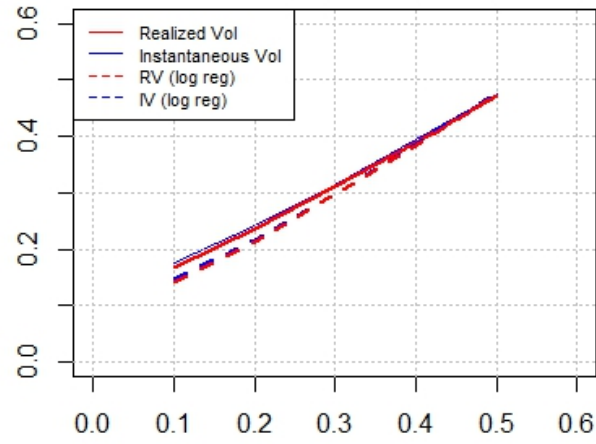


Figure 26: Estimated roughness indexes for realized and instantaneous volatility from a RSFV model (19) plotted for price paths generated by different values of H . The solid lines are roughness indexes via p -th variation $\hat{H}_{L,K}$ with $L = 3600$ and $K = 60$ while the dashed lines represents estimated roughness by logarithmic regression.

5 Conclusion

In this thesis we have investigated roughness of volatility processes in financial time series and taken a closer look at the claim 'volatility is rough'. More precisely, we have investigated the robustness of roughness estimates based on realized volatility for three different roughness estimators. Most roughness estimators in literature rely on approximating true spot volatility σ_t by realized volatility and then estimating the roughness of these proxy values $\hat{\sigma}_t$. We introduce two roughness estimators based on this approach, the roughness estimator via normalized p -th variation statistics $\hat{H}_{L,K}$ and the roughness estimator by logarithmic regression \hat{H}_R . We have in this thesis shown that such estimates based on realized volatility can be unreliable. We have conducted numerical experiments, and as described in section 4.1 and 4.2 we find that for two stochastic volatility diffusion models driven by a Brownian motion, realized volatility exhibit rough behaviour with an estimated roughness index significantly smaller than the true roughness index $H = \frac{1}{2}$. For our two stochastic diffusive examples we estimate the roughness of realized volatility in the range $\hat{H} \approx 0.3$ and $\hat{H} \approx 0.15$ respectively. However, when estimating directly from instantaneous volatility we obtain accurate results very close to the true roughness of the underlying model $H = \frac{1}{2}$. The two estimators $\hat{H}_{L,K}$ and \hat{H}_R widely coincides, and generate very similar results in all cases. As these are simulation examples, this apparent rough behaviour of realized volatility is entirely caused by the difference between realized volatility and instantaneous volatility also known as the estimation error. These results suggest that it cannot necessarily be concluded that observed rough behaviour in realized volatility is an indicator of similar behaviour in true spot volatility.

Moreover, we show in section 4.3 and 4.4 that for volatility processes driven by a fractional Ornstein-Uhlenbeck model, realized volatility always exhibit rough behaviour (i.e. $\hat{H} < \frac{1}{2}$). For both of these examples even when the true underlying model has diffusive or smooth behaviour (i.e. $H \geq \frac{1}{2}$) the roughness estimates based on realized volatility always stay below 0.3. If we estimate roughness directly from instantaneous volatility we obtain accurate estimates, and this again indicates that the estimation error can substantially distort the outcome of the roughness estimation. More precisely, these results show that stochastic diffusive models or even smooth volatility processes are perfectly compatible with rough behaviour observed in realized volatility. The conclusions from 'rough volatility' literature such as Gatheral et al. [2018] stating that 'volatility is rough' because

In this thesis we also introduce a third roughness estimator, the sequential scale estimator $\hat{\mathcal{H}}_n^s$ first introduced by Han and Schied [2023]. This estimator estimates the roughness of a path from discrete observations of integrated variance or realized variance and is developed with the purpose of minimizing the possible distortion caused by estimation errors. We find that by modifying the estimator

into the absolute sequential scale estimator $|\hat{\mathcal{R}}_n^s|$ we do obtain more accurate estimates from realized data compared to our other two roughness estimators. For our two stochastic diffusive models in section 4.1 and 4.2, $|\hat{\mathcal{R}}_n^s|$ performs very well when using realized variance as input, and it accurately estimates the true roughness of the underlying model $H = \frac{1}{2}$. Thus, in these two examples $|\hat{\mathcal{R}}_n^s|$ completely remedies the apparent rough behaviour of realized volatility. However, for our models driven by a fractional Ornstein-Uhlenbeck model in section 4.3 and 4.4, the absolute sequential scale estimator does not perform as well as before. Here we find that $|\hat{\mathcal{R}}_n^s|$ tends to always estimate the roughness of realized variance near 0.5 when the underlying model is actually smooth (i.e. $H > \frac{1}{2}$). When the underlying model is rough (i.e. $H < \frac{1}{2}$) $|\hat{\mathcal{R}}_n^s|$ tends to overestimate the true roughness such that it in all cases estimates the path to be nearer to diffusive behaviour (i.e. $H = \frac{1}{2}$) than the true model actually is. When we instead of realized variance use integrated variance computed directly from instantaneous volatility σ_t as input, the sequential scale estimator performs very well, and does in all cases estimate the true roughness of the model accurately. Thus, $|\hat{\mathcal{R}}_n^s|$ is also vulnerable to estimation errors where estimation error in this case means the difference between realized variance and integrated variance.

Comparing our three roughness estimators we find that they all perform very well in our examples when based on instantaneous volatility. $\hat{H}_{L,K}$ and $|\hat{\mathcal{R}}_n^s|$ work in a model-free setting and are therefore very flexible and applicable to many different models. Estimators like \hat{H}_R only works well if the class of models is well-specified. However, \hat{H}_R seems to converge faster, and it requires less data input for the estimate to be accurate.

For realized volatility/variance $|\hat{\mathcal{R}}_n^s|$ does in some cases perform much better than the other two estimators, but it does not estimate the roughness of the underlying model accurately in all cases. Furthermore, $|\hat{\mathcal{R}}_n^s|$ has much stronger requirements to the input data, and it needs observations at very specific time points. That makes the estimator less applicable to real world high-frequency data where the discrete grid of observation times might not be a uniform partition sequence.

For future investigation it could be relevant to investigate how the absolute sequential scale estimator $|\hat{\mathcal{R}}_n^s|$ could be further improved. The estimation error for this estimator is as shown very small, and if the estimator could fully handle this estimation error it has the potential to perform reliable roughness estimates from realized data. Another thing that could be further investigated is the estimation error itself. If it turns out that the estimation error is insignificant when using a certain approach to calculate it or a certain set of models, it would shed light on when to expect reliable roughness estimates. Both of these things could be a helping tool in determining whether volatility is rough or not.

6 Appendix

6.1 Simulation of fractional Brownian motion

Throughout this thesis we will be simulating from the fractional Brownian motion by using a spectral method which can be used for stationary processes. The idea is to analyse the stochastic process in the spectral domain rather than the time domain. The spectral density is computed as follows for frequencies $-\pi \leq \lambda \leq \pi$:

$$f(\lambda) := \sum_{j=-\infty}^{\infty} \gamma(j) \exp(ij\lambda) \quad (21)$$

where $\gamma(\cdot)$ represent the autocovariance function. It can be shown that the spectral density of fractional Gaussian noise is given by

$$f(\lambda) = 2 \sin(\pi H) \Gamma(2H + 1) (1 - \cos \lambda) [|\lambda|^{-2H-1} + B(\lambda, H)] \quad (22)$$

where $\Gamma(\cdot)$ denotes the Gamma function and

$$B(\lambda, H) := \sum_{j=1}^{\infty} ((2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1}) \quad (23)$$

for $-\pi \leq \lambda \leq \pi$. The infinite sum makes direct numerical evaluation almost impossible. However, a useful approximation of the sum is made by [Paxson \[1997\]](#). They show that by using

$$\tilde{B}_3(\lambda, H) := \sum_{j=1}^3 \left((a_j^+)^{-2H-1} + (a_j^-)^{-2H-1} \right) + \frac{(a_3^+)^{-2H} + (a_3^-)^{-2H} + (a_4^+)^{-2H} + (a_4^-)^{-2H}}{8H\pi}$$

where $a_j^{\pm} = 2\pi j \pm \lambda$, $f(\lambda)$ is approximated quite well.

Now, consider a stationary Gaussian discrete-time process $X = \{X_n : n = 0, \dots, N-1\}$ where N is the required sample size. The spectral theorem states that it can be represented in terms of the spectral density as

$$X_n \stackrel{d}{=} \int_0^{\pi} \sqrt{\frac{f(\lambda)}{\pi}} \cos(n\lambda) dB_1(\lambda) - \int_0^{\pi} \sqrt{\frac{f(\lambda)}{\pi}} \sin(n\lambda) dB_2(\lambda) \quad (24)$$

where B_1 and B_2 are two mutually independent Brownian motions [Dieker and Mandjes \[2003\]](#). We wish to approximate (8). The integrand is replaced by a simpler function. Fix some integer l and set $t_k = \pi k/l$ for $k = 0, \dots, l$. Now, define a simple function $\xi_n^{(l)}$ on $[0, \pi]$ for $0 \leq n \leq N-1$ by

$$\xi_n^{(l)}(\lambda) = \sqrt{\frac{f(t_1)}{\pi}} \cos(nt_1) \mathbf{1}_{\{0\}}(\lambda) + \sum_{k=0}^{l-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) \mathbf{1}_{(t_k, t_{k+1}]}(\lambda). \quad (25)$$

The first integral in (8) can be approximated by $\int_0^\pi \xi_n^{(\ell)}(\lambda) dB_1(\lambda)$. Since $\xi_n^{(\ell)}$ is a simple function the stochastic integral can be computed as

$$\int_0^\pi \xi_n^{(\ell)}(\lambda) dB_1(\lambda) = \sum_{j=0}^{\ell-1} \xi_n^{(\ell)}(B(t_{j+1}) - B(t_j)). \quad (26)$$

Thus, we obtain

$$\begin{aligned} \int_0^\pi \xi_n^{(\ell)}(\lambda) dB_1(\lambda) &= \sum_{j=0}^{\ell-1} \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) \mathbf{1}_{(t_k, t_{k+1}]}(B_1(t_{j+1}) - B_1(t_j)) \\ &= \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) (B_1(t_{k+1}) - B_1(t_k)) \\ &= \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) U_k^{(0)} \sqrt{\frac{\pi}{\ell}} = \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\ell}} \cos(nt_{k+1}) U_k^{(0)} \end{aligned}$$

where $U_k^{(0)}$ is an i.i.d. standard normal random variable for $k = 0, \dots, \ell-1$. The $U_k^{(0)} \sqrt{\pi/\ell}$ represent the Brownian motion increments which per definition are normally distributed with mean 0 and variance $t_{k+1} - t_k$.

The second integral in (8) can be approximated in a similar way by replacing the cosine terms with sine terms. Thus, we obtain the following approximation of X_n :

$$\hat{X}_n^{(\ell)} := \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\ell}} \left(\cos(nt_{k+1}) U_k^{(0)} - \sin(nt_{k+1}) U_k^{(1)} \right). \quad (27)$$

The two vectors $U^{(0)}$ and $U^{(1)}$ are mutually independent since B_1 and B_2 are independent as well. In order to calculate $\hat{X}_n^{(\ell)}$ efficiently we will be using the fast Fourier transform (FFT). To this end, we define the sequence $(a_k)_{k=0, \dots, 2\ell-1}$ by

$$a_k := \begin{cases} 0 & k = 0; \\ \frac{1}{2} \left(U_{k-1}^{(0)} + i U_{k-1}^{(1)} \right) \sqrt{\frac{f(t_k)}{\ell}} & k = 1, \dots, \ell-1; \\ U_{k-1}^{(0)} \sqrt{\frac{f(t_k)}{\ell}} & k = \ell; \\ \frac{1}{2} \left(U_{2\ell-k-1}^{(0)} - i U_{2\ell-k-1}^{(1)} \right) \sqrt{\frac{f(t_{2\ell-k})}{\ell}} & k = \ell+1, \dots, 2\ell-1. \end{cases}$$

It is shown in Appendix 6.1.1 that the Fourier transform of a_k is indeed real and equals $\hat{X}_n^{(\ell)}$. From this approximated fractional Gaussian noise we can generate the fBm.

Dieker and Mandjes [2003] shows that the finite-dimensional distributions of $\hat{X}^{(\ell)}$ converge in probability to the corresponding finite-dimensional distributions of X as $\ell \rightarrow \infty$. The rate of convergence is, however, quite slow. Therefore, we will be using a $\ell \geq 30000$ when simulating from the fBm by the spectral method in this thesis in order to make sure that our simulations are reliable. Figure 27 shows three simulations of fractional Brownian motions with different Hurst parameters generated by the spectral method.

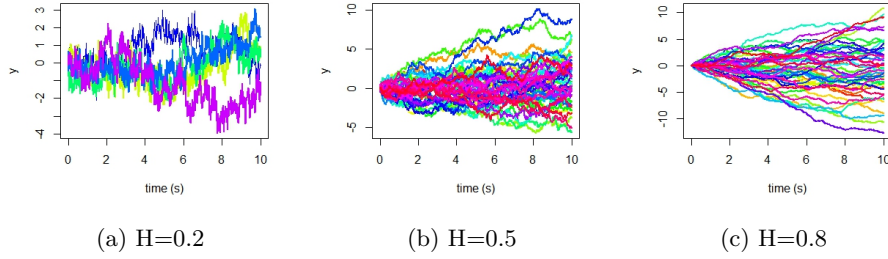


Figure 27: Fractional Brownian motions generated by the spectral method.

6.1.1 Proof of Fourier transform in spectral method

Consider

$$a_k := \begin{cases} 0 & k = 0; \\ \frac{1}{2} \left(U_{k-1}^{(0)} + iU_{k-1}^{(1)} \right) \sqrt{\frac{f(t_k)}{\ell}} & k = 1, \dots, \ell - 1; \\ U_{k-1}^{(0)} \sqrt{\frac{f(t_k)}{\ell}} & k = \ell; \\ \frac{1}{2} \left(U_{2\ell-k-1}^{(0)} - iU_{2\ell-k-1}^{(1)} \right) \sqrt{\frac{f(t_{2\ell-k})}{\ell}} & k = \ell + 1, \dots, 2\ell - 1. \end{cases}$$

We want to take the Fourier transform of this sequence. The Fourier transform of $(a_k)_{k=0}^{2\ell-1}$ is $\lambda_n = \sum_{k=0}^{2\ell-1} a_k \exp(2\pi i \frac{nk}{2\ell})$. By Euler's formula we obtain

$$\lambda_n = \sum_{k=0}^{2\ell-1} a_k \left(\cos \left(\pi \frac{nk}{\ell} \right) + i \sin \left(\pi \frac{nk}{\ell} \right) \right).$$

We will now insert a_k and split the sum into the four different cases of a_k . Hence, we obtain

$$\begin{aligned}\lambda_n &= 0 + \sum_{k=1}^{l-1} \frac{1}{2} \left(U_{k-1}^{(0)} + iU_{k-1}^{(1)} \right) \sqrt{\frac{f(t_k)}{\ell}} \left(\cos\left(\pi \frac{nk}{\ell}\right) + i \sin\left(\pi \frac{nk}{\ell}\right) \right) \\ &+ U_{\ell-1}^{(0)} \sqrt{\frac{f(t_\ell)}{\ell}} \left(\cos\left(\pi \frac{n\ell}{\ell}\right) + i \sin\left(\pi \frac{n\ell}{\ell}\right) \right) \\ &+ \sum_{k=l+1}^{2l-1} \frac{1}{2} \left(U_{2\ell-k-1}^{(0)} - iU_{2\ell-k-1}^{(1)} \right) \sqrt{\frac{f(t_{2\ell-k})}{\ell}} \left(\cos\left(\pi \frac{nk}{\ell}\right) + i \sin\left(\pi \frac{nk}{\ell}\right) \right).\end{aligned}$$

By definition $t_k = \frac{\pi k}{\ell}$. Thus, what is inside the cosine and sine functions is simply nt_k . Note that for $k = \ell$ we have $t_k = \pi$. Thus, $\sin(nt_\ell) = 0$ since n is an integer. By inserting this and changing the indexes of the two sums, we obtain

$$\begin{aligned}\lambda_n &= U_{\ell-1}^{(0)} \sqrt{\frac{f(t_\ell)}{\ell}} \cos(nt_\ell) \\ &+ \sum_{k=0}^{l-2} \frac{1}{2} \left(U_k^{(0)} + iU_k^{(1)} \right) \sqrt{\frac{f(t_{k+1})}{\ell}} (\cos(nt_{k+1}) + i \sin(nt_{k+1})) \\ &+ \sum_{k=0}^{l-2} \frac{1}{2} \left(U_{\ell-k-2}^{(0)} - iU_{\ell-k-2}^{(1)} \right) \sqrt{\frac{f(t_{\ell-k-1})}{\ell}} (\cos(nt_{\ell-k-1}) + i \sin(nt_{\ell-k-1})).\end{aligned}$$

Now note that $\sin(nt_{\ell-k-1}) = -\sin(nt_{\ell-(k+1)})$ since $t_\ell = \pi$. Similarly, $\cos(nt_{\ell-k-1}) = \cos(nt_{\ell-(k+1)})$. By inserting that in the last sum in the above expression, the sum becomes

$$\begin{aligned}&\sum_{k=0}^{l-2} \frac{1}{2} \left(U_{\ell-k-2}^{(0)} - iU_{\ell-k-2}^{(1)} \right) \sqrt{\frac{f(t_{\ell-k-1})}{\ell}} (\cos(nt_{\ell-k-1}) - i \sin(nt_{\ell-k-1})) \\ &= \sum_{k=0}^{l-2} \frac{1}{2} \left(U_k^{(0)} - iU_k^{(1)} \right) \sqrt{\frac{f(t_{k+1})}{\ell}} (\cos(nt_{k+1}) - i \sin(nt_{k+1}))\end{aligned}$$

where we have reversed the sum order. Now we can combine the two sums into one. Several terms cancel out such that we obtain

$$\begin{aligned}\lambda_n &= U_{\ell-1}^{(0)} \sqrt{\frac{f(t_\ell)}{\ell}} \cos(nt_\ell) + \sqrt{\frac{f(t_{k+1})}{\ell}} \sum_{k=0}^{l-2} \left(U_k^{(0)} \cos(nt_{k+1}) - U_k^{(1)} \sin(nt_{k+1}) \right) \\ &= \sum_{k=0}^{l-1} \sqrt{\frac{f(t_{k+1})}{\ell}} \left(U_k^{(0)} \cos(nt_{k+1}) - U_k^{(1)} \sin(nt_{k+1}) \right)\end{aligned}$$

which equals the definition of $\hat{X}_n^{(\ell)}$ from (27) as desired.

References

- Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.
- F Biagini. Stochastic calculus for fractional brownian motion and applications. *Probability and its Applications (New York)/SpringerVerlag London, Ltd*, 2008.
- Anine E Bolko, Kim Christensen, Mikko S Pakkanen, and Bezirgen Veliyev. A gmm approach to estimate the roughness of stochastic volatility. *Journal of Econometrics*, 235(2):745–778, 2023.
- Tim Bollerslev and Hans Ole Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of econometrics*, 73(1):151–184, 1996.
- Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.
- Rama Cont and Purba Das. Rough volatility: fact or artefact? *Sankhya B*, 86(1):191–223, 2024.
- Antonius Bernardus Dieker and Michael Mandjes. On spectral simulation of fractional brownian motion. *Probability in the Engineering and Informational Sciences*, 17(3):417–434, 2003.
- Omar El Euch, Jim Gatheral, and Mathieu Rosenbaum. Roughening heston. *Risk*, pages 84–89, 2019.
- Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance*, 32(4):1086–1132, 2022.
- Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative finance*, 18(6):933–949, 2018.
- Xiyue Han and Alexander Schied. The hurst roughness exponent and its model-free estimation, 2022.
- Xiyue Han and Alexander Schied. Estimating the roughness exponent of stochastic volatility from discrete observations of the realized variance. *arXiv preprint arXiv:2307.02582*, 2023.
- Jean Jacod. *Discretization of processes*. Springer, 2012.
- Vern Paxson. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*, 27(5):5–18, 1997.