

# Multiscale Volatility Analysis for Noisy High-Frequency Prices

Tim Leung\* and Theodore Zhao

Applied Mathematics Department, University of Washington, Seattle, WA 98195, USA; zdzhao16@uw.edu

\* Correspondence: timleung@uw.edu

**Abstract:** We present a multiscale analysis of the volatility of intraday prices from high-frequency data. Our multiscale framework includes a fractional Brownian motion and microstructure noise as the building blocks. The proposed noisy fractional Brownian motion model is shown to possess a variety of volatility behaviors suitable for intraday price processes. Algorithms for numerical estimation from time series observations are then introduced, with a new Hurst exponent estimator proposed for the noisy fractional Brownian motion model. Using real-world high-frequency price data for a collection of U.S. stocks and ETFs, we estimate the parameters in the noisy fractional Brownian motion and illustrate how the volatility varies over different timescales. The Hurst exponent and noise level also exhibit an intraday pattern whereby the noise ratio tends to be higher near market close.

**Keywords:** multiscale volatility; high-frequency data; fractional Brownian motion; noisy data

## 1. Introduction

The study of multiscale properties and scaling laws has long been a topic in finance. One of the key questions is: how does the distribution of returns behave at different timescales? Various stochastic models and statistical methods have been proposed to examine this problem using financial data (see [Capobianco \(2004\)](#); [Di Matteo \(2007\)](#); and [Leung and Zhao \(2022\)](#), among others). While many multiscale studies have relied on daily data, the increasing availability of high-frequency data calls for more sophisticated models and statistical methods (see [Müller et al. \(1990\)](#); [Fan and Wang \(2007\)](#); and [Mensi et al. \(2021\)](#)), especially for volatility analysis, as studied by [Andersen et al. \(2003\)](#) and [Fan and Wang \(2007\)](#). Volatility and correlation properties at middle or high frequency are crucial components for multiscale analysis.

As studied by [Osborne \(1959\)](#), the Brownian motion stock price model intrinsically determines return distribution at any timescale due to the independent increments and statistical properties of the model. The random walk model's empirical validity, especially its agreement with the efficient market hypothesis (EMH), is advocated by [Fama \(1965\)](#). This standard model, however, has been questioned and modified afterward, starting with the work of [Mandelbrot \(1963\)](#). The generalization to fractional Brownian motion by [Mandelbrot and Van Ness \(1968\)](#) leads to a much larger class of stochastic processes, allowing for new scaling properties and a long-memory process. At the same time, it also inherently rejects EMH and gives room for arbitrage opportunities, as shown by [Rogers \(1997\)](#).

As price dynamics and volatility modeling become more complex when observing data at a higher frequency, researchers have turned to empirical and non-parametric approaches to study the statistical properties of returns, including the measuring of realized volatility, as studied by [Andersen et al. \(2003\)](#); [Barndorff-Nielsen and Shephard \(2002\)](#). Multiple studies (see e.g., [Ait-Sahalia et al. \(2005\)](#); [Bandi and Russell \(2008\)](#)) have shown that it is not optimal to use all the data directly at the highest available frequency due to the presence of market microstructure noise. Different techniques have been designed for an accurate measurement of realized volatility for the "true price", e.g., the work by [Podolskij and Vetter \(2009\)](#). One of the key ideas was to measure realized volatility at different timescales



**Citation:** Leung, Tim, and Theodore Zhao. 2023. Multiscale Volatility Analysis for Noisy High-Frequency Prices. *Risks* 11: 117. <https://doi.org/10.3390/risks11070117>

Academic Editor: Mogens Steffensen

Received: 31 May 2023

Revised: 20 June 2023

Accepted: 20 June 2023

Published: 26 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

and combine the results to discover the latent value (see Zhang et al. (2005)). This multiscale technique was generalized to a wider class of models and other return statistics in the work by Aït-Sahalia et al. (2010); Bibinger et al. (2014); Fan and Wang (2007).

In this paper, we define functions to describe multiscale volatility in asset prices and investigated their behaviors in common classes of financial models. We then propose models for intraday high-frequency prices with their multiscale properties. Numerical estimation of model parameters is then proposed, including a modified algorithm to evaluate the Hurst exponent. Introduced by Hurst (1951), the Hurst exponent describes the dependency structure of the time series and how its variability depends on the timescale. The value of the Hurst exponent is associated with the smoothness of the random process, long memory, and fractal dimension (see Mandelbrot and Hudson (2007)), and therefore is widely used in finance (see Granero et al. (2008); Matos et al. (2008); Tzouras et al. (2015), among others). We analyze a modified Hurst exponent under microstructure noise and present a new method to estimate it using real-world high-frequency price data. The corresponding multiscale volatility behavior is examined and illustrated.

The fractional Brownian motion model and multiscale volatility estimator are typically used separately for different purposes. Interestingly, they share some very similar forms for showing multiscale behaviors. This similarity inspires us to formally propose definitions to describe the multiscale properties of the price process and explore a wider class of models for high-frequency financial time series. We also find that the market microstructure noise can potentially bias the estimation of the scaling parameters. In a related recent study, Cont and Das (2022) show that one form of microstructure noise due to the discrepancy between the realized and instantaneous volatility can lead to bias in the Hurst exponent estimation. The necessity of new multiscale models is also justified by real-world observations in high-frequency data.

The rest of the paper is structured as follows. In Sections 2 and 3 we discuss our multiscale approach and the stochastic model for noisy price data. In Section 4, a series of numerical experiments are presented to illustrate the multiscale volatility in asset prices. In Section 5, we apply our multiscale model to real-world high-frequency data and illustrate the empirical multiscale volatility and Hurst exponent. Concluding remarks are provided in Section 6.

## 2. Multiscale Variance

The price process of a financial asset, denoted by  $(P_t)_{t \geq 0}$ , is observed over time. The log price at time  $t$  is denoted by  $X_t := \log(P_t)$ . In turn, the log return from time  $t$  to  $t + \tau$  is defined by

$$r_{t,\tau}^X = X_{t+\tau} - X_t. \quad (1)$$

For any given  $\tau$ ,  $(r_{t,\tau})_{t \geq 0}$  is a time series. If one assumes stationary increments, the distribution of  $r_{t,\tau}$  depends on  $\tau$  only, which is the main focus of this study. As a key feature of the return distribution, the scaling properties of variance are of great interest. To that end, we introduce the following definitions.

**Definition 1.** Let the log price process  $(X_t)_{0 \leq t \leq T}$  be a stochastic process with stationary increments on  $[0, T]$ . Define the following multiscale functions of  $\tau \in \mathbb{R}^+$ :

- *Multiscale variance:*

$$\mathcal{V}[X](\tau) := \mathbb{V}[r_{t,\tau}^X] = \mathbb{E}[(r_{t,\tau}^X)^2] - (\mathbb{E}[r_{t,\tau}^X])^2, \quad (2)$$

- *Multiscale volatility:*

$$\mathcal{S}[X](\tau) := \sqrt{\frac{\mathcal{V}[X](\tau)}{\tau}}. \quad (3)$$

- Growth rate:

$$\Lambda[X](\tau) := \frac{1}{2} \frac{d(\log \mathcal{V}[X](\tau))}{d(\log \tau)}. \tag{4}$$

The functions  $\mathcal{V}[X]$  and  $\mathcal{S}[X]$  describe the variance and volatility, respectively, of the time series at different timescales. The growth rate,  $\Lambda$ , is a unitless operator measuring the speed of risk accumulation on different scales. Next, we summarize some useful properties of multiscale variance.

**Property 1.** Let  $X_t^{(1)}, \dots, X_t^{(n)}$  be a sequence of independent processes, then the multiscale variance is linearly additive:

$$\mathcal{V} \left[ \sum_{i=1}^n X^{(i)} \right] (\tau) = \sum_{i=1}^n \mathcal{V}[X^{(i)}](\tau). \tag{5}$$

In general, the multiscale functions above may take on many forms. However, for a wide class of processes, the multiscale functions have certain useful properties.

**Property 2.** Suppose  $X_t$  has stationary increments that are independent, that is,

$$\mathbb{E}[(X_{t_2} - X_{t_1})(X_{t_3} - X_{t_2})] = \mathbb{E}[X_{t_2} - X_{t_1}] \cdot \mathbb{E}[X_{t_3} - X_{t_2}], \quad \forall t_1 < t_2 < t_3.$$

Then, Property 1 and Definition 1 imply that the multiscale functions must satisfy

$$\mathcal{V}[X](\tau) \propto \tau, \tag{6}$$

$$\mathcal{S}[X]'(\tau) = 0, \tag{7}$$

$$\Lambda[X](\tau) = \frac{1}{2}. \tag{8}$$

### 3. High-Frequency Multiscale Models

Scaling behavior is directed by the dynamics of the underlying stochastic process. In this section, we examine the multiscale volatility of high-frequency price data. We first review the classic fractional Brownian motion model introduced by Mandelbrot and Van Ness (1968), which is widely used in finance to model long-range dependency. In our study, we incorporate market microstructure noise as a new stochastic component into the fractional Brownian motion model for high-frequency time series. This combination gives rise to some new scaling behaviors.

#### 3.1. Fractional Brownian Motion

First, we recall the definition of the fractional Brownian motion (fBm) model according to Mandelbrot and Van Ness (1968).

**Definition 2.** A fractional Brownian motion (fBm)  $B_t^H, t \geq 0$ , is a continuous-time Gaussian process satisfying that  $\mathbb{E}[B_t^H] = 0, \forall t \geq 0$ , and giving the following covariance function:

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

where  $t, s \geq 0$ , and  $H \in (0, 1)$  is called the Hurst exponent.

When  $H = 1/2$ , fBm reduces to the standard Brownian motion with independent increments. When  $H \in (0, 1/2)$ , the process is anti-persistent or mean-reverting with negatively correlated increments. When  $H \in (1/2, 1)$ , the process is persistent or trending with positively correlated increments.

**Property 3.** A fractional Brownian motion satisfies the self-similar property:

$$B_{\lambda t}^H \stackrel{\text{dist.}}{=} \lambda^H B_t^H, \quad \forall t \geq 0. \quad (9)$$

with the Hurst exponent serving as the scaling exponent,  $H$ .

Following Definition 1, a fractional Brownian motion with Hurst exponent,  $H$ , has the following multiscale functions:

$$\mathcal{V}[B^H](\tau) = \tau^{2H}, \quad \mathcal{S}[B^H](\tau) = \tau^{H-1/2}, \quad \Lambda[B^H](\tau) = H. \quad (10)$$

When  $H > 1/2$ , the functions reduce to the forms given by Proposition 2. The volatility decreases with the scale if  $H < 1/2$  and increases with  $\tau$  if  $H > 1/2$ . Note also that the growth rate,  $\Lambda$ , is a constant for any self-similar process, which can serve as a useful indicator.

### 3.2. Incorporating Microstructure Noise

For intraday high-frequency financial time series, there have been a variety of studies on the market microstructure noise effect. The market microstructure noise has been recognized to play an important role in finance, with earlier research dating back to Black (1986), and recent studies on various types of assets such as cryptocurrency (see Bouri et al. (2022)). At any time point, the asset price is the result from trading activity. Various frictions in the trading process (e.g., bid–ask spread, the discreteness of price, etc.) may lead to “noise” in the price process, as shown by Jacod et al. (2017). Several models have been proposed to formulate the market microstructure noise problem in different ways.

To avoid any confusion, we use the same definition of market microstructure noise as in the work by Ait-Sahalia et al. (2005); Podolskij and Vetter (2009); Zhang et al. (2005). In most of these studies, the aim is to improve realized volatility estimation, and the noise is regarded as a random deviation from the unobserved true price. However, a price observed is the price that was traded at, thus even the “noisy” price is financially meaningful. Therefore, in our study herein, our objective is not to simply remove noise for better estimation of other quantities. Instead, we seek to understand the critical role of noise on price dynamics and its impact on multiscale behaviors.

**Definition 3.** Let  $(X_t)_{t \geq 0}$  be some real-valued latent stochastic process. The noisy price process  $(P_t)_{t \geq 0}$  is defined by

$$P_t = \exp(Y_t),$$

where

$$Y_t = X_t + \epsilon_t,$$

and  $\epsilon$  is the microstructure noise. Here, the two processes  $\epsilon$  and  $X$  are independent. In addition,  $\mathbb{E}[\epsilon_t] = 0$ ,  $\mathbb{V}[\epsilon_t] = \sigma_\epsilon^2$ , and  $\epsilon_t$  are i.i.d. for  $\forall t \geq 0$ .

**Remark 1.** The noise can be viewed as an effect acting on the price whenever there is a trade. Unfortunately, the noisy price process defined in this way will be nowhere continuous. However, from the market microstructure viewpoint, the price is not necessarily a continuous-time process, as the trades occur at discrete time points. This is the reason that we should always keep  $\tau > 0$ , though a detailed discussion is beyond the scope of this paper.

Directly applying Property 1 and using the fact that

$$\mathcal{V}[\epsilon](\tau) = \mathbb{V}[\epsilon_{t+\tau} - \epsilon_t] = 2\sigma_\epsilon^2, \quad \forall \tau > 0,$$

we can derive the multiscale functions in Definition 1 for the noisy price process:

$$\mathcal{V}[Y](\tau) = \mathcal{V}[X](\tau) + 2\sigma_\epsilon^2, \quad (11)$$

$$\mathcal{S}[Y](\tau) = \sqrt{\mathcal{S}^2[X](\tau) + \frac{2\sigma_\epsilon^2}{\tau}}. \quad (12)$$

Note that the volatility  $\mathcal{S}[Y](\tau) \rightarrow \infty$  as  $\tau \rightarrow 0^+$ , for any  $\forall \sigma_\epsilon > 0$ . The undefined volatility at  $\tau = 0$  is the result of the non-continuity of the noisy price path. Unlike the sensitivity of volatility to the noise, the following proposition shows a conservation law for the variance and the growth rate.

**Proposition 1.** Define the growth-adjusted variance operator

$$\tilde{\mathcal{V}} := \mathcal{V} \cdot \Lambda \quad (13)$$

Then, for any noisy process,  $Y_t = X_t + \epsilon_t$ , as defined in Definition 3,

$$\tilde{\mathcal{V}}[Y] \equiv \tilde{\mathcal{V}}[X]. \quad (14)$$

**Proof.** Using the variance Equation (11), we can write the growth rate associated with the noisy process  $Y$  as

$$\begin{aligned} \Lambda[Y](\tau) &= \frac{1}{2} \frac{d}{d \log(\tau)} \log(\mathcal{V}[X](\tau) + 2\sigma_\epsilon^2) \\ &= \frac{1}{2(\mathcal{V}[X](\tau) + 2\sigma_\epsilon^2)} \frac{d}{d \log(\tau)} (\mathcal{V}[X](\tau) + 2\sigma_\epsilon^2) \\ &= \frac{1}{2\mathcal{V}[Y](\tau)} \frac{d\mathcal{V}[X](\tau)}{d \log(\tau)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \tilde{\mathcal{V}}[Y](\tau) &= \mathcal{V}[Y](\tau) \cdot \Lambda[Y](\tau) \\ &= \mathcal{V}[Y](\tau) \cdot \frac{1}{2\mathcal{V}[Y](\tau)} \frac{d\mathcal{V}[X](\tau)}{d \log(\tau)} = \frac{1}{2} \frac{d\mathcal{V}[X](\tau)}{d \log(\tau)}. \end{aligned}$$

On the other hand, the multiscale variance for  $X$  is given by

$$\begin{aligned} \tilde{\mathcal{V}}[X](\tau) &= \mathcal{V}[X](\tau) \cdot \Lambda[X](\tau) \\ &= \mathcal{V}[X](\tau) \cdot \frac{1}{2} \frac{d \log(\mathcal{V}[X](\tau))}{d \log(\tau)} \\ &= \mathcal{V}[X](\tau) \cdot \frac{1}{2\mathcal{V}[X](\tau)} \frac{d\mathcal{V}[X](\tau)}{d \log(\tau)} \\ &= \tilde{\mathcal{V}}[Y](\tau). \end{aligned}$$

□

### 3.3. Noisy Fractional Brownian Motion

In order to describe the scaling property of high-frequency price dynamics under microstructure noise, we propose the noisy fractional Brownian motion defined as follows:

**Definition 4.** For the Hurst exponent,  $H \in (0, 1)$ , drift,  $\mu \in \mathbb{R}$ , volatility parameter,  $\sigma \in \mathbb{R}^+$ , and initial value,  $Y_0 \in \mathbb{R}$ , the noisy fractional Brownian motion is defined as:

$$Y_t = Y_0 + \mu t + \sigma B_t^H + \epsilon_t, \quad (15)$$

where  $\epsilon_t$  is the microstructure noise with variance  $\mathbb{V}(\epsilon_t) = \sigma_\epsilon^2$ .

**Property 4.** The multiscale functions of the noisy fractional Brownian motion  $Y_t$  are

$$\mathcal{V}[Y](\tau) = \sigma^2(\tau^{2H} + 2\eta), \quad (16)$$

$$\mathcal{S}[Y](\tau) = \sigma \cdot \sqrt{\tau^{2H-1} + \frac{2\eta}{\tau}}, \quad (17)$$

$$\Lambda[Y](\tau) = H \cdot \frac{\tau^{2H}}{\tau^{2H} + 2\eta}, \quad (18)$$

where  $\eta := \sigma_\epsilon^2 / \sigma^2$  is the noise ratio.

To obtain the expressions for  $\mathcal{V}[Y](\tau)$  and  $\mathcal{S}[Y](\tau)$  in Property 4, we apply (10) to (11) and (12), respectively. In turn, a direct differentiation yields the formula for  $\Lambda[Y](\tau)$ .

**Remark 2.** The effect of the microstructure noise on the multiscale functions is only through the noise ratio,  $\eta$ . One should be careful that the ratio is not unitless. The reason is that while  $\sigma_\epsilon^2$  is the variance of the noise,  $\sigma^2$  is not the variance of the latent process itself, but contributes in the form of  $\sigma^2 \tau^{2H}$ . The unit of  $\eta$  is thus  $[\text{Time}]^{2H}$ , depending on  $H$  as well.

### 3.4. Asymptotic Behaviors

We now discuss the asymptotic behaviors of the multiscale functions for a noisy fBm. These functions and their behaviors will be useful for the empirical estimation presented in the next section.

First, the asymptotic behavior under a big scale limit is approximating the latent fractional Brownian motion itself because the microstructure noise effect dissipates at a long range. Therefore, we have

- As  $\tau \rightarrow \infty$ :

$$\begin{aligned} \mathcal{V}[Y](\tau) &= \sigma^2 \tau^{2H} + \mathcal{O}(1), \\ \mathcal{S}[Y](\tau) &= \sigma \tau^{H-1/2} + \mathcal{O}(\tau^{-1/2}), \\ \Lambda[Y](\tau) &\rightarrow H. \end{aligned}$$

Note that the big  $O$  terms are dominated by the asymptotic terms by rates controlled by  $H > 0$ . The larger the Hurst exponent is, the faster the convergence rate is.

At a very small scale, the microstructure noise dominates the process, making the path very rough. This results in the following limits:

- As  $\tau \rightarrow 0$ :

$$\begin{aligned} \mathcal{V}[Y](\tau) &= 2\sigma_\epsilon^2 + \mathcal{O}(\tau^{2H}), \\ \mathcal{S}[Y](\tau) &\rightarrow \infty, \\ \Lambda[Y](\tau) &\rightarrow 0. \end{aligned}$$

The variance does not vanish but converges to a constant determined by the microstructure noise level. As discussed before, the volatility is not well-defined at zero. It is also worth noting that, even for a plain fractional Brownian motion without noise, the volatility is not well-defined at zero for any  $H \in (0, 1/2)$ . The variance growth is dominated by the noise at the very short scale, making the growth rate close to zero.

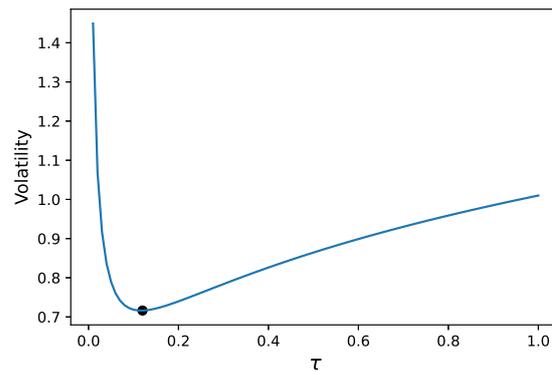
For the Hurst exponent,  $H \in (1/2, 1)$ , the volatility goes to infinity both at zero and at infinity. Interestingly, the multiscale volatility function is not monotonic and there exists

a critical timescale minimizing the volatility. As illustrated in Figure 1, the minimum is achieved at the scale

$$\tau^* = \left( \frac{\eta}{H - 1/2} \right)^{\frac{1}{2H}} \tag{19}$$

so that

$$\mathcal{S}[Y](\tau^*) = \sigma \cdot \sqrt{2H} \left( \frac{\eta}{H - 1/2} \right)^{\frac{H-1/2}{2H}}. \tag{20}$$



**Figure 1.** Volatility curve of a noisy fractional Brownian motion with  $H = 0.75, \sigma = 1, \eta = 0.01$ . The minimal volatility is marked with a black dot.

#### 4. Numerical Estimation

In reality, we are only able to observe a sequence of prices at discrete time points  $t_1, t_2, \dots, t_n$ . For simplicity, we assume that the time series is sampled over a uniform time grid. In other words, each sample path of price  $(P_1, P_2, \dots, P_n)$  is observed discretely at equal-spaced time points  $(\delta t, 2\delta t, \dots, n\delta t)$ .

##### 4.1. Multiscale Volatility

First, we convert the price observations to the log price time series

$$X_i = \log(P_i), \quad i = 1, \dots, n.$$

For  $m \in \mathbb{N}^+ < n$ , we can estimate the multiscale functions in Definition 1 at scale  $\tau = m\delta t$  as follows. Let  $\mathcal{P}_m \subset \{(i, j) | 1 \leq i, j \leq n, j - i = m\}$  be a subset of index pairs spanning a range of length  $m$ .

- Variance:

$$V_m = \frac{1}{|\mathcal{P}_m|} \sum_{(i,j) \in \mathcal{P}_m} (X_j - X_i - m\mu)^2. \tag{21}$$

where  $\mu = \frac{X_n - X_1}{n-1}$  is the estimated drift for detrending the process.

- Volatility:

$$S_m = \sqrt{\frac{V_m}{m\delta t}}. \tag{22}$$

After computing a sequence of  $V_m$  values, the growth rate can be estimated by

$$\Lambda_m = \frac{\log(V_{m+1}/V_m)}{\log(1 + 1/m)}. \tag{23}$$

The index pair set,  $\mathcal{P}$ , is made flexible. In practice, there are some common choices. The continuous rolling window  $\{(1, m + 1), (2, m + 2), \dots, (n - m, n)\}$  is a popular one, which forms the largest possible set. In some cases, people use non-overlapping windows  $\{(1, m + 1), (m + 1, 2m + 1), \dots, ((k - 1)m + 1, km + 1)\}$ .

We remark that the estimator of the variance in (21) is biased, even if the index pair set,  $\mathcal{P}$ , consists of non-overlapping windows. The reason is that the log returns on disjoint windows may be correlated. However, if the true value of the drift,  $\mu$ , is known, the variance estimator in (21) is unbiased. In practice, there are many ways to detrend the process, such as fitting via ordinary least squares (OLS), or using empirical mode decomposition (see [Leung and Zhao \(2021\)](#)). Nevertheless, even if we have an unbiased estimator of the variance,  $S_m$  is still biased because of the concavity in the square root operator.

#### 4.2. Hurst Exponent Estimation

The Hurst exponent is the key parameter in self-similar processes and the noisy fractional Brownian motion model. Over the past few decades, several estimation methods have been proposed, including the original rescaled range (R/S) analysis, as introduced by [Hurst \(1951\)](#), and the popular detrended fluctuation analysis (DFA) by [Peng et al. \(1994\)](#). [Barabási and Vicsek \(1991\)](#) propose a generalized Hurst exponent to estimate the scaling exponent for each order of the moment. Moment scaling is an important tool to understand the properties of time series, with a number of applications in finance (see [Di Matteo \(2007\)](#); [Górski et al. \(2002\)](#); [Mandelbrot \(2013\)](#)). It is important to note that no method is perfect; in fact, different methods can produce significantly different results (see [Weron \(2002\)](#)).

In this study, we discuss a modification of the generalized Hurst exponent algorithm. We first review the original rescaled range (R/S) method, as it is the standard when comparing different algorithms. After a review of the generalized Hurst exponent estimation, we discuss how the estimator can be biased under market microstructure noise and how to remove the bias.

##### 4.2.1. Rescaled Range Analysis

The original rescaled range (R/S) analysis algorithm using log prices is summarized below.

###### Rescaled Range

For time series observation  $X_1, X_2, \dots, X_n$ , take a consecutive sub-sequence:  $X_k, X_{k+1}, \dots, X_{k+m}$ , and apply the following operations:

- Estimate the drift,  $\mu = \frac{1}{m}(X_{k+m} - X_k)$ .
- Detrend the partial time series:  $Z_t = X_{k+t} - \mu t$ , for  $t = 1, 2, \dots, m$ .
- Compute the range,  $R$ , at scale  $m$ :

$$R(m) = \max(Z_1, Z_2, \dots, Z_m) - \min(Z_1, Z_2, \dots, Z_m).$$

- Compute the return standard deviation,  $S$ , at scale  $m$ :

$$S(m) = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_{k+i} - X_{k+i-1} - \mu)^2}.$$

- Estimate the rescaled range at length  $m$  as  $(R/S)_m = R(m)/S(m)$ , and average over all non-overlapping length- $m$  sub-sequences.

Following the same procedure, estimate the rescaled range at different values of  $m$ . The Hurst exponent,  $H$ , is then estimated by fitting  $(R/S)_m \sim m^H$ . A typical OLS is fitted on the log-log plot,  $\log(R/S)_m \sim \log m$ , taking the slope as the estimator.

The R/S algorithm, however, will produce a biased result when the time series is noisy. Here we briefly show it by considering how the range (R) and the standard deviation (S) are affected by the noise. Denote the noisy time series as  $\tilde{X}_t = X_t + \epsilon_t$ . It is easier to consider the return standard deviation first:

$$\tilde{S}(m) \approx \sqrt{S^2(m) + 2\text{Var}(\epsilon)} \approx S(m) + \frac{\text{Var}(\epsilon)}{S(m)}.$$

The return standard deviation will be larger. However, it is the scaling property of  $S(m)$  that is important. In that sense, because all  $S(m)$ s are computed from the finest timescale returns, their expectations can be assumed to be the same if the process is stationary. The noise just makes the same normalization factor smaller, and thus it will not affect the final scaling behavior from this part.

On the other hand, the range,  $R(m)$ , depends on  $m$  in most cases and will experience a change that is due to the noise.  $R(m)$  is the range of the detrended log price path over  $m$  consecutive intervals. Imagine a white noise is added to the path of length  $m$ , and new maximum and minimum values are obtained. The longer the length of  $m$  is, the higher is the chance for the maximal value to be higher and the minimal value to be lower. Therefore, the range,  $R(m)$ , will increase under the noise, and the level of increment depends on  $m$ , potentially biasing the slope estimation. A detailed analytical analysis of the amount and direction of bias would be overly complicated, so we only provide a numerical evaluation of the deviation in Table 1.

**Table 1.** Hurst exponent estimation on synthesized noisy fractional Brownian motion ( $\sigma = 1 \times 10^{-2}$ ,  $\sigma_\epsilon = 2 \times 10^{-5}$ ). The proposed growth-adjusted (GrAd) variance estimation algorithm is compared against the two standard algorithms, variance analysis and rescaled range analysis, under three different scenarios:  $H = 0.5$  (random walk),  $H = 0.4$  (mean-reverting), and  $H = 0.6$  (trending). For each scenario, 1000 sample paths are generated to evaluate the mean, confidence interval corresponding to plus/minus one standard deviation of the mean estimator, and mean absolute deviation from the ground truth value.

<b><math>H = 0.5</math> (Random Walk)</b>			
Algorithm	Mean	$\pm 1$ mean std	Mean abs. dev.
GrAd Variance	0.5003	(0.4998, 0.5009)	0.01474
Variance analysis	0.4846	(0.4843, 0.4848)	0.01558
Rescaled range	0.5302	(0.5294, 0.5311)	0.03410
<b><math>H = 0.4</math> (Mean-Reverting)</b>			
Algorithm	Mean	$\pm 1$ mean std	Mean abs. dev.
GrAd Variance	0.4057	(0.4050, 0.4064)	0.01799
Variance analysis	0.4192	(0.4189, 0.4194)	0.01923
Rescaled range	0.4518	(0.4510, 0.4526)	0.05202
<b><math>H = 0.6</math> (Trending)</b>			
Algorithm	Mean	$\pm 1$ mean std	Mean abs. dev.
GrAd Variance	0.6035	(0.6030, 0.6041)	0.01373
Variance analysis	0.5070	(0.5068, 0.5073)	0.09296
Rescaled range	0.6077	(0.6067, 0.6087)	0.02673

#### 4.2.2. Generalized Hurst Exponent

By fitting the moments of increments against the scale, the original Hurst exponent can be generalized to be order-based (see [Barabási and Vicsek \(1991\)](#)). The generalized Hurst exponent, denoted by  $H_q$ , studies the scaling property of the increment distribution at the  $q$ -th order moment. For a time series observation  $X_i$ ,  $i = 1, 2, \dots, n$ , compute the  $q$ -th order structure function as

$$G_q(m) := \langle |X_{i+m} - X_i|^q \rangle_i,$$

where the  $\langle \cdot \rangle_i$  operator takes average over all feasible index,  $i$ .  $H_q$  is obtained by fitting

$$G_q(m) \sim m^{qH_q}.$$

For self-similar processes,  $H_q$  is the same for all order,  $q$ . In our study, we are mostly interested in the scaling of the second moment, thus fixing  $q = 2$ . The algorithm is to compute

$$G_2(m) := \left\langle (X_{i+m} - X_i)^2 \right\rangle_i \sim m^{2H_2}.$$

In the later part of the paper, we will set  $H = H_2$  for simplicity. This is the same as the multiscale variance function,  $V_m$ , estimator in Equation (21) (with data centered). The formal variance analysis algorithm is as follows.

Variance Analysis

- Let  $\mathcal{M} \subset \mathbb{N}^+$  be the set of scales for estimating the variance.
- For  $m \in \mathcal{M}$ , compute  $V_m$  using Equation (21).
- Fit OLS linear regression of  $\log(V_m) \sim \log(m)$ . Denote the slope as  $\alpha$ .
- Estimator  $\hat{H} = \alpha/2$ .

Similar to the original rescaled range method, the Hurst exponent is estimated by taking the log of both  $V_m$  and  $m$ , and fitting a linear regression to find the slope. However, as we have shown in Equation (11), there will be a constant term in the structure function if the data are subject to microstructure noise. Therefore, directly taking the log and fitting linear regression will cause inevitable bias in the estimation of  $H$ . The following proposition provides an analytical estimate of the deviation from the true value.

**Proposition 2.** *Let  $X_t$  be a self-similar process with Hurst exponent,  $H$ , and volatility,  $\sigma$ . The noisy observation  $Y_t = X_t + \epsilon_t$  has noise level  $\sigma_\epsilon$ . Suppose we know the variance  $\mathcal{V}[Y](\tau)$  for  $\tau \in \mathcal{T} \subset \mathbb{R}^+$  on a finite set of scales. Then the Hurst exponent estimation,  $\hat{H}$ , by fitting  $\log(\mathcal{V}[Y]) \sim \log(\tau)$  will be*

$$\left( 1 - \frac{2\eta}{\tau_{min}^{2H} + 2\eta} \right) \cdot H \leq \hat{H} \leq \left( 1 - \frac{2\eta}{\tau_{max}^{2H} + 2\eta} \right) \cdot H,$$

where  $\eta := \sigma_\epsilon^2 / \sigma^2$  is the noise ratio, and  $\tau_{min} := \inf \mathcal{T}$ ,  $\tau_{max} := \sup \mathcal{T}$ .

**Proof.** The slope of the OSL regression is bounded by the range of the derivative

$$\frac{d \log(\mathcal{V}[Y](\tau))}{d \log(\tau)} = 2\Lambda[Y](\tau).$$

Use the growth rate function in Property 4, and note that  $\Lambda[Y](\tau)$  is a monotonically increasing function, we can obtain

$$2H \cdot \frac{\tau_{min}^{2H}}{\tau_{min}^{2H} + 2\eta} \leq \alpha \leq 2H \cdot \frac{\tau_{max}^{2H}}{\tau_{max}^{2H} + 2\eta},$$

$$\left( 1 - \frac{2\eta}{\tau_{min}^{2H} + 2\eta} \right) \cdot H \leq \hat{H} \leq \left( 1 - \frac{2\eta}{\tau_{max}^{2H} + 2\eta} \right) \cdot H.$$

□

**Remark 3.** *The proposition shows that the estimated Hurst exponent will be strictly lower than the true value if there exists microstructure noise. The deviation amount, as calculated of the form  $\frac{2\eta}{\tau_{min}^{2H} + 2\eta}H$ , is determined by the noise level only through the noise ratio,  $\eta$ . Choosing  $\mathcal{T}$  at large timescales can reduce, but not completely remove, the bias. Lastly, the Hurst exponent itself can also affect the amount of deviation, with a smaller  $H$  suffering greater bias.*

In order to remove the microstructure noise bias, we propose the growth-adjusted (GrAd) variance analysis algorithm to estimate the Hurst exponent. Instead of directly

fitting the variance, the growth-adjusted variance function, defined in Proposition 1, is used as it has good conservation properties. Notice that, under the noisy fractional Brownian motion model, the growth-adjusted variance recovers the variance function of the latent process, up to a constant of  $H$  (see Proposition 1 and Equation (10)). The formal algorithm is as follows:

#### Growth-Adjusted Variance Analysis

- For  $m = 1, 2, \dots, M$ , compute  $V_m$  using Equation (21).
- Compute growth rate,  $\Lambda_m$ , using Equation (23).
- Compute the growth-adjusted variance:

$$\tilde{V}_m = \Lambda_m \cdot \frac{V_m + V_{m+1}}{2}. \quad (24)$$

- Fit least squares linear regression of  $\log \tilde{V}_m \sim \log(m + 1/2)$ . Denote the slope as  $\alpha$ .
- Estimator  $\hat{H} = \alpha/2$ .

Since we approximate differentiation by finite difference when calculating the growth rate, all other values are adapted to the average between scales  $m$  and  $m + 1$  for better accuracy.

#### 4.3. Noise Estimation

For the noisy fractional Brownian motion model, we can further estimate the noise level,  $\sigma_\epsilon$ .

Case 1: Brownian motion assumption.

If one is to assume that the latent process is a Brownian motion, then the noise level can be evaluated as follows. Using Property 4 and fix  $H = 0.5$ , we propose the following estimator for the noise level:

$$\min_{\beta_0 \geq 0, \beta_1 \in \mathbb{R}} \sum_{m \in \mathcal{M}} (\beta_1 m + \beta_0 - V_m)^2. \quad (25)$$

Then the noise level is estimated by  $\hat{\sigma}_\epsilon = \sqrt{\beta_0/2}$ , and the volatility is estimated by  $\hat{\sigma} = \sqrt{\beta_1}$ . The noise ratio can be computed as  $\hat{\eta} = \beta_0/(2\beta_1)$ . Here,  $\mathcal{M} \subset \mathbb{N}^+$  is a pre-specified scale set, and  $V_m$ s are computed using Equation (21). This is actually an OLS with the constraint on the intercept to be non-negative, because the variance of the noise  $\sigma_\epsilon^2 \geq 0$ .

Case 2: Fractional Brownian motion assumption.

In the general case, the latent process is assumed to be a fractional Brownian motion with  $H \in (0, 1)$ . In this case, we first need to use growth-adjusted variance analysis to estimate  $\hat{H}$ , and then solve

$$\min_{\beta_0 \geq 0, \beta_1 \in \mathbb{R}} \sum_{m \in \mathcal{M}} (\beta_1 m^{2\hat{H}} + \beta_0 - V_m)^2. \quad (26)$$

Then, the noise level is estimated by

$$\hat{\sigma}_\epsilon = \sqrt{\beta_0/2}.$$

**Remark 4.** In practice, the scale set,  $\mathcal{M}$ , can be chosen to consist of only very small  $m$ s, because  $V_m$  is most sensitive to the noise at a small scale. On the other hand, the error of the variance estimator becomes larger as  $m$  increases. Usually even  $\mathcal{M} = \{1, 2\}$  can be a good choice.

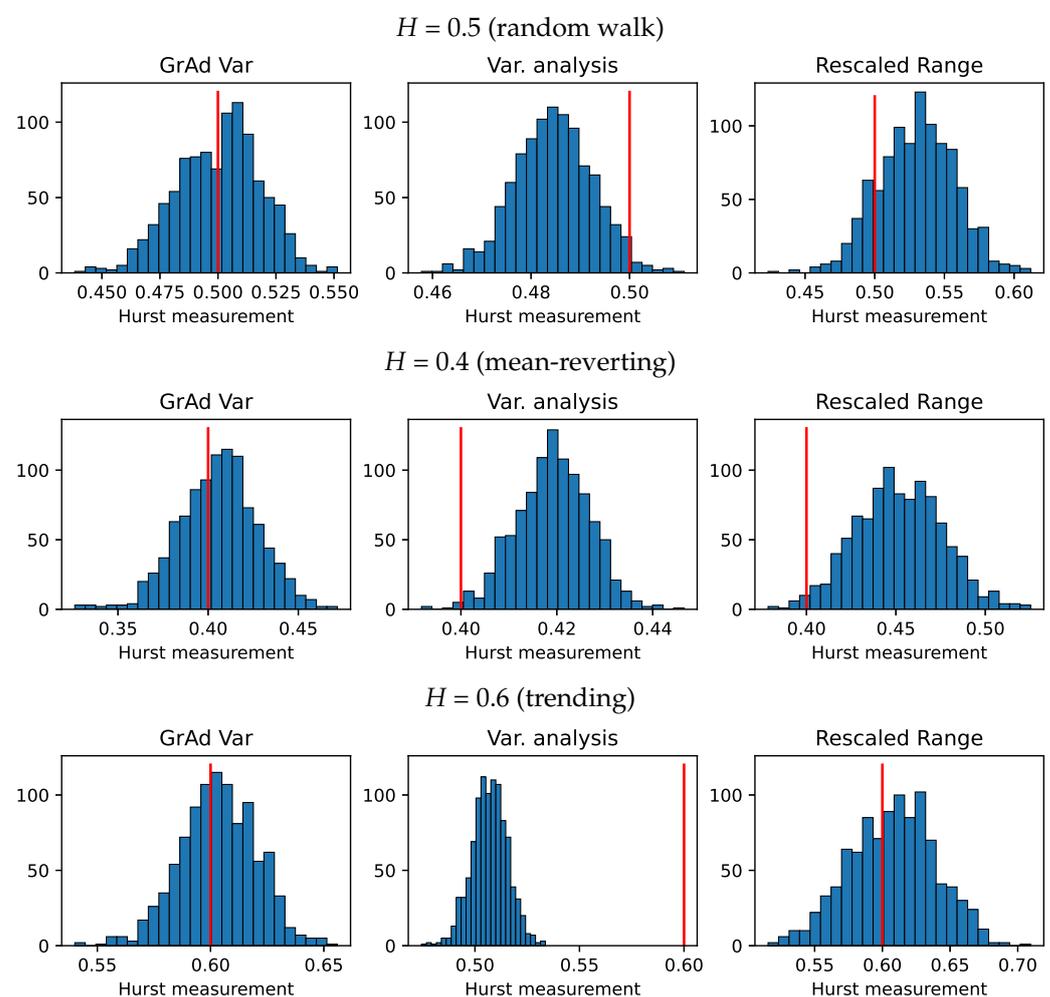
#### 4.4. Monte Carlo Simulation

To evaluate the performance of the proposed algorithm, we test on simulated noisy fractional Brownian motion sample paths. Three scenarios with Hurst exponents 0.5, 0.4, and 0.6 are examined with noise level  $\sigma_\epsilon = 2 \times 10^{-5}$  and volatility parameter  $\sigma = 1 \times 10^{-2}$ .

To mimic the 3 s intraday time series to be used in the next section, sample paths are generated on  $t \in [0, 1]$ , with 7800 equal-spaced points.

In Table 1, we present the simulation results of the Hurst exponent estimation, using the proposed growth-adjusted (GrAd) variance analysis, along with a comparison against the standard variance analysis and the original rescaled range analysis. Under three different scenarios, with the Hurst exponents corresponding to random walk, mean-reverting, and trending behaviors, we see that the proposed algorithm provides consistently unbiased results and has the lowest mean absolute deviation from the ground truth values.

Algorithms not handling microstructure noise cannot produce unbiased estimates for all of the scenarios. The histograms of the simulation result are in Figure 2 below. The red vertical line in each histogram indicates the ground truth value. We can see the distribution of the GrAd variance analysis estimator is centered around the ground truth values, while the other two algorithms yield biased results.



**Figure 2.** Histograms of Hurst exponent measurement using different algorithms. Three different scenarios are tested with ground truth value  $H = 0.5$  (random walk),  $H = 0.4$  (mean-reverting), and  $H = 0.6$  (trending). The red vertical line in each histogram indicates the ground truth value.

Table 2 shows the noise level estimation results on simulated noisy fractional Brownian motion data, with parameters the same as above. From the result, we can see that the noise estimation is most accurate when  $H$  is large and the fractional Brownian motion path is smoother. The accuracy is low when  $H$  is small. The reason might be that the fractional Brownian motion path at  $H < 0.5$  is rough and non-continuous, making the separation of noise extremely difficult.

**Table 2.** Noise estimates on simulated noisy fractional Brownian motion data. The ground truth value is  $2 \times 10^{-5}$ . Each scenario is estimated with 1000 sample paths.

$(10^{-5})$	$H = 0.5$	$H = 0.4$	$H = 0.6$
Mean	1.8109	0.7962	2.1046
St. Dev.	0.9877	1.6801	0.1194

## 5. Experiments on Intraday Data

In this section, we apply real-world data to the algorithms introduced in Section 4. Specifically, we consider a collection of exchange-traded funds (ETFs) and stocks (tickers: SPY, IWM, QQQ, XLK, AAPL, and MSFT), with dates ranging from January 2020 to February 2023, and conduct our experiments using 3 s high-frequency intraday price data. Prices are recorded on each day from open to close, spanning 6.5 h, making 7800 equal-spaced time points. We choose a time frame such that  $t = 1$  corresponds to 1 min (i.e.,  $\delta t = 0.05$ ). The interpretation is that any volatility estimates will reflect equivalently the standard deviation of a 1 min price change.

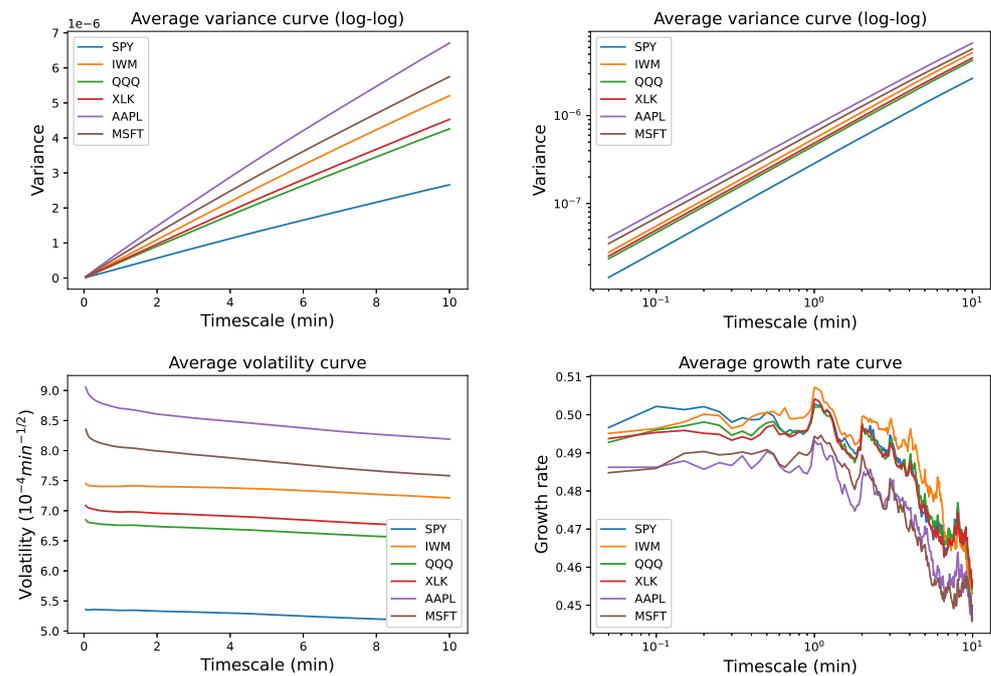
### 5.1. Ensemble Multiscale Curves

We start by showing the multiscale behaviors of real-world high-frequency price data. Figure 3 shows the plots of the multiscale functions estimated on a 3 s intraday time series, averaged over all available dates in the dataset. Starting with the top two plots showing the average variance curves, we see that for all tickers the variance scales in an approximately linear form, with very slight concavity. This is close to many base models in finance such as Brownian motion or other types of random walk with uncorrelated increments, as shown in Proposition 2. The log–log plot shows that the variance scales with an approximately constant growth rate, which looks like the behavior of self-similar processes (see Equation (10)). The slopes of the lines on the log–log plot can be used to estimate the Hurst exponents.

However, as we move on to the two bottom plots showing finer details of the curves, there are observations unexplained by a standard model. Starting with the bottom left plot showing the volatility curves, we see that the volatility is actually not constant, violating the form given by Proposition 2. There is a consistent downward slope across all six tickers. Moreover, for many of the tickers, especially the two stocks AAPL and MSFT, we see a clear sharp drop in volatility when the timescale is very small. This phenomenon matches the asymptotic behavior of the noisy fractional Brownian motion model, showing clear evidence of the microstructure noise. The finer sampling frequency may show a larger effect of the noise.

Turning to the lower right plot, we show the growth rate curves calculated from average variance functions. Even though the upper right plot indicates that the prices follow self-similar processes with seemingly constant growth rates, the bottom right plot shows that they are not.

For all tickers, if we look at the timescale within the 1 min range, the growth rate stays on an almost constant level, with a very slight increase. This can be explained by the noisy fractional Brownian motion model, as derived in Property 4. However, the growth rate declines obviously when going beyond the 1 min threshold, which means that self-similarity no longer holds. This is also not explained by the microstructure noise, which only contributes to an increasing growth rate. This phenomenon suggests that there might be some other dynamics in the mid-frequency region.



**Figure 3.** Average variance, volatility, and growth rate curves estimated on the 3 s intraday dataset. For the intraday time series on each day, one variance curve is estimated for each asset using Equation (21). The top two plots show the average variance curves over all dates in the dataset. The bottom two plots are volatility curves and growth rate curves derived from the average variance curves, using Equations (22) and (23), respectively.

We then proceed to estimate the parameters in the noisy fractional Brownian motion setting, using the average variance curves in Figure 3. Because we have seen in the growth rate plot that the model assumption may not be valid beyond the 1 min limit, we only take the  $V_m$  for  $m = 1, \dots, 20$  (from 3 s to 1 min). The GrAd variance analysis algorithm is used to fit the Hurst exponent,  $H$ , with results shown in Table 3. Since the Hurst exponents are very close to  $1/2$ , we take the Brownian motion assumption for noise fitting (Equation (25)). Noise level,  $\sigma_\epsilon$ , volatility,  $\sigma$ , and noise ratio,  $\eta := \sigma_\epsilon^2 / \sigma^2$ , estimations are also shown in Table 3.

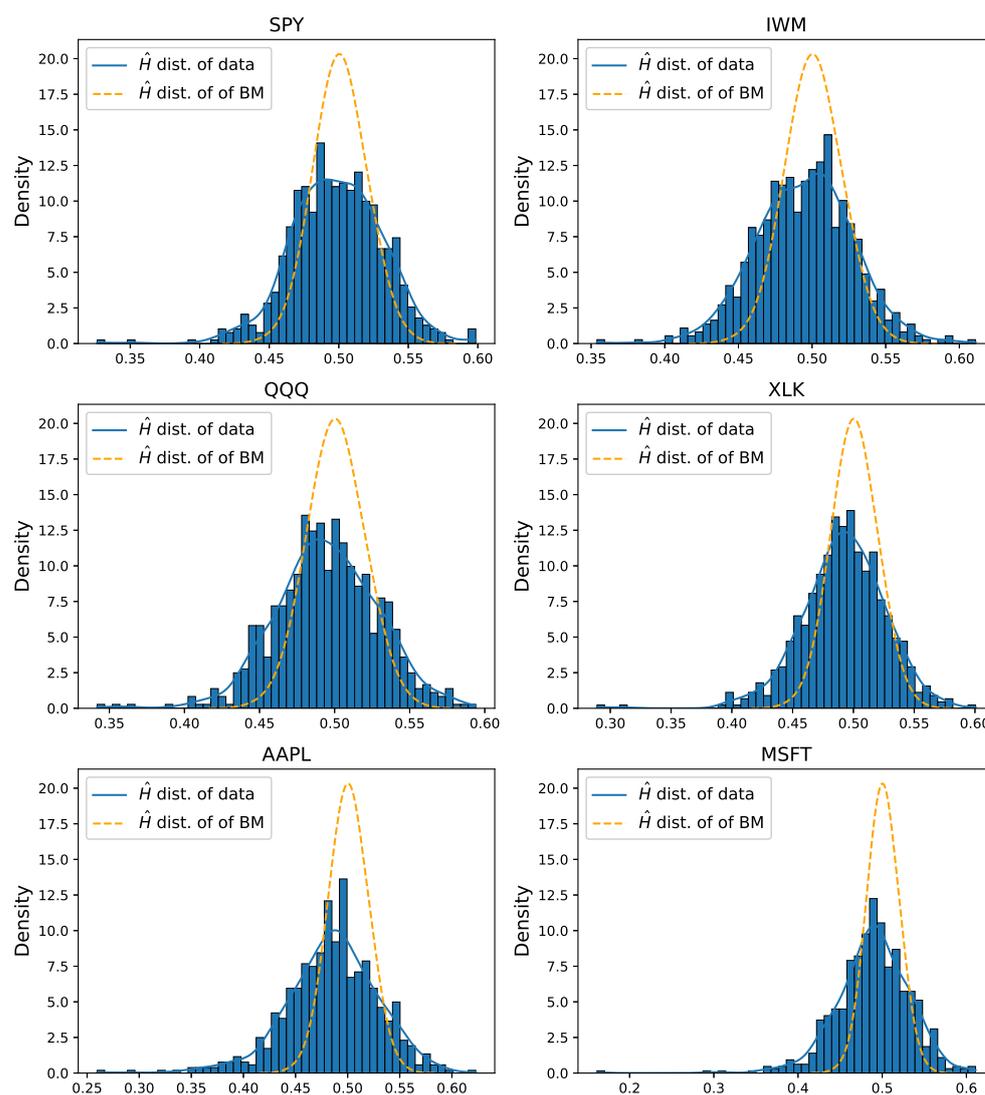
**Table 3.** Hurst exponent,  $H$ , noise level,  $\sigma_\epsilon$ , volatility,  $\sigma$ , and noise ratio,  $\eta := \sigma_\epsilon^2 / \sigma^2$ , estimations on the average variance curves in Figure 3. Equation (21) estimates variance curves for each intraday time series. The average (over all available dates) is taken as the  $V_m$  in the GrAd variance analysis algorithm to fit the Hurst exponent,  $H$ . Noise level,  $\sigma_\epsilon$ , volatility,  $\sigma$ , and noise ratio,  $\eta$ , are estimated using the Brownian motion assumption estimator in Equation (25), also using the averages,  $V_m$ .

	SPY	IWM	QQQ	XLK	AAPL	MSFT
$H$	0.4975	0.5005	0.4967	0.4957	0.4869	0.4902
$\sigma_\epsilon (10^{-5})$	1.3458	1.1871	2.5640	3.0032	6.1282	5.1814
$\sigma (10^{-4} \text{ min}^{-\frac{1}{2}})$	1.1949	1.6554	1.5098	1.5581	1.9400	1.7967
$\eta (10^{-2} \text{ min})$	1.2686	0.5142	2.8838	3.7149	9.9779	8.3168

### 5.2. Daily Hurst Exponent

Although the ensemble multiscale behavior suggests that the price process is very close to a random walk, the true dynamics of each intraday path may be mean-reverting, trending, or otherwise. In this section, we employ the proposed growth-adjusted variance analysis to estimate Hurst exponents for each intraday price path. The time series on each day for each ticker will yield a single Hurst exponent estimate.

Figure 4 shows the histograms of daily Hurst exponent results, compared against the distribution of estimating the Hurst exponent from Brownian motion simulation. Although the Hurst exponent of the intraday data is centered around  $1/2$ , we can clearly see that its distribution is not the same as Brownian motion.  $H$  estimated on the real-world data is in general distributed to the left of a Brownian motion. The kernel density curves show a much wider spread distribution and feature long tails. There is a significant portion of days that have very low or high Hurst exponents, which is unlikely if the Brownian motion assumption holds. This implies that the price dynamics can be different on different days. There are days when the price movement is mean-reverting, and on other days the movements are trending or like a random walk.



**Figure 4.** Hurst exponent estimated on 3 s intraday data from 2020 to 2023. The intraday time series on each date for each ticker yield one Hurst exponent estimation. This figure shows the histogram of the estimated values over all dates. The orange dash-line shows the distribution of the Hurst exponent estimator,  $\hat{H}$ , from a standard Brownian motion in comparison with that from the dataset.

There have been ongoing debates over whether a small or large Hurst exponent estimated from price data is able to reject the Brownian motion null hypothesis. While the statistical analysis of  $H$  estimates is complicated, there are studies on its estimation error, including the error bounds of the  $H$  estimator on standard Brownian motions (see Couillard and Davison (2005)).

With our proposed algorithm, we now employ a more empirical way to determine whether the result is statistically significant to reject the null hypothesis  $H = 1/2$ . For any algorithm discussed herein, we start by evaluating via simulations of standard Brownian motions. Then, estimate  $\mu_0$  as the mean of the null hypothesis, and  $\sigma_0$  as the standard deviation of the null hypothesis. For any new  $\hat{H}$  estimated on real-world data, we compute

$$t = \frac{\hat{H} - \mu_0}{\sigma_0}$$

as the  $t$ -statistic. One can check if the  $t$ -statistic is below or above a certain significance threshold to decide if the Brownian motion null hypothesis should be rejected.

In Table 4, we show the distributional statistics of daily Hurst exponent values, with the last three columns showing the percentage of days when the estimated value shows mean-reverting, trending, or random walk values, at the 5% significance level. We observe that the portion is much larger, which could be from standard Brownian motion. Similar to Table 3, the overall level of the estimated Hurst exponents is slightly below 0.5. Note also that the standard deviation is much larger than that from a (fractional) Brownian motion, which we have seen from Table 1.

**Table 4.** Hurst exponent estimated on 3 s intraday data from 2020 to 2023. The intraday time series on each date for each ticker yield one Hurst exponent estimation. This table shows statistics of the estimates over all dates. The mean and the standard deviation of  $H$  are in the first two columns, and the last three columns show the percentage of dates with the Hurst exponent estimated to be mean-reverting, trending, or random walk values, with a threshold at the 5% significance level.

	Mean	St.Dev	Mean-Reverting (%)	Trending (%)	Random Walk (%)
SPY	0.4986	0.0327	12.5523	12.9707	74.4770
IWM	0.4948	0.0326	17.8522	9.0656	73.0823
QQQ	0.4949	0.0340	17.5732	11.9944	70.4324
XLK	0.4919	0.0342	19.1074	9.3445	71.5481
AAPL	0.4854	0.0443	29.0098	12.2734	58.7169
MSFT	0.4891	0.0437	24.5467	13.6681	61.7852

### 5.3. Intraday Evolution

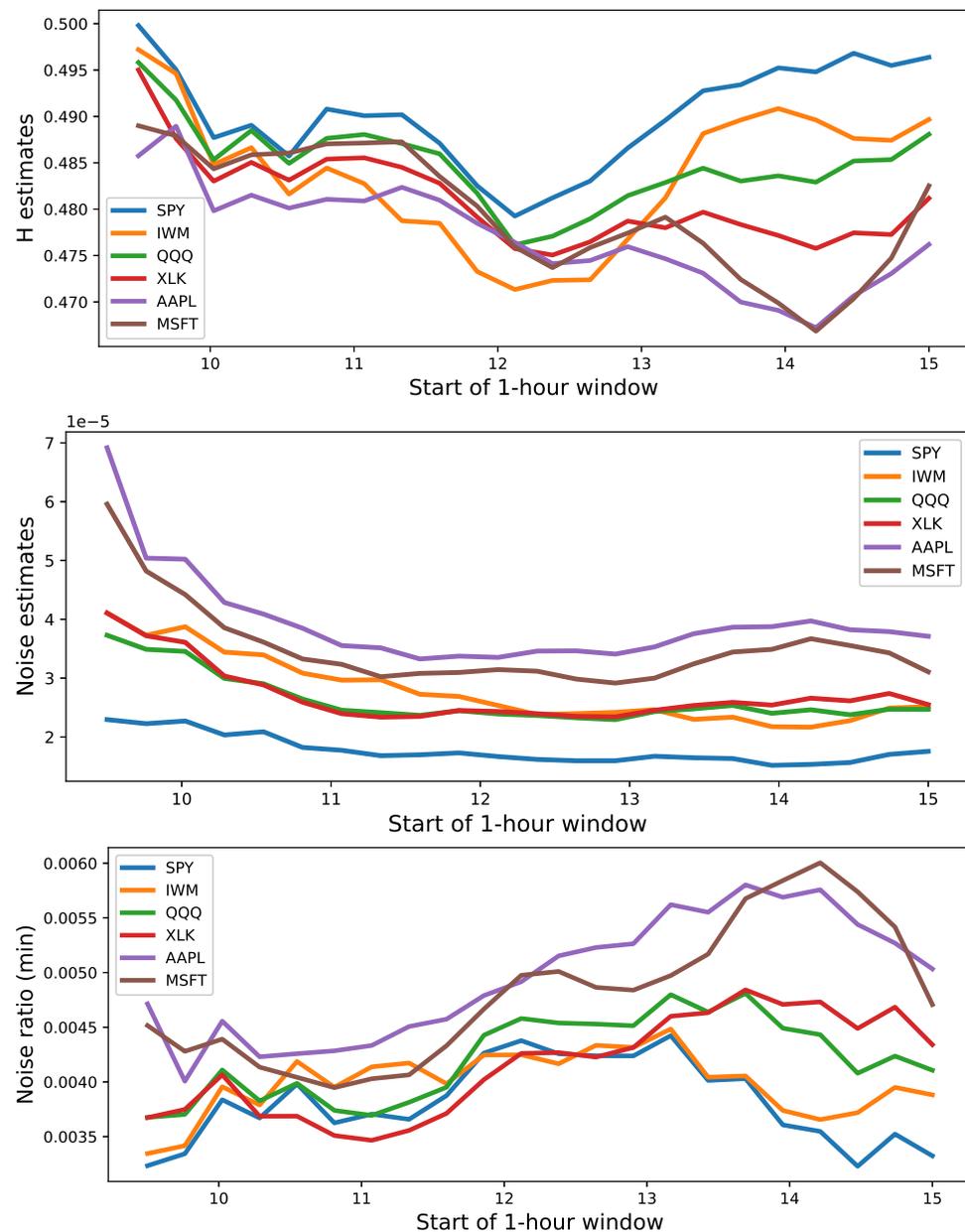
In reality, asset prices may exhibit different behaviors at different times of the day. There is well-documented evidence that shows patterns and seasonality of return distribution during trading hours (see [Allez and Bouchaud \(2011\)](#) and references therein). One well-known observation is the U-shape volatility, as examined by [Admati and Pfleiderer \(1988\)](#) and [Andersen and Bollerslev \(1997\)](#), which states that the volatility is highest near market open and close, and lowest at midday.

To test if the multiscale parameters are constant, we show the intraday rolling Hurst exponent, noise level, and noise ratio in Figure 5. The result shows varying Hurst exponents at different times of day, i.e., the scaling exponent is not constant. For all of the tickers, the exponent is highest at market open, and then decreases rapidly.

The market indexes ETFs SPY and IWM have the smallest values in the middle of the day, showing mean-reverting behavior. Their Hurst exponents at market open and in the last hours are close to 0.5, indicating weak dependency. The “U” shape of their Hurst exponents can help understand how the commonly observed U-shape volatility pattern is formed. In the middle of the day, the price movement is more mean-reverting. Even though the volatility is lower at midday, the smaller Hurst exponents indicate that the return can be even noisier as the noise ratio is higher.

Unlike the index ETFs, the two stocks AAPL and MSFT have Hurst exponents always below 0.5, and continue decreasing in the afternoon. Near the last hour before market close, their Hurst exponents notably increase towards 0.5, which might be due to increasing trading activities. The decreasing Hurst exponent can be compared with the increasing

kurtosis result from the work by Allez and Bouchaud (2011), suggesting that the more long-tail movement towards the end of the day actually may happen at small timescales. For QQQ and XLK, which have a larger correlation with the tech stocks, their behavior is in between SPY, IWM, and AAPL, MSFT. Different Hurst exponent patterns clearly show different trading dynamics for ETFs and stocks.



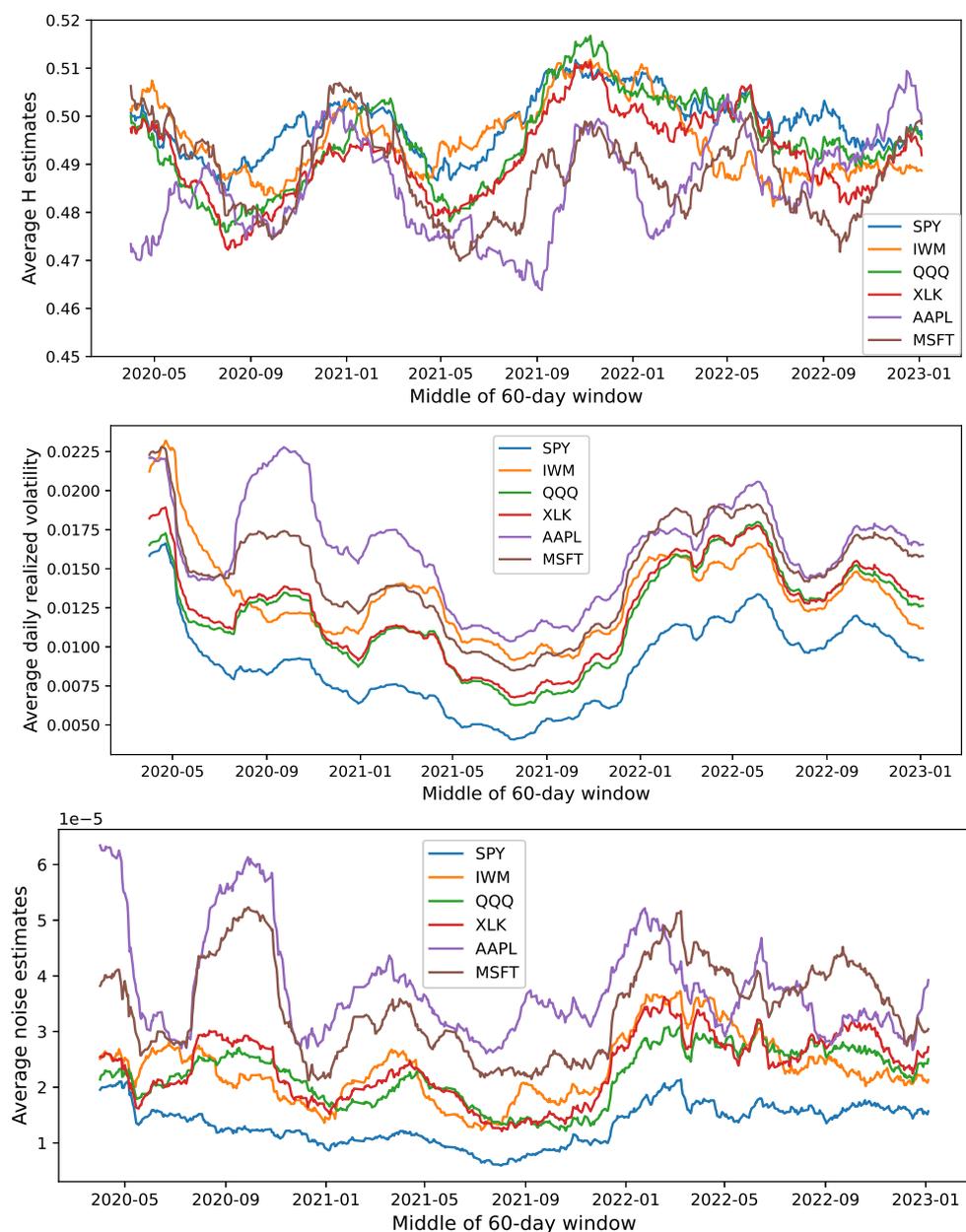
**Figure 5.** Hurst exponent,  $H$ , noise level,  $\sigma_\epsilon$ , and noise ratio,  $\eta := \sigma_\epsilon^2 / \sigma^2$ , on a 1 hour rolling window. The x-axis marks the time of the day (e.g., 10 = 10 a.m. EST and 13 = 1 p.m. EST). The starting point of the window shifts on a 15 min basis. For each day, one rolling Hurst curve is estimated for each asset. The plot shows the average over all dates in the dataset.

The noise level in the noisy fractional Brownian motion model provides another lens to inspect the intraday volatility pattern. The middle plot in Figure 5 shows the microstructure noise estimation on a 1 hour rolling window. We see that all of the assets have the highest noise value at market open, and decrease or flat out into the day. The two stocks AAPL and MSFT have the sharpest trend. The bottom plot in Figure 5 shows the rolling noise ratio is defined as  $\eta := \sigma_\epsilon^2 / \sigma^2$ . From the plot, we see that the two market indexes ETFs

SPY and IWM have noise ratios that increase in the morning and then decrease in the afternoon. In comparison, the two stocks AAPL and MSFT have increasing noise ratios almost throughout the day, but drop sharply near market close. Different noise ratio patterns may indicate different trading activities.

#### 5.4. Evolution of Parameters

The Hurst exponent can be different on different days. In Figure 6 we show the 60-day rolling average of the Hurst exponent from 2020 to 2023, in comparison with the realized volatility and noise level change. We can see that, at the beginning of the pandemic in 2020, most assets have a Hurst exponent close to 0.5, corresponding to a random walk. The Hurst exponent then trends lower towards mean-reverting values and fluctuates for a period of time. Overall, the ETFs tend to have a higher Hurst exponent, which stays closer to 0.5, while the Hurst exponent for the two stocks AAPL and MSFT tends to fluctuate below 0.5.



**Figure 6.** Time series of the Hurst exponent, realized volatility, and noise from 2020 to 2023. Values are estimated from intraday data sampled every 3 s for each ticker and averaged over 60-day rolling windows.

As for the volatility level, we see that the noise pattern moves together with the realized volatility time series. The difference between the noise among different assets is smaller than that of the volatility. For the two stocks AAPL and MSFT, their microstructure noise is higher than that for the ETFs for a long period from early 2020 to early 2021. Even though all of the tickers have different volatility levels, the noise level difference within the stock group and within the ETF group is very small, suggesting two distinct types of trading activities for stocks and ETFs.

## 6. Conclusions

We have presented a multiscale analysis of the volatility of noisy high-frequency data. A key feature of our framework is a fractional Brownian motion with microstructure noise. The proposed noisy fractional Brownian motion model is shown to possess a variety of volatility behaviors suitable for intraday price processes. For empirical estimation, we have proposed a new Hurst exponent estimator for the noisy fractional Brownian motion model and applied it to a collection of major US stocks and ETFs. This allows us to understand not only the intraday evolution of asset prices but also their long-term behaviors better.

For future research, one practical application of our approach is to incorporate the Hurst exponent and microstructure noise estimates into machine learning models for price prediction at different timescales. It is of practical interest to investigate the driving factors for the Hurst exponent for different assets, ranging from equities to cryptocurrencies, over time. This can potentially be useful for developing related trading and risk management strategies.

**Author Contributions:** Conceptualization, T.L.; Methodology, T.L. and T.Z.; Validation, T.Z.; Formal analysis, T.L. and T.Z.; Investigation, T.L. and T.Z.; Data curation, T.Z.; Writing – original draft, T.L. and T.Z.; Writing – review and editing, T.L.; Visualization, T.Z.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data source: <https://www.kaggle.com/datasets/brtnsmth/intraday-market-data> (accessed on 10 June 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Admati, Anat R., and Paul Pfleiderer. 1988. A theory of intraday patterns: Volume and price variability. *Review of Financial Studies* 1: 3–40. [CrossRef]
- Ait-Sahalia, Yacine, Per A. Mykland, and Lan Zhang. 2005. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18: 351–416. [CrossRef]
- Ait-Sahalia, Yacine, Jianqing Fan, and Dacheng Xiu. 2010. High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105: 1504–17. [CrossRef]
- Allez, Romain, and Jean-Philippe Bouchaud. 2011. Individual and collective stock dynamics: Intra-day seasonalities. *New Journal of Physics* 13: 025010. [CrossRef]
- Andersen, Torben G., and Tim Bollerslev. 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4: 115–58. [CrossRef]
- Andersen, Torben G., Tim Bollerslev, and F. X. Diebold. 2003. *Some like It Smooth, and Some like It Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility*. Working Paper. Philadelphia: Financial Institutions Center, Wharton School, University of Pennsylvania. [CrossRef]
- Bandi, Federico M., and Jeffrey R. Russell. 2008. *Microstructure Noise, Realized Volatility, and Optimal Sampling*. *The Review of Economic Studies* 75: 339–69. [CrossRef]
- Barabási, Albert-László, and Tamas Vicsek. 1991. Multifractality of self-affine fractals. *Physical Review A* 44: 2730. [CrossRef]
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 253–80. [CrossRef]
- Bibinger, Markus, Nikolaus Hautsch, Peter Malec, and Markus Reiß. 2014. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *The Annals of Statistics* 42: 1312–46. [CrossRef]
- Black, Fisher. 1986. Noise. *Journal of Finance* 41: 528–43. [CrossRef]

- Bouri, Elie, Ladislav Kristoufek, Tanveer Ahmad, and Syed Jawad Hussain Shahzad. 2022. Microstructure noise and idiosyncratic volatility anomalies in cryptocurrencies. *Annals of Operations Research*. [\[CrossRef\]](#)
- Capobianco, Enrico. 2004. Multiscale stochastic dynamics in finance. *Physica A: Statistical Mechanics and Its Applications* 344: 122–27. [\[CrossRef\]](#)
- Cont, Rama, and Purba Das. 2022. Rough volatility: Fact or artefact? *arXiv* arXiv:2203.13820.
- Couillard, Michel, and Matt Davison. 2005. A comment on measuring the Hurst exponent of financial time series. *Physica A: Statistical Mechanics and Its Applications* 348: 404–18. [\[CrossRef\]](#)
- Di Matteo, Tiziana. 2007. Multi-scaling in finance. *Quantitative Finance* 7: 21–36. [\[CrossRef\]](#)
- Fama, Eugene F. 1965. The behavior of stock-market prices. *Journal of Business* 38: 34–105. [\[CrossRef\]](#)
- Fan, Jianqing, and Yazhen Wang. 2007. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association* 102: 1349–62. [\[CrossRef\]](#)
- Górski, A., S. Drożdż, and J. Speth. 2002. Financial multifractality and its subtleties: An example of DAX. *Physica A: Statistical Mechanics and Its Applications* 316: 496–510. [\[CrossRef\]](#)
- Granero, M. Sanchez, J. Trinidad Segovia, and J. Garcia Pérez. 2008. Some comments on Hurst exponent and the long memory processes on capital markets. *Physica A: Statistical Mechanics and Its Applications* 387: 5543–51. [\[CrossRef\]](#)
- Hurst, Harold Edwin. 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116: 770–99. [\[CrossRef\]](#)
- Jacod, Jean, Yingying Li, and Xinghua Zheng. 2017. Statistical properties of microstructure noise. *Econometrica* 85: 1133–74. [\[CrossRef\]](#)
- Leung, Tim, and Theodore Zhao. 2021. Financial time series analysis and forecasting with Hilbert–Huang transform feature generation and machine learning. *Applied Stochastic Models in Business and Industry* 37: 993–1016. [\[CrossRef\]](#)
- Leung, Tim, and Theodore Zhao. 2022. Adaptive complementary ensemble emd and energy-frequency spectra of cryptocurrency prices. *International Journal of Financial Engineering* 9: 2141008. [\[CrossRef\]](#)
- Mandelbrot, Benoit. 1963. New methods in statistical economics. *Journal of Political Economy* 71: 421–40. [\[CrossRef\]](#)
- Mandelbrot, Benoit. 2013. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. *Selecta Volume E*. Berlin: Springer Science & Business Media.
- Mandelbrot, Benoit, and John W. Van Ness. 1968. Fractional Bownian motions, fractional noises and applications. *SIAM Review* 10: 422–37. [\[CrossRef\]](#)
- Mandelbrot, Benoit, and Richard L. Hudson. 2007. *The Misbehavior of Markets: A Fractal View of Financial Turbulence*. New York: Basic Books.
- Matos, Jose, Silvio M. Gama, Heather J. Ruskin, Adel Al Sharkasi, and Martin Crane. 2008. Time and scale Hurst exponent analysis for financial markets. *Physica A: Statistical Mechanics and Its Applications* 387: 3910–15. [\[CrossRef\]](#)
- Mensi, Walid, Mobeen Ur Rehman, Muhammad Shafiullah, Khamis Hamed Al-Yahyaee, and Ahmet Sensoy. 2021. High frequency multiscale relationships among major cryptocurrencies: Portfolio management implications. *Financial Innovation* 7: 1–21. [\[CrossRef\]](#)
- Müller, Ulrich A., Michel M. Dacorogna, Richard B. Olsen, Olivier V. Pictet, Matthias Schwarz, and Claude Morgeneegg. 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance* 14: 1189–208.
- Osborne, M. F. 1959. Brownian motion in the stock market. *Operations Research* 7: 145–73. [\[CrossRef\]](#)
- Peng, Chung-Kang, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. 1994. Mosaic organization of DNA nucleotides. *Physical Review E* 49: 1685. [\[CrossRef\]](#)
- Podolskij, Mark, and Mathias Vetter. 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15: 634–58. [\[CrossRef\]](#)
- Rogers, Chris. 1997. Arbitrage with fractional Brownian motion. *Mathematical Finance* 7: 95–105. [\[CrossRef\]](#)
- Tzouras, Spilios, Christoforos Anagnostopoulos, and Emma McCoy. 2015. Financial time series modeling using the Hurst exponent. *Physica A: Statistical Mechanics and Its Applications* 425: 50–68. [\[CrossRef\]](#)
- Weron, Rafal. 2002. Estimating long-range dependence: Finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and Its Applications* 312: 285–99. [\[CrossRef\]](#)
- Zhang, Lan, Per. A. Mykland, and Yacine Aït-Sahalia. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100: 1394–411. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.