3rd Report, Computer Simulations

Katarzyna Macioszek and Ada Majchrzak

June 30, 2023

In this report, we will shortly describe, and then compare, two methods of estimation of the Hurst index for Fractional Brownian Motion (further called FBM). Before we proceed with analysis results, let us first start with introducing FBM and the Hurst index. It is also important to note that all the computations were performed in Python.

1 Fractional Brownian Motion

Fractional Brownian Motion is a generalization of Brownian Motion, where increments are not independent. It can be described by the following stochastic representation

$$B_H(t) = \frac{1}{\Gamma(H+\frac{1}{2})}\left(\int_{-\infty}^{0}\left[(t-s)^{H-1/2}-(-s)^{H-1/2}\right]dB(s)+\int_0^t (t-s)^{H-1/2}dB(s)\right), \tag{1}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)dx$ and $0 < H < 1$ is the Hurst parameter, which we will describe in net section

Often the Fractional Brownian Motion is also described by it's autocovariance function γ

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

To get a better understanding of how this process behaves, we will show a few trajectories and basic properties of FBM. To simulate the sample, due to performance and accuracy of methods, we decided to use Davies and Harte method.

On Figure 1 we presented some example trajectories to be able to observe behavior of FBM. On the second plot we see empirical quantile lines generated from $N = 1000$ samples. They are compared to the theoretical values that are described by the following formula

$$q(\tau) \sim \tau^{2H} Z_q, \tag{3}$$

where Z_q is q-th quantile of the standard normal distribution. As the process is random, to calculate quantile lines we used $N = 1000$ trajectories. We can see that empirical and theoretical lines are almost the same for analyzed number of tracks, which thanks to Davies and Harte method is really quick to compile. Moreover we can immediately see that Fractional Brownian Motion is not a stationary process as quantile lines are not parallel nor constant.

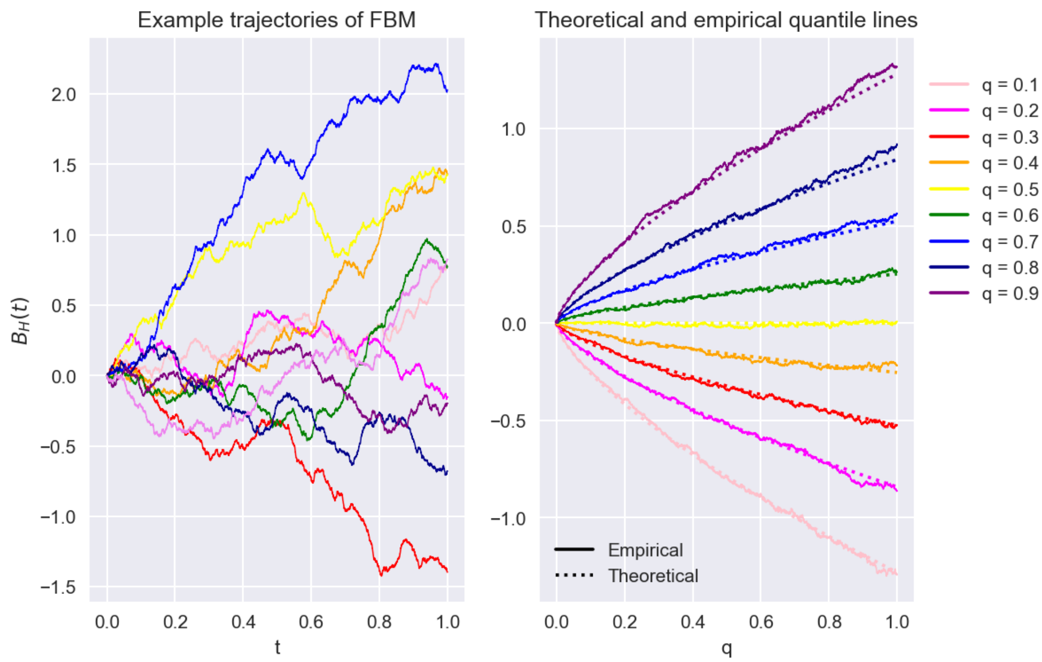

Figure 1: Trajectories and quantile lines for FBM

2 Hurst index

Hurst parameter H describes the memory of a process. Different values of this attribute correspond to particular processes

- $0 < H < \frac{1}{2}$ – short memory process,
- $H = \frac{1}{2}$ – geometric random walk,
- $\frac{1}{2} < H < 1$ – long-range dependent process.

In particular the fractional Brownian Motion with $H = \frac{1}{2}$ is the classical Brownian Motion.

This parameter is also closely related to a specific property of a process – self-similarity. That can give a better understanding to what H really is. A process $X = (X(t))_{t>0}$ is called self-similar (ss) if it satisfies

$$X(at) \overset{d}{=} a^H X(t), \tag{4}$$

where H is the Hurst parameter.

3 Lag variance method

Let's denote an arbitrary lag of the time series X_t by τ. Then, the variance for lag τ can be expressed as

$$\mathrm{Var}(\tau) = \mathrm{Var}(X_{t+\tau} - X_t). \tag{5}$$

We can write down the relationship between the variance and the lag

$$\mathrm{Var}(\tau) \approx \tau^{2H}, \tag{6}$$

where H — Hurst index. Having this relationship we can easily estimate the Hurst index by simply computing the slope of $\log(\mathrm{Var}(\tau))$ against $\log(\tau)$. Obtained slope, divided by 2, is our H estimator.

4 Rescaled range method

This method is based on the self-similarity property of considered process $X = (X(t))_{t>0}$. First we have to find the mean ($M = \frac{1}{n}\sum_1^n X_i$) and standard deviation $S(n) = \frac{1}{n}\sum_1^n (X_i - M)^2$). Next we construct a mean adjusted series $Y_i = X_i - M$ and a cumulative deviation $Z_t = \sum_1^n Y_i$. We use that to calculate the range

$$R(n) = \max(Z_1, Z_2, \ldots, Z_n) - \min(Z_1, Z_2, \ldots, Z_n) \tag{7}$$

Finally from that we can calculate the rescaled range $P(n) = \frac{R(n)}{S(n)}$ and from that obtain the estimation of Hurst parameter

$$\widehat{H} = \frac{\log P(n)}{\log n} \tag{8}$$

5 Results

In this section, we will compare the methods of estimation of the Hurst index, described above. Let us first take a look at the mean variance and the mean bias for each method. Both statistics were computed 100 times for values of H from 0.1 to 0.9 with 0.01 step, and then averaged.

Table 1: Mean variance and bias for both estimation methods

	Lag variance	Rescaled range
Variance	0.00011	0.00202
Bias	0.01007	0.04454

From Table 1 we observe that for both measures, Lag variance method performs better. Both estimation variance and bias are bigger for rescaled range method. Later we will see that both methods permorm similarly for long-range dependent processes, however lag variance method is significantly better for processes with short memory.

Next thing we want to compare is the estimation error. Figure 2 shows scatter plots of real and estimated Hurst parameter for the lag variance method. We can see here that the method estimates H really well as scatter plots are very close to each other and estimation errors vary from 0.00 to around 0.04. It is nice to observe that the errors are much smaller for $H < 0.7$ and above this value they tend to be even twice bigger. Figure 3 shows the same plots for rescaled range method. We clearly see that this method is less accurate — scatter plots don't overlap so well and the values of error reach even 0.125. Moreover, for $H < 0.5$ the errors tend to reach higher values, while for $H > 0.5$ they are more scattered. It shows that the rescaled range method works better for long memory processes.

Now we will look at the boxplots of estimated values of Hurst index, for $H = 0.3, 0.7$ and both methods. On Figure 4 we see the boxplots for lag variance method. We can observe that for $H = 0.3$ the median is much closer to the theoretical value than in the case of $H = 0.7$. Also the whiskers are less

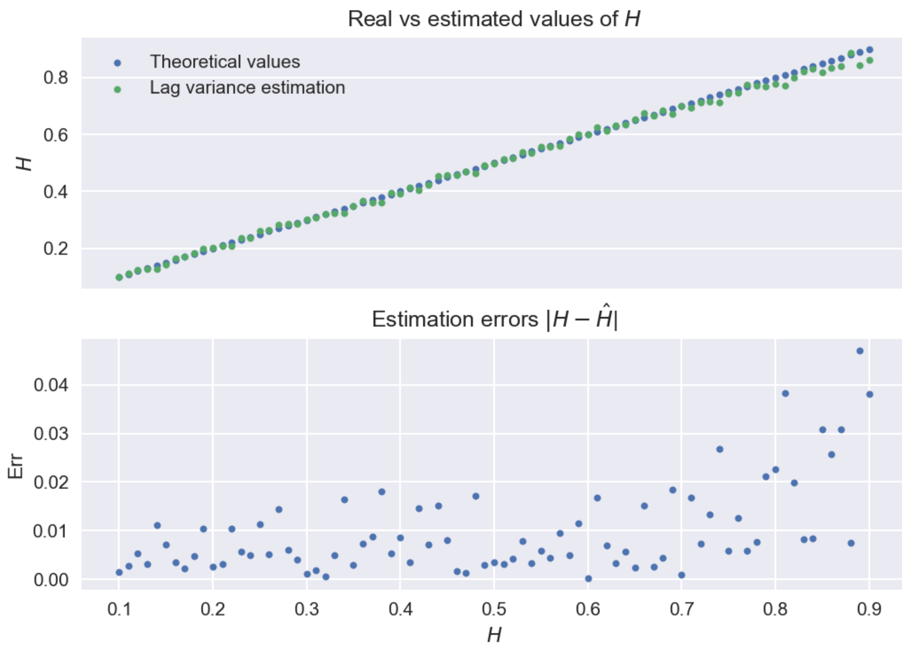

Figure 2: Real vs estimated values of H and estimation errors for lag variance method

stretched for smaller H, which indicated smaller discrepancy of values outside the interquartile range. On the other hand, outliers can only be seen on the first boxplot, so we can not definitely decide for which H the method performs better.

Figure 5 displays the boxplots for rescaled range method. This time we can clearly say that the method performs better for $H = 0.7$. Not only is the median much closer to the theoretical value, but also we observe no outliers. Moreover, for $H = 0.3$ almost all observations are stretched above the theoretical values of H. So again, we can see that the rescaled range method is not that good with short memory processes. Last thing we will use to collate the two methods is quantile line plot. Here we will see how the quantile lines with estimated H compare to those with theoretical values of H. Figure 6 shows that for the lag variance method, the differences are almost invisible, which indicates that the estimations are very good. We also see no significant difference between different values of H. From Figure 7 we can see that the situation is quite different in the case of rescaled range method. Here, the supremacy of $H = 0.7$ is very explicit — while for this value of H the empirical and theoretical quantile lines are very close to each other, or even overlap, for $H = 0.3$ we can observe very visible differences. So again, we showed that the rescaled range method performs better for long memory processes.

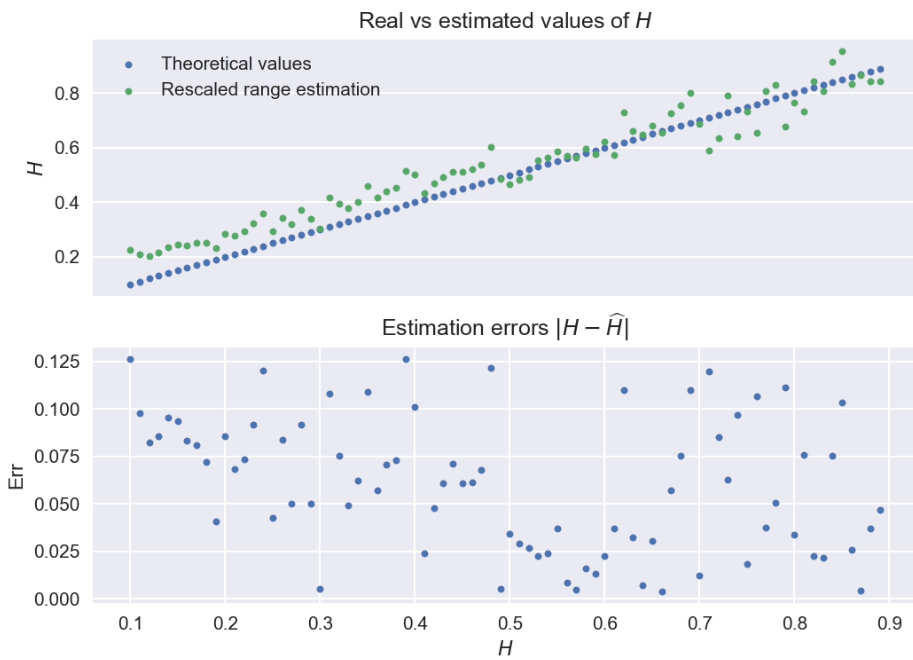

Figure 3: Real vs estimated values of H and estimation errors for rescaled range method

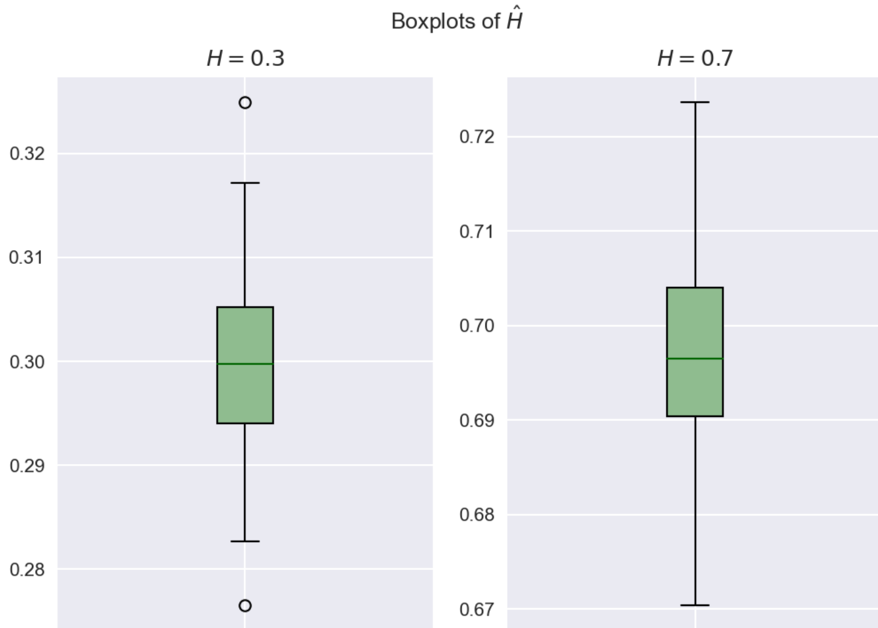

Figure 4: Boxplots of estimated values of H for lag variance method

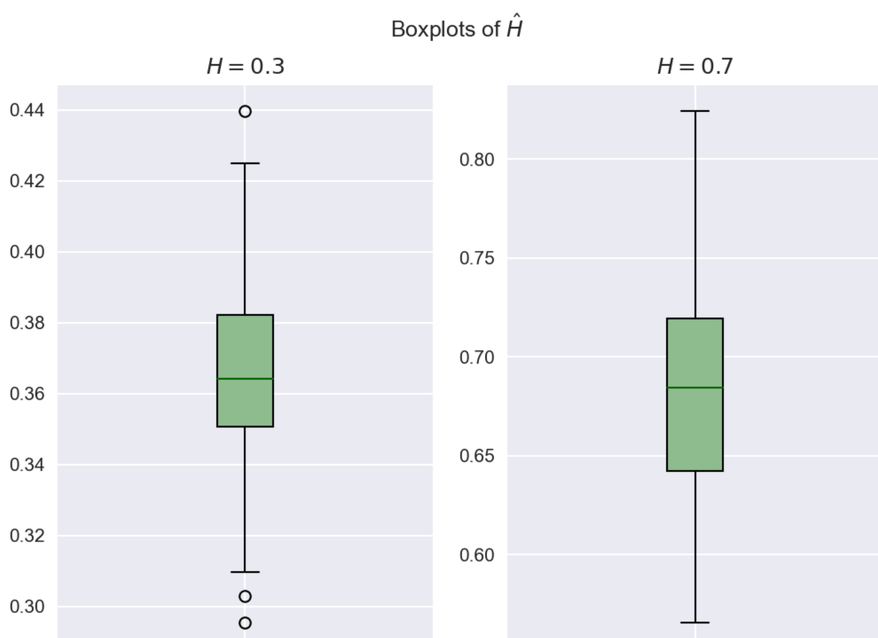

Figure 5: Boxplots of estimated values of H for rescaled range method

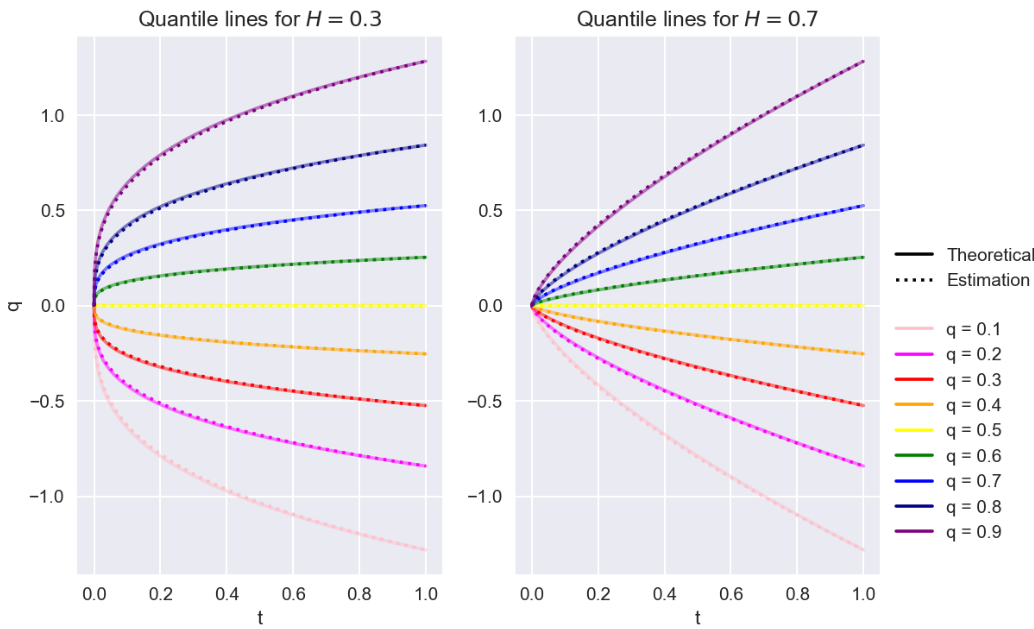

Figure 6: Quantile lines for lag variance method

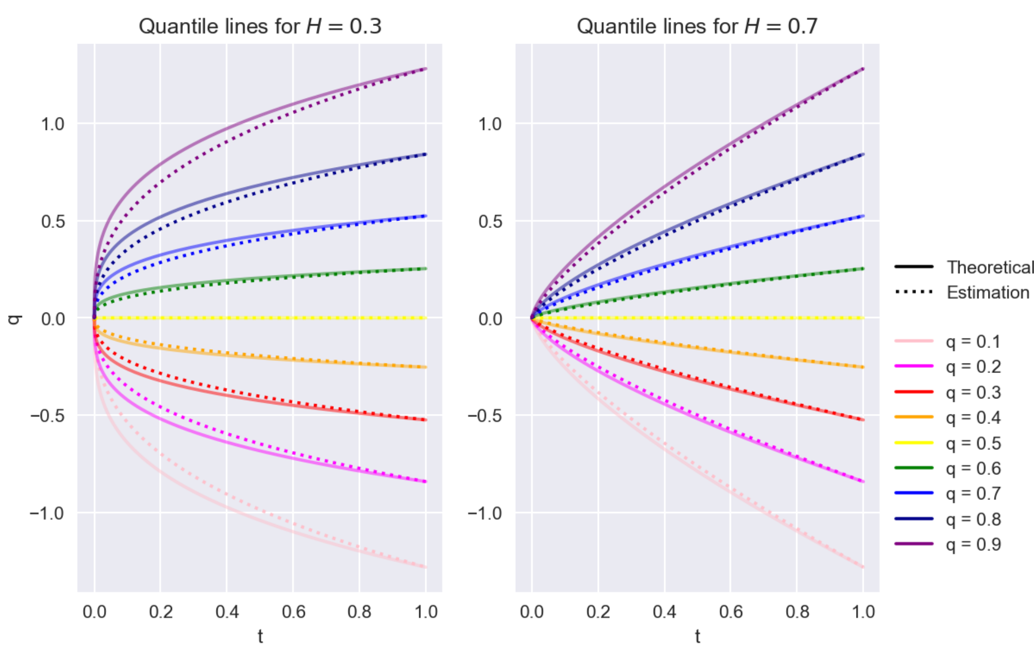

Figure 7: Quantile lines for rescaled range method

6 Conclusions

Comparing the overall performance of the two presented methods, we can surely say that the lag variance method is a better choice for estimating the Hurst index value. It is much simpler and more accurate, especially in the case of short memory processes. However, if we are only interested in long memory processes, both methods will stand for a good choice as they are similarly accurate for $H > 0.5$. Of course, in this report we only touched the surface of Hurst parameter estimation. There are multiple methods that we didn't cover here, some of which can be found in [1].

References

[1] Amjat H. Hamza and Munaf Y. Hmood (2021) *Comparison of Hurst exponent estimation methods*.

[2] `http://epchan.blogspot.com/2016/04/mean-reversion-momentum-and-volatility.html`.