Computer Simulations of Stochastic Processes

Report III

Piotr Kobus 185677

1 Methods of Approximations

Three methods of approximations will be compared. To describe them the top-down approach will be used. Two of them are based on detrending and basic idea is similar. In first step we need to have a trajectory of self-similar process. In our case it will be fractional Brownian motion. Then, we have to divide our series into windows of the same length. In every window we have to approximate series by some function. For every method this function will be different. Having this approximation we need to calculate a fluctuation function which will look differently for every method. Now we need to change size of windows, approximate series and calculate one more time fluctuation function. After repeating this procedure many times we will obtain fluctuation function depending on size of the window. In every of methods undermentioned equality holds:

$$F(s) \sim s^H.$$

F(s) is fluctuation function, s is timescale and H is scaling exponent. In order to obtain H parameter we need to calculate the slope of fluctuation function versus timescale on double logarithmic scale.

1.1 Fluctuation Analysis (FA)

This method is not based on detrending. This is the simples of methods from the programing point of view. Fluctuation function is given by formula:

$$F^2(s) = \overline{\Delta y(s)^2} - \overline{\Delta y(s)}^2,$$

where

$$\Delta y(s) = y(s_0 + s) - y(s_0).$$

Having this we need to progress as was mentioned above. Just repeat calculating fluctuation function for different values of s and obtain value of slope on double logarithmic scale.

1.2 Detrended Fluctuation Analysis (DFA)

In each box (interval) we need to determine a trend function. After this we need to determine residual function:

$$\varepsilon(t) = y(t) - g(t).$$

Where y(t) is series and g(t) is trend function. In case of detrending fluctuation function it is polynomial function. In case of this report the second order polynomial were chosen. Fluctuation function is given by the following formula:

$$F(s) = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \varepsilon(t)^2}.$$

1.3 Detrended Moving Average (DMA)

In this case procedure is almost the same as for Detrended fluctuation analysis. The only difference is that instead of fitting and subtracting polynomial we are subtracting a moving average function.

2 Analysis

In case of every measurements we have to remember about the errors which can be caused by facility we are using. In case of computational calculations and especially in case of some approximations algorithms we have to be aware of errors which results directly from the approximation model. But in both case, measurements (physical, biological, chemical etc.) and results of computations have to be normally distributed. In other case, there is a possibility of making some very expensive mistake.

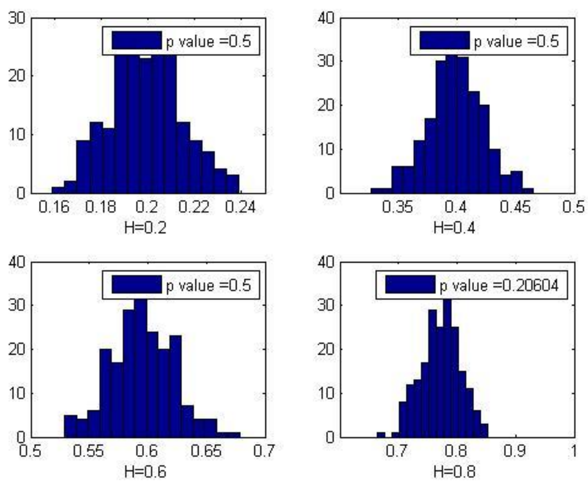

Figure 2.1 Histogram for FA.

In order to check if prepared algorithms approximate correctly the distribution of outcomes were checked. For initial value of Hurst exponent equal to 0.2, 0.4, 0.6 and 0.8 computations were repeated 200 times. For observations in each group of Hurst exponent and for each method the Lillie test was performed to check whether the distribution pass normality test. Each figure presents p-value of Lillie test and histograms. As we can see in each case approximations pass normality test. Figures 2.1, 2.2 and 2.3 present the distribution of approximations of Hurst exponent calculated via fluctuation analysis, detrended fluctuation analysis and detrended moving average respectively.

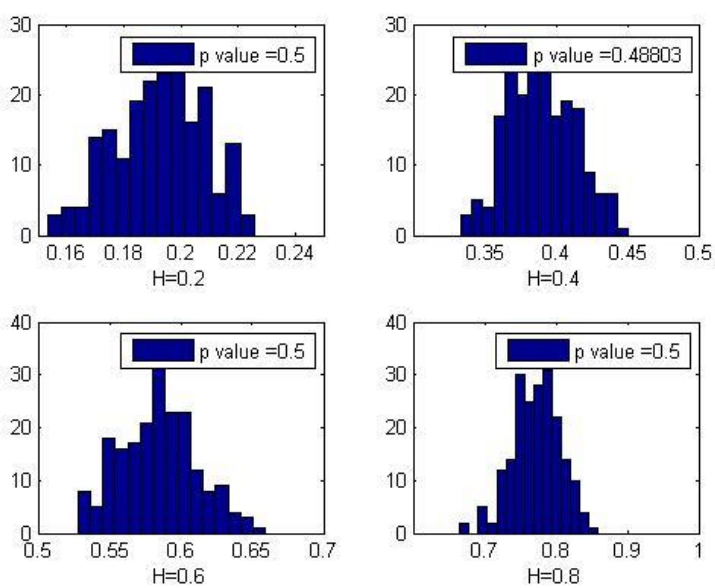

Figure 2.2 Histograms for DFA.

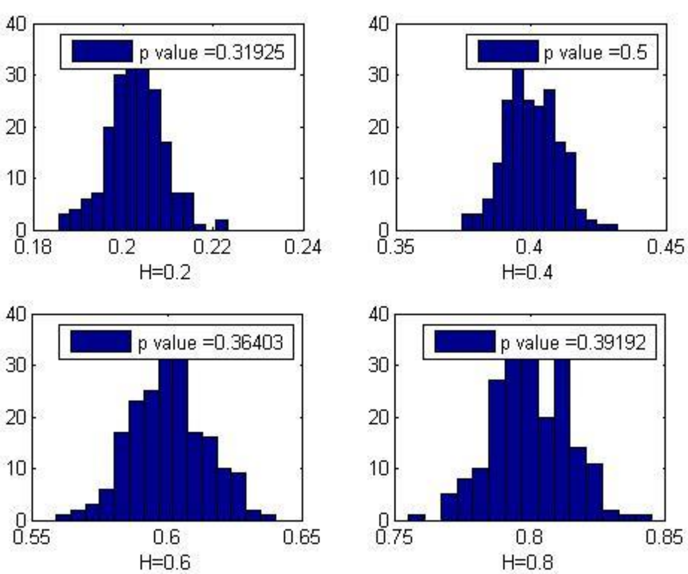

Figure 2.3 Histograms for DMA.

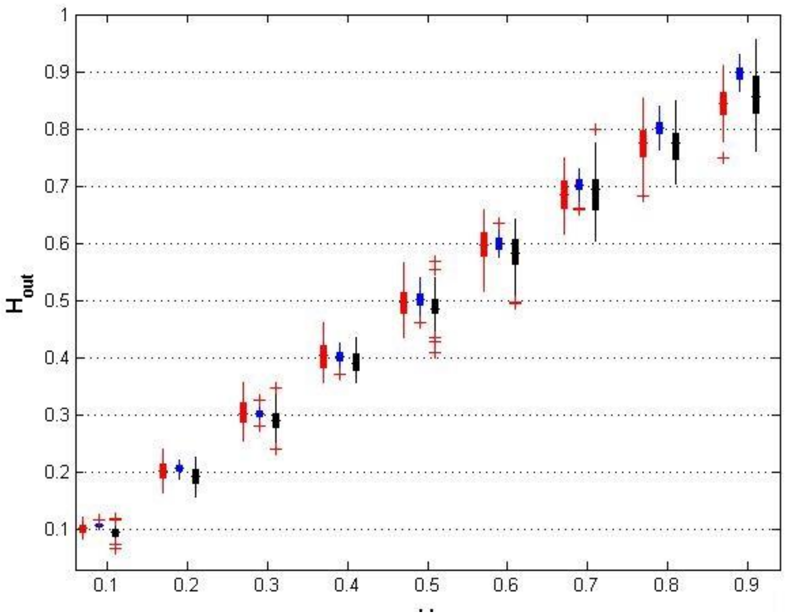

Figure 2.4 Boxplots of FA (red), DFA (blue) and DMA (black).

Figure 2.4 presents boxplots for three methods. As we can see for increasing values of Hurst exponent the variance of measurements increases. With no hesitation we can say that detrended fluctuation analysis approximate with the best accuracy even for large values of Hurst exponent. As we can see fluctuations analysis gives just slightly worst results than detrending moving average.

Figure 2.5 presents mean relative error for all of methods. It corresponds to results obtained in figure 2.4. as we can see for almost all the cases DFA produces the best results. FA and DMA produce similar results. For H < 0.75 FA is slightly better than DMA and for H > 0.75 FA gets worst results than DMA. Figure 2.6 presents mean square error. All results corresponds to those obtained before.

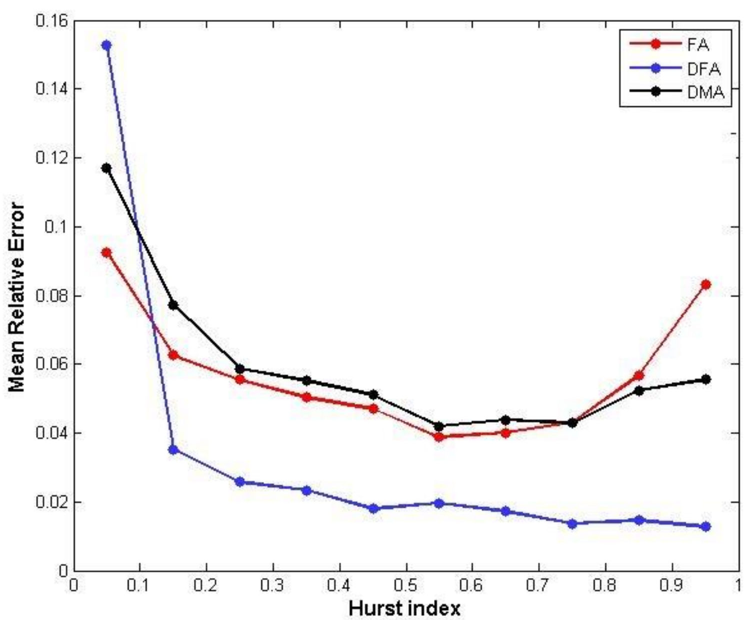

Figure 2.5 Mean relative error for all methods.

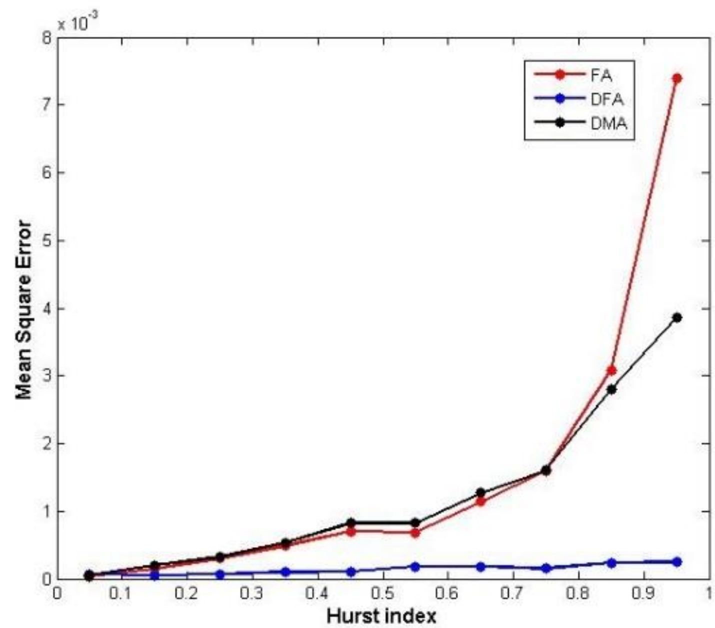

Figure 2.6 Mean square error for all methods.

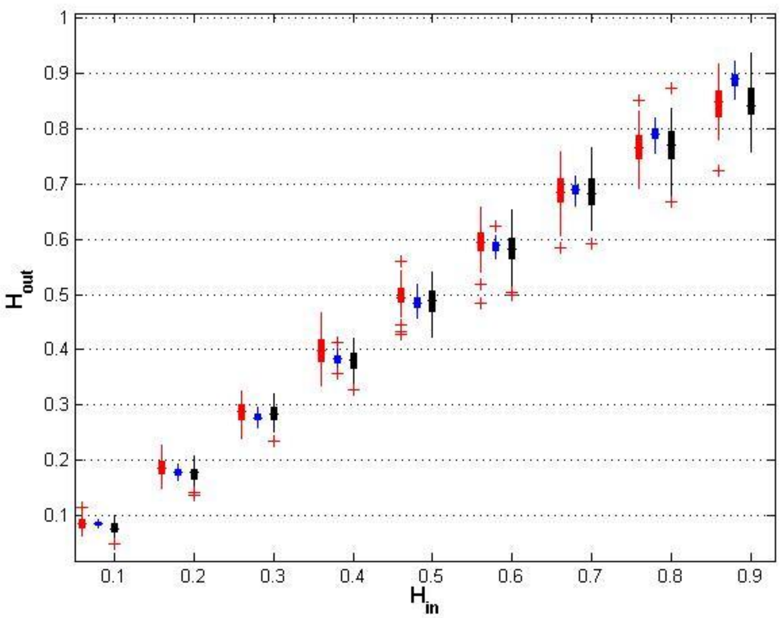

Figure 2.7 Boxplots of FA (red), DFA (blue) and DMA (black) with added N (0, 0.5) noise.

For a real life data we have to remember that nothing is pure in statistical sense. Each measurement brings some error or there might be present some noise. That is why it is important to check weather methods work with data into which noise was added. Figure 2.7 presents boxplots of data with random normal noise with mean 0 and variance 0.5. As we can see results still maintaine the same trend. DFA gives the best results but every method gets worst for increasing value of H.

Figure 2.8 and 2.9 presents mean relative and mean square error respectively. As we can see DFA remains the best method for almost all values of H. the goodness of FA and DMA is still changing above H > 0.75.

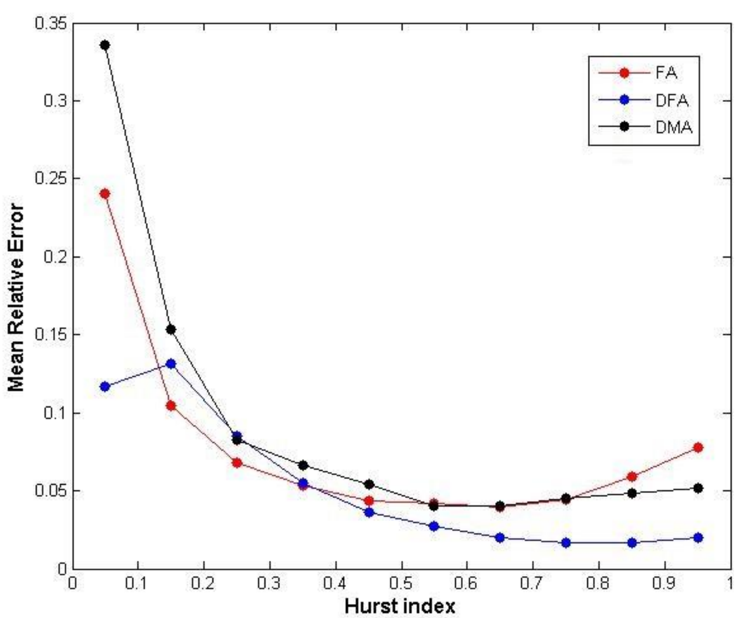

Figure 2.8 Mean relative error for all methods with added N(0, 0.5) noise.

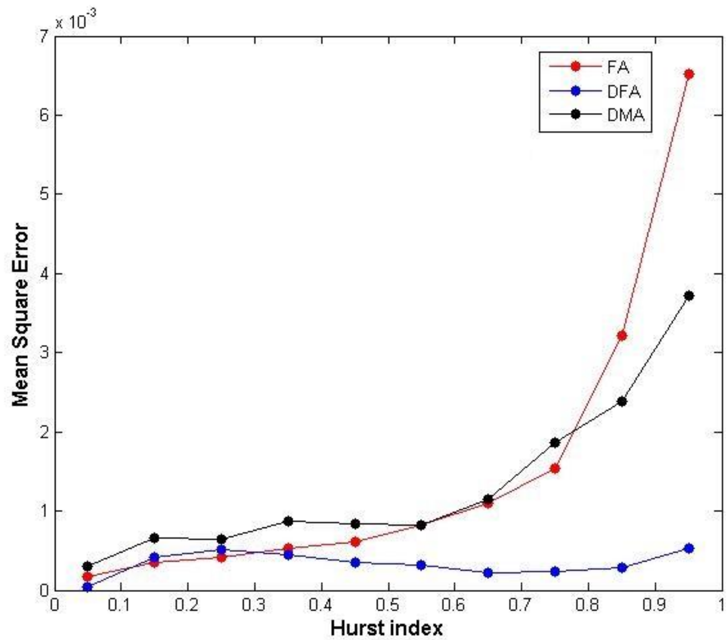

Figure 2.9 Mean square error for all methods with added N(0, 0.5) noise.

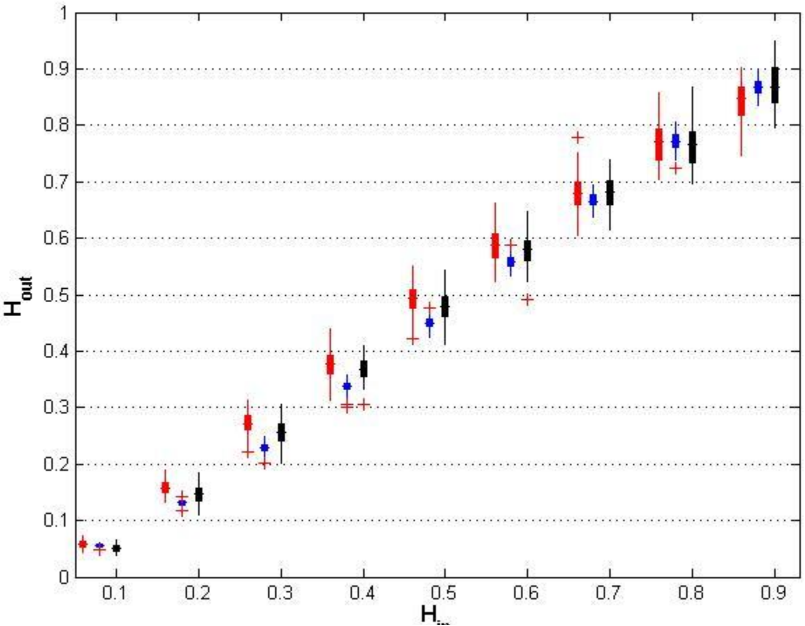

Figure 2.10 Boxplots of FA (red), DFA (blue) and DMA (black) with added N (0, 1) noise.

Figure 2.10 presents boxplots of data with N (0, 1) noise. As we can see approximations get worst. But we cannot say for sure which method is the best in which interval. This judge figures 2.11 and 2.12 which present mean relative and mean square error respectively for data with noise. As we can see for H < 0.75 DFA produces worse results than FA and DMA. For extreme values of H DFA approximate better than other methods. The tendency with FA and DMA still holds. For H < 0.75 FA produces better results and for H > 0.75 MSD is more accurate.

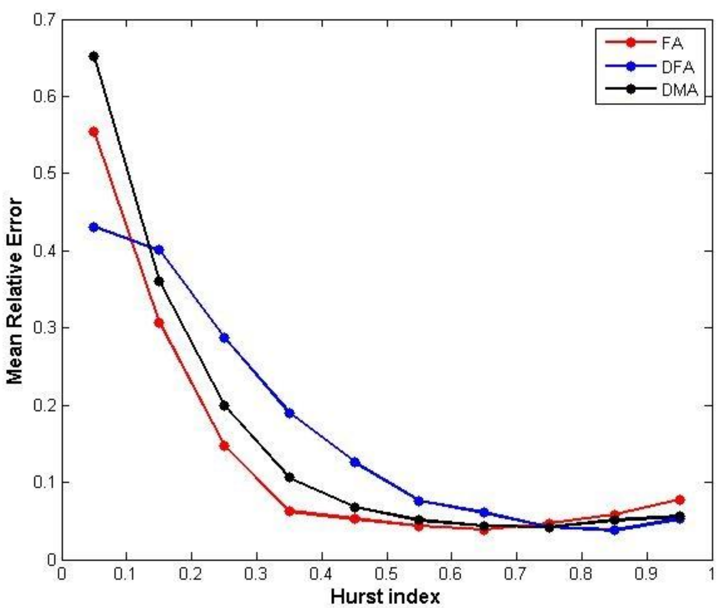

Figure 2.11 Mean relative error for all methods with added N(0, 1) noise.

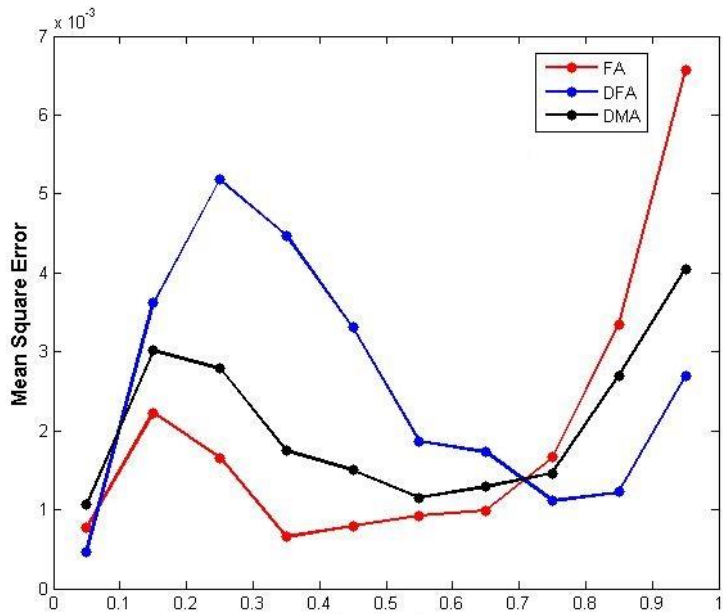

Figure 2.12 Mean square error for all methods with added N(0, 1) noise.

3 Bibliography

Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series, Ying-Hui Shao et.al.

Second-order moving average and scaling of stochastic time, E. Alessio et.al.

Long-range correlations in nucleotide sequences, Ch. K. Peng et al.