Local Scaling Analysis with Continuous DFA in High Frequency Financial Time Series

M. Chorowski[a,*] and Z.R. Struzik[a,b,c]

[a]*Faculty of Physics, University of Warsaw,*
L. Pasteura 5, PL-02093 Warsaw, Poland
[b]*Advanced Center for Computing and Communication,*
RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[c]*Graduate School of Education, the University of Tokyo,*
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Doi: 10.12693/APhysPolA.139.407   *e-mail: ma.chorowski@student.uw.edu.pl

The crucial question every investor has to answer before buying stock is: what is the risk associated with it? In this work, we applied the Continuous Detrended Fluctuation Analysis (CDFA) to uncover the point-wise Hölder exponent of the time series at various time scales. The Hölder exponent is a time-dependent measure of the autocorrelation of the variability in data, and it can be used to measure the market risk. CDFA is an extension of the popular Detrended Fluctuation Analysis (DFA) method, but — contrary to DFA — it allows for estimating the Hölder exponent locally. We used this method to study the two-minute frequency data of the U.S. S&P 500 index from 1984 to 1995. We found that the value of the exponent increases as lower resolution data was used, which means the financial time series appear more predictable at higher scales than at lower scales. Moreover, we found that correlations between the Hölder exponent and the log returns and absolute log returns are scale-dependent. That means that the CDFA method could potentially be used in the future to identify time scales at which the series are the most predictable which can have practical applications.

topics: econophysics, financial time series, Hölder exponent, detrended fluctuation

1. Introduction

Economic systems are immensely complex, and their analysis must rely mostly on time series data. The dynamic of economic or financial time series is often described by a power scaling. A time series characterized by a single scaling exponent is called a fractal, and a multifractal if multiple different scaling exponents are required to describe its behavior [1]. In the case of stochastic processes, we call a process X self-similar if $X(ct) \sim c^H X(t)$ in terms of probability distributions and Hurst exponent H measures then the self-similarity and the intensity of long-range dependence [2]. The local scaling behavior of time series can also be measured using the Hölder exponent. It is defined [3] as the supremum of all the alphas of the kind that

$$|f(x) - f_n(x - x_0)| \leq C |x - x_0|^\alpha, \tag{1}$$

where $f_n(x)$ is the polynomial of degree $n < \alpha$.

Notably, both the Hölder and the Hurst exponents describe the same system behavior. The difference, however, is that Hurst is related to the global scaling of the time series, while Hölder is a local point-wise measure. We further show that the exponents are indeed related by comparing the time-averaged Hölder with the known Hurst exponent of artificially generated time series.

Studying the long-term memory of the time series is important in many fields and applications. In this paper, however, we focus on its application to financial time series. The scaling exponent plays an important role in the behavior of financial markets. It can reflect the randomness (persistence or anti-persistence) of prices. Moreover, a lower Hurst exponent is tied to higher volatility of prices [4]. This is true both mathematically and when the risk is assessed by humans [5]. The scaling exponent plays a role in the Value-at-Risk (VaR) calculations, as not accounting for $H \neq 0.5$, it can lead to risk under- and overestimation in cases when the VaR value of d days is derived by multiplying the one-day VaR value by $\sqrt{d}$ [6]. The Hurst exponent has often been applied in the study and prediction of the financial time series and in the classification of FOREX Securities [7]. It has been found that periods in the foreign exchange markets with

a large Hurst exponent are more predictable [9]. Also, in some FX Markets, the Hurst exponent changed dynamically over time [9]. This suggests that a time-dependent scaling exponent might help assess the predictability and riskiness of the financial time series.

The Detrended Fluctuation Analysis (DFA) is a popular method for estimating the statistical self-similarity of non-stationary time series. It was first introduced by Peng et al. [10]. Contrary to Hurst's Rescaled Range (R/S) Analysis [11], DFA can be used in the presence of extrinsic non-stationarities due to the detrending operation. The protection against the non-stationarities is, nonetheless, disputed [12]. Still, DFA performs well — as compared to other methods — when tested against a simulated series of fractional Gaussian noise and fractional ARIMA(0,d,0) [2]. It has also been generalized, and extended to study multifractals [13], or multivariate time series.

As already outlined, earlier studies suggest the scaling exponent changes dynamically in time, and its value is tied to the predictability of the time series. However, the traditional DFA method does not allow for a local (point-wise) scaling analysis. In this paper, we aim to present, study, and apply the method of obtaining the point-wise time-dependent scaling exponent to financial time series. This method, called the Continuous Detrended Fluctuation Analysis (CDFA), was first proposed by Struzik [3] in 2004. As far as we know, it has not been used yet to study the financial time series which we do in this paper. It has to be mentioned that there are also alternatives to CDFA. For example, a similar detrending moving average (DMA) method was used to study the time-dependent Hurst exponent $H(t)$ of the German market on a minute scale [14]. Another possibility to extract local scaling exponents is to use the wavelet transform based methods [15, 16].

In our paper, we have provided the analysis of the scaling exponent of the high-frequency (two minutes) S&P 500 index returns at an arbitrary scale and time. A similar analysis for this index was done using the wavelet transform based method [17]. In our study, a different method is used, allowing thus the comparison of the results. To calculate the Hölder exponents $h(t)$, we only used the data available up to time t. In other words, we did not use future data to calculate the scaling exponent at any time t. Such a case would be important for economic analysis when, for example, one is interested in predicting future volatility, or constructing a predictor of economic crashes. Therefore, our results can be potentially useful in practical applications. In addition, we studied the relationship between the estimated scaling exponent and a different measure of volatility. In the future, those relationships might be potentially used for building risk estimators for various time scales.

2. Methods

To perform our analysis, we used the S&P 500 index two minute $\Delta t = 2$ min data. First, the Index value was transformed according to

$$P(t) = \log\left(100 SPX(t)\right), \tag{2}$$

where $SPX(t)$ is the raw index value. The transformed data is shown in Fig. 1.

To obtain the exponents at various time-scales, we transformed the data in the following way. First, a lagged difference of scale s was taken as in

$$R(t,s) = P(t) - P(t-s). \tag{3}$$

As shown earlier in (2), $P(t)$ is a logarithm thus $R(t,s)$ is a log return with lag s. Next, we transform the log returns back to a random walk using

$$P(t,s) = \sum_{i=1}^{t} R(t,s). \tag{4}$$

This step is important, since the Detrended Fluctuation Analysis is normally done on the unbounded process.

Now, the CDFA was performed on the index $P(t,s)$ data for various resolutions ϵ (possible window sizes). We keep the window sizes constant at all scales, but the scale of the analyzed time series varies. A different approach would be to keep the scale of the time series constant and instead vary the possible window sizes ϵ to extract the scaling exponent at different scales. Obviously, the scaling exponents calculated in those two approaches are not equivalent and would be different. In this work, we focus only on the first approach, that is, we study how the scaling relations change when the resolution of the underlying data changes.

The extraction of the point-wise scaling exponent $h(t,s)$ at a chosen scale s goes as follows:

1. Pick time t.

   - Pick a window size ϵ, starting with $\epsilon_{\min}$.
   - Fit a local polynomial $f_n(t)$ of order n on the $R(t)$ data between $(t-\epsilon)$ and t.
   - Calculate the mean fluctuations in the analyzed window

$$F(\epsilon) = \frac{1}{\epsilon} \sum_{i=t-\epsilon}^{t} |R(i) - f_n(i)|. \tag{5}$$

2. Pick a larger window size $\epsilon + 1$ and go back to item (a) until $\epsilon_{\max}$ is reached.

3. Perform linear fit in log–log coordinates of $F(\epsilon)$ versus ϵ. The slope is the point-wise scaling exponent $h(t)$.

4. Pick the next data point $t = t + 1$.

For our analysis, we have chosen $\epsilon_{\min} = 50$ and $\epsilon_{\max} = 500$, which correspond to windows of sizes from 100 to 1000 min for the minimum scale $s = 1$, and goes up to windows of sizes from 24000 to 240000 min for scale $s = 240$. It has to be underlined that the number of observations in a window
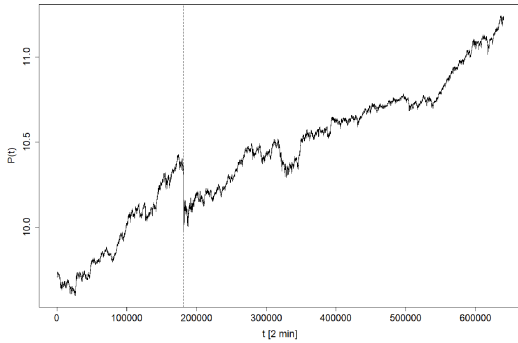
Fig. 1. The plot shows the empirical two-minute S&P 500 index transformed (with (2)) data. The lowest point of the Black Monday crash of 1987 is marked with the vertical line.
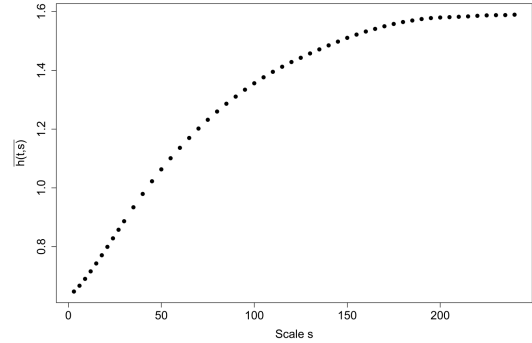

Fig. 2. The Hölder exponent values are represented by color, with lower values going towards red, and higher — towards yellow. The exponents for a given scale s were calculated using the data transformed with (4). The horizontal axis represents time from the beginning of the dataset. The vertical axis is the scales, ranging from 1 to 240 (2 to 480 min).

is independent of scale s in the current approach and is always between 50 and 500 observations in every scale. Different, possibly larger windows can be used in the future. For $f_n(t)$, we have picked a first-order polynomial.

In some cases, the linear fit in log–log coordinates (see item No. 2 of the algorithm procedure) was very poor, and the function was very far from linear. For those points, we removed the outlying Hölder exponents by choosing the threshold. It is defined as

$$\chi^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} > 0.3. \tag{6}$$

The threshold value applied to remove outliers that worked optimally in our tests is chosen to be 0.3. If χ^2 of the fit for some point (t, s) was larger than 0.3, then the calculated exponent $h(t, s)$ was removed from further analysis. Note that the value of the threshold is arbitrary. In general, a value which is too high will cause the removal of too many points and a threshold which is too low will fail to correctly identify the outliers.

The algorithm we used is the left-side CDFA. It is also possible to extract the Hölder exponent using windows centered on the analyzed points [3]. We employed the left-side version since usually in practical circumstances — when analyzing financial time series and predicting volatility — the future is not known, and a centered window could not be used.

3. Results and discussion

Firstly, we checked if our algorithm gives correct results on artificial data. We have generated sample paths (using R package *somebm* [18]), each consisting of twenty thousand observations, for each of the following Hurst exponents $H = 0.3, 0.5, 0.7, 0.9$. We then calculated the mean of the Hölder exponents as described in Sect. 2 and compared them to the known theoretical ones that were used to generate the random walks. Indeed, the mean exponents values $\overline{h(t)} = 0.31, 0.51, 0.70, 0.87$ were close to the theoretical ones. Since CDFA gives

point-wise exponents, their variability is quite significant. The standard deviation was, respectively, $SD(h(t)) = 0.17, 0.23, 0.30, 0.34$. Those results could probably be further improved with a longer sample and a larger window size $\epsilon_{\max}$.

After testing the algorithm, we studied our empirical data. In Fig. 1, we present the empirical price data we used. We marked the 1987 Black Monday crash with a vertical line. We used high-frequency, two-minute data. In that way, we were able to conduct the analysis, and extract the Hölder exponents at multiple scales s. The data at various scales were given by (4). We calculated the Hölder exponents at scales ranging from $s = 3$ to $s = 240$. The results are presented as a heatmap in Fig. 2. Each point in the heatmap is a calculated Hölder exponent at a given scale and time. White color marks the points we removed from the analysis due to the poor fit (as with (6)). The number of removed points depended on the scale, but was generally low. With the lowest number of outliers, $\approx 0.1\%$ of our dataset for $s = 3$, and the highest number of outliers, about 5% of data for $s = 185$.

No clear pattern is visible for long scales. As expected, the exponents are arranged in vertical lines which indicate that the shocks in the time series are visible across multiple scales. There is a shift to red (lower $h(t)$ values) at shorter time scales, meaning that the returns are more volatile at shorter time scales. However, we observe high values of $h > 2$ for almost all time scales. It might indicate the presence of local correlations at various times and resolutions (scales). A higher scaling exponent can improve the quality of forecasting methods [8]. Thus, our method might be potentially used in the future to build short-term (point-wise) predictors.

The shift from red to yellow as we move to higher scales is confirmed in Fig. 3, where the time-averaged Hölder exponent $\overline{h(t)}$ at different scales is presented. A higher scale means that the input
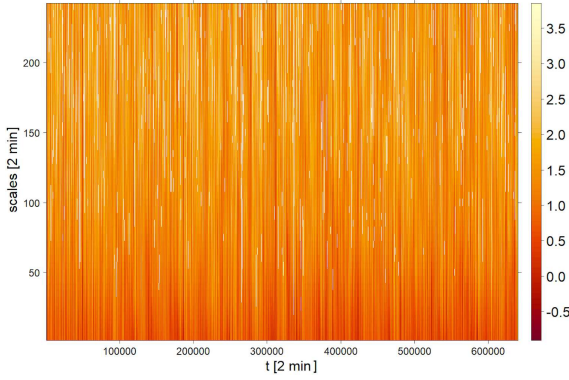
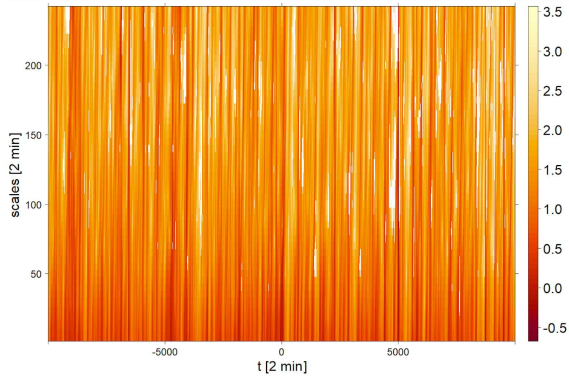Fig. 3. The plot shows the time-averaged Hölder exponent as a function of scale s.

Fig. 4. The Hölder exponent values are represented by color, with lower values going towards red, and higher — towards yellow. The exponents for a given scale s were calculated using the data transformed with (4). The horizontal axis represents time from the beginning of the dataset. The vertical axis is the scales, ranging from 1 to 240 (2 to 480 min). The heatmap is centered on the lowest point of the 1987 Black Monday crash.

data is a smoother result of the transformation defined in (3) and (4). Hence, the corresponding exponents tend to be higher as the time series smooths out. That means that the random walk time series and its trends become more predictable at higher scales (even though the increments probably still tend to be unpredictable). This suggests that the unpredictability of the system is visible only in high resolutions. In other words, looking at it only at high resolutions (for example daily) might give an impression that the dynamic of the system is predictable, while at lower scales the series remain unpredictable. Moreover, our results suggest that the financial time series might be more predictable at higher time scales where long-term trends are visible. That is consistent with how the evolution of stock prices is usually modeled, for example using the Geometric Brownian Motion with constant drift.

Next, we wanted to look further into the behavior of the exponents around times of high market stress. We present those exponents in Fig. 4

centered on the lowest point of the 1987 Black Monday Crash. The heatmap unveils a characteristic pattern of vertical lines which indicate that the volatility of the underlying time series can be visible under many different scales. Especially, as time approaches the Black Monday Crash, the heatmap turns red at various scales, indicating severe irregularity and volatility of returns.

We wanted to study closer the relationship between the volatility and the estimated Hölder exponent $h(t, s)$ at different scales s. We used the absolute log returns $|R(t, s)|$ as the measure of volatility (absolute value obtained with (3)) since when measuring volatility we are interested in how the series deviates from the mean in time, and not necessarily in the sign of the changes.

In Fig. 5, we presented the correlations between the Hölder exponent $h(t, s)$ and the absolute log returns $|R(t, s)|$ at various scales. One can read therefore that the sign of the correlation changes at some scale. The correlation is initially negative, as expected, which indicates significant irregularity of the time series. The correlation changes sign around $s = 205$ (410 min). Increasing s means that the log differences are taken over longer time periods, as defined in (3). This only means that $|R(t, s)|$ tends to increase with s, while the time series described with (4) itself becomes smoother. At the same time, as shown in Fig. 3, in higher scales the Hölder exponent becomes saturated. Above $s = 205$, a further increase in s does not affect $h(t, s)$ significantly, but $|R(t, s)|$ will still increase. That means that increasing s above 205 does not significantly smooth the time series any further. Hence, we observe a positive correlation between $|R(t, s)|$ and $h(t, s)$ in Fig. 6.

To understand our results better, we looked at the correlations between the Hölder exponent $h(t, s)$ and the log returns $R(t, s)$ in Fig. 6. The correlation is positive across all scales. It implies that on average the prices are more likely to increase than
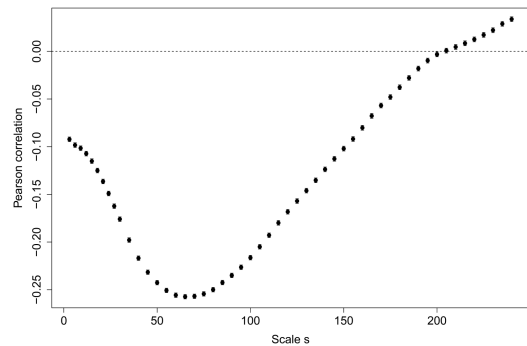

Fig. 5. The vertical axis shows the Pearson correlation value between the Hölder exponent and the absolute log returns at a given scale $|R(t, s)|$ (horizontal axis). The bars represent the 95% confidence intervals.
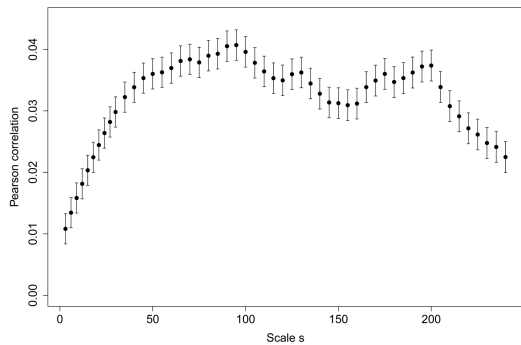
Fig. 6. The vertical axis shows the Pearson correlation value between the Hölder exponent and the log returns at a given scale $R(t, s)$ (horizontal axis). The bars represent the 95% confidence intervals.

decrease during times of low volatility (times when the market is described by high h). It has to be noted that the correlations are very small, but statistically significant due to the very large sample.

In both cases, the correlation of $h(t, s)$ with log returns $R(t, s)$ and absolute log returns $|R(t, s)|$ is obviously non-linear in scale s which suggests that the relationship between the local scaling exponent and the scale at which we analyze the data is not trivial. Most importantly, the correlation between the exponent and returns has a maximum in analyzed scales around $s = 95$ (190 min) and is in general scale dependent. That has important implications for practical investors. Our results in Fig. 6 show that the investor should trade at the time scale at which the predictability (measured by the correlation between $h(t, s)$ and $R(t, s)$) of the time series is the highest. Our method could be helpful in identifying the optimal time scales at which the investor is most likely to be able to predict the direction of the prices.

Given the observed correlations of the Hölder exponent with the absolute log returns at various scales, this measure might be potentially used in the future as an additional indicator of market risk. A great advantage of the CDFA method is its speed. However, before it could be applied in finance in practice, much more work needs to be done. First of all, the estimated Hölder exponent was often above one. It is not uncommon for the DFA to find the scaling exponents above one. It might, however, indicate that the detrending of non-stationarities was not entirely successful [12]. In the future, perhaps a higher polynomial used in detrending could be helpful. Furthermore, the most important parameters that should be investigated in the future are $\epsilon_{\min}$ and $\epsilon_{\max}$, the minimum and the maximum window size. It is possible that the chosen window sizes were too small to accurately estimate the Hölder exponent. What is more, the analysis at various scales was done by using different lags of returns and transforming the differences back to a random walk (4). A different approach would be

to keep the original SPX data and instead change the possible window sizes $\epsilon_{\min}$ and $\epsilon_{\max}$ to extract the exponents at different scales. Those issues and aspects could be further studied in the future.

4. Conclusions

We used the Continuous Detrended Fluctuations Analysis to extract the point-wise Hölder exponent from the high-frequency S&P index data at different scales. The Hölder exponent measures the autocorrelation of noise and can be used to measure risk. A higher value of the exponent corresponds to more regular functions, and low values mean that a function is more irregular. Therefore, the risk of losses may appear to be predictable (due to high h) at least in a point-wise fashion discussed here — the fact probably not used to date.

The correlation showed that, on average, the estimated Hölder exponent was negatively correlated with the absolute log returns, and positively correlated with the returns. Interestingly, the exponent showed saturation at higher scales and was not increasing further after $s = 205$ (410 min), and its correlation with the absolute log returns changed its sign.

Our results show that since on average the exponent h grows with the scale at which we look at the data, the system may appear to be unpredictable and uncorrelated only for the highest resolutions (low s). It might serve as a warning against using only low resolution data. At the same time, it is a hint that long-term trends give information about predictability. Finally, due to the dependence of the predictability of the time series on scale, our method could be used as a tool helping investors find optimal time frames at which they are most likely to make correct predictions. Of course, the practicality of this approach requires further testing.

Acknowledgments

References

[1] H. Xiong, P. Shang, *Commun. Nonlinear Sci. Numer. Simul.* **42**, 12 (2017).

[2] M.S. Taqqu, V. Teverovsky, W. Willinger, *Fractals* **3**, 785 (1995).

[3] Z. Struzik, "Local Scaling Analysis with Continuous DFA Method", CWI Research Report, 2003.

[4] B. Mandelbrot, R.L. Hudson, *The Misbehavior of Markets: A Fractal View of Financial Turbulence*, Basic books, 2007.

[5] D. Sobolev, N. Harvey, *Risk Anal.* **36**, 2216 (2016).

[6] J. Miller, D. Edelman, J. Appleby, *Numerical Methods for Finance*, CRC Press, 2007.

[7] M.S. Raimundo, J. Jr. Okamoto, *Int. J. Model. Optim* **8**, 116 (2018).

[8] R. Szóstakowski, *Przegląd Statystyczny* **65**, 200 (2018).

[9] M. Corazza, A.T.G. Malliaris, *Multinatl. Finance J.* **6**, 65 (2002).

[10] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).

[11] H.E. Hurst, *Trans. Amer. Soc. Civil Eng.* **116**, 770 (1951).

[12] R. Bryce, K. Sprague, *Sci. Rep.* **2**, 315 (2012).

[13] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley , *Physica A* **316**, (2002).

[14] A. Carbone, G. Castelli, H.E. Stanley, *Physica A* **344**, 267 (2004).

[15] Z.R. Struzik, *Fractals* **8**, 163 (2000).

[16] P. Goncalves, P. Abry, in: *1997IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Munich (Germany) 1997*, Vol. 5, 1997, p. 3433.

[17] Z.R. Struzik, *Physica A* **296**, 307 (2001).

[18] J. Huang, somebm: some Brownian motions simulation functions, 2013, R-package ver. 0.1.