Department of Economics and Finance

Working Paper No. 2215

Economics and Finance Working Paper Series

Guglielmo Maria Caporale, and Alex Plastun

Persistence in High Frequency Financial Data

October 2022

http://www.brunel.ac.uk/economics

PERSISTENCE IN HIGH FREQUENCY FINANCIAL DATA

Guglielmo Maria Caporale
Brunel University London

Alex Plastun
Sumy State University

October 2022

Abstract

This paper investigates persistence in high-frequency, intraday data (and also daily and monthly ones) in the case of the EuroStoxx 50 futures over the period from 2002 to 2018 (720 million trade records) using R/S analysis and the Hurst exponent as a measure of persistence. The results indicate that persistence is sensitive to the data frequency. More specifically, monthly data are highly persistent, daily ones follow a random walk, and intraday ones are anti-persistent. In addition, persistence varies over time. These findings imply that the Efficient Market Hypothesis (EMH) only holds in the case of daily data, whilst it is possible to make abnormal profits using trading strategies based on reversal strategies at the intraday frequency.

Keywords: *Persistence, Long Memory, R/S Analysis, high-frequency data*

JEL Classification: *C22, G12*

Corresponding author: Professor Guglielmo Maria Caporale, Department of Economics and Finance, Brunel University London, Uxbridge, UB8 3PH, UK. Email: Guglielmo-Maria.Caporale@brunel.ac.uk; https://orcid.org/0000-0002-0144-4135

Alex Plastun gratefully acknowledges financial support from the Ministry of Education and Science of Ukraine (0121U100473).

1. Introduction

Since the seminal paper by Fama (1970) a huge number of studies have examined empirically the Efficient Market Hypothesis (EMH) according to which asset prices should follow a random walk and thus not exhibit long memory, trends, or mean-reversion. Many papers have found instead that (high) persistence is a typical property of financial data (Lo, 1991; Mynhardt et al., 2014; Bariviera, 2017; Caporale et al., 2018; Phillip et al., 2018 and many others). Most of this evidence is based on daily, weekly or monthly data and suggests that the data frequency matters; for example, Caporale and Gil-Alana (2010) showed (through Monte Carlo experiments and an empirical application to the S&P500) that it affects the estimates of the fractional integration parameter measuring persistence; more precisely, if the true DGP is I (d) with d lying between 0 and 1, d tends to be underestimated at lower frequencies, and this bias tends to be bigger for bigger deviations from 0 or 1. In another paper, Caporale et al. (2019) investigated persistence in financial time series at daily, weekly and monthly frequencies for various financial markets (stock markets, FOREX, commodity markets) over the period from 2000 to 2016 using both R/S analysis and fractional integration; their results indicate that in most cases persistence is higher at lower frequencies, for both returns and their volatility, which is inconsistent with the EMH and implies that abnormal profits can be made by using trading strategies based on trend analysis.

A few studies have also analysed long-memory behaviour at higher frequencies. For instance, Caporale and Gil-Alana (2013) focused on the US dollar/British exchange rate and found several cases of mean-reverting behaviour when the data are collected every 10 minutes whilst for even higher frequencies the unit root null cannot be rejected; in other words, persistence tends to be lower at higher frequencies characterised by more noise in price dynamics.

The present study goes further by measuring persistence in the case of data collected at a much higher frequency (namely, micro seconds) using R/S analysis based on the Hurst exponent method; specifically, the series examined is the EuroStoxx 50 futures prices over the period 2002-2018. The analysis is then repeated at lower (monthly and daily) frequencies to examine whether persistence is sensitive to the data frequency, and finally a sliding-window approach is used to investigate whether it varies over the time.

The layout of the paper is the following: Section 2 briefly reviews the relevant literature; Section 3 describes the data and outlines the empirical methodology; Section 4 presents the empirical results; Section 5 offers some concluding remarks.

2. Literature Review

Numerous papers have analysed persistence in various financial assets such as stocks (Lo, 1991; Los, 2006), exchange rates (Kim and Yoon, 2004; Da Silva et al., 2007), commodity prices (Alvarez-Ramirez et al., 2002; Serletis and Rosenberg, 2007), and cryptocurrencies (Urquhart 2016; Bouri et al., 2016; Bariviera et al., 2017; Caporale et. al, 2018), often providing international evidence (e.g., Jacobsen, 1995; Lento, 2009; Zunino et al., 2009; Niere, 2013). A variety of methods have been applied to estimate persistence: R/S analysis (Glenn, 2007; Lento, 2009; Caporale et. al, 2018), fractional integration (Caporale and Gil-Alana, 2013), the generalized Hurst exponent approach (Barunik and Kristoufek, 2010), detrended moving average (Grech and Mazur, 2005), multifractal generalization (Kantelhardt et al., 2002), detrended fluctuation analysis (Taqqu et al., 1995), etc.

Most studies have used daily data (e.g., Zunino et al., 2009; Niere, 2013), considerably fewer have focused on weekly (MacDonald and Taylor, 1992) or monthly (Caporale et al., 2019) ones, and an even smaller number on high-frequency ones. In particular, Andersen and Bollerslev (1997) analysed persistence in 5-minutes returns in the FOREX and US stock market and found long-memory properties in their volatility, whilst

Cotter (2005) reported similar findings in the case of UK futures using data at 5-minute intervals; finally, as already mentioned, Caporale and Gil-Alana (2013) examined high frequency data (collected every 1, 2, 3, 5 and 10 minutes) on the US dollar-British pound spot exchange rate and found lower persistence at higher frequencies; this is consistent with Caporale et al. (2019), who investigated persistence at three different frequencies (daily, weekly and monthly) in various stock, FOREX and commodity markets using both R/S analysis and fractional integration and reached the same conclusion.

It is noteworthy that the highest frequency considered by the papers discussed above is 5 minutes, none of them investigating persistence at higher frequencies. The present study aims to fill this gap by providing evidence on persistence in the case of data collected at micro seconds as detailed below.

3. Data and Methodology

High-frequency data on EuroStoxx 50 futures prices over the period 2002-2018 are used for the empirical analysis; they consist of 720 million trade records being collected every tenth of a second or even more frequently. Daily and monthly series are also examined. The data source is Eurex Exchange, the leading platform for Eurozone equity and equity index derivatives (https://www.eurex.com/ex-en/).

The measure of persistence used is the Hurst exponent estimated by carrying out R/S analysis. This approach dates back to Hurst (1951) and was extended by Mandelbrot and Wallis (1969) and Mandelbrot (1972). Despite newer methods having been developed since then (such as detrended fluctuation analysis (DFA), multifractal generalization (MF-DFA), stabilogram diffusion analysis (SDA), and others), R/S analysis remains very popular and is very often used for the purpose of estimating persistence in financial data (Mynhardt et al., 2014; Raimundo and Okamoto, 2018; Danylchuk et al., 2020; Metescu, 2022). The rationale for choosing this method is that it is relatively simple and suitable for programming as well as visual interpretation, and it appears to capture accurately the properties of the data.

The Hurst exponent (H) lies in the interval [0, 1]. Persistence is found when $H > 0.5$. Random data are characterised instead by $H = 0.5$. Anti-persistence is detected when $H < 0.5$. The algorithm for R/S analysis is constructed as follows. For each sub-period range R (the difference between the maximum and minimum index within the sub-period), the standard deviation S and their average ratio are calculated. The length of the sub-period is increased and the calculation repeated until the size of the sub-period is equal to that of the original series. As a result, each sub-period is determined by the average value of R/S. The least squares method is applied to these values and a regression is run, obtaining an estimate of the slope of the following regression: $log\ (R / S) = log\ (c) + H^{*}log\ (n)$. This estimate is a measure of the Hurst exponent, which is an indicator of market persistence. More details are provided below.

1. One starts with a time series of length M and transform it into one of length $N = M - 1$ using logs and converting prices into returns (or volatility):

$$N_i = \log\left(\frac{Y_{t+1}}{Y_t}\right), \quad t = 1,2,3,\ldots (M-1) \qquad (1).$$

2. One then divides this period into contiguous A sub-periods with length n, such that $A_n = N$, then each sub-period is defined as I_a for $a = 1, 2, 3. . . , A$. Each element I_a is represented as N_k with $k = 1, 2, 3. . . , N$. For each I_a with length n the average e_a is defined as:

$$\mathrm{e_a} = \frac{1}{\mathrm{n}} \sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{N_{k,a}}, \quad \mathrm{k} = 1,2,3,\ldots\mathrm{N}, \quad \mathrm{a} = 1,2,3\ldots,\mathrm{A} \qquad (2).$$

3. Accumulated deviations $X_{k,a}$ from the average e_a for each sub-period I_a are defined as:

$$X_{k,a} = \sum_{i=1}^{k}(N_{i,a} - e_a). \qquad (3)$$

The range is defined as the maximum index $X_{k,a}$ minus the minimum $X_{k,a}$, within each sub-period (I_a):

$$R_{Ia} = \max(X_{k,a}) - \min(X_{k,a}), \quad 1 \le k \le n. \qquad (4)$$

4. The standard deviation S_{Ia} is calculated for each sub-period I_a:

$$S_{Ia} = \left(\left(\frac{1}{n}\right) \sum_{k=1}^{n}(N_{k,a} - e_a)^2 \right)^{0,5}. \qquad (5)$$

5. Each range R_{Ia} is normalised by dividing by the corresponding S_{Ia}. Therefore, the re-normalised scale during each sub-period I_a is R_{Ia}/S_{Ia}. In the step 2 above, one obtains adjacent sub-periods of length n. Thus, the average R/S for length n is defined as:

$$(R/S)_n = (1/A)\sum_{i=1}^{A}(R_{Ia}/S_{Ia}). \qquad (6)$$

6. The length n is increased to the next higher level, $(M - 1)/n$, and must be an integer number. In this case, we use n-indexes that include the initial and end points of the time series, and Steps 1 - 6 are repeated until $n = (M - 1)/2$.

7. Next least squares can be used to estimate the equation $log\ (R / S) = log\ (c) + H log\ (n)$. The slope of the regression line is an estimate of the Hurst exponent H. This can be defined over the interval [0, 1], and is calculated within the boundaries specified below:

- $0 \le H < 0.5$ – the data are fractal, the distribution has fat tails, the series are anti-persistent, returns are negatively correlated, there is pink noise with frequent changes in the direction of price movements, trading in the market is riskier for individual participants.
- $H = 0.5$ – the data are random, asset prices follow a random Brownian motion (Wiener process), the series are normally distributed, returns are uncorrelated (no memory in the series), they are a white noise, traders cannot «beat» the market using any trading strategy.

- $0.5 < H \leq 1$ – the data are fractal, the distribution has fat tails, the series are persistent, returns are highly correlated, there is black noise and a trend in the market.

To analyse the dynamics of market persistence we use a sliding-window approach. The procedure is the following: having obtained the first value of the Hurst exponent (for example, for the date 01.04.2004 using data for the period from 01.01.2004 to 31.03.2004), each of the following ones is calculated by shifting forward the "data window", where the size of the shift depends on the number of observations and a sufficient number of estimates is required to analyse the time-varying behaviour of the Hurst exponent. For example, if the shift equals 10, the second value is calculated for 10.04.2004 and characterises the market over the period 10.01.2004 till 09.04.2004, and so on.

We analyse returns computed as follows:

$$R_i = \left(\frac{Close_i}{Open_i} - 1\right) \times 100\% , \tag{9}$$

where R_i – returns on the *i-th* period in percentage terms;

$Open_i$ – open price on the *i-th* period;

$Close_i$ – close price on the *i-th* period.

3. Empirical Results

Descriptive statistics for the R/S analysis in the case of the high-frequency data collected at intervals of microseconds are presented in Table 1.

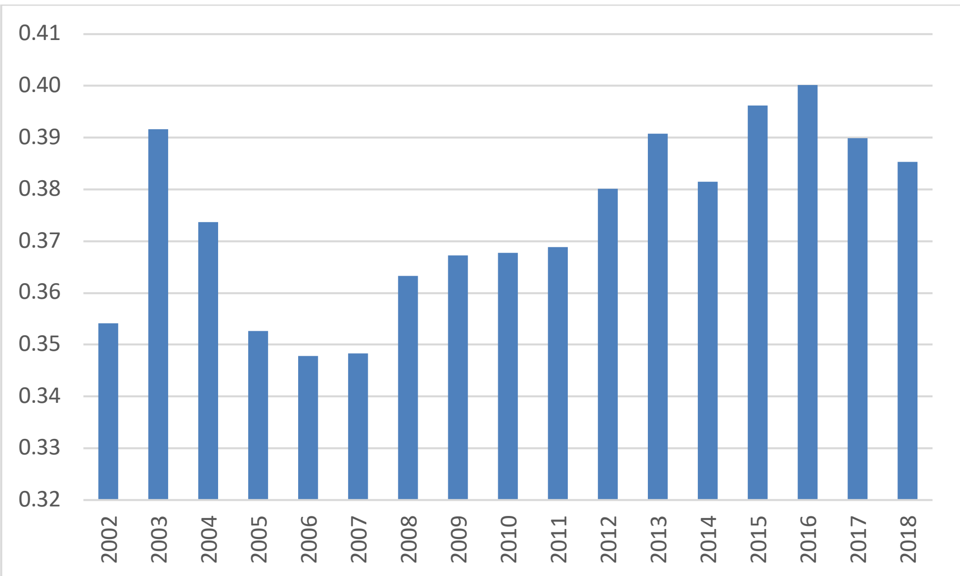
Table 1: Descriptive statistics for the R/S analysis

Year/ Parameter	Mean	Standard deviation	Interval	Min	Max	Count
2002	0.35	0.06	0.33	0.12	0.45	167
2003	0.39	0.06	0.37	0.08	0.45	219
2004	0.37	0.05	0.29	0.15	0.44	113
2005	0.35	0.08	0.42	0.03	0.45	187
2006	0.35	0.08	0.43	0.04	0.47	254

2007	0.35	0.09	0.45	0.01	0.45	233
2008	0.36	0.08	0.43	0.03	0.46	233
2009	0.37	0.07	0.39	0.06	0.45	229
2010	0.37	0.08	0.44	0.02	0.46	252
2011	0.37	0.08	0.46	0.03	0.49	235
2012	0.38	0.08	0.42	0.05	0.47	229
2013	0.39	0.08	0.41	0.08	0.49	247
2014	0.38	0.09	0.43	0.04	0.47	252
2015	0.40	0.09	0.46	0.04	0.50	186
2016	0.40	0.09	0.49	0.02	0.51	251
2017	0.39	0.09	0.46	0.04	0.50	247
2018	0.39	0.10	0.43	0.05	0.48	238
All	0.37	0.08	0.50	0.01	0.51	3772

As can be seen, the mean values are in the range [0.35-0.40], which implies the presence of anti-persistence in the data. However, there is also evidence of time variation in the Hurst exponent; this is apparent from Figure 1, which displays the results from the dynamic analysis for the high-frequency data based on a sliding-window approach.

Figure 1: Hurst exponent during 2002-2018



To examine whether the differences between the estimated values of the Hurst exponent for different years are statistically significant we perform t-tests (see Table 2);

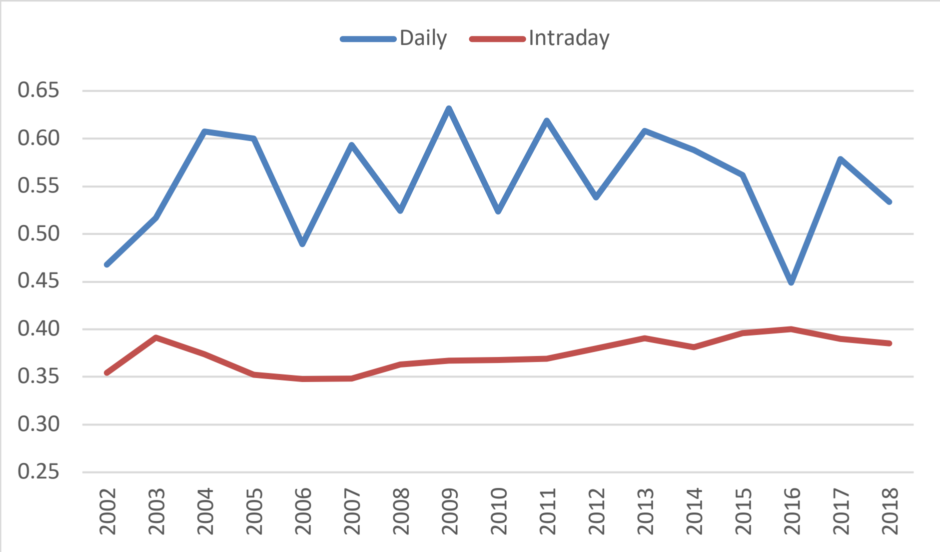
these confirm that indeed they are in the majority of cases, which implies that the degree of market efficiency changes over time.

Table 2: t-tests for differences in the R/S analysis results

Year/ Parameter	t-test	Difference is statistically significant
2002	4.00	yes
2003	3.79	yes
2004	0.15	no
2005	3.70	yes
2006	5.40	yes
2007	4.21	yes
2008	1.98	yes
2009	1.43	no
2010	1.29	no
2011	1.02	no
2012	1.10	no
2013	3.13	yes
2014	1.25	no
2015	3.12	yes
2016	4.39	yes
2017	2.69	yes
2018	1.71	no

Next we replicate the dynamic analysis using daily data (see Figure 2); this is not feasible in the case of monthly data as the sample (12 observations in each case) would be too small.

Figure 2: Dynamics of the Hurst exponent during 2002-2018: daily and intraday data



As can be seen, the only similarity between the two sets of estimates is the instability of persistence in both cases. This is confirmed by the very low correlation coefficient between them (-0.06). It is also apparent that persistence in daily data is much higher (see Table 3): the higher data frequency is characterised by a lower Hurst exponent and by anti-persistence rather than persistence.

Table 3: R/S analysis results for the cases of daily and high-frequency data

Year	Daily	Intraday	Difference, %
2002	0.47	0.35	-32%
2003	0.52	0.39	-32%
2004	0.61	0.37	-62%
2005	0.60	0.35	-70%
2006	0.49	0.35	-41%
2007	0.59	0.35	-70%
2008	0.52	0.36	-44%
2009	0.63	0.37	-72%
2010	0.52	0.37	-42%
2011	0.62	0.37	-68%
2012	0.54	0.38	-42%
2013	0.61	0.39	-56%
2014	0.59	0.38	-54%
2015	0.56	0.40	-42%
2016	0.45	0.40	-12%
2017	0.58	0.39	-48%
2018	0.53	0.39	-38%

Table 4 reports the estimates of the Hurst exponent for the whole sample at the three frequencies considered (monthly, daily and intraday) – these confirm that the behaviour of the series is very different at different frequencies. More specifically, monthly returns are the most persistent. This implies that they contain information about their future values, and thus autoregressive models can be estimated to predict them and develop strategies to "beat" the market. By contrast, intraday data are anti-persistent, which suggests that contrarian trading strategies should be applied.

Table 4: Hurst exponent values for different data frequencies, 2002-2018

Data frequency	Hurst
Month	0.72
Day	0.54
Intraday	0.37

4. Conclusions

This paper uses the Hurst exponent (calculated by means of R/S analysis) to explore the long-memory properties of high-frequency financial data for the case of EuroStoxx 50 futures prices over the period 2002-2018. The aim of the analysis is to establish whether or not persistence is sensitive to the data frequency (intraday, daily, monthly) and whether or not it varies over the time. Although such issues had already been examined in some earlier studies (see, e.g., Caporale and Gil-Alana, 2010), none had used data collected at micro seconds as the present one does.

The findings indicate that the series exhibit very different properties at different frequencies. More specifically, the higher the data frequency is, the lower is persistence. Monthly data are highly persistent, daily ones follow a random walk, and intraday ones are anti-persistent. The dynamic R/S analysis also shows that persistence varies over the time. The implication of these results is that the EMH only holds in the case of daily returns. In particular, in the case of intraday data it appears to be possible for traders, and more specifically scalpers (who enter and exit financial markets quickly, usually within seconds, using high levels of leverage to place large-sized trades in the hopes of achieving greater profits from minuscule price changes) to make abnormal profits by adopting mean-reverting trading strategies (sell after price increases, buy after price declines).

References

Alvarez-Ramirez, J., Cisneros, M., Ibarra-Valdez, C., Soriano A., (2002), Multifractal Hurst analysis of crude oil prices, Physica A, Vol. 313, 651-670.

Andersen, T.G., and T. Bollerslev, (1997), Intraday Periodicity and Volatility Persistence in Financial Markets, Journal of Empirical Finance, 4, 115-158.

Bariviera, Aurelio F., (2017), The inefficiency of Bitcoin revisited: A dynamic approach. Economics Letters, 161, 1–4.

Barunik, J., and Kristoufek, L., (2010), On Hurst exponent estimation under heavy-tailed distributions. Physica A: Statistical Mechanics and its Applications 389(18), 3844-3855.

Bouri, Elie, Georges Azzi, and Anne Haubo Dyhrberg, (2016), On the Return-volatility Relationship in the Bitcoin Market Around the Price Crash of 2013. Economics Discussion Papers, No 2016-41, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2016-41.

Caporale G. M., Gil-Alana L., Plastun A., (2018), Persistence in the cryptocurrency market. Research in International Business and Finance, Volume 46, Pages 141-148.

Caporale, GM. and Gil-Alana, L., (2010), Fractional integration and data frequency. Journal of Statistical Computation and Simulation, 80, 2, 121-132.

Caporale, GM. and Gil-Alana, L., (2013), Long memory and fractional integration in high frequency data on the US dollar/British pound exchange rate. International Review of Financial Analysis, 29. 1- 9

Caporale, GM., Gil-Alana, L. and Plastun, A., (2019), Long memory and data frequency in financial markets. Journal of Statistical Computation and Simulation, 89 (10). 1763 – 1779.

Cotter, J., (2005), Uncovering long memory in high frequency UK futures, European Journal of Finance 11, 325-337.

Da Silva S., Matsushita R., Gleria I., Figueiredo A., (2007), Hurst exponents, power laws, and efficiency in the Brazilian foreign exchange market, Economics Bulletin, 7 (1), 1-11.

Danylchuk, H., Kovtun, O., Kibalnyk, L., Sysoiev O., (2020), Monitoring and modelling of cryptocurrency trend resistance by recurrent and R/S-analysis. E3S Web Conf. 166 13030 DOI: 10.1051/e3sconf/202016613030

Fama, E., (1970), Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, 25, 383-417.

Glenn, L., (2007), On Randomness and the NASDAQ Composite, Working Paper, Available at SSRN: http://ssrn.com/abstract=1124991.

Grech, D. and Mazur, Z., (2004), Can one make any crash prediction in finance using the local Hurst exponent idea? Physica A: Statistical Mechanics and its Applications 336, 133-145.

Hurst, H., (1951), Long-term Storage of Reservoirs. Transactions of the American Society of Civil Engineers, 799 p.

Jacobsen, B., (1995), Are Stock Returns Long Term Dependent? Some Empirical Evidence. Journal of International Financial Markets, Institutions and Money 5 (2/3), 37-52.

Kantelhardt, J., Zschiegner S., Koscielny-Bunde E., Bunde A., Havlin S., and Stanley E., (2002), Multifractaldetrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications 316, 1-4.

Kim, K. Yoon, Seong-Min, (2004), Multifractal features of financial markets, Physica A: Statistical Mechanics and its Applications, 344 (1), 272-278.

Lento, C., (2009), A Synthesis of Technical Analysis and Fractal Geometry - Evidence from the Dow Jones Industrial Average Components, Available at SSRN: http://ssrn.com/abstract=1263615.

Lo, A.W., (1991), Long-term memory in stock market prices, Econometrica, 59, 1279–1313.

Los, C. A. and Yalamova, R. M., (2006), Multi-Fractal Spectral Analysis of the 1987 Stock Market Crash, International Research Journal of Finance and Economics, 1 (4), 106-133.

MacDonald, R. and Taylor, M.P., (1992), Persistence in stock market returns: estimates using a weekly data base. Journal of Business Finance and Accounting, 19, 23-31.

Mandelbrot, B., (1972), Statistical Methodology ForNonperiodic Cycles: From The Covariance To Rs Analysis. Annals of Economic and Social Measurement 1, 259-290.

Mandelbrot, B. and Wallis J.R., (1969), Computer experiments with fractional noises, Water Resources Research 5, 228.

Metescu, Ana-Maria, (2022), Fractal market hypothesis vs. Efficient market hypothesis: applying the r/s analysis on the Romanian capital market. Journal of Public Administration, Finance and Law. 11. 199-209. 10.47743/jopafl-2022-23-17.

Mynhardt, R., Plastun A., and Makarenko I., (2014), Behavior of financial markets efficiency during the financial market crisis: 2007 – 2009. Corporate Ownership and Control 11(2), 473-488.

Niere, Harvey M., (2013), A Multifractality Measure of Stock Market Efficiency in Asean Region, European Journal of Business and Management 5 (22), 13-19.

Phillip A, J.S.K. Chan, Peiris M.S., (2018), A new look at Cryptocurrencies, Economics Letters, 163, 6–9.

Raimundo, M., & Okamoto Jr, Jun, (2018), Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities. International Journal of Modeling and Optimization. 8. 116-124. 10.7763/IJMO.2018.V8.635.

Serletis, A., and Rosenberg, A., (2007), The Hurst exponent in energy futures prices. Physica A 380, 325-332.

Taqqu, M., Teverovsky, V. and Willinger W., (1995), Estimators for long-range dependence: an empirical study, Fractals 3 785–798.

Urquhart, A., (2016), The Inefficiency of Bitcoin. Economics Letters, 148, 80-82.

Zunino, L., Tabak B., Garavaglia M. and Rosso O., (2009), Multifractal structure in Latin-American market indices. Chaos, Solitons and Fractals; v. 41(5), 2331-2340.