FORECASTING THE BEHAVIOR OF FRACTAL TIME SERIES: HURST EXPONENT AS A MEASURE OF PREDICTABILITY

Vladimir KULISH[*], Vladimír HORÁK[**]
[*]School of Mechanical & Aerospace Engineering, Nanyang Technological University, Singapore (mvvkulish@ntu.edu.sg)
[**] Faculty of Military Technology, University of Defense, Brno, Czech Republic (vladimir.horak@unob.cz)

DOI: 10.19062/1842-9238.2016.14.2.8

Abstract: *The Hurst exponent (H) is a statistical measure used to classify time series. H = 0.5 indicates a random series while H > 0.5 indicates a trend reinforcing series. The larger the H value is the stronger trend. In this paper we investigate the use of the Hurst exponent to classify series of financial data representing different periods of time. In this paper we show that series with large values of the Hurst exponent can be predicted more accurately than those series with H value close to 0.5. Thus the Hurst exponent provides a measure for predictability.*

Keywords: *Hurst exponent, time series analysis, forecasting*

1. INTRODUCTION

The Hurst exponent, proposed by H. E. Hurst [1] for use in fractal analysis, has been applied to many research fields, ranged from vibration and control, to biomedical signal processing, to temperature and velocity fluctuations in viscous fluid flows, and to climate change studies [2, 3]. It has recently become popular in the finance community [4, 5, 6] largely due to Peters' work [7, 8]. The Hurst exponent provides a measure for long-term memory and fractality of a time series. Since it is robust with few assumptions about underlying system, it has broad applicability for time series analysis, i.e., the origin of the time series is unimportant for this analysis. In view of this, the conclusions drawn in this study are universal and can be employed in any area of research, in which forecasting of the time series behavior is necessary.

The values of the Hurst exponent range between 0 and 1. Based on the Hurst exponent value H, a time series can be classified into three categories: (1) $H = 0.5$ indicates a random series; (2) $0 < H < 0.5$ indicates an anti-persistent series; (3) $0.5 < H < 1$ indicates a persistent series. An anti-persistent series has a characteristic of "mean-reverting", which means an up value is more likely followed by a down value, and vice versa. The strength of "mean-reverting" increases as H approaches 0. A persistent series is trend reinforcing, which means the direction (up or down compared to the last value) of the next value is more likely the same as current value. The strength of trend increases as H approaches 1. Most financial time series are persistent with $H > 0.5$.

In time series forecasting, the first question we want to answer is whether the time series under study is predictable. If the time series is random, all methods are expected to fail. We want to identify and study those time series having at least some degree of predictability. We know that a time series with a large Hurst exponent has strong trend, thus it is natural to believe that such time series are more predictable than those having a Hurst exponent close to 0.5.

In this paper we show that series with large values of the Hurst exponent can be predicted more accurately than those series with H value close to 0.50. Thus the Hurst exponent provides a measure for predictability.

In this study, we chose a financial time series, because of its data easy availability in the public domain. Yet, it is noteworthy to emphasize here once again that the conclusions drawn in this study are universal and can be employed in any area of research, in which forecasting of the time series behavior is necessary, including any risk management analysis.

2. HURST EXPONENT AND R/S ANALYSIS

The Hurst exponent can be calculated by rescaled range analysis (R/S analysis). For a time series, $X = X_1, X_2, \ldots, X_n$, R/S analysis method is as follows:

(1) Calculate mean value m:

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

(2) Calculate mean adjusted series Y:

$$Y_t = X_t - m, \qquad t = 1, 2, \ldots, n \tag{2}$$

(3) Calculate cumulative deviate series Z:

$$Z_t = \sum_{i=1}^{n} Y_i, \qquad t = 1, 2, \ldots, n \tag{3}$$

(4) Calculate range series R:

$$R_t = \max(Z_1, Z_2, \ldots Z_t) - \min(Z_1, Z_2, \ldots Z_t), \qquad t = 1, 2, \ldots, n \tag{4}$$

(5) Calculate standard deviation series S:

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(X_i - u)^2}, \qquad t = 1, 2, \ldots, n \tag{5}$$

Here u is the mean value from X_1 to X_t.

(6) Calculate rescaled range series (R/S):

$$(R/S)_t = R_t / S_t, \qquad t = 1, 2, \ldots, n \tag{6}$$

Note $(R/S)_t$ is averaged over the regions $[X_1, X_t]$, $[X_{t+1}, X_{2t}]$ until $[X_{(m-1)t+1}, X_{mt}]$, where m = floor(n/t). In practice, to use all data for calculation, a value of t is chosen that is divisible by n.

Hurst found that (R/S) scales by power-law as time increases, which indicates:

$$(R/S)_t = c^* \, t^H \tag{7}$$

Here c^* is a constant and H is called the Hurst exponent. To estimate the Hurst exponent, we plot (R/S) versus t in log-log axes. The slope of the regression line approximates the Hurst exponent. For $t < 10$, $(R/S)_t$ is not accurate, thus we shall use a region of at least 10 values to calculate rescaled range. Fig. 1 shows an example of R/S analysis.

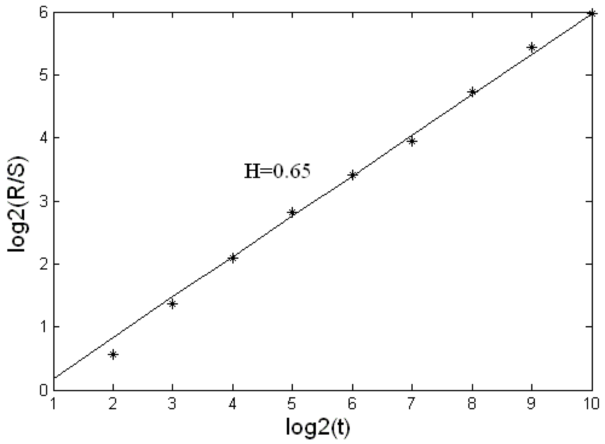

FIG. 1. R/S analysis for Dow-Jones daily return from 11/18/1969 to 12/6/1973

In our experiments, we calculated the Hurst exponent for each period of 1024 trading days (about 4 years). We use $t = 2^4, 2^5, \ldots, 2^{10}$ to do regression. In the financial domain, it is common to use log difference as daily return. This is especially meaningful in R/S analysis since cumulative deviation corresponds to cumulative return. Fig. 2 shows the Dow-Jones daily return from Jan. 2, 1930 to May 14, 2004. Fig. 3 shows the corresponding Hurst exponent for this period. In this period, Hurst exponent ranges from 0.4200 to 0.6804. We also want to know what the Hurst exponent would be for a random series in our condition.

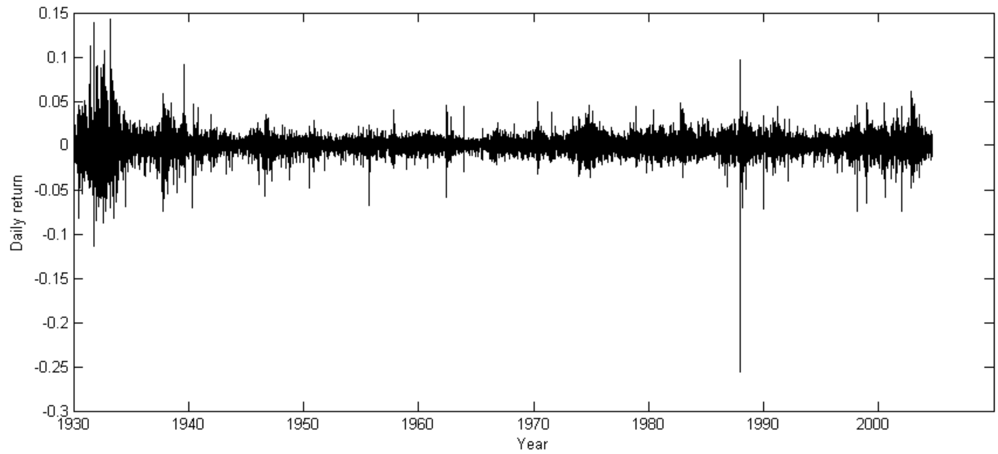

FIG. 2. Dow-Jones daily return from 1/2/1930 to 5/14/2004

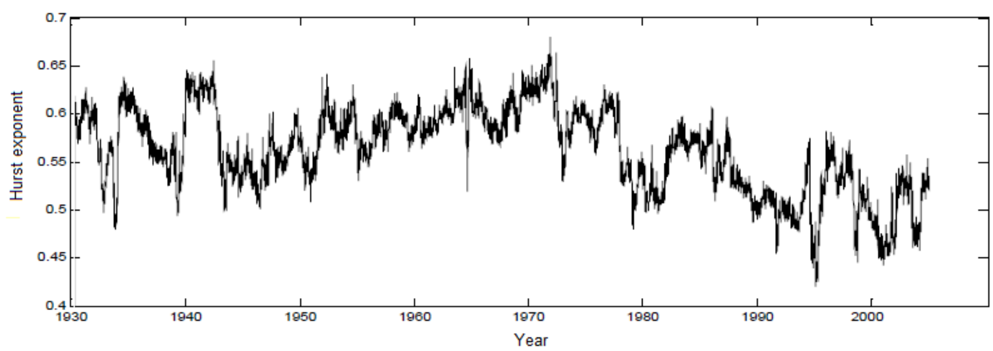

FIG. 3. Hurst exponent for Dow-Jones daily return from 1/2/1930 to 5/14/2004

3. MONTE CARLO SIMULATION

For a random series, Feller [13] gave expected $(R/S)_t$ formula as (8):

$$E[(R/S)_t] = (n\,\pi/2)^{0.50} \tag{8}$$

However, this is an asymptotic relationship and is only valid for large t. Anis and Lloyd [14] provided the following formula to overcome the bias calculated from (8) for small t:

$$E\,[(R/S)_t] = \left\{\Gamma\,[0.5\,(t-1)]\big/\left[\sqrt{\pi}\;\;\Gamma\,(0.5\,t)\right]\right\}\sum_{r=1}^{t-1}\sqrt{(t-r)/r} \tag{9}$$

For $t > 300$, it is difficult to calculate the gamma function by most computers. Using Sterling's function, formula (9) can be approximated by:

$$E\,[(R/S)_t] = (t\,\pi/2)^{-0.50}\sum_{r=1}^{t-1}\sqrt{(t-r)/r} \tag{10}$$

Peters [8] gave equation (11) as a correction for (9):

$$E\,[(R/S)_t] = [(t-0.5)/t]\,(t\,\pi/2)^{-0.50}\sum_{r=1}^{t-1}\sqrt{(t-r)/r} \tag{11}$$

We calculate the expected (R/S) values for $t = 2^4, 2^5, \ldots, 2^{10}$ and do least squares regression at significance level $\alpha = 0.05$. Results are shown in table 1.

Table 1. Hurst exponent calculation from Feller, Anis and Peters formula

log2(t)	log2[E(R/S)]		
	Feller	*Anis*	*Peters*
4	0.7001	0.6059	0.5709
5	0.8506	0.7829	0.7656
6	1.0011	0.9526	0.9440
7	1.1517	1.1170	1.1127
8	1.3022	1.2775	1.2753
9	1.4527	1.4345	1.4340
10	1.6032	1.5904	1.5902
Regression Slope (H)	**0.5000** ±5.5511e-16	**0.5436** ±0.0141	**0.5607** ±0.0246

From table 1, we can see that there are some differences between Feller's, Anis' and Peters' formulae. Moreover, their formulae are based on large numbers of data points. In our case, the data is fixed at 1024 points. So what is the Hurst exponent for random series in our case?

Fortunately, we can use Monte Carlo simulation to derive the result. We generate 10,000 Gaussian random series. Each series has 1024 values. We calculate the Hurst exponent for each series and then average them. We expect the average number to approximate the true value. We repeated this process 10 times. Table 2 below gives the simulation results.

From table 2, we can see that in our situation, the Hurst exponent calculated from Monte Carlo simulations is 0.5454 with standard deviation 0.0485. Our result is very close to Anis' formula.

Based on the above simulations, with 95% confidence, the Hurst exponent is in the interval 0.5454 ± 1.96·0.0485, which is between 0.4503 and 0.6405. We choose those periods with Hurst exponent greater than 0.65 and expect those periods to be bearing some structure different from random series. However, since these periods are chosen from a large sample (total 17651 periods), we want to know if there exists true structure in these periods, or just by chance. We run a scramble test for this purpose.

Table 2. Monte Carlo simulations for Hurst exponent of random series

	Simulated Hurst exponent	*Standard deviation (Std.)*
1	0.5456	0.0486
2	0.5452	0.0487
3	0.5449	0.0488
4	0.5454	0.0484
5	0.5456	0.0488
6	0.5454	0.0481
7	0.5454	0.0487
8	0.5457	0.0483
9	0.5452	0.0484
10	0.5459	0.0486
Mean Sdt.	**0.5454** ±2.8917e-4	**0.0485**

4. SCRAMBLE TEST

To test if there exists true structure in the periods with Hurst exponent greater than 0.65, we randomly choose 10 samples from those periods. For each sample, we scramble the series and then calculate the Hurst exponent for this scrambled series. The scrambled series has the same distribution as the original sample except that the sequence is random. If there exists some structure in the sequence, after scrambling the structure will be destroyed and the calculated Hurst exponent should be close to that of a random series. In our experiment, we scramble each sample 500 times and then the average Hurst exponent is calculated. The results are shown in table 3 below.

Table 3. The average Hurst exponent on 500 scrambling runs

	Hurst exponent after scrambling	*Standard deviation*
1	0.5492	0.046
2	0.5450	0.047
3	0.5472	0.049
4	0.5454	0.048
5	0.5470	0.048
6	0.5426	0.048
7	0.5442	0.051
8	0.5487	0.048
9	0.5462	0.048
10	0.5465	0.052
Mean	**0.5462**	**0.048**

From table 3, we can see that the Hurst exponents after the scrambling of samples are all very close to 0.5454 which is the number from our simulated random series. Given this result, we can conclude that there must exist some structure in those periods making them different from random series and that scrambling destroys the structure. We hope this structure can be exploited for prediction.

5. HURST EXPONENTS AS A MEASURE OF PREDICTABILITY

The efficient market hypothesis (EMH) asserts that financial markets are "efficient", or prices on traded assets, e.g. stocks, bonds, or property, already fully reflect all available information and therefore are unbiased in the sense that they reflect the collective beliefs of all investors about future prospects. In other words, the efficient market hypothesis implies that it is not possible to consistently outperform the market – appropriately adjusted for risk – by using any information that the market already knows, except through luck. The efficient market hypothesis follows from the assumption that the price formation is a random walk process. The random walk theory asserts that price movements will not follow any patterns or trends and that past price movements cannot be used to predict future price movements.

The paradox hidden behind the EMH is the same as the paradox of instantaneous energy propagation behind the classical diffusion (heat transfer) model. In physics, this paradox is overcome by assuming a finite time lag between the onset of a disturbance upon a physical system and the system's response to it. A similar assumption of phase-lagging behavior must be made to realistically describe the behavior of markets.

Introducing a finite time lag between receiving a new piece of information and response to it, allows one, in turn, to introduce a quantitative measure of the market *inefficiency* at any given moment of time (the inefficiency coefficient). The more inefficient market is the more predictable its behavior. Hence, computing the inefficiency coefficient as a function of time and, from its value, the predictability measure, one should be able to forecast the market behavior. The forecast will be more exact for larger values of the inefficiency coefficient and less exact for smaller values, becoming zero at those instances when market becomes fully efficient (see [15] for details).

In the preceding sections we showed that the Hurst exponent, a measure of the time series persistency, can be viewed as a predictability measure of those time series. It is hypothesized here that the time series in question are in fact solutions to fractional (non-integer order) partial differential equations, in which the Hurst exponent is the order of the time derivative and is itself a function of time. This presents the main mathematical challenge of the model: to develop methods for solving fractional partial differential equations with time-variable order [16].

From the practical point of view, the knowledge of the value of the inefficiency coefficient provides one with an edge with respect to an average (unknowledgeable or uninformed) market participant. Hence, the Kelly theorem can be used to maximize one's gain (the expected rate of return) from trading an asset. The Kelly theorem asserts that, to maximize one's return, the fraction of the current bankroll to wager must be equal to the ratio of the expected net winnings to the net winnings if one wins [17].

Indeed, the amount of information, obtained from one's knowledge of the Hurst exponent's value is

$$S = -H\log(H) - (1-H)\log(1-H) \tag{12}$$

The function, given by (12), reaches its maximum at H = 0.5, S(0.5) = 1 and is symmetric with respect to its maximal value, that is, $S(0) = S(1) = 0$. At the same time, it is obvious that the behavior of time series becomes totally predictable at S = 0, while the case of S = 1 represents a totally unpredictable time series. From this, a predictability measure can be introduced merely as

$$P = 1 - S \tag{13}$$

At the same time, P merely plays the role of the probability to win and the Kelly criterion gives

$$f = \frac{P(b+1)-1}{b} \tag{14}$$

where f is the fraction of the current bankroll to wager, i.e. how much to bet and b is the net odds received on the wager ("b to 1"), that is, one could win $b (on top of getting back your $1 wagered) for a $1 bet.

CONCLUSIONS

In this paper, we analyze the Hurst exponent for all 1024-trading-day periods of the Dow-Jones index from January 2, 1930 to May 14, 2004. We find that the periods with large Hurst exponents can be predicted more accurately than those with H values close to random series. This suggests that stock markets are not totally random in all periods. Some periods have strong trend structure and this structure can be used to benefit forecasting.

Since the Hurst exponent provides a measure for predictability, we can use this value to guide data selection before forecasting. We can identify time series with large Hurst exponents before we try to build a model for prediction. Furthermore, we can focus on the periods with large Hurst exponents. This can save time and effort and lead to better forecasting.

It is noteworthy to emphasize here once again that the conclusions drawn in this study are universal and can be employed in any area of research, in which forecasting of the time series behavior is necessary, including any risk management analysis [18, 19].

AKNOWLEDGMENT

This work has been supported by the grant from Ministry of Education, Singapore. Grant number RG119-15. It also has been supported by the institutional funding DZRO K 201 “VÝZBROJ” and by the specific research project of Faculty of Military Technology SV16-216 of University of Defence, Czech Republic.

REFERENCES

[1] H. E. Hurst, Long-term storage of reservoirs: an experimental study, *Transactions of the American society of civil engineers,* 116, pp. 770-799, 1951;

[2] B. B. Mandelbrot and J. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Review,* 10, pp. 422-437, 1968;

[3] B. Mandelbrot, *The fractal geometry of nature*, New York, W. H. Freeman, 1982;

[4] C. T. May, *Nonlinear pricing: theory & applications*, New York : Wiley, 1999;

[5] M. Corazza and A. G. Malliaris, Multi-Fractality in Foreign Currency Markets, *Multinational Finance Journal,* vol. 6, no. 2, pp. 65-98, 2002;

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

[7] E.E. Peters, *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*, New York, Wiley, 1991;
[8] E. E. Peters, *Fractal market analysis: applying chaos theory to investment and economics,* New York, Wiley, 1994;
[9] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural networks,* vol. 2, no. 5, pp. 259-366, 1989;
[10] S. Walczak, An empirical analysis of data requirements for financial forecasting with neural networks, *Journal of management information systems,* vol. 17, no. 4, pp. 203-222, 2001;
[11] E. Gately, *Neural networks for financial forecasting*, New York, Wiley, 1996;
[12] A. Refenes, *Neural networks in the capital markets*, New York, Wiley, 1995;
[13] W. Feller, The asymptotic distribution of the range of sums of independent random variables, *The annals of mathematical statistics,* 22, pp. 427-432, 1951;
[14] A. A. Anis and E. H. Lloyd, The expected value of the adjusted rescaled Hurst range of independent normal summands, *Biometrika,* 63, pp. 111-116, 1976;
[15] V. V. Kulish, Market Efficiency and the Phase-Lagging Model of Price Fluctuations, *Physica A*, vol. 387, issue 4, pp. 861-867, 2008;
[16] V. V. Kulish and G. V. Nosovskiy, Applicability of Laplace transform for fractional PDE with variable rate of derivatives, *Mathematics in Engineering, Science and Aerospace*vol, 4, no. 4, pp. 415-427, 2013;
[17] J. L. Kelly, A New Interpretation of Information Rate, *Bell System Technical Journal*, vol. 35 no. 4, pp. 917–926, 1956.
[18] M. Milandru, The Risk Management, Determinant Factor of Increasing the Logistics Performance, *Review of the Air Force Academy*, vol. 13, no. 3, pp. 139–142, 2015;
[19] F.-C. Olteanu and C. Gheorghe, Aspects Regarding the Qualitative Analysis of Risks Due to the Occurrence of Low Probability and Very High Impact Events, *Review of the Air Force Academy*, vol. 14, no. 1, pp. 133–140, 2016.