A Different Statistic for the Management of Portfolios - the Hurst Exponent: Persistent, Antipersistent or Random Time Series?

Ana-Maria CALOMFIR (METESCU)[1]

Abstract
In recent years, research in the capital markets and management of portfolios has been producing more questions than it has been answering: the need for a new paradigm or a new way of looking at things has become more and more concludent. The existing and classical view of capital markets, based on efficient market hypothesis, has a definite theory for the last six decades, but it is still not capable of significantly increase the understanding of how capital markets function. The purpose of this article is to theoretically describe a less used statistic coefficient, having a vast area of applicability due to its robustness, and which can easily divide the random series from a non-random series, even if the random series is non-Gaussian: the Hurst exponent.

Key words: *random walk, Hurst exponent, fractal dimension, capital markets.*

JEL classification: *G11, G170, G14.*

INTRODUCTION

Efficient Market Hypothesis states, basically, the following phenomenon: because current prices reflect all available or public information, then future price changes can be determined only by the arrival of new information. And it must be mentioned that the process of new information arrival is random. With all prior information already reflected in prices, the markets follow a random walk. This means that each day's price movement is unrelated to previous day's activity, and that all investors react immediately to new information, so that the future is unrelated to past or present.

But here it may appear an obvious question: do people really make decisions in this manner? Typically, there are people who do react to information as it is received. Still, most people wait for confirming information and do not react until a trend is clearly established. The amount of information needed to validate a new trend varies, and the uneven assimilation of information may cause a biased random walk. Biased random walks were extensively studied by Hurst in the 1940s and by Mandelbrot in the 1960s and 1970s. Mandelbrot called those mathematical processes as fractional Brownian motions, and the time series described by those movements can be named as fractal time series.

[1] Romanian Academy, Romania, anametescu@yahoo.com

First empiric research in this area has primarily been made by hydrologist Hurst, who studied for 4 decades the manner in which quantity of water from Nile river fluctuated. In order to reach to make estimates with a certain accuracy degree, Hurst tried to create a mathematical model for describing this variable. While implementing and testing the model, he acknowledged that a he had to assume an uncontrollable part of the system, in his case, the influx of water from rainfall, and which he agreed to follow a random walk. This assumption was common, when dealing with a large system that has many degrees of freedom.

When Hurst decided to test the assumption, he gave us a new statistic: the Hurst exponent (H), which he later tested on most natural phenomena, i.e. river discharges, temperatures, rainfalls and sunspots, and he discovered that this new statistic has a vast area of applicability, due its robustness. With the discovery of this new statistic coefficient, one can distinguish a random series from a non-random series, even if the random series is non-Gaussian (not normally distributed). Hurst found that most natural systems do not follow a random walk, Gaussian or otherwise. This conclusion was reached by analyzing the historic time series of the values, and taking into account the “square root of time rule”, meaning that if the series were random, the range would increase with the square root of time. By creating this new statistic, Hurst succeeded to demonstrate and “remove” the old perception according to which the natural phenomena follow random walks, proving that a natural phenomenon follows a “biased random walk” – meaning the time series follows a trend with noise.

1. HURST EXPONENT - DEFINITION AND RELEVANCE

The Hurst exponent represents a powerful statistical tool for dividing time series into persistent, antipersistent or ergodic series, or, in very rare situations, applied on natural world phenomena, random series. Its importance has increased due to appearance of interdisciplinary and transborder sciences. As capital markets, or economic process, being assimilated to real world processes, were thought to follow random walks, and testing those processes conducted to no precise conclusion regarding those processes, the Hurst exponent gives more clues about the nature of the time series, but still remains a “black hole” about the nature of the distribution that characterizes the process itself. Economic and capital market time series have, as a general characteristic, long but finite memories, depending from market to market, as well as from security to security. Applying the R/S analysis to capital markets, we find fractal structures and nonperiodical cycles, conclusive evidence that the capital markets are nonlinear systems and that the EMH is questionable.

1.1. Computation methodology for the Hurst exponent

Having as a starting point the mathematical model implemented by Hurst, the intention of different researchers was to extend Hurst’s work area of applicability from the world of natural phenomena to the world of economic processes and capital markets, in order to observe if the later follow the same biased random walks. To reformulate Hurst’s work for a general time series, we must define a range that would be comparable to the fluctuations of the reservoir height levels.

The mathematical model is presented below: We begin with an existing time series, *t*, with *u* observations:

$$X_{t,N} = \sum_{u=1}^{t}(e_u - M_N) \qquad (1)$$

where: $X_{t,N}$ = cumulative deviation over *N* periods;

e_u = value in year *u*;

M_N = average e_u over *N* periods.

In this way, the range then becomes the difference between the maximum and minimum levels attained in equation (1).

$$R = Max(X_{t,N}) - Min(X_{t,N}) \qquad (2)$$

where: *R* = Range of *X*;
Max(X) = Maximum level of *X*;
Min(X) = Minimum level of *X*..

In order to compare different types of time series, Hurst divided this range by the standard deviation of the original observations. This "rescaled range" should increase with time. Hurst formulated the following relationship:

$$R/S = (a * N)^H \qquad (3)$$

where: *R/S* = rescaled range;
N = number of observations;
a = a constant;
H = Hurst exponent.

According to statistical mechanisms, H should equal 0.5 for series following a random walk. In other words, the range of cumulative deviations should increase with the square root of time, N. When Hurst applied his statistic to the Nile River discharge record, he found H=0.90. Then he tried other rivers, and H was usually greater than 0.50. Then, he recalculated to different natural phenomena, and, in all cases, he found H greater than 0.5. The explanation resides in the fact that when H differentiates from 0.50, the observations were not independent. Practically, each observation carried a "memory" of all the events that preceded it. But this memory was not a short-term memory, often called "Markovian", this was a different memory, long-term one, which is supposed to last forever. The translation of this fact is the following: more recent events have a greater impact that distant events, but there still exists residual influence. On a broader scale, a system that exhibits Hurst statistics is the result of a long stream of interconnected events, meaning that what happens today influences the future.

From the point of view of the values that can be obtained for the Hurst exponent, we can identify three categories: (1) $H=0.5$, (2) $0 \leq H \leq 0.5$, si (3) $0.5<H<1$. The first category refers to random series. Events are random and uncorrelated. The present does not influence the future, nor the past influences the present. The density probability function can be a normal curve, but it does not have to be. Typically, in a classical statistical manner, nature should follow a normal distribution. But Hurst's findings refute that teaching. H for natural phenomenon is typically greater than 0.5, and its probability distribution is not normal.

The second case, for $0 \leq H \leq 0.5$, corresponds to antipersistent or ergodic series. This kind of processes are often referred to as "mean reverting processes". This means that if the system has been up in the previous period, it is more likely to be up in the next period. The strength of this antipersistent behavior depends on how close H is to zero. This kind of series is more volatile than a random series, due to the fact that it consists of frequent reversals. Still, despite of the prevalence of the mean reversal concept in economic and financial literature, few antipersistent series have been found yet.

In the third case, for which $0.5<H<1$, we have a persistent or trend-reinforcing series. This can be translated in the following manner: if the series has been up or down in the last period, then the chances are that it will continue to be positive or negative in the next period. Trends are apparent, and the strength of the trend-reinforcing behavior, or persistence, increases as H approaches to 1. On the other hand, the closer H is to 0.5, the noisier it will be, and the less defined its trends will be. Persistent series are fractional Brownian motions or biased random walks.

The most important class of series to study is persistent time series, due to the fact that we may find plentiful of them in nature, as in economic processes and in capital markets.

Persistent time series, defined as $0.5<H<1$, are fractal because they can also be described as fractional Brownian motion. As a general characteristic for this type of process, in fractional Brownian motion there is correlation between events across time scales. Because of this relationship, the probability of two events following one another is not 50/50. The Hurst exponent (H) describes the likelihood that two consecutive events are likely to occur. For example, if $H=0.6$, this means that there is 60% probability that, if the last move was positive, the next move will also be positive. As H draws closer to 1, the series becomes less noisy and has more consecutive observations with the same signs. Persistent time series are the more interesting class because, as Hurst found, they are plentiful in nature, as are the capital markets. Still, a question arrives: what causes persistence? Why does it involve a memory effect?

For answering those questions, it is vital to understand the way Hurst's statistics is constructed, and Hurst's method simulating technique for determining this statistic.

Mandelbrot (1972) has shown that the inverse of the H coefficient is the fractal dimension. Based on this relationship, a pure random walk, with $H=0.5$, would have a fractal dimension of 2. If, for example, H would increase to 0.7, the fractal dimension would become 1/0.7, meaning 1.43. As a direct consequence, it means that a random walk is truly two-dimensional and would fill up a plane.

As H increases, there are more positive increments followed by positive increments, and negative increments followed by negative increments. The correlation of the signs in the series is increasing.

In order to estimate the Hurst exponent, we must logarithm the basic equation, and we will obtain:

$$\log(R/S) = H * \log(N) + \log(a) \quad (4)$$

Identifying the slope of the log/log graph of R/S versus N will provide an estimate of H. But, this estimate of H makes no assumptions about the shape of the underlying distribution.

For very long N, we might expect the series to converge to the value H=0.5, due to the fact that the memory effect diminishes to a point where it becomes unmeasurable. In other words, observations with long N can be expected to exhibit properties similar to regular brownian motion, or a pure random walk, as the memory effect dissipates.

The following fact is to be mentioned: the R/S analysis is an extremely robust tool. But it does not assume that the underlying distribution is Gaussian. Finding H=0.50 does not prove a Gaussian random walk, but it represents only a certitude for the inexistence of the long memory process. In other words, any independent system, Gaussian or otherwise, would produce H=0.5.

2. COMPUTATION METHODOLOGY FOR THE HURST EXPONENT IN THE CAPITAL MARKETS

In the case of a real closing price time series of an index or share, the computation methodology follows few simple steps, as indicated below:

1. Obtain the daily returns time series from the real closing price time series.
2. Then compute the mean of the daily returns time series, which will be named as M, after the formula:

$$M = \frac{1}{n}[r(1) + r(2) + r(3) + ...r(n)]. \quad (5)$$

3. The next step is to compute the deviations from the mean of the daily returns, which will be named with *x(1), x(2), x(3),..., x(n)*, following the simple relationship: *x(1)=r(1)-M; x(2)=r(2)-M; x(3)=r(3)-M,..., x(n)=r(n)-M*. It is important to mention that the mean of the deviations is zero.
4. Then, the *Y*'s should be computed, as *Y(1)=x(1), Y(2)=x(1)+x(2), ..., Y(n)=x(1)+x(2)+...+x(n)*. For those values obtained, there should be found the maximum and minimum levels of Y, and substracting them, to find the range:

R=Max[Y]-Min[Y] (6)

5. Compute the standard deviations of the original observations, *s=STDEV[r(k)]*, and divide the range at the standard deviation of the original observation, to obtain the value of the Rescaled Range:

$$R/S = (a * N)^H \quad (7)$$

6. For a certain n, this will generate a point on our *log(R/S)* vs *log(n)* chart. Further on, the method should be repeated for *n+10*, *n+20*, each time generating another point chart, and computing until *n+100*. The slope of the graph represents an estimate of the Hurst exponent.

4. THE VALIDITY OF THE H ESTIMATE

Even if a significantly anomalous value of H is found, there may still be a question as to whether the estimate itself is valid. Perhaps there were not enough data, or there may even be a question as to whether R/S analysis works at all. In a basical sense, a value of the H that significantly differes from 0.5 may offer two possible explanations:

1. There is a long memory component in the time series being studied, meaning that each observation is correlated to some degree with observations that follow.
2. The analysis itself is flawed, and an flawed value for the Hurst exponent does not mean that there is a long memory effect at all.

The time series under study represents an independent series of random variables, which scale according to a value different of 0.5, or represents an independent process with fat tails, as suggested by Cootner (1964).

The validity of the results can be tested by randomly scrambling the data, so that the order of the observations is completely different from the data of the original time series. Due to the fact that the actual observations are still there, the frequency distribution should remain unchanged. The H statistics should be repeated again as calculus on the scrambled data. If the series is truly an independent series, then the Hurst exponent calculation should remain virtually unchanged, due to the fact that there were no long memory effects, or correlations between the observations. As a direct consequence, scrambling the data would have no effect on the qualitative aspects of the data.

On the opposite side, if there exists a long memory effect in place, the order of the data is important, and the scrambling process destroys the structure of the system. The value of the H estimate calculated should be much lower, and close to 0.5, even though the frequency distribution of the observations remains unchanged. An important decrease of value of the Hurst exponent is justified by the explanation that the long memory effect of the original time series was destroyed by the scrambling process, and the correlation of the successive events are completely lost, the scrambling process producing an independent time series. An important consequence is the proof of the Mandelbrot's assertion that R/S analysis is a robust tool with respect to the distribution of the underlying series. It is important to mention that this scrambling process of data was applied also to time series of real data from natural systems, and one of the most widely known and accepted natural systems with a nonperiodic cycle is the sunspot cycle. The sunspot numbers start being recorded since the middle of the 18^{th} century, when Wolf began a daily routine of examining the sun's face through telescope and counting daily the number of black spots on the sun's surface.

Sunspots offer a highly appropriate time series for R/S analysis, due to their long registered history. Harlan True Stetson published in 1938 the paper „*Sunspots and Their Effects*", which also provided a table with montly data of sunspots numbers from January 1749 until December 1937. The real data time series was alluring enough to offer study material for Mandelbrot, Wallis (1969) and Hurst, who analysed the series. Peters (1992) was convinced to redo the analysis, incorporating the advances in technology since the last study offered by his colleagues. The R/S analysis was applied to the logarithmic first difference in the monthly sunspot numbers. The results of the analysis are interesting: periods shorter than 12 or 13 years have a Hurst exponent of 0.55, meaning a persistent behavior, and the slope of the log/log drops drastically after this point, showing the dissipation or dissapearance of the long memory effect after this period. Also, by scrambling the data, all traces of correlation between original data series have been destroyed. As a logical conclusion that can be applied to natural systems providing data, we can notice that all natural systems may have long memories as postulated by the fractional brownian motion model. But, still, the memory length is not infinite, it is long and finite. The result has similarity with the relationship between natural and mathematical fractals, meaning that mathematical fractals scale forever, both infinitely small and large, whereas natural fractals stop scaling after a point.

In a similar manner, fractal time series have long, but finite memories. Economic and capital market time series have, as a general characteristic, long but finite memories, depending from market to market, as well as from security to security. Applying the R/S analysis to capital markets, we find fractal structures and nonperiodical cycles, conclusive evidence that the capital markets are nonlinear systems and that the EMH is questionable. For capital market appliance, there will be used logarithmic returns, defined as:

$$S_t = \ln(P_t / P_{t-1}) \qquad (8)$$

where: S_t =logarithmic return at time „*t*";

P_t = price at time „*t*".

For this particular type of analysis, logarithmic returns are more appropriate than the commonly used percentage change in prices. The range used in R/S analysis is the cumulative deviation from the average, while logarithmic returns sum to cumulative return, while percentage change not. The first step is to convert the price or yield series into logarithmic returns. The next step is to apply equations described above for various increments of time. A reasonable increment of time may be monthly time series covering a number of years, which may be converted into number of years *12 (number of months) logarithmic returns. It it important to mention that the periods are nonoverlapping, and, as a result, there should be independent observations. There should be also mentioned that the independence condition is neglected in the case of short-term Markovian dependence. Then, the computation methodology described above can be applied, and calculate the range of each period. Then, each range should be rescaled by its standard deviation, and obtain a series of R/S observations. The slope of the log/log graph of R/S versus N will provide an estimate of the Hurst exponent. It is important to mention that the process can be continued for extended time intervals. But, as a direct consequence of extending the non-overlapping time intervals, the stability of the estimate is expected to decrease. In theory, long memory processes are supposed to last forever, but, according to Chaos theory, there is a point in any nonlinear system where memory of initial conditions is lost. This point of loss corresponds to the end of the natural period of the system. The visual inspection of data is a key factor for this process, to see whether such a transition phenomenon occurs. A regression can then be run over the range of the data, to show any evidence of a long memory process. In a similar manner, fractal scaling in other natural systems produces same effects. In a theoretical manner, all fractals scale forever, as for example the Sierpinski triangle. But natural fractals, like the human vascular system, do not scale forever. The natural fractal systems have a limit, and the long memory process underlying most systems is not is not infinite, but finite. Another important issue is the length of the time series needed to obtain an accurate view about the natural period of the system. In this respect, Chaos theory suggests that data from 10 cycles are enough. Applying the scrambling analysis, we may obtain the fact that the scrambling destroyed the long memory structure of the original series and turned it into an independent series.

CONCLUSIONS

The fractal nature of the capital markets contradicts the EMH and all the quantitative models that derive from it, and numerous other models that depend on the normal distribution and/or finite variance. The reasoning behind the failure of all those models resides in the fact that they simplify reality by assuming random behavior, and ignore the influence of time on decision making. By assuming randomness, the problem is simplified and „neat". Using random walk theory, we can identify „optimal portfolios" and „fair price". On the contrary, fractal analysis makes the mathematics more complicated for the modeler, but it brings the understanding closer to reality, due to its admitance of cycles, trends and many possible „fair values".

ACKNOWLEDGEMENT

This paper has been financially supported within the project entitled "*Routes of academic excellence in doctoral and post-doctoral research,* contract number *POSDRU/159/1.5/S/137926*, beneficiary: Romanian Academy, the project being co-financed by European Social Fund through Sectorial Operational Programme for Human Resources Development 2007-2013.

REFERENCES

Cootner, P.H. (1964). *The Random Character of Stock Market Prices*, Cambridge, MA: M.I.T. Press.

Cootner, P. (1964). Comments on the Variation of Certain Speculative Prices", in P. Cootner, ed., *The Random Character of Stock Market Prices. Cambridge,* MA:M.I.T. Press.

Hurst, H.E. (1951). Long-Term Storage of Reservoirs, *Transactions of the American Society of Civil Engineers*, 116.

Mandelbrot, B. (1960). The Pareto-Levy Law and the Distribution of Income, *International Economic Review* 1.

Mandelbrot, B. (1964). The Variation of Certain Speculative Prices, in P. Cootner, ed., *The Random Character of Stock Prices*, Cambridge, MA: M.I.T. Press.

Mandelbrot, B. (1972). Statistical Methodology for the Non-Periodic Cycles: From the Covariance to R/S Analysis, *Annals of Economic and Social Measurement*, 1.

Mandelbrot, B., & Van Ness, J. (1968). Fractional Brownian Motion, Fractional Noises and Applications, *SIAM Review*, 10.

Mandelbrot, B., & Wallis, J.R. (1969). Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence, *Water Resources Research*, 5.

Peters, E. E. (1991). *Chaos and Order in the Capital Markets – A New View of Cycles, Prices, and Market Volatility*, John Wiley & Sons, Inc.