IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 9, September 2019

Investigating Time Series Data for Self-Similarity Estimation

Satavisha Mitra[1], Vivekananda Mukherjee[2]

Department of Electronics & Communication Engineering, Techno International New Town, New Town, Kolkata[1,2]

Abstract: An investigation has been prepared to explore concept of self-similarity and stationarity nature of time series data of a stock market. Two time series data namely the Bombay Stock Exchange (BSE) as well as, the Stock Exchange of Hong-Kong (SEHK), the average parameters have been observed from January 2007 to November 2017. Analysis of both parameters has been carried out, through statistical techniques, to establish the nature of scaling pattern and non-stationarity. The calculation of Hurst exponent done by WVA and VGA showed that the time series is anti-persistent. Augmented Dicky Fuller Test (ADF), Kwiatkowski–Phillips-Schimdt-Shin test (KPSS) and Continuous Wavelet Transform (CWT) have been used to test for non-stationarity.

Keywords: Stock Market; Hurst Exponent; Fractality; Stationarity; Continuous Wavelet Transform

I. INTRODUCTION

The stock market, like any other market, is a place where goods and commodities are traded against prices those are determined by the twin forces of demand and supply. Stock markets and its operations are facilitated by stock exchanges. Stock markets are not just for retail traders and investors but are also for corporations and conglomerates. The stock exchanges in addition facilitate transparency in trades and set limits and regulations to restrict sophisticated entities to exploit the markets. In this context it is important to talk about two major players and their respective stock exchanges, one a major economic household and the other on the course to becoming the world's next superpower. We take the instance of Stock Exchange of Hong Kong Limited (SEHK) and India's BSE.

BSE, short for the Bombay Stock Exchange is India's largest and Asia's very first stock exchange located at Dalal Street, Mumbai. The BSE has a claim to being the world's fastest stock exchange with a median trading speed of just about 6 microseconds. The BSE Sensex a very popular stock index, comprising of the most established and financially viable companies of India, it is often used as a benchmark in comparing the country's economic growth. The Stock Exchange of Hong Kong Limited (SEHK) is a Hong Kong based stock exchange. It is the 3rd largest stock exchange in East Asia and Asia as per capitalization of market. Computer-assisted trading system was first introduced on 2.04.1986 by the exchange. The top 3 largest stocks by market capitalization listed on this exchange are Tencent Holdings, China Mobile, and HSBC Holdings. The last major stock disaster to hit this exchange happened in 2016 which was due to the EU referendum.

Many researchers have studied about the stock market and proposed suitable methods to detect the trends of stock market data. An evaluation has been carried out within the multidimensional framework of stock market volatility [1][2]. By carrying out experiments using neural networks it was concluded that with higher values of Hurst exponent, series can be foreseen better with respect to with the series having lower Hurst exponent value [3][6]. The ups and downs in stock market can be predicted by data mining. There is an influence of real-time news on the stocks. The SEHK stock markets are a favourable destination for best global investors and among all the Indian market shows the most dominance which was concluded using ADF, KPSS test for stationarity. The Indian stock markets do have some relation with some other markets around the world which was inferred using the Engle-Granger test for co integration [7]. Calculation of Hurst exponent by DFA and R/S fractal study serves as a measure of LRD in series [4][8][11]. Resolve of persistence of any trends and its predictability can be done by using the Hurst of Rescaled Range methodology on the stock market data of India for 19 years [5][9]. The multidimensional framework of stock market volatility has been evaluated [10].

In this document a proposal was undergone to expose the scaling nature and time reliance of the frequency of average value of sensex index of Bombay Stock Exchange (http://www.bseindia.com/) and Stock Exchange of Hong Kong Limited (SEHK) (https://finance.yahoo.com/) during January 2007 to November 2017 which may be regarded as a delegation of any Stock market analysis. Two diverse schemes like Visibility Graph Algorithm (VGA) and Wavelet Variance Analysis (WVA) have been utilized to gauge the Hurst Exponents to realize signal character with value to scales of a dissimilar nature for signal classification as in terms of Brownian motion of a fractional nature in order to check the stationarity or non-stationarity of the signals. For getting an incontestable termination concerning the time series' scaling, approximation of the Hurst Exponent by more than one technique is expected to be more helpful. Hence two processes (mentioned above) were selected to work out the Hurst Exponent for confirming the legitimacy of the finale taken away from the outcome. Two binary based methods namely ADF and KPSS tests and Time Frequency Representation (TFR) based method Continuous Wavelet Transform (CWT) have been incorporated for getting the unarguable conclusion regarding the stationarity/nonstationarity behavior of that specified profile.

II. MATERIAL & METHODOLOGY

A. Data

Analysis of stock market trends is extremely important for investors as well as business owners in the current scenario, since it gives an overview of the nature of classes of stock, which in turn can provide important market analysis insights for all concerned parties. It also has the general advantage that, rather than trying to predict market characteristics by following several major stocks, time-series analysis of index data offers a more probable set of market predictions based on index data variations. The current paper makes an effort to unravel the trends of stock market index data and investigate the nature of stationarity of the dataset. The average values of Hong Kong Stock Exchange Hang Seng Index, Hong Kong (https://finance.yahoo.com/) and BSE sensex, India (http://www.bseindia.com/) have been observed and examined for a period of 10 years from January 2007 to December 2017, for ensuring statistical significance of trend data. The summary statistics and plotting of both the time series data are represented in Table 1 and Figure 1 respectively. Summary statistics represents the statistical nature of that specified signal.

Table 1: Summary Statistics for HK and BSE stock exchange data

Score	SEHK	BSE
Mean	21684.83	20405.59
Median	22077.7	19285.91
Mode	27629.04	16248
Standard Deviation	3058.3	5304.53
Variance	9353400	28138042
Maximum	31619.77	31301.44
Minimum	11636.23	8184.66
Skewness	0.7938	.079264
Kurtosis	4.8189	2.2991

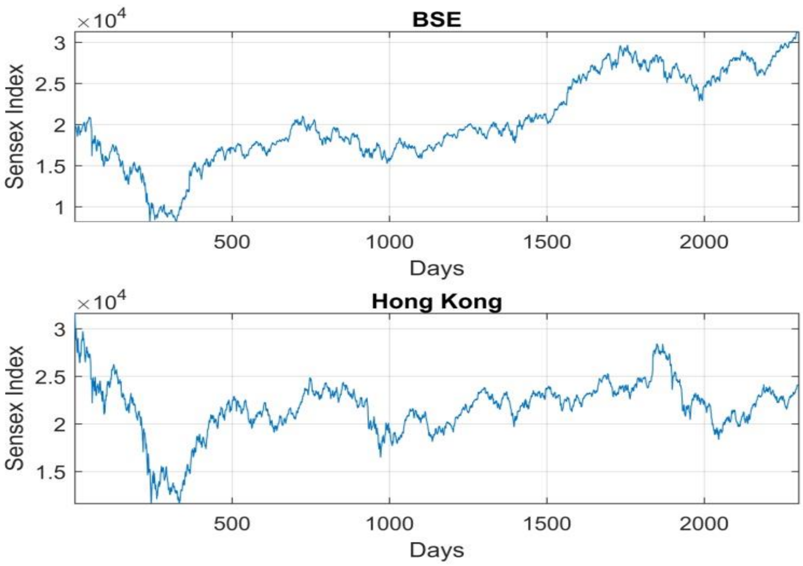

Fig.1. Plot of Daily average Sensex value for SEHK and BSE

B. Method of Analysis

A Hurst exponent assessment between 0 and 0.5 is analytical of anti-persistent activities and as its value approaches 0, the time series has a greater tendency to relapse to its long-term mean values. A Hurst value between 0.5 and 1 shows persistent activities. Superior the H, stronger is the trend. An anti-persistent sequence has an attribute of "mean reverting", which connotes an up value is likely to be succeeded by a down value, and so on. The strength of "mean reverting" increases as H approaches 0.0. A persistent series is tendency emphasizing, which means the direction (up or down evaluated to the last value) after that value is more probable to be same as existing value. The power of drift augments as H move towards 1 [12][13][14].

In regulation to calculate Hurst Exponent for any time series data we know how to adapt any estimator as of a large number of estimators. For this paper, two major methods have been used for the calculation of Hurst value i.e. Visibility Graph Analysis (VGA) and the Wavelet Variance Analysis.

1. Visibility Graph Algorithm (VGA)

In the approach developed in this paper the time series is converted to a graph based on visibility criteria. The basic idea of visibility graph algorithm can be described as follows: $X = \{x_i\}_{=1,\dots,n}$ are time series with n nodes; if any two nodes (t_a, x_a) and (t_b, x_b) are visible with each other, then $\forall(t_i, x_i) \in X$ with $t_a<t_i<t_b$ should satisfy the constraint of

$$x_i < x_b + (x_a - x_b)\frac{(t_b - t_i)}{(t_b - t_a)} \qquad (1)$$

The complex networks after transformation can ensure that each node except the first and the last time point can at least connect with two adjacent ones. The nature of variation inherent in the series can be determined based on the visibility graph. For example, periodic series map into regular graphs, random series into random graphs and fractal series into scale free graphs. Hence a fractal series having a power-law distribution $M(x)= x^{-\alpha}$ can be found to be representable in the form of a scale free graph. The mathematical form of the exponent α is

$$\alpha = 1 + N\left[\sum_{i=1}^{N} \log\frac{x_i}{x_{min}}\right]^{-1} \qquad (2)$$

Here N represents the total numbers of values, x_i $(i = 1,2,\dots,N)$ are the values of the time series and x_{min} is the minimum value of x exhibiting power law behavior. α is linearly related with the Hurst Exponent (H) [15]. Wavelets are well localized in both time and frequency domain whereas the standard Fourier transform is only localized in frequency domain.

2. Wavelet Variance Analysis (WVA)

In order to get the characteristics of time series, there is an entirely different research approach that is Wavelet analysis [15]. This method is explained below:

Continuous wavelet transform, or coefficient, for time series $f(t)$ is given by:

$$W(t,a) = \frac{1}{a^{\frac{1}{2}}}\int_{-\infty}^{\infty} \dot{g}\ t,a(t°)f(t°)dt° \qquad (3)$$

Here, $\dot{g}$ represents the complex conjugate form of g which is the mother wavelet. Considering scale or dilatation parameter $\dot{g}t$ and translation parameter t, a $(t°)$is given by:

$$\dot{g}t, a\ (t°) = \frac{1}{a^{\frac{1}{2}}}\ g\left(\frac{t°-t}{a}\right) \qquad (4)$$

The Scalogram is obtained as the plot of the square of the modulus of $W(t,a)$ in $t-a$ plane.

$$\text{Variance of } (t,a), v(a) = E(W^2) - E(W)^2) \qquad (5)$$

If the scaling of the time series is linear, v(a) is found to follow a power law of a, represented as $v(a) \sim a^{\delta}$ considering δ to be the exponent of variance of the wavelet.

Double log plot of $v(a)$ versus a is used to determine δ and its values are found to range between 1 and 3, that is, $1 \leq \delta \leq 3$. Considering a fractional Gaussian noise (FGN) type signal, $-1 \leq \delta \leq 1$ while considering Fractional Brownian Motion (FBM), $1 \leq \delta \leq 3$.

The wavelet Hurst exponent H_w can be obtained from δ as:

$$H_\omega = \begin{cases} \frac{\delta+1}{2} & -1 \leq \delta \leq 1 \text{ (FGN)} \\ \frac{\delta-1}{2} & 1 \leq \delta \leq 3 \text{ (FBM)} \end{cases} \quad (6)$$

In order to perform the wavelet analysis we also need to select what family of wavelet filters to use. In our case we used the Haar (discrete) filter [16].

C. Test for Stationarity / Non-Stationarity

A stationary process is one where invariance is observed in the mean, variance and autocorrelation structures over time. A non-stationary process has a variable variance and tends to assume values far from a long-run mean estimate, and does not converge to the estimated long-run mean over time. A stationary process on the other hand converges to a constant long-term mean, and is also observed to have a constant time-independent variance. We implemented the visibility-graph construction algorithm in a brute-force manner. If algorithms such as sweep line are used, the implementation will be much more efficient.

1. Kwiatkowski–Phillips–Schmidt–Shin (KPSS)

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test are used for testing a null hypothesis for confirming the stationarity of a given time series. A stationary time series is one wherever statistical properties remain invariant over time. The null hypothesis for the test assumes stationarity of the given dataset. Hence the alternate hypothesis for the test infers that the dataset under consideration is non-stationary. The KPSS test works on the principle of linear regression. It breaks up a series into a deterministic trend (βt) and a random walk (r_t) with a stationary error (ε_t), represented by the equation of regression shown in equation 7 below.

$$x_t = r_t + \beta t + \varepsilon_t \quad (7)$$

This test can be considered a Lagrange multiplier test of the hypothesis that a random walk along the dataset given will be found to have the variance parameter equal to zero.

2. Augmented Dickey Fuller (ADF)

Augmented Dickey Fuller test was developed by David Dickey and Wayne Fuller. The Augmented Dickey–Fuller test (ADF) tests the presence of a unit root in a time series sample.
The procedure of application of the ADF test is similar to the Dickey–Fuller test; however it is applied to the model:

$$\nabla Y(t) = \alpha + \beta t + \gamma Yt - 1 + \cdots + \delta p - 1 + \nabla Yt - p + \quad 1 + \varepsilon t \quad (8)$$

Here α is assumed to be a constant, β represents the time trend coefficient with p having the lag order of the autoregressive process. The test for obtaining unit root is carried out under the null hypothesis $\gamma=0$ against the alternative hypothesis of $\gamma<0$. The statistic corresponding to the test is obtained from equation 8 shown below.

$$DF\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (9)$$

The value of the statistic obtained in this manner can be compared to the relevant critical value for the Dickey–Fuller Test.

ADF is utilized over DF for a larger and more complicated set of time series models. There are 2 hypotheses for the experiment: 1) The null hypothesis for this test assumes the existence of a unit root. The alternate hypothesis differs slightly according to the equation that is used. The alternate hypothesis generally used is that the time series is stationary (or trend-stationary). The major difference between ADF and KPSS is that their null hypotheses are complementary i.e. for ADF the null hypothesis states that the series should be non-stationary and for KPSS it should be stationary. As a result of KPSS complementary result it is often used in conjunction with ADF to reinforce the results obtained. For our time series data when these tests are performed complementary results from ADF test and KPSS test are obtained. Both of them agree upon the fact of non-stationarity of the series. The null hypothesis is rejected by KPSS while ADF accepts it. Therefore a consensus is achieved from this that the times series are non-stationary.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 9, September 2019

3. Continuous Wavelet Transform (CWT) Method

Actual global information or signals normally expose the slow altering trend or oscillations scattered with transient. Though Fourier Transform (FT) is a potential method for statistical examination, but it doesn't characterize sudden transforms proficiently. FT indicates data as computation of sine waves that are not contained in time or space. Waves fluctuate eternally, thus to precisely scrutinize signals that have sudden changes, must apply fresh set of functions which are contained with time and frequency. This introduced the subject of wavelets.

The primary goal of the Continuous Wavelet Transform (CWT) [17] [18] is concerned with obtaining the signal's energy allocation concurrently in the time domain as well as the frequency domain. The continuous wavelet transform is an oversimplification of the Short-Time Fourier Transform (STFT) that allows for the analysis of non-stationary signals at multiple scales. Key features of CWT are time frequency investigation and straining of time localized frequency components. The mathematical equation for CWT is given below:

$$C(a,\tau) = \int \frac{1}{\sqrt{a}} \psi\left(t - \frac{\tau}{a}\right) x(t)\, dt \qquad (10)$$

Where C (a, τ) is the function of the parameter a, τ .

The parameter a is the dilation of wavelet (scale) and τ defines a translation of the wavelet and indicates the time localization; ψ (t) is the wavelet. The coefficient $\frac{1}{\sqrt{a}}$ is an energy normalized factor (the energy of the wavelet must be the same for different a value of the scale).

III. RESULTS

The values of Hurst exponents for the two time series like HK and BSE has been calculated using the two methods, WVA and VGA which are being tabulated below in TABLE 2.

Table 2: Calculated Hurst Exponent values for BSE and HK.

Methods	Hurst exponent (H)	
	HK	BSE
WVA	0.4568	0.3546
VGA	0.4340	0.0678

From the table it is found that Hurst value for both the case is less than 0.5. Now this indicates that the stock prices are highly volatile and anti-trended. Also as the values are less than 0.5 and greater than 0 they show anti-persistent behaviour.

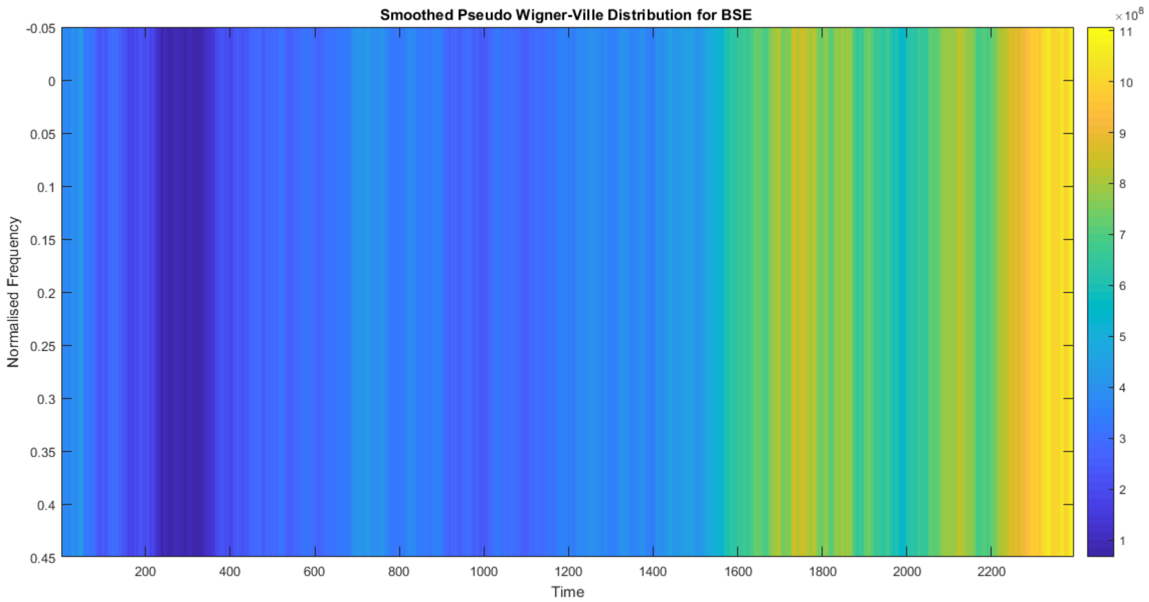

Fig. 2. CWT analysis for SEHK data

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 9, September 2019

This anti-persistent behaviour show that time series has a tendency to revert back to its long term value. Both the series shows a tendency to revert back this means that some sort of function drives them both to return to its long term value. Now the most stable long term value would be that of the mean as it is most stable point for any kind of change. This function consistently drives them to get stabilised. The calculated binary value of ADF for both SEHK and BSE are logical 0 and KPSS are logical 1 respectively. Both of them agree upon the fact that the series is non-stationary. KPSS rejects the null hypothesis while ADF accepts it. Therefore a consensus is achieved from this that the times series are non stationarity in nature. Scalogram Percentage energy for each wavelet coefficient for the two time series are shown in Figure 2 and Figure 3 respectively.

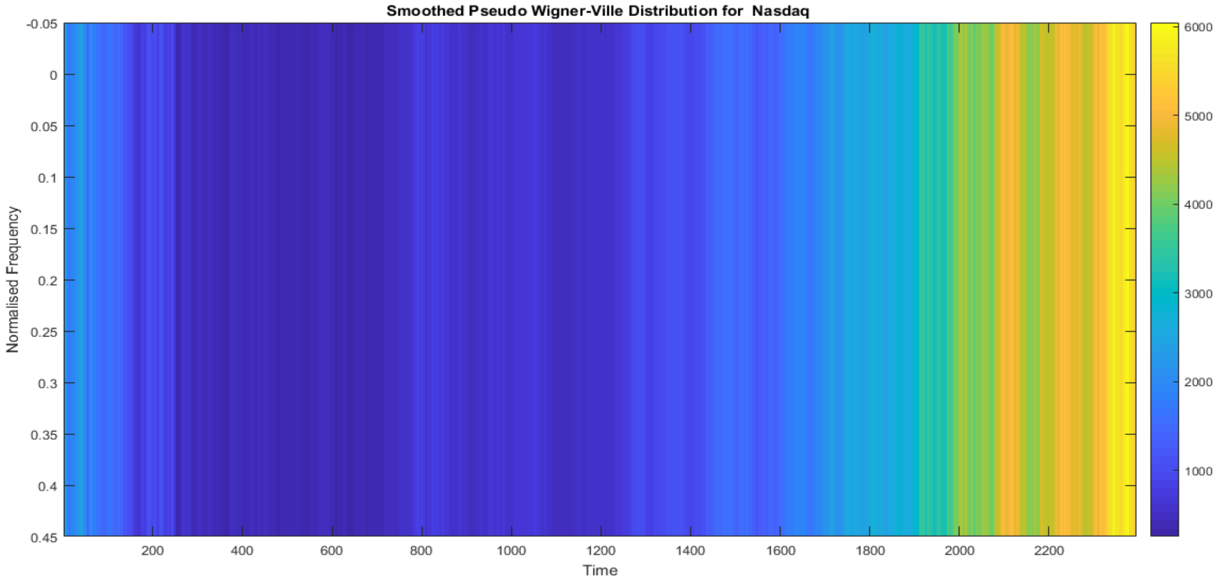

Fig. 3. CWT analysis for BSE data

It can thus be inferred that the small CWT coefficients dominate the majority of the time-scale plane and also that density of the various CWT coefficients vary considerably all through, giving an indication of non-stationarity in the input time series. Relatively small portions of the Scalogram are occupied by the intermediate sized CWT coefficients towards the beginning and large sized CWT coefficients towards the end. Both the time series of stock market data are non stationary in nature.

IV. CONCLUSION

The value of Hurst Exponent of any method is larger than 1/2 or may be fewer than 1/2 however unequal to 1 which normally confirms that the dynamics of the system are nonlinear. Hence in conclusion we may consider both time series to be nonlinear in terms of the dynamics since the Hurst values obtained are less than 1/2. In addition the anti-persistent behavior of average value of sensex of both BSE and SEHK give the outline of the subsistence of a little unconstructive response system that desires to be exposed more in the successive work. The small value of H denotes more steadiness of the BSE than that of the SEHK. Again it is found that there is non-Stationary in both the time series. So, it can be concluded that the SEHK and BSE are not a random trend somewhat it is multifaceted and non-linear, stable process. As the sensex value of SEHK and BSE are used as the vital figures of worth to evaluate the economical stability of country, the consequent mechanism is to discover the character of the non-linear dynamics and originate a representation following current effort conclusions. A natural extension to this work is the analysis and comparison between the remaining stock markets which would also help in understanding the markets and their behavior to setting up a model for the prediction of the market's future growth. The accuracy of the model would help in analyzing future worth in a better way.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 9, September 2019

REFERENCES

[1]. Stock Market Volatility: An Evalution. Bhowmik, Dr. Debesh. 10, October 2013, International Journal of Scientific and Research Publications, Vol. 3. ISSN 2250-3153.
[2]. Volatility patterns of stock returns in India. Kalra, Rosy & Pandey, Mr.Piyuesh. 2015, NMIMS Mngmnt Review, Vol.XXVII. ISSN:0971-1023
[3]. Predicting drastic drop in Chinese stock market with local Hurst exponent. Xu, Shao-jun & Jin, Xue-jun. 2009, Management Science & Engg.
[4]. Long-range dependence in Indian stock market: a study of Indian sectoral indices. Kumar, Dilip. 4, 2014, International Journal of Emerging Markets, Vol. 9, pp. 505-519.
[5]. Identification of persisting trend in the Indian stock markets using hurst exponent. Jamwal, Naina Jamwal and Malhotra, Akshay. Bangalore , 2017. Recent Trends in Electronics, Information & Communication Technology (RTEICT)
[6]. The Hurst Exponent: Predictability of Time Series. Mansukhani, S. s.l., 2012. Analytics magazine.org. Vol.4, pp. 29-46.
[7]. Stationarity and Engle-Granger test of co integration.FaikBilgili, , posted 4 January 2017. Online at https://mpra.ub.uni-muenchen.de/75967/ MPRA Paper No.75967.
[8]. The characteristics of wireless network traffic : fractality and staionarity. Mukherjee, S, et al. [ed.] 2017. ICRCICN 17,Kolkata : IEEE.
[9]. Approach to An Irregular Time Series on the Basis of the Fractal Theory. Higuchi, T. 1988, Physica D, Vol. 31, pp. 277–283.
[10]. Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. Gómez, C., et al. 3, 2008, Medical Engineering Physics, Vol. 31, pp. 306-313.
[11]. ‘Comparative Analysis for Estimating of the Hurst exponent for Sstationary and nonstaionary time series’ Ludmila Kirichenko, Tamara Radivilova, Zhanna Deineko, 2011. International Journal "Information Technologies & Knowledge" Vol.5 page-371.
[12]. Features Affecting Hurst Exponent estimation on time series. Wairimu, M J. 2013. Jomo Kenyatta University Of Agriculture and Technology. Juja : University of Kenya.
[13]. Fractals and scaling in finance. Mandelbrot, Benoit. New York, 1997 : springer.
[14]. Dynamical Generalized Hurst exponent as a Tool to Monitor Unstable Periods in Financial Time Series. Morales, Raffaello, Matteo, T. Di and Gramatica, Ruggero. 2012. Physica A. Vol. 391, pp. 3180–3189.
[15]. Research on the Weak Signal Detection Based on Adaptive Filtering Of Wavelet Transform: JingLia,HongLib,ZhiyongLeib, 2011. ELSEVIER, Procedia Engineering ,Volume 15, , Pages 2583-2587.
[16]. Document Text Extraction from Document Images Using Haar Discrete Wavelet Transform: S.Audithan, 2009. European Journal of Scientific Research ISSN 1450-216X Vol.36 No.4, pp.502-512.
[17]. Continuous wavelet transforms: Shi Yunhui , 2004. Signal Processing Proceedings. ICSP '04. 7th International Conference.
[18]. Applications of the continuous wavelet transform in the processing of musical signals : P.De.Gersem, 1997.Digital Signal Processing Proceedings.13th International Conference.