Application of Multiscale entropy analysis to verification of the applicability of Efficient Market Hypothesis.

B.G. Sharma[1], D.P. Bisen[2], Ravi Sharma[3] and Malti Sharma,[4]

1. Department of Physics and Computer science, Govt. Science College Raipur 492010 India
   Email - sharma_balgopal@rediffmail.com
2. School of studies in Physics, Pt.Ravishankar Shukla University Raipur 492010 India
   Email - dpbisen@rediffmail.com
3. Arts & Commerce Girls College Devendranagar, Raipur,(C.G.) 492001, India
   Email - rvsharma65@gmail.com
4. WQ-1,Govt. Science College Raipur 492010 India
   Email - sharma_balgopal@rediffmail.com

Abstract - *Traditional entropy based algorithms used in the analysis of time series data quantify the regularity of the time series. But there is no straight forward correspondence between regularity and complexity. Neither completely predictable (e.g. periodic) signal, which have minimum entropy, nor completely unpredictable (e.g. uncorrelated random) signals which have maximum entropy, are truly complex, since they can be described very compactly. Entropy increases with disorder, however, an increase in entropy may not always be associated be with increase in dynamic complexity. Thus the traditional algorithms may generate misleading results because the algorithms are based on single time scale .However, the multiscale entropy (MSE) approach measures the complexity of the system taking into account the multiple time scales. This computational tool can be quite effectively used to quantify the complexity of a given time series.*

In this paper the behavior of Indian stock market index NIFTY is studied using MSE approach and it is shown that the market exhibits different MSE patterns at different level of information received. However this difference in MSE profile disappears when the time scale is increased. Since the complexity of the market changes with information this approach can be used to verify the applicability of Efficient Market Hypothesis(EMH) and to test the time scales above (below) which market behave efficiently(inefficiently).

1. INTRODUCTION:

Nonlinear dynamic analysis techniques empowered by progress in the tools of complexity, deterministic chaos and fractals etc. is clearly outperforming the linear methods in analysis of real time series data [1]. Non linear methods consist of an array of toolkits ranging from all kind of correlation dimension calculations, Hurst dimension, Lyapunov exponents, fractal dimensions - both in time as in phase space domain, several types of entropy and complexity measurements and algorithms/methods to estimate the state space embedding

dimension ("embedology") and time delays (false nearest neighbors, autocorrelation, mutual information) etc. Yet these approaches might not be without some problem and pitfalls. Hence a cautious approach is required otherwise a blind application of the methodology on non linear signal may give misleading results.
It is estimated that for Grassbergers and Proccacia correlation dimension a long enough data segment is necessary of length n so that $D_2 < 2$ 1ogn [2]. Such long segments will probably contain artifacts and non-stationarities that could jeopardize the whole analysis. Takens Theorem states that if there is deterministic chaos then the fractal attractor can be reconstructed in a time delay space of appropriate dimension constructed from only the one dimensional realization (the signal at hand), this does not imply the reverse[3]. Calculating some non integer dimensionality is evidently no sure proof of existence of deterministic chaos [4]. The linear noise filtering of true random series also suggests the presence of deterministic chaos [5-6]. Hence the correlation dimension cannot be used as an absolute detector of chaos. Thus while using these techniques it is imperative to prevent false "positive" conclusions.

Thus it will be useful to focus on model free approach and try to demonstrate alternation in signal complexity using time domain fractal dimension. Multiscale entropy analysis (MSE) is such a technique that is robust, less model dependent (can be applied to deterministic chaos, stochastic as well as periodic signals), can be used on relative short signal segments and is less noise sensitive[4].
This method signals (white-noise), which are highly unpredictable but not structurally “can effectively be used in measuring the complexity of finite length time series. This computational tool can be applied to various types of physical data sets and can be used with variety of measures of entropy. Whereas traditional methods quantify the degree of regularity of a time series by evaluating the appearance of repetitive patterns, there is no straight forward correspondence between regularity and complexity. Neither completely predictable (e.g. periodic) signal, which have minimum entropy, nor completely unpredictable (e.g. Uncorrelated random) signals which have maximum entropy, are truly complex, since they can be described very compactly. There is no consensus definition of complexity. Intuitively complexity is associated with “meaningful structural richness”, which in contrast to the outputs of random phenomena, exhibits relatively higher regularity. Entropy based measures, such as the entropy rate and the Kolmogorov complexity, grow monotonically with the degree of randomness. Therefore their measures assign the highest values to uncorrelated random complex” and at a global level, admit a very simple description.

Thus, when applied to a given time-series, traditional entropy-based algorithms may lead to misleading results. For example, they may assign higher entropy value to the data set that generate erratic outputs than to the output of the systems that are exquisitely regulated by multiple interacting control mechanisms. Substantial attention, therefore, has been focused on defining a quantitative measurement of complexity that assign minimum values to both deterministic/predictable and uncorrelated random / unpredictable signals.
The remainder of this paper is organized as follows: section 2 provides a theoretical background and a review of related previous research. In Section 3, we discuss the methodology applied to different data sets and its analysis. In Section 4 we present the result and the discussion of the study and finally section 5 contains References.

2.1. EFFICIENT MARKET HYPOTHESIS(EMH):

The efficient-market hypothesis was developed by Professor Eugene Fama[8]. Although the efficient-market hypothesis has become controversial because substantial and lasting inefficiencies are observed, Beechey et al. consider that it remains a worthwhile starting point [9]. Fama published a review of both the theory and the evidence for the hypothesis. The paper extended and refined the theory, included the definitions for three forms of financial market efficiency: weak, semi-strong and strong [10]. Further to this evidence that the UK stock market is weak-form efficient, other studies of capital markets have pointed toward their being semi-strong-form efficient. A study by Khan indicated semi-strong form efficiency following the release of large trader position information [11]. Studies by Firth shows that the share prices were fully and instantaneously adjusted to their correct levels, thus concluding that the UK stock market was semi-strong-form efficient[12,13].

Beyond the normal utility maximizing agents, the efficient-market hypothesis requires that agents have rational expectations; that on average the population is correct (even if no one person is) and whenever new relevant information appears, the agents update their expectations appropriately. Note that it is not required that the agents be rational. EMH allows that when faced with new information, some investors may overreact and some may underreact. All that is required by the EMH is that investors' reactions be random and follow a normal distribution pattern so that the net effect on market prices cannot be reliably exploited to make an abnormal profit, especially when considering transaction costs (including commissions and spreads). Thus, any one person can be wrong about the market—indeed, everyone can be—but the market as a whole is always right. There are three common forms in which the efficient-market hypothesis is commonly stated—weak-form efficiency, semi-strong-form efficiency and strong-form efficiency, each of which has different implications for how markets work.

In weak-form efficiency, future prices cannot be predicted by analyzing prices from the past. Excess returns cannot be earned in the long run by using investment strategies based on historical share prices or other historical data. Technical analysis techniques will not be able to consistently produce excess returns, though some forms of fundamental analysis may still provide excess returns. Share prices exhibit no serial dependencies, meaning that there are no "patterns" to asset prices. This implies that future price movements are determined entirely by information not contained in the price series. Hence, prices must follow a random walk. This 'soft' EMH does not require that prices remain at or near equilibrium, but only that market participants not be able to systematically profit from market 'inefficiencies'. However, while EMH predicts that all price movement (in the absence of change in fundamental information) is random (i.e., non-trending), many studies have shown a marked tendency for the stock markets to trend over time periods of weeks or longer[14] and that, moreover, there is a positive correlation between degree of trending and length of time period studied[15] (but note that over long time periods, the trending is sinusoidal in appearance). The problem of algorithmically constructing prices which reflect all available information has been studied extensively in the field of computer science [16].

In semi-strong-form efficiency, it is implied that share prices adjust to publicly available new information very rapidly and in an unbiased fashion, such that no excess returns can be earned by trading on that information. Semi-strong-form efficiency implies that neither fundamental analysis nor technical analysis techniques will be able to reliably produce excess returns. To test for semi-strong-form efficiency, the adjustments to previously unknown news must be of a reasonable size and must be instantaneous. To test for this, consistent upward or downward adjustments after the initial change must be looked for. If there are any such adjustments it would suggest that investors had interpreted the information in a biased fashion and hence in an inefficient manner.

In strong-form efficiency, share prices reflect all information, public and private, and no one can earn excess returns. If there are legal barriers to private information becoming public, as with insider trading laws, strong-form efficiency is impossible, except in the case where the laws are universally ignored. To test for strong-form efficiency, a market needs to exist where investors cannot consistently earn excess returns over a long period of time. Even if some money managers are consistently observed to beat the market, no refutation even of strong-form efficiency follows: with hundreds of thousands of fund managers worldwide, even a normal distribution of returns (as efficiency predicts) should be expected to produce a few dozen "star" performers.

It has been established that that complexity changes with information. Thus entropy approach can be a useful one for the test of the efficiency of markets.

2.2 TESTS OF MARKET EFFICIENCIES:

A number of different approaches were used to test the efficient market hypothesis. One of the most obvious ones was to perform studies on serial correlation of security prices[17]. A variation of this approach was to test various trading strategies recommended by technical analysts to see if they have any investment value. Both have been tried, and invariably came back with mostly negative results. It turned out that stock returns are not normally distributed. They follow some sort of distribution, but, to our knowledge, no one has figured out exactly what kind of distribution it is. On several occasions, stable Paretian distribution and Student *t*-distribution or Levy stable distribution were found to be better approximations than the normal distribution. Needless to say, this poses a huge methodological problem for researchers who, for lack of a better assumption, are still assuming normal distributions for drawing statistical inferences. An important breakthrough in testing market efficiency came with the advent of the "event study" methodology [18]. According to FFJR findings, the market begins to anticipate an event like a stock split more than two years before it actually happens and figures out the consequences of the split the day it is announced. The event study techniques were further refined by other researchers. Some of the research designs are quite clever [19]. By 1975, the preponderance of evidence argued that markets were efficient. Statistical studies showed that technical analysis did not add value (consistent with the weak form of market efficiency). Event studies found that the market quickly reacts to new information (consistent with the semi-strong form of market efficiency). And studies of professional investors' performance made a strong case for the strong form market efficiency. However as more and more researchers tested the efficient market hypothesis, some rather controversial evidence also began to appear [20].
This is a good point at which to consider the efficient market hypothesis and identify those assumptions that may be inconsistent with reality as we know it. First of all, as ironic as it sounds, there is no way to test market efficiency per se. We can only test a joint hypothesis stating that, first, the market is efficient in equating asset prices with their intrinsic values, and, second, we know what the intrinsic values are. Whenever an anomaly is found, we don't know (and have no way of knowing) which part of this joint hypothesis did not work. Returning to Fama's definition of an efficient market, he assumes that important current information is almost freely available to all participants. This appears to be an accurate assumption. Ultimately the information has to be available to everyone,

there is matter of time only. Thus if fluctuations are ignored, an inefficient market at small scale seems to be efficient at larger time scales.

2.3 MULTISCALE ENTROPY ANALYSIS:

The MSE method is based on the estimation of sample entropy (SampEn) [21] which is a refinement of the approximate entropy family of statistics introduced by Pincus[22]. Sample entropy is a statistical measure proposed by Richman and Moorman which quantifies the variability of time-series by comparing sequence of consecutive data point. It provides a measure of the regularity or predictability of a time-series (high complexity). Sample entropy is derived from the conditional probability that sequence of data-point is within a certain tolerance range for a number of steps. Sample entropy depends on the length of the series. Though can't be used to distinguish between signals of similar form but different frequency. A signal which contains noise and has a certain period no more complex than the same quantity of data but with a different periodicity.

Due to the interrelationship of entropy and scale, which is incorporated in the MSE analysis, the results are consistent with the consideration that both completely ordered and completely random signals are not really complex. In particular, the MSE method shows that uncorrelated random signals (white noise) are less complex than correlated random signal. When the MSE Result for white noise is compared with 1/f noise (pink noise) it is found that for scale one, a higher value of SampEn is obtained for white noise than pink noise. Although the value of entropy for the coarse-grained 1/f noise series remains constant for all scales, the value of entropy for the coarse-grained white noise time series monotonically such that for scales above 4, it becomes smaller than the corresponding value for 1/f noise. In contrast with single-scale entropy based analysis, the MSE results are consistent with the fact that unlike white noise, 1/f noise contains correlations across multiple time scales and is, 1/f therefore, more complex than white noise.

3. MSE METHOD

MSE method depends on coarse-graining procedure of the given series. It incorporates two steps:

1.Consider a given time series $\{x_i\} = \{x_1 x_2 x_3 \dots \dots x_n\}$

The length of the series is N. Then we construct consecutive coarse-grained time series by averaging a successively increasing number of data points in nonoverlapping windows. Figure1 shows an schematic illustration of the coarse-graining procedure for scale 2 and 3. Each element of the coarse-grained time series, $y_j^{(\tau)}$ is calculated accordingly to the equation

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{i\tau} x_i \qquad (1)$$

Where, τ represents the scale the factor and $i \leq j \leq \frac{N}{\tau}$ The length of each coarse-grained time-series is N/τ . For scale one, the time series $\{y^{(1)}\}$ is simply the original time-series.

2. Finally, we calculate sample entropy (Samp En) for each coarse-grained time-series, and then SampEn is plotted as a function of the scale-facto

Let

$$U_m(i) = \{x_i, x_{i+1} \dots \dots x_{i+m-1},\} \qquad 1 \leq i \leq N-m \qquad (2)$$

be vectors of length m. Let $n_{im}(r)$ represent the number of vectors $U_{m(i)}$ Within distance r of $U_{m(i)}$ where j range from 1 to (N-m) and j≠ 1 to exclude self matches.

$$C_i^m(r) = \frac{n_{im}(r)}{N-m-1} \qquad (3)$$

Is probability that any vector $U_m(i)$ is within tolerance range r of $U_m(i)$ We then define

$$U^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} ln\, C_i^m(r) \qquad (4)$$

The parameter Sample entropy (Samp En) is defined as

$$SampEn\ (m, r) = \lim_{N\to\infty}\left\{-ln\frac{U^{m+1}(r)}{U^m(r)}\right\} \qquad (5)$$

For a Time Series of finite length (N), the sample entropy is estimated by statistics,

$$SampEn\ (m, r, N) = \left\{-ln\frac{U^{m+1}(r)}{U^m(r)}\right\} \qquad (6)$$

Sample entropy is the natural logarithm of the ratio of the total number of two components templates matches to the total number of three components templates matches.

For scale one, the value of entropy is higher for the white noise time series in comparison to the 1/f noise. This result explains the facts that the 1/f noise contains complex structures across multiple scales in contrast to the white noise.

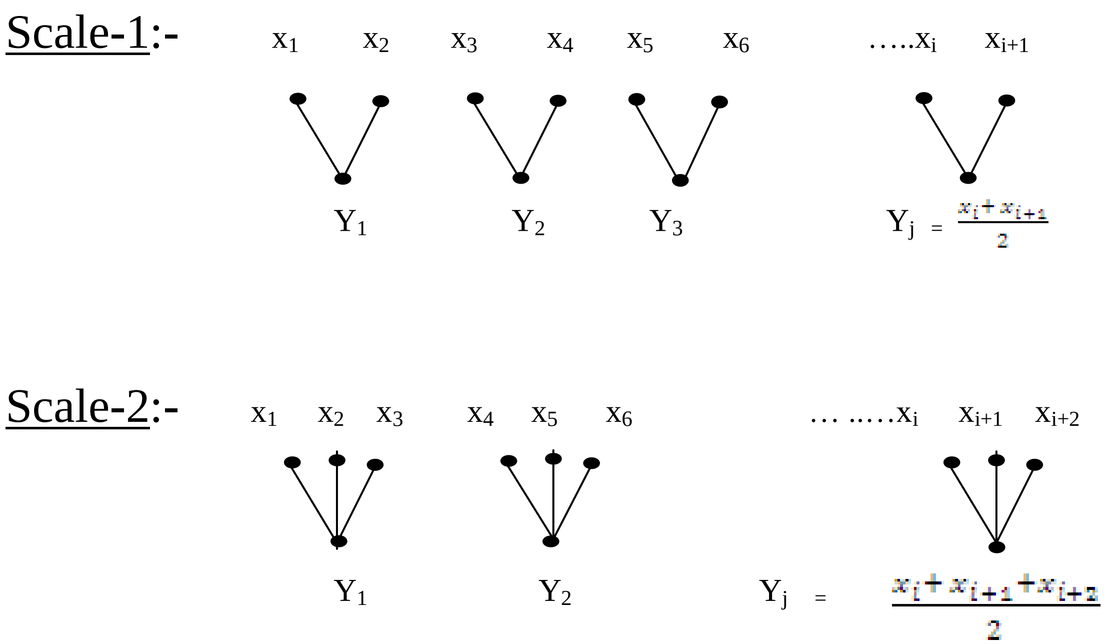


Fig.1- Course - Graining procedure of time series data.

4. DATA ANALYSIS :

The whole study is divided into three parts. In first part we analyzed the data base of various sectoral indices of National Stock Exchange of India from tt5 (Advance) of India Infoline Sec[7]. Daily closing values of indices are taken from 17^{th} june 2003 to 9^{th} feb 2010. Fig 2(a)-2(b) show the daily variations of various secrtoral indices and fig 3(a)-3(b) show their MSE profile respectively. Fig 4(a)-4(b) shows the variation of tick value of NIFTY for pre budget hours, just after the budget announcement and at later hours of the market. Fig 5(a) - 5(b) show their MSE profile respectively. Fig 6(a) shows the variation in the tick value of NIFTY for 24th feb 2010 to 2 march 2010 in which there are two pre- budget days one budget day and one post- budget day. Fig 6(b) shows their MSE profiles.

The data base of various sectoral indices of National Stock Exchange of India from tt5 (Advance) of India Infoline Sec. Daily closing values of indices are taken from 17th june 2003 to 9th feb 2010. Total no of data points are taken to be 1623.

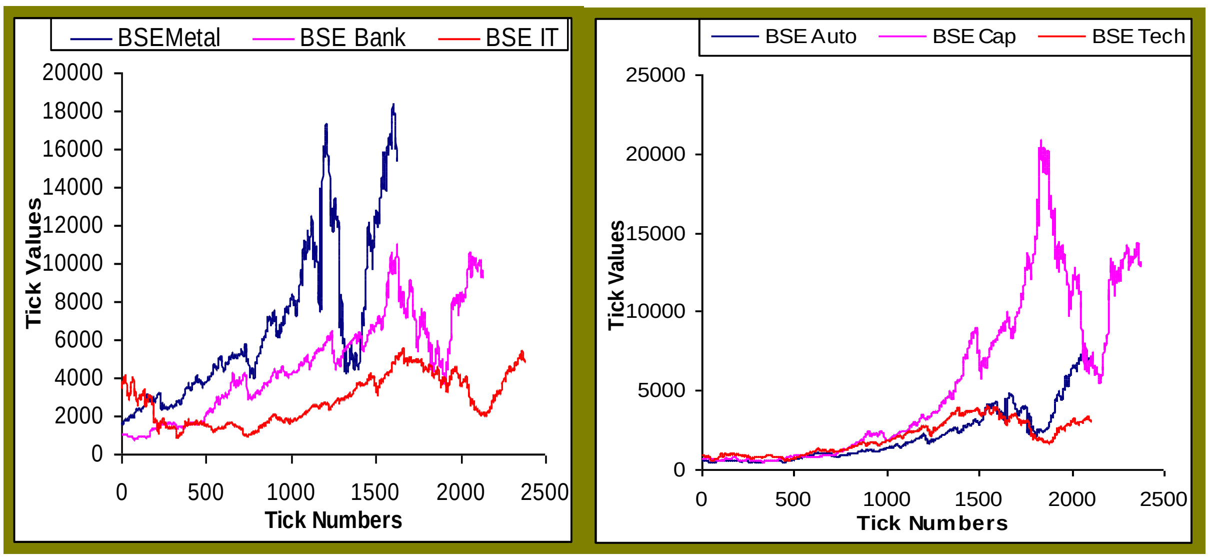
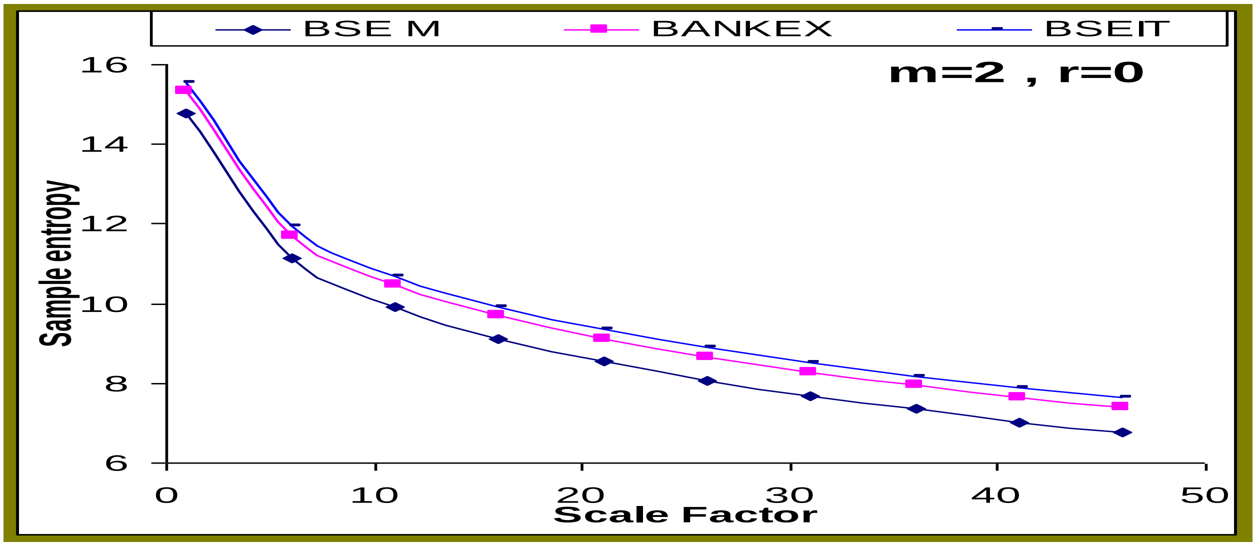


Fig 2(a): Values of sectoral indices of Indian stock market (Metal, Bank and IT sectors); Fig 2(b): Values of sectoral indices of Indian stock market (Auto, Cap. Good and Tech. sectors)



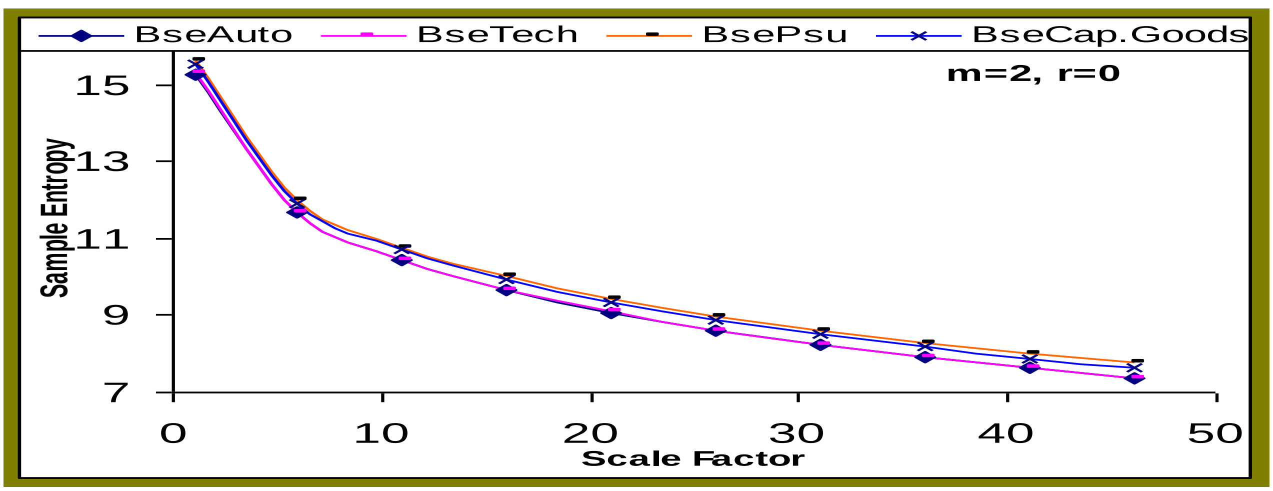


Fig3(a-b) MSE Profile of Various Sectoral Indices of Bombay Stock Exchange of India

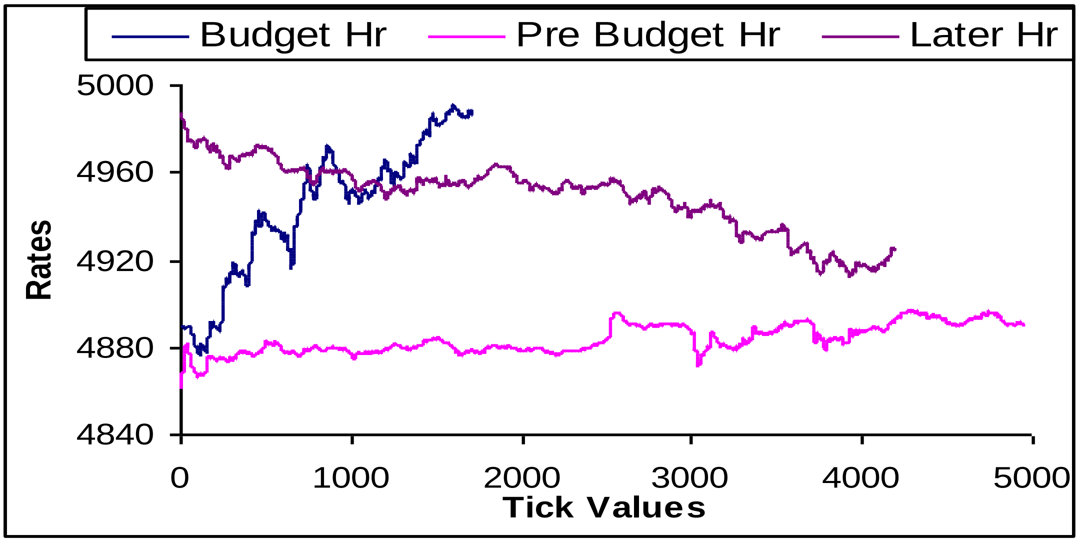


Fig.4. Tick Value of NIFTY for Pre Budget, Budget hours & post hours of the Market.

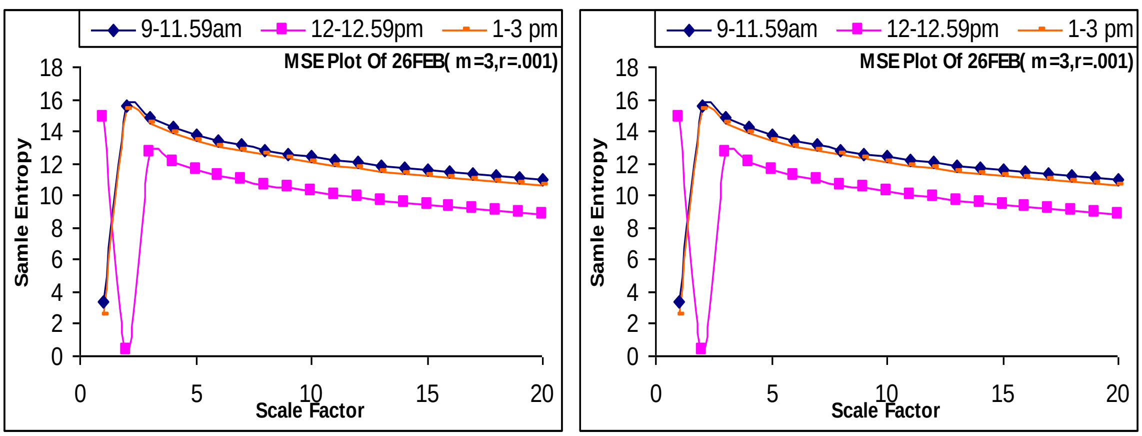


Fig 5(a-b) MSE Profile of Pre Budget, Budget, and Post budget hours of the market.

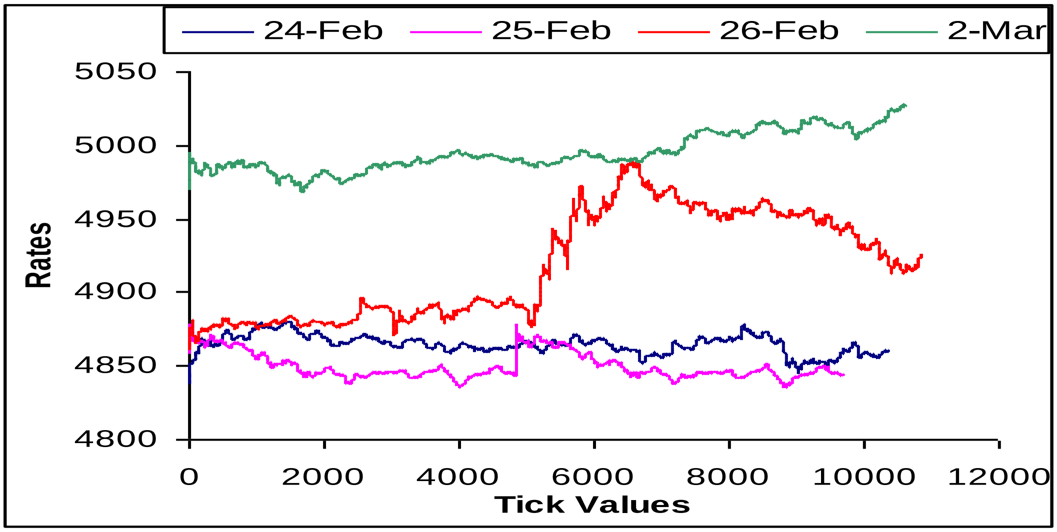


Fig 6(a) Variation in the Tick value of NIFTY for 24th feb 2010 to 2 march 2010

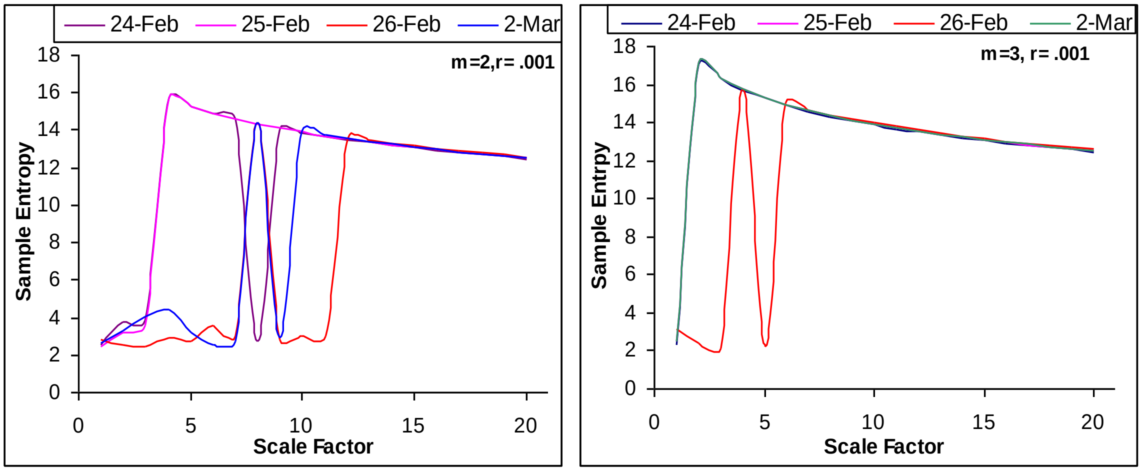


Fig 6(b) MSE Profile of Tick value of NIFTY for 24th feb 2010 to 2 march 2010

5. RESULT AND DISCUSSION:

Study of the MSE profile of daily variations of different indices show that the BSE Metal Index exhibit lower MSE pattern as compared to other sectoral indices. This is due to the cyclical nature of the constituents of the index. Thus a pattern emerging out of the cyclical nature gives rise to a lower MSE pattern at all scales. Similarly MSE study is performed for pre budget hours, just after budget, and later hours of market timing. The MSE profile shows that the entropy of market was the maximum at pre budget hours as if the market was behaving like a closed system with low degree of orderliness. As the provisions of the budget were made public market started interacting with the information showing lower MSE profile indicating higher orderliness (less complexity). After the market responded sufficiently to the information received again it shows a higher MSE profile but still lower than the pre- budget hours.

However the same study performed on the NIFTY tick value for the period from 24th feb. 2010 to 2nd march 2010 in which there are two days of pre budget, budget day and one post budget day indicates that the difference among their sample entropy profile disappear and their profiles converge showing the identical behavior. This indicates that the market responds to received information with a higher degree of order and adjust itself interacting with the information. As the information has been received the market behaves like an isolated system with higher entropy.he findings of this study demonstrates that multiscale entropy measurements could be an effective alternative nonlinear approach for analyzing the Efficient behavior of the stock market at different time scales. This study shows that time scale is important as for as the efficiency of market is concerned. When the behavior is studied using interday time series, market shows Efficient behavior. However when the time scale is even shorter the time lag should be taken into account since market requires some time to absorb the information (showing inefficiency in its behavior.

References:

[1]. Holger Kantz, Thomas Schreiber , Nonlinear Time Series Analysis , Cambridge Press-1999.
[2]. Grassberger and Proccacia , Pliysica D 9. (1983), 189-208.
[3]. Takens F., Lect. Notes. Math. 898, (1981) 366-381.
[4]. G. Otte, Acta. Neurol. Belg, 107, (2007) 61-62.
[5]. Rapp P. E. Biologist,40(1993), 89-94.
[6]. Theiler j., Physics Letters A, 196.(1995)335-341.
[7]. www. Indiainfoline.com
[8]. Eugen Fama, Financial Analyst Journal 21(5), (1965) 55-59
[9]. Beechey M., Gruen D., Vickrey J.,citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.pdf

[10]. Fama, Eugene. Journal of Finance 25 (2), (1970) 383-417.doi:10.2307/2325486.
[11].Khan Arshad, Atlantic Economic Journal Volume 14, Number 3, (1986) 51-55
[12]. Firth M.A., Farnborough, Saxon House (1976).
[13]., Firth M. A. Mcmillon, London (1977);Jr. of Finance, XXXIII (1978) 670-72
[14]. Saad, Emad W.; Prokhorov, Danil V ; and Wunsch,II, Donald C., IEEE Transactions on Neural Networks 9 (6): (November, 1998)1456–1470.
[15]. Granger, Clive W. J. & Morgenstern, Oskar (May 2007). Kyklos 16 (1)(2007)1–27
[16]. Kleinberg, Jon; Tardos, Eva (2005). Algorithm Design. Addison Wesley. ISBN 0321295358.
[17]. chuvakhin ncbase.com/papers/EMH-BF.pdf
[18]. Fama, Eugene, Lawrence Fisher, Michael Jensen, and Richard Roll, International Economic Review, vol. 10, (1969), 1-21.
[19] Johnson, W. Bruce, Robert P. Magee, Nandu J. Nagarajan, and Harry A. Newman Journal of Accounting and Economics, (1985), 151-174.
[20]. Rozeff, M. and W. Kinney,Journal of Financial Economics 3, (1976) 379-402.
[21]. Costa M., Goldberger A.L., Peng C.-K., Phys Rev Lett, 89 (2002):062102.
[22].Pincus SM. ,Proc Natl. Acad Sci USA;88(1991) 2297-2301.