Articles

NON-LINEAR TIME SERIES INVARIANTS TO STUDY PRICE MANIPULATION IN STOCK MARKET

Y.V.Reddy[a] **A.Sebastin**[b]

ABSTRACT

Several studies have found the successive prices at which trades of shares are executed in a stock exchange to form time series with non-linear characteristics. Invariants of a non-linear time series of share prices contain valuable information about the patterns in the share prices, which factor may be useful in the analysis of share prices for various purposes. Of the commonly used invariants, entropy is found to be a useful tool in studying price manipulation in stock market. Among the various versions of entropy, sample entropy computed for the price data of a share, for various trading days in the period during which the share was reported to be subject to price manipulation, proves to be potential evidence for manipulation of the share's price. This paper studies the suitability of sample entropy of non-linear time series of share prices and also the selection of appropriate values for the parameters used for computation of sample entropy of a short time series of share prices, for investigating into price manipulation in stock market.

Keywords: Non-linear invariants, stock price manipulation, sample entropy, template size, tolerance limit, mutual information, relative error

INTRODUCTION

Stock price manipulation has been studied by various authors over periods under different situations like insider trading, asymmetric information, corners, short squeezes, imperfect competition, financial signaling, equity offerings, takeover bids, 'talking down' the firm, bluffing and front running. Though literature abound the study of market microstructure in general and market manipulation in particular, there is a lot of scope for in-depth study of manipulation of prices in the stock market, using the concepts of stochastic calculus, game theory and information theory.

Mathematical modeling and statistical analysis of stock price movements has become a field of its own, starting from the Louis Bachelier's Brownian motion model of 1900 for pricing warrants traded on the Paris bourse, to the recent dynamical systems theory and neural networks. However, pollutants such as fraud and market manipulation seem nearly impossible to be modeled, yet are real and significantly alter price movements without any economic reasons. Simply incorporating fraud into a random effects component of a model fails because the extent of fraud is rarely chronic, but is much more interrupted with the outcomes of a complicated game between regulatory efforts and corruptive creativity. So, a model-independent (providing qualitative inferences across diverse model configurations) analytic tool to study price manipulation will be of effective utility. Since a stock's prices are observed to form a time series with non-linear characteristics also, in this article, the suitability of a non-linear time series invariant viz. entropy, a concept originated in thermodynamics and developed in information theory, as a tool to study stock price manipulation, is considered.

II Recent studies on entropic analysis in stock market

A brief history of the entropic concept is given in annexure. The theory of entropy has been applied throughout financial economics and specifically, in the stock market domain. A few

[a]Departnment of Commerce, Goa University
[b]National Stock Exchange of India Limited

important studies are mentioned below, in order to illustrate the pervasiveness of entropy.

David Nawrocki (1984) has described an application of entropy theory to financial market disequilibrium. John Conover (1994) has applied entropy theory in devising a methodology for programmed trading of equities. David T. Marantette (1998) has found the analysis of entropy tops and bottoms as an addition to the buy and sell points of cyclic analysis.

Marco Frittelli (2000) has studied the characterization of the density of the minimal entropy martingale measure, which suggests that the equivalence between the maximization of expected exponential utility and the minimization of the relative entropy.

Esfandiar Maasoumi and Je Racine (2002) have examined the predictability of stock market returns by employing a new metric entropy measure which is capable of detecting non-linear dependence within the returns series.

Steve Pincus and Rudolf E.Kalman (2004) have demonstrated the utility of approximate entropy to assess subtle and potentially exploitable changes in serial structure of a financial variable.

Jing Chen (2005) has shown that most empirical evidences about market behaviour may be explained by a new information theory generalised from Shannon's entropy theory of information.

Dionisio, *et al.,* (2005) have presented the advantages of entropy over variance as a measure of uncertainty and have also shown the good performance of entropy in comparison with systematic risk and specific risk, with respect to diversification effect in portfolio management.

Wolfgang Kispert has used the prices of stock options to find a probability measure for the underlying stock and has derived that the probability vector with maximal entropy seems to be theoretically more justified than others.

Thus, the existing literature on the applications of entropy in the financial markets covers areas as wide ranging as market equilibrium and predicting market returns. Apparently, there has been no attempt to utilise the concept of entropy to the study of price manipulation in stock market. Further, there is a rich literature available on stock price manipulation, starting from the insider trading model of Kyle (1985), through the classification of market manipulation strategies proposed by Cherian and Jarrow (1995) to the recent study of front running by Fang Cai (2003). However, most of these publications portray stock price manipulation in market maker model of the American stock markets and analyse the conditions for equilibrium of the market, with price manipulation as a characteristic. It seems that no empirical tests have been suggested to study such manipulation. Since entropy measures the degree of irregularity in a data set, which is directly affected by manipulative efforts, it is proposed to verify the potential of entropy as an empirical tool in the identification of stock price manipulation.

III Entropy and stock price manipulation

Evidence of non-linear dependence in financial time series has been reported, very often in finance literature and non-linear structure has been observed in the stock returns of Indian stock exchanges also. Poshakwale (2002) examined the random walk hypothesis in the emerging Indian stock market using daily data on individual stocks traded in Bombay Stock Exchange (BSE) and an equally weighted portfolio and found statistical evidence to reject the random walk hypothesis and also to suggest that daily returns earned by individual stocks show significant non-linear dependence and persistent volatility effects. Sathe (2005) has found using serial correlation, mutual information, correlation dimension and method of close returns, that daily returns of various stock indices of National Stock Exchange of India Limited (NSE) display small amounts of linear and significant amounts of non-linear dependence. Hence, non-linear dynamics may be quite useful in studying the microstructure of Indian stock markets and other markets, which display non-linear dependence in share prices, in order to identify price manipulation.

A brief explanation of the invariants or parameters which characterise the non-linear dynamics of time series is given in annexure. Of the commonly used non-linear invariants, (i) entropy measures irregularity or complexity of the data (ii) Lyapunov exponent measures sensitive dependence on initial conditions and (iii) correlation dimension measures spatial correlation in the phase space. The last two are used to study the presence of chaotic structure in very large data sets and perform poorly with small samples. Entropy is used to study the presence of repetitive patterns or shifts in data structure and some versions of entropy may be measured for very short and noisy time series also. Hence, entropy may be suited to study price manipulation in stock market, using the time series of successive prices of a scrip on any day.

In the electronic stock trading system, as market participants place orders for buying or selling the shares of a scrip at different prices and for various quantities, trades are effected by matching these orders according to price–time priority. A scrip's price is expected to change from time to time, based on the fundamental factors of the scrip, its past history and the demand for the scrip. The prices at which, the times at which and the quantities for which, orders are placed by a participant, are expected to be in accordance with the prevalent market conditions and towards investment/speculative purpose. So long as a participant places orders in the normal course of business, the entropy values of these time series will be in some ranges. However, when a participant repeatedly places orders for buying/selling a scrip according to some pattern in the price or time or quantity, with a motive of manipulating the price of the scrip, more regularity or persistence will be induced in the time series and consequently, entropy is likely to decrease. Large decrease in the entropy value from usual ranges may lead to potential evidence of price manipulation by a participant.

Of course, regularity of such nature in the order related data may occur, by chance, rarely. However, repeated drops in the entropy values of the order related time series of a scrip in a span of a few trading days point to likely manipulation in the price of the scrip. For any scrip, the only publicly available information are trade price, trade time and trade quantity, for all trades on any day and without the identity of the participants who are parties to the trades. Under the circumstances, versions of entropy like approximate entropy (ApEn) and sample entropy (SampEn) are available for computing the entropy of short and noisy time series of trade related information.

In practice, for investigating into potential manipulation cases, stock exchanges maintain surveillance systems which monitor trading activities of the participants in the various securities and throw alerts based on intra-day price movements and abnormal trade quantities. For this purpose, different measures of variation like high – low variation and consecutive trade price variation are generally used. It may be noted that although a trading pattern which crosses the limits stipulated for such variations may be identified on-line, there may be many manipulative trading patterns well within the stipulated limits, which will go undetected. Further, these measures of variation identify only linear relationships. Hence, advanced techniques are required to detect complex manipulative strategies involving non-linear relationships also. *The merits of entropy over the measures of variation are given in annexure.*

IV Computational aspects

The concepts of ApEn and SampEn have been briefly explained in the annexure. Considering the advantages of SampEn over ApEn, it is proposed that SampEn be computed for the time series of trade price related information of a scrip chosen for the study. Since daily settlement is followed in the stock market, day-wise time series, each consisting of the prices of all trades on a day, are to be formed and SampEn is to be computed for every trading day with appropriate values for the parameters m and r. Although no guidelines exist for optimising their values, generally values between 0.1 and 0.25 for r and values of 1 or 2 for m are used for data sets with length ranging from 100 to 5000. Informed selection of values for the parameters m and r is to be preferred to unguided use of the parameters based on unquestioned acceptance of the idea that differences in entropy estimates are always the result of differences in irregularity of the data. Optimal selection of the parameters is an unexplored area of paramount importance.

Template size 'm'

For picking the value of m, some authors have suggested the use of auto-regression (AR) models. The motivation for this approach is that if a data set is an AR process of order x, then $m \geq x$. To estimate m for a data set, AR model of various orders is fit to the data and the order corresponding to the minimum value of Schwarz Bayesian Criterion (SBC) or Akaike Information Criterion (AIC) is considered to be the order x of the process. In this article, it is proposed that the information theoretic concept of mutual information be used to estimate appropriate value for the parameter m. *The definition of mutual information of a time series is given in annexure.*

It may be noted that the computation of SampEn of a time series involves construction of templates of size m from the scalar observations forming the time series and counting the number of such templates in the neighbourhood (i.e. within a distance of r) of each such template. A good choice for m is such that contiguous templates of size m constructed from the time series are not within the neighbourhood of one another. Such a choice is provided by the value of m corresponding to which the mutual information of the time series with delay m viz. I(m) is small and consequently, the contiguous templates are independent to a large extent. As m is increased, I(m) decreases and may rise again and hence, the first minimum of I(m) may be considered to choose the value of m. It may also be noted that Fraser and Swinney (1986) have suggested that in the construction of multidimensional phase portrait from a scalar time series, the time delay T that produces the first local minimum of the mutual information of the time series may be used. Since mutual information measures the general dependence between two variables or between two time series of the same variable with time delay, it provides a better criterion for the choice of m than the autocorrelation function which measures only linear dependence.

Tolerance limit 'r'

If A is the number of templates of size m which match within a tolerance level of r and B is the number of templates out of A which matches for template size m+1 also, then B / A estimates the conditional probability P of match of template size m+1, given that there is a match of template size m and SampEn is – log(P). We know that for any differentiable function f of a random variable X, the standard deviation $\sigma_{f(X)}$ is approximated by | f ′(X) | σ_X. Hence we have σ_{SampEn} = σ_P / P and thus the standard error of SampEn is the relative error of P.

r is to be chosen in such a way that it is neither so stringent that the number of matches is too near 0 (low confidence), nor so relaxed that P is too near 1 (low discrimination). r may be selected so as to minimise the quantity max $[\frac{\sigma_P}{P}, \frac{\sigma_P}{-P\log(P)}]$ which is the maximum of the relative errors of the estimate of P and SampEn. This efficiency metric favours estimates with low variance and thus reflects the efficiency of the entropy estimate. This criterion also represents a trade-off between accuracy and discrimination capability, as it simultaneously penalises P near 0 and near 1. Given a value of m based on the first minimum of the mutual information of the stock price time series, an optimal value of r can be selected to minimise the efficiency metric.

V Case study

The scrip of the company M/s Surya Rooshni Ltd., which has been reported to be subject to price manipulation on various days during the period April 2000 - October 2000 (www.sebi.com) is chosen for the study. Day-wise time series are formed by taking the prices of all intra-day trades (tick by tick trade prices) executed in the scrip on the National Stock Exchange of India Limited, on each of the trading days during this period. The differences in the prices of successive intra-day trades are computed in each time series. By taking such first differences, stationary character of the time series may be assumed so that meaningful analysis may be made. Further, a manipulator is always interested in price differences in order to gain as much as possible and hence places successive orders, with artificial prices carrying a manipulative intent.

For deciding the template size m in the computation of SampEn for each time series, the mutual information of each time series for the various trading days during the period April 2000 - October 2000 was calculated, with time lag ranging from 1 to 10 and the values are presented in Tables 1 to 3. It may be observed that the first minimum of mutual information occurred for the time lag between 2 and 4 in almost all the cases and hence, SampEn may be calculated for template size m = 2, 3 or 4. For time series with a few thousands of data points, the optimum value of r has been observed to lie between 0.15 and 0.20 if the template size m is assigned a value 2, 3 or 4.

TABLE 1

Mutual information for the trade price time series – April and May 2000

Delay	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16
0	2	5	3	4	5	3	4	4	2	1	6	4	4	4	1	2
1	0	1.125	1	1	1.875	1	1	1.5	0	1	1.28125	1.5	0.75	1	1	0
2	0	2.8125	1	1.125	1.1875	1	0.5	0.5	2	1	1.53125	1.25	1.125	0.5	1	2
3	2	1.375	1	1.25	1.375	1	1.125	2	0	1	1.25	0.5	0.5	1	1	0
4	0	1.375	1.5	1.25	1.1875	0.5	1.375	1.875	0	1	1.5625	0.75	1	1.5	1	0
5	0	1.125	1	1.5	1.125	0.5	0.75	1	2	1	1.1875	0.75	0.875	1.25	1	0
6	0	1.1875	0.5	1.875	1.0625	1	1.125	0	0	1	1.46875	1.25	1	0.75	1	2
7	2	1.8125	1	0.875	1.0625	1	0.875	1.375	2	1	1	1	0.5	0.75	1	0
8	0	1.625	1	0.875	1.125	1	1.5	1.125	0	1	1.1875	1.5	0.5	2	1	0
9	0	1.25	1	1.25	1.4375	1	1	0.75	0	1	1.34375	1	0.5	1.125	1	2
10	0	1.5625	1	1.5	1.4375	0.5	1	1.25	0	1	1.0625	1	1.25	0.75	1	0

TABLE 2
Mutual information for the trade price time series – June, July and August 2000

Delay	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14
0	3	2	2	2	3	0	0	3	1	2	1	2	2	3
1	1	0	2	2	1	0	0	1	1	0	1	0	2	1.5
2	1	2	0	0	1	0	0	1	1	2	1	2	2	1
3	1	0	0	0	1	0	0	1.5	1	0	1	2	2	1
4	1	2	0	2	0.5	0	0	1	1	2	1	0	0	1
5	1	2	0	2	1	0	0	1	1	0	1	0	0	0.5
6	1	0	2	2	1.5	0	0	1.5	1	0	1	0	0	1
7	1	0	0	0	1	0	0	1	1	0	1	2	0	0.5
8	0.5	0	0	0	1	0	0	1	1	2	1	0	0	1.5
9	1	0	0	2	1	0	0	0.5	1	0	1	2	2	0.5
10	1	0	0	0	1	0	0	1	1	0	1	0	0	0.5

TABLE 3
Mutual Information for the Trade Price Time Series – September and October 2000

Delay	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11
0	4	6	4	3	2	3	3	2	4	2	3
1	0.75	1.40625	1	0.5	0	1	1	2	1	0	1
2	0.75	1.375	1	0.5	0	1	1	2	1.375	0	0.5
3	0.75	1.09375	0.5	0.5	2	1	1	0	1	0	1
4	2	1.3125	1.5	1	2	1	2	0	1.125	2	1
5	1.625	1.40625	1	1	2	0.5	1	2	1	0	1
6	1.25	1.03125	1	1	2	1	0.5	2	1.25	0	1
7	1	1.375	1	3	2	1	1	2	0.5	0	0.5
8	1	1.40625	0.5	1	0	1	1	0	1.375	2	0.5
9	1.375	1.125	1	1	0	1	1	0	1	0	1
10	1	1.125	0.75	1	0	1	1	0	1.75	0	0.5

Accordingly, SampEn of the time series consisting of the differences in successive trade prices of the scrip of M/s Surya Rooshni Ltd. was computed for m = 2, 3, 4 and r = 0.15, 0.16, 0.17, 0.18, 0.19, 0.20 These SampEn values for the various trading days during the period April 2000 - October 2000 are given in Tables 4 to 10 respectively. It may be observed that SampEn is very low on days 2, 5, 8 and 10 in April 2000; nil in the month of May; on days 8, 12, 13, 14, 17, 18 and 21 in June 2000; on days 2 and 3 in July 2000; on days 5 and 14 in August 2000; on days 4, 7, 12 and 13 in September 2000 and on days 2, 7, 8, 11, 12 and 13 in October 2000, for all values of m and r. Specifically, SampEn for all these days is utmost 0.85 for m = 4 and r = 0.20*SD. The above mentioned days in April 2000 - October 2000, in respect of which SampEn is very low, are days of potential manipulation in the price of the scrip of M/s Surya Rooshni Ltd. Further, it appears that manipulation has been rampant in the months of June and October 2000 since sample entropy has remained at very low level continually for many days.

TABLE 4
SampEn values for April 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16	Day17
2,0.15	1.253	0.441	2.996	4.025	0.142	0.916	1.609	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.946
2,0.16	1.253	0.441	2.996	4.025	0.142	0.916	1.609	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.946
2,0.17	1.253	0.441	2.996	4.025	0.142	0.916	1.609	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.946
2,0.18	1.312	0.441	2.996	4.025	0.142	0.916	1.609	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.355
2,0.19	1.312	0.441	2.996	4.025	0.142	1.204	1.014	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.355
2,0.20	1.312	0.441	2.996	4.025	0.142	1.204	1.014	0.51	1.386	0.764	1.213	4.7	4.025	4.025	2.485	0.288	1.355
3,0.15	1.792	0.362	2.996	4.025	0	0.693	1.386	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	6.136
3,0.16	1.792	0.362	2.996	4.025	0	0.693	1.386	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	6.136
3,0.17	1.792	0.362	2.996	4.025	0	0.693	1.386	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	6.136
3,0.18	1.946	0.362	2.996	4.025	0	0.693	1.386	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	1.386
3,0.19	1.946	0.362	2.996	4.025	0	1.099	1.308	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	1.386
3,0.20	1.946	0.362	2.996	4.025	0	1.099	1.308	0.264	6.477	0.718	0.606	4.7	4.025	4.025	2.485	0	1.386
4,0.15	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	6.136
4,0.16	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	6.136
4,0.17	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	6.136
4,0.18	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	0.693
4,0.19	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	0.693
4,0.20	6.04	0.321	2.996	4.025	0	5.347	7.451	0.1	6.477	0.357	0.693	4.7	4.025	4.025	2.485	1.099	0.693

TABLE 5
SampEn values for May 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16	Day17	Day18	Day19	Day20
2,0.15	5.835	1.792	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.204	3.401	4.025	0.174	6.962	1.23	5.835	2.485	4.883	4.277	1.609
2,0.16	5.835	1.54	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.609	3.401	4.025	0.174	6.962	1.23	5.835	2.485	4.883	4.277	1.609
2,0.17	0.717	1.54	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.609	3.401	4.025	0.174	6.962	1.23	5.835	2.485	4.883	4.277	1.609
2,0.18	0.717	1.54	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.609	3.401	4.025	0.174	6.962	1.23	5.835	2.485	4.883	4.277	1.609
2,0.19	0.717	1.54	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.609	1.099	4.025	0.174	6.962	1.23	5.835	2.485	4.883	4.277	1.609
2,0.20	0.717	1.54	0.56	5.204	1.228	5.204	4.883	4.277	3.738	1.609	1.099	4.025	0.174	6.962	1.23	5.835	2.485	0.693	4.277	1.609
3,0.15	5.835	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	4.883	4.277	6.136
3,0.16	5.835	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	4.883	4.277	6.136
3,0.17	0.647	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	4.883	4.277	6.136
3,0.18	0.647	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	4.883	4.277	6.136
3,0.19	0.647	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	4.883	4.277	6.136
3,0.20	0.647	5.204	0.693	5.204	1.319	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.186	6.962	1.335	5.835	2.485	1.099	4.277	6.136
4,0.15	5.835	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136
4,0.16	5.835	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136
4,0.17	0.606	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136
4,0.18	0.606	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136
4,0.19	0.606	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136
4,0.20	0.606	5.204	1.099	5.204	1.216	5.204	4.883	4.277	3.738	5.05	3.401	4.025	0.205	6.962	7.082	5.835	2.485	4.883	4.277	6.136

Table 6
SampEn values for June 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12
2,0.15	1.099	0.693	6.835	6.136	3.401	1.946	1.204	0.511	5.05	4.277	1.792	0.693
2,0.16	1.099	0.693	6.835	1.386	3.401	1.946	1.204	0.511	5.05	4.277	1.792	0.693
2,0.17	1.099	0.693	6.835	1.386	3.401	1.946	0.657	0.511	5.05	4.277	1.792	0.693
2,0.18	1.099	0.693	6.835	1.386	3.401	1.946	0.657	0.511	5.05	4.277	1.792	0.693
2,0.19	1.099	0.693	6.835	1.386	3.401	1.946	0.657	0.511	5.05	4.277	1.792	0.693
2,0.20	1.099	0.693	6.835	1.386	3.401	1.946	0.657	0.511	5.05	4.277	1.792	0.693
3,0.15	5.05	5.204	6.835	6.136	3.401	6.04	1.099	0.693	5.05	4.277	1.792	0.693
3,0.16	5.05	5.204	6.835	1.386	3.401	6.04	1.099	0.693	5.05	4.277	1.792	0.693
3,0.17	5.05	5.204	6.835	1.386	3.401	6.04	1.03	0.693	5.05	4.277	1.792	0.693
3,0.18	5.05	5.204	6.835	1.386	3.401	6.04	1.03	0.693	5.05	4.277	1.792	0.693
3,0.19	5.05	5.204	6.835	1.386	3.401	6.04	1.03	0.693	5.05	4.277	1.792	0.693
3,0.20	5.05	5.204	6.835	1.386	3.401	6.04	1.03	0.693	5.05	4.277	1.792	0.693
4,0.15	5.05	5.204	6.835	6.136	3.401	6.04	4.883	1.099	5.05	4.277	1.792	0.693
4,0.16	5.05	5.204	6.835	6.136	3.401	6.04	4.883	1.099	5.05	4.277	1.792	0.693
4,0.17	5.05	5.204	6.835	6.136	3.401	6.04	1.609	1.099	5.05	4.277	1.792	0.693
4,0.18	5.05	5.204	6.835	6.136	3.401	6.04	1.609	1.099	5.05	4.277	1.792	0.693
4,0.19	5.05	5.204	6.835	6.136	3.401	6.04	1.609	1.099	5.05	4.277	1.792	0.693
4,0.20	5.05	5.204	6.835	6.136	3.401	6.04	1.609	1.099	5.05	4.277	1.792	0.693

m,r	Day13	Day14	Day15	Day16	Day17	Day18	Day19	Day20	Day21
2,0.15	0.693	0.665	5.204	1.099	0.693	5.204	4.883	0.693	1.792
2,0.16	0.693	0.665	5.204	1.099	0.693	5.204	4.883	0.693	1.792
2,0.17	0.693	0.665	5.204	1.099	0.693	5.204	4.883	0.693	1.792
2,0.18	0.693	0.665	5.204	1.099	0.693	1.792	4.883	0.693	1.792
2,0.19	0.693	0.665	5.204	1.099	0.693	1.792	4.883	0.693	1.792
2,0.20	0.693	0.665	5.204	1.099	0.693	1.792	4.883	0.693	1.792
3,0.15	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
3,0.16	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
3,0.17	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
3,0.18	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
3,0.19	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
3,0.20	0.693	0.492	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.15	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.16	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.17	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.18	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.19	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792
4,0.20	0.693	0.606	5.204	5.05	0.693	5.204	4.883	0.693	1.792

TABLE 7
SampEn values for July 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16	Day17	Day18	Day19	Day20
2,0.15	3.738	0.811	0.693	1.792	1.792	2.996	4.7	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
2,0.16	3.738	0.811	0.693	1.792	1.792	2.996	4.7	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
2,0.17	3.738	0.811	0.693	1.792	1.792	2.996	1.253	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
2,0.18	3.738	0.811	0.693	1.792	1.792	2.996	1.253	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
2,0.19	3.738	0.811	0.693	1.792	1.792	2.996	1.253	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
2,0.20	3.738	0.811	0.693	1.792	1.792	0	1.253	2.996	1.792	2.565	5.481	5.204	3.401	4.277	0.405	4.5	1.386	3.401	1.792	2.485
3,0.15	3.738	0.693	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
3,0.16	3.738	0.693	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
3,0.17	3.738	0.693	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
3,0.18	3.738	0.693	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
3,0.19	3.738	0.693	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
3,0.20	3.738	0.693	0.693	1.792	1.792	0.405	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	1.099	4.5	5.204	3.401	1.792	2.485
4,0.15	3.738	0.405	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485
4,0.16	3.738	0.405	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485
4,0.17	3.738	0.405	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485
4,0.18	3.738	0.405	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485
4,0.19	3.738	0.405	0.693	1.792	1.792	2.996	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485
4,0.20	3.738	0.405	0.693	1.792	1.792	0	4.7	2.996	5.724	6.628	5.481	5.204	3.401	4.277	5.05	4.5	5.204	3.401	1.792	2.485

TABLE 8
SampEn values for August 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16	Day17
2,0.15	1.386	3.738	4.7	5.204	1.041	1.792	4.025	1.253	0.693	0.606	1.099	4.277	1.099	0.847	0.693	1.946	2.639
2,0.16	1.386	3.738	4.7	5.204	1.041	1.792	4.025	1.253	0.693	0.606	1.099	4.277	1.099	0.847	0.693	1.946	2.639
2,0.17	1.386	3.738	4.7	5.204	1.041	1.792	4.025	1.253	0.693	0.606	1.099	4.277	1.099	0.847	0.693	1.946	2.639
2,0.18	1.386	3.738	4.7	5.204	1.041	1.792	4.025	0.342	0.693	0.606	1.099	4.277	1.099	0.847	0.693	0.838	2.639
2,0.19	1.386	3.738	1.099	5.204	1.041	1.792	4.025	0.342	0.693	0.606	1.099	4.277	1.099	0.847	0.693	0.838	2.639
2,0.20	1.386	3.738	1.099	5.204	1.041	1.792	4.025	0.342	0.693	0.606	1.099	4.277	1.099	0.847	0.693	0.838	2.639
3,0.15	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.693	1.099	0.693	5.05	4.277	4.277	1.099	1.099	6.227	6.7
3,0.16	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.693	1.099	0.693	5.05	4.277	4.277	1.099	1.099	6.227	6.7
3,0.17	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.693	1.099	0.693	5.05	4.277	4.277	1.099	1.099	6.227	6.7
3,0.18	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.351	1.099	0.693	5.05	4.277	4.277	1.099	1.099	0.981	6.7
3,0.19	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.351	1.099	0.693	5.05	4.277	4.277	1.099	1.099	0.981	6.7
3,0.20	5.835	3.738	4.7	5.204	0.693	1.792	4.025	0.351	1.099	0.693	5.05	4.277	4.277	1.099	1.099	0.981	6.7
4,0.15	5.835	3.738	4.7	5.204	0.405	1.792	4.025	5.94	4.277	6.04	5.05	4.277	4.277	0	5.204	6.227	6.7
4,0.16	5.835	3.738	4.7	5.204	0.405	1.792	4.025	5.94	4.277	6.04	5.05	4.277	4.277	0	5.204	6.227	6.7
4,0.17	5.835	3.738	4.7	5.204	0.405	1.792	4.025	5.94	4.277	6.04	5.05	4.277	4.277	0	5.204	6.227	6.7
4,0.18	5.835	3.738	4.7	5.204	0.405	1.792	4.025	0.379	4.277	6.04	5.05	4.277	4.277	0	5.204	1.099	6.7
4,0.19	5.835	3.738	4.7	5.204	0.405	1.792	4.025	0.379	4.277	6.04	5.05	4.277	4.277	0	5.204	1.099	6.7
4,0.20	5.835	3.738	4.7	5.204	0.405	1.792	4.025	0.379	4.277	6.04	5.05	4.277	4.277	0	5.204	1.099	6.7

TABLE 9
SampEn values for September 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15	Day16	Day17	Day18
2,0.15	1.386	1.411	1.106	0.693	1.558	0.847	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
2,0.16	1.386	1.411	1.106	0.693	1.558	0.847	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
2,0.17	1.386	1.411	1.106	0.693	1.558	0.847	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
2,0.18	1.386	1.411	1.106	0.693	1.558	0.693	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
2,0.19	1.386	1.411	1.106	0.693	1.558	0.693	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
2,0.20	1.386	1.411	1.106	0.693	1.558	0.693	0.818	2.996	0.869	1.099	1.227	0.693	0.693	2.996	2.485	2.398	1.792	4.5
3,0.15	5.05	1.609	1.093	0.796	1.386	6.04	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
3,0.16	5.05	1.609	1.093	0.796	1.386	6.04	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
3,0.17	5.05	1.609	1.093	0.796	1.386	6.04	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
3,0.18	5.05	1.609	1.093	0.796	1.386	0.693	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
3,0.19	5.05	1.609	1.093	0.796	1.386	0.693	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
3,0.20	5.05	1.609	1.093	0.796	1.386	0.693	0.762	2.996	1.466	6.04	1.447	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.15	5.05	7.635	1.488	0.651	6.477	6.04	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.16	5.05	7.635	1.488	0.651	6.477	6.04	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.17	5.05	7.635	1.488	0.651	6.477	6.04	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.18	5.05	7.635	1.488	0.651	6.477	0.405	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.19	5.05	7.635	1.488	0.651	6.477	0.405	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5
4,0.20	5.05	7.635	1.488	0.651	6.477	0.405	0.847	2.996	0.405	6.04	0.693	0.693	0.693	2.996	2.485	6.314	1.792	4.5

Table 10
SampEn values for October 2000 (M/s Surya Rooshni Ltd.)

m,r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14
2,0.15	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.253	2.996	0.693	0.693	0.693	3.401
2,0.16	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.253	2.996	0.693	0.693	0.693	3.401
2,0.17	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.012	2.996	0.693	0.693	0.693	3.401
2,0.18	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.012	2.996	0.693	0.693	0.693	3.401
2,0.19	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.012	2.996	0.693	0.693	0.693	3.401
2,0.20	2.485	0.693	4.025	1.792	5.05	3.401	0.894	0.693	1.012	2.996	0.693	0.693	0.693	3.401
3,0.15	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	5.204	2.996	0.693	0.693	0.693	3.401
3,0.16	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	5.204	2.996	0.693	0.693	0.693	3.401
3,0.17	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	1.386	2.996	0.693	0.693	0.693	3.401
3,0.18	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	1.386	2.996	0.693	0.693	0.693	3.401
3,0.19	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	1.386	2.996	0.693	0.693	0.693	3.401
3,0.20	2.485	0.693	4.025	1.792	5.05	3.401	0.405	0.693	1.386	2.996	0.693	0.693	0.693	3.401
4,0.15	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401
4,0.16	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401
4,0.17	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401
4,0.18	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401
4,0.19	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401
4,0.20	2.485	0.693	4.025	1.792	5.05	3.401	0.182	0.693	5.204	2.996	0.693	0.693	0.693	3.401

CONCLUSION

Entropic analysis is a novel area in the Indian stock market and there is a lot of scope for the application of entropic analysis in the Indian stock market. This paper applies entropic analysis to study price manipulation in the Indian stock market and sample entropy is found to be suited for this study. SampEn values for the trade price data related to the scrip of M/s Surya Rooshni Ltd., for various trading days in the period during which it was reported to be subject to price manipulation, are found to support such reporting. The values of the parameters required in the computation of SampEn - the template size m and the tolerance limit r – are not to be assigned arbitrarily but m may be chosen on the basis of minimum mutual information of the trade price time series so as to enhance the independence of the templates to a large extent and r may be chosen on the basis of minimum relative error of sample entropy so as to reduce the variance of the entropy estimate to a large extent. This empirical analysis is to be done for many other scrips in order that entropic analysis is considered as an effective tool to study stock market manipulation. Further, entropic analysis of order related data (if available) will ensure more efficiency in the study of price manipulation attempts in the stock market.

REFERENCES

Cai, Fang, "Was there front running during the LTCM crisis?" *Working Paper no. 758, , International Finance Discussion Paper*s, www.federalreserve.gov/pubs/ifdp, February 2003.

Chen, Jing, "Information Theory and Market Behaviour", *ICFAI Journal of Behavioral Finance*, Vol. 2, No. 4, (2005): 25 – 45.

Cherian, Joseph A. and Robert A. Jarrow, "Market Manipulation", In *Handbooks in Operations Research and Management Science* Vol. 9 Elsevier Science B.V. , 1995.

Conover, John, "NtropiX", www.johncon.com, (1994).

Dionisio, Andreia, Rui Menezes and Diana A. Mendes, "Entropy and Uncertainty Analysis in Financial Markets" XIV:0709.0668, (2005).

Ebrahimi, Maasoumi and Soofi, "Ordering Univariate Distributions by Entropy and Variance", *Journal of Econometrics*, Vol. 90 No.2 (1999): 317 – 336.

Esfandiar, Maasoumi and Jeff Racine, "Entropy and Predictability of Stock Market Returns", *Journal of Econometrics*, Vol. 107 (1-2) (March 2002): 291-312.

Fraser, Andrew M. and Harry L. Swinney, "Independent coordinates for strange attractors from mutual information", *Physical Review* , Vol. 33, No. 2 (February 1986): 1134 – 1140.

Frittelli, Marco, "The minimal entropy martingale measure and the valuation problem in incomplete markets", *Mathematical Finance*, Vol. 10 No. 1 (January 2000): 39 – 52.

Joshua, Richman S., and J. Randall Moorman, "Physiological Time-Series Analysis using Approximate Entropy and Sample Entropy", *American Journal of Physiology*, Vol. 278, Issue 6 (2000): 2039 – 2049.

Kyle, Albert S., "Continuous auctions and insider trading", *Econometrica*, , Vol. 53, No. 6 (November 1985) : 1315-1336.

Marantette, David T. "*The entropy danger*", www.goldstock.com, 1998.

McCauley, J., "Thermodynamic analogies in Economics and Finance: Instability of markets" *Physica* A, Vol. 329 (2003): 199 – 212.

Nawrocki, David, "Entropy, Bifurcation and Dynamic Market Disequilibrium", *The Financial Review*, Vol. 19 No. 2 (May 1984): 266 – 284.

Pincus, Steve and Rudolf E. Kalman, "Irregularity, Volatility, Risk and Financial Market Time Series", *Proceedings of the National Academy of Sciences of USA*, Vol. 101, No. 38 (2004): 13709 – 13714.

Poshakwale, Sunil S., "The random walk hypothesis in the emerging Indian stock market" *Journal of Business Finance & Accounting*, Vol. 29, No. 9 & 10 (November/ December 2002) : 1275-1299.

Sathe, Saket, "Testing efficiency of Indian capital markets with applications to risk management" *Project report, Indian Institute of Technology, Mumbai*, (2005).

Wolfgang, Kispert "Implicit Distributions in Stock Markets - Estimation with a Maximum Entropy Procedure" *LRZ-Account: b5101aa* .

www.sebi.com

ANNEXURE

Entropy – a brief history

The concept of entropy arose in physical sciences during the 19th century. Clausius, building on the previous intuition of Carnot, introduced for the first time in 1867, a mathematical quantity S, which he called entropy that describes heat exchanges occurring in thermal processes via the relation dS = dQ / T, where Q denotes the amount of heat and T is the absolute temperature at which the exchange takes place. Ludwig Boltzmann derived that the Clausius entropy S associated with a system in equilibrium is proportional to the logarithm of the number W of microstates which form the macrostate of this equilibrium i.e. S = k * ln (W). Since then, the concept of entropy was extended to study microscopically unpredictable processes in a number of fields like stochastic processes and random fields, information and coding, data analysis and statistical inference, partial differential equations and rational mechanics. This led to the employment of diverse mathematical tools in dealing with the concept of entropy.

The theoretical foundation of entropic methods used in modern finance was formalised by the mathematicians Jacob Bernoulli and Abraham de Moivre. The concept of entropic analysis of equity prices was first proposed by Louis Bachelier in 1900, which anticipated many of the mathematical discoveries made later by Norbert Wiener and A.A.Markov in early nineties. J.L.Kelly, Jr. established the relationship between the information rate in a binary symmetric channel and speculation under uncertainty and made the large mathematical infrastructure of information theory, which was further developed by Claude Shannon in the mid 1940's. The Shannon's definition of entropy of a random variable X with p(x) as the probability mass function, is

$$H(X) = H(p) = -\sum_{x} p(x) \log p(x) = E\,[\,\log \{1/\,p(x)\}]$$

where the base of the logarithm is 2 and 0 log 0 is taken as 0. Entropy is measured in bits and $0 \leq H(X) < \infty$. If logarithm is taken to the base e, then entropy is measured in nats. $H_a(X)$ denotes the entropy of X when logarithm is taken to some base a.

The introduction of metric entropy and the extension of classification theory of measure-preserving transformations, by Kolmogorov in the 1950's, led to significant advances. The uncertainty about the actual state of a system or process is measured by Shannon entropy. If the uncertainty is about predictions concerning the future of a process, it may be decreased by gaining information from the evolution of time itself. However, the dynamics of the process may go on producing new information at each successive stage so that forecasting is not made more reliable by knowledge of the past and this kind of uncertainty about the future is measured by Kolmogorov – Sinai (KS) entropy.

Invariants of a non-linear time series

(i) Entropy of a dynamical system is the amount of disorder in the system, as described in Thermodynamics and is also the amount of information needed to predict the next measurement with a certain precision, as described in Information Theory. Entropy does not measure the shape of the distribution of the realizations of a system, but provides information about how the system fluctuates with time – in frequency space or phase space. The concept of embedding the one-dimensional signal in a phase space, which is achieved by comparing the time series of realizations with itself but lagged by a specified time interval, is used to estimate the entropy of a system. In an irregular signal, the prediction of the next point using the knowledge of previous points is not easy and in a regular signal, such prediction is more reliable. The number of previous (lagged) points required to make the prediction is the embedding dimension. Using these embeddings, various versions of entropy such as Shannon entropy and KS entropy are estimated.

(ii) Lyapunov exponent (LE) is the inverse of time scale and quantifies the rate by which two typically nearby trajectories converge or diverge in time. In a predominantly periodic system, this divergence will be very slow, whereas in a chaotic system, this separation will be exponentially fast. LE is a properly averaged exponent of this increase and is characteristic of the system underlying the data and quantifies the strength of chaos. There are as many LEs for a dynamical system as there are phase space dimensions and generally, the maximal LE is measured to study a dynamical system. The maximal LE of a dissipative system may be negative, indicating the existence of a stable fixed point and two trajectories approaching a fixed point also approach each other exponentially fast. If the maximal LE is zero, the system is marginally stable and settles down to a limit cycle and two trajectories can separate or approach each other only slower than exponentially. A predominantly deterministic system, perturbed by small scale random noise behaves like a diffusion process and the corresponding maximal LE is large and positive. LE is invariant under all smooth transformations of shifting, rescaling or otherwise processing of data since it describes long term behaviour.

(iii) The portrayal of a data set as a geometrical object in phase space, as represented by the trajectories of the system, leads to the concept of dimension of the data set. Non-integer dimensions are assigned to geometrical objects which exhibit unusual kind of self-similarity (a part of an object when magnified, resembling the object itself) and which show structure on all length scales. The trajectories of a dissipative (contraction of volume elements) dynamical system do not fill the phase space and are confined to lower dimensional sub-sets which possess a fractal structure i.e. which are self-similar in a non-trivial way. Generalised dimensions are a class of quantities to characterize the fractal nature of a data set. Hausdorff dimension is the most natural concept to characterise a fractal set, from mathematical point of view. Information dimension is more attractive for physical systems, since it takes into account the relative visitation frequencies. Correlation dimension is useful for the characterization of measured data. Dimensions are invariant under smooth transformations and thus computable in time delay embedding spaces.

Entropy Vs measures of variation

Stock market analysts normally study shifts in mean levels and in variation (in various notations) to understand the state of the market. However, the persistence of certain patterns or shifts may provide critical information. It may be noted that formulae to directly quantify randomness have not been used in market analysis, perhaps due to the lack of a quantification technology, until recently. So, excluding sequential patterns or features which presented themselves, subtler changes in serial structure would remain largely undetected. Volatility is generally equated to the variability of a scrip's price, with large swings normally denoted as highly volatile or unpredictable. However, there are two fundamentally distinct means by which data deviate from central tendency – (i) they have high variation (as may be measured by standard deviation or variance) (ii) they appear highly irregular or unpredictable (as may be measured by entropy). These two non-redundant means have important consequences. The point is that the extent of variation in scrip prices is generally not feared, but what concerns is the unpredictability in time and quantity of the variation. If a market participant is assured of a typical model, with large amplitude for future changes in the price of a scrip, it will not be frightening because future prices and resultant strategies may be planned. Thus, a quantification technology to separate the concepts of classical variation and irregularity is of paramount importance.

Entropy is a measure of disparity of the probability mass function of a distribution from the uniform distribution, whereas, variance measures the average distance of the various realizations from the mean of a distribution. According to Ebrahimi, *et al.,* (1999), both these measures reflect concentration. However, unlike variance which measures concentration only around the mean, entropy measures diffuseness of the density, irrespective of the location of concentration. They also show, using a Legendre series expansion, that entropy depends on many parameters of a distribution and may be related to high order moments of a distribution. Therefore, entropy could offer a closer characterization of the probability mass function since it uses more information about the distribution than that used by variance and hence is more general than the traditional methods based on variance. McCauley (2003) propounds that entropy represents the disorder and uncertainty of a stock market or a particular stock, since entropy has the ability to capture the complexity of the systems, without requiring rigid assumptions which could bias the results. While volatility is an estimate of the variation of a scrip's price, entropy is concerned with the irregularity or randomness of the price fluctuations. Hence, entropy is more suited than any measure of variation, to study manipulation of stock market.

Approximate entropy and Sample entropy

Traditional methods for estimating the entropy of a system, represented by a time series, are not well-suited to analysis of short and noisy data sets. The calculation of Shannon entropy requires the probability density (mass) function of the random variable, which denotes the time series. However, Kolmogorov - Sinai (KS) entropy may be a useful parameter to characterise system dynamics. Though KS entropy measures the mean rate of creation of information, it cannot be estimated with reasonable precision for real world time series of finite length. Hence ApEn, a set of measures of serial irregularity and closely related to the entropy measure, has been introduced for typically short and noisy time series. ApEn grades a continuum that ranges from totally ordered to maximally irregular (completely random). ApEn attempts to distinguish data sets on the basis of regularity and not to construct an accurate model of the data.

ApEn measures the logarithmic likelihood that runs of patterns that are close, or remain close on next incremental comparisons. The intuition motivating ApEn is that if joint probability measures that describe each of two systems are different, then their marginal distributions on a fixed partition are likely to be different. ApEn assigns a non-negative number to a sequence or time series, with a larger value corresponding to greater apparent serial randomness or irregularity and a smaller value corresponding to more instances of recognizable features in the data. Two input parameters – a block or run length m and a tolerance window r, are required to be specified to compute ApEn. Thus, ApEn(m,r) of a time series computes the logarithmic frequency that runs of patterns that are within r% of the SD (standard deviation) of the time series for m contiguous observations, remain within the same tolerance width r for m+1 contiguous observations. Normalising r to the SD of the time series makes ApEn translation and scale invariant, in that ApEn remains unchanged under uniform process magnification. Precisely, ApEn is defined using the formula

$$\frac{\text{average of the log of the ratios of (m+1) tuples within r - distance of any m+1 tuple}}{\text{average of the log of the ratios of m - tuples within r - distance of any m tuple}}$$

Thus, $0 \leq \text{ApEn} < \infty$ with ApEn = 0 implying perfect regularity. Small values of ApEn imply strong regularity or persistence in a sequence and large values of ApEn imply substantial fluctuation or irregularity. ApEn algorithm counts each sequence as matching itself, in order to avoid the occurrence of log 0 in the calculations. This leads to a bias, which causes ApEn to lack two important properties

- ApEn is heavily dependent on the record length and is uniformly lower than expected for short records
- ApEn lacks relative consistency ie. if ApEn of a data set is higher than that of another, it should, but does not, remain higher for all conditions.

Hence, sample entropy has been introduced by Richman and Moorman (2000) to quantify irregularity in short and noisy time series. SampEn is a new family of statistics which is free of the bias caused by self-matching. SampEn is largely independent of record length and displays relative consistency under circumstances where ApEn does not. The name refers to the applicability to time series data sampled from a continuous process. There are two major differences between SampEn and ApEn statistics :

- SampEn does not count self-matches, which is justified on the ground that entropy being a measure of the rate of information production, comparing data with themselves is meaningless
- SampEn does not use a template-wise approach when estimating conditional probabilities

SampEn(m,r) is precisely the negative natural logarithm of the conditional probability that a data set, having repeated itself within a tolerance r% for m points, will also repeat itself for m+1 points, without allowing self-matches. SampEn displays the property of relative consistency i.e. if SampEn of a data set is higher than that of another data set, for a set of values of the parameters m and r, it remains higher for any other values of the parameters also. Since sample entropy addresses the drawbacks of approximate entropy, sample entropy of the time series of trade price, trade time and trade quantity of a scrip over a period may be used to discern serial irregularity and to study manipulation in the price of the scrip.

Mutual information

The mutual information I(X;Y) between two random variables X and Y with a joint probability mass function p(x,y) and marginal mass functions p(x) and p(y), is defined as the relative entropy D between the joint distribution p(x,y) and the product distribution p(x) p(y).

i.e. $I(X;Y) = D(p(x,y) \mid\mid p(x)\, p(y)) =$

$$\sum_{x} \sum_{y} p(x,y) \log \{p(x,y) / p(x)\, p(y)\}$$

It may be noted that $I(X;Y) \geq 0 = 0$ if X and Y are independent.

Also, $I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X)$ where H denotes the entropy, i.e. mutual information is the reduction in the uncertainty of X due to the knowledge of Y and vice-versa. Due to symmetry, X says as much about Y as Y says about X.

Also, $I(X;Y) = H(X) + H(Y) - H(X,Y)$ and $I(X;X) = H(X)$.

Thus, the mutual information of a random variable with itself is the entropy of the random variable. That is why, entropy is referred to as self-information.

The mutual information I(m) between a time series $\{u_1, u_2, \ldots, u_N\}$ and itself with a delay of m viz. $\{u_{m+1}, u_{m+2}, \ldots, u_N\}$ measures the information carried over by the delayed time series from the original time series. If I(m) is small or around 0, then the two time series are essentially independent and if I(m) is very large, then the delayed series is related to the original series. If the delay m is too short, then the delayed series is similar to the original series and when the data are plotted, most of the observations will lie near the line $u_{m+i} = u_i$ and I(m) will tend to be large. If the delay m is too long, then the data are independent and no information can be gained from the plot and I(m) will tend to be small.

ABOUT THE AUTHORS

Y.V. Reddy is a Reader at the Department of Commerce, Goa University. He can be reached at yvreddy@unigoa.ac.in and yvreddy3@rediffmail.com

A. Sebastin is Asst. Vice President at the National Stock Exchange of India Limited. He can be reached at asebastin@nse.co.in