Published in: NICE Journal of Business 2(1); 2007; 41-53.

PRICE MANIPULATION IN STOCK MARKET
An Entropic Analysis

Y.V. Reddy* and A. Sebastin**

Abstract

This paper seeks to apply information-theoretic concept of entropy to study suspected price manipulation of scrip in the context of the Indian stock market. With the sample entropy (SampEn) methodology, the authors analysed the trade data of scrip reported to have been manipulated in a certain period of time. The values so computed, revealed potential evidence of manipulation.

Key Words: ***Stock price manipulation, Variation, Irregularity, Approximate entropy, Sample entropy***

INTRODUCTION

STOCK price manipulation has been studied by various authors in different situations, such as continuous auction, insider trading, asymmetric information, corners, short squeezes, imperfect competition, financial signaling, equity offerings, takeover bids, 'talking down' the firm, no information, nested information, bluffing and front running. Though there is plenty of literature on market microstructure in general and market manipulation in particular, there is still a scope for an in-depth study of manipulation of stock market prices, using the concepts of stochastic calculus, game theory and information theory among others.

Mathematical modeling and statistical analysis of stock price movements has become a field of its own, beginning with Louis Bachelier's Brownian Motion Model of 1900, to price warrants traded on the Paris bourse (Stock Exchange), and the recent dynamic systems theory and neural networks. Though, pollutants, such as fraud and market manipulation appear nearly impossible to be modelled, they are real and significantly alter price movements without any seeming economic reasons. Simply incorporating fraud into a random effects component of a model fails, as the extent of fraud is rarely chronic, similar to a complicated game, it moves between regulatory efforts and corruptive creativity. Hence, a model independent analytic tool (i.e., a tool providing qualitative inferences across diverse model configurations) tracking stock price movements will be of immense use in order to study price manipulation.

* Reader, Department of Commerce, Goa University, Goa - 403 306; e-mail: *yvreddy@unigoa.ac.in*

** Assistant Vice-President, National Stock Exchange of India Limited, Mumbai - 400 056; e-mail: *masebastin@yahoo.com*

The inspiration for the present study comes from two sources, namely (1) Vila Jean Luc (1989) which presents two examples of market manipulation as games with asymmetric information and derives the conditions for equilibrium, treating the gains and losses as payoffs associated with the various strategies of the players and (2) Gilad Bavly and Abraham Neyman (2003) which presents the feasibility of online correlation in the strategies adopted by a group of players in a repeated game with perfect monitoring, which is concealed from a boundedly rational player and discusses the conditions for the existence of a concealed strategy in terms of entropy, in a theoretical game environment. Entropy is defined as the probability mass or density function of a random variable and a mixed strategy in a game is a probability distribution on the set of all pure strategies available to a player. Hence, entropy of the mixed strategies of the players in a game is defined naturally.

ENTROPY - CONCEPT AND HISTORY

The concept of entropy arose in physical sciences in 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. It 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 forms the macrostate of this equilibrium, i.e., $S = k^* \ln (W)$. Since then, the concept of entropy has been extended to study microscopically unpredictable processes in fields, such as stochastic processes and random fields, information and coding, data analysis and statistical inference, partial differential equations and rational mechanics. This has 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 two mathematicians, Jacob Bernoulli and Abraham de Moivre. The concept of entropic analysis of equity prices was first proposed by Louis Bachelier in 1900. It anticipated many of the mathematical discoveries made later by Norbert Wiener and A.A. Markov in the 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 forties. 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 the classification theory of measure-preserving transformations, by Kolmogorov in the fifties, led to significant advances. The uncertainty about the actual state of a system or a process measured by Shannon is entropy. If the uncertainty is about predictions of a process, it may be decreased by gaining information from the passage of time itself. However, the dynamics of the process may produce new information at each successive stage making forecasting unreliable despite knowledge of the past. This kind of uncertainty about the future is measured by Kolmogorov – Sinai (KS) entropy.

SELECTED STUDIES OF ENTROPIC ANALYSIS IN STOCK MARKET

Nawrocki (1984) used an application of entropy theory for the first time to describe financial market disequilibrium. Ten years later, John Conover (1994) applied entropy theory in devising a methodology for programmed trading of equities. Then, the analysis of entropy tops and bottoms was found by Marantette (1998) as an addition to the buy and sell points of cyclic analysis. In this regard, Pincus and Singer (1998) suggested techniques to produce irregular finite series and normal infinite series, based on constructions and properties derived from approximate entropy (ApEn), a computable formulation of irregularity for a sequence of arbitrary length. Working in the same direction, Maasoumi and Racine (2000) examined the predictability of stock market returns by employing a new metric entropy measure capable of detecting non-linear dependence within the returns series. In addition, Richman and Moorman (2000) introduced a new statistic tool called sample entropy to quantify irregularity in short and noisy time series. The utility of approximate entropy to assess subtle and potentially exploitable changes in serial structure of a financial variable was determined by Pincus and Kalman (2004). Chen (2005) showed that most empirical evidences about market behaviour may be explained by a new information theory generalised from Shannon's entropy theory of information. Finally, Kispert 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.

BASIC CONCEPTS OF ENTROPY

First of all, it is imperative to take a glance at a few basic concepts of entropy as per Shannon's approach, which are widely used in information theory.

Entropy of a Random Variable

Let X be a random variable with p(x) as the probability mass function. Then the entropy of X is defined as

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

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 to some base a.

Joint Entropy

The joint entropy of a pair of random variables X and Y with a joint probability mass function p(x,y) is defined as

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

Conditional Entropy

The conditional entropy of a random variable Y given another variable X is defined as

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

Then we get the chain rule:

$$H(X,Y) = H(X) + H(Y\,/\,X) = H(Y) + H(X\,/\,Y)$$

and more generally,

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \,|\, X_{i-1}, \ldots, X_1)$$

It follows that $H(X_1, X2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$ with equality iff X_i are independent.

Conditioning reduces entropy, i.e., $H(X\,/\,Y) \leq H(X)$, with equality iff X and Y are independent.

Also, $H((X,Y)\,/\,Z) = H(X\,/\,Z) + H(Y\,/\,(X,Z))$.

Relative Entropy

The relative entropy or cross entropy or the Kullback – Leibler (KL) distance between two probability functions p(x) and q(x) is $D(p \| q) =$ $p(x) \log \{p(x)\,/\,q(x)\} = \sum_{x} E[\log \{p(x)\,/\,q(x)\}]$.

It may be noted that $D(p \| q) \geq 0$ $= 0$ if p = q.

However, D(pPq) D(qPp) in general.

Q relative entropy is not symmetric and does not satisfy the triangle property, it is not a true distance between distributions.

Mutual Information

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

i.e., $I(X;Y) = \sum_x \sum_y p(x,y) \log \{p(x,y) / p(x)\, p(y)\}$

$= D(p(x,y) \, P \, p(x)\, p(y))$

It may be noted that $I(X;Y) \geq 0$

$= 0$ if X and Y are independent.

Also, $I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X)$

That is to say, 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.

ENTROPY AND STOCK PRICE MANIPULATION

In the electronic stock trading system, as participants place orders for buying or selling the shares of a scrip at different prices and for various quantities, trades is effected by matching these orders according to price - time priority. Prices 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 or speculative purpose. As the information related to and the perception on the price of a scrip change with time, a participant assigns values to these variables - price, time and quantity - with certain probabilities, while placing orders. Hence, price, time and quantity of the order in respect of scrip may be considered as random variables with probability distributions.

Since Shannon's entropy is defined for a random variable with probability mass function p(x) as $H(p) = - \sum_x p(x) \log p(x)$, we may compute the entropy for the random variables of order price, order time and order quantity in respect of a scrip for every participant, only if a probability distribution can be fitted for each of these variables. As long as a participant places orders in the normal course of business, the entropy values of these variables will be in some range. Just as volatility of price differs from scrip to scrip, and from time to time, the entropy will also vary from scrip to scrip, depending on trading activity.

However, when a participant repeatedly places orders for buying/selling according to some pattern in the price or time or quantity, to manipulate the price of the scrip, the probability distributions of these variables undergo changes which are reflected in the corresponding entropy values. Such orders placed for manipulating the stock market will induce more regularity or persistence in the distributions and consequently, entropy is likely to decrease. Large decrease in the entropy value from normal ranges for the respective variables may lead to potential evidence of price manipulation by a participant. Though this happens rarely, repeated drops in the entropy values of the order-related variables of scrip, within a span of a few trading-days point to likely manipulation in the price of that scrip.

ENTROPY VERSUS VARIANCE

Stock market analysts usually study shifts in mean levels as also variations (in various notations) to understand the state of the market. However, the persistence of certain patterns or shifts may provide critical information. So far, formulae to directly quantify randomness have not been used in market analysis, possibly due to lack of a quantification technology. So, excluding sequential patterns or features which presented themselves, subtler changes in serial structure would remain undetected largely. Volatility is generally equated to the variability of 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 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, Maasoumi and Soofi (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 J. (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 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 the stock market.

APPROXIMATE ENTROPY AND SAMPLE ENTROPY

In the absence of publicly available, participant-wise order data for buying or selling a stock, fitting probability distributions for order price, order time and order quantity is not possible and hence there is no way of computing Shannon entropy values. For any scrip, the only publicly available information is on trade price, trade time and trade quantity, not the identity of the participants who are parties to the trades. Thus, one can construct a time series of each scrip based on price, time and quantity, on a daily basis and under such circumstances, tools for computing the entropy of short and noisy time series. Approximate entropy and sample entropy are advances made in this direction.

Traditional methods for estimating the entropy of a system represented by a time series are not suited to the analysis of short and noisy data sets. The calculation of Shannon's entropy requires the probability density (mass) function of the random variable which denotes the time series. However, Kolmogorov - Sinai's (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 approximate entropy (ApEn), a set of measures of serial irregularity, has been introduced for typically short noisy time series. ApEn, a family of statistics closely related to the entropy measure, provides an appropriate tool to grade the extent of irregularity. 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.

Approximate Entropy Estimation

ApEn measures the logarithmic likelihood of patterns that are close on next incremental comparisons. The intuition motivating ApEn is that if joint probability measures that describe sets 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 the larger value corresponding to greater apparent serial randomness or irregularity, and the 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. To be very precise, ApEn computes the logarithmic frequency of patterns that run within r% of the SD (standard deviation) of a time series for m contiguous observations and 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.

Let the given time series be $u_1, u_2, \ldots, u_N$.

For any m s.t. $1 \le m < N$, define in $\Re^m$ the following m-tuples

$x_{m,1} = (u_1, u_2, \ldots, u_m)$

$x_{m,2} = (u_2, u_3, \ldots, u_{m+1})$

$x_{m,2} = (u_i, u_{i+1}, \ldots u_{i+m-1})$

$x_{m,N-m+1} = (u_{N-m+1}, \ldots, u_N)$

Define $d(x_{m,i}, x_{m,j}) = \max |u_{i+k} - u_{j+k}|$ where $k = 0,1,\ldots,m-1$

This gives the distance between any two m-tuples.

For $r > 0$, the r-neighbourhood of $x_{m,i}$ is defined as $\{x_{m,j} \in \Re^m / d(x_{m,i}, x_{m,j}) \le r\}$

i.e. the neighbourhood of a particular m-tuple $x_{m,i}$ consists of all m-tuples which are within a distance of r units from $x_{m,i}$.

$\because d(x_{m,i}, x_{m,i}) = 0 \;\forall\, i$, the r-neighbourhood of any $x_{m,i}$ is never empty for any r, which is chosen generally as a % of the standard deviation of the data series $\{u_i\}$.

Let $C_{m,i}(r) = \dfrac{\text{Number of } j \text{ s.t. } d(X_{m,i}, X_{m,j}) \le r}{N-m+1}$ = ratio of $x_{m,i}$'s in the r-neighbourhood of $x_{m,i}$

and $\Phi_m(r) = \dfrac{1}{N-m+1}\displaystyle\sum_{i=1}^{N-m+1} \log C_{m,i}(r)$

= average of the log of the ratios of $x_{m,j}$'s in the r-neighbourhood of any $x_{m,i}$

$$\text{Then, } \Phi_{m+1}(r) - \Phi_m(r) = \log \frac{\left[\prod_{i=1}^{N-m} C_{m+1,i}(r)\right]^{\frac{1}{N-m}}}{\left[\prod_{i=1}^{N-m+1} C_{m,i}(r)\right]^{\frac{1}{N-m+1}}} \quad \ldots(1)$$

$$= \frac{\text{average of the log of the ratios of (m+1)tuples in the r-neighbourhood of any } X_{m=1,i}}{\text{average of the log of the ratios of m-tuples in the r-neighbourhood of any } X_{m,i}}$$

The ratio in (1) is always ≤ 1 so that $-\infty < \Phi_{m+1}(r) - \Phi_m(r) \le 0 \;\forall\, r \ge 0$ and $m = 1,2,\ldots,N$.

Fixing m and r, ApEn $= Lt_{N\to\infty}\{\Phi_m(r) - \Phi_{m+1}(r)\}$

Given N data points, this formula is implemented by defining

$$ApEn(m,r,N) = \Phi^{m}(r) - \Phi^{m+1}(r)$$

$\therefore\ \Phi_{m}(r) \geq \Phi_{m+1}(r)$ it is clear that $0 \leq ApEn < \infty$ 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.

For finite N, the largest possible value of ApEn is $-\Phi_{m+1}(r) \leq -\log(N-m)^{-1}$

$$\therefore\ 0 \leq ApEn(m,r,N) \leq \log(N-m)$$

Demerits of Apen

It may be noted that 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.

Let $A_{m,i}(r)$ = Number of $j \neq i$ s. t. $d(x_{m,i}, x_{m,j}) \leq r$

The ApEn algorithm, then assigns to each template $x_{m,i}$, a biased conditional probability of $\frac{A_{m+1,i}(r)+1}{A_{m,i}(r)+1}$ which will be always > the unbiased probability of $\frac{A_{m+1,i}(r)}{A_{m,i}(r)}$. The largest deviation occurs when a large number of templates $x_{m,i}$ have $A_{m,i}(r) = A_{m+1,i}(r) = 0$, since a conditional probability of 1 is assigned to these templates (corresponding to perfect order). The difference between the biased and the unbiased conditional probabilities assigned to individual templates makes the calculations sensitive to record length in a way that depends on the conditional probability. As $N \rightarrow \infty$, $A_{m,i}(r)$ and $A_{m+1,i}(r)$ will be generally large, this makes the biased and the unbiased probabilities asymptotically equivalent. Hence, the bias is evident only for finite data sets and the expected value of ApEn(m,r,N) is less than the parameter ApEn(m,r). This bias cannot be eliminated by simply removing self-matches, unless $C_{m,i}(r) > 0 \forall$ i. One way of reducing the bias is to redefine $A_{m+1,i}(r) = \varepsilon_1$ when $A_{m+1,i}(r) = 0$ and $A_{m,i}(r) = \varepsilon_2$ when $A_{m,i}(r) = 0$ where ε_1 and ε_2 are infinitesimal. However this is arbitrary and hence another family of statistics called Sample entropy (SampEn) has been introduced.

Sample Entropy Estimation

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.

$\because\ x_{m+1,i}$ is not defined for $i = N - \overline{m-1}$, only the first (N - m) vectors of length m and all the (N -m) vectors of length m+1 are considered without self-matches in the calculation of SampEn.

$$\text{Let } B_{m,i}(r) = \frac{\text{Number of } j \neq i \text{ s.t. } d(x_{m,i}, x_{m,j}) \leq r}{N - \overline{m-1}},$$

where j = 1, 2,…,N -m

and

$$B_{m+1,i}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_{m,i}(r)$$

Similarly,

$$\text{let } B_{m+1,i}(r) = \frac{\text{Number of } j \neq i \text{ s.t. } d(x_{m+1,i}, x_{m+1,j}) \leq r}{N - \overline{m-1}},$$

where j = 1, 2,…,N -m

and

$$B_{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_{m+1,i}(r)$$

Then, SampEn(m,r) = $\text{Lt}_{N \to \infty} \log \left[\frac{B_{m+1}(r)}{B_m(r)} \right]$,

which is estimated by the statistic

$$\text{SampEn}(m,r,N) = -\log \left[\frac{B_{m+1}(r)}{B_m(r)} \right]$$

When two (m+1)-tuples are within a distance of r, the corresponding m-tuples are also within the same distance. Hence $B_m(r) \geq B_{m+1}(r)$ always implying that SampEn ≥ 0.

ApEn versus SampEn

There are two major differences between SampEn and ApEn statistics

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

If $B(m,r) = \frac{(N-m-1)(N-m)}{2} B_m(r)$ is the total number of template matches of length m and

$B(m+1,r) = \frac{(N-m-1)(N-m)}{2} B_{m+1}(r)$ is the total number of forward matches of length m+1, then the conditional probability that two sequences within a tolerance of r for m points remain within r at the next point is given by

$$\frac{B(m+1,r)}{B(m,r)} = \frac{B_{m+1}(r)}{B_m(r)}$$ and hence

SampEn(m,r,N) = - log [ ].

In contrast to ApEn(m,r,N) which calculates probabilities in a template-wise fashion, SampEn(m,r,N) calculates the negative logarithm of a probability associated with the time series as a whole. SampEn(m,r,N) is defined except when B(m,r) = 0 which implies that no regularity has been detected or when B(m+1,r) = 0 which corresponds to a conditional probability of 0 and an infinite value of SampEn(m,r,N). The lowest non-zero conditional probability as per this algorithm is $\frac{2}{(N-m-1)(N-m)}$ and, therefore, the upper bound of SampEn(m,r,N) is given by

$$-\log \left[\frac{2}{(N-m-1)(N-m)} \right] = \log \left[\frac{(N-m-1)(N-m)}{2} \right]$$

which is almost double the upper bound of ApEn(m,r,N) viz. log (N- m).

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.

A CASE STUDY

The scrip of Lupin Laboratories Ltd., which has been reported to be the subject to price manipulation on various days during the period from October 1999 to January 2000 (www.sebi.com) was chosen for the study. The prices of all the trades executed in the scrip on the National Stock Exchange were taken for the various trading days during this period. The differences in the prices of successive trades were taken as time series. By taking such first differences, stationary character of the time series may be assumed safely 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. The computation of SampEn for each time series requires two input parameters, viz., the template size m and the tolerance window r. For time series with a few thousand data points, a value between 2 and 5 for m and a value between 0.1 and 0.25 for r are used generally for computing SampEn. In this analysis, it was observed empirically that the results were not too dependent on specific values of m and r within these ranges respectively.

SampEn of the time series consisting of the differences in successive trade prices of the scrip of Lupin Laboratories Ltd., was computed for m = 2, 3, 4, 5 and r = 0.15, 0.16, 0.17, 0.18, 0.19, and 0.20. These SampEn values for the trading days in October 1999, November, December and January 2000 are given in Tables 1, 2, 3 and 4, respectively. It may be observed that SampEn is very low on days 4, 5, 6, 7, 8, 9, 10 and 13 in October, on days 5, 17 and 19 in November, on days 15, 16 and 18 in December 1999 and on days 1 and 7 in January 2000, for all values of m and r. Specifically, SampEn for all these 16 days is utmost 0.20 for m = 5 and r = 0.20*SD. Also, SampEn was 0 on days 6, 7 and 13 in October and on day 17 in November 1999 implying maximum regularity in the data pertaining to these days. The fact that the trades on each of these days were executed at the same price lends credence to the maximum possible regularity in the time series and hence the least possible value of SampEn for each of these days.

The above mentioned 16 days in October 1999 - January 2000, when the SampEn is very low, are days of potential manipulation in the price of the Lupin Laboratories Ltd. scrip. The order related data for the scrip, pertaining to these 16 days may be analysed further to observe the trading patterns of the participants and discern price manipulation attempt by any participant.

CONCLUSION

Entropic analysis is a novel area for the Indian stock market and with a near vacuum in research efforts. This paper applies entropic analysis to study price manipulation. Due to non-availability of participant-wise, order-related data, sample entropy was found suitable for this study. SampEn values for the trade price data related to the scrip of Lupin Laboratories Ltd., on various trading days in the period during which it was reported to be subject to price manipulation, supported such suspicion. This empirical analysis should be done for many other scrips to reach the conclusion that entropic analysis is an effective tool to study stock market manipulation. Further, since all orders placed by a participant may not result in trades due to non-availability of matching counter-orders and some orders may result in multiple trades, the data may not reveal the entire information contained in all the orders. Hence entropic analysis of order related data (if available) will ensure more efficiency in the study of price manipulation attempts in the stock market.

REFERENCES

Bavly, Gilad, and Abraham Neyman (2003), "Online Concealed Correlation by Boundedly Rational Players", Discussion Paper No. 336, www.ratio.huji.ac.il/neyman/publications (accessed on 22.04.2005, 08.03.2007)

Chen, Jing (2005), "Information Theory and Market Behaviour", University of Northern British Columbia, Working Paper

Cover, Thomas M., and Joy A. Thomas (1991), Elements of Information Theory (New York: John Wiley and Sons)

Greven, Andreas, Gerhard Keller, and Gerald Warnecke (2003), Entropy (Princeton: Princeton University Press)

John Conover (1994), "NtropiX", www.johncon.com (accessed on 31.05.2006, 31.01.2007)

Luc, Vila Jean (1989), "Simple Games of Market Manipulation", Economic Letters, Vol. 29, Issue 1, pp. 21-26

Maasoumi, Esfandiar, and Je Racine (2000), "Entropy and Predictability of Stock Market Returns", University of California San Diego, Working Paper

Marantette, David T. (1998), The Entropy Danger, www.goldstock.com (accessed on 21.03.2006)

Nawrocki, David (1984), "Entropy, Bifurcation and Dynamic Market Disequilibrium", The Financial Review, Vol. 19, Issue 2, May 1984, pp. 266 - 284.

Pincus, Steve, and Burton H. Singer (1998), "A Recipe for Randomness", Proceedings of the National Academy of Sciences of USA, Vol. 95, No. 4, pp. 1363-1368.

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

Richman, Joshua S., and J. Randall Moorman (2000), "Physiological Time-Series Analysis using Approximate Entropy and Sample Entropy", American Journal of Physiology, Vol. 278, No. 6, pp. 2039 - 2049

www.sebi.gov.in (accessed on 30.11.2005, 30.12.2006)

Table 1

Sample Entropy Values for Lupin Laboratories Ltd.

m, r	*Day1*	*Day2*	*Day3*	*Day4*	*Day5*	*Day6*	*Day7*	*Day8*	*Day9*	*Day10*	*Day11*	*Day12*	*Day13*	*Day14*	*Day15*
2, 0.15	1.031	1.184	0.961	0.132	0.05	0	0	0.391	0.22	0.239	0.644	0.905	0	0.482	0.551
2, 0.16	1.031	1.025	0.961	0.132	0.05	0	0	0.391	0.22	0.241	0.644	0.905	0	0.482	0.551
2, 0.17	0.94	1.025	0.961	0.132	0.05	0	0	0.391	0.22	0.241	0.644	0.838	0	0.482	0.551
2, 0.18	0.94	1.025	0.961	0.132	0.05	0	0	0.371	0.221	0.241	0.63	0.838	0	0.482	0.551
2, 0.19	0.94	1.025	0.961	0.132	0.05	0	0	0.371	0.221	0.241	0.63	0.838	0	0.492	0.551
2, 0.20	0.94	1.025	0.961	0.132	0.05	0	0	0.371	0.222	0.24	0.63	0.838	0	0.492	0.517
3, 0.15	0.993	1.006	0.887	0.098	0.051	0	0	0.276	0.15	0.176	0.541	0.843	0	0.352	0.431
3, 0.16	0.993	0.9	0.887	0.098	0.051	0	0	0.276	0.15	0.179	0.541	0.843	0	0.352	0.431
3, 0.17	0.855	0.9	0.887	0.098	0.051	0	0	0.276	0.15	0.179	0.541	0.775	0	0.352	0.431
3, 0.18	0.855	0.9	0.887	0.098	0.051	0	0	0.273	0.15	0.179	0.534	0.775	0	0.352	0.431
3, 0.19	0.855	0.9	0.887	0.098	0.051	0	0	0.273	0.15	0.179	0.534	0.775	0	0.362	0.431
3, 0.20	0.855	0.9	0.887	0.099	0.051	0	0	0.273	0.151	0.184	0.534	0.775	0	0.362	0.406
4, 0.15	0.828	0.719	0.871	0.077	0.053	0	0	0.236	0.063	0.16	0.462	0.673	0	0.288	0.298
4, 0.16	0.828	0.738	0.871	0.077	0.053	0	0	0.236	0.063	0.159	0.462	0.673	0	0.288	0.298
4, 0.17	0.746	0.738	0.871	0.077	0.053	0	0	0.236	0.063	0.159	0.462	0.632	0	0.288	0.298
4, 0.18	0.746	0.738	0.871	0.077	0.053	0	0	0.245	0.063	0.159	0.449	0.632	0	0.288	0.298
4, 0.19	0.746	0.738	0.871	0.077	0.053	0	0	0.245	0.063	0.159	0.449	0.632	0	0.295	0.298
4, 0.20	0.746	0.738	0.871	0.078	0.053	0	0	0.245	0.063	0.164	0.449	0.632	0	0.295	0.299
5, 0.15	0.668	0.633	0.874	0.062	0.054	0	0	0.178	0.048	0.124	0.419	0.587	0	0.236	0.235
5, 0.16	0.668	0.745	0.874	0.062	0.054	0	0	0.178	0.048	0.123	0.419	0.587	0	0.236	0.235
5, 0.17	0.617	0.745	0.874	0.062	0.054	0	0	0.178	0.048	0.123	0.419	0.567	0	0.236	0.235
5, 0.18	0.617	0.745	0.874	0.062	0.054	0	0	0.192	0.048	0.123	0.404	0.567	0	0.236	0.235
5, 0.19	0.617	0.745	0.874	0.062	0.054	0	0	0.192	0.048	0.123	0.404	0.567	0	0.245	0.235
5, 0.20	0.617	0.745	0.874	0.063	0.054	0	0	0.192	0.048	0.123	0.404	0.567	0	0.245	0.229

Table 2

Sample Entropy Values for Lupin Laboratories Ltd.

m, r	Day1	Day2	Day3	Day4	Day5	Day6	Day7	Day8	Day9	Day10	Day11	Day12	Day13	Day14	Day15
2, 0.15	0.785	0.677	0.95	1.021	0.368	1.089	0.957	0.863	0.744	1.055	0.893	0.901	0.574	1.235	1.139
2, 0.16	0.785	0.677	0.95	1.021	0.368	1.089	0.957	0.863	0.744	1.055	0.893	0.901	0.574	1.235	1.139
2, 0.17	0.785	0.677	0.95	0.913	0.368	1.089	0.957	0.863	0.744	1	0.849	0.901	0.574	1.045	0.968
2, 0.18	0.786	0.677	0.95	0.913	0.368	1.089	0.957	0.863	0.744	1	0.849	0.901	0.574	1.045	0.968
2, 0.19	0.786	0.658	0.901	0.913	0.367	1.089	0.893	0.863	0.744	1	0.849	0.901	0.546	1.045	0.968
2, 0.20	0.786	0.658	0.901	0.913	0.367	1.019	0.893	0.863	0.744	1	0.849	0.901	0.546	1.045	0.968
3, 0.15	0.639	0.544	0.817	1.009	0.257	1.026	0.938	0.731	0.665	1.064	0.755	0.821	0.433	1.224	1.043
3, 0.16	0.639	0.544	0.817	1.009	0.257	1.026	0.938	0.731	0.665	1.064	0.755	0.821	0.433	1.224	1.043
3, 0.17	0.639	0.544	0.817	0.857	0.257	1.026	0.938	0.731	0.665	1.024	0.739	0.821	0.433	1.044	0.927
3, 0.18	0.639	0.544	0.817	0.857	0.257	1.026	0.938	0.731	0.665	1.024	0.739	0.821	0.433	1.044	0.927
3, 0.19	0.639	0.538	0.757	0.857	0.26	1.026	0.873	0.731	0.665	1.024	0.739	0.821	0.463	1.044	0.927
3, 0.20	0.639	0.538	0.757	0.857	0.26	0.976	0.873	0.731	0.665	1.024	0.739	0.821	0.463	1.044	0.927
4, 0.15	0.54	0.42	0.669	1.005	0.194	0.909	0.812	0.61	0.552	1.031	0.612	0.835	0.338	0.971	0.889
4, 0.16	0.54	0.42	0.669	1.005	0.194	0.909	0.812	0.61	0.552	1.031	0.612	0.835	0.338	0.971	0.889
4, 0.17	0.54	0.42	0.669	0.817	0.194	0.909	0.812	0.61	0.552	1.012	0.592	0.835	0.338	0.908	0.852
4, 0.18	0.528	0.42	0.669	0.817	0.194	0.909	0.812	0.61	0.552	1.012	0.592	0.835	0.338	0.908	0.852
4, 0.19	0.528	0.425	0.64	0.817	0.189	0.909	0.776	0.61	0.552	1.012	0.592	0.835	0.388	0.908	0.852
4, 0.20	0.528	0.425	0.64	0.817	0.189	0.868	0.776	0.61	0.552	1.012	0.592	0.835	0.388	0.908	0.852
5, 0.15	0.436	0.347	0.686	1.094	0.177	0.92	0.757	0.555	0.495	1.025	0.493	0.726	0.266	0.794	0.892
5, 0.16	0.436	0.347	0.686	1.094	0.177	0.92	0.757	0.555	0.495	1.025	0.493	0.726	0.266	0.794	0.892
5, 0.17	0.436	0.347	0.686	0.833	0.177	0.92	0.757	0.555	0.495	1.061	0.506	0.726	0.266	0.864	0.833
5, 0.18	0.461	0.347	0.686	0.833	0.177	0.92	0.757	0.555	0.495	1.061	0.506	0.726	0.266	0.864	0.833
5, 0.19	0.461	0.366	0.623	0.833	0.172	0.92	0.741	0.555	0.495	1.061	0.506	0.726	0.31	0.864	0.833
5, 0.20	0.461	0.366	0.623	0.833	0.172	0.899	0.741	0.555	0.495	1.061	0.506	0.726	0.31	0.864	0.833

Table 3

Sample Entropy Values for Lupin Laboratories Ltd.

m, r	*Day1*	*Day2*	*Day3*	*Day4*	*Day5*	*Day6*	*Day7*	*Day8*	*Day9*	*Day10*	*Day11*	*Day12*	*Day13*	*Day14*	*Day15*	*Day*
2, 0.15	0.87	0.992	0.763	0.576	0.962	0.864	0.857	0.91	0.815	0.775	1.085	0.641	0.805	0.821	0.183	0.03
2, 0.16	0.87	0.992	0.763	0.576	0.962	0.864	0.855	0.91	0.766	0.775	1.049	0.641	0.805	0.804	0.175	0.03
2, 0.17	0.87	0.992	0.763	0.526	0.962	0.864	0.855	0.834	0.766	0.735	1.049	0.641	0.805	0.804	0.175	0.03
2, 0.18	0.87	0.992	0.763	0.526	0.962	0.864	0.855	0.834	0.766	0.735	1.049	0.641	0.806	0.804	0.175	0.03
2, 0.19	0.814	0.915	0.763	0.526	0.909	0.852	0.855	0.834	0.766	0.735	1.042	0.641	0.806	0.804	0.175	0.03
2, 0.20	0.814	0.915	0.763	0.526	0.909	0.852	0.855	0.834	0.716	0.735	1.042	0.624	0.806	0.787	0.175	0.03
3, 0.15	0.816	0.994	0.715	0.477	0.904	0.76	0.79	0.655	0.835	0.691	1.096	0.534	0.718	0.749	0.123	0.03
3, 0.16	0.816	0.994	0.715	0.477	0.904	0.76	0.8	0.655	0.774	0.691	1.077	0.534	0.718	0.736	0.118	0.03
3, 0.17	0.816	0.994	0.715	0.453	0.904	0.76	0.8	0.616	0.774	0.652	1.077	0.534	0.718	0.736	0.118	0.03
3, 0.18	0.816	0.994	0.715	0.453	0.904	0.76	0.8	0.616	0.774	0.652	1.077	0.534	0.716	0.736	0.118	0.03
3, 0.19	0.748	0.889	0.715	0.453	0.844	0.755	0.8	0.616	0.774	0.652	1.067	0.534	0.716	0.736	0.118	0.03
3, 0.20	0.748	0.889	0.715	0.453	0.844	0.755	0.8	0.616	0.718	0.652	1.067	0.513	0.716	0.726	0.118	0.03
4, 0.15	0.793	1.002	0.665	0.393	0.772	0.704	0.696	0.67	0.725	0.604	0.998	0.503	0.692	0.679	0.109	0.0
4, 0.16	0.793	1.002	0.665	0.393	0.772	0.704	0.722	0.67	0.686	0.604	0.956	0.503	0.692	0.665	0.106	0.0
4, 0.17	0.793	1.002	0.665	0.392	0.772	0.704	0.722	0.598	0.686	0.556	0.956	0.503	0.692	0.665	0.106	0.0
4, 0.18	0.793	1.002	0.665	0.392	0.772	0.704	0.722	0.598	0.686	0.556	0.956	0.503	0.69	0.665	0.106	0.0
4, 0.19	0.746	0.905	0.665	0.392	0.764	0.689	0.722	0.598	0.686	0.556	0.946	0.503	0.69	0.665	0.106	0.0
4, 0.20	0.746	0.905	0.665	0.392	0.764	0.689	0.722	0.598	0.667	0.556	0.946	0.485	0.69	0.659	0.106	0.0
5, 0.15	0.749	1.129	0.611	0.358	0.716	0.628	0.674	0.632	0.776	0.658	0.82	0.462	0.644	0.634	0.098	0.0
5, 0.16	0.749	1.129	0.611	0.358	0.716	0.628	0.692	0.632	0.768	0.658	0.775	0.462	0.644	0.611	0.096	0.0
5, 0.17	0.749	1.129	0.611	0.355	0.716	0.628	0.692	0.581	0.768	0.611	0.775	0.462	0.644	0.611	0.096	0.0
5, 0.18	0.749	1.129	0.611	0.355	0.716	0.628	0.692	0.581	0.768	0.611	0.775	0.462	0.633	0.611	0.096	0.0
5, 0.19	0.696	1.033	0.611	0.355	0.746	0.601	0.692	0.581	0.768	0.611	0.768	0.462	0.633	0.611	0.096	0.0
5, 0.20	0.696	1.033	0.611	0.355	0.746	0.601	0.692	0.581	0.757	0.611	0.768	0.477	0.633	0.615	0.096	0.0

Table 4

Sample Entropy Values for Lupin Laboratories Ltd.

(for the month of January 2000)

m, r	*Day1*	*Day2*	*Day3*	*Day4*	*Day5*	*Day6*	*Day7*
2, 0.15	0.339	0.835	0.787	1.031	0.733	1.065	0.376
2, 0.16	0.339	0.835	0.787	1.031	0.733	1.065	0.376
2, 0.17	0.316	0.779	0.781	0.979	0.733	0.99	0.376
2, 0.18	0.316	0.779	0.781	0.979	0.721	0.99	0.374
2, 0.19	0.316	0.779	0.781	0.979	0.721	0.981	0.374
2, 0.20	0.315	0.77	0.773	0.963	0.721	0.965	0.374
3, 0.15	0.267	0.839	0.705	1.166	0.646	0.972	0.31
3, 0.16	0.267	0.839	0.705	1.166	0.646	0.972	0.31
3, 0.17	0.256	0.793	0.698	1.096	0.646	0.903	0.31
3, 0.18	0.256	0.793	0.698	1.096	0.637	0.903	0.316
3, 0.19	0.256	0.793	0.698	1.096	0.637	0.893	0.316
3, 0.20	0.255	0.778	0.69	1.072	0.637	0.889	0.316
4, 0.15	0.228	0.733	0.656	1.182	0.575	0.992	0.272
4, 0.16	0.228	0.733	0.656	1.182	0.575	0.992	0.272
4, 0.17	0.218	0.713	0.647	1.111	0.575	0.935	0.272
4, 0.18	0.218	0.713	0.647	1.111	0.566	0.935	0.272
4, 0.19	0.218	0.713	0.647	1.111	0.566	0.93	0.272
4, 0.20	0.218	0.694	0.636	1.108	0.566	0.926	0.272
5, 0.15	0.206	0.736	0.688	1.027	0.423	1.085	0.203
5, 0.16	0.206	0.736	0.688	1.027	0.423	1.085	0.203
5, 0.17	0.187	0.683	0.675	0.91	0.423	0.98	0.203
5, 0.18	0.187	0.683	0.675	0.91	0.43	0.98	0.205
5, 0.19	0.187	0.683	0.675	0.91	0.43	0.964	0.205
5, 0.20	0.187	0.698	0.66	0.922	0.43	0.951	0.205