Applications of statistical physics and stochastic processes in economics: A changing paradigm

Ashok Jain[1] and M.K.Das[2]

[1]527, Pocket-B, Sarita Vihar, New Delhi-110076.

[2] Institute of Informatics & Communication, University of Delhi South Campus, New Delhi-110021.

Abstract

The emergent paradigm of Econophysics involves incorporation of methods of statistical physics and stochastic processes in macro-economic studies. Recognizing that Market Efficiency Hypothesis (MEH) is often used in economics to gain insight into market dynamics, in this paper we present application of these methods to estimation of crude oil market efficiency. The objective is to illustrate application of Econophysics approaches in estimation of market efficiency; analysis of MEH, its merits or limitations, does not come under the overview or scope of our work. Based on the paper by Ortiz-Cruz et al. (2012), efficiency of crude oil market has been estimated using Brent and West Texas Intermediate (WTI) crude oils price (in US $/barrel) time series data for the period 1986-2017 within the framework of both approximate and sample entropy. The method also throws some light on dynamics of relationship between unusual fluctuations in oil prices attributable to 'critical events' and changes in market efficiency.

The Raison d'être for New Paradigm

“ We are at the beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on.”

Richard Feynman

“The blame game continues over who is responsible for the worst recession since the Great Depression- the financiers who did such a bad job of managing risk or the regulators who failed to stop them. But the economics profession bears more than a little culpability…….”

Joesph Stieglitz, ‘ Needed: A New Economic Paradigm’, *Financial Times, August 20, 2010*)

1. Introduction

While search for a New Economic Paradigm is an on going project, Aoyama et.al (2017) have introduced Macro-Econophysics as a new paradigm in the study of macro economic activity; a paradigm that essentially calls for looking at macro economic activities from the perspective of statistical physics and stochastic processes. To illustrate the applicability of Econophysics in economics, we have chosen MEH approach used in economics as an example of one of the areas of application. Estimation of Informational efficiency, an aspect of MEH, has often been suggested as providing useful inputs to regulators in taking decisions regarding intervention in market. Use of statistical physics approaches and viewing dynamics of economic activity as an stochastic process, it becomes possible to estimate informational efficiency that is efficiency with which information gets embedded and becomes operational in the market in terms of entropy. Gulko Les (1999) had postulated endogenous emergence of entropy as a condition for **seeking** price information and maximization of entropy as an estimate of o**perational** aspect that is of informational efficiency. Gulko Les had applied their approach to study efficiency of stock prices market.

I this paper, following Ortiz-Cruz et al. (2012), we have analyzed efficiency of crude oil market within the frameworks of both approximate and sample entropy using Brent and West Texas Intermediate (WTI) crude oils price (in US $/barrel) time series data for the period 1986-2017.

Section 2 describes the mathematical formulation of Entropy based analysis and WTI data. The WTI price data is presented graphically to show major variations and related events and as logarithmic price differences over the entire period and in various time segments to reveal time taken for market to operationalize information embedded as entropy. Results are given in Section 3 and concluding concluding remarks in Section 4.

2. Entropy based analysis:

(a)Methodology Used

To analyze a time series within the frame work of entropy, several measures e.g. Shannon entropy, Spectral entropy, Kolmogorov entropy, permutation entropy , approximate entropy and sample entropy have been used in the literature to characterize it as regular , chaotic or stochastic. It is known that the spectral entropy and Shannon entropy do not involve the dynamics of the system and hence are of limited application for a nonlinear system. In order to incorporate the system dynamics, Grassberger and Procaccia (1983) proposed the idea of computing Kolmogorov-Sinai entropy (KS) that uses the correlation integral defined as:

$$C(R) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq 1}^{N} \Theta(R - |x[i] - x[j]|)$$

where $C(R)$ represents the total number of points contained within hyper-sphere of radius R and Θ correspond to Heavyside operator. Further N refers to number of points in the m dimensional phase space. Using the correlation integral as above, the average of log correlation sum for all points along the attractor is defined as:

$$\Phi^m(R) = \frac{1}{N-m+1} \sum_{n=1}^{N-m+1} \log C(R)$$

The KS is thus defined as:

$$KS = \lim_{R\to 0} \lim_{m\to\infty} \lim_{N\to\infty} [\Phi^m - \Phi^{m+1}]$$

and describe how often a trajectory reaches a point along the attractor in the limit $R \to 0$. It may be noted that complicated complex trajectories do not visit the attractor regularly. The foregoing expression for KS can be evaluated only approximately because of the limited data size and also due to the presence of noise, the limit $R \to 0$ cannot be realized. Pincus(1994) suggested the need to examine only samples at fixed N, m and R and coined the term approximate entropy ($ApEn$) to correspond the regularity with which a given time series template gets repeated on next increment. Therefore for more repeatable segments, the time series is predictable and hence has low entropy while for less predictable time series, the repeatable segments are few and the situation corresponds to high entropy.

In the following we provide some details of the algorithm.

Step:1 Consider a time series

$$X = x_1, x_2, \dots \dots, x_N$$

of length N. If the x_i's are sampled at time T_s, then a scale τ is defined as $\tau = NT_s$.

Step:2 Select two vectors

$$u^m(i) = \{x_i, x_{i+1}, x_{i+2}, \dots x_{i+m-1}\} \,;\; v^m(j) = \{x_j, x_{j+1}, x_{j+2}, \dots x_{j+m-1}\}; i \neq j, 1 \leq i \leq j \leq N-m+1.$$

Step:3 These vectors are similar if their distance

$$d_{u,v}(i,j) = max\{u(i+k) - v(j+k)| : 0 \leq k \leq m-1\}$$

is smaller than a specified tolerance ϵ.

Step:4 For every $N-m+1$ vectors $u^m(i)$, the number of similar vectors $v^m(j)$ is given by their respective distances. If ${B_i}^m$ is he number of vectors $v^m(j)$ similar to $u^m(i)$, the relative frequency of finding $v^m(j)$ is given by

$$C_i(m, \epsilon, \tau) = \frac{{B_i}^m}{N-m};$$

where

$$B_i{}^m = \sum_{j=1}^{N-m} \Theta[|u^m(i) - v^m(j)| < \epsilon]$$

Here $N - m$ is the total potential number of vectors $v^m(j) \neq u^m(i)$. The value of ϵ is usually taken as 0.15 times the standard deviation of the time series X.

Step:5 Next compute the relative frequency of the logarithm of $C_i(m, \epsilon, \tau)$,

$$\Phi(m, \epsilon, \tau) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i(m, \epsilon, \tau)$$

Step:6 For finite N , on repeating the foregoing steps by letting $m \rightarrow m + 1$, the approximate entropy is given by

$$ApEn(m, \epsilon, \tau, N) = \frac{1}{T_s}[\Phi(m, \epsilon, \tau) - \Phi(m + 1, \epsilon, \tau)]$$

In fact the complexity of a time series is not confine to a single scale and in fact may depend on different scales . Such dependency arises because a signal / time series may be more certain for some scales while it is uncertain for other scales. In other words, a multi-scale approach may provide an index corresponding to the mean rate of creation of information at a given time scale. A moving average method is generally used to obtain a time series at scale $(= n\Delta t$, Δt is the sampling period) from a given time series X:

$$Y_i = \frac{1}{n}\sum_{k=1}^{n} x_{i+k-1}$$

In this way , if say a financial market, is considered and daily price of a commodity is noted, then the foregoing procedure yields a daily varying time series for $n = 1$, weekly varying time series for $n = 5$ and monthly times series $n = 20$ etc. For each such time series and therefore scale , we may compute the scale dependent or multi-scale $ApEn$.

Richman and Moorman (2000) suggested an improved method of estimating entropy and named it as sample entropy (*SampEn*) which is free of bias due to self matching in *ApEn* computation. In their formulation , defining

$$B = \sum_{i=1}^{N-m} {B_i}^m \Big/ N - m \ ; \ A = \sum_{i=1}^{N-m} {A_i}^m \Big/ N - m - 1$$

where

$$A_i{}^m = \sum_{j=1}^{N-m} \Theta[|u^{m+1}(i) - v^{m+1}(j)| < \epsilon] ;$$

the sample entropy for a time series of length N, is defined by

$$SampEn(m, r, N) = -\ln\left[\frac{A}{B}\right].$$

Using moving average method , described earlier, the multi-scale sample entropy (*SampEn*) can be computed for a given time series X at different scale τ.

(b) Crude oil WTI data for the period 1986-2017

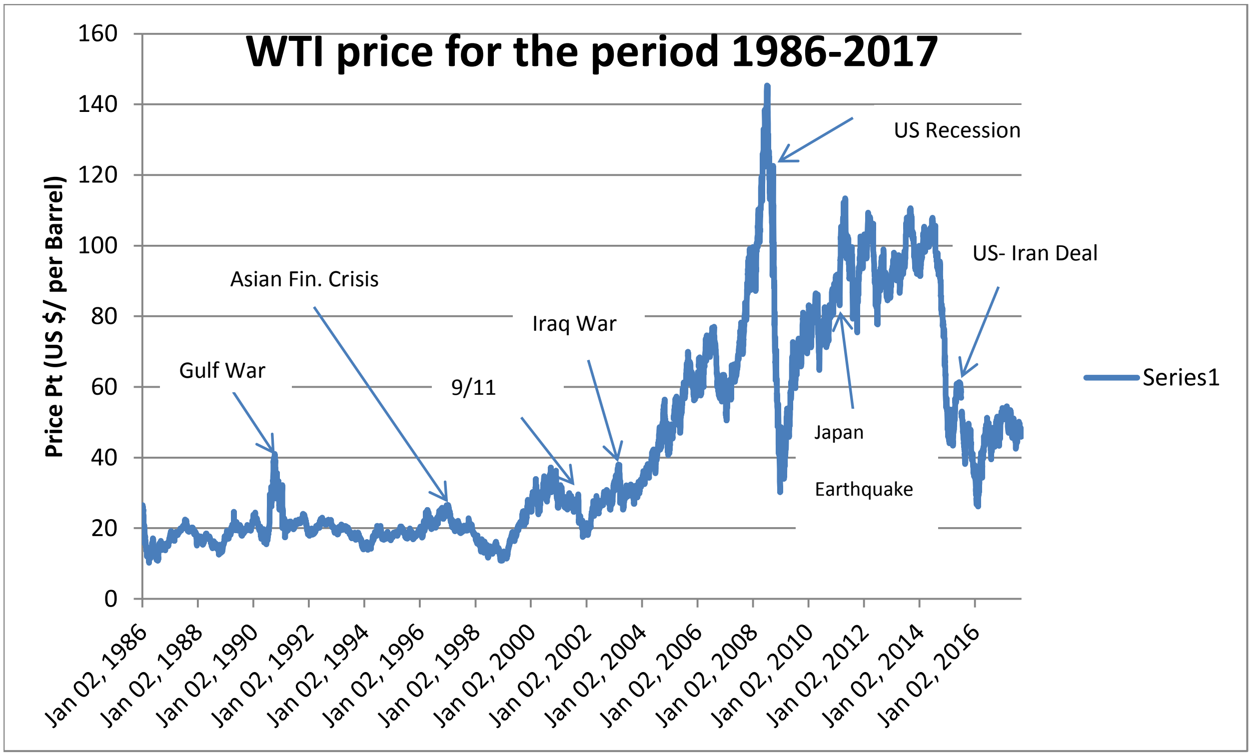


Fig.1 Time series of the WTI price for the period 1986-2017

In this work , we have used the crude oil price data obtained freely from the Energy Information Administration (EIA) at http://tonto.eia.doe.gov for the period 1986-2017. The span of the data is Jan 02,1986 Jan 02,2017 as shown in Fig.1 and it also indicate the various important events which might have affected the global economy. Fig.2 illustrates the logarithmic price differences time series for the entire period and used for further analysis .

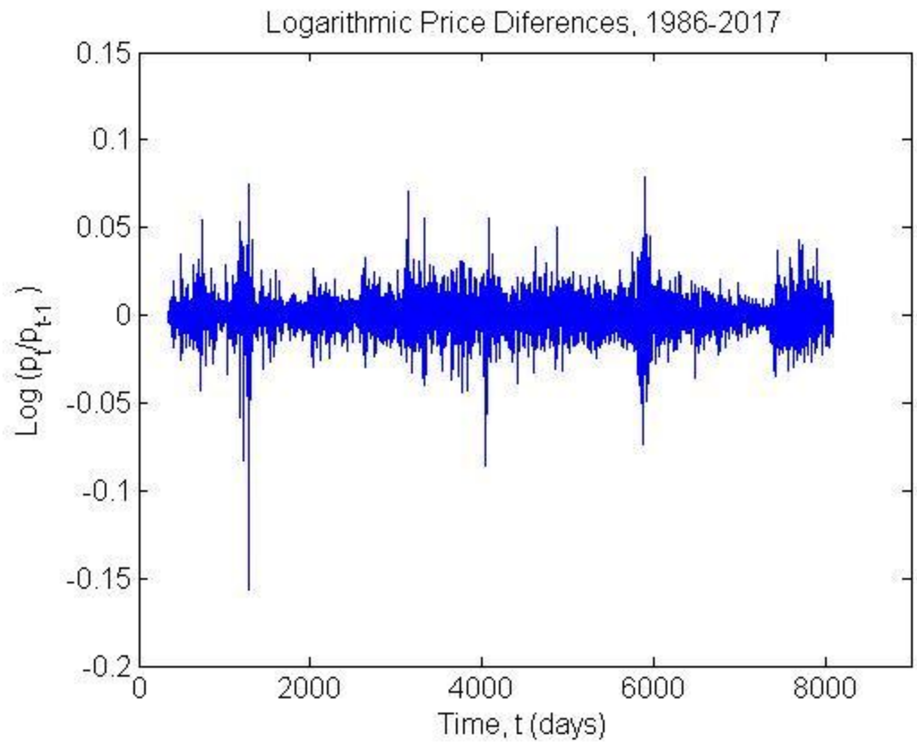
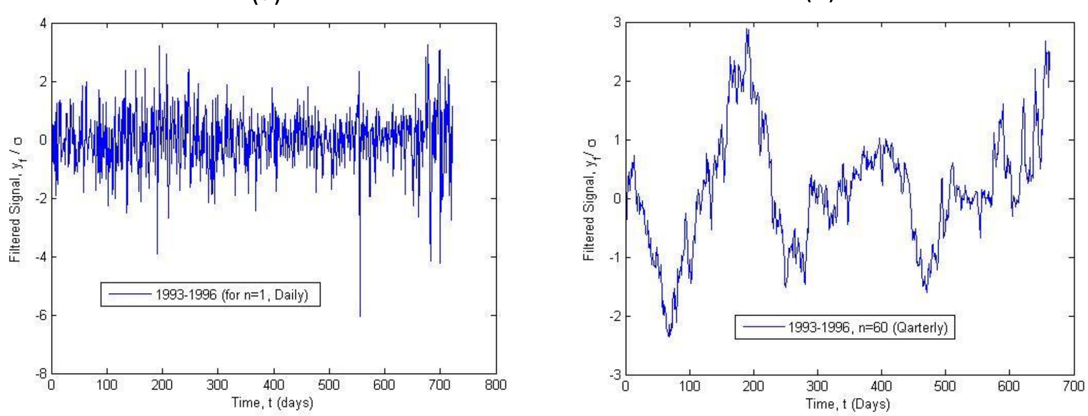


Fig.2, Time series of logarithmic price differences for the period 1986-2017.

The representative daily $(n = 1)$, quarterly $(n = 60)$ and six monthly $(n = 120)$ time series derived on using the moving average or low pass filter described earlier are shown in Fig.3.

(a) (b)



(c)

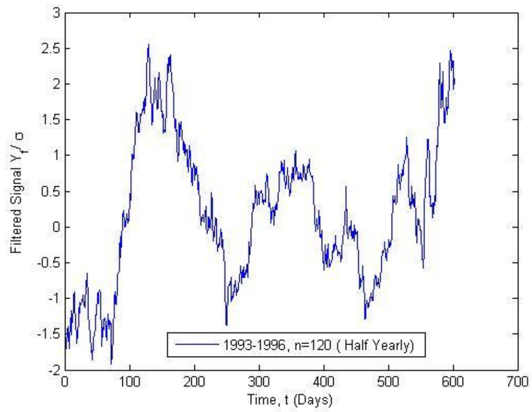


Fig.3 Daily ($n = 1$) , quarterly ($n = 60$) and six monthly ($n = 120$) time series for the year 1993-1996.

3.Results

In this work , we restricted ourselves to compute the multi-scale approximate (*ApEn*) and multi-scale sample entropy (*SampEn*) of WTI crude oil price data for the period 1986-2014. The primary objective is to relate the results of the computation to socio-economic events marked in Fig.1. The analysis is further divided into 3-yr periods namely 1986-1989, 1993-1996, 1999-2002, 2004-200, 2007-2010 and 2011-2014. The behavior of multi-scale approximate entropy (*ApEn*) and sample entropy (*SampEn*) with scale is shown in Fig.4, for the 1986-1989 WTI subset data, wherein we observe a general decreasing trend in the entropies with scale. Our computational results for *ApEn* are in agreement with those reported earlier by Ortiz-Cruz et al.(2012). However , for time scales up to one quarter (~60 days), the persistent decreasing trend of both *ApEn* and *SampEn* is interspersed with local maximum (Fig.4). The observed decreasing

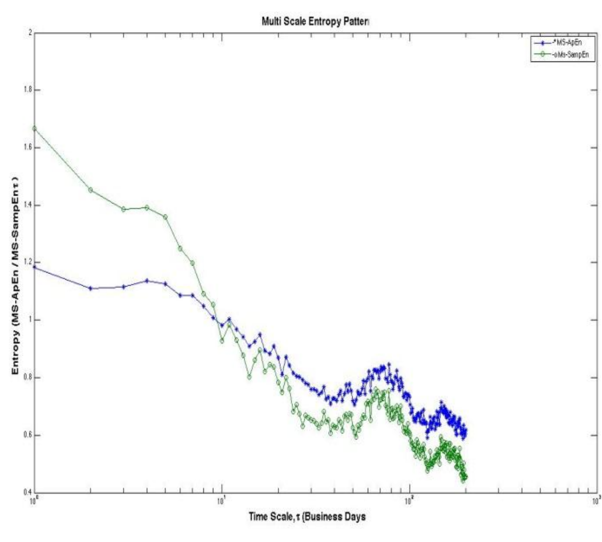
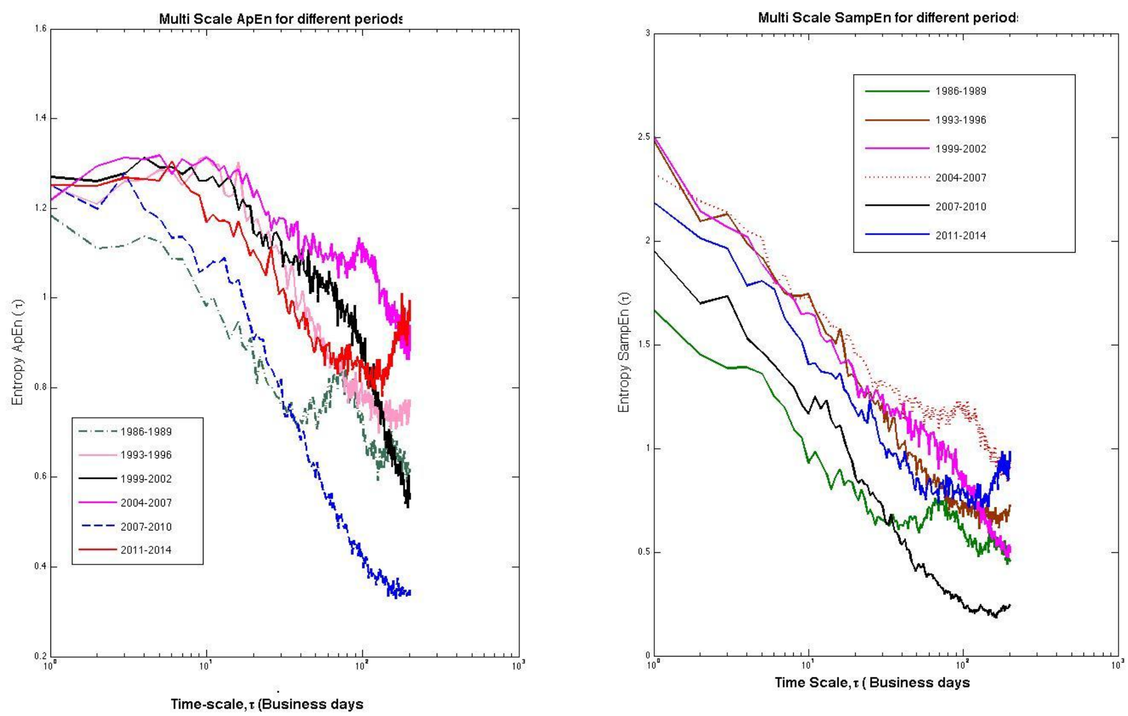


Fig.4 Multi-scale *SampEn* and *ApEn* of the low - pass filtered data for the period 1986-1989.

decrease in entropy suggests a decreased information content in the first quarter. Beyond the first quarter, say around ~70-80 business days and also ~160 business days , a conspicuous entropy maximum occurs thereby making it harder to predict the ongoing price dynamics. Therefore crude oil market seems to be of higher complexity because of the higher entropy levels than it is when the entropy is lower. It is thus possible to conjecture that the crude oil price predictability is higher at larger time scale, τ. We extended the analysis to different periods e.g., 1993-1996, 1999-2002, 2004-200, 2007-2010 and 2011-2014 and the results are summarized in Fig.5.

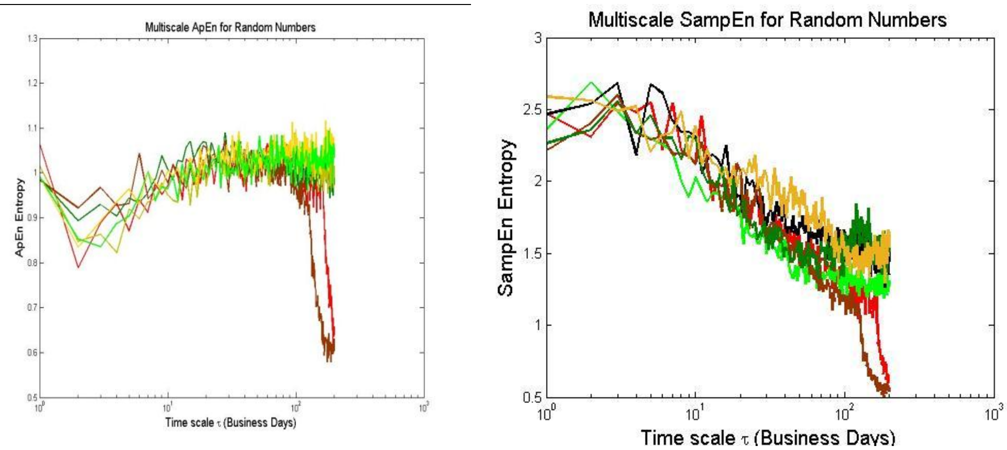


(a) (b)

Fig.5, Multi-scale entropy of the low - pass filtered data for the six selected subperiods, (a) *ApEn*, (b) *SampEn*.

The run of both multi-scale sample entropy and approximate entropy for the period 2004-2007 with scale, τ, suggests a higher level of entropy and also increased pattern diversity than the other five subperiods . Such observations may be attributed to increased growth in US economy following the 9/11 (2001) crisis ,Iraq war and crude oil demand from Asian Pacific countries. However, for the recession period 2007-2010, the crude oil market has been more predictable due to continuous decrease in both *ApEn* and *SampEn* at all scales indicating reduced market efficiency. Similarly, the period 2011-2014 involving Japanese tsunami and US-Iran deal , the increased fluctuation in crude oil price were observed (cf.Fig.1). As a result, the market seems to be more efficient at all scale compare to that in 2007-2010 period. The increased diversity of patterns at all scales , in both *ApEn* and *SamEn* further lends support to more efficient market scenario than in 2007-2010.

For efficient market , empirical test involve the rejection of the hypothesis that logarithmic price differences follow a random walk model. In case the prices follow a random walk process then the prices remains unpredictable as no regular pattern will exist in the market evolution. Assuming , a random walk model, we obtained both the multi-scale *ApEn* and *SampEn* for six different realizations of random numbers and the results are shown in Fig.6. Increasing the number of realizations of random samples



(a) (b)

Fig.5, Multi-scale entropy of the low - pass filtered data for the six selected subperiods, (a) *ApEn*,

(b) *SampEn*.

obviously will provide a lower (L) and upper (U) envelope of various run of the scale dependent curves for *ApEn* and *SampEn*. In case the scale dependent *ApEn* curve (C) for the crude oil price market falls below the lower envelope (L) , the informational market efficiency, $I_{IME}(\tau)$, defined as (cf. Ortiz-Cruz et al., (2012):

$$I_{IME}(\tau) = \left[\frac{ApEn(\tau)}{L(\tau)}\right] \times 100$$

is expected to be less than 100% i.e., partially efficient, whereas 100% informational efficiency occurs if it is bounded by the curves (L) and (U).

4: Concluding remarks

From (a) and (b) Figures 4 and 5 we note that the entropy representing informational efficiency is bounded by limits above and below the value for random walk process in all the time segments. The drop from high entropy value (low informational efficiency) to low values of entropy (higher informational efficiency) in a time segment is an indication of time it takes for information efficiency to get operationalize either by regulations or through wider and rapid diffusion of information amongst the agents. It is also noted that after a 'crisis' the post facto introduction of regulation shows up in a decreasing trend of entropy for some time, the decrease is seen as dependent on the scale of crisis.

Acknowledement: Authors would like to thank Mr.Manis Shailani, IIC, University of Delhi South Campus for assisting us in various computational task.

Reference

Aoyama, H., Fujiwara, Y., Ikeda, Y., Iyetomi,H., Souma,W. and Yoshikawa, H., 2017, “Macro-Econophysics”, Cambridge University Press.

Grassberger,P. and Procaccia ,I., 1983, Phys.Rev.A, **28**,2591.

Gulko Les ,1999, Int. J. Theo. Appl. Fin., **2**, 293.

Ortiz-Cruz , A., Rodriguez, E., Ibarra-Valdez, C. and Alvarez-Ramirez, J., 2012, Energy Policy, **41**, 365.

Pincus, S.M., 1991, PNAS (USA), **88**, 2297.

Richman, J.S. and Moorman,J.R., 2000, **278**, 2039.