World Applied Sciences Journal 12 (6): 835-839, 2011
ISSN 1818-4952

Predictability of Dow Jones Index via Chaotic Symbolic Dynamics

[1]M. Alvo, [2]E. Firuzan and [2]A.R. Firuzan

[1]Deparment of Mathematics and Statistics, Science Faculty,
Ottawa University, Ontario, Canada K1N 6N5
[2]Deparment of Statistics, Science Faculty,
Dokuz Eylul University, 35160 Buca-Izmir, Turkey

Abstract: We define alignment scores, the Hurst exponent and root mean square variation and use them along with Shannon entropy to analyze the Dow Jones index for the years 1985-2010. It is seen that the dynamical behavior of the US stock market is characterized by the temporal variations of the Hurst exponents, the Shannon entropy, the scores and the root mean square variation. We conclude that these measures can be used as precursors to describe the health of the US market.

Key word: Schaotic time series · Symbolic dynamic analysis · Dow Jones index · Hurst exponent · Shannon entropy

INTRODUCTION

Equity markets exhibit a natural daily fluctuation sometimes marked by large surges in both price and volume of trades. An investor is often interested in determining when to buy and when to sell an individual stock. Most securities follow the trend of the major indices such as the DJIA or the NASDAQ which provide information about the overall market trend. However, the level of an index by itself does not provide sufficient information for an investor to make a decision. The other side of the coin is information about the volume of shares trading hands. Exogenous and endogenous impacts or political and economic impacts can have a considerable effect on the dynamics on financial markets.

This article attempts to detect and monitor the chaos and to assess their utility in flagging crisis events. Different complexity measures quantify different kinds of patterns of the given time series. The Hurst Exponent provides a measure which represents the long-range correlation properties of the data and can be viewed as a self-similarity parameter characteristic of fractals.

We represent the original time series as sequence of symbols of a finite alphabet defined by the turning points of a map. Finding the symbolic dynamics of a time series is a mathematical-statistical technique aimed at the detection of the underlying topological and metrical structures in that time series [1]. Information about the periodic orbits of a dynamical system can often be captured in terms of these sequences of symbols. The probabilities of occurrence for different symbol sequences constitute the symbol sequence statistics. In this study, the symbol sequence statistics are used as the target for detecting interesting features in the series.

The concept of entropy and the tools derived from chaotic symbolic dynamics are used together in an attempt to detect and monitor the chaos. All complexity measures based on static entropy (Shannon entropy, pattern entropy, renormalized entropy) quantify statistical order in the time series. At the core of these measures is the probability distribution of the objects defined from the data. Entropy serves as an indicator for the complexity of the underlying processes that gave rise to the variability in the time series data. Entropy, especially Kolmogorov-Sinai entropy, characterizes the degree of randomization of chaotic orbits.

The aim of this article is to examine the variability in a chaotic time series using symbolic dynamic analysis. For this purpose, sequences of symbols are examined using information entropies, alignment scores and fluctuations of alignments. The scores obtained for each sequence as in Yamano *et al*. abnormalities occurring in the series. Scores alone however do not provide sufficient information about abnormalities occurring in the series. Hence, in order to compare pairs of sequences, we perform a detrended fluctuation analysis (DFA) and compute the Shannon entropy for each sequence.

Corresponding Author: Esin Firuzan, Department of Statistics, Science Faculty, Dokuz Eylul University, 35160 Tinaztepe Campus, Buca-Izmir, Turkey, Tel: +2324128557, E-mail: esin.firuzan@deu.edu.tr.

We define alignment scores, the Hurst exponent and root mean square variation and use them along with Shannon entropy to analyze the Dow Jones Industrial Average (DJIA) index for the time period 1985-2010. This time series was chosen in order to illustrate the ability of impact analysis to flag critical economic events. It is seen that the dynamical behavior of the US stock market is characterized by the temporal variations of the Hurst exponents, the Shannon entropy, the scores and the root mean square variation.

In the following section, exogenous and endogenous impacts on DJIA will be introduced. In section 3 gives the review of chaotic symbolic dynamic analysis. In section 4 presents the application of symbolic dynamic analysis to the DJIA and comparatively evaluates the results for different symbolic sequence alignment. Section 5 summarizes and concludes the paper along with recommendations for forecasting.

Critical Events (IMPACTS) on DJIA (1985-2010): Exogenous and endogenous impacts can have a considerable effect on the dynamics on financial markets. An endogenous impact in a financial system can be generated as a result of the behavior of investors. As an example, it may be argued that the 1987-market crash was partially induced by the overreaction of uninformed investors, resulting in panic selling and abrupt, dramatic price slips [2]. The United States (U.S.) savings and loan crisis of the 1980s led to a credit crunch which is seen as a major factor in the U.S. recession of 1990-91 [3]. Early 2000 to 2001 witnessed the striking effect of the dot-com bubble. Excitement over the prospects of the internet has led to huge increases in the Dow Jones index. On January 14, 2000 the Dow Jones peaked at 11722.98. The present on-going financial crisis has been linked to reckless and unsustainable lending practices resulting from the deregulation and securitization of real estate mortgages in the United States [4].

Exogenous shocks, such as the Gulf war had a positive effect on investor dynamics. The Gulf war of 1990 was a unique war in American history in terms of its effect on the budget economy, both because it was fought out of arms inventories and because America's allies paid for the war, at least in the short run [5]. Unlike the Gulf war, the attack on the World Trade Centre in 2001 had a negative effect on investor dynamics [4]. Major stock markets did not open from September 11 to September 17. When the stock markets reopened, the Dow Jones Industrial Average fell 684 points, or 7.1%, to 8921, a record-setting one-day point decline [6]. Although most of the exogenous and endogenous events usually produce a short-term significant effect on financial markets, their impact is diminished within a relatively short time after. For example, after a short time period following the 9/11-event, the financial market recovered and reached 10,200 point by December of that year.

Review of Chaotic Symbolic Dynamic Measures: This section presents a brief introduction to methods for analyzing a chaotic time series in order to extract relevant physical information on the dynamical system that generated the observed data. Chaotic behavior in deterministic dynamical systems is an intrinsically nonlinear phenomenon.

There are several dynamic measures which can be used for identification and classification of chaotic time series. Briefly information about these measures will be introduced in the following subsections. For further information, references to comprehensive presentations of those fields are also included.

Detrended Fluctuation Analysis and Hurst Exponent: Detrended fluctuation analysis quantifies the fractal like correlation properties of the time series and uncovers both short-range and long-range correlations. This approach allows us to quantify the scaling properties in processes of various origins. Numerical analysis of a time series $i = 1,...,N$ within the framework of the detrended fluctuation technique is performed as follows: First, the mean value $\bar{x}$ is removed from each x_i and we calculate the cumulative sum $y_k = \sum_{i=1}^{k}(x_i - \bar{x})$. The y_k's are then regresses against time and the resulting regression yields an estimate $\hat{y}_k$. The root-mean-square fluctuation of this integrated and detrended time series is calculated as

$$RMS(w_S) = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(y_i - \hat{y}_i)^2}$$

This process is repeated for each sliding window. It is expected that the root-mean square fluctuations $RMS\ (w_s)$ exhibit the power-law dependence $RMS\ (w_S) \sim S^H$, where H is called the scaling or Hurst exponent. The latter is a self-affinity parameter representing the long-range correlation properties of the data. We may plot the log of the root-mean square fluctuation against the log of the window size. The slope of this line is then the value of H. An analysis of the local exponents as a function of S may be used to characterize the detailed structure of complex time series.

The scaling exponent has been viewed as a self-similarity parameter related to fractal Brownian motion [7]. The numerical values of H (Hurst Exponent) characterize the persistence behavior: a value of H equal to 0.5 indicates that the sequence is either random or uncorrelated. A value in the range $0.5<H<1$ corresponds to power- law correlations while the range $0 < H < 0.5$ reflects the presence of anti-correlations in the data series.

Shannon Entropy: Shannon entropy quantifies statistical disorder in a time series. At the core of this measure is the probability distribution of the objects defined from the data. Entropy serves as an indicator of the complexity of the underlying processes that gave rise to the variability in the time series data. Shannon entropy is calculated for each window by equation 1.

Let $W_L = \{w_{L,1},\ldots,w_{1,k}\}$ be the set of all distinct sequences of symbols of length L that can be extracted from a window of length S. Let $P_L = \{p_{L,1},\ldots,p_{1,k}\}$ denote the corresponding probabilities of occurrence of each sequence. The Shannon entropy of symbol sequence is defined as follows

$$\mathrm{H}_{S,L} = -\sum_{i=1}^{K} P_{L,i_{w1,w2,\ldots,wK}} \log_2 P_{L,i_{w1,w2,\ldots,wK}} \qquad (1)$$

Sequence Scores: The scores are obtained for each sequence of length S with -1 for each D, +1 for each U and 0 for each F. The scores are summed for each sequence in order to compare them. No account was taken of gaps in the sequence since this does not significantly affect the variation [8]. As an example, the sequence DDUUUFUU is assigned score 3.

Application of Chaotic Symbolic Dynamics to DJIA (1985-2010): Chaotic Symbolic Dynamic analysis applied to the Dow Jones Industrial Average (DJIA) Index. Historical data from January 2, 1985 to February 17, 2010 were extracted from the internet site www.yahoo.com .The data contain opening, highs, lows and closing daily values. In this work, as a representative daily index, the mean between opening and closing values was considered. The data, presented in Figure 1 (a, b), correspond to daily records and its index. The data over two consecutive days are symbolized by using a 3-letter alphabet. D represents the decrease of the index, U represents an increase of index and F represents the stability of the index as "Flat". Yamano *et al*. and Pichl *et al*. [9] compared sequences using pre-defined scores. However, sequence score do not contain sufficient information to take into account sequence fluctuations.

Fig. 1: Daily DJIA

In addition to scores, we compared the sequences using sequence statistics, Hurst exponent, RMS and Shannon entropy. As an added bonus, the tools developed in this article were used to flag some abnormalities in the sequences from the time window.

The motivation for the application of Chaotic Symbolic Dynamics is the irregular variation of DJIA since 1985, as depicted Figure 1. This is formally verified by calculation of the autocorrelation and partial autocorrelation functions. Hence, the series cannot be modeled by the usual calculation of the linear time series models.

DJIA index is examined by daily records which are defined as $I_i = x_i / x_{i-1}$ with I_i being the index at time i and defined 3 symbols. The use of the index adjusts for the changes in magnitude of the DJIA over time.

$$Sym(x_i) = \begin{cases} D & I_i < 0.995 \\ F & 0.995 \le I_i \le 1.005 \\ U & I_i > 1.005 \end{cases} \qquad (2)$$

In this work, root mean square, Shannon entropy and score have been employed as additional criteria for the selection of flags for detecting a crisis. These measures were computed for 100 business day window size and 5 business day alphabet size. Figure 2 presents the Dow Jones Index, Shannon entropy, the fluctuations of RMS, scores and Hurst Exponent, respectively. An examination reveals that these measures fluctuate around a benchmark over time. We have as well superimposed the occurrence of critical events.

According to Figure 2, for 1987-crisis period the slope declined from about 0.60 to 0.10 and for the period after 1987 the slope is 0.55. This indicates that the long term trend of HE is increasing after 1987, suggesting that the 1987-crisis introduced a delay in the stabilization of the market functioning around conditions close to efficient operation. The HE has decreased to achieve values smaller than 0.5 during the 2005 time period.

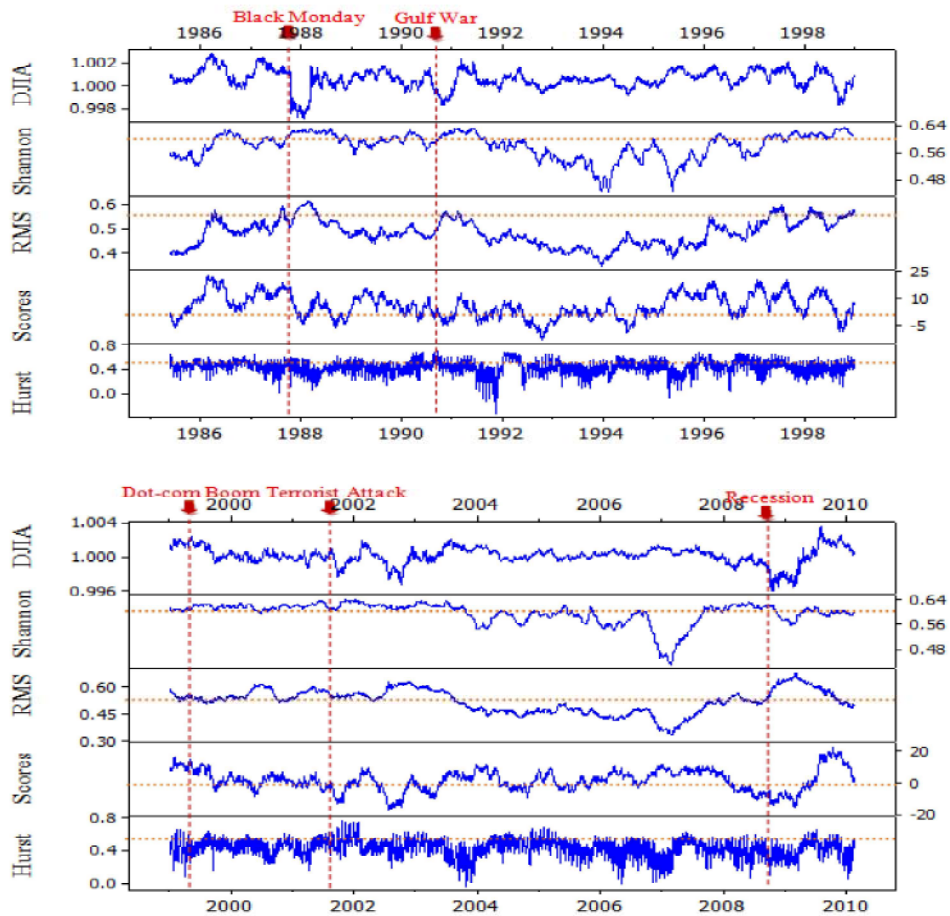

Fig. 2: Fluctuations of the DJ index, the Shannon entropy, the RMS, scores and HE

This behavior has not been previously observed in the US market. The shift of the HE from persistence to anti-persistence values in the beginning of the 2005 till the end of the 2007 seems to be linked to the overreaction of stock markets.

CONCLUSION

Accounting for risk and uncertainty are essential to forecasting and prediction. In time series forecasting, the first issue to investigate is whether or not the time series under study is predictable. This question is answered by referring to Hurst exponents.

We observe that the Hurst exponent (HE) lies mostly below 0.5 pointing to anti-persistence behavior for the index as shown in Figure 2. This means that drops in the index may be followed by rises. When HE exceeds 0.5, we have persistence; pointing to the existence of trends. The HE characterizes cyclic behavior around the linear trend.

Historical events can be identified with local maximum and minimum values of the HE. For example, the 1987-crisis induced a local minimum corresponding to less predictable price variations. On the other hand, the 9/11 events correspond to a local maximum.

Score values below 0 appear to point to crisis events whereas values exceeding 0 correspond to better economic conditions. The score is more indicative of a trend.

We now turn our attention to the RMS which is sensitive to the variation in the index. It can be seen that the benchmark for the RMS is 0.55 above which we witness the occurrence of Black Monday in 1987, the dot-com crisis in 2000, the period preceding the terrorist attack in 2001 and finally the recession of 2009. The same behavior occurs for the Shannon entropy but with a benchmark value of 0.60.

Both scores and the HE series exhibit decreasing trends over the 1987-crash time period. HE appears to provide a warning on 08/1987 preceding the 1987 crash. Following the Gulf war, the RMS and Shannon entropy show peaks in 08/1990 while the scores and HE exhibit a decrease in a linear fashion. Shortly after the dot-com crisis and terrorist attack, the economy had recovered and their effect was short-lived as presented in Figure 2 for all the measures.

The present recession appears to have had its beginning in 2008 as can be seen from the rise in RMS and Shannon entropy and in the decrease in score and HE. The recession has indeed persisted. As shown in Figure 2, HE has its lowest value four months before and Shannon entropy points out also alarm point one month before. Hence it appears that the measures are good indicators of an economic crisis. At present it appears that the Dow Jones index exhibits good stability with values of entropy and RMS below the threshold. The score value is at 0 while the Hurst index is less than 0.5.

The dynamic measures or tools were then applied to study the behavior of the DJIA which is seen to show irregular behavior resulting from nonlinear chaotic dynamics rather than from random external forces. It appears that at the present the Dow Jones index exhibits good stability with values of entropy and RMS below the threshold. The score value is at 0 while the Hurst index is less than 0.5. Some benchmarks are obtained for different chaotic dynamics. Obtained these benchmarks makes it more easy for predictability of Dow Jones Index.

REFERENCES

1. Mendesa, D., V.M. Mendesb, N. Ferreirac and R. Menezes, 2009. Symbolic Shadowing and the Computation of Entropy for Observed Time Series.
2. Stewart, J. and D. Hertzbeg 1987. How the stock market almost disintegrated a day after the crash. The Wall Street J.,
3. Fratianni, M. and F. Marchionne, 2009. The role of banks in the subprime financial crisis. Review of Economic Conditions in Italy, 2009/1: 11-48.
4. Mishkin, F., 2008. How Should We Respond to Asset Price Bubbles?. Speech at the Wharton Financial Institutions Center and Oliver Wyman Institute's Annual Financial Risk Roundtable, Philadelphia, May 15, 2008 http://www. federalreserve. gov/newsevents/ speech/mishkin 20080515a.htm
5. Silk, L., 1991. Economic Scene; The broad impact of the Gulf War. A version of this article appeared in print on August 16, 1991, New York Times http://www. nytimes. com/ 1991/08/16/ business/ economic-scene-the-broad-impact-of-the-gulf-war.html?pagewanted=1
6. Barnhart, B., 2001. Markets reopen, plunge. Chicago Tribune. http://www. chicagotribune. com/business/ chi-010917 markets,1,2882341.story
7. Mandelbrot, B.B. and J.R. Wallis, 1969. Some long-run properties of geophysical records. Water Resources Res., 5: 321.
8. Yamano, T., K. Sato, T. Kaizoji, J. Rost and L. Pichl, 2008. Symbolic analysis of indicator time series by quantitative sequence alignment. Computational Statistics and Data Analysis, 53: 486-495.
9. Pichl, L., T.K. Yamano and T. Kaizoji, 2006. On the symbolic analysis of market indicators with the dynamic programming approach. Lecture Notes in Computer Sci., 3973: 432.

00. Makinen, G., 2002. The Economic Effects of 9/11: A Retrospective Assessment. Congressional Research Service. pp. CRS-4. http://www. fas. org/irp/crs/ RL31617. pdf.