Theoretical and Applied Economics
Volume XXI (2014), No. 8(597), pp. 19-26

Wavelet based sample entropy analysis: A new method to test weak form market efficiency

Anoop S. KUMAR
University of Hyderabad, Hyderabad, India
taichichuaan@gmail.com
B. KAMAIAH
University of Hyderabad, Hyderabad, India
kamaiahbandi@gmail.com

Abstract. *In this article, we analyze informational efficiency in daily returns of NASDAQ, DJIA and S&P 500 indices ranging from 04-01-1980 to 12-09-2013.We replace the traditional coarse graining method used in multi-scale entropy analysis by a Maximal Overlap Discreet Wavelet Transform decomposition and extract Sample entropy measure across different timescales. To compare against efficient market behavior, we simulate an i.i.d. normal series with the same mean and variance of the underlying series and repeat the procedure. Next, we plot both of these estimates to see how the values differ from each other across the scales. It is found that the three markets under study are not weak form efficient at high to medium frequencies (up to semi-annual period). They are informationally efficient in the long run (annual-biannual period). Here, efficiency of a financial market is closely related with the time horizons under which the agents operate. As time horizon increases, the markets move towards an informationally efficient state. It could be due to the fact that agents with a long investment horizon make use of the information set available in a comparatively efficient manner due to their comparatively high tolerance for price fluctuations as opposed to their high-frequency counterparts.*

Keywords: EMH, efficiency, transfer entropy, wavelets, equity markets.

JEL Classification: G01, G14, G15.
REL Classification: 11B.

1. Introduction

Efficient market hypothesis in its weak form (EMH henceforth), proposed by (Fama, 1965) makes the assumption that all past information related to an asset is reflected in its price. It talks about of investors who are rational in behavior, taking decisions alone and able to process all the information available to them instantly. Here, an implicit assumption is that all investors operate under same time horizon and use the same information set. In such a scenario, it is not possible to have supernormal returns in the long run. In other words, returns follow a normal distribution. If this hypothesis is to hold, there are no chances of extreme fluctuations in a market such as an asset bubble or a financial crisis. The notion of efficient markets has been refuted by many empirical evidences over the years as well as the major financial crises occurred during the past years, the global financial crisis of 2007 being the most recent one. Still, EMH considered as one of the bedrocks of main stream finance theory.

In a market, there are different type of investors with different time horizons and risk preferences. A good example will be an intra-day trader (chartist) and a mutual fund manager. While a chartist is mostly concerned with the trends in the data, the mutual fund manger would be looking into the information s/he has about the performance of the economy. In other words, even though both may have access to the complete information set, they consider only parts of it that they feel relevant to them.

It could be assumed that their risk preference will be also difference. An investor with a long investment time horizon would treat the variance in an asset price differently from a chartist. If we take standard deviation as a measure of risk, the σ_m of the mutual fund manager calculated based on his/her transactions will be higher than the same of a chartist (σ_c) for the same asset under normal circumstances as the former trades keeping a longer time-frame in mind and therefore accept larger variations in asset prices compared to the chartist. Each of them would have a threshold level beyond which they will not hold the asset.

Consider the following scenario: a chartist decides to sell of an asset where the σ_m crosses the ‘acceptable threshold’. At the same time, the mutual fund manager could purchase this asset if s/he is interested because of the longer investment horizon and thereby the corresponding larger value of σ. Here, the selling signal made by the chartist is considered to be a buying signal by the mutual fund manager. A volume of such transactions keep the market stable. However, we cannot be sure if the market is efficient under such circumstances.

There are many econometric/statistical tests such as variance ratio tests by (Lo and McKinlay, 1988), (Chow and Denning, 1993), (Chen and Deo, 2005) as well as unit root tests formulated by (Phillips and Perron, 1998), (Kwiatkowski et al., 1992) etc.

to test weak form market efficiency. However, it is to be noted that these tests deals with the level data and thereby implicitly accepting the assumption that all traders operate under the same time horizon and with the same information set.

Here we propose a new method to test weak form market efficiency with the help of information theory. Information theory could be traced back to the seminal work of (Shannon, 1948) where he introduced the concept of Shannon entropy to extract the information content of a time series. (Pincus et al., 1991) bought the concept of Approximate Entropy. (Richman and Moorman, 2000) introduced the concept of sample entropy, an improvement over Approximate Entropy. Sample entropy in its initial form also dealt with information content of a time series at its level. (Costa et al., 2002) introduced the idea of multiscale entropy analysis where he implemented a coarse graining procedure to create different realizations of the same time series over a number of scales and extract the sample entropy for all the realizations. Here, we propose an alternative to the coarse graining method to suit our purposes. The details are given in the following section.

2. Data and Methodology

We have use log returns calculated from the daily closing data for 3 major US equity indices namely NASDAQ, S&P 500 and DJIA from 04-Jan-1980 to 12-Sep-2013 for the analytical purpose.

Multiscale entropy (MSE) analysis is a relatively new method of measuring the complexity of finite length time series. This method was developed by (Costa et al., 2002). Here entropy was estimated using the sample entropy (SampEn) measure, developed by (Richman and Moorman, 2002). SampEn is a refinement of the approximate entropy (ApEn) family of statistics. SampEn is preferred because it avoids self counting and it is suitable for a short-lengthed sample.

Traditional entropy measures quantify only the regularity (predictability) of time series on a single scale. However, a single value cannot explain the underlying dynamics of the given data. The MSE method tries to address this problem. The estimation incorporates two steps:

1. A "coarse-graining" process is applied to the time series. For a given time series, multiple coarse-grained time series are constructed by averaging the data points within non-overlapping windows of increasing length, τ. Each element of the coarse-grained time series y_j^{τ}, is calculated according to the equation $y_j{}^{\tau} = \frac{1}{T}\sum_{i=(j-1)\tau+1}^{j\tau} x_i$.

Where τ represents the scale factor and $1 \leq j \leq N/\tau$. The length of each coarse-grained time series is N/τ. For scale 1, the coarse-grained time series is simply the original time series.

2. SampEn is calculated for each coarse-grained time series, and then plotted as a function of the scale factor. SampEn is a regularity statistic. It looks for patterns in a time series and quantifies its degree of predictability or regularity.

For a time series with enough data points, application of the coarse graining procedure will be of no issue. However if we are a shorter time series like IIP, the estimation of SampEn after a particular scale will not be possible. Our method is implemented keeping this issue in mind.

Here, we propose to use a MODWT decomposition instead of coarse graining.

The MODWT (Percival and Walden, 2000) is similar to the discrete wavelet transform (DWT) in that the high-pass and low-pass filters are applied to the input signal at each level. However, in the MODWT, the output signal is never sub-sampled (not decimated). Instead, the filters are up sampled at each level. Suppose we are given a signal $s[n]$ of length N where $N = 2^J$ for some integer J. Let $h_1[n]$ and $g_1[n]$ be the low-pass filter and the high-pass filter defined by an orthogonal wavelet. At the first level of MODWT, the input signal $s[n]$ is convolved with $h_1[n]$ to obtain the approximation coefficients $a_1[n]$, and with $g_1[n]$ to obtain the detail coefficients $d_1[n]$:

$$a_1[n] = h_1[n] \text{ x } s[n] = \sum_k h_1[n-k]s[k] \quad (1)$$

$$d_1[n] = g_1[n] \text{ x } s[n] = \sum_k g_1[n-k]s[k] \quad (2)$$

Without sub-sampling, $a_1[n]$ and $d_1[n]$ are of length N instead of N/2 as in the DWT. At the next level of the MODWT, $a_1[n]$ is filtered using the same scheme, but with modified filters $h_2[n]$ and $g_2[n]$ obtained by dyadic up-sampling $h_1[n]$ and $g_1[n]$. This process is continued recursively. For J=1, 2, ..., $J_0 - 1$,where $J_0 \leq J$, define

$$a_{j+1}[n] = h_{j+1}[n] \text{ x } a_j[n] = \sum_k h_{j+1}[n-k]a_j[k] \quad (3)$$

$$d_{j+1}[n] = g_{j+1}[n] \text{ x } a_j[n] = \sum_k g_{j+1}[n-k]a_j[k] \quad (4)$$

Where $h_{j+1}[n] = U(h_j[n])$ and $g_{j+1}[n] = U(g_j[n])$. Here U is the up-sampling operator that inserts a zero between every adjacent pair of elements of the time series. The output of the MODWT is the sum of the detailed coefficients $d_1[n]$, $d_2[n]$, $d_3[n]$.....$d_{J0}[n]$ and the approximate coefficients $a_{J0}[n]$.

The wavelet decomposition makes sense because here we can decompose the data into different timescales and thereby effectively isolates the information associated with all timescales. Also, here all the decomposed time series have same length where in the original procedure; hence we could be sure about get robust estimates of SampEn.

Here in the first step as explained before, we decompose the given time series into J levels using a MODWT. Next we compute transfer entropy for all the decomposed wavelet coefficients.

The next step involves simulating an efficient behavior against which we can test the measures obtained in the first step. If a market is efficient, the prices are said to follow a geometrical Brownian motion, or random walk. In other words, there is no temporal dependence structure among the time series under analysis. Moreover, the returns follow normal distribution.

Therefore in order to verify EMH, we simulate an i.i.d. normal series with same mean and standard deviation of the analyzed dataset and apply the two steps explained above. Then we plot the values of entropy at different scales for both series. If the market is efficient, the entropy curve of the original should coincide with the entropy curve of the simulated series. We have used an LA (least asymmetric) wavelet filter with a length of 8 for the wavelet decomposition.

3. Analysis

Figures 1 to 3 represent entropy curves for DJIA, NASDAQ and S&P 500. The X axis depicts the wavelet scales while Y axis shows the SampEn values.

Figure 1. *Entropy curve for DJIA*

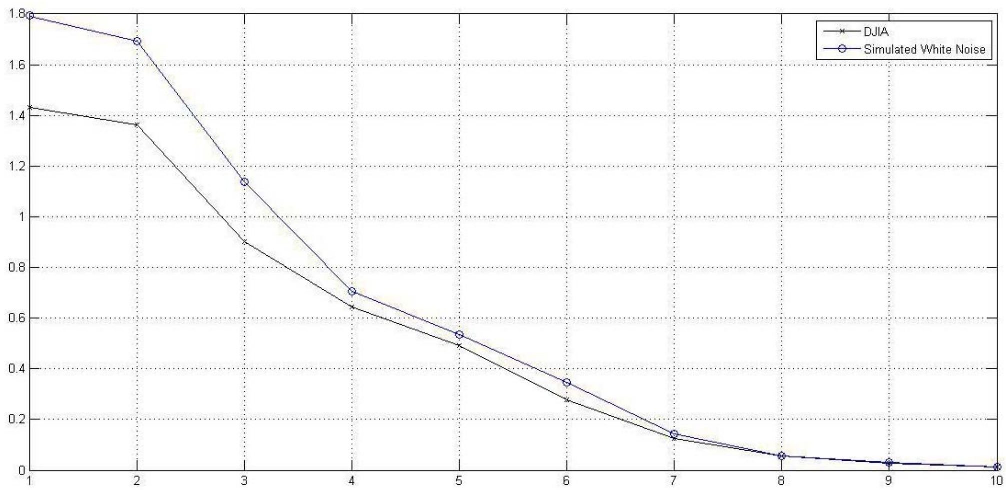


Figure 2. *Entropy curve for NASDAQ*

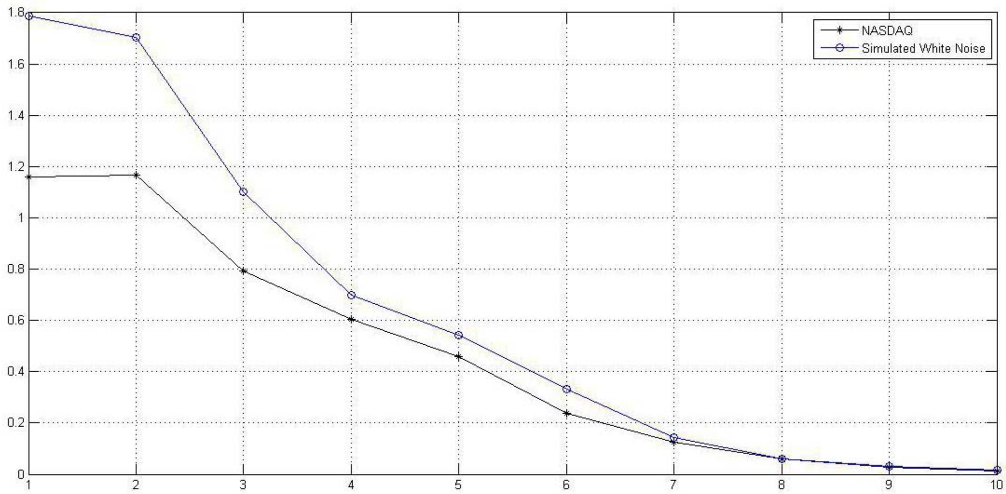
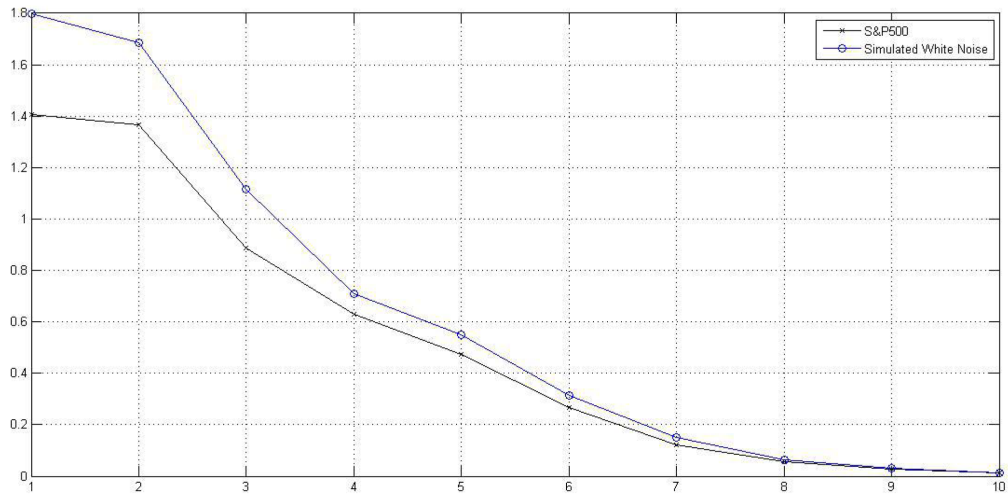


Figure 3. *Entropy curve for S &P 500*



From the graphs, certain common features are visible. First, the corresponding i.i.d. normal series have a greater value of sample entropy compared to its counterpart at short frequencies. As we move towards coarser scales, (low frequency), this divergence becomes small, and the curves coincide around scale 8 for all the series. Scale 8 means the timescale of 256-512 trading days, or annual-biannual period.

Here, the market does not reflect the informationally efficient behavior in the finer scales i.e. high to medium frequency areas. The actual market behavior converges towards the efficient behavior as the timescale increases. From scale 8 onwards, the curves converge; indicating informational efficiency in the long run.

As bulk of the transactions in an equity market happen within the high to medium frequency timescales; it could be safely assumed that the market behavior will be influenced by activities taking place across those timescales. Hence, the informational efficiency at the long run might not have any effect on turbulences that can happen in an equity market. If it had any effect at all, the markets should be free from the extreme events such as financial crash of 1987, the dot-com bubble and 2008 meltdown that occurred during the period of analysis.

Conclusion

In this article, we tried to analyze weak form market efficiency using a new method, namely wavelet based sample entropy analysis. We worked under the assumption that a market consists of different types of traders with different risk preferences and investment time horizon. Due to these factors, while it is possible that the market may remain stable, it need not be true that it is informationally efficient. We have used daily log returns of three major US equity indices namely DJIA, NASDAQ and S&P 500 from the period 03-01-1980 to 13-09-2013 for the analysis. In the first stage, we decomposed the give data series into different timescales using a MODWT procedure. MODWT was preferred over the traditionally used coarse graining method in order to avoid down-sampling. To see the degree of efficiency, an i.i.d. normal series with the same mean and standard deviation of the underlying indice were simulated and entropy measures were calculated after decomposing it using the wavelet procedure. From the entropy curves, it was visible that the actual market behavior diverges from the expected efficient market behavior at high to medium frequencies. At low frequencies (annual-biannual timescales), the market becomes informationally efficient. However, this efficiency does not prevent the markets from undergoing extreme events like financial crises.

References

Chen, W., Deo, R. (2005). "The Variance Ratio Statistic at large Horizons", *Econometrics,* 0501003 EconWPA

Chow, K.V., Denning, K.C. (1993). "A simple multiple variance ratio test", *Journal of Econometrics,* 58(3), pp. 385-401

Costa, M., Goldberger, A.L., Peng, C.K. (2002). "Multiscale Entropy Analysis of Complex Physiologic Time Series", *Physical Review Letters,* 89(6): 068102-1-4

Fama, E. (1965). "The Behavior of Stock Market Prices", *Journal of Business*, 38, pp. 34-105

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", *Journal of Econometrics,* 54(3), pp. 159-178

Lo, A.W., McKinlay, A.C. (1988). "Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies,* 1, pp. 41-66

Percival, D., Walden, A. (2000), *Wavelet Methods for Time Series Analysis,* Cambridge University Press

Pincus, S., Gladstone, I., Ehrenkranz, R. (1991). "A regularity statistic for medical data analysis", *Journal of Clinical Monitoring*, 7 (4), pp. 335-345

Richman, J.S., Moorman, J.R. (2000). "Physiological time-series analysis using approximate entropy and sample entropy", *American Journal of Physiology*, Heart and Circulatory Physiology, 278: H2039-H2049, 2000

Shannon, C.E. (1948). "A Mathematical Theory of Communication", *Bell System Technical Journal*, 27, pp. 379-423,623-656