Entropy analysis of energy price movement

Jacek Kapica

University of Life Sciences in Lublin, Doświadczalna 50A, 20-280 Lublin; email: jacek.kapica@up.lublin.pl

Summary. The paper presents a novel method of analysis of energy price movement with use of sample entropy, which is used to measure the complexity or information content of a given data set (time series). This technique is used to verify the efficient market hypothesis. The results show that although the price movement is nearly random, there is a statistically significant difference from a completely random walk.

Key words: energy price, efficient market hypothesis, sample entropy.

INTRODUCTION

Prices of fossil energy sources are crucial for the energy industry and the economy of countries, regions and the world. The crude oil cost increase observed over the last few years has been a challenge not only to individual consumers but even more, to companies whose economy is highly dependent on the cost of energy. In the crisis conditions there is a need to find new approaches towards increasing the efficiency of finance management [12]. The energy cost has a great impact on people life quality: International Monetary Fund estimates that a crude oil price growth of 5 USD per barrel results in economy development rate decrease by 0.3 % [16]. Increasing fuel costs result in increased production costs in the whole industry, especially in agriculture [17].

Another group of people interested in crude oil price behavior are investors, who try to profit from price movement. Recently, many internet-based trading platforms have been made available to individual investors, who can implement their trading strategies easily. For all the groups: individual investors, institutional investors and company management, to be able to forecast the future cost of energy means profit. With this in mind, the author would like to look at the efficient market hypothesis, which in general states, that the price of a given asset responds quickly and accurately to relevant information, so that the future price is unpredictable [3].

The classical definition of market efficiency from 1976 by Jensen is as follows: *A market is efficient with respect to information set Ω_t if it is impossible to make economic profits by trading on the basis of information set Ω_t.* [11] quoted after [19].

Many authors try to address the question of market efficiency with unconclusive results [6]. In this paper the author employs sample entropy method to verify the hypothesis.

SAMPLE ENTROPY

Entropy is a way to measure complexity or dynamics of a given time series. Shannon entropy is classically defined as [5,4]:

$$H(X) = -\sum_{x_i \in \Theta} p(x_i) \log p(x_i) = -E[\log p(x_i)], \quad (1)$$

where: X represents a random variable with a set of values Θ and $p(x_i)$ is a probability that X will be equal x_i. For a time series representing output of a stochastic process, joint entropy is calculated. A major disadvantage of this definition of entropy is that its value strongly depends on the length of the time series.

In 2000, Richman and Moorman proposed a new algorithm for the calculation of entropy, called sample entropy (S_E) [15], which is a function of the time series $X=\{x_1, x_2, ..., x_N\}$, template length m and tolerance r. In order to calculate S_E, the algorithm finds the first m-length sequence (template) of data points and then looks for matching pattern through the rest of the series. The pattern is considered as matching if the consequent data points are within a distance of r to the corresponding points of the template. The number of matches is counted as n^m. Then the same procedure is performed for the m+1 - length sequence and the number of matches counted as n^{m+1}. The whole process is repeated until the end of the time series is reached. In this way S_E is defined as:

$$S_E(m,r,N) = \ln \frac{\sum_{i=1}^{N-m} n'^{m}_{i}}{\sum_{i=1}^{N-m} n'^{m+1}_{i}}. \tag{2}$$

The difference between n^m and n'^m is such that in n'^m self-matches are not counted. For an illustration of how the S_E is calculated, the reader is referred to reference [5].

S_E is precisely equal to the negative of the natural logarithm of the conditional probability that sequences close to each other for m consecutive data points will also be close to each other when one more point is added to each sequence [5]. When entropy is calculated in the previously stated method, it is less dependent on the length of the time series.

Sample entropy has been widely used in analysis of bioelectrical signals, mainly ECG and EEG [2,18,13,2]. Recently some attempts have been made to employ this technique in analysis of various markets (stocks and commodities) [1,14,8]. This paper aims to use Sample entropy to verify the Efficient Market Hypothesis.

DATA, METHODS AND TOOLS

The data that was used for the analysis was daily spot price of brent type crude oil for the years 1988 - 2011 (6091 points). The data is freely available from the Internet [20].

There is no consensus on how the prices of energy sources should be modeled [9]. In the presented study the author decided to take subsets of the original data points in two ways. The first method is to simply divide the data into periods of calendar years (approximately 255 points, depending of the year) and then calculate sample entropy for each period as described below. The second is to create subsets consisting of z data points continually for all points of the main dataset starting at point z until the end of the set according to the following expression:

$$Y_i = \{x_{i-z}, x_{i-z+1}, ..., x_i\}, \tag{3}$$

in which Y_i is the subset calculated for the i-th point of input data. It is clear from looking at the equation that the smallest i for which the subset Y_i can be created equals z+1. In analogy to a moving average indicator used widely by technical traders, S_E calculated for subsets Y_i will be called “moving entropy” in this paper. The moving entropy was obtained for z=200.

Once the subsets are obtained, the data must be prepared for entropy calculation. As, according to the description presented in the previous subsection, the entropy algorithm looks for patterns and the price has various ranges, the value will be recalculated to obtain the day-to-day price difference Z_j using the equation (4):

$$Z_j = Y_j - Y_{j-1}. \tag{4}$$

Then the data must be normalized so that the mean value of the subset is equal to 0 and standard deviation is 1. This can be illustrated by equation (5):

$$v_{si} = \frac{v_i - \mu(v)}{\sigma(v)}, \tag{5}$$

where:
v_i is the price difference at point i,
v_{si} is the normalized value of v at point i,
$\mu(v)$ is arithmetic mean of all points of v,
$\sigma(v)$ is standard deviation of all points of v.

For each of the subsets S_E was calculated with typical parameters used widely: template length m=2 and tolerance r=0.15. The scripts used were based on the examples found on the Physionet website [10] and executed in the Octave programming environment [7].

In order to verify the Efficient Market Hypothesis, sets of random data points with normal distribution were generated. Its length depended on the subset length and for the yearly entropy comparison it consisted of 248 samples (lowest number of data points in the years analyzed) and for the moving entropy - 200. The number of test sets was equal to number of subsets of the original data according to each method: 24 for the 24 years used for yearly analysis and 5891 subsets for the moving entropy comparison. Additionally, to see some statistical properties (especially possibly maximum and minimum values) of S_E calculated for normal distribution 50 000 data sets of lengths respectively 248 and 200 samples were generated. Than, a statistical t-test was performed to verify whether the difference between entropies calculated for price data subsets and random sets is statistically significant.

RESULTS

Table 1 presents the S_E values for daily crude oil price in the years 1988-2011, table 2 summarizes its statistical properties and table 3 shows statistical properties of S_E calculated for 50000 samples calculated for data set of 248 points with normal distribution.

The p-value for statistical two-tailed t-test performed for 24 S_E values obtained for yearly oil prices and 24 sets of 248 samples of random numbers with normal distribution is equal to 0.018 which proves that there is statistically significant difference between the groups.

Figure 1. presents daily brent price for the years 1988 - 2011 along with moving sample entropy. Because the first 200 points are needed to calculate the first value of the indicator, there first value is presented for point 201 on the chart.

The p-value for statistical two-tailed t-test performed for 5891 subsets of 200 points of daily oil price and 5891

Table 1. Sample Entropy values for crude oil prices for the years 1988–2011.

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
S_E	1.990	2.451	1.604	1.545	2.329	2.475	2.533	2.356	2.569	2.327	2.345	2.456
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
S_E	2.322	2.350	2.547	2.448	2.425	2.467	2.483	2.465	2.374	2.451	2.405	2.321

Table 2. Main statistical properties of SE calculated for crude oil prices for the years 1988 - 2011.

Maximum	Minimum	Arithmetic mean
2.569	1.545	2.335

Table 3. Main statistical properties of SE calculated for 50 000 sets of 248 normally distributed points.

Maximum	Minimum	Arithmetic mean
2.857	2.138	2.485

sets of 200 points of normally distributed random values is well below 0.0001.

CONCLUSIONS

At the first look, the results of the analysis would support the Efficient Market Hypothesis: in most cases the values of the sample entropy calculated for lie within the range of SE values for random data sets. This would suggest that the price movements are random and it is not possible to forecast its future direction.

However if we take into consideration the results of the t-tests, it is clear that the behavior of the price is not random. The p-value for the yearly entropies for oil price and random data sets groups reflects the fact that the probability of the null hypothesis (i.e. that the two groups are not statistically different) is 0.018 which is very low. Traditionally, a level of 0.05 is considered as proving statistical significance. The value p-value for the moving sample entropy/sample entropy for random data sets is even lower.

The results show that the sample entropy value for crude oil price is lower than for randomly generated data with normal distribution, which means that the price change are more predictable than a purely random walk.

REFERENCES

1. **Alvarez-Ramirez J., Rodriguez E., AlvarezJ. 2012:** A multiscale entropy approach for market efficiency, International Review of Financial Analysis, Volume 21, Pages 64-69, ISSN 1057-5219, 10.1016/j.irfa.2011.12.001.
2. **Bosl W., Tierney A., Tager-Flusberg H., Nelson C. 2011:** EEG complexity as a biomarker for autism spectrum disorder risk. BMC Medicine 9:18.
3. **Chien-Chiang Lee, Jun-De Lee, 2009:** Energy prices, multiple structural breaks, and efficient market hypoth-

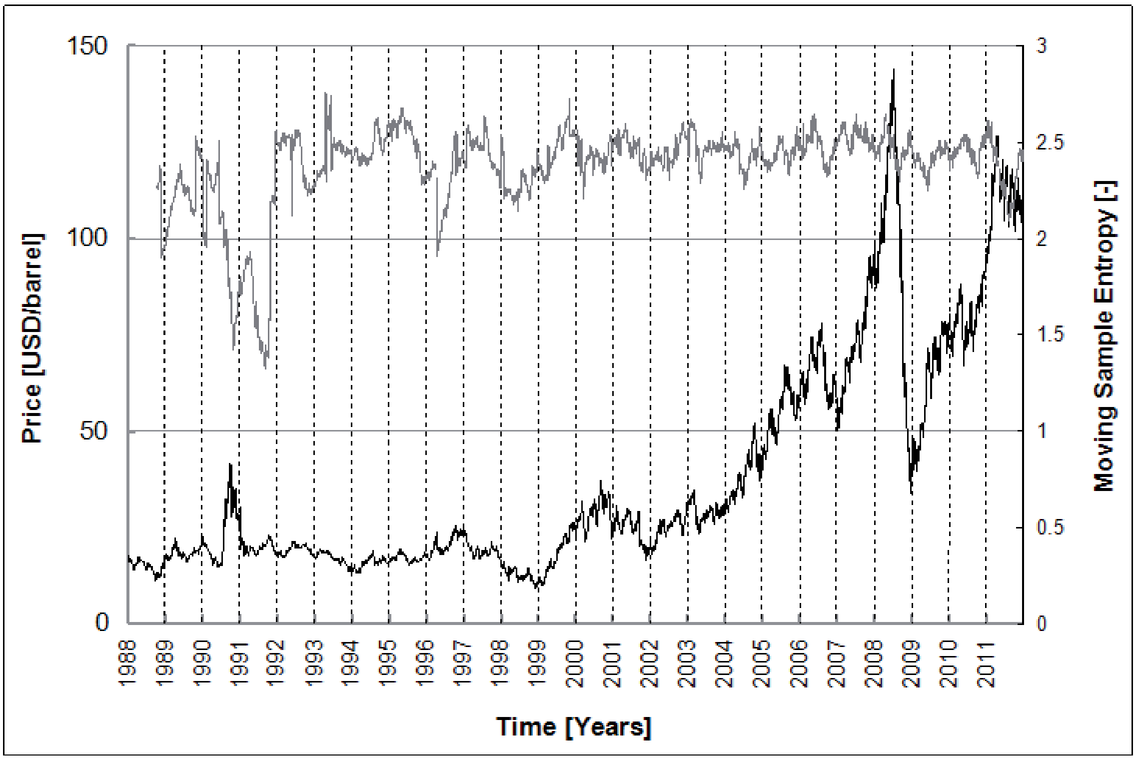

Fig. 1. Daily brent price (black) and moving sample entropy indicator (grey) for years 1988-2011.

esis, Applied Energy, Volume 86, Issue 4, Pages 466-479, ISSN 0306-2619, 10.1016/j.apenergy.2008.10.006.
4. **Costa M., Goldberger A. L., Peng C.-K. 2002:** Multiscale Entropy Analysis of Complex Physiologic Time Series. Physical Review Letters, Vol 89 number 6.
5. **Costa M., Goldberger A.L., Peng C.-K. 2005:** Multiscale entropy analysis of biological signals. PHYSICAL REVIEW.
6. **Doran J.S., Peterson D.R., Wright C., 2010:** Confidence, opinions of market efficiency, and investment behavior of finance professors, Journal of Financial Markets, Volume 13, Issue 1, Pages 174-195, ISSN 1386-4181, 10.1016/j.finmar.2009.09.002.
7. **Eaton J.W. 2000:** GNU Octave Manual. Network Theory Limited, ISBN: 0-9541617-2-6.
8. **Gabjin Oh, Seunghwan Kim, Cheoljun Eom. 2007:** Market efficiency in foreign exchange markets, Physica A: Statistical Mechanics and its Applications, Volume 382, Issue 1, 1 August 2007, Pages 209-212, ISSN 0378-4371, 10.1016/j.physa.2007.02.032.
9. **Ghoshray T., Johnson B. 2010:** Trends in world energy prices, Energy Economics, Volume 32, Issue 5, Pages 1147-1156, ISSN 0140-9883.
10. **Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. 2000:** PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220. PMID: 10851218; doi: 10.1161/01.CIR.101.23.e215.
11. **Jensen M. 1976:** Some anomalous evidence regarding market efficiency. Journal of Financial Economics, 6, 95-10.
12. **Kostyrko L., 2010:** The mechanism of the stable development financial strategy of an entrprise: methodology and organization. TEKA Kom. Mot. i Energ. Roln. – OL PAN, 10A, pp. 303-310.
13. **Mizuno T., Takahashi T., Cho R. Y., Kikuchi M., Murata T., Takahashi K., Wada Y. 2010:** Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy. Clinical Neurophysiology 121.
14. **Ortiz-Cruz A., Rodriguez E., Ibarra-Valdez C., Alvarez-Ramirez J. 2012:** Efficiency of crude oil markets: Evidences from informational entropy analysis, Energy Policy, Volume 41, February 2012, Pages 365-373, ISSN 0301-4215, 10.1016/j.enpol.2011.10.057.
15. **Richman J. S., Moorman J. R. 2000:** Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 278: H2039–H2049.
16. **Roszkowski A. 2006:** Agriculture and fuels of the future. TEKA Kom. Mot. i Energ. Roln. – OL PAN, 6, pp. 131-134.
17. **Shebanin V., Pastushenko S., Havrysh V. 2006:** Prospect od gaseous fuel application in plant-growing in Ukraine. TEKA Kom. Mot. Energ. Roln., 2006, 6A, pp. 156-162.
18. **Takahashi T. Cho R. Y., Murata T., Mizuno T., Kikuchi M., Mizukami K., Kosaka H., Takahashi K., Wada Y. 2009:** Age-related variation in EEG complexity to photic stimulation: A multiscale entropy analysis. Clinical Neurophysiology 120. 476-483.
19. **Timmermann A., Granger C.W.J. 2004:** Efficient market hypothesis and forecasting, International Journal of Forecasting, Volume 20, Issue 1, pp. 15-27, ISSN 0169-2070, 10.1016/S0169-2070(03)00012-8.
20. US Energy Information Administration. 2012. Internet: http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=D, accessed on March 7th, 2012.

ANALIZA ENTROPII ZMIENNOŚCI CEN ENERGII

Streszczenie. Artykuł przedstawia nowoczesną metodę analizy zmienności cen energii przy wykorzystaniu sample entropy, która to wielkość jest używana do określania złożoności oraz zawartości informacji w danym zbiorze danych (szeregu czasowym). Technika ta jest wykorzystana do weryfikacji hipotezy efektywności rynku. Wyniki wskazują, że pomimo że ruchy cen są niemalże losowe, istnieje statystycznie istotna różnica w porównaniu z czysto losowym ruchem cen.

Słowa kluczowe: cena energii, hipoteza efektywności rynku, sample entropy.