Physica A 450 (2016) 104–114

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A comprehensive segmentation analysis of crude oil market based on time irreversibility

Jianan Xia *, Pengjian Shang, Dan Lu, Yi Yin

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China

HIGHLIGHTS

- Perform a comprehensive entropic segmentation analysis of crude oil future prices.
- Apply time irreversibility analysis of each segment to divide all segments into two groups.
- Rearrange the segments by combining with time irreversibility analysis.

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 11 November 2015
Available online 14 January 2016

Keywords:
Segmentation
Time irreversibility
Complexity
Crude oil market

ABSTRACT

In this paper, we perform a comprehensive entropic segmentation analysis of crude oil future prices from 1983 to 2014 which used the Jensen–Shannon divergence as the statistical distance between segments, and analyze the results from original series $\mathcal{S}$ and series begin at 1986 (marked as $\mathcal{S}^*$) to find common segments which have same boundaries. Then we apply time irreversibility analysis of each segment to divide all segments into two groups according to their asymmetry degree. Based on the temporal distribution of the common segments and high asymmetry segments, we figure out that these two types of segments appear alternately and do not overlap basically in daily group, while the common portions are also high asymmetry segments in weekly group. In addition, the temporal distribution of the common segments is fairly close to the time of crises, wars or other events, because the hit from severe events to oil price makes these common segments quite different from their adjacent segments. The common segments can be confirmed in daily group series, or weekly group series due to the large divergence between common segments and their neighbors. While the identification of high asymmetry segments is helpful to know the segments which are not affected badly by the events and can recover to steady states automatically. Finally, we rearrange the segments by merging the connected common segments or high asymmetry segments into a segment, and conjoin the connected segments which are neither common nor high asymmetric.

1. Introduction

The analysis of data structure, which contains correlation, complexity, fractal and irreversibility analysis, is required in diverse areas such as DNA sequences, physiological signals, stock markets and traffic systems. A popular problem related to the analysis of sequence structure is the series segmentation algorithm. Generally speaking, segmentation is a basic method of data analysis to identify change-points and is a partition of the sequence into non-overlapping segments which are as

* Corresponding author.
E-mail addresses: 08271110@bjtu.edu.cn (J. Xia), pjshang@bjtu.edu.cn (P. Shang), 13121597@bjtu.edu.cn (D. Lu), 09271087@bjtu.edu.cn (Y. Yin).

http://dx.doi.org/10.1016/j.physa.2015.12.156
0378-4371/

homogeneous as possible. Of course, the homogeneity property can be defined in different ways to satisfy our practical need, the only request is to make sequence as simple as it can.

There are several basic segmentation methods introduced and developed to a high-level observation of the sequence's structure, such as dynamic programming [1], heuristic algorithms (like top-down [2,3], randomized methods [4]), approximation algorithms (like approximate dynamic programming [5], divide and conquer method [6]) and inner composition alignment [7,8]. The cubic complexity makes the dynamic programming algorithm prohibitive to use in practice. Then, the approximation algorithms are presented to reduce the running time and give high quality results, and the heuristic algorithms also attempt to speed up. Both algorithms have their advantages and disadvantages, but they are fairly helpful to provide a more accurate classification and clustering, especially in DNA sequence analysis. The results are capable of uncovering a complex fractal structure for DNA, and presenting a novel view of the compositional heterogeneity of a DNA sequence [9–12]. This observation motivates researchers to develop optimal algorithms and broaden application scope. In recent years, statistical segmentation and clustering analysis are performed to analyze financial time series which is useful to calculate the cross correlations between various economic sectors [13–16].

In the present work we adopt the entropic segmentation method [9,10,17,18] developed by Bernaola-Galván and coworkers for biological sequence segmentation, and apply it to analyze the crude oil market. Though, there are plenty of tools applied to analyze the crude oil price dynamics, such as detrended fluctuation analysis (DFA) [19–21], RS analysis [22], asymmetry analysis [23] and multiscale entropy analysis (MSE) [24]. In our previous work, we apply these popular methods to study and investigate global properties of time series [25–29]. We consider the fixed window length $[N/s]$ with given scale s in DFA or the coarse-grained procedure in MSE may lose sight of inner structure of recorded series. So we apply the entropic segmentation method to study the structure of oil price series (from April 4, 1983 to December 31, 2014), and figure out the simplicity of entropic algorithm makes it easy to understand splitting procedure which can be concluded in 3 steps: (i) Start with unsegmented series and introduce one cut point at a time; (ii) At the k th step introduce the k th cut position by splitting one of the existing k segments in two; (iii) Repeat splitting until the stopping condition is satisfied. And its fast running also makes it attractive to use in massive real-data (such as oil price series).

Obviously, the presence of inner structure can cause the segmentation at various significance levels r. Then, if all of the segments tend to unique, the segmentation analysis is successful. As far as we know, the noise in original series may cause over splitting that some segments are too short. Therefore, we find the right subseries $D_1^{(2)}$ of original daily series D after the first splitting with distinct r which always begin at November 14, 1985 and $W_1^{(2)}$ of weekly series W which start from November 22, 1985. The time of first cut point coincides with that of the price mechanism reformed which happened at 1986 and can be viewed as a change point. On account of the effect of the price mechanism revolution, we analyze the daily and weekly series begin at 1986, marked as D^* and W^* respectively. Compare both results to find the noteworthy segments which are the common segments with same boundaries. Meanwhile, we employ two indices, *Porta's* index [30–32] and *Guzik's* index [33], to measure the asymmetry degree of each segment. Although, there are several other indices developed and applied to analyze the artificial series [34] and physiological signals [35,36]. They actually have the similar function to describe the property of time irreversibility in time sequence. In order to increase the information gained from the oil price series, we construct a two-dimensional plane [27,37] of *Porta's* and *Guzik's* index to identify the asymmetry degree of segments, and compare results to seek significant and important segments which can reflect the characteristics of the price series.

This paper is organized as follows. In the next section, we briefly review the segmentation method based on the Jensen–Shannon divergence [38], and time irreversibility measure indices. In Section 3, we present our main findings from statistical segmentation analysis and calculate the asymmetry degree for every segment to identify its structure. In the final section, we conclude the paper with a summary and discussion of our results.

2. Methods

2.1. Segmentation procedure

For a given sequence $X = \{x_1, x_2, \ldots, x_N\}$, let $Y = \{a_1, a_2, \ldots, a_N\}$ be a symbolic series and each a_i from the alphabet $\mathcal{A} = \{A_1, A_2, \ldots, A_k\}$. To divide series Y into homogeneous segments, we move a sliding pointer from left to right along the series and, at each position n we consider the two subseries:

$$Y^{(1)} = \{a_1, a_2, \ldots, a_n\}, \qquad Y^{(2)} = \{a_{n+1}, a_{n+2}, \ldots, a_N\}$$

and calculate the relative frequencies of each symbol in $Y^{(1)}$ and $Y^{(2)}$, respectively:

$$\mathcal{F}^{(1)} = \{f_1^{(1)}, \ldots, f_k^{(1)}\}, \qquad \mathcal{F}^{(2)} = \{f_1^{(2)}, \ldots, f_k^{(2)}\}$$

where $f_i^{(j)}$ is the relative proportion of symbol A_i in $Y^{(j)}$ $(i = 1, \ldots, k;\ j = 1, 2)$.

Then we use the Jensen–Shannon divergence measure [38] to compare adjacent subsequences and decide whether they are different. For two distributions,

$$JS(\mathcal{F}^{(1)}, \mathcal{F}^{(2)}) = H(\pi_1\mathcal{F}^{(1)} + \pi_2\mathcal{F}^{(2)}) - (\pi_1 H(\mathcal{F}^{(1)}) + \pi_2 H(\mathcal{F}^{(2)})) \tag{1}$$