

Distinguishing Stock Indices and Detecting Economic Crises Based on Weighted Symbolic Permutation Entropy

Ningning Zhang*, Yupeng Sun^{*,‡}, Yongbo Zhang[†], Pengbo Yang*,
Aijing Lin* and Pengjian Shang*

**School of Science, Beijing Jiaotong University
Beijing 100044, P. R. China*

*†Teaching and Research Section of Mathematics
Aviation University of Air Force
Changchun 130022, P. R. China*

‡ypsun@bjtu.edu.cn

Received 9 January 2019

Accepted 13 March 2019

Published 10 May 2019

Communicated by Wei-Xing Zhou

Permutation entropy (PE) is evolved as an important way to reveal the intrinsic characteristics of complex systems. While PE does not consider that the amplitudes of identical order sequences might be different, weighted permutation entropy (WPE), the extended technique, weighs each vector to alleviate this problem. Furthermore, experimental investigation shows that symbolizing the time series contributes to detecting sudden dynamical changes. As an exemplary measure, we study weighted symbolic permutation entropy (WSPE) which is a combination of WPE and threshold-dependent symbolic entropy to mount the intrinsic information of four stock indices from China and America stock markets, and six sector indices of banking, aviation industry and pharmacy from China stock markets. One thing needs to be especially pointed out is that symbolic method we use in this work is not the most common method which divides time series into some small bins, but a two-step process, including symbolic process and coding process. The statistical analysis in stock indices shows that WSPE method is capable of distinguishing stock indices and detecting economic crises.

Keywords: Weighted symbolic permutation entropy (WSPE); weighted permutation entropy (WPE); threshold-dependent symbolic method; financial time series.

1. Introduction

Nowadays, it is an increasingly important and active topic to describe the characteristics of the system by studying the nature of time series data [1–8]. Many complex measures are introduced in order to reveal the complexity of complex systems, such as dimensional analysis [9], largest Lyapunov exponents [10, 11],

[‡]Corresponding author.

statistical complexity measures [11] and various entropies [12–15], in which permutation entropy (PE) [4] is the most useful and easily implemented tool, and used for biological signals [16–21] and financial time series [22, 23], etc. The intrinsic features of complex systems can be estimated by mapping a fixed-length sequence to its ordinal pattern and then measuring the complexity of the ordinal sequence. The advantages of PE are its simplicity and fast calculation speed.

Although PE is effective in solving practical problems, it ignores the amplitude information of the identical order sequences [24–27]. Therefore, [28] introduced weighted permutation entropy (WPE), which addressed the limitation of PE by integrating amplitude contents from the corresponding sequential structure. By introducing a “weight” mechanism, WPE is not only superior to PE in distinguishing abrupt dynamics of the signals, but also able to detect the amplitude distinction between the identical order sequences.

Generally, it is necessary to extract symbol sequences from original time series when we analyze their properties to understand the systems. Various symbolic methods [29–35] that help reflect dynamical features are used for the complex systems. In this work, we adopt the symbolic method from threshold-dependent symbolic entropy [29, 36] and apply it to weighted symbolic permutation entropy (WSPE). To demonstrate the efficiency of WSPE based on symbolic process and coding process, we test the intrinsic features of four stock indices from America and China stock markets, and six sector indices of banking, aviation industry and pharmacy from China stock markets. The experimental results show WSPE could classify the stock time series from various countries and capture approximate period of economic crises.

The overall arrangement of this paper is as follows: Section 2 simply describes the methodologies of PE, WPE and WSPE. The data source used in this paper will be illustrated in Sec. 3. In Sec. 4, detailed experimental analysis by WSPE and WPE is given. Section 5 offers the conclusions and future studies.

2. Methods

2.1. Permutation entropy (PE)

In this section, we give a brief definition of PE, which is widely used to measure the complexity of dynamical systems [4]. For any time series $X = \{x_1, x_2, \dots, x_n\}$, we use phase-space reconstruction method to obtain its representation $X_i^{m,\tau} = \{x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\}$, where i is changed in the range of $[1, N - (m-1)\tau]$, τ represents time delay and m denotes embedding dimension. Next, we arrange the points of the X_i in ascending order $\{x_{i+(l_1-1)\tau} \leq x_{i+(l_2-1)\tau} \leq \dots \leq x_{i+(l_m-1)\tau}\}$, and define $\pi = [l_1, l_2, \dots, l_m]$ as the original pattern. Then, we calculate the probability of each order sequence as

$$p(\pi_l^m) = \frac{\#\{i | i = 1, 2, \dots, N - m + 1; X_i^m \text{ has type } \pi_l^m\}}{N - m + 1}. \quad (1)$$

PE, which is an extension of the Shannon entropy and can be calculated according to the following formula:

$$H(\tau, m) = - \sum_{l=1}^{m!} (\pi_l) \log p(\pi_l). \quad (2)$$

PE values change from 0 to $\log m!$. Although PE is effective in solving practical problems, it ignores the amplitude contents of the signals. In order to resolve this problem, Fadlallah proposed the WPE [28], which will be shown in the next subsection.

2.2. Weighted permutation entropy (WPE)

The differences between WPE and PE are that it is not same for the weight of WPE between the adjacent vectors of the identical order patterns with different amplitude changes. First, the weight value w_i is used to weight each vector X_i^m . Then, calculate w_i with the variances of all adjacent vectors X_i as

$$w_i = \frac{1}{m} \sum_{k=1}^m [x_{i+(k-1)\tau} - \bar{X}_i]^2. \quad (3)$$

The following formula shows how to compute the weighted relative frequencies:

$$p_w(\pi_l^m) = \frac{\sum_{l: X_i^m \text{ has } \pi_l^m} w_i}{\sum_{l \leq N-m+1} w_i}. \quad (4)$$

Next, WPE is calculated as

$$H_w(\tau, m) = - \sum_{l: \pi_l^m \in \Pi} p_w(\pi_l^m) \log p_w(\pi_l^m). \quad (5)$$

2.3. Weighted symbolic permutation entropy (WSPE)

In the following, we introduce the WSPE based on threshold dependent symbolic method, which mainly includes two processes, the symbolization process and the encoding process. At first, for a given time series $X = \{x_1, x_2, \dots, x_n\}$, it can be mapped to symbolic patterns $Y = \{y_1, y_2, \dots, y_n\}$ with ξ values from zero to $\xi - 1$. For the sake of simplicity, we mainly use binary signification in this article. In this way, the time series can be symbolized as a sequence consisting of symbols '0' and '1' by the following formula:

$$x^\xi = \begin{cases} 1 & |x_i - \bar{x}| \geq \alpha \\ 0 & |x_i - \bar{x}| < \alpha \end{cases}, \quad (6)$$

where α demotes the threshold value, and \bar{x} represents the average value of x . For the value of α , we can set it to a multiple of the variance of the sequence.

Unlike traditional symbolic methods [23], we then implement the encoding process on the symbol sequence. The symbol sequence is divided with L -length and then code sequences can be obtained. In this paper, we set $L = 3$. Firstly, we implement an overlapping partition of the symbol sequences according to the following formulas:

$$\psi_{L,i}^{\xi} = x_i^{\xi}, x_{i+1}^{\xi}, x_{i+2}^{\xi}, \dots, x_{i+L-1}^{\xi}. \quad (7)$$

Therefore, the code series is generated as

$$y_{L,i} = x_i^{\xi} \xi^{L-1} + x_{i+1}^{\xi} \xi^{L-2} + x_{i+2}^{\xi} \xi^{L-3} + \dots + x_{i+L-1}^{\xi} \xi^0, \quad (8)$$

where ξ denotes the quantization level and ξ^L is all possible number of codes. In fact, the encoding process is to convert the fixed length of L 's ξ -ary number into decimal number, which is equivalent to reconstructing the state space of the symbol sequence. For example, given a time series $X = \{4, 5, 3, 8, 9, 3\}$, we can get the symbolic sequences $Y = \{0, 0, 0, 1, 1, 0\}$ and the code sequences $\{0, 0, 1, 3, 2\}$ with $L = 2$. In this way, we can get the coding sequence of the original time series. Finally, we can estimate the complexity of the original sequence by calculating the WPE of the coded sequence.

3. Data Source

To verify the validity of our algorithm, we use real-world data to perform a series of experiments. Financial time series are quite complex and fluctuating frequently, so we focus on the field of finance. Hence, four stock indices and six sector indices from America and China stock markets are chosen, and we list these four stock indices which span the period from January 3, 2006 to March 11, 2016 in Table 1, and six sector indices of banking, aviation industry and pharmacy during the period from 5 November, 2007 to 15 July, 2016 in Table 2. The data used in this paper can be downloaded from the website of <http://finance.yahoo.com>.

As is known to all, since stock indices from different countries may have different trading dates, we use cubic spline interpolation to complete these data to obtain the financial time series of the same length. In addition, for x_t , the price of indices on day t , we calculate the logarithmic daily price return r_t , $r_t = \ln(x_t) - \ln(x_{t-1})$. Figure 1 shows the graphical representation of returns of these four stock indices. We can find the returns have mutual zone of fluctuations in Fig. 1 which might be the worst days.

Table 1. The list of four stock indices.

Num	Country	Index	Abbreviation
1	America	Nasdaq	NAS
2	America	Standard and Poors 500 Index	S&P500
3	China	Shanghai Stock Exchange Composite Index	SSE
4	China	Shenzhen Component Index	SZSE

Table 2. The list of six sector indices.

Num	Country	Index	Abbreviation	Industry
1	China	China Construction Bank	CCB	Banking
2	China	AIR CHINA	CA	Aviation
3	China	TRT Health International	TRT	Pharmacy
4	America	Bank of America Corporation	BAC	Banking
5	America	American Airlines Group Inc.	AAL	Aviation
6	America	Vertex Pharmaceuticals Incorporated	VRTX	Pharmacy

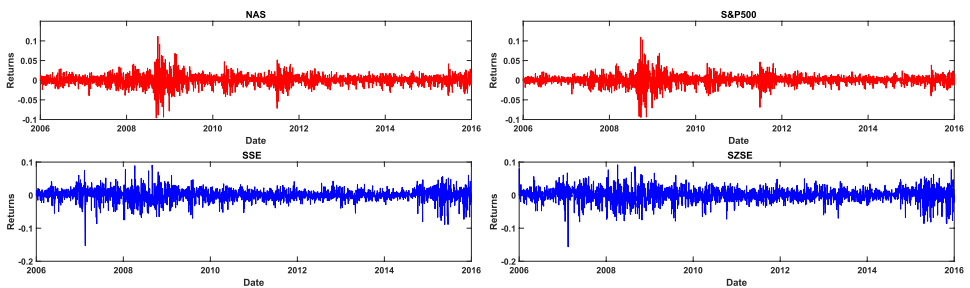


Fig. 1. Returns of NAS, S&P 500, SSE and SZSE. We can find that the returns from the same country have a mutual zone of large fluctuations, which may be the worst days of the economy.

4. Results

Here, we will compare the experimental results of the WSPE and the WPE. These two algorithms are applied to investigate the complexity of the four stock indices (NAS, S&P500, SSE and SZSE) and six sector indices (CCB, CA, TRT, BAC, AAL and VRTX). We will not only observe the fluctuation of the two entropy measures at different times, but also give the corresponding analysis of the fluctuation based on some economic events. The relevant parameters in the experiments shown in this section are set as follows: the embedding dimensions m is in the range of 3–7 according to [4] and the time delay τ is in the range of 1–20.

Firstly, we employ the WSPE to investigate the complexity of the stock indices (see Fig. 2). Moreover, the WPE analysis is also given as a comparison. From Fig. 2, we can see the values of the WSPE and WPE of the stock indices from the same countries are similar with the increase of time delay at any m values. Although the WPE could classify the stock indices according to the country, the values of the WPE from the same countries are very similar, making it difficult to distinguish between stock indexes. In contrast, the WSPE performs better than the WPE in classifying the financial time series as WPE ignores subtle differences of stock indices from the same group. Figure 3 presents the WSPE and WPE results when τ changes from 1 to 20 and m varies from 3 to 7. In Fig. 3, we can see that the values of the WSPE rise greatly as the embedding dimension m increases than those of the WPE.

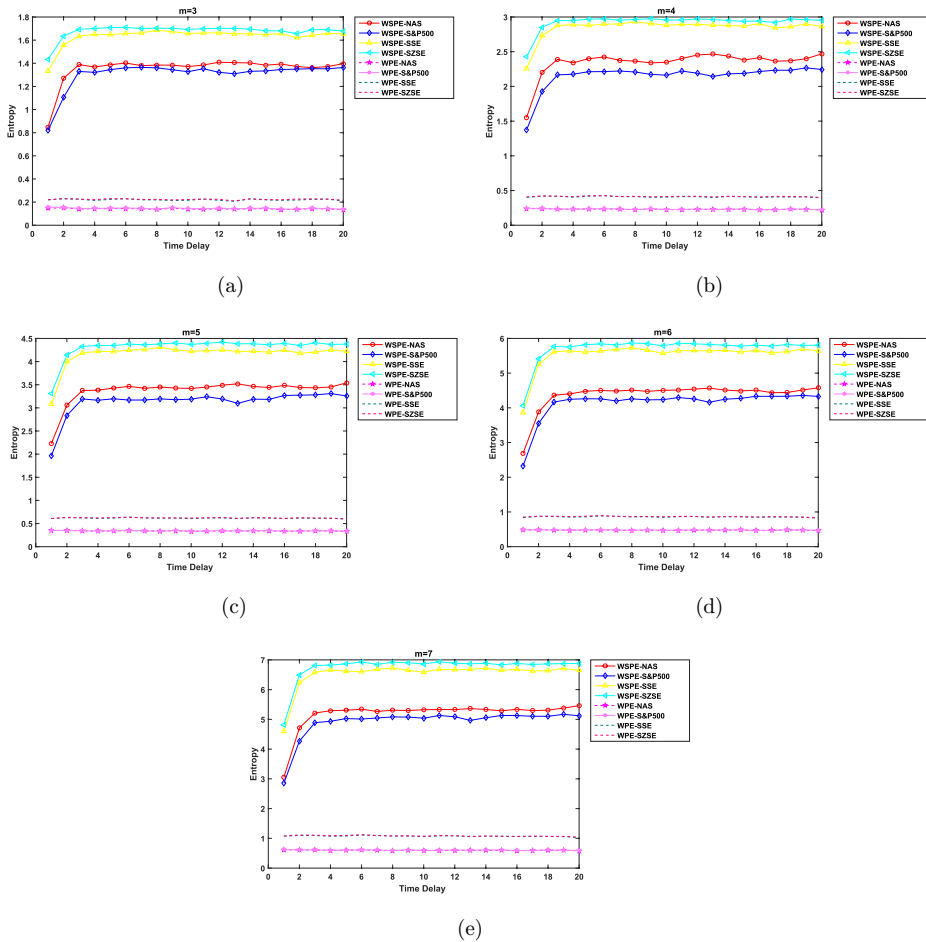


Fig. 2. The WSPE analysis and WPE analysis of NAS, S&P500, SSE and SZSE at different m . (a) The values of the WSPE and WPE for NAS, S&P500, SSE and SZSE stock indices when $m = 3$ versus time delay. (b) The values of the WSPE and WPE for NAS, S&P500, SSE and SZSE stock indices when $m = 4$ versus time delay. (c) The values of the WSPE and WPE for NAS, S&P500, SSE and SZSE stock indices when $m = 5$ versus time delay. (d) The values of the WSPE and WPE for NAS, S&P500, SSE and SZSE stock indices when $m = 6$ versus time delay. (e) The values of the WSPE and WPE for NAS, S&P500, SSE and SZSE stock indices when $m = 7$ versus time delay.

Thus, WSPE is more accurate than WPE since WSPE easily detects the differences among different time series. In order to fully prove WSPE can classify different stock indices, we employ WSPE to different level sector indices as a control group. As shown in Fig. 4, as the time delay increases, the values of WSPE for CCB, CA, TRT from China fluctuate around 1 to 1.1, while those for BAC, AAL and VRTX are little changed from 0.6 to 0.7. It is obvious that the WSPE values of the sector indices from the same country are much closer. This result is consistent with the division of objective fact.

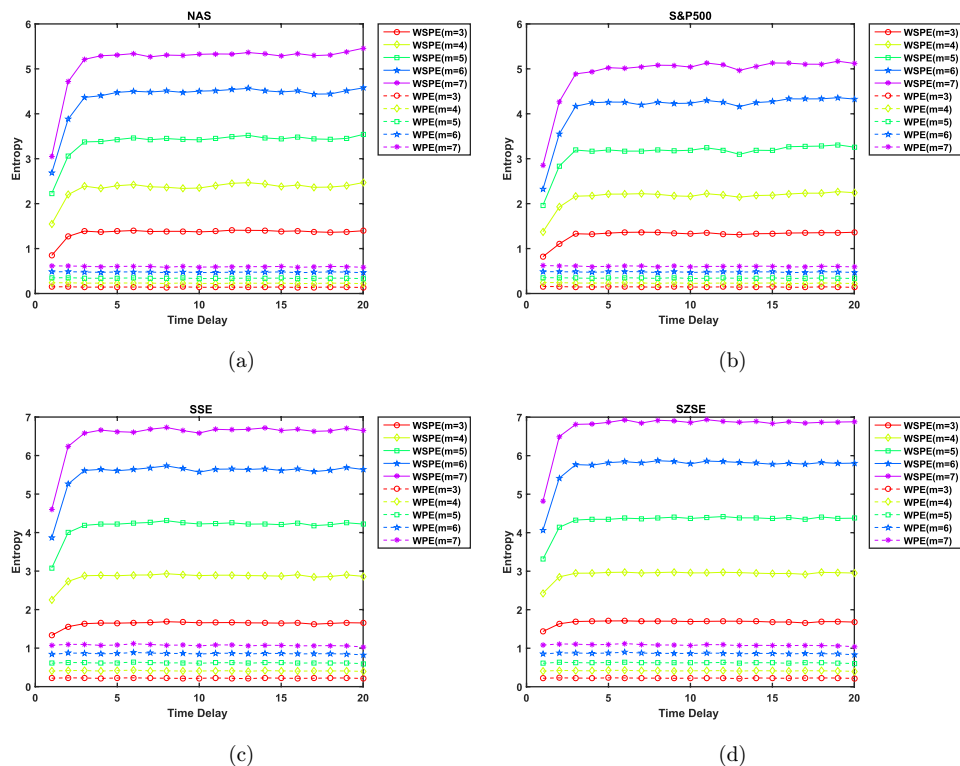


Fig. 3. The WSPE analysis and WPE analysis of NAS, S&P500, SSE and SZSE stock indices at different m . (a) The values of the WSPE and WPE for NAS stock index at different m versus time delay. (b) The values of the WSPE and WPE for S&P500 stock index at different m versus time delay. (c) The values of the WSPE and WPE for SSE stock index at different m versus time delay. (d) The values of the WSPE and WPE for SZSE stock index at different m versus time delay.

Next, we consider using the rolling windows method to analyze the changes of the WSPE for the four stock indices. Here we fix the embedding dimension m . Because the results are not clear-cut, we go to work on the changes of the variances of the WSPE for those four stock indices under varying time delay (see Fig. 5). With regard to the choice of the size of the rolling window, the fact is that a large window size is suitable for testing the general trends of long-range market dynamics, while a small window size will be better for observing the short-term dynamics. Therefore, the window size is fixed to be 780 trading days in our study, which roughly equals to three years. And the window step is 260 single trading days, which roughly equals to one year. As shown in Figs. 5(a) and 5(b), the WSPE values greatly fluctuate during the period from 2008–2011 when stock indices from America suffer economic crisis. We observe the variances of the WSPE have a mutual zone of large fluctuations—2008–2011 and 2009–2012 in Figs. 5(c) and 5(d), which may be the worst days of the economy.

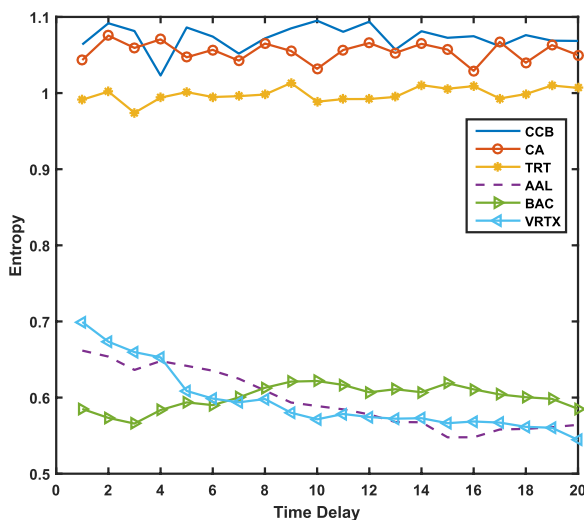


Fig. 4. The WSPE analysis of different sector indices from China and America at $m = 3$ versus time delay.

Finally, we study the WSPE during the period 2008–2012 and during the period 2012–2016 when m is in the range of 3–7 and τ is in the range of 1–20 in Fig. 6. It can be seen in Fig. 6 that the values of the WSPE during the period 2008–2012 are larger than those during the period 2012–2016 no matter what the values of embedding

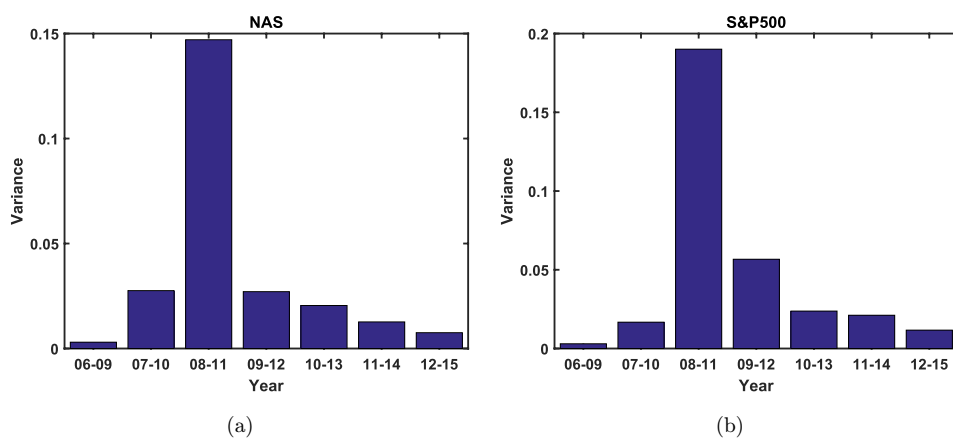


Fig. 5. The variance of the WSPE of NAS, S&P500, SSE and SZSE stock indices. (a) The values of WSPE of NAS stock index for 2006–2009, 2007–2010, 2008–2011, 2009–2012, 2010–2013, 2011–2014, 2012–2015 and 2013–2016. (b) The values of WSPE of S&P500 stock index for 2006–2009, 2007–2010, 2008–2011, 2009–2012, 2010–2013, 2011–2014, 2012–2015 and 2013–2016. (c) The values of WSPE of SSE stock index for 2006–2009, 2007–2010, 2008–2011, 2009–2012, 2010–2013, 2011–2014, 2012–2015 and 2013–2016. (d) The values of WSPE of SZSE stock index for 2006–2009, 2007–2010, 2008–2011, 2009–2012, 2010–2013, 2011–2014, 2012–2015 and 2013–2016.

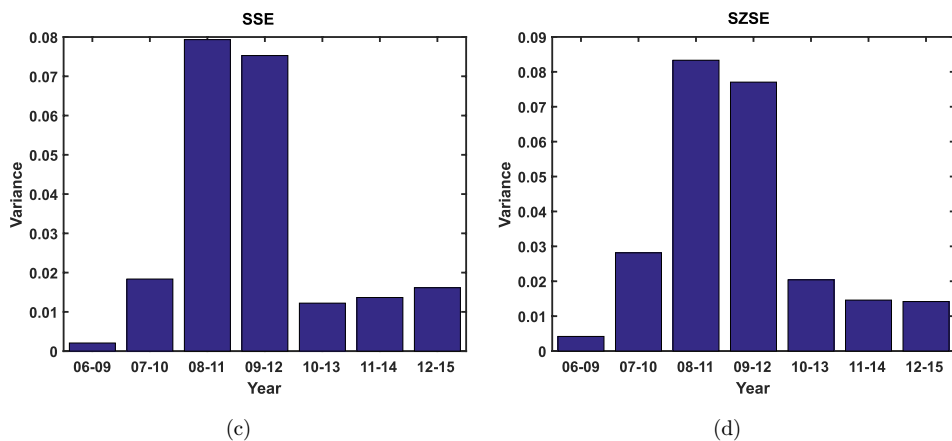


Fig. 5. (Continued)

dimension m and time delay τ , which suggests the complexity of the financial time series of economic crisis is higher than those of other periods. It is possible to verify that in the period between January 2008 and May 2012, significant changes are caused by the sub-prime crisis in America and the sovereign debt crisis in European countries [37].

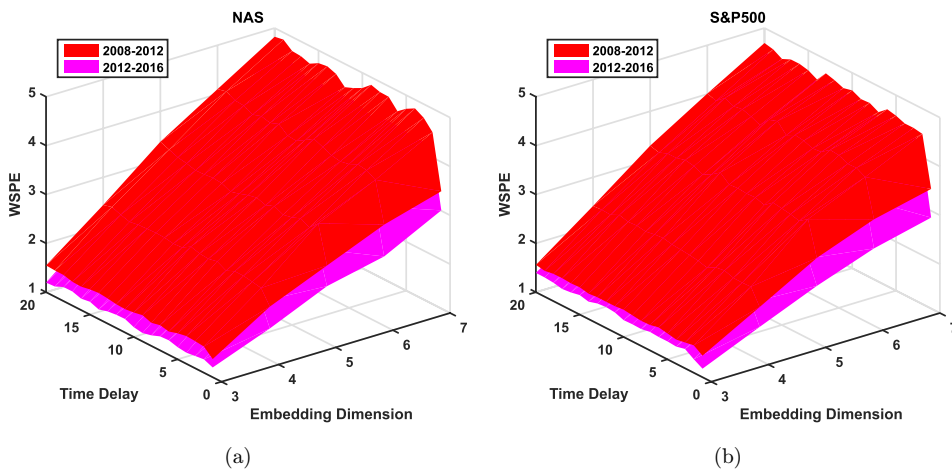


Fig. 6. The WSPE analysis for NAS, S&P500, SSE and SZSE stock indices from the period 2008–2012 and from the period 2012–2016. (A) The values of WSPE for NAS stock indice from 2008–2012 and from the period 2012–2016 versus time delay and embedding dimension. (B) The values of WSPE for S&P500 stock indice from 2008–2012 and from the period 2012–2016 versus time delay and embedding dimension. (C) The values of WSPE for SSE stock indice from 2008–2012 and from the period 2012–2016 versus time delay and embedding dimension. (D) The values of WSPE for SZSE stock indice from 2008–2012 and from the period 2012–2016 versus time delay and embedding dimension.

5. Conclusions

In this work, we explore an improved methodology — WSPE which combines WPE and threshold dependent symbolic entropy to explore the complexity of America stock indices (NAS and S&P500) and China stock indices (SSE and SZSE), six sector indices of banking, aviation industry and pharmacy from China markets. In order to verify the superiority of WSPE we compare it with WPE on the analysis of complexity of the stock indices. The first thing we find is that the WSPE is capable of distinguishing four stock indices from America and China stock markets while the WPE exhibits a poor performance. Moreover, the WSPE is sensitive to the changes of embedding dimension. Then we analyze the complexity changes of the four stock indices over time using the WSPE combined with rolling windows. The results show that the WSPE successfully detects two economic crises: the sub-prime crisis in America and the sovereign debt crisis in European countries.

In summary, the WSPE can be an effective approach to study the complexity of nonlinear time series. WSPE has an extensive prospect on complexity analysis of time series in various fields, such as transportation, physiology, geography.

Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. 2017YJS195).

References

- [1] T. J. McCabe, A complexity measure, *IEEE Trans. Softw. Eng.* **2**(4) (1976) 308–320.
- [2] S. Pincus, Approximate entropy (apen) as a complexity measure, *Chaos: An Interdiscip. J. Nonlinear Sci.* **5**(1) (1995) 110–117.
- [3] F. Kaspar and H. Schuster, Easily calculable measure for the complexity of spatiotemporal patterns, *Phys. Rev. A* **36**(2) (1987) 842.
- [4] C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.* **88**(17) (2002) 174102.
- [5] W. B. Arthur, Complexity in economic and financial markets: Behind the physical institutions and technologies of the marketplace lie the beliefs and expectations of real human beings, *Complexity* **1**(1) (1995) 20–25.
- [6] M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (Simon and Schuster, 1993).
- [7] N. E. Hynes and H. A. Lane, ErbB receptors and cancer: The complexity of targeted inhibitors, *Nat. Rev. Cancer* **5**(5) (2005) 341.
- [8] S. M. Pincus, Approximate entropy as a measure of system complexity, *Proc. National Academy of Sciences* **88**(6) (1991) 2297–2301.
- [9] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Phys. D: Nonlinear Phenom.* **9**(1–2) (1983) 189–208.
- [10] M. T. Rosenstein, J. J. Collins and C. J. De Luca, A practical method for calculating largest lyapunov exponents from small data sets, *Phys. D: Nonlinear Phenom.* **65**(1–2) (1993) 117–134.

- [11] H. Kantz, A robust method to estimate the maximal lyapunov exponent of a time series, *Phys. Lett. A* **185**(1) (1994) 77–87.
- [12] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.* **85**(2) (2000) 461.
- [13] S. Leung-Yan-Cheong and T. Cover, Some equivalences between shannon entropy and kolmogorov complexity, *IEEE Trans. Inf. Theory* **24**(3) (1978) 331–338.
- [14] W. Harrison, An entropy-based measure of software complexity, *IEEE Trans. Softw. Eng.* **18**(18) (1992) 1025–1029.
- [15] M. U. Ahmed and D. P. Mandic, Multivariate multiscale entropy: A tool for complexity analysis of multichannel data, *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* **84**(6 Pt 1) (2011) 061918.
- [16] X. Sun, Y. Zou, V. Nikiforova, J. Kurths and D. Walther, The complexity of gene expression dynamics revealed by permutation entropy, *Bmc Bioinf.* **11**(1) (2010) 607–607.
- [17] B. Frank, B. Pompe, U. Schneider and D. Hoyer, Permutation entropy improves fetal behavioural state classification based on heart rate analysis from biomagnetic recordings in near term fetuses, *Med. Biol. Eng. Comput.* **44**(3) (2006) 179.
- [18] D. Labate, I. Palamara, N. Mammone, G. Morabito, F. L. Foresta and F. C. Morabito, SVM classification of epileptic eeg recordings through multiscale permutation entropy, in *Int. Joint Conf. Neural Networks* (Dallas, Texas, USA, 2014), pp. 4–9.
- [19] W. Aziz and M. Arif, Multiscale permutation entropy of physiological time series, in *Int. Multitopic Conf.* (Pakistan, 2005), pp. 1–6.
- [20] S. Pigolotti, S. Krishna and M. H. Jensen, Symbolic dynamics of biological feedback networks, *Phys. Rev. Lett.* **102**(8) (2009) 088701.
- [21] R. Shalhaf, H. Behnam, J. W. Sleigh, A. Steynross and L. J. Voss, Monitoring the depth of anesthesia using entropy features and an artificial neural network, *J. Neurosci. Methods* **218**(1) (2013) 17–24.
- [22] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez and O. A. Rosso, Forbidden patterns, permutation entropy and stock market inefficiency, *Phys. A: Stat. Mech. Appl.* **388**(14) (2009) 2854–2864.
- [23] Y. Yin and P. Shang, Weighted permutation entropy based on different symbolic approaches for financial time series, *Phys. A: Stat. Mech. Appl.* **443** (2016) 137–148.
- [24] X. Chen, N.-D. Jin, A. Zhao, Z.-K. Gao, L.-S. Zhai and B. Sun, The experimental signals analysis for bubbly oil-in-water flow using multi-scale weighted-permutation entropy, *Phys. A: Stat. Mech. Appl.* **417** (2015) 230–244.
- [25] Y. Yin and P. Shang, Weighted multiscale permutation entropy of financial time series, *Nonlinear Dyn.* **78**(4) (2014) 2921–2939.
- [26] R. Li and J. Wang, Interacting price model and fluctuation behavior analysis from lempel–ziv complexity and multi-scale weighted-permutation entropy, *Phys. Lett. A* **380**(1) (2016) 117–129.
- [27] P. L. Vuong, A. S. Malik and J. Bornot, Weighted-permutation entropy as complexity measure for electroencephalographic time series of different physiological states, in *2014 IEEE Conf. Biomedical Engineering and Sciences (IECBES)* (IEEE, 2014), pp. 979–984.
- [28] B. Fadlallah, B. Chen, A. Keil and J. Principe, Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information, *Phys. Rev. E* **87**(2) (2013) 022911.
- [29] W. Aziz and M. Arif, Complexity analysis of stride interval time series by threshold dependent symbolic entropy, *European J. Appl. Physiol.* **98**(1) (2006) 30–40.
- [30] F. López, M. Matilla-García, J. Mur and M. R. Marín, A non-parametric spatial independence test using symbolic entropy, *Reg. Sci. Urban Econ.* **40**(2) (2010) 106–115.

- [31] C. Renu Madhavi and A. Ananth, Quantification of heart rate variability (hrv) data using symbolic entropy to distinguish between healthy and disease subjects, *Thyroid* **1**(5.6) (2010) 3–4.
- [32] N. Zhang, A. Lin and P. Shang, Symbolic phase transfer entropy method and its application, *Commun. Nonlinear Sci. Numer. Simul.* **51** (2017) 78–88.
- [33] K. Park and S. Yi, Accessing physiological complexity of hrv by using threshold-dependent symbolic entropy, *J. Korean Phys. Soc.* **44** (2004) 569.
- [34] W. Aziz Loun and M. Arif, Complexity analysis of heart beat time series by threshold based symbolic entropy, in *3rd Kuala Lumpur Int. Conf. Biomedical Engineering 2006* (Springer, 2007), pp. 369–373.
- [35] Y. Wu and L. Shi, Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation, *Med. Eng. Phys.* **33**(3) (2011) 347–355.
- [36] W. Aziz, M. Rafique, I. Ahmad, M. Arif, N. Habib and M. Nadeem, Classification of heart rate signals of healthy and pathological subjects using threshold based symbolic entropy, *Acta Biologica Hungarica* **65**(3) (2014) 252–264.
- [37] B. A. Goncalves and A. P. F. Atman, Visibility graph combined with information theory: An estimator of stock market efficiency, *J. Network Theory Finance* **3**(2) (2017) 1–15.