Physica A 484 (2017) 215–224

Contents lists available at ScienceDirect

Physica A

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

Multiscale Shannon entropy and its application in the stock market

Rongbao Gu

School of Finance, Nanjing University of Finance and Economics, Nanjing 210046, PR China

HIGHLIGHTS

- A new concept of the multi-scale Shannon entropy is proposed.
- The DJIA is studied by the multi-scale entropy analyses.
- The predictive power of the multi-scale entropy for DJIA are analyzed.
- The useful information has predictive power for DJIA in the long-term.
- The noise has predictive power for DJIA in the short-term.

ARTICLE INFO

Article history:
Received 25 January 2017
Received in revised form 9 April 2017
Available online 3 May 2017

Keywords:
Stock market
Prediction
Singular value decomposition
Multi-scale Shannon entropy

ABSTRACT

In this paper, we perform a multiscale entropy analysis on the Dow Jones Industrial Average Index using the Shannon entropy. The stock index shows the characteristic of multi-scale entropy that caused by noise in the market. The entropy is demonstrated to have significant predictive ability for the stock index in both long-term and short-term, and empirical results verify that noise does exist in the market and can affect stock price. It has important implications on market participants such as noise traders.

1. Introduction

Efficient market hypothesis (EMH) proposed by Fama [1] in 1960s is the cornerstone of modern financial research. It states that if all the information available in the past can be reflected in the stock price, the market is efficient. In this way, the stock price would follow a random walk behavior and is unpredictable. However, it has been well documented that EMH cannot be established in real market, implying that stock price is predictable to some extent.

Entropy is an important notion in nonlinear science, it is a measurement for the uncertainty and complexity of dynamic system. Recently, entropy has been used to study the predictability of stock market. For instance, Maasoumi et al. [2] used metric entropy to detect the predictability of stock market, and found that compared with traditional predicting method, metric entropy can capture more nonlinear relations. Eom et al. [3] used metric entropy and Hurst index to study the predictability of several stock indices, they found that the predictability of a stock index is positively related to the Hurst index but negatively related to the value of metric entropy.

Shannon entropy is a measurement of information contained in a system. The greater value of Shannon entropy indicates the more information is needed for people to understand this system. Caraiani [4] introduced the singular value decomposition entropy based on the correlated coefficient matrix, and the entropy turns out to have predictive power for

E-mail address: rbgu@sina.com.

http://dx.doi.org/10.1016/j.physa.2017.04.164
0378-4371/

the Dow Jones Industrial Average Index. Following Caraiani, Gu et al. [5] studied the predictive power of singular value decomposition entropy on Chinese Shenzhen stock market. They found that the predictive power is affected by the structural breaks in the market, and it only works on the Shenzhen component index after the reform of non-tradable shares. This is an interesting result that differ from Caraiani [4] to some degree.

Behavioral finance theory holds that irrational investors exist in the stock market. Both the useful information and noise is the fundamentals of investors' decision making. Though the traditional singular value decomposition technique can effectively distinguish different information, the singular value decomposition entropy proposed by Caraiani [4] is a composite measurement for all kinds of information. Then, the more efficient measurement of the different information and further analysis of its predictive power for stock market index is an available topic that can promote a deeper understanding of stock market.

In this paper we introduce a new notion of entropy, the multi-scale Shannon entropy, and apply it on the Dow Jones Industrial Average Index to detect the predictive power of the singular value decomposition multi-scale entropy for the index. Caraiani [6] suggests that what has been observed in terms of entropy as a systemic measure is also observable in terms of local properties. We contribute in this direction by analyzing the implications of entropy for different local scales for stock markets.

The remainder of paper is organized as follows: Section 2 is dedicated to an introduction about the singular value decomposition multi-scale entropy. Section 3 is the multi-scale entropy analysis for the Dow Jones Industrial Average Index. The predictive power of the singular value decomposition multi-scale entropy for the index is presented in Section 4. Section 5 is a brief conclusion.

2. Singular value decomposition multi-scale entropy

2.1. Shannon entropy and its generalization

The conception of entropy stems comes from physics. In 1856, German physicist Clausius [7] originally put forward the concept of entropy, which is used to describe the complexity of energy distribution in the space. Then Shannon [8] applied the entropy into the field of information science, using it to measure the amount of information transmission.

In a system, if P_i stands for the occurrence probability of one event, then $-\log(P_i)$ is taken as the amount of transitive information from the event. Thus, the statistical average value of the transitive information from all single events, $-\sum_i P_i \log(P_i)$ will be definited as the transitive information of the system. The statistical average value is called the information entropy or Shannon entropy of a system and wrote as *ent*. That is:

$$ent = -\sum_i P_i \log(P_i). \tag{1}$$

A system with higher Shannon entropy has more transitive information, which indicates greater uncertainty. It is noticed that the event with higher occurrence probability contributes less transitive information to the system. On the contrary, the event with lower occurrence probability contributes more transitive information to the system. In order to distinguish the amount of transitive information from the event with different occurrence probabilities, we introduce the following multi-scale Shannon entropy.

Assuming that $P_1, P_2, \ldots, P_n$ is the probability distribution, for any scale $q \neq 0$, define a qth-order Shannon entropy as:

$$ent_q = \left[\sum_i P_i \left(\log P_i^{-1}\right)^q\right]^{1/q}. \tag{2}$$

For scale $q = 0$, we define qth-order Shannon entropy as

$$ent_q = \prod_i e^{P_i} \left(\log P_i^{-1}\right). \tag{3}$$

Eqs. (2) and (3) are jointly called as the generalized Shannon entropy or the multiscale Shannon entropy. Especially, the generalized Shannon entropy is the normal Shannon entropy when q equals to 1.

The generalized Shannon entropy describes the multi-scale characteristic of a system from the perspective of information transmission, which is similar with generalized Hurst exponent [9]. The generalized Shannon entropy shares some similar characteristics with the generalized Hurst exponent. For instance, the form of generalized Shannon entropy ent_q depends on scale q. For negative q, ent_q mainly describes those transitive information from events with higher occurrence probability. For positive q, ent_q mainly describes those transitive information from events with lower occurrence probability.

2.2. Singular value decomposition multi-scale entropy

Caraiani [4] proposed the singular value decomposition entropy and investigated in its predictive power for the Dow Jones Industrial Average Index. The definition of the singular value decomposition entropy is presented as follows:

Assuming S_k is the kth constituent stock from one stock index, $S_{k,t}$ is the closing price of stock k at moment t, $y_{kt} = \log(S_{k,t}/S_{k,t-1})$ is the logarithmic return series of stock k, and $A = (R_{i,j})$ is the correlation matrix of the stock index. Where $R_{i,j}$ stands for the Pearson correlations between stock i and j, namely by:

$$R_{i,j} = \frac{\langle (y_{it} - \langle y_{it} \rangle)(y_{jt} - \langle y_{jt} \rangle) \rangle}{\sigma_i \sigma_j} \tag{4}$$

where $\langle \cdot \rangle$ stands for the mean of the returns of the stock, while σ_k is the standard deviation of the logarithmic return series of stock k.

We pick the unitary matrix U and V to make the following equation come into existence.

$$A = USV^T \tag{5}$$

where V^T is the transpose of matrix V, $S = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, is a diagonal matrix. The Eq. (5) is the singular value decomposition of matrix A. Where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are singular values of matrix A, which is also called singular values of the stock index.

Set $\bar{\lambda}_i = \lambda_i / \sum_j \lambda_j$, where $\sum_i \bar{\lambda}_i = 1$. The following Shannon entropy is:

$$Ent = -\sum_i \bar{\lambda}_i \log\left(\bar{\lambda}_i\right). \tag{6}$$

It is called singular value decomposition entropy of correlation matrix of the stock index, or singular value decomposition entropy of the stock index. It is a measurement for those information contained in the correlation matrix.

In Eq. (5), the singular values in the diagonal matrix are ranked in order of numerical values. Singular values are always consistent with the importance of information in correlation matrix. It is known from [10] that the bigger singular values ranked ahead reflect the amount of useful information, and the smaller singular values ranked behind reflect the amount of noise. Thus, the singular value decomposition technique can effectively distinguish different kinds of information in correlation matrix.

The singular value decomposition entropy defined by Eq. (6) is a composite measurement for all kinds of information. In order to measure the amount of useful transitive information and noise in correlation matrix respectively, we introduce the notion of singular value decomposition multi-scale entropy based on correlation matrix.

For any scale s, the following sth-Shannon entropy is:

$$ENT_s = \left(\sum_i \bar{\lambda}_i \left(\log \bar{\lambda}_i^{-1}\right)^s \right)^{\frac{1}{s}} \tag{7}$$

for $s \neq 0$:

$$ENT_s = \prod_i e^{\bar{\lambda}_i} \left(\log \bar{\lambda}_i^{-1}\right) \tag{8}$$

for $s = 0$, it is called the singular value decomposition multi-scale entropy of the correlation matrix of the stock index, briefly, the singular value decomposition multi-scale entropy of the stock index. (Note: This is different from the multi-scale singular value decomposition entropy proposed by Gu and Shao in [11].)

If ENT_s are the same for different scale s, then the stock index has the characteristic of single-scale entropy. Otherwise, the stock index has the characteristic of multi-scale entropy. For negative scale s, entropy ENT_s mainly reflect the amount of useful transitive information, and for positive scale s, entropy ENT_s mainly reflect the amount of noise in correlation matrix.

The following singular value decomposition multi-scale entropy mentioned in this paper particularly means this multi-scale Shannon entropy.

3. The multi-scale entropy analysis for the *DJIA*

3.1. Description of data

As the application of singular value decomposition multi-scale entropy on stock market, we consider one of the main stock market indices of America—the Dow Jones Industrial Average Index (here after, *DJIA*). We collect the daily closing prices of the Dow Jones Industrial Average Index and its component stocks. The *DJIA* index is composed of thirty component stocks listed in Table 1, among which Visa (March 19, 2008), Goldman Sachs Group, Inc. (Maybe 4, 1999) and Cisco Systems, Inc. (Feb. 16, 1990) are three latest listed stocks, so the price data of these three stocks are eliminated. We choose the sample covering from February 16, 1990 to June 30, 2016 (totally 6645 observations). The stock prices are treated by forward right in database and all data comes from the Yahoo finance website.

3.2. Multi-scale entropy of the DJIA

Andreadis and Serietis [12] investigated in the *DJIA* and found that the index has the characteristic of multi-fractal. In the following part, we will reveal the multi-scale characteristic of the index from the perspective of information entropy.

Table 1
The component stocks of the Dow Jones Industrial Average Index.

Company	Abbrev.	Company	Abbrev.
Apple Inc.	AAPL	The Coca-Cola Company	KO
American Express Company	AXP	McDonald's Corporation	MCD
The Boeing Company	BA	3M Company	MMM
Caterpillar, Inc.	CAT	Merck & Company, Inc.	MRK
Cisco Systems, Inc.	CSCO	Microsoft Corporation	MSFT
Chevron Corporation	CVX	Nike, Inc.	NKE
E.I. du Pont de Nemours and Company	DD	Pfizer, Inc.	PFE
The Walt Disney Company	DIS	Procter & Gamble Company	PG
General Electric Company	GE	The Travelers Companies, Inc.	TRV
The Goldman Sachs Group, Inc.	GS	United Health Group incorporated	UNH
The Home Depot, Inc.	HD	United Technologies Corporation	UTX
International Business Machines Corporation	IBM	Visa	V
Intel Corporation	INTC	Verizon Communications Inc.	VZ
Johnson & Johnson	JNJ	Wal-Mart Stores, Inc.	WMT
JP Morgan Chase & Co.	JPM	Exxon Mobil Corporation	XOM

Table 2
Singular values of *DJIA*.

Order number	Singular value	Order number	Singular value	Order number	Singular value	Order number	Singular value
1	8.3128	8	0.6437	15	0.2203	22	0.0794
2	6.5013	9	0.5652	16	0.1827	23	0.0710
3	2.7504	10	0.5430	17	0.1592	24	0.0597
4	2.1868	11	0.4548	18	0.1355	25	0.0517
5	1.7089	12	0.3833	19	0.1201	26	0.0396
6	1.1769	13	0.2889	20	0.1117	27	6.8e−16
7	0.9147	14	0.2514	21	0.0872	28	6.8e−16

Table 3
The singular value decomposition multi-scale entropy of *DJIA*.

s	ENT_s	s	ENT_s
−150	1.2243	1	2.2548
−100	1.2292	5	3.3625
−50	1.2443	10	4.1898
−20	1.2892	20	5.8607
−10	1.3552	50	18.0479
−5	1.4610	100	26.2753
−1	1.7944	150	29.7797

In order to test the multi-scale characteristic of *DJIA*, we construct the correlation matrix of the index using Eqs. (4) and (5), and conduct the singular value decomposition. The empirical results are shown in Table 2.

From Table 2, we can find that the first six singular values of *DJIA* are on single digit level, and from the 7th to 22th singular value, the number falls on decile level. From the 21th to 26th number, singular values of *DJIA* are on percentile level, followed by the 27th and 28th singular values falling on 10^{-16} level. It obviously shows that the useful information is mainly reflected by the first six singular values, and the noise is centrally reflected by the last ten singular values, especially the last two. We can measure the amount of useful information and noise of *DJIA* using singular value decomposition multi-scale entropy based on different scales, the two highest singular values are 8.3128 and 6.5013 and the lowest singular values is 6.8e−16.

Table 3 shows singular value decomposition multi-scale entropy *ENTs* of *DJIA* when scale s is set from −150 to 150. We can find that multi-scale entropy ENT_s monotonically increase with the increase of scale s, which means the index *DJIA* has the obvious characteristic of multi-scale entropy.

With further observation, we can find that *ENT* stays between 1.2 and 1.8 for every negative scale s, while it increases from 2.2548 to 29.7797 when s increases from 1 to 150. This indicates that singular value decomposition multi-scale entropy ENT_s of *DJIA* are insensitive to negative scale, but sensitive to positive scale. In other words, the multi-scale characteristic of ENT_s is only presented on the positive scale part. For negative s, entropy ENT_s mainly measure the amount of useful transitive information. For positive s, entropy ENT_s mainly measure the amount of transitive noise. Thus, the multi-scale characteristic of *DJIA* is mainly caused by the transmission of noise.

4. The predictive power of multi-scale entropy for *DJIA*

From the discussion above, we have known that the useful information and noise have different impacts on the multi-scale characteristic of singular value decomposition entropy of *DJIA*. In the following part, we will study the predictive power of useful information and noise on the *DJIA* respectively.

Table 4
Descriptive statistic test.

Series	Mean	Std. Dev.	Skewness	Kurtosis	Jarque–Bera	*P*-value
DJIA	9.0781	0.5041	−0.8154	2.6235	746.2235	0.0000
ENT^{y}_{-50}	−0.4186	0.3739	−0.7819	3.4111	696.4823	0.0000
ENT^{y}_{50}	2.8891	0.0018	0.0996	2.5669	60.5317	0.0000
ENT^{hy}_{-50}	−0.4201	0.3214	−0.8378	3.2784	783.7229	0.0000
ENT^{hy}_{50}	2.8891	0.0016	0.0329	2.6662	31.4316	0.0000
ENT^{q}_{-50}	−0.4280	0.3187	−0.7569	3.3468	431.963	0.0000
ENT^{q}_{50}	2.8891	0.0016	0.0729	2.3788	558.836	0.0000
ENT^{m}_{-50}	−0.4442	0.3657	−1.1625	4.7958	2381.966	0.0000
ENT^{m}_{50}	2.9125	0.0017	−0.2202	2.6924	79.3442	0.0000

4.1. Construction of singular value decomposition multi-scale entropy series

We compute singular value decomposition multi-scale entropy series employing moving time windows. In Caraiani [4], the width of moving time window was set as one year (about 252 observations). In order to make comparison, we set the moving time window to be half of one year (about 126 observations), one quarter of one year (about 63 observations) and one month (about 22 observations). Then we mark the window width as y (one year), hy (half of a year), q (one quarter of a year) and m (one month). Form Eq. (7), we can see that $ent(s)$ related to larger positive scale s reflects more noise and $ent(s)$ related to larger negative scale s reflects more useful information. In Table 3, we can see that the $ent(s)$ increases monotonously with the scale s. The differences between $ent(50)$ and $ent(-50)$ is large enough to differentiate noise and information. So, to simplify the measurement, we fixedly set scale s as −50 and 50.

For each width w in $\{y, hy, q, m\}$ and each scale s in $\{-50, 50\}$, we compute singular value decomposition multi-scale entropy of *DJIA* using moving time windows with formulas (4)–(8), which is written as ENT^{w}_{s}. With t denoting the ending date of the window, we can obtain singular value decomposition entropy series with width w and scale s, $ENT^{w}_{s}(t)$. For example, ENT^{y}_{s} (Jan. 1, 1977) denotes the singular value decomposition multi-scale entropy with scale s calculated in the one-year window from Jan. 2, 1976 to Jan. 1, 1977.

Through the singular value decomposition multi-scale entropy series calculated by moving time windows, we can observe the dynamic changes of amount and complexity of information in the market under different time scales.

4.2. Basic statistical analysis

We conduct the logarithmic processing to *DJIA* and these entropy series ENT^{w}_{s}, and we mark the logarithmic series as log(*DJIA*) and $\log(ENT^{w}_{s})$ representing *DJIA* and ENT^{w}_{s} respectively. Table 4 exhibits the descriptive statistics of these series.

It shows that all entropy series have smaller mean value than *DJIA*, the mean value of entropy series ENT^{w}_{s} is negative when $s = -50$ and positive when $s = 50$ for all time windows w in $\{y, hy, q, m\}$. All entropy series have smaller volatility than *DJIA*, and volatility of entropy series ENT^{w}_{-50} is significantly smaller than ENT^{w}_{50} for all time windows w. The entropy series ENT^{w}_{s} is left-skewed when $s = -50$ and right-skewed when $s = 50$ for time windows w in $\{y, hy, q\}$, while ENT^{m}_{s} are always left-skewed. The entropy series ENT^{w}_{-50} have kurtosis greater than three for all time windows w and other series have lower kurtosis. All series follow the non-normal distribution, since the null hypothesis of normal distribution is rejected at 1% significant level in Jarque–Bera test.

4.3. Stationarity test

We examine the stationarity of all series before doing Granger causality test, using the ADF unit root test. Table 5 presents the results of the ADF unit root test.

It is seen from Table 5 that all the entropy series ENT^{w}_{s} are stationary, because the null hypothesis of having unit root is rejected at 1% significant level by the ADF test with intercept or trend and intercept. However, the *DJIA* is integrated with one-order difference, because the null hypothesis of having unit root falls to be rejected at 10% significant level, but it can be rejected at 1% significant level with the first order difference.

4.4. Linear Granger causality test

4.4.1. Test method

Granger causality test is a useful tool in testing the predictive power of one economic variable on another one. Granger [13] investigated in the question that whether variant x causes y based on the prediction theory. The variant y

Table 5
Unit root test.

Series	Level			Difference		
	None	Intercept	Trend and intercept	None	Intercept	Trend and intercept
DJIA	2.2184	−1.6783	−2.0031	−60.447***	−60.517***	−6.5175***
ENT^{y}_{-50}	−3.6460***	−5.5987***	−5.7280***			
ENT^{y}_{50}	−0.1026	−4.3912***	−4.5054***			
ENT^{hy}_{-50}	−4.8759***	−8.3102***	−8.3393***			
ENT^{hy}_{50}	−0.1039	−7.0377***	−7.1313***			
ENT^{q}_{50}	−7.5341***	−13.293***	−13.455***			
ENT^{q}_{50}	0.0324	−11.578***	−11.794***			
ENT^{m}_{50}	−7.4245***	−20.925***	−21.679***			
ENT^{m}_{50}	−0.0550	−18.151***	−18.992***			

Notes: Values in the table are t-statistics. *, ** and *** denote that the statistic is significant at 10%, 5% and 1% level respectively.

cannot be caused by x, if

$$MSE\left[\hat{E}(Y_{t+j}|Y_t, Y_{t-1}, \ldots)\right] = MSE\left[\hat{E}(Y_{t+j}|Y_t, Y_{t-1}, \ldots, x_{t-1}, x_{t-2}, \ldots)\right] \tag{9}$$

for each $j = 1, 2, \ldots,$ where $MSE = \frac{1}{j}\sum_{k=1}^{j}\left(\hat{y}_{t+k} - y_{t+k}\right)^2$ is the mean-square error.

The vector auto-regression (VAR) model is a tool introduced by Sims [14,15] to capture the linear correlation among multiple time series. The following VAR (p) can be used to test Granger causality of two stationary series x_t and y_t:

$$x_t = a_1 + \sum_{j=1}^{p} \beta_{1j} x_{t-j} + \sum_{j=1}^{p} \gamma_{1j} y_{t-j} + a_{1t} \tag{10}$$

$$y_t = \alpha_2 + \sum_{j=1}^{p} \beta_{2j} x_{t-j} + \sum_{j=1}^{p} \gamma_{2} y_{t-j} + a_{2t}. \tag{11}$$

The a_{1t} and a_{2t} are two random disturbance terms, p is the optimal lag length obtained by the AIC criterion. If $\beta_{2j} = 0$ is established for all $j = 1, 2, \ldots, p$ in Eq. (11), then the null hypothesis that x_t does not Granger cause y_t is true.

It is noted that if any sequence between x_t and y_t is not stationary, then this method cannot work. Then Toda and Yamamoto [16] proposed an extensive VAR model which does not require the stability of time series in operation.

$$x_t = \alpha_1 + \sum_{j=1}^{p+d_{\max}} \beta_{1j} x_{t-j} + \sum_{j=1}^{p+d_{\max}} \gamma_{1j} y_{t-j} + a_{1t} \tag{12}$$

$$x_t = \alpha_2 + \sum_{j=1}^{p+d_{\max}} \beta_{2j} x_{t-j} + \sum_{j=1}^{p+d_{\max}} \gamma_{2j} y_{t-j} + a_{2t}. \tag{13}$$

The $d_{\max}$ is the maximal integration order of x_t and y_t. If x_t does not Granger cause y_t, then $\beta_{2i} = 0$ is established for all $j = 1, 2, \ldots, p$ in Eq. (13), which can be tested by the Wald coefficient test. For convenience, we call this test T–Y Granger causality test. As for the test that based on VAR model and is affected by the linear relations between variables, we call it linear Granger causality test.

4.4.2. Test result

From the ADF unit root test, we know that all entropy series ENT_s^w are stationary, but the index series *DJIA* is integrated with one-order difference. So, we will employ T–Y Granger causality test to study the causality from entropy to the index. The results are displayed in Table 6, where the optimal lag orders are ascertained by the AIC criterion.

We can find that for all time windows w, the singular value decomposition multi-scale entropy ENT_s^w are not the linear Granger causes of the *DJIA* for scale $s = -50$ or 50, as the null hypothesis that ENT_s^w do not Granger cause *DJIA* is not rejected at 10% significant level, no matter judging from F-statistic or Chi-sq statistic. This indicates that, useful information and noise do not have the predictive power on the *DJIA*.

It is noticed that linear Granger causality test is based on the assumption that there exist linear relation between variables. Bake and Brock [17] pointed out that the linear Granger causality test can occur deviations if the relation between variables is nonlinear. Therefore, it is necessary for us to further conduct nonlinear tests on the correlation between *DJIA* and entropy series ENT_s^w, including nonlinear Granger causality tests.

4.5. Nonlinear Granger causality test

4.5.1. Test method

A basic assumption in linear Granger causality test is that there exist linear correlations between variables [18]. However, the correlation between financial variables often turns out to be nonlinear due to the impact of financial crisis, policy change and other events [19]. If the nonlinearity is neglected, the result of linear Granger causality test may have significant deviations [20,21]. So Diks and Panchenko [22] introduced a new nonparametric test for nonlinear Granger causality and it can be stated as following:

Considering $\{x_t\}$ and $\{y_t\}$ are two stationary series, and $X_t^{l_x} = \left(x_{t-l_x+1}, \ldots, x_t\right)$, $Y_t^{l_y} = \left(y_{t-l_x+1}, \ldots, y_t\right)$ are two delay vectors, with $l_x, l_y \geq 1$. The null hypothesis that y_t cannot be nonlinear Granger cause of x_t means that the observations $X_t^{l_x}$ contain no additional information (except that in $Y_t^{l_y}$) about y_{t+1}.

$$H_0 : y_{t+1} \left| \left(X_t^{l_x}; Y_t^{l_y}\right) \sim y_{t+1} \right| Y_t^{l_y}. \tag{14}$$

We set $W_t = \left(X_t^{l_x}, Y_t^{l_y}, z_t\right)$ with $z_t = y_{t+1}$. Then formula (14) implies that the distribution of $\left(X_t^{l_x}, Y_t^{l_y}, z_t\right)$ is invariant. If we ignore the time index and suppose that $l_x = l_y = 1$, then the formula (14) implies that the distribution of z when giving $(x, y) = (u, v)$ is the same as the z when giving $y = v$. Considering the joint distribution, the formula (14) is reconstructed, the joint probability density function $f_{x,y,z}\,(u, v, w)$ and its marginal density function should satisfy the following equation:

$$\frac{f_{x,y,z}(u, v, w)}{f_y(v)} = \frac{f_{x,y}(u, v)}{f_y(v)} \cdot \frac{f_{y,z}(v, w)}{f_y(v)}. \tag{15}$$

Diks and Panchenko [22] pointed out that the null hypothesis implies:

$$q \equiv E\left[f_{x,y,z}(x, y, z) f_y(y) - f_{x,y}(x, y) f_{y,z}(y, z)\right] = 0. \tag{16}$$

Supposing $\hat{f}_W\,(W_i)$ is the local density estimator of a d_W-variate random vector W at W_i. It is defined as $\hat{f}_W\,(W_i) = (2\varepsilon_n)^{-dw}\,(n-1)^{-1} \sum_{j:j\neq i} I_{ij}^W$ where $I_{ij}^W = I\left(\left\|W_i - W_j\right\| < \varepsilon_n\right)$, with the indicator function $I(\cdot)$, the bandwidth ε_n, and the sample size n. Then, the following formula is the scaled sample version of q in (16):

$$T_n\,(\varepsilon_n) = \frac{n-1}{n\,(n-2)} \sum_i \left(\hat{f}_{x,y,z}\,(x_i, z_i, y_i)\,\hat{f}_y\,(y_i) - \hat{f}_{x,y}\,(x_i, y_i)\,\hat{f}_{y,z}\,(y_i, z_i)\right). \tag{17}$$

Diks and Panchenko [22] found that for $l_x = l_y = 1$, if $\varepsilon_n = C_n^{-\beta}$ $(C > 0, (1/4) < \beta < (1/3))$, then the Eq. (17) satisfies the following formula:

$$\sqrt{n}\frac{(T_n\,(\varepsilon_n) - q)}{S_n} \xrightarrow{D} N\,(0, 1) \tag{18}$$

where $\xrightarrow{D}$ denotes convergence in distribution and S_n is an estimator of the asymptotic variance of $T_n(\cdot)$. Adopting Diks and Panchenko's method, we implement a single-tailed test. The null hypothesis will be rejected if the left part of the formula (18) is too large. For convenience, we call this test D–P Granger causality test.

4.5.2. Test result

In order to analyze that whether nonlinear relation exists between *DJIA* and the entropy series ENT_s^w, we conduct linear filtrations for the *DJIA* and ENT_s^w using the VAR model. Then we examine that whether the residuals of VAR are independent and identically distributed (here after, i.i.d.) through BDS statistic. Table 7 presents the results of the BDS statistic test for residuals from VAR(1).

From Table 7, we can find that results of BDS statistic test to all residual series of *DJIA* significantly reject the null hypothesis at 1% level. The same results occur when the BDS statistic test is applied on residuals of ENT_s^w. This shows that, for each time window w and each scale s, there does exist nonlinear relation between entropy series ENT_s^w and index *DJIA*. Therefore, we can further detect the nonlinear causality relation between entropy series ENT_s^w and *DJIA*.

First, we directly conduct nonlinear causality between ENT_s^w and *DJIA*, using the D–P Granger causality test. Since ENT_s^w are stationary and *DJIA* is integrated with one-order difference, the D–P Granger causality test is applied on the series of ENT_s^w and the first-order difference of *DJIA*. The results are presented in Table 8.

Table 8 shows that, for scale $s = -50$, the null hypothesis that ENT_{-50}^{hy} does not nonlinear Granger cause *DJIA* is significantly rejected at 10% level for lag 1, and the null hypothesis is significantly rejected at 1% level for other lags,

Table 6
T–Y Granger causality test.

Null hypothesis	Lag	F-statistic (probability)	Chi-sq (probability)
ENT^{y}_{-50} does not Granger cause *DJIA*	6	1.0049 (0.4200)	6.0298 (0.4199)
ENT^{y}_{50} does not Granger cause *DJIA*	7	0.2966 (0.9555)	2.0765 (0.9555)
ENT^{hy}_{-50} does not Granger cause *DJIA*	6	1.0785 (0.3726)	6.4712 (0.3725)
ENT^{hy}_{50} does not Granger cause *DJIA*	8	0.4610 (0.8840)	3.6884 (0.8841)
ENT^{q}_{-50} does not Granger cause *DJIA*	4	1.0665 (0.3713)	4.2662 (0.3712)
ENT^{q}_{50} does not Granger cause *DJIA*	8	0.4297 (0.9039)	3.4383 (0.9039)
ENT^{m}_{-50} does not Granger cause *DJIA*	3	1.6945 (0.1659)	5.0837 (0.1658)
ENT^{m}_{50} does not Granger cause *DJIA*	6	0.6494 (0.6907)	3.8965 (0.6907)

Table 7
Nonlinearity test for VAR-residuals of *DJIA*.

VAR(1)	BDS statistic				
	m				
	2	3	4	5	6
DJIA ENT^{y}_{-50}	0.0175***	0.0396***	0.0565***	0.0678***	0.0734***
DJIA ENT^{y}_{50}	0.0174***	0.0396***	0.0564***	0.0678***	0.0734***
DJIA ENT^{hy}_{-50}	0.0175***	0.0396***	0.0564***	0.0676***	0.0731***
DJIA ENT^{hy}_{50}	0.0175***	0.0396***	0.0565***	0.0676***	0.0731***
DJIA ENT^{q}_{-50}	0.0176***	0.0396***	0.0563***	0.0675***	0.0731***
DJIA ENT^{q}_{50}	0.0176***	0.0396***	0.0563***	0.0675***	0.0731***
DJIA ENT^{m}_{-50}	0.0173***	0.0389***	0.0553***	0.0662***	0.0717***
DJIA ENT^{m}_{50}	0.0173***	0.0390***	0.0553***	0.0663***	0.0717***

Note: m is embedding dimension. *, ** and *** denote rejection of the null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 8
Nonlinear D–P Granger causality test for level series.

Null hypothesis	Level series			
	Lag			
	1	2	3	4
ENT^{y}_{-50} does not Granger cause *DJIA*	2.815 (0.002)	2.803 (0.002)	1.819 (0.034)	1.275 (0.101)
ENT^{y}_{50} does not Granger cause *DJIA*	2.105 (0.017)	1.955 (0.025)	1.048 (0.147)	0.828 (0.203)
ENT^{hy}_{-50} does not Granger cause *DJIA*	1.419 (0.077)	0.810 (0.208)	−0.332 (0.630)	−0.471 (0.681)
ENT^{hy}_{50} does not Granger cause *DJIA*	0.739 (0.229)	0.567 (0.285)	0.030 (0.487)	0.094 (0.462)
ENT^{q}_{-50} does not Granger cause *DJIA*	3.928 (0.000)	3.521 (0.000)	2.913 (0.001)	2.773 (0.002)
ENT^{q}_{50} does not Granger cause *DJIA*	3.515 (0.000)	2.955 (0.001)	2.018 (0.021)	1.713 (0.043)
ENT^{m}_{-50} does not Granger cause *DJIA*	4.199 (0.00)	3.608 (0.000)	2.621 (0.004)	2.005 (0.022)
ENT^{m}_{50} does not Granger cause *DJIA*	3.903 (0.000)	3.121 (0.000)	1.849 (0.032)	1.582 (0.056)

Note: 1. The index series is the differential series of *DJIA*. 2. The values in table are T-statistic, and values in brackets are p-value of associated statistics.

suggesting that the useful information has some predictive power on the *DJIA* index in the long-term (at least one year). However, for scale $s = 50$, the null hypothesis that ENT^{hy}_{50} does not nonlinear Granger cause *DJIA* cannot be rejected from lag 1 to 4 at 10% level, but other null hypothesis can be significantly rejected at 5% level for some lags. These unstable results may be caused by the deviations due to the difference of series.

Although some predictive power of useful information on the *DJIA* index has been found, the difference calculation may lead to the loss of some useful information in series, which may cause deviation in results. So we conduct the D–P Granger causality test for residuals of VAR in order to confirm the predictive power of singular value decomposition multi-entropy ENT^{w}_{s} on the *DJIA*. Here, we use VAR(1) to filter the linear relation between ENT^{w}_{s} and *DJIA*, then the D–P Granger causality test is applied on the residuals of VAR(1) for capturing the nonlinear causality relation between these two series. The results are presented in Table 9.

Table 9 shows that, for scale $s = 50$, only the null hypothesis that ENT^{m}_{50} does not nonlinear Granger cause *DJIA* is significantly rejected at 5% level for lag 1, 2 and 4 and at 10% level for lag 3, suggesting that the noise only has predictive power on the *DJIA* index in the short-term (about one month). However for scale $s = -50$, the null hypothesis that ENT^{hy}_{-50} does not nonlinear Granger cause *DJIA* cannot be rejected at 10% level for lag from 1 to 4, but other null hypothesis can

Table 9
Nonlinear Granger causality test for VAR(1)-residuals.

Null hypothesis	T-statistic (probability)			
	Lag			
	1	2	3	4
ENT^{y}_{-50} does not Granger cause *DJIA*	3.114 (0.000)	2.308 (0.010)	1.020 (0.153)v	0.593 (0.276)
ENT^{y}_{50} does not Granger cause *DJIA*	−2.426 (0.992)	−2.759 (0.997)	−2.938 (0.998)	−2.602 (0.995)
ENT^{hy}_{-50} does not Granger cause *DJIA*	0.388 (0.348)	0.504 (0.307)	−0.042 (0.517)	0.162 (0.435)
ENT^{hy}_{50} does not Granger cause *DJIA*	−1.017 (0.845)	−1.348 (0.911)	−1.262 (0.896)	−1.085 (0.861)
ENT^{q}_{-50} does not Granger cause *DJIA*	2.469 (0.006)	2.952 (0.001)	3.580 (0.000)	3.877 (0.000)
ENT^{q}_{50} does not Granger cause *DJIA*	−1.083 (0.860)	−0.517 (0.697)	−0.537 (0.704)	−0.220 (0.587)
ENT^{m}_{-50} does not Granger cause *DJIA*	3.003 (0.001)	2.810 (0.002)	2.679 (0.003)	2.317 (0.010)
ENT^{m}_{50} does not Granger cause *DJIA*	1.728 (0.041)	2.118 (0.017)	1.495 (0.067)	1.681 (0.046)

Table 10
Nonlinear Granger causality test for VAR(k)-residuals.

Null hypothesis	T-statistic (probability)			
	Lag			
	1	2	3	4
ENT^{y}_{-50} does not Granger cause *DJIA*	2.853(0.002)	1.876(0.030)	2.224(0.013)	1.645(0.049)
ENT^{y}_{50} does not Granger cause *DJIA*	0.157(0.437)	−0.038(0.515)	−0.772(0.780)	−0.490(0.687)
ENT^{hy}_{-50} does not Granger cause *DJIA*	2.254(0.012)	2.254(0.017)	2.342(0.009)	3.036(0.001)
ENT^{hy}_{50} does not Granger cause *DJIA*	−1.207(0.886)	01.519(0.908)	−1.109(0.866)	−0.974(0.835)
ENT^{q}_{-50} does not Granger cause *DJIA*	2.928(0.001)	2.986(0.001)	3.148(0.000)	3.592(0.001)
ENT^{q}_{50} does not Granger cause *DJIA*	−0.927(0.823)	−0.217(0.586)	−0.217(0.645)	−0.218(0.586)
ENT^{m}_{-50} does not Granger cause *DJIA*	2.143(0.016)	2.769(0.002)	1.873(0.030)	2.532(0.005)
ENT^{m}_{50} does not Granger cause *DJIA*	1.598(0.054)	1.247(0.106)	1.335(0.090)	1.686(0.045)

be rejected at 5% level for some lags. These unstable results are due to the linear relation contained in residuals series. Furthermore, we use VAR(k), where k is the optimal lag order, to filter linear relation between ENT^{w}_{s} and *DJIA*, then the D–P Granger causality test is applied on the residuals of VAR(k), The results are presented in Table 10.

According to the Table 10, we find that the test for scale $s = 50$ also shows that the noise only has predictive power on the *DJIA* index in the short-term (about one month). The null hypothesis that ENT^{w}_{-50} does not nonlinear Granger cause *DJIA* is significantly rejected at 5% level for lag from 1 to 4, suggesting that the useful information has the predictive power on the *DJIA* index in the long-term (at least one year).

5. Conclusion

Caraiani [4] previously investigated in the predictive power of singular value decomposition entropy on stock market. He found that the entropy has predictive power on the Dow Jones Industrial Average Index (*DJIA*), employing linear Granger causality test. Gu et al. [5] found that the predictive power of the entropy for the Shenzhen component index is affected by the structural breaks in the market when employing linear Granger causality test to the Chinese stock market, and the entropy has predictive power on the Shenzhen component index after the reform of non-tradable shares. On the one hand, Caraiani's singular value decomposition entropy is calculated by the Pearson correlation coefficient matrix which can only reflect simple (linear) correlation information between the price of stocks. On the other hand, stock markets always display some nonlinear characteristics. Therefore, some significant deviations could appear in the results obtained by linear Granger causality test [18,20,21].

We introduce a new concept of singular value decomposition multi-scale entropy. We make multi-scale entropy analysis for the *DJIA* index, and find that this index has multi-scale entropy characteristic, which is mainly caused by the transmission of the noise in the stock market.

Employing linear and nonlinear Granger causality analysis, we study the predictive power of the singular value decomposition multi-scale entropy on the *DJIA* index. From the perspective of linearity, useful information and noise do not have the predictive power on the *DJIA* index. However, from the perspective of nonlinearity, the useful information has the predictive power on the index in the long-term (at least one year), and noise only has the predictive power on the index in the short-term (about one month). This means that both useful information and noise have predictive power on stock index, but their capacity of predicting (predictive term) is different, and these predictive power are presented through nonlinear mechanism rather than the simple linear mechanism.

The notion of singular value decomposition multi-scale entropy proposed in this paper creatively describes the multi-scale characteristic of complex system from the perspective of informatics. Through the multi-scale entropy analysis on the *DJIA* index, we not only learn about the multi-scale characteristic but also find that the predictive power of useful information and noise for index have obvious differences. Our results not only verify the noise trading theory that noise exists in the market and can affect stock price, but also have reference value for investors of stock market, especially for noise traders.

Higher entropy implies the higher degree of market uncertainty. The existence of nonlinear Granger causality from entropy to stock return has some important implications for asset pricing. First, it suggests that the market uncertainty may be taken as the predictor of stock return. Forecasting stock return is notoriously difficult. It has been well documented that economic models perform worse in forecasting stock return. This finding indicates that the entropy measure may provide a potential way to resolve the problem of forecasting stock returns. Second, the predictability from entropy to stock return also implies the market inefficiency. Third, the existence of nonlinear causality signals that the relationship between risk and return is not linear, challenging the overwhelming idea of linear risk–return relationship in the area of modern finance.

Acknowledgments

The author was supported by the National Natural Science Foundation of China (71471081), the Humanities and Social Science Project of Ministry of Education of China (12YJAZH020), the Social Science Foundation of Jiangsu Province (15eyb013), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and the Project Funded by Jiangsu Modern Service Institute (PMS).

References

[1] E.F. Fama, The behavior of stock-market prices, J. Bus. 38 (1965) 34–105.
[2] E. Maasoumi, J. Racine, Entropy and predictability of stock market returns, J. Econometrics 107 (2002) 291–312.
[3] C. Eom, G. Oh, W.S. Jung, Relationship between efficiency and predictability in stock price change, Physica A 387 (2008) 5511–5517.
[4] P. Caraiani, The predictive power of singular value decomposition entropy for stock market dynamics, Physica A 393 (2014) 571–578.
[5] R.B. Gu, W. Xiong, X.J. Li, Does the singular value decomposition entropy have predictive power for stock market? — Evidence from the Shenzhen stock market, Physica A 439 (2015) 103–113.
[6] P. Caraiani, The predictive power of local properties of financial networks, Physica A 466 (2017) 79–90.
[7] R. Clausius, Über verschiedene für die anwendung bequeme formen der hauptgleichungen der mechanischen wärmetheorie, Ann. Phys. 125 (1865) 353–400.
[8] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[9] J.W. Kantelhart, Multifractal detrended fluctuation analysis of nonstationary time series, Physica A 316 (2002) 87–114.
[10] J. Shlens, A tutorial on principal component analysis, Cornell University Library, 2014, https://arxiv.org/abs/1404.1100.
[11] R.B. Gu, Y.M. Shao, How long the singular value decomposed entropy predicts the stock market?—Evidence from the Dow Jones Industrial Average Index, Physica A 453 (2016) 150–161.
[12] L. Andreadis, A. Serietis, Evidence of a random multifractal turbulent structure in the Dow Jones Industrial Average, Chaos Solitons Fractals 13 (2002) 1309–1315.
[13] C.W.J. Granger, Investigating causal relations by econometric models and cross spectral methods, Econometrica 37 (1969) 424–438.
[14] C.A. Sims, Money, income, and causality, Amer. Econ. Rev. 62 (1972) 540–552.
[15] C.A. Sims, Marcroeconomics and reality, Econometrica 48 (1980) 1–48.
[16] H.Y. Toda, T. Yamamoto, Statistical inference in vector autoregression with possibly integrated processes, J. Econometrics 66 (1995) 225–250.
[17] E. Baek, W.A. Brock, General test for nonlinear Granger causality: bivariate model, Working Paper, Iowa State University and University of Wisconsin-Madison, 1992.
[18] C. Hiemstra, J.D. Jones, Testing for linear and nonlinear Granger causality in the stock price-volume relation, J. Finance 49 (1994) 1639–1664.
[19] T.H. Lee, H. White, C.W.J. Granger, Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, J. Econometrics 56 (1993) 269–290.
[20] Y. Hong, H. Lin, C. Wu, Are corporate bond market returns predictable, J. Bank. Finance 36 (2012) 2216–2232.
[21] M. Billio, M. Getmansky, A.W. Lo, L. Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, J. Financ. Econ. 104 (2012) 535–559.
[22] C.G.H. Diks, V. Panchenko, A new statistic and practical guidelines for nonparametric Granger causality testing, J. Econom. Dynam. Control 30 (2006) 1647–1669.