Digital Signal Processing 63 (2017) 56–71

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Symbolic complexity of volatility duration and volatility difference component on voter financial dynamics

Rui Li *, Jun Wang

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

ARTICLE INFO

Article history:
Available online 4 January 2017

Keywords:
Volatility duration series
Volatility difference component
Voter financial price dynamics model
Symbolic sequence
Zipf distribution
Permutation Lempel–Ziv complexity

ABSTRACT

The nonlinear complexity of volatility duration and volatility difference component based on voter financial dynamics is investigated in this paper. The statistic – volatility difference component is first introduced in this work, in an attempt to study the volatility behaviors comprehensively. The maximum change rate series and the average change rate series (both derived from the volatility difference components) are employed to characterize the volatility duration properties of financial markets. Further, for the proposed series model and the proposed financial statistic series (which are transformed to symbolic sequences), the permutation Lempel–Ziv complexity, a novel complexity measure, is introduced to study the corresponding randomness and complexity behaviors. Besides, Zipf analysis is also applied to investigate the corresponding Zipf distributions of the proposed series. The empirical study shows the similar complexity behaviors of volatility between the proposed price model and the real stock markets, which exhibits that the proposed model is feasible to some extent.

1. Introduction

Recent researches on the financial field demonstrate that the financial market is a complex and dynamical system whose fluctuations often represent strong nonlinear and dynamical characteristics, and the interactions of financial market participants have attracted many financial researchers' attentions. Over the past ten years, many new interacting particle models have been proposed to study the financial markets [1–11], such as Bertrand and Cournot competitions in continuous time [1], reduced-form point process model [2], correlated default model [3] and so on. Many financial behaviors, including large pools of loan [4], portfolio losses [5–8], inter-bank lending and borrowing [9–11], are studied with these models. The stock market is an important part of financial markets, where there are some common properties called stock market stylized empirical facts, including fat tails, absence of autocorrelation, volatility clustering and so on [12]. In addition, with the governments' deregulation of stock markets all over the world, it is becoming a vital topic to capture the dynamics of the forward prices of stock markets in risk management, derivatives pricing and physical assets valuation. The modeling of stock markets, aiming at understanding price fluctuation dynamics, demands to establish a mechanism for the formation of the stock price. In the past years, considering the similarity between stock markets and physical systems, some scholars apply the statistical physics theories and methods to perform the empirical research on the stock markets [13–21]. Some agent-based interacting models from the percolation networks, the Potts dynamic system, etc. [13,14,18–21], have been established attempting to reproduce the complex and dynamical behaviors of stock markets. For example, Stauffer and Penna [14] developed a price model by the lattice percolation system and exhibited the existence of the fat tails for the return process; Hong and Wang [19] modeled the stock dynamics by the Potts model and explored the correlation of the logarithmic returns. The voter model, one of discrete agent-based models of opinion dynamics, is a stochastic interacting Markov process [20]. The voter process represents a voter's attitude affected by his neighbors' opinions at times distributed on a particular topic according to a stochastic rule [22–26]. Taking into account most of agents in the stock market trade stocks basing on their opinions to the investment information, we suppose that the interaction among the stock market agents is random, then utilize the voter interaction system to model the dynamics of the agents' opinions attempting to reproduce financial price fluctuations and volatility behaviors. Then we investigate the nonlinear phenomena of volatilities of the voter price model.

It is very important to understand the volatility behavior of financial markets, since it helps investors quantify the risk, optimize the portfolio and so on. The absolute returns, which is also called volatility series, is the key target for financial volatility be-

* Corresponding author.
E-mail address: lirui1401@bjtu.edu.cn (R. Li).

http://dx.doi.org/10.1016/j.dsp.2016.12.013
1051-2004/

havior research, and many new methods of volatility analysis are developed [27–29] for better understanding volatility behaviors, for example, the return interval analysis, characterizing the occurrence of volatilities within a certain range, are proposed in Refs. [27,28]. In this paper, we introduce a new statistic – volatility difference component (VDC) to characterize the volatility behaviors of the financial model in the volatility duration length, and develop a new method to explore the nonlinear phenomena of volatilities for the proposed price model, by transferring the volatility series to three kinds of volatility duration statistical series, the volatility duration series, the maximum change rate series and the average change rate series. The volatility duration series reflects local rising or falling volatility duration length, while the rest two series (which are derived from the volatility difference components) record the maximum change rate and the average change rate of volatility difference components in every volatility duration length. The exploration on these three series is useful for further understanding the volatility behaviors of financial markets and the proposed price model. In the empirical research, the randomness and complexity of the proposed volatility duration statistical series are investigated by introducing a novel approach – permutation Lempel–Ziv complexity (PLZC). The dynamic behaviors of volatilities are also studied by the Zipf analysis with different parameters of thresholds and timescales. The daily closing price data of Shanghai Stock Exchange (SSE) Composite Index and Shenzhen Stock Exchange (SZSE) Component Index are selected as the real empirical data for comparison.

2. Voter interacting financial price model

2.1. Voter interaction system

The stochastic voter interacting particle system was introduced independently by Clifford and Sudbury [22] and Holley and Liggett [23]. The evolution mechanism of the voter system starts with voters located at the nodes of lattice $\mathbb{Z}^d$, which might have one of two possible opinions on a political issue (in favor "1" or against "0") at independent exponential times. A voter reassesses his opinion by choosing a neighbor at random with certain probabilities and adopting his position. Let ξ_τ be the set of voters in favor, which is a continuous time Markov process. The dynamics of the process is specified by the collection of transition rates $c(x,\xi)$ [24–26]. For any $\xi \in \{0,1\}^{\mathbb{Z}^d}$, the state of $x \in \mathbb{Z}^d$ flips following the transition rates

$$0 \to 1 \text{ at rate } \lambda \sum_{y \in \mathbb{Z}^d} p(x,y)\, \mathbb{I}_{\{\xi(y)=1\}} \tag{1}$$

$$1 \to 0 \text{ at rate } \sum_{y \in \mathbb{Z}^d} p(x,y)\, \mathbb{I}_{\{\xi(y)=0\}} \tag{2}$$

where $\mathbb{I}$ is the indicator function, $p(x,y) \ge 0$ for $x, y \in \mathbb{Z}^d$, and $\sum_{y\in\mathbb{Z}^d} p(x,y) = 1$ for all $x \in \mathbb{Z}^d$. The transition probability $p(x,y)$ is translation invariant and symmetric, and the voter process with those transition probabilities is irreducible. If a node $x \in \mathbb{Z}^d$ is occupied by 1 (respectively, 0), then, at rate 1 (respectively, λ), it picks a node $y \in \mathbb{Z}^d$ with probability $p(x,y)$, then adopts the state of the voter at y. The stochastic dynamics of voter model ξ_τ on a configuration space $\{0,1\}^{\mathbb{Z}^d}$ is given as the form of generator by

$$\mathcal{A}g(\xi) = \sum_{x\in\mathbb{Z}^d} c(x,\xi)[g(\xi^x) - g(\xi)] \tag{3}$$

where the function g on $\{0,1\}^{\mathbb{Z}^d}$ depends on the finitely many coordinates, and $\xi^x(z) = \xi(z)$ if $z \ne x$, $\xi^x(z) = 1 - \xi(x)$ if $z = x$, for $x, z \in \mathbb{Z}^d$. In details: (i) if $x \in \xi_\tau$, then x becomes vacant at a rate equal to the number of vacant neighbors; (ii) if $x \notin \xi_\tau$, then x becomes occupied at a rate equal to λ times the number of occupied neighbors, where λ is an intensity, which is called the carcinogenic advantage in the voter process. When $\lambda = 1$, the model is called the voter model, and for $\lambda > 1$ it is called the biased voter model. Let $\xi_\tau^{\{A\}}$ denote the state at time τ with the initial state set $\xi_0^{\{A\}} = \{A\}$, and let $\xi_\tau^{\{0\}}(x)$ be the state of $x \in \mathbb{Z}^d$ at time τ with the initial state $\xi_0^{\{0\}} = \{0\}$, which means that only the original point $\{0\}$ of $\mathbb{Z}^d$ is occupied in the initial state (at $\tau = 0$) of the process $\xi_\tau^{\{0\}}$. More generally, the initial distribution is considered as υ_ρ, the product measure with density ρ (each node is independently occupied by probability ρ), and let $\xi_\tau^{\upsilon_\rho}$ be the voter model with initial distribution υ_ρ.

For the biased voter model ($\lambda > 1$), there is a "critical value" for the process on $\Omega = \{0,1\}^{\mathbb{Z}^d}$, the critical value λ_c is defined as [13,14]

$$\lambda_c = \inf\{\lambda : P(|\xi_\tau^{\{0\}}| > 0, \text{ for all } \tau \ge 0) > 0\} \tag{4}$$

where $|\xi_\tau^{\{0\}}|$ is the cardinality of $\xi_\tau^{\{0\}}$. Suppose $\lambda > \lambda_c$, then there is convex set C so that on $\Omega_\infty = \{\xi_\tau^{\{0\}} \ne \emptyset, \text{ for all } \tau\}$, for any $\epsilon > 0$ and for all τ sufficiently large

$$(1-\epsilon)\tau C \cap \mathbb{Z}^d \subset \xi_\tau^{\{0\}} \subset (1+\epsilon)\tau C \cap \mathbb{Z}^d. \tag{5}$$

If $\lambda \le \lambda_c$, for some positive $\gamma(\lambda)$, then

$$P(\xi_\tau^{\{0\}} \ne \emptyset) \le e^{-\gamma(\lambda)\tau}. \tag{6}$$

The above theory shows that, on d-dimensional lattice, the process becomes vacant exponentially for $\lambda < \lambda_c$, and survives with positive probability for $\lambda > \lambda_c$.

2.2. Construction of financial price model

The financial price dynamics based on the voter process is formulated as follows. Suppose that the investment information leads to the fluctuation of a stock price, and there are three kinds of information including buying, selling and neutral, which classify the investors into their corresponding groups. Assume that each trader can trade the stock several times at each day $t \in \{1, 2, \cdots, N\}$, but at most, one unit number of the stock at each time. Let l be the time length of one trading day, we denote the stock price at time τ in the tth trading day by $P_t(\tau)$, where $\tau \in [0, l]$. Suppose that the stock market is made up of $2m+1$ (m is large enough) invertors, who are located in a line $\{-m, \cdots, -1, 0, 1, \cdots, m\} \subset \mathbb{Z}$ (similarly for a d-dimensional lattice $\mathbb{Z}^d$). At the starting of each trading day, only the investor at the origin site "0" receives some information. And a random variable ζ_t with values 1, −1, 0 represents that this investor holds buying opinion, selling opinion or neutral opinion with probabilities p_1, p_{-1} or $1 - p_1 - p_{-1}$ respectively. Then this investor sends bullish, bearish or neutral signal to his nearest neighbors. According to the voter dynamical system, investors can affect each other or the information can be disseminated, which is considered as the main factor of price fluctuations for the stock market.

For a trading day $t \le N$ and $\tau \in [0, l]$, let

$$B_t(\tau) = \zeta_t \times \frac{|\xi_\tau^{\{0\}}|}{2m+1}, \quad \tau \in [0, l] \tag{7}$$

where $|\xi_\tau^{\{0\}}| = \sum_{w=-m}^{m} \xi_\tau^{\{0\}}(w)$. The stock price process at tth trading day is given as [30,31]

$$P_t(\tau) = e^{\alpha_t B_t(\tau)} P_{t-1}(\tau), \quad \tau \in [0, l] \tag{8}$$

$$P_t(\tau) = P_0 e^{\sum_{i=1}^{t} \alpha_i B_i(\tau)}, \quad \tau \in [0, l] \tag{9}$$

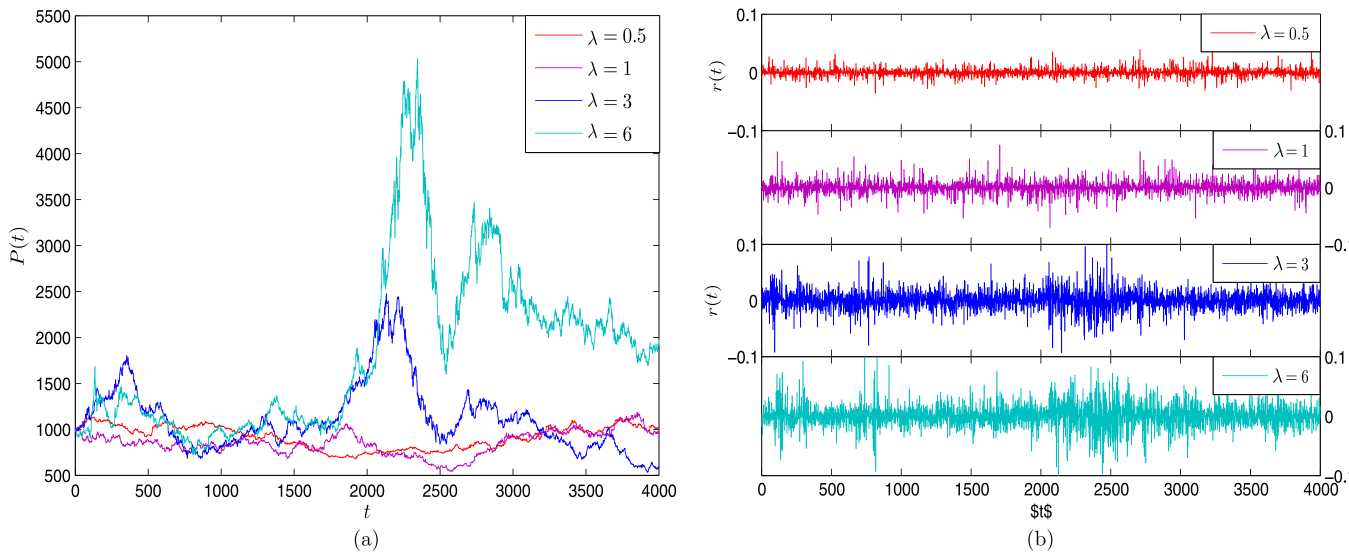

Fig. 1. (a) Fluctuation plots of prices and (b) returns respectively for the proposed model with different values of λ.

where $\alpha_t > 0$ denotes the depth parameter of the market, and P_0 is the initial stock price at time 0. The corresponding logarithmic returns defined as following

$$r(t) = \ln P_t(\tau) - \ln P_{t-1}(\tau), \quad \tau \in [0, l] \text{ and } t = 1, 2, \cdots, N. \quad (10)$$

From the dynamics of the voter process, if $\lambda > \lambda_c$, the investment information will be disseminated widely, so this will affect the investors' opinions, and at last will affect the fluctuation of the stock price. If $\lambda \le \lambda_c$, the influence on the stock price by the investors is limited. For the above price model on $\mathbb{Z}$, its critical value is $\lambda_c = 1$. Fig. 1 presents the plots of fluctuations of prices and the corresponding returns for four groups of the simulation data with $\lambda \in \{0.5, 1, 3, 6\}$. There is an evident that $r(t)$ shows significant volatility clustering behaviors for the simulation data with $\lambda > \lambda_c$, while for the simulation data with $\lambda \in \{0.5, 1\}$, their volatility clustering is not significant in Fig. 1(b). Besides, the volatility clustering of the simulation data with $\lambda = 6$ is more significant than that of the simulation data with $\lambda = 3$. In the following, we mainly investigate the volatility behaviors with $\lambda > \lambda_c$, since their return series show relatively significant volatility clustering behaviors.

2.3. Volatility duration and volatility difference component

The investigation on the volatility behaviors of financial markets is a crucial topic in financial research. The return interval analysis which analyzes the return interval between the daily volatilities of price changes is one of the methods applied in the volatility analysis. Refs. [27–29] explored the distribution function scales with mean return interval by this method. Inspired by the return interval analysis, we consider the duration of stock volatilities consistently above or below a given data point in the volatility series [32], and some quantity relationships which are worth to be taken into account in the duration period of time. Then we introduce three volatility duration statistical series derived from the volatility series to characterize the financial volatility behaviors by embodying the intensity–duration–quantity relationship in the volatility series.

We begin by generating the volatility duration length series $I(t)$ of $|r(t)|$ at day t. At trading day t, if $|r(t+1)| - |r(t)| > 0$, we say the volatility series is locally running up at t. On the opposite, if $|r(t+1)| - |r(t)| < 0$, we say the volatility series displays locally sliding down trend at t. $I(t)$ is set to record the duration length of the local trend of volatility intensity, defined as follows

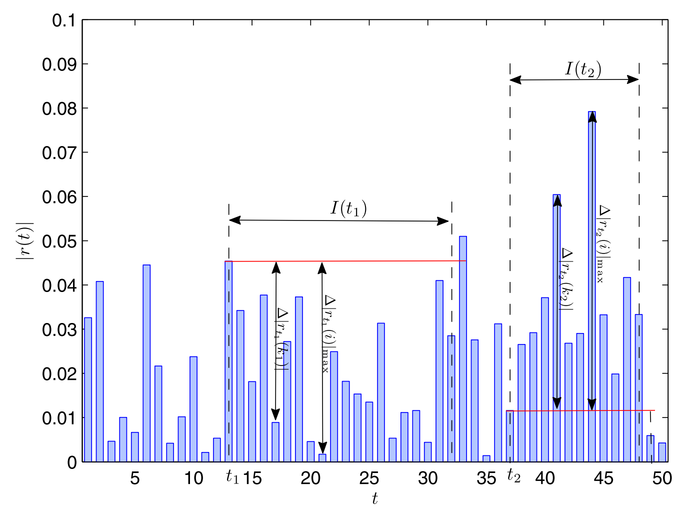

Fig. 2. Illustrations of $I(t)$, $\Delta|r_t(i)|$ and $\Delta|r_t(i)|_{\max}$ in the volatility series.

$$I(t) = \begin{cases} \max\{\tau : |r(t+i)| > |r(t)|, \text{ for } i \le \tau\}, \\ \quad \text{if } |r(t+1)| - |r(t)| > 0 \\ \max\{\tau : |r(t+i)| < |r(t)|, \text{ for } i \le \tau\}, \\ \quad \text{if } |r(t+1)| - |r(t)| < 0. \end{cases} \quad (11)$$

For $|r(t+1)| = |r(t)|$, let $I(t) = 0$. An illustration of $I(t)$ is presented in Fig. 2. Then, the volatility duration series $D(t)$ $(t = 1, 2, \cdots, T)$ of $|r(t)|$, combining the local volatility trend and duration length of volatility intensity, is defined as follows

$$D(t) = \text{sign}(|r(t+1)| - |r(t)|) \times \sqrt{I(t)}. \quad (12)$$

From the above definition, $D(t)$ reflects the time length of volatility series's local rising or falling duration at day t.

Albeit $D(t)$ characterizes the volatility series's local rising or falling duration time length, there are some other interesting statistics in accompany with the duration time $I(t)$. In this paper, we develop a new statistic, which is called volatility difference component (VDC) in the duration length, in an attempt to investigate the volatility behaviors comprehensively. Let $\Delta|r_t(i)|$ denote the volatility difference component at time i in the duration time $I(t)$ at day t, which is defined as

$$\Delta|r_t(i)| = \Big| |r(t+i)| - |r(t)| \Big|, \quad 1 \le i \le I(t). \quad (13)$$

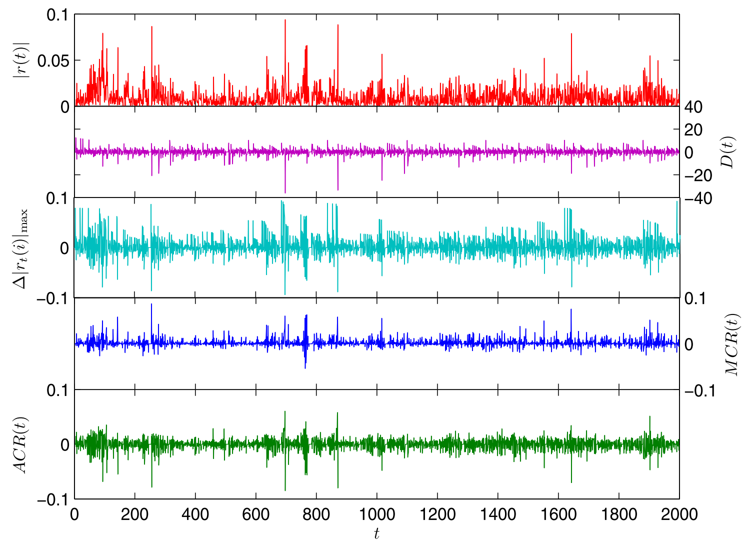

Fig. 3. Plots of $|r(t)|$, $D(t)$, $\Delta|r_t(i)|_{\max}$, $MCR(t)$ and $ACR(t)$ for the SSE.

And let $\Delta|r_t(i)|_{\max} = \max\{\Delta|r_t(i)|, 1 \le i \le I(t)\}$ represent the largest volatility difference component which is also the maximum drawdown/drawup in the duration length $I(t)$. The illustrations of $\Delta|r_t(i)|$ and $\Delta|r_t(i)|_{\max}$ are also shown in Fig. 2, and the volatility series and the $\Delta|r_t(i)|_{\max}$ series of the SSE are present in Fig. 3. Then, we consider two kinds of change rates of volatility difference components, the maximum change rate and the average change rate of the volatility series in the duration length $I(t)$ at day t. The maximum change rate series ($MCR(t)$) is defined as follows

$$MCR(t) = \text{sign}(|r(t+1)| - |r(t)|) \times \frac{\Delta|r_t(i)|_{\max}}{\min\{i: \ \Delta|r_t(i)| = \Delta|r_t(i)|_{\max}, 1 \le i \le I(t)\}}. \quad (14)$$

The other rate, the average change rate series ($ACR(t)$) is defined by

$$ACR(t) = \text{sign}(|r(t+1)| - |r(t)|) \times \frac{1}{I(t)} \sum_{i=1}^{I(t)} \Delta|r_t(i)|. \quad (15)$$

Note that, when $I(t) = 0$, we set $MCR(t) = 0$ and $ACR(t) = 0$. According to the above definitions, $MCR(t)$ accompanying with the information of maximum drawdown/drawup $\Delta|r_t(i)|_{\max}$ reflects the speed of the volatility series rising to the peak value or falling to the bottom value in the duration length time at day t, while $ACR(t)$ records the average change rate of VDC during the local rising or falling time. Three kinds of volatility duration statistical series $D(t)$, $MCR(t)$ and $ACR(t)$, which are also shown in Fig. 3, reflect different properties of the volatility series accompanying with the volatility duration, so the research on them is helpful to take a further step to understand the volatility behaviors in financial markets. In the following sections, we will mainly explore statistical and complex properties of these three time series for the real data and the simulation data.

3. Zipf distribution for volatility duration

3.1. Symbolic dynamic by Zipf analysis

Zipf analysis, originally introduced in the context of natural languages by George Kingsley Zipf [33,34], is one of the methods applied in bibliometrics. This technique is processed by counting the frequency of occurrence of each word in a given text, and reveals that the frequency of occurrence of each word f and its symbol ranking R display a power law, i.e., $f \sim R^{-\omega}$, for any natural language [35]. Recently, Zipf analysis has been applied to various area of physical and social sciences as a tool for quantifying time series symbolic complexity [36–39]. The core matter of Zipf analysis applied in time series analysis is based on converting a given time series into a symbol sequence. In this section, we covert the three volatility duration statistical series into 3-alphabet sequence, then explore the frequencies of each symbol and the corresponding symbolic dynamics of the volatility behaviors for the real data and simulation data by Zipf analysis.

We start with the definition of k-return series of stock prices in this part. The k-return series is obtained by the following definition

$$r_k(t) = \ln P(t+k) - \ln P(t), \quad t = 1, 2, \cdots, N-k \quad (16)$$

where $P(t)$ $(t = 1, 2, \cdots, N)$ is the daily closing price series and the parameter k is the timescale. For $k \in \{1, 5, 20, 60, 250\}$, k is called the characteristic timescale, it approximately stands for one transaction day, one transaction week, one transaction month, one transaction quarter and one transaction year, respectively, in terms of business time units (with weekends and holidays eliminated). For the symbolic conversion of original series, the extended random 3-alphabet conversion is applied to the three volatility duration statistical series $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of the k-volatility series $|r_k(t)|$. For the time series $s_k(t)$ $(t = 1, 2, \cdots, N-k)$, which is one of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$, there exists a symbolic sequence $y(k, \theta, s_k(t))$ obtained according to the following formula

$$y(k, \theta, s_k(t)) = \begin{cases} \text{u}, & \text{if } s_k(t) \ge \theta \\ \text{s}, & \text{if } |s_k(t)| < \theta \\ \text{d}, & \text{if } s_k(t) \le -\theta \end{cases} \quad (17)$$

where "u", "s" and "d" denote "sequence-up", "sequence-stable" and "sequence-down" respectively in $s_k(t)$. θ (the variation threshold) is a nonnegative random variable on a probability space with the probability distribution function $F_\theta(x)$. On the basis of the above conversion rule, we obtain the corresponding symbolic sequences $y(k, \theta, D_k(t))$, $y(k, \theta, MCR_k(t))$ and $y(k, \theta, ACR_k(t))$ of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$, respectively. Now, we compute the absolute frequency and the relative frequency to analyze the statistical dynamics of their symbolic sequences. Let $n_{\text{u}}(k, \theta, s_k(t))$, $n_{\text{s}}(k, \theta, s_k(t))$ and $n_{\text{d}}(k, \theta, s_k(t))$ denote the number of occurrences for "sequence-up", "sequence-stable" and "sequence-down" in $s_k(t)$ respectively. The corresponding absolute frequencies of symbolic sequence $y(k, \theta, s_k(t))$ are given as follows

Table 1

		SSE					Simulation data with $\lambda = 3$				
		$k=1$	$k=5$	$k=20$	$k=60$	$k=250$	$k=1$	$k=5$	$k=20$	$k=60$	$k=250$
(a) The numbers of "u" of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$											
$D(t)$	$\theta = 2.5$	480	559	767	802	844	481	543	690	755	793
	$\theta = 5$	136	158	266	379	424	134	149	211	347	394
	$\theta = 7.5$	54	60	102	200	271	49	63	84	201	239
	$\theta = 10$	23	28	52	78	196	17	37	42	76	148
$MCR(t)$	$\theta = 0.005$	1124	1363	1430	1432	1345	1079	1383	1437	1433	1320
	$\theta = 0.01$	622	774	824	742	776	531	841	870	835	796
	$\theta = 0.015$	354	449	458	407	405	288	500	504	491	441
	$\theta = 0.02$	209	266	251	226	221	163	300	286	288	277
$ACR(t)$	$\theta = 0.05$	11	65	243	412	484	2	49	166	352	404
	$\theta = 0.1$	0	1	48	131	240	0	0	23	108	185
	$\theta = 0.15$	0	0	6	48	166	0	0	4	19	94
	$\theta = 0.2$	0	0	0	11	121	0	0	0	3	33
(b) The numbers of "d" of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$											
$D(t)$	$\theta = 2.5$	505	653	744	780	787	529	635	704	779	798
	$\theta = 5$	157	216	319	364	436	173	214	274	386	392
	$\theta = 7.5$	78	110	147	205	306	89	109	141	228	236
	$\theta = 10$	48	61	84	131	225	59	70	86	134	176
$MCR(t)$	$\theta = 0.005$	487	592	566	595	563	456	595	585	582	580
	$\theta = 0.01$	195	256	260	270	283	148	256	279	293	277
	$\theta = 0.015$	80	107	133	133	144	39	118	134	158	148
	$\theta = 0.02$	26	54	76	75	82	16	59	72	89	80
$ACR(t)$	$\theta = 0.05$	26	117	241	357	463	9	87	193	349	388
	$\theta = 0.1$	0	13	61	128	237	0	7	39	125	182
	$\theta = 0.15$	0	1	14	56	165	0	0	10	53	111
	$\theta = 0.2$	0	0	9	27	126	0	0	4	21	69
(c) The numbers of "s" of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$											
$D(t)$	$\theta = 2.5$	3014	2783	2469	2316	2161	2989	2817	2586	2387	2178
	$\theta = 5$	3706	3621	3395	3197	2890	3692	3632	3495	3207	2964
	$\theta = 7.5$	3867	3825	3731	3535	3173	3861	3823	3755	3511	3275
	$\theta = 10$	3923	3911	3844	3731	3329	3923	3888	3852	3730	3426
$MCR(t)$	$\theta = 0.005$	2388	2040	1984	1913	1842	2464	2017	1958	1925	1850
	$\theta = 0.01$	3182	2965	2896	2928	2691	3320	2898	2831	2812	2677
	$\theta = 0.015$	3565	3439	3389	3400	3201	3672	3377	3342	3291	3161
	$\theta = 0.02$	3764	3675	3653	3639	3447	3820	3636	3622	3563	3393
$ACR(t)$	$\theta = 0.05$	3962	3813	3496	3171	2803	3988	3859	3621	3239	2958
	$\theta = 0.1$	3999	3981	3871	3681	3273	3999	3988	3918	3707	3383
	$\theta = 0.15$	3999	3994	3960	3836	3419	3999	3995	3966	3868	3545
	$\theta = 0.2$	3999	3995	3971	3902	3503	3999	3995	3976	3916	3648

$$f_{\mathrm{u}}(k,\theta,x,s_k(t)) = \frac{n_{\mathrm{u}}(k,\theta,s_k(t))}{N-k} \times \frac{1-F_\theta(x)}{2} \tag{18}$$

$$f_{\mathrm{d}}(k,\theta,x,s_k(t)) = \frac{n_{\mathrm{d}}(k,\theta,s_k(t))}{N-k} \times \frac{1-F_\theta(x)}{2} \tag{19}$$

$$f_{\mathrm{s}}(k,\theta,x,s_k(t)) = \frac{n_{\mathrm{s}}(k,\theta,s_k(t))}{N-k} \times F_\theta(x) \tag{20}$$

where $n_{\mathrm{u}}(k,\theta,s_k(t)) + n_{\mathrm{s}}(k,\theta,s_k(t)) + n_{\mathrm{d}}(k,\theta,s_k(t)) = N-k$, and $F_\theta(x) = P(\theta \le x)$. Here, x is the expected threshold of $s_k(t)$ for the investor, and $F_\theta(x) = P(\theta \le x)$ is the probability that he can bear the expected volatility in $s_k(t)$. $1 - F_\theta(x)$ is the probability of the investing risk (the investor has) when the volatility of $s_k(t)$ exceeds the max expected threshold. The corresponding relative frequencies are given as

$$g_{\mathrm{u}}(k,\theta,x,s_k(t)) = \frac{n_{\mathrm{u}}(k,\theta,s_k(t))}{n_{\mathrm{u}}(k,\theta,s_k(t)) + n_{\mathrm{d}}(k,\theta,s_k(t))} \times (1 - F_\theta(x)) \tag{21}$$

$$g_{\mathrm{d}}(k,\theta,x,s_k(t)) = \frac{n_{\mathrm{d}}(k,\theta,s_k(t))}{n_{\mathrm{u}}(k,\theta,s_k(t)) + n_{\mathrm{d}}(k,\theta,s_k(t))} \times (1 - F_\theta(x)) \tag{22}$$

In this part, we let θ be a uniform distribution and x equal to the mean value of θ. According to the numeric ranges of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$, θ is on interval $(0, 50)$ and $x = 25$ for $D_k(t)$, θ is on $(0, 0.1)$ and $x = 0.05$ for $MCR_k(t)$, and θ is on $(0, 1)$ and $x = 0.5$ for $ACR_k(t)$. We denote the corresponding frequency functions as $f_{\mathrm{u}}(k,\theta,s_k(t))$, $f_{\mathrm{d}}(k,\theta,s_k(t))$, $f_{\mathrm{s}}(k,\theta,s_k(t))$, $g_{\mathrm{u}}(k,\theta,s_k(t))$ and $g_{\mathrm{d}}(k,\theta,s_k(t))$ respectively for convenience. In this following section, we will mainly investigate the statistical behaviors of frequency functions of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ for different timescale k and threshold θ, and draw a parallel of symbolic dynamics of three volatility duration series between the real data and the simulation data by Zipf analysis.

3.2. Empirical Zipf analysis

We calculate the Zipf distributions of absolute frequency and relative frequency functions of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of k-volatility series for the SSE and the simulation data with $\lambda = 3$ for the various values of θ and $k = 1, 5, 20, 60, 250$. The numbers of "u", "d" and "s" in the symbolic sequences of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ with $k = 1, 5, 20, 60, 250$ are also counted for some fixed values of θ in Table 1. At first, we focus on the "sequence-up" frequency function in Fig. 4. It is evident to find that "sequence-up" frequency functions are undergoing exponent decrease as θ increases for both the real data and the simulation data, which indicates that there will be fewer "sequence-up" occurring in

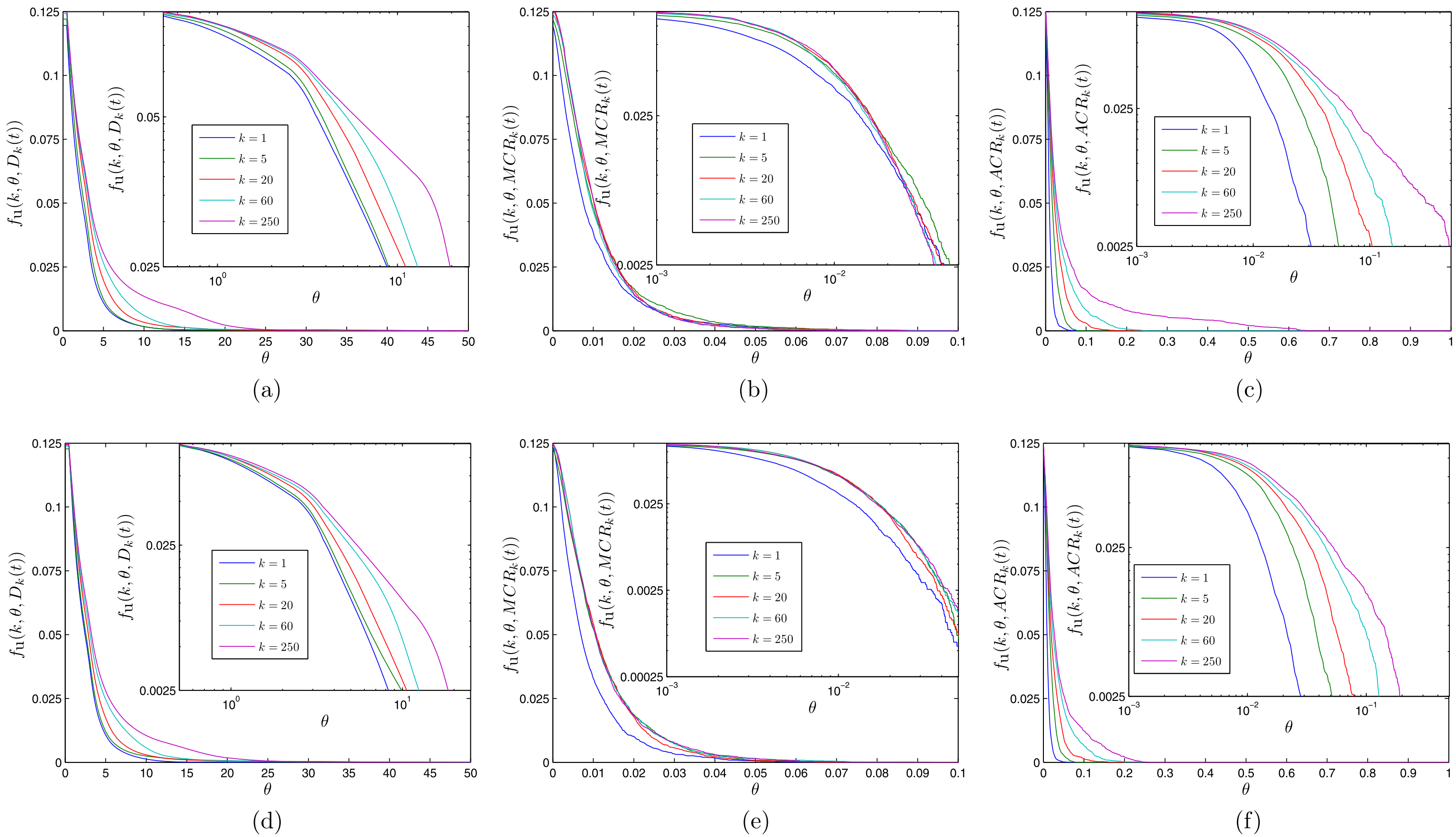

Fig. 4. (a)–(c) Plots of $f_{\mathrm{u}}(k, \theta, D_k(t))$, $f_{\mathrm{u}}(k, \theta, MCR_k(t))$ and $f_{\mathrm{u}}(k, \theta, ACR_k(t))$ for SSE; (d)–(f) Plots of $f_{\mathrm{u}}(k, \theta, D_k(t))$, $f_{\mathrm{u}}(k, \theta, MCR_k(t))$ and $f_{\mathrm{u}}(k, \theta, ACR_k(t))$ for the simulation data with $\lambda = 3$. The inner figure is the corresponding log–log plot.

$D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$. Moreover, for $f_u(k,\theta,D_k(t))$ and $f_u(k,\theta,ACR_k(t))$ of the SEE and the simulation data in Figs. 4(a), (c), (d), (f), the figures display significant distinct decaying traits for decreasing timescale k, i.e., the curve with the larger timescale k lies above the one with the smaller value k. For $\theta = 2.5, 5, 7.5, 10$ in Table 1(a), the numbers of "u" in the symbolic sequences of $D_k(t)$ and $ACR_k(t)$ with a large k are more than that with a small k. So for fixed θ, the increase of timescale k will lead more "sequence-up" happening in $D_k(t)$ and $ACR_k(t)$. Although the curve of $f_u(k,\theta,MCR_k(t))$ with $k = 1$ almost lies below the rest curves in Figs. 4(b), (e), the curves except the one with $k = 1$ are very close when θ increases. Besides, the numbers of "u" in the symbolic sequences of $MCR_k(t)$ with $k > 1$ are closed to each other for fixed θ in Table 1(a). Thereby, there are not obvious distinct decaying traits of $f_u(k,\theta,MCR_k(t))$ for decreasing k in Figs. 4(b), (e).

As for the "sequence-down" and "sequence-up" frequency functions of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of the SSE and the simulation data, the curves in Fig. 5 show the similar dynamic behaviors with those in Fig. 4, while the curves in Fig. 6 exhibit the opposite dynamic behaviors as θ and k vary in the comparison with Fig. 4. Therefore, for both the real data and the simulation data, $f_d(k,\theta,D_k(t))$, $f_d(k,\theta,MCR_k(t))$ and $f_d(k,\theta,ACR_k(t))$ decrease to 0 exponentially with the increase of θ, and the plots of $f_d(k,\theta,D_k(t))$ and $f_d(k,\theta,ACR_k(t))$ show distinct decaying traits for decreasing k, while the distinct decaying traits of $f_d(k,\theta,MCR_k(t))$ are not significant. On the contrary, the "sequence-stable" frequency functions of $D_k(t)$, $MCR_k(t))$ and $ACR_k(t))$ increase exponentially to 0.5 with the increase of threshold θ, and there exit distinct decaying traits for the increase of timescale k in Figs. 6(a), (c), (d), (f), while the distinct decaying traits for increasing k in Figs. 6(b), (e) are not evident.

In a word, for both the real data and the simulation data, "sequence-up" and "sequence-down" in $D_k(t)$, $MCR_k(k)$ and $ACR_k(t)$ happen less when the threshold θ increases. Whereas, "sequence-stable" in the three series occurs more with the increase of θ. In addition, for the given threshold θ, the increase of timescale k will make "sequence-up" and "sequence-down" happen less but lead "sequence-stable" happening more in $D_k(t)$ and $ACR_k(t)$ of $|r_k(t)|$. However, the increase of k does not have significant effect on the absolute frequency functions of $MCR_k(t)$. Furthermore in Figs. 4–6, the plots of the simulation data are very similar with the corresponding plots of the real data when threshold θ increases. Thus, the dynamic behaviors of absolute frequency functions of $D_k(t)$, $MCR_k(k)$ and $ACR_k(t)$ of $|r_k(t)|$ of the simulation data are analogous to those of the real date for the changes of timescale k and threshold θ.

Next, we study the properties of relative frequency functions of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ when threshold θ varies for the SSE and the simulation data with different timescales. Fig. 7 demonstrates the plots of relative frequency functions versus $\log(\theta)$ with $k = 1, 5, 20, 60, 250$. For the $D_k(t)$ in Figs. 7(a), (d), it is evident to find that there exist two phases as the relative frequency functions evolve with θ for all $k = 1, 5, 20, 60, 250$. Within the first phase when θ is relatively small, both of $g_u(k,\theta,D_k(t))$ and $g_d(k,\theta,D_k(t))$ are approximately almost equal to 0.25 (since x is set to be the mean value of θ, the maximum value of the relative frequencies is equal to 0.5 in this paper), which means that "sequence-up" and "sequence-down" in $D_k(t)$ occur with almost equal probability. With the increase of θ, a significant deviation of two relative frequency functions of $D_k(t)$ comes out. $g_d(k,\theta,D_k(t))$ taking advantage over $g_u(k,\theta,D_k(t))$ increases to 0.5 at relative large θ for any timescale k, while $g_u(k,\theta,D_k(t))$ shows opposite dynamic behaviors. Since θ is a psychological threshold of the investors' expected duration, investors are willing to trade stocks actively leading relatively high volatility intensity when θ is less than the max expected threshold. So "u" and "d" in $D_k(t)$ happen with almost equal probability. When θ exceeds the investors' max expected threshold, investors will face with relatively high investing risk, which causes them not to actively participate in stock trading. The volatility intensity will decrease while the length of the volatility series's local falling duration will increase, that is, there are more "d" occurring than 'u' in $D_k(t)$. Therefore, $g_d(k,\theta,D_k(t))$ becomes greater than $g_u(k,\theta,D_k(t))$ increasing to 0.5 at relatively large θ. For the $MCR_k(t)$ in Figs. 7(b), (e), there also exist significant deviations of $g_u(k,\theta,MCR_k(t))$ and $g_d(k,\theta,MCR_k(t))$, but the deviation of relative frequency functions of $MCR_k(t)$ presents opposite properties to that of $D_k(t)$, i.e., $g_u(k,\theta,MCR_k(t))$ is bigger than $g_d(k,\theta,MCR_k(t))$ as θ increases for all timescale k. This indicates that there are more "sequence-up" occurring in the maximum change rate of k-volatility series during the increase of θ.

For the rest series $ACR_k(t)$ of the SSE and the simulation data, its relative frequencies show the similar behaviors with those of $D_k(t)$ in Figs. 7(c), (f). $g_u(k,\theta,ACR_k(t))$ and $g_d(k,\theta,ACR_k(t))$ are almost equal at relative small θ, then the deviation of them comes out at relative large θ. Besides, the deviation of the relative frequencies of $ACR_k(t)$ for small timescale k is more significant than that for large k, for example, $g_u(k,\theta,ACR_k(t))$ and $g_d(k,\theta,ACR_k(t))$ for $k = 250$ are almost equal when θ increases in Figs. 7(c), (f). In fact, $|r_k(t)|$ is the annual volatility series when $k = 250$ and it has relatively large fluctuation range, so that the interval of θ (on $(0,1)$) may not be enough to the relative frequency function of $ARC_k(t)$ with $k = 250$. We set the interval of θ as $(0,10)$ for $ARC_k(t)$ with $k = 250$ and calculate its relative frequency functions with θ increasing from 0 to 10. Fig. 8 presents the plots of relative frequency functions of $ARC_k(t)$ with $k = 250$ for the SSE and the simulation data. The figures of $g_u(k,\theta,ACR_k(t))$ and $g_d(k,\theta,ACR_k(t))$ with $k = 250$ show similar behaviors with those with $k = 1$ by comparing Figs. 7(c), (f) and Figs. 7(a), (b), that is, a significant deviation of two relative frequency functions of $ACR_k(t)$ with $k = 250$ comes out with the increase of θ from 0 to 10, and $g_d(k,\theta,ACR_k(t))$ also taking advantage over $g_u(k,\theta,ACR_k(t))$ increases to 0.5 at relative large θ, while $g_u(k,\theta,ACR_k(t))$ shows opposite dynamic behavior. Besides, in the comparison of the plots of the real data and the simulation data in Figs. 7–8, the relative frequency functions of three series of k-volatility series of the simulation data show the similar dynamic behaviors with those of the real data during the increase of θ.

4. Permutation Lempel–Ziv complexity for volatility duration

4.1. PLZC analysis of symbolic sequence

The Lempel–Ziv complexity (LZC), proposed by Lempel and Ziv [40,41], is a non-parametric measure of complexity used as a technique to evaluate the randomness of a finite symbolic sequence. Recently, this measure has been extensively employed to evaluate the complexity of the series of discrete-time in many fields [42–45]. For calculating the LZC, times series must be transformed into a finite sequence $s(t)$ whose elements are only a few symbols [46], this process is called coarse-graining. For example, the binary conversion method processed by complaining each element of the original series $x(t)$ with a threshold (commonly the mean value or the median value of $x(t)$) is a typical one [47]. The coarse-graining process in the LZC plays an important role because the conversion process determines how much information retained of the original series, and more types of elements in $s(t)$ means more information of the original series remained [45–48]. In this section, we apply a novel method called permutation conversion to the coarse-graining process of $D(t)$, $MCR(t)$ and $ACR(t)$, then calculate their permutation Lempel–Ziv complexity (PLZC) to investigate the sym-

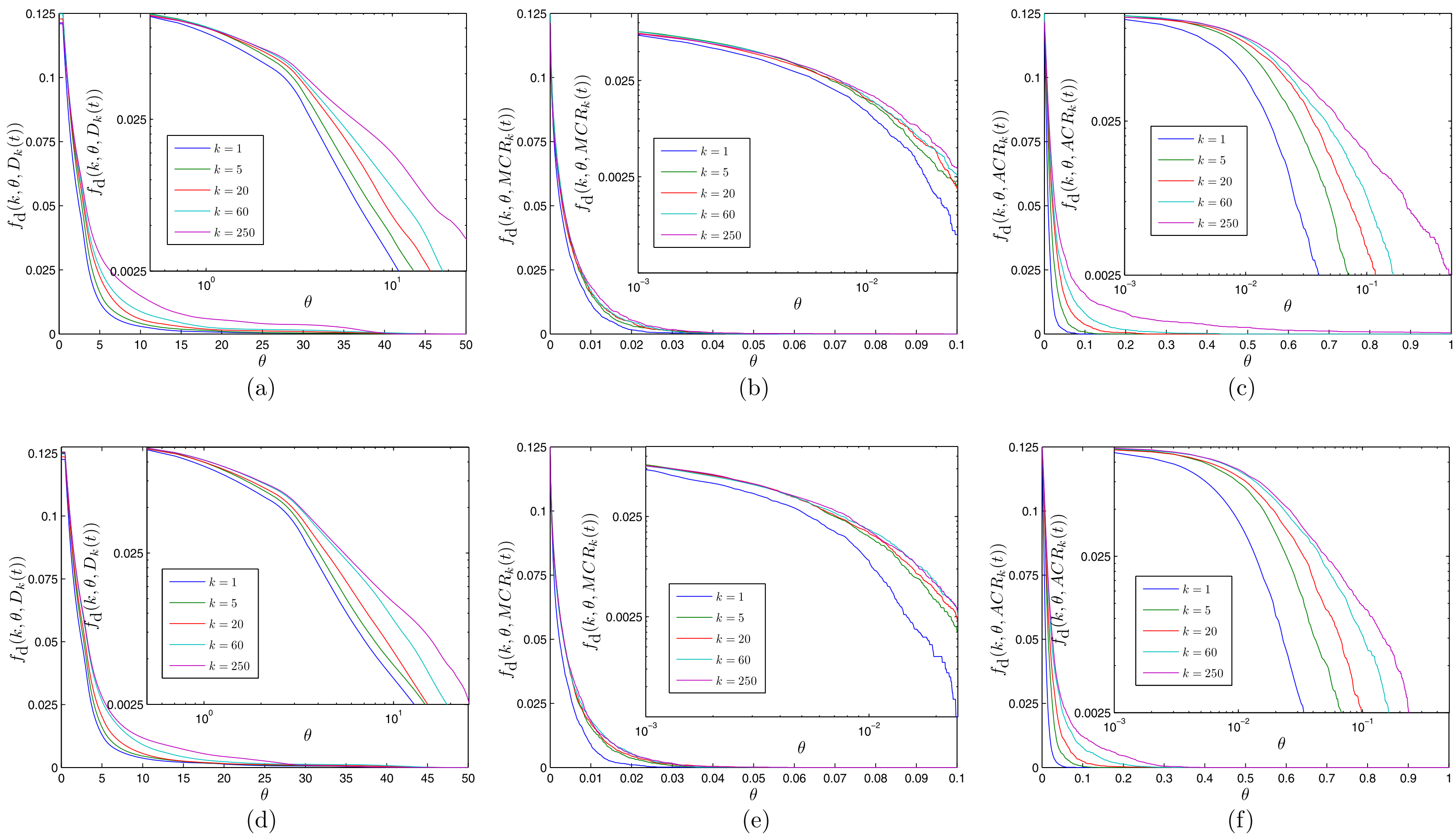

Fig. 5. (a)–(c) Plots of $f_d(k, \theta, D_k(t))$, $f_d(k, \theta, MCR_k(t))$ and $f_d(k, \theta, ACR_k(t))$ for SSE; (d)–(f) Plots of $f_d(k, \theta, D_k(t))$, $f_d(k, \theta, MCR_k(t))$ and $f_d(k, \theta, ACR_k(t))$ for the simulation data with $\lambda = 3$. The inner figure is the corresponding log–log plot.

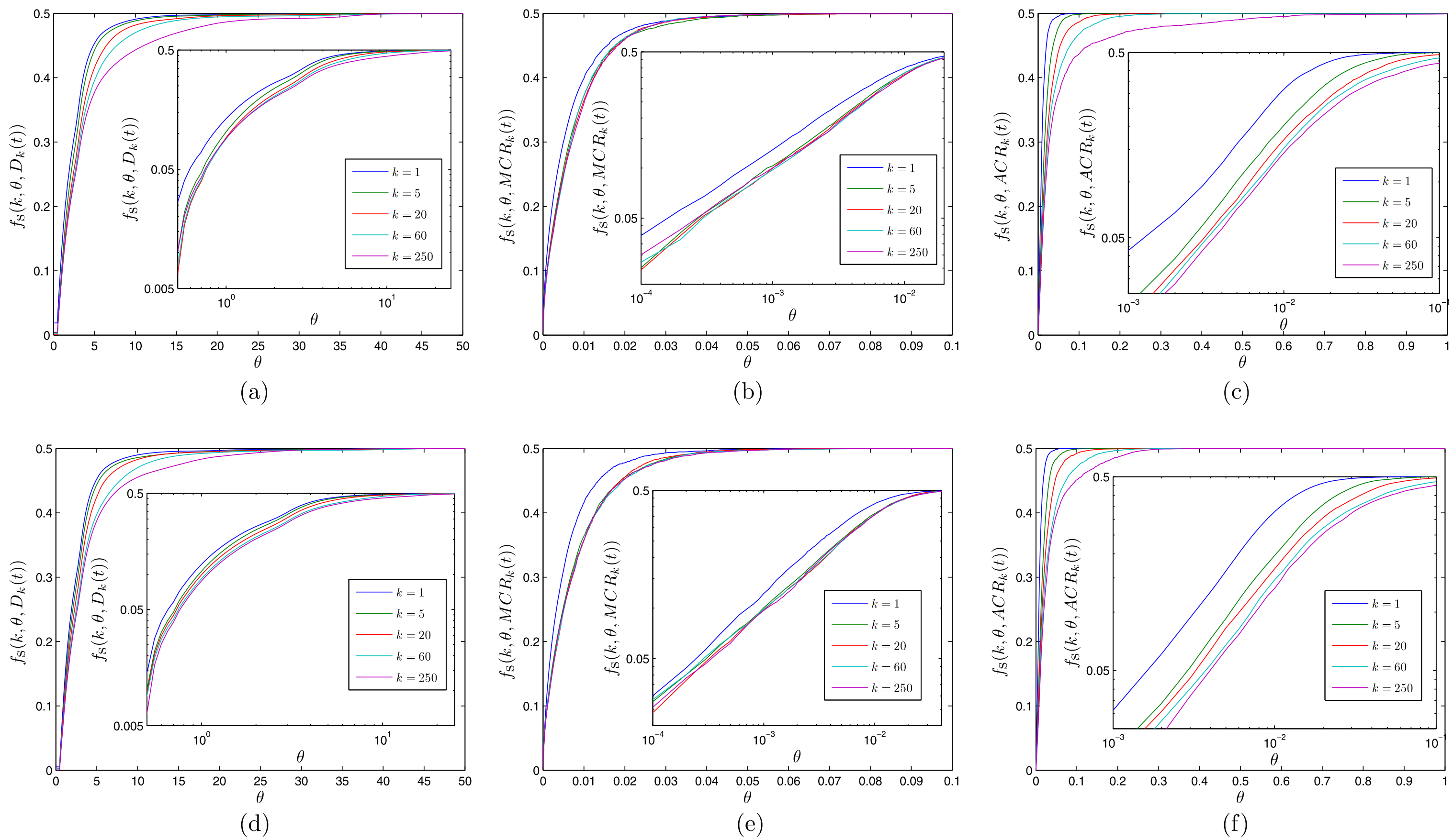

Fig. 6. (a)–(c) Plots of $f_s(k,\theta,D_k(t))$, $f_s(k,\theta,MCR_k(t))$ and $f_s(k,\theta,ACR_k(t))$ for SSE; (d)–(f) Plots of $f_s(k,\theta,D_k(t))$, $f_s(k,\theta,MCR_k(t))$ and $f_s(k,\theta,ACR_k(t))$ for the simulation data with $\lambda=3$. The inner figure is the corresponding log–log plot.

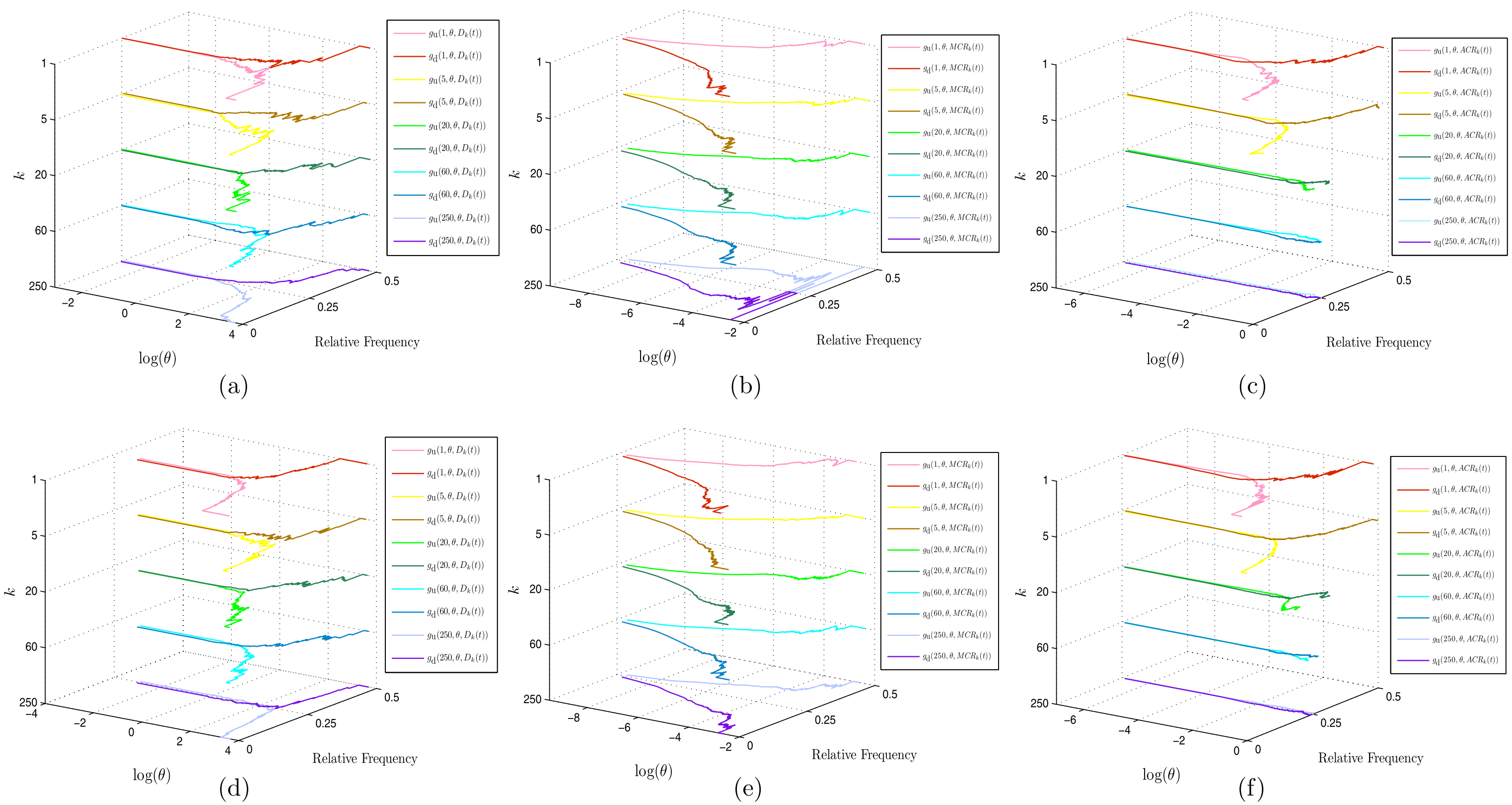

Fig. 7. (a)–(c) Plots of $g_{\mathrm{u}}(k, \theta, s_k(t))$ and $g_{\mathrm{d}}(k, \theta, s_k(t))$ versus $\log(\theta)$ of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ for SSE with $k = 1, 5, 20, 60, 250$; (d)–(f) Plots of $g_{\mathrm{u}}(k, \theta, s_k(t))$ and $g_{\mathrm{d}}(k, \theta, s_k(t))$ versus $\log(\theta)$ of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ for the simulation data with $k = 1, 5, 20, 60, 250$.

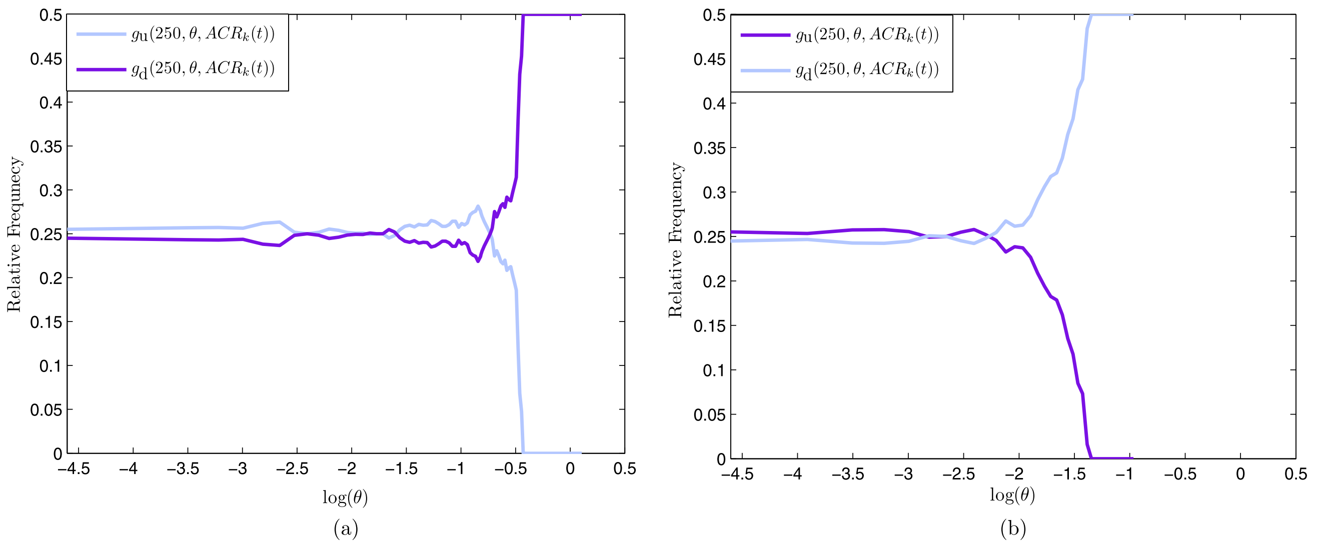

Fig. 8. (a) Plots of $g_u(k, \theta, ACR_k(t))$ and $g_d(k, \theta, ACR_k(t))$ versus $\log(\theta)$ of SSE with $k = 250$ for θ varies in $(0, 10)$; (b) Plots of $g_u(k, \theta, ACR_k(t))$ and $g_d(k, \theta, ACR_k(t))$ versus $\log(\theta)$ of the simulation data with $k = 250$ for θ varies in $(0, 10)$.

bolic complexity of volatility behaviors for the real stock market and the proposed price model.

Firstly, we introduce the permutation conversion method [45,48,49]. For a time series $x(t)$ $(t = 1, 2, \cdots, N)$, given an embedding dimension m and a time delay η, $x(t)$ is embed to an m-dimensional space $X(t) = [x(t), x(t+\eta), \cdots, x(t+(m-1)\eta)]$ $(t \in \{1, 2, \cdots, N-(m+1)\eta\})$. Then, arrange the components of $X(t)$ in an increasing order

$$y^s(t + (q_1 - 1)\eta) \le y^s(t + (q_2 - 1)\eta) \le \cdots \le y^s(t + (q_m - 1))\eta). \tag{23}$$

When an equality coming out, e.g., $x(t+(q_i-1)\eta) = x(t+(q_j-1)\eta)$ $(i, j \in \{1, 2, \cdots, m\})$, the quantity order depends on the q values, namely if $q_i \le q_j$, let $x(t+(q_i-1)\eta) \le x(t+(q_j-1)\eta)$. So, any vector $X(t)$ has a permutation $\pi_t = [q_1, q_2, \cdots, q_m]$, which is one of permutations of m distinct symbol set $\{1, 2, \cdots, m\}$. The distinct symbol set $\{1, 2, \cdots, m\}$ has $m!$ different permutations, which can be corresponding to $m!$ different characters $\{c_1, c_2, \cdots, c_{m!}\}$. For each permutation π_l $(l \in \{1, 2, \cdots, m!\})$ of symbol set $\{1, 2, \cdots, m\}$, it has only one corresponding character c_l $(c_l \in \{c_1, c_2, \cdots, c_{m!}\})$. Therefore, the vector $X(t)$ $(t \in \{1, 2, \cdots, N-(m+1)\eta\})$ whose permutation is π_t can be converted to character c_t which is corresponding to π_t. In this way, the time series $x(t)$ is converted to a symbolic sequence $s(t)$ with no more than $m!$ kinds of different characters in. Moreover, $s(t)$ records the local structure of $x(t)$, since every character in $s(t)$ reflects a local permutation pattern of $x(t)$.

Now, we introduce the permutation Lempel–Ziv complexity algorithm. For a time series $x(t)$, the PLZC of $x(t)$ is measured in the following steps [42–45,47,48]. Let S and Q represent two different character sequences, and SQ represents the concatenation of S and Q. $SQ\pi$ is the sequence obtained from SQ in which its last character is deleted. Let $\nu(SQ\pi)$ represent the set comprising all different subsequences of $SQ\pi$.

(i) Set dimension parameter m and time delay parameter η. Convert the original series $x(t)$ to a symbol sequence $s(t)$ whose length is n according to the permutation conversion method.

(ii) At the beginning, set $c(n) = 1$, $S = s(1)$, $Q = s(2)$, then $SQ = s(1), s(2)$, and $SQ\pi = s(1)$.

(iii) In general, suppose that $S = s(1), s(2), \cdots, s(r)$, $Q = s(r+1)$, so $SQ\pi = s(1), s(2), \cdots, s(r)$. If $Q \in \nu(SQ\pi)$, then Q is a subsequence of $SQ\pi$, not a new sequence.

(iv) Renew Q by adding $s(r+2)$ to Q, that is, $Q = s(r+1), s(r+2)$, then judge if Q belongs to $\nu(SQ\pi)$.

(v) Repeat steps (iii) and (iv) until Q does not belong to $\nu(SQ\pi)$. Now $Q = s(r+1), s(r+2), \cdots, s(r+i)$ is not a subsequence of $SQ\pi = s(1), s(2), \cdots, s(r+i-1)$, but a new sequence, so increase $c(n)$ by one.

(vi) Then, S is renewed to be $SQ = s(1), s(2), \cdots, s(r+i)$, and $Q = s(r+i+1)$.

(vii) Repeat the previous steps until Q contains the last character of $s(t)$. At that time, $c(n)$ is the complexity of symbol sequence $s(t)$ which denotes the number of distinct new patterns in the original time series.

For the $c(n)$, which is the total number of new subsequences in above $s(t)$, Lempel and Ziv [40,50] has proved that its upper bound $L(n)$ is given by

$$c(n) < L(n) = \frac{n}{(1-\varepsilon_n)\log_{m!}(n)} \tag{24}$$

where

$$\varepsilon_n = 2\frac{1 + \log_{m!}\log_{m!}(m!n)}{\log_{m!}(n)} \tag{25}$$

and $\varepsilon_n \to 0$ $(n \to \infty)$. In general, the limit of $L(n)$, i.e.,

$$\lim_{n\to\infty} L(n) = \frac{n}{\log_{m!}(n)} \tag{26}$$

is the upper bound of $c(n)$ [43], so $c(n)$ can be normalized by $n/\log_{m!}(n)$

$$C(n) = \frac{c(n)\log_{m!}(n)}{n}. \tag{27}$$

$C(n)$, the normalized PLZC of $x(t)$, reflects the arising rate of new pattern generation along with the sequence, capturing the temporal structure of a time series [43,51]. In the PLZC algorithm, dimension parameter m and time delay parameter η are crucial for calculating PLZC because these will generate different symbolic sequences for different parameters m and η. Traditional permutation process recommends m to be 3–7 [49,52] for when $m < 3$, there will be too few possible patterns to make the permutation no sense. But, the calculation of PLZC will be very complex for $m > 7$. Therefore, a large value of m and n should not be selected in order to maintain sensitivity to the instantaneous characteristic changes of nonlinear dynamic systems and economize computation time.

Table 2
PLZC values of $D(t)$, $MCR(t)$ and $ACR(t)$ of $|r(t)|$.

	SSE	SZSE	$\lambda = 3$	$\lambda = 6$	$\lambda = 9$	$\lambda = 12$
$D(t)$	0.4432	0.4396	0.4612 ± 0.0120	0.4456 ± 0.0130	0.4420 ± 0.0117	0.4480 ± 0.0091
$MCR(t)$	0.5055	0.5115	0.5139 ± 0.0124	0.5055 ± 0.0052	0.5091 ± 0.0100	0.5187 ± 0.0177
$ACR(t)$	0.4971	0.4768	0.5007 ± 0.0156	0.5019 ± 0.0060	0.5139 ± 0.0106	0.5031 ± 0.0108

Furthermore, to ensure every possible character of $\{c_1, c_2, \cdots, c_{m!}\}$ occurs in $s(t)$ with length n, $m!$ has to be less than or equal to $n-(m-1)\eta$. And n needs to satisfy $n \gg m! + (m-1)\eta$ to avoid under sampling [52,53]. Bai et al. [48] recommends $m = 4$ for $n \geq 1000$ or $m = 5$ for $n \geq 2000$. In this paper, we choose a low dimension $m = 4$ when calculating the PLZC. The another parameter η is chosen according to the autocorrelation function (ACF) e^{-1} rule, i.e., the ACF of $x(t)$ decays to e^{-1} of its peak value, the corresponding lag is chosen as the time delay parameter η [54,55].

We calculate the PLZC values of $D(t)$, $MCR(t)$ and $ACR(t)$ of daily volatility series for the SSE, the SZSE and the simulation data with $\lambda = 3, 6, 9, 12$ in Table 2. The standard errors of the PLZC values for the simulation data are also presented in Table 2. For the real data and the simulation data, the PLZC values of $D(t)$, $MCR(t)$ and $ACR(t)$ are close to 0.5 less than 1, which indicates that $D(t)$, $MCR(t)$ and $ACR(t)$ exhibit regularity and randomness at the same time. For $D(t)$, the PLZC values are all less than 0.5, so the regularity and periodicity in $D(t)$ are more significant than its randomness for both the real data and the simulation data. On the contrary, the randomness of $MCR(t)$ is more significant than the regularity and periodicity, since the PLZC values of $MCR(t)$ in Table 2 are more than 0.5. As for the $ACR(t)$, although its PLZC values of the real data are less than 0.5 while those of the simulation data are more than 0.5, the difference of PLZC values of $ACR(t)$ between the SSE and the simulation data with $\lambda = 3, 6, 12$ is less than the corresponding standard error respectively. So $ACR(t)$ of the real data and the simulation data shows similar randomness and periodicity. Generally speaking, the PLZC values of $D(t)$, $MCR(t)$ and $ACR(t)$ of daily volatility series for the real data and the simulation data are very close respectively in Table 2, so the volatility behaviors of the real data and the simulation data show similar symbolic complexity.

4.2. PLZC analysis for different timescale

Now we investigate the PLZC dynamics of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of k-volatility series. We calculate the corresponding PLZC values when timescale k varies from 1 to 250. Fig. 9 displays the plots of PLZC versus timescale k for the real data and the simulation data, the corresponding box plots of the PLZC values are shown in Fig. 10. Table 3 contains the 25% percentile q_1, the median value q_m and 75% percentile q_3 of the PLZC values. In Figs. 9–10, we find that the PLZC values of $D_k(t)$ almost concentrate in three intervals $(0.4, 0.45)$, $(0.6, 0.65)$ and $(0.75, 0.8)$ for both the real data and the simulation data. When timescale k increases from 1 to 250, the PLZC of $D_k(t)$ swings between the three intervals, which means that $D_k(t)$ shows three levels of complexity during the increase of k. When the PLZC is in the interval $(0.4, 0.45)$, the regularity and periodicity of $D_k(t)$ are more significant, while for the $D_k(t)$ whose PLZC in the other two intervals, its randomness is more obvious.

In Table 3, q_m and q_3 of PLZC of $D_k(t)$ for both the real data and the simulation data are all in $(0.6, 0.65)$. For the SZSE and the simulation data with $\lambda = 6, 9$, q_1 of PLZC of $D_k(t)$ is also in this interval, while that for the SSE and the rest simulation data is in $(0.4, 0.45)$. In general, most of PLZC of $D_k(t)$ are between 0.6 and 0.65 during timescale's increase, so the randomness of $D_k(t)$ is more significant than its regularity and periodicity. Moreover in Fig. 9, the PLZC values of $D_k(t)$ swing in $(0.4, 0.45)$ at relative small k, while at relative larger timescale, most of PLZC values of $D_t(k)$ fluctuate in the interval $(0.6, 0.65)$, which indicates that the increasing timescale will cause the increasing complexity of $D_k(t)$ of k-volatility series. For $MCR_k(t)$ of k-volatility series, the PLZC swings between 0.5 to 0.55 during the timescale's increase in Fig. 9. In the corresponding box plots, the PLZC values of $MCR_k(t)$ lie almost in the interval $(0.5, 0.55)$, and q_1, q_m and q_3 of PLZC of $MCR_k(t)$ are between 0.52 to 0.54 in Table 3. Therefore, the randomness of $MCR_k(t)$ is more obvious than its regularity and periodicity, and the timescale k will not affect the complexity of $MCR_k(t)$ so much. For the $ACR_k(t)$ from Figs. 9–10, we find that its PLZC values also concentrate three intervals $(0.5, 0.55)$, $(0.65, 0.7)$ and $(0.8, 0.85)$, and the PLZC shows the similar behaviors with the PLZC of $D_k(t)$ during the timescale's increase, which indicates that the $ACR_k(k)$ displays three level of complexity during the increase of k. The q_1, q_m and q_3 of PLZC of $ACR_k(t)$ are all in $(0.65, 0.7)$ in Table 3, so that most PLZC values of $ACR_k(t)$ are between 0.65 and 0.7 when k varies from 1 to 250. Furthermore, the randomness of $ACR_k(t)$ is more significant than regularity and periodicity since the PLZC values of $ACR_k(t)$ are more than 0.5 during timescale's increase, and the randomness of $ACR_k(t)$ becomes more obvious with the increase of k.

Table 3
Percentile and median of PLZC values of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$.

		SSE	SZSE	$\lambda = 3$	$\lambda = 6$	$\lambda = 9$	$\lambda = 12$
$D_k(t)$	q_1	0.4432	0.6105	0.4217	0.6045	0.6069	0.4277
	q_m	0.6201	0.6237	0.6285	0.6249	0.6213	0.6357
	q_3	0.6309	0.6321	0.6393	0.6369	0.6321	0.6465
$MCR_k(t)$	q_1	0.5307	0.5295	0.5271	0.5283	0.5283	0.5259
	q_m	0.5343	0.5331	0.5319	0.5319	0.5319	0.5307
	q_3	0.5391	0.5379	0.5355	0.5367	0.5355	0.5343
$ACR_k(t)$	q_1	0.6729	0.6765	0.6753	0.6801	0.6753	0.6849
	q_m	0.6849	0.6885	0.6897	0.6897	0.6843	0.6987
	q_3	0.6957	0.7053	0.6969	0.6981	0.6957	0.7089

Moreover, the PLZC of $MCR_k(t)$ is larger than that of $D_k(t)$ and $ACR_k(t)$ at relative small k, while at large k, the PLZC of $MCR_k(t)$ becomes the smallest one in Fig. 9. However, for both the real data and the simulation data, the median value in Table 3 of PLZC of $ACR_k(t)$ is the largest among D_k, $MCR_k(t)$ and $ACR_k(t)$, while that of $MCR_k(t)$ is the smallest one. In general, $ACR_k(t)$ is more random and complex than $D_k(t)$ and $MCR_k(t)$, and the regularity and periodicity of $MCR_k(t)$ is more significant than those of $D_k(t)$ and $ACR_k(t)$. In Fig. 9, although the plots of PLZC values of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of the simulation data do not display significant regular change as λ increases, the plots of some simulation data are similar with those of the real data. For example, Figs. 9(d), (f) are similar with Figs. 8(a), (b) in the comparison of dynamical behaviors of PLZC values of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ comprehensively as k increases. From Figs. 9–10, as k increases, the dynamical behavior of complexity and randomness of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of the simulation data with $\lambda = 6$ is similar with that of the SZSE, while that of the simulation data with $\lambda = 12$ is similar with that of the SSE. Therefore, as λ increases, $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ of some simulation data shows similar complex and random properties with those of the real data when the timescale k varies.

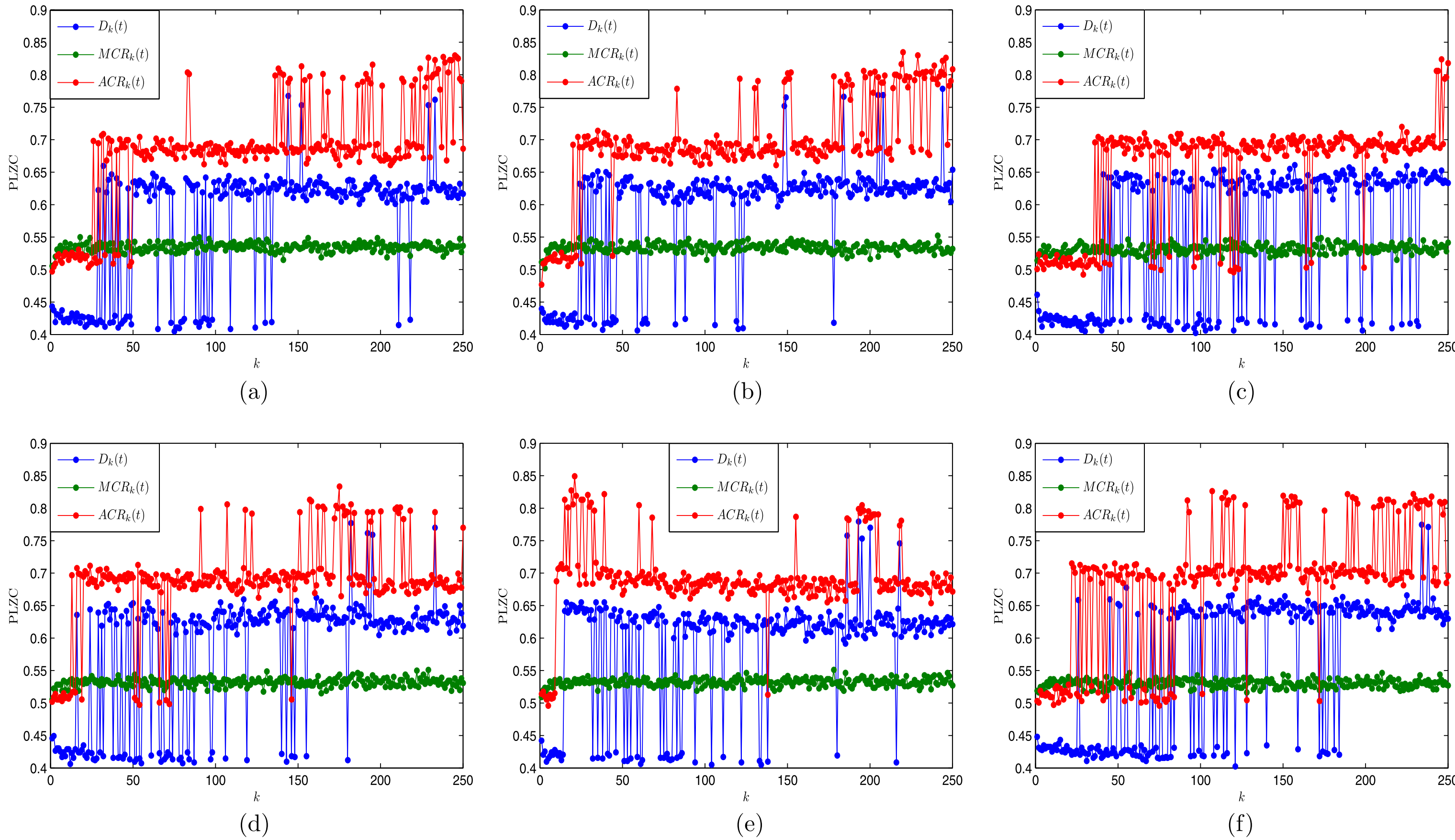

Fig. 9. (a)–(f) Plots of PLZC versus k of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ for SSE, SZSE and the simulation data with $\lambda = 3, 6, 9, 12$.

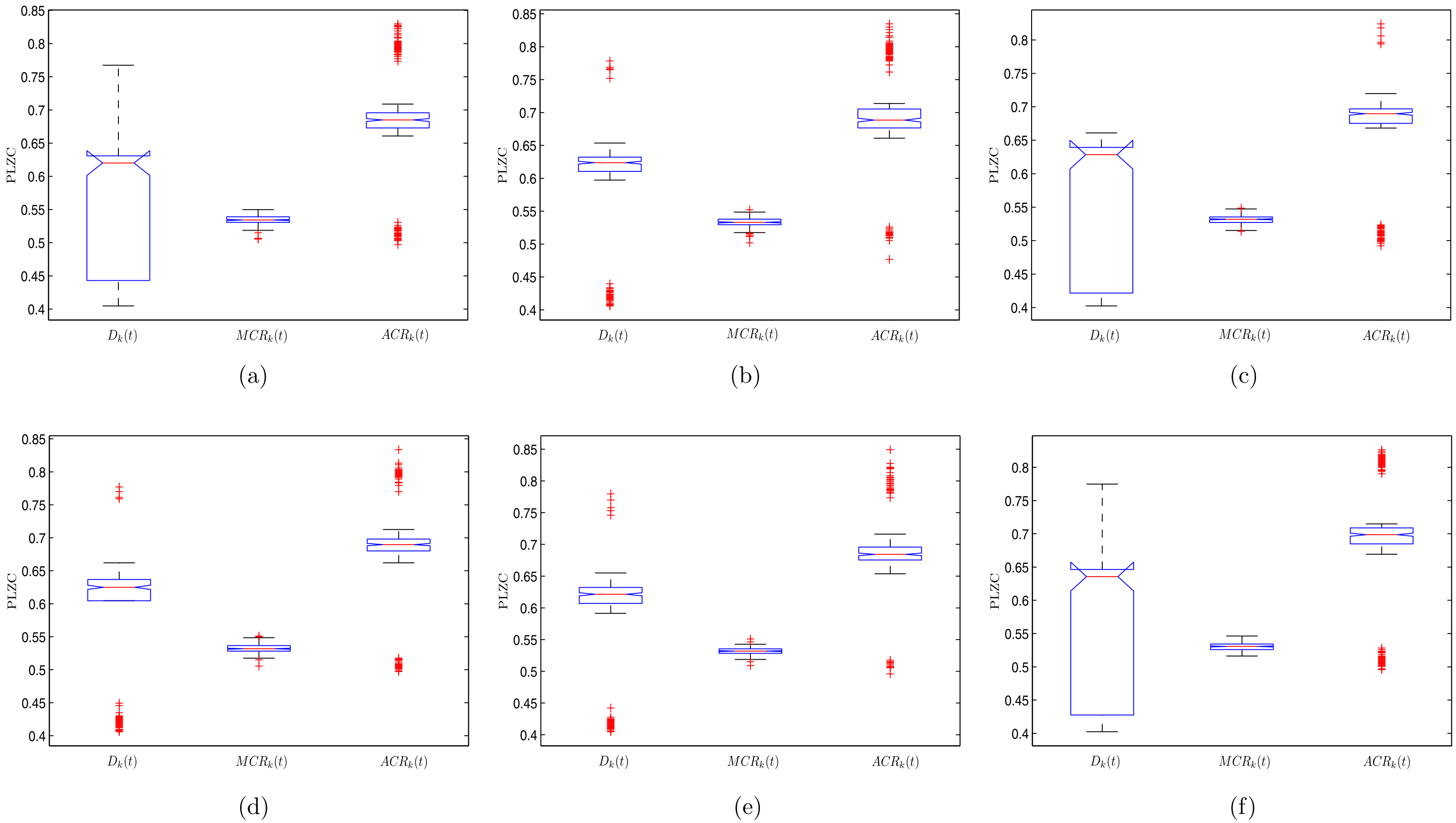

Fig. 10. (a)–(f) Box plots of PLZC values of $D_k(t)$, $MCR_k(t)$ and $ACR_k(t)$ for SSE, SZSE and the simulation data with $\lambda = 3, 6, 9, 12$, where k increases from 1 to 250.

5. Conclusion

In the present paper, two new concepts of series about volatility duration and volatility difference component are introduced, which are the maximum change rate series and the average change rate series of the volatilities in the volatility duration. The proposed volatility return duration series and the above two series are transformed to symbolic sequences, and then the corresponding symbolic complexity analysis are performed by the permutation Lempel–Ziv complexity (which is a novel complexity measure) analysis and Zipf analysis with different timescales and thresholds. Meanwhile, a stochastic voter financial dynamics model is proposed to utilize making comparison on the volatility behaviors with the real market data. The PLZC empirical results display that $D(t)$, $MCR(t)$ and $ACR(t)$ show regularity and randomness, and the regularity and periodicity of $D(t)$ are more significant than the randomness, while those of $MCR(t)$ and $ACR(t)$ show the opposite properties. The changing value of timescale k causes the fluctuation of randomness and complexity of $D_k(t)$ and $ACR_k(t)$ between three different levels, but does not affect those of $MCR_k(t)$ so much. Generally speaking, $MCR_k(t)$ is more random than $D_k(t)$ and $ACR_k(t)$ at relative small timescale. But at relative large timescale, $ACR_k(t)$ is more random than the rest two series, and $MCR_k(t)$ become the most regular series. Furthermore, Zipf analysis on the proposed three series shows that, the increasing threshold θ will lead to the exponential increase or decrease of absolute frequency functions of the three series, and cause the deviation of the relative frequency functions. The change of timescale k will lead significant change of the frequency functions of $D_k(t)$ and $ACR_k(t)$ while does not affect those of $MCR_k(t)$ so much. Through the comparison of the above analyses on the three volatility duration series, the simulation data derived from the financial price model has the similar symbolic complex properties of volatilities with the real data, which indicates that the presented financial price model is reasonable for the real stock market to some extent.

Acknowledgment

The authors were supported in part by National Natural Science Foundation of China Grant No. 71271026.

References

[1] P. Chan, R. Sircar, Betrand and Cournot mean field games, Appl. Math. Optim. 71 (2015) 533–569.
[2] K. Spiliopoulos, J.A. Sirignano, K. Giesecke, Fluctuation analysis for the loss from default, Stoch. Process. Appl. 124 (2014) 2322–2362.
[3] J. Cvitancić, J. Ma, J.F. Zhang, The law of large numbers for self-exciting correlated defaults, Stoch. Process. Appl. 122 (2012) 2781–2810.
[4] J.A. Sirignano, K. Giesecke, Risk Analysis for Large Pools of Loans, working paper, Stanford University, 2014.
[5] L. Bo, A. Capponl, Bilateral credit valuation adjustment for large credit derivatives portfolios, Finance Stoch. 18 (2014) 431–482.
[6] P.D. Pra, M. Tolotti, Heterogeneous credit portfolios and the dynamics of the aggregate losses, Stoch. Process. Appl. 199 (2009) 2913–2944.
[7] P.D. Pra, W.J. Runggaldier, E. Sartori, M. Tolotti, Large Portfolio losses: a dynamic contagion model, Ann. Appl. Probab. 19 (2009) 347–394.
[8] J.A. Sirignano, G. Tsoukalas, K. Giesecke, Large-Scale Loan Portfolio Selection, working paper, Stanford University, 2016.
[9] R. Carmona, J.P. Fouque, L.H. Sun, Mean field games and systemic risk, Quant. Finance 16 (2016) 1219–1235.
[10] J.P. Fouque, T. Ichiba, Stability in a model of inter-bank lending, SIAM J. Financ. Math. 4 (2013) 784–803.
[11] L.J. Bo, A. Capponi, Systemic risk in interbanking networks, SIAM J. Financ. Math. 6 (2015) 386–424.
[12] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quant. Finance 1 (2001) 223–236.
[13] S. Bornholdt, F. Wagner, Stability of money: phase transitions in an Ising economy, Physica A 316 (2002) 453–500.
[14] D. Stauffer, T.J.P. Penna, Crossover in the Cont–Bouchaud percolation model for market fluctuations, Physica A 256 (1998) 284–290.
[15] R.N. Mantegna, H.E. Stanley, Scaling behaviour in the dynamics of an economic index, Nature 376 (2002) 46–49.
[16] T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397 (1999) 498–500.
[17] F. Black, M. Scholes, The pricing of options and corporate liabilities, J. Polit. Econ. 81 (1973) 637–654.
[18] Y. Yu, J. Wang, Lattice oriented percolation system applied to volatility behavior of Stock market, J. Appl. Stat. 39 (4) (2012) 785–797.
[19] W.J. Hong, J. Wang, Nonlinear scaling analysis approach of agent-based Potts financial dynamical model, Chaos 24 (2014) 043113.
[20] H.L. Niu, J. Wang, Volatility clustering and long memory of financial time series and financial price model, Digit. Signal Process. 23 (2013) 489–498.
[21] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Econophysics: financial time series from a statistical physics point of view, Physica A 279 (2000) 443–456.
[22] P. Clifford, A. Sudbury, A model for spatial conflict, Biometrika 60 (1973) 581–588.
[23] R.A. Holley, T.M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, Ann. Probab. 3 (1975) 643–663.
[24] T.M. Liggett, Interacting Particle Systems, Springer-Verlag, New York, 1985.
[25] T.M. Liggett, Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer-Verlag, New York, 1999.
[26] R. Durrett, Lecture Notes on Particle Systems and Percolation, Wadsworth & Brooks, Pacific Grove, CA, 1998.
[27] F. Wang, K. Yamasaki, S. Havlin, H.E. Stanley, Scaling and memory of intraday volatility return intervals in stock markets, Phys. Rev. E 73 (2006) 026117.
[28] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, H.E. Stanley, Scaling and memory in volatility return intervals in stock and currency markets, Proc. Natl. Acad. Sci. USA 102 (2005) 9424–9428.
[29] Y.F. Dong, J. Wang, Fluctuation behavior of financial return interval series model for percolation on Sierpinski carpet lattice, Fractals 21 (2013) 1350023.
[30] S.M. Ross, An Introduction to Mathematical Finance, Cambridge University Press, Cambridge, 1999.
[31] D. Lamberton, B. Labeyre, Introduction to Stochastic Calculus Applied to Finance, Chapman and Hall/CRC, 2000.
[32] G. Yang, J. Wang, W. Deng, Nonlinear analysis of volatility duration financial series model by stochastic interacting dynamic system, Nonlinear Dyn. 80 (1–2) (2015) 701–713.
[33] G.K. Zipf, Human Behaviour and the Principle of Least Effort, Addison–Wesley Press, Cambridge, 1949.
[34] G.K. Zipf, The Psycho-Biology of Language: An Introduction to Dynamic Psychology, Addison–Wesley Press, Cambridge, 1968.
[35] D.M.W. Powers, Applications and explanations of Zipf's law, Adv. Neural Inf. Process. Syst. 5 (4) (1998) 595–599.
[36] J. Alvarez-Ramirez, A. Soriano, M. Cisneros, R. Suarez, Symmetry/anti-symmetry phase transitions in crude oilmarkets, Physica A 322 (2003) 583–596.
[37] Y.L. Guo, J. Wang, Simulation and statistical analysis of market return fluctuation by Zipf method, Math. Probl. Eng. 2011 (2011) 253523.
[38] H.L. Niu, J. Wang, Power-law scaling behavior analysis of financial time series model by voter inter interacting dynamic system, J. Appl. Stat. 40 (2013) 2188–2203.
[39] N. Vandewalle, M. Ausloos, The n-Zipf analysis of financial data series and biased data series, Physica A 268 (1999) 240–249.
[40] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory 22 (1976) 75–81.
[41] J. Ziv, N. Merhav, Estimating the number of stats a finite-state source, IEEE Trans. Inf. Theory 38 (1992) 61–65.
[42] A. Fernández, M. López-Ibor, A. Turrero, J. Santos, M. Morón, R. Hornero, C. Gómez, M.A. Méndez, T. Ortiz, J.J. López-lbor, Lempel–Ziv complexity in shizophrenia: a MEG study, Clin. Neuropharmacol. 122 (2011) 2227–2235.
[43] D. Abásolo, R. Hornero, C. Gómez, M. García, M. López, Analysis of EEG background activity in Alzheimer's disease patients with Lempel–Ziv complexity and central tendency measure, Med. Eng. Phys. 28 (2006) 315–322.
[44] J. Szczepański, J.M. Amigó, E. Wajnryb, M.V. Sanchez-Vives, Characterizing spike trains with Lempel–Ziv complexity, Neurocomputing 58–60 (2004) 79–84.
[45] R. Li, J. Wang, Interacting price model and fluctuation behavior analysis from Lempel–Ziv complexity and multi-scale weighted-permutation entropy, Phys. Lett. A 380 (2016) 117–129.
[46] E. Estevez-Rams, R.L. Serrano, B.A. Fernandez, I.B. Reyes, On the non-randomness of maximum Lempel–Ziv complexity sequences of finite size, Chaos 23 (2013) 023118.
[47] S.J. Zhou, Z.C. Zhang, J. Gu, Interpretation of coarse-graining of Lempel–Ziv complexity measure in ECG signal analysis, IEEE Eng. Med. Biol. (2011) 2716–2719.
[48] Y. Bai, Z.H. Liang, X.L. Li, A permutation Lempel–Ziv complexity measure for EEG analysis, Biomed. Signal Process. Control 19 (2015) 102–114.
[49] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (2002) 174102.

[50] J. Hu, J. Gao, J.C. Príncipe, Analysis of biomedical signals by the Lempel–Ziv complexity: the effect of finite data size, IEEE Trans. Biomed. Eng. 53 (2006) 2606–2609.
[51] X.S. Zhang, R.J. Roy, E.W. Jensen, EEG complexity as a measure of depth of anesthesia for patients, IEEE Trans. Biomed. Eng. 48 (2001) 1424–1433.
[52] X.L. Li, G.X. Ouyang, Estimating coupling direction between neuronal populations with permutation conditional mutual information, NeuroImage 52 (2010) 497–507.
[53] J.M. Amigó, M.B. Kennel, L. Kocarev, The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems, Physica D 210 (2005) 77–95.
[54] G. Ouyang, C. Dang, D.A. Richards, X. Li, Ordinal pattern based similarity analysis for EEG recordings, Clin. Neurophysiol. 121 (2010) 694–703.
[55] M. Shelhamer, Nonlinear Dynamics in Physiology, World Scientific, Singapore, 2006.

Rui Li is a researcher from School of Science at Beijing Jiaotong University, China. His research interests are in Stochastic Systems, Statistical Physics, Stochastic Processes, Modeling and Computer Simulation, Probability Theory and Statistics, Financial Mathematics and Financial Statistics.

Jun Wang is a Full Professor in School of Science, and Director of Institute of Financial Mathematics and Financial Engineering at Beijing Jiaotong University, China. His research interests include Large Scale Interacting Systems, Statistical Physics, Stochastic Systems, Stochastic Process, Artificial Intelligence, Neural Networks, Modeling and Computer Simulation, Probability Theory and Statistics, Financial Mathematics, Financial Engineering, and Financial Statistics.