Simulation and Analysis of the Chinese Stock Market based on the Oriented Percolation Model

Zhuoheng Wang

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Zhuoheng_wang@sina.com

Abstract

Financial Engineering is the vital research topic in the field of finance at present. This thesis studies the issue of price fluctuation and yield in financial market. Based on the theory of Oriented Percolation Model to conduct a modeling simulation analysis, a comparison research is conducted between the simulated data and real data and the result turns out to share a high similarity in their data images. The simulated data can generally present the phenomenon of sharp peaks and fat tail in probability distribution of financial time sequence. It indicates that he Herd Behavior (or Sheep-Flock Effect) among the investors can be generally portrayed by building model of financial time sequence with percolation theory. In the part of empirical analysis, this thesis studies the recent data from Shanghai Composite Index and Shenzhen Component Index. In the last part, the simulated data is added into the real financial time sequence to study their dynamic relations and further analysis is conducted to verify the possibility of replacing a set of real data with simulated data by utilizing VAR model. After the analysis, the good correlation is found between the simulated data and real data. The author makes further study to demonstrate the fact that the building model partly simulates the price fluctuation in China's security market.

Keywords

Two-Dimension Oriented Percolation Model, Financial Time Sequence, VAR (Vector Auto-Regression).

1. Introduction

1.1 Fundamental Theory of Random Process

1.1.1 Random Process

Firstly, the fundamental concept of random variable should be given. Random variable refers to select a certain numerical number previously unknown with the probability of certain from the results of every experiment of many experiments. In the real process, some occasions may involve random variables changing with time (t). For instance, the stock market price index changes randomly with the change of time (t). These random variables changing with the time (t) are usually named the random process. Furthermore, when studying a case of a stock market price index changing with time (t) in a certain period of time, if the time is t , the price index is possible to be $x_1(t)$, $x_2(t)$, $x_3(t)$, etc,. Though the number selected at a time cannot previously be known, but it is definitely to be one from all possible selection numbers. That is to say, the number at the certain moment is randomly selected. From another perspective, those selected numbers, $x_1(t)$, $x_2(t)$, $x_3(t)$, are made up to a time function. This function is usually named sample function (short for ample) of the random process. In conclusion, random process can be defined as following:

Definition 2.1 We suppose that (Ω, F, P) is a given probability space and T is an index set. There exists definition in (Ω, F, P) for random value $t \in T$ that is corresponding to value of random value $X(\omega,t)(\omega \in \Omega)$ form E . Then the group of random variables from t can be named the random process and simplified as $\{X_t(\omega)\}$ or $\{X(t)\}$.

Indicator set T is usually expressed as time set, phase space of the random process. Random process $\{X(\omega,t):\omega\in\Omega,t\in T\}$ is the function of two variables of time parameter t and sample point ω; for given time $t_0\in T$, $X(\omega,t_0)$ is the random variables on probability space (Ω,F,P). For given sample point $\omega_0\in\Omega$, $X(\omega_0,t)$ is the real function on T. Then it can be named as a sample function of random process corresponding to ω_0. It can also be named as sample track or item. Phase space E is named state space. Normally, $X_t=x$ indicates that X_t is in the state of x.

1.1.2 Markov Chain

Suppose that $\left\{X_n,n=0,1,2,\cdots\right\}$ is a random process, and the condition of $t=u$ is known when $t=u$, if the future condition of $X_v\left(v>u\right)$ is free from the influences of the past condition of $X_w\left(w<u\right)$, and then $\left\{X_n\right\}$ is a process with Markov property. Markov property refers to the following situation: based on *present* known conditions, both the *future* and the *past* are independent. This property is also named non-aftereffect property. Discrete parameter Markov Chain and continuous parameter Markov Chain are respectively introduced in following passage.

Definition 2.2 If we only sample limited or countable values of random process $\left\{X_n,n=0,1,2,\cdots\right\}$ then entire possible values in the process can be named the state space and marked as E. Normally, E is supposed to be $E=\left\{0,1,2\cdots\right\}$. If $X_n=i$, it can be explained "the process is in the i state at the time of n". Suppose that when the process is in the state of i, and the probability is fixed for j to be the next state of i, thus for random time n

$$P\left(X_{n+1}=j\,|\,X_n=i\right)=P_{ij}$$

If to any state $i_0,i_1,\cdots,i_{n-1}$, (i, j and n≥0), there exists

$$P\left(X_{n+1}=j\,|\,X_{n-1}=i_{n-1},\cdots,X_1=i_{1,}X_0=i_0\right)=P\left(X_{n=1}=j\,|\,X_n=i\right)$$

Random process like this can be named Markov Chain.

Definition 2.3 Suppose that E is the state space of random process $\left\{X_t,t\geq 0\right\}$, for any arbitrary integer $i_k\in E,0\leq k\leq n+1$, there exists

$$P\left(X_{t_{n+1}}=j\,|\,X_{t_0}=i_0,X_{t_1}=i_1,\cdots,X_{t_n}=i_n\right)=P\left(X_{t_{n+1}}=i_{n+1}\,|\,X_{t_n}=i_n\right)$$

Then $\left\{X_t,t\geq 0\right\}$ can be named as Markov Chain of continuous time. The formula above is named as transition probability of the chain.

1.2 Relative Concept of Financial Time sequence

1.2.1 Stock Price Fluctuation Process

The studies on price process of security are not only the basis of studying the price fluctuation and yield of securities but also the basis of pricing model and numeric analysis of related financial derivatives. The main methods utilized in this field are binomial model, finite difference method and Monte Carlo Simulation. In this thesis, the study of stock price fluctuation is conducted via a model of studying the price process and yield of securities based on a model similar to binomial model and with adding percolation model.

Given that the stock price conforms to the following equation:

$$dS_t=\mu Sdt+\sigma SdW_t \tag{1}$$

Concluded from the Ito Integral:

$$S_t = S_0 e^{\mu t - \frac{1}{2}\sigma^2 t + \sigma W_t} \tag{2}$$

It is easy to know:

$$E(S_T) = S_0 e^{\mu T} \tag{3}$$

In real monetary and securities market, the price fluctuation is so complicated that it is usually simplified into the following process:

$$P\left(S_n = \mu_n S_{n-1}\right) = \alpha\,,\ P\left(S_n = S_{n-1}\right) = 1-\alpha-\beta,\ P\left(S_n = d_n S_{n-1}\right) = \beta$$

1.2.2 Yield

Price, yield and risk are three influential variables in the field of finance. The processes of price, yield and risk will be established respectively with the application of random process.

Suppose that the price process of a stock at the moment ***t*** is $\{ S_t : t = 1,2...T \}$, and the initial price at t=0 is S_0.

In real life, the data that the researchers analyzed is a yield sequence derived from the calculation of price sequences instead of a price sequence itself. There are two reasons hidden by this method: firstly, the purpose of conducting work of statistics and analysis is to offer some references to investors. So the collected statistics are from the perspective of investors. For an investment project, compared to the price, the investment yield attracts more attention of the investors. Secondly, the yield holds better statistical properties and features than price for those statisticians. Based on these two points, the thesis will conduct analysis focusing on yield.

There are usually two kinds of yields: relative yield and logarithmic yield. The latter refers to the compounding of the relative yield and is named as the continuous composite yield. (See reference [10, 11])

In the defined price process above, the relative yield from ***t*** to ***t+r*** is:

$$R'_\tau = \frac{P_{t+\tau} - P_t}{P_t} \tag{4}$$

And the logarithmic yield is:

$$R_\tau = \log P_{t+\tau} - \log P_t \tag{5}$$

Viz, if it is counted to be *m* phase of stock, then the formula above can be interpreted as:

$$P_{t+\tau} = P_t\left(\frac{1+R_\tau(t)}{m}\right)^m \tag{6}$$

And when $m \to \infty$, then

$$P_{t+\tau} = P_t \exp(R_\tau(t)) \tag{7}$$

After comparing these two yield calculation formula, the researchers find that the former has great defects though it is comparatively intuitive and simple. For instance, it is not practical for multi-phase yield calculation. Because multi-phase yields is the product of every single phase yield, and the random variation, which causes contradictions. In other words, the single phase yield follows the Gaussian distribution but turn out not to follow it after the product operation. However, logarithmic yield is free from those defects. So unless stated in this thesis, the yield indicates the logarithmic yield.

1.2.3 Yield Process

Under the definition of stock price, the yield process can be derived.

The yield of the time ***t*** can be defined as R_t **,** then

$$R_t = \ln S_t - \ln S_{t-1}, t = 1,2...T \tag{8}$$

And $R_0 = 0$, then { $R_t : t = 0,1...T$ } is the yield process of the price process { $S_t : t = 1,2...T$ }. It is also called yield process.

From the formula above, the relation between yield process and price process can be derived:

$$\log S_t = \log S_{t-1} + R_t \quad \text{t=1,2,....T} \tag{9}$$

Furthermore,

$$S_t = S_{t-1} e^{R_t} \quad \text{t=1,2,....T} \tag{10}$$

$$\text{To sum up, } S_t = S_0 \exp(R_t + R_{t-1} + ... + R_1) = S_0 \exp(\sum_{k=1}^{t} R_k) \tag{11}$$

2. Stock Price Model based on Percolation Theory

2.1 Percolation Theory

2.1.1 Fundamental Definition of Percolation Theory

Firstly, the fundamental definition of Percolation Theory is given in this passage.

$\mathbb{Z}^d$ is considered to be the vector in dimension *d*, and $\mathbb{Z}^d$ is a set of vector $x = (x_1, x_2, \ldots, x_d)$. On condition that $d \geq 1$, for $x \in \mathbb{Z}^d$, x_i is marked to be the No. i coordinate of x.

Definition 3.1 $\delta(x, y) = \sum_{i=1}^{d} |x_i - y_i|$ is the distance from x to y.

Definition 3.2 $\delta(0, x) = |x|$ is the distance from origin to x.

When $\delta(x, y) = 1$,, a side is added to Dot x and Dot y and the distance between these two dots is 1, then Z^d can be regarded as a lattice figure with *d* dimension(s). This figure can be marked as L^d.All sides are marked as E^d. $L^d = (\mathbb{Z}^d, E^d)$

Definition 3.3 If $\delta(x, y) = 1$, then Dote x and Dote y are defined to be connected and can be marked as $\langle x, y \rangle$.

Then, the next part focuses on the introduction of communication probability.

For those figures meet the conditions of '$0 < p \leq 1$, $p + q = 1$,' p and q are respectively the open interval and closed interval of all sides of Ld. Meanwhile, $\Omega = \prod_{e \in E^d} \{0,1\}$,the probability for all sides of being open and closed intervals are mutual independent from each other. A dot is taken as the sample marked as $\omega = (\omega(e) : e \in E^d)$.Then as $\omega(e) = 0$, it indicates that e is closed; as w (e) =1, it indicates that e is open. If F is marked as the algebra σ derived from subset of Ω and p is the measure on (Ω, F), then $p_p = \prod_{e \in E^d} \mu_e$, Here μ_e is the Bernoulli Measure at $\{0,1\}$ and $\mu_e(\omega(e) = 0) = q, \mu_e(\omega(e) = 1) = p$. On the basis of different sides, for the probability space of every side e $\mu_e(\omega(e) = 0) = q, \mu_e(\omega(e) = 1) = p$, it can be concluded that $p_p = \prod_{e \in E^d} \mu_e$ and $\mu_e(\omega(e) = 0) = q(e), \mu_e(\omega(e) = 1) = p(e)$.

Definition 3.4 $x_0, e_0, x_1, e_1 \cdots, e_{n-1}, x_n$ is a sequence constructed on different peaks of x_i, e_i. If it meets the condition of $e_i = \langle x_i, x_{i+1} \rangle$, then it can be named a route from x_0 to x_n with the length of *n*. A part of sides and peaks of L^d are taken into consideration. The connected parts in the diagram can be defined as open string. $c(x)$ is the open string including x. $c(x)$ is the number of points on the string.

Definition 3.5 If $|c(x)| = \infty$, then percolation occurs.

Definition 3.6 $\theta(p) = p_p(|c| = \infty)$ is named the probability of side percolation.

Lemma 3.1 There exists pc, if $0 < p_c < 1$, then θ (p) =0; if p<pc, then θ (p); if p>pc, then pc is named the critical value of percolation.

Definition 3.7 $p_c = \sup\{p : \theta(p) = 0\}$.

2.1.2 Fundamental Definition of Percolation Theory

Z^2 is defined to be the two-dimensional space, and another definition is given as following: $T = \left\{\left(x, y\right) \in Z^2 : x + y \text{ is even}, x \text{ and } y \text{ are integers}, y \geq 0\right\}$. For a better understanding, the Diagram One is offered to explain this process. Firstly, two-dimensional space Z2 can be depicted by a square lattice figure like the left part in Diagram One. The right part is from the left one with a counter-clockwise rotation of 45 degree. Every node in the diagram is every point in set T. The nodes are connected by the sides. The arrows on the sides are the directions of percolation.

For any point $\left(x_i, y_i\right) \in T$, the water flows to the point $\left(x_i + 1, y_i + 1\right)$ or $\left(x_i - 1, y_i + 1\right)$ with the probability p_{ij}. p_{ij} Stand for the connectivity probability of every side. That is to say, p_{ij} is the probability of percolation from one point to another, and then $1 - p_{ij}$ is the probability of percolation failure form the point to next point.

The following definition is given: if there is an open circuit from point to x to point y, then these two points are connected. Moreover, there is point sequence $x = x_0, x_1, \cdots, x_j = y$ in set T. And to any value $k < j$, there is the open circuit from point x_{k-1} to x_k that can be marked as $x \mapsto y$. All route that includes point x can be expressed by $\mathbb{F} = \left\{y : x \mapsto y\right\}$. Additionally, all points that can be connected from the original point $0 = (0,0)$ can be marked as $\mathbb{F}\left(0\right)$.

$\xi_n^{(0)} = \left\{m : \left(m, n\right) \in \mathbb{F}\left(0\right)\right\}$ is marked as the state of the point that the original point goes after n steps. $\mathbb{F}_\infty = \left\{\left|\mathbb{F}\left(0\right)\right| = \infty\right\}$ is defined as the infinite open string event $\mathbb{F}_\infty$. Namely, there exists the infinite string that initially starts from the original point 0. Meanwhile, there is an important concept in Percolation Theory called percolation critical value that for any point n there sexist the probability of infinite open string marked as $p_c = \inf\left\{p : p\left(\xi_n^{(0)} \neq \varnothing, \forall n\right) > 0\right\}$

When p>pc, the percolation string has the tendency to last infinitely. Under this circumstance, the percolation occurs. From references 4 and 5, it is known that the critical value of oriented percolation is 0.64≤pc≤0.68.

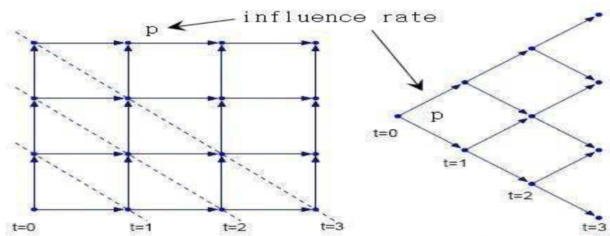


Diagram One: Schematic diagram of two-dimension lattice oriented percolation; the right is from the left one with a counter-clockwise rotation of 45 degree; p is the percolation probability.

2.2 Constructing Model of Stock Price Based on Percolation Theory

2.2.1 Hypotheses

Hypotheses One: In two-dimension Oriented Percolation Model, there is an investor on every single lattice point. When there is some information spread, the mutual communication among the investors turns into a percolation string.

Hypotheses Two: Once an investor receive information and is involved into one percolation string, he or she will no longer receive any other information. That is to say, the investment decision that the investor will make is free from the interference of any other information.

Hypotheses Three: The quantity of investors can be limited into a two-dimension lattice figure with the side length of L, and then the percolation model can be structured on a L×L two-dimension plane network.

Hypotheses Four: For a random side i, pi is the probability of the flow percolating to next point, then 1-pi the probability of the flow failing to percolate to next point.

Hypotheses Five: During the propagation of information, suppose that all percolation actions happen at the same moment.

Hypotheses Six: In the initial state, the investors have their investment propensities and choose to buy or sell with equal probabilities.

Hypotheses Seven: An investor has three choices: buy, sell or neutral that are respectively marked as 1,-1 and 0 with α, β and 1- α-β as their probabilities.

Hypotheses Eight: In the initial state, the transaction behaviors of investors happen at the same probability and α=β=0.5.

2.2.2 Structuring Model

Despite the fact that Percolation Theory firstly originated from the study of nature physics, now it is widely applied in all kinds of scientific studies and it can portrays any incident that holds features of percolation. In stock market, the investors usually respond to the same information with the same transaction decisions and can influence each other because of mutual communication. This is the cause of price fluctuation in stock market. This process of spreading information is quite similar to the water percolation, so the percolation model is used to portray and display the Herd Behavior of the investors and thus to depict the price fluctuation in the stock market.

Yield and financial price fluctuation in financial market are always the focuses in the field of financial mathematics. Through structuring and analyzing model of financial market price fluctuations, the conclusions on yield can be drawn. From the Graph One, a vivid impression of oriented percolation process can be gained, namely the flow of water among the two-dimension lattice points. Here the flow is limited to only two directions that are mutually perpendicular. On the basis of definitions above, a specific financial model based on Oriented Percolation Theory is structured to study the price fluctuations and yield in financial market. In this two-dimension oriented percolation model of Graph One, the hypothesis goes that there is an investor at one lattice point. When there is some information spread, the mutual communication among the investors turns into a percolation string. (Hypotheses One) This process is corresponding to the percolation string derived from all lattice points that the water flow starts from the original point and passes a route forward in the percolation model. Meanwhile, once an investor receive information and is involved into one percolation string, he or she will no longer receive any other information. That is to say, the investment decision that the investor will make is free from the interference of any other information.(Hypotheses Two) Furthermore, the quantity of investors can be limited into a two-dimension lattice figure with the side length of L, then the percolation model can be structured on a L×L two-dimension plane network. (Hypotheses Three)

Firstly, suppose that the water starts from the original point and flows to two sides that are mutually perpendicular. If the side is open, then the water flows to next lattice point; if the side is closed, then the water cannot flow to next lattice point. And the water flows to every other lattice point in this mode. For a random side ***I***, ***p_i*** is the probability of the flow percolating to next point, then **1-*p_i*** the probability of the flow failing to percolate to next point. (Hypotheses Four) This is corresponding to the situation that two investors spread the information in the stock market with a probability of ***p_i*.** The whole model is corresponding to the following situation in financial market: in a specific group of investors, one of them firstly get the information, and it is possible for this investor to spread this information to other investors in specific directions; then the investors of successful communication continue this process and spread the information to people of specific directions at a specific probability. During the propagation of information, suppose that all percolation actions happen at the same moment. (Hypotheses Five) So when percolation occurs, there is a determined quantity of investors on the percolation string in the L×L two-dimension plane network. That is to say, the quantity of investors influenced by the information is determined as well. Only this group of investors can influence the transaction results that refer to the fluctuations of stock price. Based on Percolation Theory, the coverage area of the percolation process, or the proportion of percolation string in whole space of the network, is decided by the probability of successful percolation of every single side. And the proportion is the determinant of setting the stock price. Thus the thesis will focus on discussion of percolation probability in analysis.

Another issue needed to be considered is activity ratio of investors or the probability of investors' transactions. Here it is defined to be the sum of buying and selling probabilities. We suppose that the investors have their investment propensities in the initial state and choose to buy or sell with equal probabilities.(Hypotheses Six) An assumption is given that an investor has three choices: buy, sell or neutral that are respectively marked as 1,-1 and 0 with α, β and 1- α-β as their probabilities. Here an indicative function is introduced to show the activity ratio of investors:

$$P\left(\operatorname{sgn}\left(x\right)=1\right)=\alpha$$

$$P\left(\operatorname{sgn}\left(x\right)=-1\right)=\beta$$

$$P\left(\operatorname{sgn}\left(x\right)=0\right)=1-\alpha-\beta$$

Besides, we suppose that the transaction behaviors of investors happen at the same probability and α=β=0.5 in the initial state. When a group of investors receive the information and conduct the same investment behavior, it will bring the fluctuations of the stock price. As a group of investors choose to buy, their behavior pushes up the price of the stock; if they choose to sell, their behavior causes to the decrease of stock price. In the modeling, we suppose that there are many investors in L×L two-dimension plane network market. Those investors on the open strings (really they are investors with the same opinion) are marked as $m_k, k=1,2,3...$, and then the proportion of those investors to total investors can be calculated with the following formula:

$$M_l=\frac{\sum m_k}{\left(L+1\right)^2}(l=0,1,2...) \tag{12}$$

We suppose that ξ_l is a random variable that can be chosen as 1or -1 at the same probability. Based on relative theory of financial mathematics [9, 15, 16, and 24], the security price can be presented by the following differential equation:

$$\frac{dS_s}{ds}=S_s\alpha\chi(s) \text{ (c >0 and c is a constant)} \tag{13}$$

Furthermore, α is a random variable parameter and $\chi(s)$ is the function of time. The following formula of security price can be concluded on a discrete time sequence:

$$\frac{S_{t+1}}{S_t}=\exp\{\alpha\chi(t)\}(t=1,2,3...) \tag{14}$$

From the relative definitions we can conclude the following formula from the model of security price:

$$S_t = S_0 \exp\left\{-\alpha \sum_{l=1}^{t} \xi_l \left|M_l\right|\right\} \tag{15}$$

S_0 is the security price at the initial time (t=0). From time t to t+1, the logarithmic yield and its absolute value are defined as following:

$$r_t = \ln S_{t+1} - \ln S_t \tag{16}$$

$$\left|r_t\right| = \left|\ln S_{t+1} - \ln S_t\right| \tag{17}$$

Until now, we have roughly finished the construction of model, the price fluctuation process and its yield. Then the thesis will cover the generation of data via the model, relative analysis and discussion on value of the model parameter.

2.3 Simulation of Yield and Price Fluctuation in Financial Market

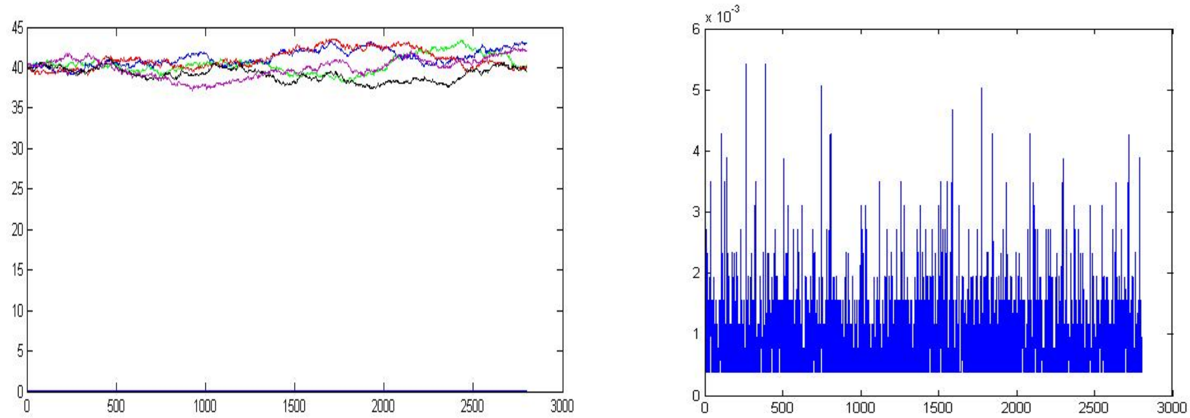
In this section, the study of price fluctuations and yield will be conducted by utilizing the previous model of stock price fluctuation. Firstly, the security price and yield are simulated via the software Matlab.

During the process of computer simulation, several important parameters are the core issues discussed in this thesis. An assigned value $L = 200$ defines the market in the model as a 200×200 plane network space. We assign the value $p = 0.4, 0.5, 0.68, 075, 0.8, 0.9$ and respectively study the price fluctuation with different percolation probabilities. The time sequence length is assigned to be 3,000, indicating that the time length of simulation will be 3,000 days. Then multiple simulations should be conducted to analyze the basic statistical features of the simulated data.

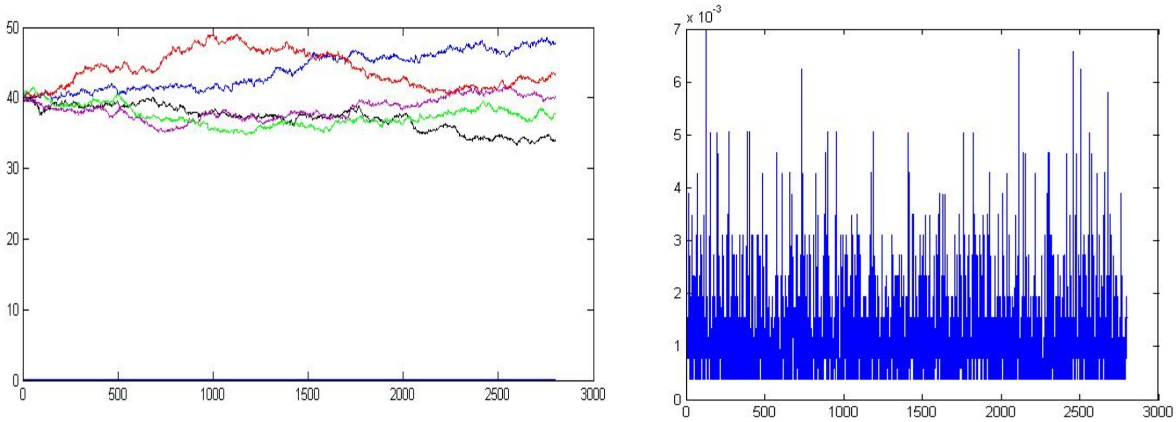
The initial stock price is assigned as 40. Firstly, we observe the situation when the percolation probability less than 0.64. Based on Percolation Theory, the critical value for the occurrence of percolation is $p_c \in (0.64, 0.68)$.If the percolation probability is less than 0.64, then there is no phenomenon of percolation in the system. From Graph Two and Three, it is obvious that the stock price fluctuation has a trend of moderation when p=0.4 or 0.5 and the absolute yield of the stock is not exceeding 6×10^{-3}. This is identical to the situation when no percolation phenomenon occurs: there is no Herd Behavior resulting into the stock price fluctuation among investors.

Then we observe the simulated result of financial time sequence model when p=0.68, 0.75 or 0.9 and there is occurrence of percolation. The situation of stock price fluctuation simulated for many times is displayed from Graph Four to Six when p is assigned to be a specific value. The curves with different colors indicate different simulations. The absolute yield of stock is shown in Graph Eight to Eleven when p is assigned with different values. Intuitively, a certain law is found. The greater the value of p, the greater the value range of absolute yield will be. When p=0.68,$\left|r\right|$,the absolute value is not exceeding 1.8%.When p is successively assigned to be 0.75, 0.8 or 0.9, the absolute value is respectively $\left|r\right| < 2.5\%$,$\left|r\right| < 5\%$and $\left|r\right| < 15\%$.Meanwhile, the fluctuation range increases with the greater assignment value of p.

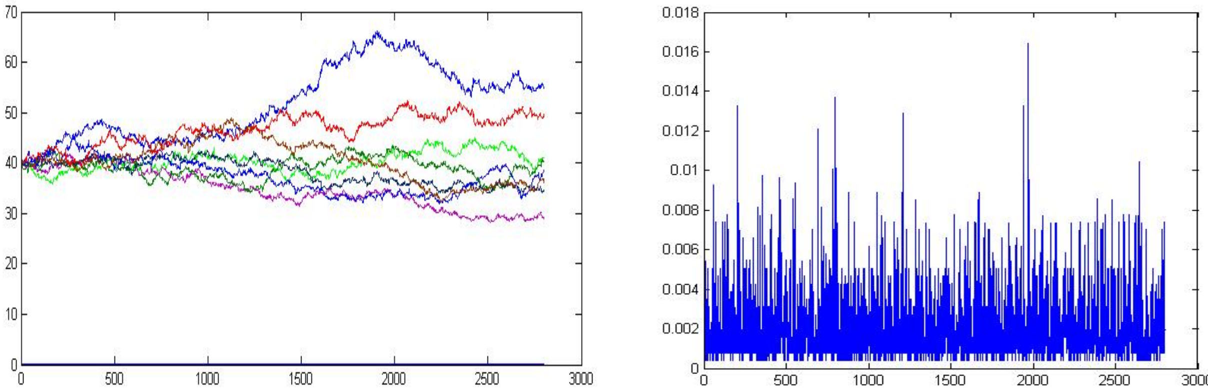
These phenomenon are consistent with previous analysis of the model constructed on the basis of Percolation Theory. In percolation model, the value of p stands for water diffusion ability. In the price model, p stands for the information communication ability among the investors. Thus p becomes an important parameter to determine the stock price simulation results. If the value of p is greater, there will be greater ability of information communication among the investors and larger information diffusion range and quantity of investors influenced by the information. Therefore the transaction operations of investors bring greater influence and larger range of fluctuations to the stock price. Whereas, if the value of p is smaller, there will be smaller range of information communication. And transaction operations of investors bring smaller influence and range of fluctuations to the stock price. This law is well-revealed by the status of the graphs of the simulate data.



Graph Two: The price fluctuation process and absolute value of yield when p=0.4; on the left is the price fluctuation process and on the right is the absolute value of yield.



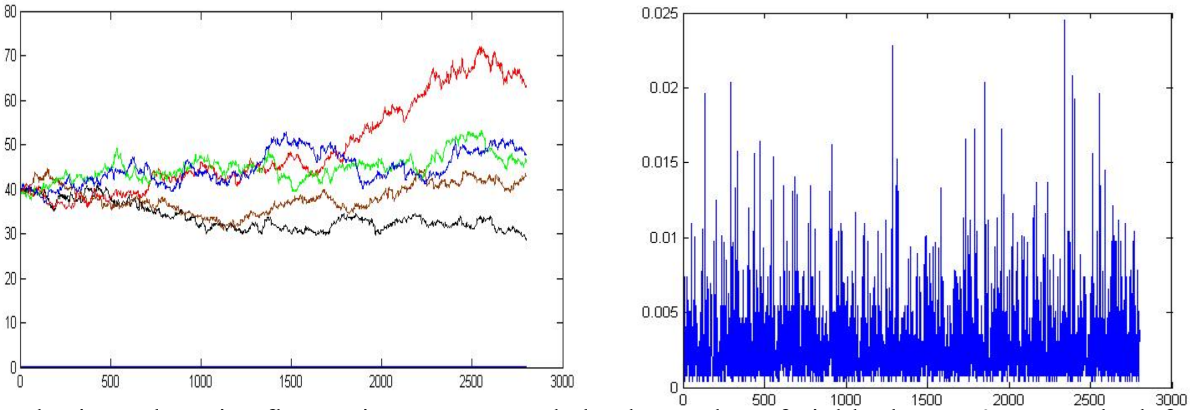
Graph Three: The price fluctuation process and absolute value of yield when p=0.5; on the left is the price fluctuation process and on the right is the absolute value of yield.



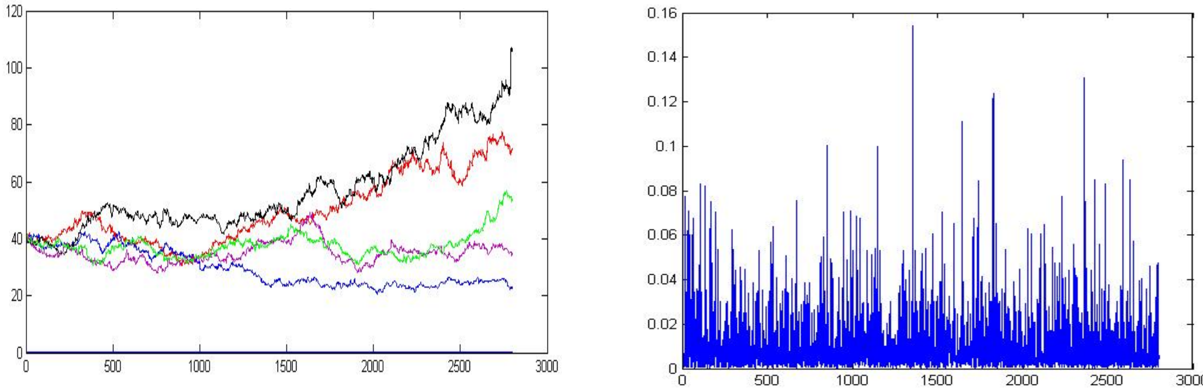
Graph Four: The price fluctuation process and absolute value of yield when p=0.68; on the left is the price fluctuation process and on the right is the absolute value of yield.

2.4 Chapter Summary

Firstly, the fundamental definition and property of Oriented Percolation Theory are covered in this chapter. Then the model of stock price fluctuation is constructed via taking advantage of two-dimension oriented percolation model. Then the model of stock price fluctuation is used to simulate the situation of stock absolute yield and price fluctuation. Further research on features of stock price fluctuation is conducted under the circumstances of different assignment values of percolation probability. The results turn out to be identical with the Percolation Theory and universal law of the real market. The rationality of financial time sequence model is preliminarily proved.



Graph Five: The price fluctuation process and absolute value of yield when p=0.75; on the left is the price fluctuation process and on the right is the absolute value of yield.



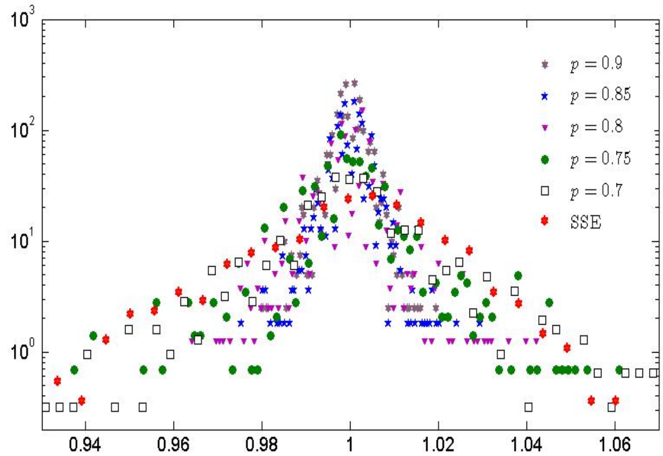
Graph Six: The price fluctuation process and absolute value of yield when p=0.9; on the left is the price fluctuation process and on the right is the absolute value of yield.

3. The Simulation and Analysis of Stock Price and Yield

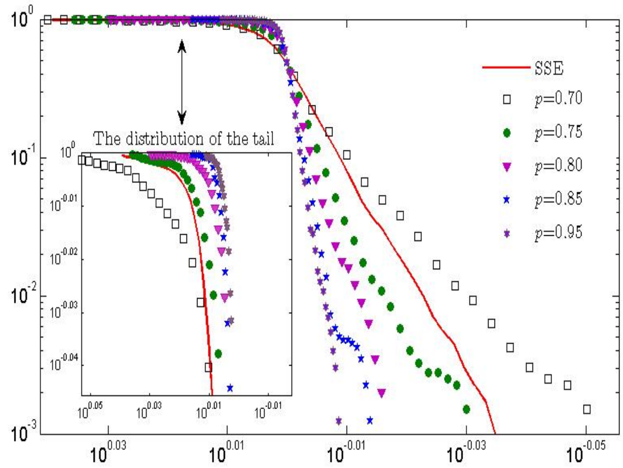
3.1 Yield Distribution of Simulated Data and Real Market Data

In this chapter, the author analyzes the distribution of stock yield from the perspective of real market data. Meanwhile, a comparison is made between distributions of stock yield simulated data and real market data. Here we take data samples of SSE from 2005 to 2015 for analysis. And the parameters of simulated data are respectively assigned as *p*=0.7,0.75,0.8,0.85,0.9.In the study, compared to the fact that financial time sequence is identical to Gaussian distribution in traditional theories, it is found that both real market data and simulated data embody the feature of sharp peaks and fat tail. Especially for simulated data, as the percolation parameter p increases, it is more obvious for the feature of sharp peaks and fat tail. Furthermore, the cumulative distribution tail of stock yield is consistent with the distribution of power law.[6] It can be portrayed by the mathematical formula $P(r_t > x) \sim x^{-\mu}$ ($\mu \approx 3$). r_t is the yield when stock price is in the given interval Δt. And the simulated data is greatly identical to this feature.

It is more obvious for the feature of sharp peaks and fat tail of simulated data. When *p* is assigned to be 0.75, the graph of simulated data is the most identical to the real data graph. In terms of the cumulative function distribution, it is found that both simulate and real data are consistent with the power law distribution. Meanwhile, when the percolation parameter *p* is 0.75, the real data is closest to simulated data from the observation of the graph.



Graph Seven: The distribution of yield probability density of simulated data and real data with different values



Graph Eight: The yield calculative distribution of simulated data and real data with different values.

3.2 Statistical Analysis of Simulated Data and Real Market Data

This section will mainly cover the discussion on statistical property differences of simulated data and real market data via two-dimension oriented percolation model. We assign percolation probability p with different values to analyze simulated data and real market data fluctuation trend and statistical properties of some eigenvalues, including logarithmic yield, absolute yield and chain price, etc,. The comparison of their differences will be presented through the statistics of their maximums, minimums, standard deviations, mean values, skewness and kurtosis, etc,. The following mathematical expressions are the parameters of skewness and kurtosis:

$$偏度(Skewness)=\frac{\sum_{i=1}^{n}\left(q_i-u_q\right)^3}{(n-1)\sigma^3},\quad 峰度(Kurtosis)=\frac{\sum_{i=1}^{n}\left(q_i-u_q\right)^4}{(n-1)\sigma^4},$$

q_i is the yield of the $i$trading day; u_q is the mean value of q_i ; n is the sum of the data; σ is the standard deviation of relative data. Skewness and kurtosis respectively depict the symmetry and aggregation of the data. It is known that the skewness and kurtosis of standard Gaussian distribution are respectively 0 and 3.Hence when the data kurtosis is greater than 3, the data turns out to be more concentrated in the central section with longer tail compared to Gaussian distribution.

Diagram One is the statistic results of calculating the logarithmic yield and its absolute value by taking simulated data with percolation probability from 0.6 to 0.9 with a interval of 0.5 form SSE and SZSE.

Diagram One is the depiction on relative statistical properties of simulated data and real market data. Firstly, the fact that all kurtosises, with every parameter value of simulated data and real data included, are greater than 3 and all skewnesses are greater than 0.1 can be discovered from the observation of Diagram One. This fact indicates that no matter it is the simulated data based on the model or the real market data, the yield distribution is different from Gaussian distribution and turns out to be more volatile. Generally, to different values of percolation probability p, when it is greater, the relevant kurtosis will be greater as well. In other words, the sharp peaks and fat tail turn out to be more obvious with the increase of p.

Normally, the greater the percolation probability p, the larger the percolation range will be. This fact can be depicted by the constructed model as following: the greater the information communication probability, the larger the communication range will be; then there will be more investors receiving the information and the influence caused by relevant investors' behavior will be greater; furthermore, the stock market price fluctuation will be more volatile. Meanwhile, it is found that all statistical factors of yield in the model is very approximate to relevant real data of SSE Index when p is 0.75. If a comparison of absolute yield and chain price is conducted, the statistical properties of simulated data and real data are approximate when minor adjustment of p is made. More simulation comparisons can be made by minor adjustment of assigning the percolation probability close to 0.75. Through the quantitative comparisons of these statistical properties, a deduction can be drawn that the price sequence in real financial security market can be comparatively better simulated through the simulated price fluctuation process of the financial model based on the application of two-dimension percolation structure. Both of them share the similar distribution and the property of sharp peaks and fat tail.

	n	r	Minimum	Maximum	Mean	Std. dev.	Skewness	Kurtosis
$p = 0.6$	1000	r	-0.013978	0.011066	1.30460e-4	2.9602079e-3	-0.199	4.543
$p = 0.65$	1000	r	-0.013978	0.011648	-1.04251e-4	3.3486298e-3	-0.124	4.100
$p = 0.7$	1000	r	-0.020384	0.019220	-2.19566e-4	4.5971004e-4	-0.205	5.193
$p = 0.75$	1000	r	-0.025044	0.029121	1.78221e-4	5.4244736e-3	0.150	5.985
$p = 0.8$	1000	r	-0.036692	0.041351	3.92544e-4	8.5229265e-3	0.138	5.911
$p = 0.85$	1000	r	-0.073966	0.068725	-6.1153e-5	1.3616295e-2	0.278	8.674
$p = 0.9$	1000	r	-0.174140	0.139780	2.23043e-4	2.5688673e-2	-0.215	12.989
$p = 0.6$	1000	$\|r\|$	5.824e-4	0.013978	2.310996e-3	1.8530565e-3	1.932	7.980
$p = 0.65$	1000	$\|r\|$	5.824e-4	0.013978	2.623750e-3	2.0816444e-3	1.745	6.684
$p = 0.7$	1000	$\|r\|$	5.824e-4	0.020384	3.434468e-3	3.0617398e-3	2.056	8.247
$p = 0.75$	1000	$\|r\|$	5.824e-4	0.029121	3.982520e-3	3.6851546e-3	2.280	9.937
$p = 0.8$	1000	$\|r\|$	5.824e-4	0.041351	6.073375e-3	5.9893001e-3	2.097	8.405
$p = 0.85$	1000	$\|r\|$	5.824e-4	0.073966	8.736738e-3	1.0440318e-2	2.614	11.591
$p = 0.9$	1000	$\|r\|$	5.824e-4	0.174140	1.494059e-2	2.0892892e-2	3.293	13.874
SSE	1000	r	-0.092558	0.090368	3.7956e-4	2.1727466e-2	-0.403	4.929
SSE	1000	$\|r\|$	0.00000	0.092558	0.01588397	1.4821407e-2	1.737	6.934
SZSE	1000	r	-0.097504	0.091647	8.0019e-4	2.3934175-2	-0.410	4.364
SZSE	1000	$\|r\|$	0.00000	0.097504	0.01787486	1.5926557e-2	1.554	6.842

Diagram One: The statistical properties of simulated data and real data

4. Further Analysis of Simulated Data Based on VAR Model

4.1 Model Introduction

Previously, the fundamental statistical properties of data simulated by the model are verified via the applications of various methods. Now the time sequence VAR Model is introduced via adding the simulated data into the real financial time sequence to study their dynamic relation and further analyze whether the simulated data embodies the property of replacing a group of real data.

Usually, the VAR Model is applied to the prediction of relative time sequence system and the study on dynamic influence of stochastic disturbance on variable system. Compared to structural modeling, the principal advantage of VAR Model lies in no need to differentiate response variable and explanatory variable. In other words, all variables in the system can be signified by other variables of their hystereses. The following is the mathematical expression of VAR Model:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + B_1 x_t + \ldots + B_r x_{t-r} + \varepsilon_t \tag{19}$$

In this expression, y_t is the endogenous variable in dimension. x_t is the exogenous variable in dimension d . $A_1 \ldots A_p$ and $B_1 \ldots B_r$ are the parameter matrices needed to be estimated. There are respectively p and r phases of hysteresis in endogenous variable and exogenous variable. ε_t is a random disturbance term. A VAR Model with phases of endogenous variables can be named VAR (p) Model. Normally, the researcher's hope that the hysteresis phase of p and r can be as long as possible that the dynamic features of the Model constructed can be comprehensively and completely reflected. However, once the hysteresis phase is longer, there will be more parameters needed to be estimated and lower degree of freedom. So the researchers need to maintain a proper balance between hysteresis phase and degree of freedom. AIC and SC are the standards commonly used.

The VAR Model is based on four groups of time sequences, including simulated data sequences from SSE Index and SZSE Index when the values of percolation probability p are respectively 0.75 and 0.8. The selected time sequences shall be stationary time sequences is the fundamental condition to structure VAR Model to analyze their dynamic relations. If this condition is not met, it is impossible to structure the VAR model. Firstly, the researchers need to analyze whether the four groups of time sequences or their difference sequences in same phase meet the fundamental condition. The method used to verify is Unit Root Test. An introduction of Unit Root Process is given in following passage.

4.2 Unit Root Test

This section will cover a further study of the distribution property and comparison of parameter β and α .The discussion and study are from the perspective of parameter difference (or parameter variation) $\Delta\beta$ and $\Delta\alpha$.

If the mean value or auto covariance function changes with the time sequence, it is can be concluded that this time sequence is not stationary time sequence.

For random process $\{y_t, t = 1, 2 \ldots\}$,if $y_t = \rho y_{t-1} + \varepsilon_t$, $\rho = 1$ and ε_t is a stationary process; moreover, $E(\varepsilon_t) = 0, \operatorname{cov}(\varepsilon_t, \varepsilon_{t-s}) = \mu_t < \infty$ .Then this process is the Unit Root Process.

If the unit root becomes a stationary process via the first difference, then

$$y_t - y_{t-1} = (1 - B) y_t = \varepsilon_t \tag{20}$$

Meanwhile, time sequence is the union root sequence that can be marked as $I(d)$. If non-stationary time sequence can turn into stationary time sequence via difference of d times, then it is named unit root sequence in phase d and can be marked as $I(d)$ .d stands for the phase and is also the number of unit roots in the sequence.

The time sequence stationarity can be estimated via Unit Root Test. The author applies the methods of Augmented Dickey and Fuller (ADF) and Phillips and Perron(PP). We respectively sample the natural logarithm from four original sequences to reduce the impact of disturbance term and get four new initial sequences: LSSE, LSZ, L0.75 and L0.8. The significant levels of 1%, 5% and 10% are tested. The selection of hysteresis is based on AIC. The following diagram involves the results of test on original sequences.

(a) LSSE

	ADF test	PP test
1% level	-3.435871	-3.435861
5% level	-2.863866	-2.863861
10% level	-2.568059	-2.568057
Prob.*	0.9232	0.9148
t-Statistic	-0.294523	-0.350371

(b) LSZ

	ADF test	PP test
1% level	-3.435861	-3.435861
5% level	-2.863861	-2.863861
10% level	-2.568057	-2.568057
Prob.*	0.9079	0.9129
t-Statistic	-0.392956	-0.362293

(c) L0.75

	ADF test	PP test
1% level	-3.435866	-3.435861
5% level	-2.863864	-2.863861
10% level	-2.568058	-2.568057
Prob.*	0.7105	0.6945
t-Statistic	-1.118418	-1.157455

(d) L0.8

	ADF test	PP test
1% level	-3.435861	-3.435861
5% level	-2.863861	-2.863861
10% level	-2.568057	-2.568057
Prob.*	0.8177	0.8316
t-Statistic	-0.802248	-0.751061

Diagram Two: Unit root test of original sequences

From the results of Diagram Two, for the original sequences, the tested statistic values are higher than the critical value of significant level 10%.So it can not reject the null hypotheses and there are unit roots in original sequences. Thus these groups of sequences are not stationary and cannot be used to structure VAR Model. Four new time sequences derive from the first difference of four original sequences. The new sequences are respectively marked as D (LSSE), D (LSZ), D (L0.75), D (L0.8). Diagram Three is the results of the unit root test of the new sequences.

(a) D(LSSE)

	ADF test	PP test
1% level	-2.566992	-2.566990
5% level	-1.941101	-1.941101
10% level	-1.616512	-1.616513
Prob.*	0.0000	0.0000
t-Statistic	-26.98089	-36.30887

(b) D(LSZ)

	ADF test	PP test
1% level	-2.566990	-2.566990
5% level	-1.941101	-1.941101
10% level	-1.616513	-1.616513
Prob.*	0.0000	0.0000
t-Statistic	-35.87788	-35.89813

(c) D(L0.75)

	ADF test	PP test
1% level	-2.566990	-2.566990
5% level	-1.941101	-1.941101
10% level	-1.616513	-1.616513
Prob.*	0.0000	0.0000
t-Statistic	-37.20145	-37.24487

(d) D(L0.8)

	ADF test	PP test
1% level	-2.566990	-2.566990
5% level	-1.941101	-1.941101
10% level	-1.616513	-1.616513
Prob.*	0.0000	0.0000
t-Statistic	-36.18504	-36.13937

Diagram Three: Unit root test of first difference sequence

It is found that the null hypotheses can be rejected by ADF and PP test results of these four groups of sequences. The fact indicates that the first difference sequence is a stationary time sequence and the original sequence is the unit root sequence of first difference. In other words, the sequence can turn into a stationary sequence via the first difference. The VAR Model is structured based on the first difference sequence.

4.3 Analysis of VAR Model

VAR Model can analyze every group of time sequence and the relations of their hysteresis via solving their parameters. Meanwhile, it can make a prediction of the future trend. Here the VAR Model is structured and solved on two groups of simulated data and real data from China's SSE Index and SZSE Index. From previous conclusion, the first difference sequence is a stationary time sequence that can be used to structure VAR Model. The results can be estimated form software Eviews and described in following way:

There is the formula Y = [D (LSSE), D (LSZ), D (L0.75), D (L0.8)]'

The parameter matrix can be concluded. The researchers can conduct quantitative analyses to know the dynamic influences of other terms caused by the variations of these groups of time sequences. It is easy to find that the simulated data sequences from the model can be added into the VAR Model to replace the real financial time sequences. Moreover, to verify the rationality of VAR Model, the statistical property tests should be conducted to every equation of D (LSSE), D (LSZ), D (L0.75), and D (L0.8). The results are represented in Diagram Four.

	D(LSSE)	D(LSZ)	D(L0.75)	D(L0.8)
R-squared	0.998471	0.998095	0.994277	0.993106
Adj. R-squared	0.998455	0.998075	0.994216	0.993032
Sum sq. resids	0.698415	0.562561	0.548942	0.688897
S.E. equation	0.024894	0.022342	0.022070	0.024724
F-statistic	61329.61	49210.81	16317.32	13528.66
Log likelihood	2599.114	2722.413	2736.381	2606.935
Akaike AIC	-4.53704	-4.75335	-4.77786	-4.55076
Schwarz SC	-4.47958	-4.69589	-4.72040	-4.49330
Mean dependent	8.819891	7.708904	4.221195	4.182464
S.D. dependent	0.633273	0.509209	0.290202	0.296192

Diagram Four: Results of statistical property tests

It is obvious that both R-square values and adjusted R-square values are close to 1. This fact indicates that the VAR Model is reasonable and can comparatively reckon the accurate parameters of each group. Four groups of data sequences in the model share high correlation. This correlation does not only share between in China's SSE Index and SZSE Index, but also between the simulated data and real data.

5. Conclusion

In this thesis, the price fluctuation of financial security market is simulated via two-dimension Oriented Percolation Model with the application of Percolation Theory. It is mainly based on the water percolation process to simulate the information communication among investors for studying the influences of investors' behavior to price fluctuation in the market. Many groups of simulated data are gained through model construction of assigning different values to parameter named percolation probability p. The statistical property and distribution of these simulated data are analyzed. The study and comparison are made on real data of China's SSE Index and SZSE Index in five recent years. It is found that when parameter p is assigned with proper value(the proper value estimated here is 0.75-0.8), the sequences of price fluctuation process, or yield sequence, share with the data the similar statistical distributions and properties, including the sharp peaks and fat tail. With the help of VAR Model structured on the sequences of simulated and real data, the method of time sequence analysis is introduced to study their correlation and dynamic influences of fluctuation. It can be safely concluded from these analyses that our two-dimension Oriented Percolation Model can offer good simulation of security price fluctuation process and the simulated data can be used to the analysis and prediction of security price fluctuation in real market.

References

[1] E.Majorana. I1 valore delle leggi statistiche nella fisica e nelle scienze sociali[J]. Scientia,1942, 36: 58-66.
[2] D.Stauffer, Can percolation theory be applied to the stock market?. Annals of physics, 1998.7, 529-538.
[3] M.F. Chen, From Markov chains to non-equilibrium particle systems, World Scientific, Singapore, 1992.
[4] R. Durrett, Lecture notes on particle systems and percolation, Wadsworth and Brooks, California, 1988.
[5] P. Grau-Carles, Long-range power-law correlations in stock returns, Physica A 299 (2001) 521-527.
[6] G. Grimmet, Percolation, Springer-Verlag, Berlin, 1999.
[7] K. Ilinski, Physics Of Finance: Gauge Modeling in Non-equilibrium Pricing, John Wiley, New York, 2001.
[8] D. Lamberton and B. Lapeyre, Introduction to Stochastic Calculus Applied to Finance, Chapman and Hall/CRC, London, 2000.
[9] Q.D. Li, J. Wang, Statistical properties of waiting times and returns in Chinese stock markets, WSEAS Transactions on Business and Economics 3 (2006) 758-765.
[10] Z. Liao, J. Wang, Forecasting model of global stock index by stochastic time effective neural network, Expert Systems with Applications 37 (2010) 834-841.
[11] T.M. Liggett, Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer-Verlag, New York, 1999.
[12] T.M. Liggett, Interacting Particle Systems, Springer-Verlag, New York, 1985.
[13] T.C. Mills, The Econometric Modeling of Financial Time Series (Second Edition), Cam-bridge University Press, Cambridge, 1999.
[14] S. Pliska, Introduction to Mathematical Finance, Blackwell, 2000.
[15] S.M. Ross, An Introduction to Mathematical Finance, Cambridge University Press, Cam-bridge, 1999.
[16] S. Sadique, P. Silvapulle, Long-term memory in stock market returns: international evidence, International Journal of Finance and Economics 6 (2001) 59-67.
[17] D. Sornette, Critical Phenomena in Natural Sciences, Springer-Verlag, Berlin, 2000.
[18] H.E. Stanley, Jr. Andrade, J.S. Havlin, S. Makse, H.A. Suki, Percolation phenomena: a broad-brush introduction with some recent applications to porous media, liquid water and city growth, Physica A 266 (1999) 5-16.
[19] D. Stauffer, T.J.P. Penna, Crossover in the Cont-Bouchaud percolation model for market fluctuation, Physica A 256 (1998) 284-290.
[20] D. Stauffer, A. Aharony, Introduction to Percolation Theory, Taylor & Francis, London, 2001.
[21] H. Tanaka, A percolation model of stock price fluctuations, Mathematical Economics 1264 (2002) 203-218.
[22] C. Tsallis, R. Mendes, A. Plastino, The role of constraints within generalized nonextensive statistics, Physica A 261 (1998) 534-554.
[23] T.S. Wang, J. Wang, Statistical analysis by statistical physics model for the stock markets. International Journal of Modern Physics C 20 (2009) 1547-1562.
[24] Rolski T, Schmidli H, Schmidt V, Teugels J. Stochastic processes for insurance and finance[M]. New York: Wiley, 1998.
[25] Stauer D, Penna T. Crossover in the Cont-Bounchaud percolation model for market uctuations[J]. Physica A, 1998, 256: 284-290.
[26] Black F, Scholes M. The pricing of options and corporate liabilities[J]. Political Economy, 1973, 81: 635-654.