Cross-correlation Network Analysis Work in Stock Market

Wang Penghao

System analysis and Integration Major of the school of system science, Beijing Normal University, BeiJing 100875 China

ph_wang@126.com

Keywords: complex network; stock index; topology structure

Abstract: Complex network theory is a powerful tool to make research on network structure and function of stock market, whose cross-correlation has already become one new perspective of research. It is a network structure of calculating network statistics characteristic index of stock index, by analyzing yield rate and transaction amount, PE. stock yield and transaction index have stronger cross-correlation and small-world property. PE index has weaker cross-correlation without small-word property; it is a random network with low efficiency, which is greatly affected by the outside factors.

1. Introduction

Complex network theory is the powerful tool to make research on internal structure and function of stock market, topology nature and clustering structure of stock cross-correlation network have important meanings for formation mechanism of understanding network and dynamic behavior takes place in network. It starts from dynamic perspective to make research on cross-correlation network in Chinese and uses regression analysis to search cross-correlation and effect among network statistics characteristic in Chinese stock market, as well as make research on relations between synchronization and network topology structure. Scholars in economy and mathematics etc start applying concepts such as network structure to make research on stock market and further on stock cross-correlation. It chooses yield index as research object and makes description on stock market through relevant parameter model and stock network model, make analysis on table division of stock market.

2. Construction method on stock cross-correlation network

Node represents stock in stock cross-correlation network, edge between nodes represents cross-correlation of stock price fluctuation (or measurement reflects other stock cross-correlation, edge weight indicates the detailed cross-correlation, so constructed stock cross-correlation network is undirected weighted network, it uses adjacent matrix to indicate in the actual operation. The detailed stock cross-correlation network construction includes 3 steps: (1) calculate related parameter matrix of stock price fluctuation; (2) converts cross-correlation matrix to distance matrix; (3) apply threshold method to construct final stock cross-correlation network on the basis of distance matrix. Suppose cross-correlation network includes N stocks, observation time window of each stock is [T_0, T_0+T]. The yield rate of stock i at any time τ in observation range is as follows:

$$r_i(\tau) = \ln p_i(\tau) - \ln p_i(\tau - \Delta T) \tag{1}$$

Of which, p_i (τ) is the stock price at time τ，ΔT is time period of calculating yield rate. This each stock forms yield rate series having L observation value, L depends on T and Δ T. According to yield rate series, it calculates cross-correlation coefficient $\rho_{ij}(\Delta T)$, of stock price fluctuation between 2 stock i and j in observation time window, which is as follows:

$$\rho_{ij}(\Delta) = \frac{E[r_i r_j] - E[r_i]E[r_j]}{\sqrt{(E[r_i{}^2] - E[r_i]^2)(E[r_j{}^2] - E[r_j]^2)}} \tag{2}$$

Of which, r_i is yield rate series of stock i, E[·] is mathematical expectation, $\rho_{ij}(\Delta T)\in[-1,\ 1]$. It constructs corresponding N×N cross-correlation coefficient matrix C, C is symmetric matrix, which is indicated as follows:

$$C = \begin{cases} c_{ij} = \rho_{ij}(\Delta t) & i \neq j \\ c_{ij} = 1 & i = j \end{cases}$$

(3)

Of which, C_{ij} is the element of cross-correlation coefficient matrix. Generally speaking, it requires cross-correlation coefficient matrix to be positive definite, which means L>N. Cross-correlation coefficient among daily average logarithm yield rate of any stock i and j is C_{ij}, $C_{ij}\in[-1,\ 1]$ when cross-correlation coefficient is bigger or equals to the designated threshold value θ, it is regarded that cross-correlation is edge weight. When cross-correlation coefficient is smaller than designated threshold value, it is regarded that it has no edge among stock yield rate, suppose edge connects node has no direction. Because cross-correlation coefficient set among stock cannot meet condition of measurement space, in order to further analyze internal nature of cross-correlation network, converts cross-correlation coefficient $\rho_{ij}(\Delta T)$ to corresponding measurement distance d_{ij}, which is as follows:

$$d_{ij} = \sqrt{2(1 - C_{ij})}$$

(4)

In the formula, the bigger of cross-correlation coefficient of stock i and stock j, the smaller of corresponding distance. In addition, measurement distance d_{ij} meets 3 conditions of measurement space, which is $d_{ij}=0$ only and when i=j, $d_{ij}=d_{ji}$, $d_{ij}\leq d_{im}+d_{mj}$.

Of which, m is any other stock different from i and j. Through formula(4), cross-correlation coefficient matrix C is converted into distance matrix D, and $D_{ij}=d_{ij}\in[0, 2]$.

(1) In the average route length network, the shortest route edge number of connecting any 2 nodes i and j is defined as distance d_{ij} between these 2 nodes. Distance average between 2 nodes is defined as average route length of network, which is marked as L, which means of which, N is network node number. Formula (5) includes distance from node to itself (this distance is 0).

$$L = \frac{1}{\frac{1}{2}N(N+1)}\sum_{i \geqslant j} d_{ij}$$

(5)

(2) One node in clustering coefficient assumption network has k_i edges connecting it with other nodes, so there are k_i nodes around nodes i, these nodes are regarded as neighbor of node i. Obviously, there are almost $k_i(k_i+1)/2$ edges between in this node, while this k_i neighborhood node including the actual edge E_i actually existed in node i as well as the total possible edge number is defined as clustering coefficient C_i of node i, which is as follows:

$$C_i = 2E_i/(k_i(k_i+1))$$

(6)

(3) Degree and degree distribution. Degree k_i of node i is defined as the number of this node connecting with other nodes. Intuitively speaking, the bigger of one node, which means this node is much more important in certain meanings. Degree of directed network is divided into out-degree and in-degree, the average of degree k_i of all the nodes in network is regarded as average degree of network node, which is marked as $<k>$.

Degree distribution condition of node in network can be described by distribution function. P(k) indicates one randomly selected node degree and its probability is just k, its distribution in average network is similar to Possion distribution. The degree distribution of actual network is obviously different from Possion distribution, it can use power law form P(k) $\propto k^{-\gamma}$ with proper power exponent(it is usually $2\leq\gamma\leq3$). Distribution curve of power exponent is decreasing slowly than that of Possion index distribution curve.

(4) Threshold value. Threshold value θ is system state control parameter of system takes place sudden changes, it is also regarded as critical value. System control parameter exceeds one threshold value, it usually has the next threshold value, which further changes control parameter, when it exceeds new threshold value, it will take place new sudden change, which makes system enter higher-level orderly state. This paper adopts threshold value to define network threshold value θ_0 through cross-correlation coefficient matrix, when $|\theta - \theta_0|$ is smaller, network is more stable, on the contrast, when $|\theta - \theta_0|$ is bigger, network is not so stable.

3. Stock network construction based on cross-correlation coefficient and optimal threshold value

Indexes related to stock include yield rate, stock price and PE ratio. Yield rate means return rate of investment, it can be calculated according to market price, coupon interest, face value and so on. The most important is it can measure yield of individual and social investment, it is concretely divided into individual yield rate and social yield rate. Stock market participant mainly pays attention to the former one. In financial market, more and more demonstration analysis result indicates that stock yield distribution has characteristics of leptokurtosis and fat-tail, it has long-time memory effect etc, as for one typical network, it uses edge in network to indicate cross-correlation among stocks. It chooses proper threshold value on the basis, which is critical value of cross-correlation coefficient, if cross-correlation coefficient of 2 stocks is bigger than this critical value then connects, from this constructs all the stocks in every market as one network system. In Ucinet6.0 software, it sets different threshold value and applies cross-correlation coefficient matrix in each stock market to get the biggest connected sub-graph node number of this market, and it uses the optimal threshold value to draw out stock network structure diagram. In market division, because division of market 1 and market 8 has clear definition, stock in it has no intersection phenomenon, so it only gives research result of market 1 and market 8. Under different threshold value, it respectively records the biggest connected sub-graph node number of corresponding market, diagram 1 show its change trend.
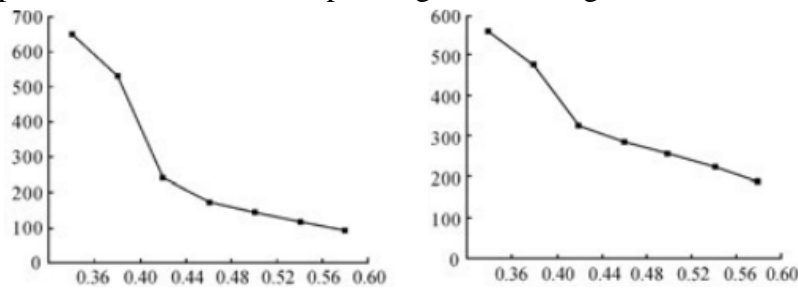

Figure 1 Curve diagram of the biggest connected sub-graph node number

From network structure of each market, it can be seen that different stocks depend on mutual cross-correlation relation to closely connected together in every market, region with relatively intensive stock indicates these stocks have bigger cross-correlation coefficient and have closer relation with each other, when region is sparse, which indicates fluctuation of this stock has little effect on other stocks. If there is one stock connects with more edges of other stocks, which indicates changes of this stock will affect much more stocks.
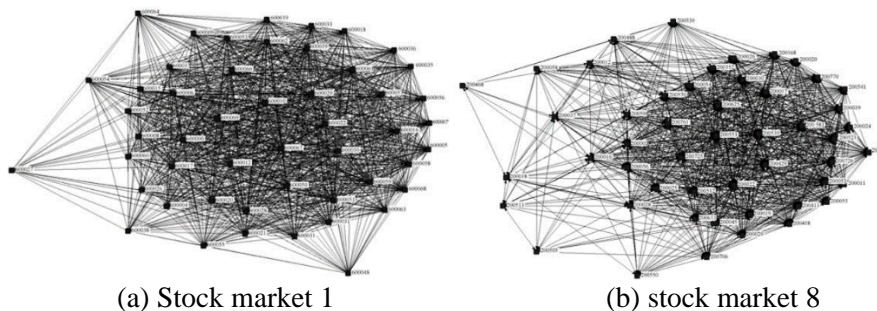

(a) Stock market 1   (b) stock market 8

Figure 2 Network structure diagram of stock market

Table 1 Table division and stock code in every market

Market category1		Market category 2		Market category 3	
Table 1	Table 2	Table 1	Table 2	Table 1	Table 2
600038	600053	900943	900933	000050	000061
600015	600037	900937	900947	000019	000016
600067	600036	900908	900907	000046	000022
…	…	…	…	…	…
600051	600052	900930	900927	000002	000033
				000049	000032

It randomly selects sample segment from all stocks in every market, applies Matlab 7 software to obtain result and consults related information of sample segment according to transaction network of relate stock, the details please see table 2. It lists related information of representative stock in every market table, from this it can be seen under precondition of different threshold value and the biggest connected sub-graph node number, the constructed stock network is mainly based on industry to divide tables of stock market, such as transportation auxiliary industry, air transportation industry, real estate development and operating industry, computer application service etc.

Table 2 Related information of each market sample

Market	Divide table	Stock code	name	place	industry
Market category 1	Table 1	600038	Hafei	Shanghai	transportation tool manufacturing industry
	Table 1	600004	Baiyun Airport	Shanghai	transportation auxiliary industry
Market category 2	Table 1	900945	B stock of Seair	Shanghai	air transportation industry
	Table 2	900911	B stock of Jinjiang	Shanghai	hotel industry
Market category 4	Table 1	000045	Shenzhen textile A	Shenzhen	clothes industry
	Table 2	000011	Shenzhen property A	Shenzhen	real estate development and operating industry
Market category 8	Table 1	200541	Guangdong lighting B	Shenzhen	mechanic, equipment and instrument
	Table 2	200025	B stock of Teli	Shenzhen	energy, material and mechanic electronic equipment wholesale industry
Market category 16	Table 1	300229	TRS	Shenzhen	computer application service industry
	Table 2	300009	Anke biotechnology	Shenzhen	medical biologics industry

4. Comparison on Division table as well as reasons for similarities and differences

From Phoenix net, NetEase Wealth Network etc we can consult the base on table division in related Chinese stock market. In reality, Chinese stock table can be divided into manufacturing, medical health, science research and technology, communication equipment etc according to industry, it can be divided into intelligent wearing, aircraft carrier shipbuilding, robot, 3D printing etc according to concept and divided into Shanxi, Sichuan, Shanghai, Hunan etc according to regions. But in reality, China usually divides stock market table according to regions, mainly considering life habit and investment convenience of stock investors. Because Chinese citizens usually like to live in

certain region and pay more attention to economic development or stock fluctuation in this region, stock market table by way of regional division is favorable for local investors spend little energy and undertake few risks to obtain higher profit.

5. Conclusions

This research confirms the optimal threshold and establishes stock network structure through related index of stock; it divides Chinese stock market into different tables according to cross-correlation index and optimal threshold of stock. From model structure it can be seen that stock network model mainly makes division on Chinese stock market from industry aspect, because cross-correlation of stock in each industry is close, cross-correlation coefficient among stocks in certain industry is usually far bigger than that between stock in this industry and stock in other industry. On participating in stock market investment, participation subject should not only pay attention to effect of other industries on stock in this industry, but also pay attention to mutual effect on stocks in the same industry.

References

[1] Yun Jian, Chang Xiqiang, Wei Wei, Niu Jaixin, Wang Chen. Research on Effect of Large-scale Photovoltaic Generation on Xinjiang Grid Relay Protection [J].Electrical Technology, 2015, 10:27-33.

[2] Lv Dan, Pan Jialong. Calculation Modeling on Grounding Transformer and Arc Eliminating Device in Photovoltaic System [J]. Renewable Energy, 2013, 09:18-20+25.

[3] Zhang Shengzhong. Analysis on Utilization Efficiency of Photovoltaic Generation in Different Access System [J].Renewable Energy, 2014, 01:29-33.

[4] Zheng Chao, Lin Junjie, Zhao Jian, Sheng Xianhui, Gao Feng. Transient Power Characteristic and Volateg Conteil in Scaled Grid-connected PV System [J]. Proceedings of the CSEE, 2015, 05:1059-1071.

[5] Zhou Lin, Ren Wei, Liao Bo, Chao Yang, Shao Nianbin, Du Xiao. Reactive Voltage Control of Grid Connected Photovoltaic Power Station [J]. Trans of China Electrotechnical Society, 2015, 20:168-175.