Chaos, Solitons and Fractals 94 (2017) 44–53

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Dynamic asset trees in the US stock market: Structure variation and market phenomena

Wei-Qiang Huang[a,*], Shuang Yao[b], Xin-Tian Zhuang[a], Ying Yuan[a]

[a] School of Business Administration, Northeastern University, Shenyang, Liaoning 110167, P. R. China
[b] School of Economics and Management, Shenyang University of Chemical Technology, Shenyang, Liaoning 110142, P. R. China

ARTICLE INFO

Article history:
Received 29 March 2016
Revised 4 October 2016
Accepted 19 November 2016
Available online 28 November 2016

Keywords:
Stock market
Minimal spanning tree
Structure variation
Market phenomena

ABSTRACT

In this work, employing a moving window to scan through every stock price time series over a period from 2 January 1986 to 20 October 2015, we use cross-correlations to measure the interdependence between stock prices, and we construct a corresponding minimal spanning tree for 170 U.S. stocks in every given window. We show how the asset tree evolves over time and describe the dynamics of its normalized length, centrality measures, vertex degree and vertex strength distributions, and single- and multiple-step edge survival ratios. We find that the normalized tree length shows a tendency to decrease over the 30 years. The power-law of vertex degree or vertex strength distribution does not hold for all trees. The survival ratio analysis reveals an increased stability of the dependence structure of the stock market as time elapses. We then examine the relationship between tree structure variation and market phenomena, such as average, volatility and tail risk of stock (market) return. Our main observation is that the normalized tree length has a positive relationship with the level of stock market average return, and it responds negatively to the market return volatility and tail risk. Furthermore, the majority of stocks have their vertex degrees significantly positively correlated to their average return, and significantly negatively correlated to their return volatility and tail risk.

1. Introduction

A quantitative description of the hierarchical structure is crucial for understanding the dynamics of complex systems [1]. In essence, the stock market is an example of a complex system consisting of many interacting components [2,3]. The correlation matrix of stock return time series, which plays a central role in investment theory and risk management, can be used to extract information about hierarchical organization of stock market. By using the correlation between pairs of elements as a similarity measure, some hierarchical clustering procedures have been proposed to select the statistically reliable information of the correlation matrix [1].

The hierarchical tree obtained by applying single linkage cluster analysis (SLCA) and average linkage cluster analysis (ALCA) to the correlation matrix, can well identify groups of stocks belonging to the same economic sector [4,5,6]. In addition to the hierarchical trees, one can also associate correlation based networks with the correlation matrix using clustering algorithm. In the correlation based networks, a subset of links which are highly informative about the hierarchical structure of the system are selected. For example, the minimum spanning tree (MST), which was firstly introduced in Mantegna (1999), is a correlation based tree associated with the SLCA. A lot of subsequent studies constructed MSTs to investigate the economic properties of stock returns [7–15]. Other examples of correlation based networks are the planar maximally filtered graph (PMFG) [16], and the average linkage minimum spanning tree (ALMST) [17]. The PMFG presents a graph structure which is richer than the one of the MST, and has been used to investigate stock return time series in Refs. [6,9,16,18,19]. To evaluate the statistical reliability of nodes in a hierarchical tree and links in a correlation based network, a bootstrap procedure of the time series has been devised in Refs. [17,20]. In order to quantify and compare the performance of different filtering procedures, a useful measure using the Kullback–Leibler distance has been proposed [21]. It was shown that the Kullback–Leibler distance is very good for comparing correlation matrices [21]. In addition to the modeling of correlation matrix for stock return time series, a lot of researches have constructed similar correlation based networks for industry indices [22], stock market indices [23,24,25], world currencies [26], and government bond market indices [27].

The empirical analysis of correlation based networks would be static or dynamic. Refs. [4,7,8,18] focused on the static network, i.e.

* Corresponding author.
E-mail address: hwqneu@hotmail.com (W.-Q. Huang).

http://dx.doi.org/10.1016/j.chaos.2016.11.007

investigating the properties of the network constructed for a long time period, such as network topology, hierarchical structure, or taxonomic studies in terms of economic sector, region, or other characteristics. However, because each stock responds differently to external information like the same economic announcements or market news, the correlation among them will vary. More and more studies constructed dynamic stock correlation networks, and investigated structural changes [9,11,12,15,28], topological stability [9,10,14,30], structural differences between crisis and non-crisis periods [11,12,14,29], and relationship between volatility and network properties [13]. For example, Micciche et al. [10] investigated the time series of the degree of minimum spanning trees obtained by using a correlation based clustering procedure which starts from asset return. The analysis showed that the degree of stocks has a very slow dynamics with a time scale of several years. Sienkiewicz et al. [12] provided empirical evidence that there is a dynamic structural and topological first order phase transition in the time range dominated by a crash. By investigating the ability of the network to resist structural or topological change, Yan et al. [19] found that the PMFG before the US sub-prime crisis has a stronger robustness against the intentional topological damage than the other two non-crisis periods. Kocheturov et al. [29] studied cluster structures of market networks constructed from correlation matrix of returns of the stocks traded in the USA and Sweden. Their main observation was that in non-crisis periods of time cluster structures change more chaotically, while during crises they show more stable behavior and fewer changes.

These existing studies mainly relate stock correlation network variations to the extreme events such as global financial crises. However, the general relationships between network structure variation and market phenomena can help us understand the interaction between network and stock market dynamics, thus it can be a good guide for risk management of stock investment. In this paper, we investigate the dynamics of correlations present between pairs of U.S. stocks traded in U.S. market by studying correlation based networks. We also investigate the general relationship between network structure variation and market dynamics. The study is performed by using stock time series during the time period from January 1986 to October 2015, which spans near 30 years. We begin with the construction of network based on raw data of stock in Section 2. In Section 3 we describe the network topology structures and market phenomena. Section 4 is empirical study and results. In the last section we present a few conclusions.

2. Network construction

A network is usually defined as a collection of vertices connected by edges. If we consider a stock correlation network, each stock will be a network vertex. Each pair of stocks is connected with an edge, with its edge weight equal to the Pearson's correlation of their corresponding stock returns in a certain time period. Furthermore, we can characterize the dynamics of stock correlation network by calculating the cross-correlation between two stocks for the moving time periods. The sample data we collected are the daily returns of N stocks traded in the U.S. stock market. The sample time period length is T days. The network evolution is analyzed by setting a time window of length w days and moving this window along time. One network is obtained by considering the time series inside each window. This window is displaced by an amount of τ days and a new network is obtained after each displacement. This process is repeated until the end of the original time series is reached. For example, the first network will be constructed from the time series starting at day $t_1^1 = 1$ and ending at day $t_2^1 = w$, the second network will be constructed from the time series starting at day $t_1^2 = 1 + \tau$ and ending at day $t_2^2 = \tau + w$, the third network will be constructed from the time series starting at day $t_1^3 = 1 + 2\tau$ and ending at day $t_2^3 = 2\tau + w$, and so on. Hence, we achieved a total of $M(M = 1 + \lceil (T - w)/\tau \rceil$, $\lceil \bullet \rceil$ denotes the ceiling function) networks. Let $R_i^m(t)$ be the log return of stock i at day t in the mth window, where $m = 1, 2, \cdots, M$.

$$R_i^m(t) = \ln P_i^m(t) - \ln P_i^m(t-1) \quad (1)$$

where $P_i^m(t)$ is the closing price of stock i at day t. The correlation between stock i and $j (i = 1, 2, \cdots, N,\ j = 1, 2, \cdots, N)$ in the mth window can be measured by Pearson's correlation of series R_i^m and R_j^m.

$$\rho_{i,j}^m = \frac{\langle R_i^m R_j^m \rangle - \langle R_i^m \rangle \langle R_j^m \rangle}{\sqrt{(\langle (R_i^m)^2 \rangle - \langle R_i^m \rangle^2)(\langle (R_j^m)^2 \rangle - \langle R_j^m \rangle^2)}} \quad (2)$$

where $\langle \cdots \rangle$ denotes the expected value, and $\rho_{i,j}^m \in [-1, 1]$. Specifically, if $i = j$ then $\rho_{i,j}^m = 1$. Thus we construct the mth stock correlation network $G^m(V, E^m)$, where $V = \{1, 2, \cdots, N\}$ denotes the node set, and the network edge set E^m can be denoted by $\{e_{ij}^m = \rho_{i,j}^m | i = 1, 2, \cdots, N, j = 1, 2, \cdots, N\}$. So e_{ij}^m reflects the edge weight between node i and j in the network, and $G^m(V, E^m)$ is an undirected and weighted network.

The correlation coefficient of a pair of stocks cannot be used as a distance between the two stocks because it does not fulfill the three axioms that define a metric. However a metric can be defined using as distance a function of the correlation coefficient. The correlation coefficient $\rho_{i,j}^m$ is transformed to a distance metric $d_{i,j}^m$ [4].

$$d_{i,j}^m = \sqrt{2(1 - \rho_{i,j}^m)} \quad (3)$$

The $d_{i,j}^m$ fulfills the three axioms of a metric distance: 1) $d_{i,j}^m = 0$ if and only if $i = j$; 2) $d_{i,j}^m = d_{j,i}^m$ and 3) $d_{i,j}^m \le d_{i,k}^m + d_{k,j}^m$ [4]. Now the edge weight e_{ij}^m can be measured by $d_{i,j}^m$, and the corresponding edge set E^m can be denoted by $\{e_{ij}^m = d_{i,j}^m | i = 1, 2, \cdots, N, j = 1, 2, \cdots, N\}$. We have $e_{ij}^m = d_{i,j}^m \in [0, 2]$. The full connected network $G^m(V, E^m)$ is then used to determine the minimal spanning tree MST^m, which is a simply connected graph that links the N vertices with the $N - 1$ edges such that the sum of all edge weights is minimum. The minimal spanning tree can provide an easy way to extract the most important correlations and information in the stock market while retaining the simplest structure and enabling the ability to visualize the relationships across stocks. A general approach to the construction of the MST^m is as follows [9,31].

Step 1: Start with an empty graph. Make an ordered list of edges in $G^m(V, E^m)$, ranking them by increasing edge weight $d_{i,j}^m$.
Step 2: Take the first element in the list and add the edge to the graph.
Step 3: Take the next element and add the edge if the resulting graph is still a tree; otherwise discard it.
Step 4: Iterate the process from Step 3 until all pairs have been exhausted.

During the whole T days period, the network construction procedure is repeated M times, and hence we have M consecutive networks.