Physica A 402 (2014) 245–254

Contents lists available at ScienceDirect

Physica A

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

Cointegration-based financial networks study in Chinese stock market

Chengyi Tu

School of Automation Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, Sichuan, PR China

HIGHLIGHTS

- Constructing financial complex network based on cointegration instead of correlation.
- Pruning to obtain Cointegration Threshold Network and Cointegration Planar Graph.
- Analyzing these non-symmetric networks by using standard methods of network analysis.
- Cointegration Planar Graph is better to represent some properties than PCPG.
- The properties of Cointegration Threshold Network approximate to complete graph.

ARTICLE INFO

Article history:
Received 28 June 2013
Received in revised form 20 October 2013
Available online 14 February 2014

Keywords:
Financial complex network
Cointegration
PMFG

ABSTRACT

We propose a method based on cointegration instead of correlation to construct financial complex network in Chinese stock market. The network is obtained starting from the matrix of *p*-value calculated by Engle–Granger cointegration test between all pairs of stocks. Then some tools for filtering information in complex network are implemented to prune the complete graph described by the above matrix, such as setting a level of statistical significance as a threshold and Planar Maximally Filtered Graph. We also calculate Partial Correlation Planar Graph of these stocks to compare the above networks. Last, we analyze these directed, weighted and non-symmetric networks by using standard methods of network analysis, including degree centrality, PageRank, HITS, local clustering coefficient, K-shell and strongly and weakly connected components. The results shed a new light on the underlying mechanisms and driving forces in a financial market and deepen our understanding of financial complex network.

1. Introduction

The financial market is a highly complex evolving system [1,2], still controversial [3], difficult to grasp and predict [1], and sensitive to economic instabilities, such as the Black Monday of October 19, 1987 [4], financial crisis of 2007–2008 caused by the subprime mortgage loan crisis [5]. So the study of it has rapidly been attracting interest within not only economics, but also mathematics and physics [2,6–9]. Recently complex network (directed or undirected, weighted or unweighted) is found useful to characterize and better understand the financial market [1,10–12], in which each stock indicates a vertex and each interaction by a pair of stocks indicates a weighted edge that encodes the distance between them. But how to quantify a distance between different stocks traded in the financial markets and then construct a network to exact useful information is still a difficult and an open problem.

E-mail address: chengyitu1986@gmail.com.

http://dx.doi.org/10.1016/j.physa.2014.01.071
0378-4371/

Mantegna [13] was possibly the first to study financial markets in the concept of networks, who encodes the distance between two stocks as a function of Pearson correlation coefficient to construct the complete graph, and then implements filtering technique Minimum Spanning Tree (MST) to simplify and extract useful information from them, such as detecting a taxonomy that organizes stocks according to their economic activity. After this seminal work, many relevant papers emerge. Bonanno et al. employed the same methodology to find a meaningful taxonomy existing in stock market indices located all over the world [14], and high-frequency cross-correlation existing between pairs of stocks in US equity markets [15]. By observing the networks obtained from real and considered artificial markets, topology of the MST of them is different was found [16]. During a stock market crisis such as Black Monday, the tree length shrinks and strong reconfiguration take place [4]. Onnela et al. investigated another network, generally called asset graph, obtained by selecting only the $N-1$ edges with lower correlation coefficient weights of the complete graph [17]. The construction of the asset graph was analyzed in Ref. [18], where the temporal evolution of the clustering and information was measured. Constructing networks from financial data also use threshold of correlations of stocks [19–22]. A complete analysis was investigated by Ref. [19], where repeatedly remove correlation links from lower to higher values or vice versa, and found that the low correlation links of the network contribute to the overall connectivity significantly more than the high correlation links.

Encoding the distance between two stocks as a function of the correlation coefficient to construct the complete network has generally been used in financial theory and practice to quantify the degree of similarity among the stocks. However, correlation time series are just demanded to move in the same direction while the magnitude of the move is unknown, so academics have long since questioned that apply correlation to represent relationships between time series. On the other hand, since these networks are usually complete graphs (all links between vertices are present), how to extract useful information and understand the behavior of them are difficult. It has not been fully investigated and only a few methods have been developed until now, such as Minimum Spanning Tree (MST) [13], also called Asset Tree, Asset Graph [17], and Planar Maximally Filtered Graph (PMFG) [23].

In this paper we firstly apply Engle–Granger cointegration test between all pairs of stocks instead of traditional correlation to construct a non-symmetric matrix of cointegration p-values. Then some tools for filtering information in complex network are implemented to prune the complete network described by the above matrix, such as setting a level of statistical significance as a threshold and Planar Maximally Filtered Graph (PMFG). To compare the above networks with network obtained by using other method, we also calculate Partial Correlation Planar Graph (PCPG) of these stocks. Last, we analyze these directed, weighted and non-symmetric networks by using standard methods of network analysis, including degree centrality, PageRank, HITS, local clustering coefficient, K-shell and strongly and weakly connected components.

This paper is organized as follows: Section 2 explains the data and methods used to empirically investigate. Next, Section 3 presents the results observed in accordance with the empirical design. Section 4 gives the advantage based on cointegration instead of correlation to construct financial complex network, the drawback of Engle–Granger cointegration test and the possibly next step research. Finally, the paper will be concluded in Section 5.

2. Method

2.1. Data

We selected 197 stocks from the Constituents List of CSI 300 index on January 04, 2013 [24], compiled the data with consecutive daily price data from April 8, 2005 (the launch day of CSI 300 index) to April 30, 2013. The CSI 300 index, compiled by the China Securities Index Company, Ltd. Refs. [25,26], is the first equity and capitalization-weighted stock market index designed to measure the performance of all the A shares traded in Chinese stock markets. The selected period includes the financial crisis of 2007 and 2008 caused by the subprime mortgage loan crisis what is considered the most severe and influential financial crisis of the second half of the 20th century. To show the trend of the stock exchange, the figure of CSI 300 index for the period is shown in Fig. 1. You could see the Chinese stock market peaked in October 2007, when the CSI 300 index approached 5688 points. It then entered a continuous and pronounced decline and by October 2008, the index had reached a trough of around 1663 points. Afterwards, the index rise to about 4000, then fluctuate in the interval between 2000 and 3000 until now.

According to the results of the previous study, the stocks often fluctuate in time and transactions occur at random times with random intensities, so great care must be taken in the selection of the most appropriate time scale to be studied [2,27]. In this paper, we study the successive differences of the natural logarithm of price and show as follows, because the merit of this approach is that the average correction of scale changes is incorporated without requiring deflators or discounting factors [2].

$$r_i(t) = \ln P_i\,(t + \Delta t) - \ln P_i(t), \tag{1}$$

where $P_i(t)$ is the daily closing price of stock i at day t, Δt is 1 in our paper.

In order to classify the stocks into different industry sectors, we apply the CICS Industry Classification [28,29] designed by CHINA SECURITIES INDEX CO., LTD. The sectors are listed as follows: 0 Energy (EN, 15 stocks), 1 Materials (MA, 41 stocks), 2 Industrials (IN, 30 stocks), 3 Consumer Discretionary (CD, 26 stocks), 4 Consumer Staples (CS, 19 stocks), 5 Health Care (HC, 17 stocks), 6 Financials (FI, 28 stocks), 7 Information Technology (IT, 7 stocks), 8 Telecommunication Services (TS, 4 stocks) and 9 Utilities (UT, 10 stocks).