ICIC Express Letters
Part B: Applications
Volume **8**, Number **2**, February **2017**

ICIC International ©2017 ISSN 2185-2766
pp. 351–356

CLUSTER FEATURES AND SYSTEMIC RISK IN CHINESE STOCK MARKET

Sumuya Borjigin

School of Economics and Management
Inner Mongolia University
No. 235, West College Road, Hohhot 010021, P. R. China
sumuya@mail.dlut.edu.cn

Received July 2016; accepted October 2016

Abstract. *In this paper, we study relation between cluster features and systemic risk in Chinese stock market. In the first step, the GARCH model was used to depict volatility of Chinese stock market. In the second step, the multi-way normalized cut spectral clustering method was used to characterize cluster features of Chinese stock market. In the third step, the Granger causality test model was used to judge the relation between cluster features and systemic risk in Chinese stock market. Empirical analysis shows that, in most of the periods, there is no direct relation between cluster features and systemic risk in Chinese stock market.*

Keywords: Cluster feature, Systemic risk, Chinese stock market

1. **Introduction.** Since there is currently no widely accepted definition of systemic risk, a comprehensive literature review of this rapidly evolving research area is difficult to provide. Like Justice Potter Stewart's description of pornography, systemic risk seems to be hard to define, but we think we know it when we see it. Such an intuitive definition is hardly amenable to measurement and analysis, a prerequisite for macro-prudential regulation of systemic risk [9]. A more formal definition is any set of circumstances that threatens the stability of public confidence in the financial system [11]. Systemic events are multi-factorial, so it is hard to measure "stability" and "public confidence" by any single metric. Instead, we focus on the four "L"s of systemic risk, which are leverage, liquidity, losses and linkages [9]. Several measures of the first three already exist. However, the one common thread running through all truly systemic events is the connections and interactions among financial stakeholders. Therefore, any measure of systemic risk must capture the degree of connectivity of market participants to some extent [9]. In this paper we focus our attention on relation between linkages and the systemic risk.

Clustering algorithms partition data into a certain number of clusters, patterns in the same cluster should be similar to each other, while patterns in different clusters should not [12]. Clustering can be used to capture the linkages and interactions among financial stakeholders, and the similarity or the dissimilarity among financial stakeholders can be seen as the connections and interactions among them. In recent years, spectral clustering has become one of the most popular modern clustering algorithms. It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm [16]. Studies show that, performance of the normalized cut criterion is better than the minimum cut criterion, minimum ratio criterion and min-max cut criterion [8]. Meanwhile, compared to recursive normalized cut spectral clustering algorithm, multi-way normalized cut spectral clustering algorithm is easier to implement.

As we all know, in recent one year, Chinese stock market presents obvious features of systemic risk. However, most of the existing papers did not pay their attention to the systemic risk of the Chinese stock market. Aldasoro and Angeloni [7] showed how elements from classic input-output analysis can be applied to banking and how to derive six indicators that capture different aspects of systemic importance, using a simple numerical example for illustration. Kang and Suh [6] examined whether emerging market financial turmoil in 2013-2014, caused mainly by the expectation of future US monetary policy tightening, and created such spillover. Kim et al. [3] investigated whether the characteristic fund performance indicators are correlated with the asset price movement using information flows estimated by the Granger causality test. Reboredo and Ugolini [5] studied systemic risk in European sovereign debt markets before and after the onset of the Greek debt crisis, taking the conditional value-at-risk (CoVaR) as a systemic risk measure, characterized and computed using copulas. Gang and Qian [4] studied the effect of domestic monetary policies on China's systemic risk after the collapse of Lehman Brothers. Vyrost et al. [14] used a sample of daily closing prices from 20 stock markets from developed countries. Granger causality networks were constructed for 94 partially overlapping sub-samples of a length of 3 months, starting from January 2006 to December 2013. In view of the mentioned above, relation between the linkage and the systemic risk by multi-way normalized cut spectral clustering method was studied. In the first step, the GARCH model was used to characterize daily volatility of shares and stock market. In the second step, the multi-way normalized cut spectral clustering method was used to depict linkage features of Chinese stock market. In the third step, the Granger causality test model was used to calculate the relation between linkage among the shares and the systemic risk.

The paper is organized as follows. In Section 2, we introduce some of the models and algorithms which will be used in the following discussions. In Section 3, we study relation between the linkage and the systemic risk by the models and algorithms. Section 4 and Section 5 provide final results.

2. Methodology.

In this section, we first introduce the GARCH model and the multi-way normalized cut spectral clustering algorithm. Then, introduce the Granger causality test model.

2.1. GARCH model.

In econometrics, autoregressive conditional heteroscedasticity models are used to characterize and model time series. They are used at any point in a series where the error terms are thought to have a characteristic size or variance. In particular, ARCH models assume the variance of the current error term or innovation to be a function of the actual sizes of the previous time periods' error terms: often the variance is related to the squares of the previous innovations. If an autoregressive moving average model is assumed for the error variance, the model is a generalized autoregressive conditional heteroscedasticity model [15].

Ordinarily, we called it GARCH for short. The GARCH(p, q) model (where p is the order of the GARCH terms σ^2 and q is the order of the ARCH terms ε^2) is given by

$$\begin{aligned} & y_t = x_t' b + \varepsilon_t \\ & \varepsilon_t \,|\psi_t \sim \mathrm{N}(0, \sigma_t^2) \\ & \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 \end{aligned}$$

2.2. Multi-way normalized cut spectral clustering method.

For a given data set $P = \{p_1, p_2, \ldots, p_n\}$ in spectral clustering, if every individual datum is regarded as a node, and the similarity between two data is defined as a weight on the edge between the nodes, then we can consider the data set $P = \{p_1, p_2, \ldots, p_n\}$ as an undirected weighted graph G. Therefore, we assume that an undirected weighted graph $G = (V, E, W)$ is given, and it can be partitioned into k disjoint sets, i.e., $V = V_1 \cup V_2 \cup \cdots \cup V_k$, $V_i \cap V_j = \Phi, 1 \le i < j \le k$ [16].

The degree of dissimilarity between two sets $\{V_i, V_j\}$ can be computed as the sum of the weights of the edges that connect them:

$$cut(V_i, V_j) = \sum_{u\in V_i, v\in V_j} \omega(u,v), \quad i,j = 1,2,\dots,k$$

Another kind of measurement for the similarity of those sets is *assoc*, which is defined as:

$$assoc(V_i, V) = \sum_{u\in V_i, v\in V} \omega(u,v), \quad i = 1,2,\dots,k$$

It is the total connection from nodes in V_i to all nodes in the graph.

The multi-way normalized cut criterion proposed by Meila and Shi [10] can be written as:

$$MNcut(k) = \frac{cut(V_1, V\backslash V_1)}{assoc(V_1, V)} + \frac{cut(V_2, V\backslash V_2)}{assoc(V_2, V)} + \cdots + \frac{cut(V_k, V\backslash V_k)}{assoc(V_k, V)}$$

Similarly, we can define $MNassoc(k)$.

For a given data set, minimizing multi-way normalized cut exactly is NP-complete. Multi-way normalized cut spectral clustering algorithm is a way to solve relaxed version of this problem. Meila and Shi found that relaxing multi-way normalized cut leads to multi-way normalized cut spectral clustering and converts the problem into the eigenproblems of block stochastic matrix $D^{-1}S$ or normalized Laplacian matrix $D^{-1}(D-S)$ [10]. Based on the above analysis, Meila and Shi proposed the multi-way normalized cut spectral clustering algorithm [10]:

Input: Data set $P = \{p_1, p_2, \dots, p_n\}$, cluster number k.

Step 1: Compute the distance matrix W, and construct similarity matrix S according to W;

Step 2: Compute the Laplacian matrix L;

Step 3: Compute the first k eigenvectors $\{v_1, v_2, \dots, v_k\}$ of the generalized eigenproblem $Sv = \lambda Dv$;

Step 4: Let $\bar{V} \in R^{n\times k}$ be a matrix composed of the vectors $\{v_1, v_2, \dots, v_k\}$ as columns;

Step 5: For $i = 1, 2, \dots, n$, let $y_i \in R^{1\times k}$ be the vector corresponding to the ith row of $\bar{V}$;

Step 6: Cluster the points $\left\{y_i \in R^{1\times k} \,|i = 1, 2, \dots, n\right\}$ with the k-means algorithm into clusters $C_1, C_2, \dots, C_k$, if $y_i \in C_j$ then $p_i \in P_j$, $1 \le i \le n$, $1 \le j \le k$;

Output: k clusters $\{P_1, P_2, \dots, P_k\}$.

For a data set which can be partitioned into k clusters, comparison of indices *MN-assoc*(k), $MNcut(k)$ and $\sum_{i=1}^{k} \lambda_k$ among different values of k is meaningless [1]. This also means these indices are not suitable for evaluating clustering with various k; furthermore, they are not appropriate for estimating cluster number directly. However, from the above discussions, the index $Ratio(k) = \dfrac{k - MNcut(k)}{k - \sum_{i=1}^{k} \lambda_i} \in [0,1]$ is suitable for evaluating clustering with various k [13]. Larger $Ratio(k)$ stands for better partition results.

For a given data set, another important cluster index used in the following discussion was the best cluster number.

2.3. **Granger causality test model.** Time series j is said to "Granger cause" time series i if past values of j contain information that helps predict i above and beyond the information contained in past values of i alone. The mathematical formulation of this test is based on linear regressions of R^i_{t+1} on R^i_t and R^j_t.

Specifically, let R_t^i and R_t^j be two stationary time series, and for simplicity assume they have zero mean. We can represent their linear inter relationships with the following models:

$$R_{t+1}^i = a^i R_t^i + b^{ij} R_t^j + e_{t+1}^i$$
$$R_{t+1}^j = a^j R_t^j + b^{ji} R_t^i + e_{t+1}^j$$

where e_{t+1}^i and e_{t+1}^j are two uncorrelated white noise processes, and a^i, a^j, b^{ij} and b^{ji} are coefficients of the model. Then, j Granger causes i when b^{ij} is different from zero. Similarly, i Granger causes j when b^{ji} is different from zero. When both of these statements are true, there is a feedback relationship between the time series [2].

3. **Data and Descriptive Statistics.** In recent one year, Chinese stock market presents obvious features of systemic risk. So our analysis focuses on relation between the cluster feature of some of the selected stocks and the systemic risk. Because of limited space, we selected 39 kinds of stocks in Shanghai Stock Exchange and then calculated volatility via GARCH model. Volatility of Shanghai Composite Index was chosen as the systemic risk of Shanghai Stock Exchange. For the stocks and index, daily return from 2015/01/05 to 2016/03/31 was chosen, and this period encompasses both tranquil boom and crisis period. Tickers of the 39 stocks are tabulated in Table 1. The first row of Figure 1 is volatility of Shanghai composite index.

TABLE 1. Stock ticker

stock ticker	stock ticker	stock ticker	stock ticker	stock ticker
600004	600005	600007	600008	600009
600010	600011	600012	600015	600016
600017	600018	600019	600020	600021
600022	600023	600027	600028	600029
600030	600031	600033	600036	600037
600038	600039	600048	600050	600051
600054	600055	600056	600057	600059
600060	600062	600063	600064	

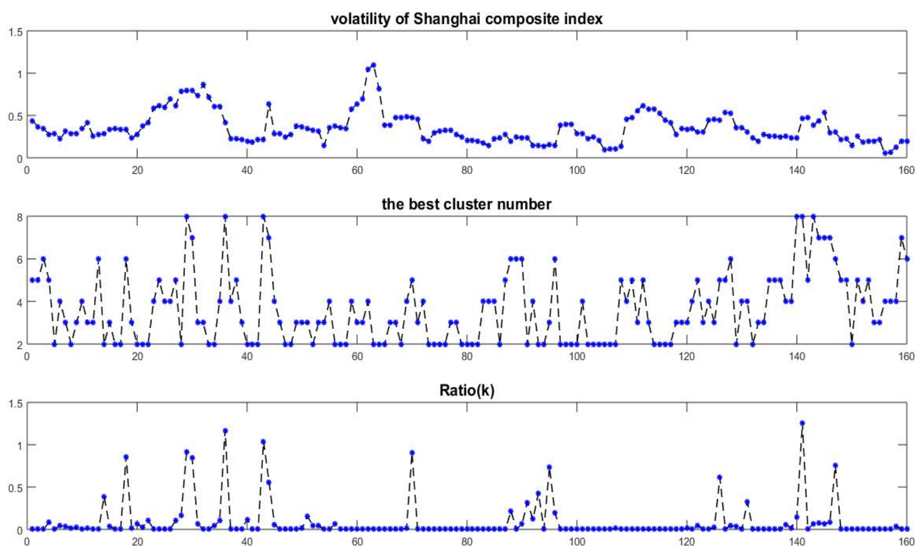

FIGURE 1. Volatility of Shanghai Composite Index and cluster features of the stocks

The multi-way normalized cut spectral clustering method was used to depict cluster features of the 39 stocks. The 20-day rolling-window sub-period cluster features were shown in the second row and the third row of Figure 1, the second row represents values of the best cluster number and the third row represents values of $Ratio(k)$.

Granger causality test was used to depict the relation between the cluster features and the systemic risk. According to the volatility of Shanghai Composite Index, the period 2015/01/05∼2016/03/31 was partitioned into 3 sub-periods: from 2015/01/05 to 2015/06/14, from 2015/06/15 to 2015/10/31 and from 2015/11/01 to 2016/03/31. We first considered relation between the cluster features and the systemic risk in the period from 2015/01/05 to 2016/03/31, then in the 3 sub-periods. The results are tabulated in Table 2-Table 5.

Table 2. Granger causality test: 2015/01/05∼2016/03/31

Null Hypothesis:	F-Statistic	Prob.
ClusterNumber does not Granger Cause SSEComposite	1.4434	0.2393
SSEComposite does not Granger Cause ClusterNumber	0.1491	0.8616
$Ratio(k)$ does not Granger Cause SSEComposite	0.1436	0.8664
SSEComposite does not Granger Cause $Ratio(k)$	4.8216	**0.0093**

Table 3. Granger causality test: 2015/01/05∼2015/06/14

Null Hypothesis:	F-Statistic	Prob.
ClusterNumber does not Granger Cause SSEComposite	1.6201	0.2069
SSEComposite does not Granger Cause ClusterNumber	0.4231	0.6571
$Ratio(k)$ does not Granger Cause SSEComposite	0.2479	0.7812
SSEComposite does not Granger Cause $Ratio(k)$	1.1483	0.3244

Table 4. Granger causality test: 2015/06/15∼2015/10/31

Null Hypothesis:	F-Statistic	Prob.
ClusterNumber does not Granger Cause SSEComposite	3.2171	**0.0509**
SSEComposite does not Granger Cause ClusterNumber	0.3430	0.7118
$Ratio(k)$ does not Granger Cause SSEComposite	0.4282	0.6547
SSEComposite does not Granger Cause $Ratio(k)$	1.1308	0.3331

Table 5. Granger causality test: 2015/11/01∼2016/03/31

Null Hypothesis:	F-Statistic	Prob.
ClusterNumber does not Granger Cause SSEComposite	0.4138	0.6637
SSEComposite does not Granger Cause ClusterNumber	1.2922	0.2851
$Ratio(k)$ does not Granger Cause SSEComposite	2.5877	**0.0869**
SSEComposite does not Granger Cause $Ratio(k)$	4.0743	**0.0240**

4. **Discussions.** The empirical results of this study show that, in most of the periods, there is no direct relation between the cluster features and systemic risk. In period when volatility of Shanghai Composite Index is small, $Ratio(k)$ and SSEComposite Granger cause each other. In the period from 2015/01/05 to 2016/03/31, ClusterNumber does not Granger Cause SSEComposite, SSEComposite does not Granger Cause ClusterNumber, $Ratio(k)$ does not Granger Cause SSEComposite, and SSEComposite Granger causes

$Ratio(k)$. In the period from 2015/01/05 to 2015/06/14, none of them Granger causes the others. In the period from 2015/06/15 to 2015/10/31, SSEComposite does not Granger Cause ClusterNumber, $Ratio(k)$ does not Granger Cause SSEComposite, SSEComposite does not Granger Cause $Ratio(k)$, and ClusterNumber Granger causes SSEComposite. In the period from 2015/11/01 to 2016/03/31, ClusterNumber does not Granger Cause SSEComposite, SSEComposite does not Granger Cause ClusterNumber, $Ratio(k)$ Granger causes SSEComposite, and SSEComposite Granger causes $Ratio(k)$.

5. **Conclusions.** In the present paper, we study relation between the cluster feature of 39 stocks and the systemic risk. Firstly, the GARCH model was used to depict volatility of the 39 stocks and Shanghai Composite Index. Secondly, the multi-way normalized cut spectral clustering model was used to characterize cluster feature of the stocks. Thirdly, the Granger causality test model was used to judge relation between the cluster features and systemic risk. Numerical experiment shows that there is no direct relation between cluster features of the stocks and the systemic risk in Chinese stock market in most of the periods.

Acknowledgments. This work was partly supported by the National Natural Science Foundation of China (Grant No. 61463039), China Postdoctoral Science Foundation (Grant No. 2015M581192) and the Natural Science Foundation of Inner Mongolia (Grant No. 2014BS0706).

REFERENCES

[1] A. Nagai, Inappropriateness of the criterion of multi-way normalized cuts for deciding the number of clusters, *Pattern Recognition Letters*, vol.28, pp.1981-1986, 2007.
[2] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, vol.37, no.3, pp.424-438, 1969.
[3] H. Kim, O. Kwon and G. Oh, A causality between fund performance and stock market, *Physica A*, vol.443, pp.439-450, 2016.
[4] J. Gang and Z. Qian, China's monetary policy and systemic risk, *Emerging Markets Finance and Trade*, vol.51, no.4, pp.701-713, 2015.
[5] J. C. Reboredo and A. Ugolini, Systemic risk in European sovereign debt markets: A CoVaR-copula approach, *Journal of International Money and Finance*, vol.51, pp.214-244, 2015.
[6] H. Kang and H. Suh, Reverse spillover: Evidence during emerging market financial turmoil in 2013-2014, *Journal of International Financial Markets, Institutions & Money*, vol.38, pp.97-115, 2015.
[7] I. Aldasoro and I. Angeloni, Input-output-based measures of systemic importance, *Quantitative Finance*, vol.15, pp.589-606, 2014.
[8] J. Shi and J. Malik, Normalized cut and image segmentation, *IEEE Trans. Pattern Analysis and Machine Learning*, vol.22, pp.888-905, 2000.
[9] M. Billio, M. Getmansky, A. Lo and L. Pelizzon, Econometric measure of connectedness and systemic risk in the finance and insurance company, *Journal of Financial Economics*, vol.104, pp.535-559, 2012.
[10] M. Meila and J. Shi, Learning segmentation by random walks, *Advances in Neural Information Processing Systems*, pp.470-477, 2000.
[11] O. De Bandt and P. Hartmann, Systemic risk: A survey, *Frankfurt: European Central Bank Working Paper35*, pp.1-77, 2000.
[12] R. Xu and D. Wunsch, *Clustering*, John Wiley & Sons, New Jersey, 2009.
[13] S. Borjigin and C. Guo, Non-unique cluster numbers determination methods based on stability in spectral clustering, *Knowledge and Information Systems*, vol.36, no.2, pp.439-458, 2013.
[14] T. Vyrost, S. Lyocsa and E. Baumohl, Granger causality stock market networks: Temporal proximity and preferential attachment, *Physica A*, vol.427, pp.262-276, 2015.
[15] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, vol.31, no.3, pp.307-327, 1986.
[16] U. Luxburg, A tutorial on spectral clustering, *Statistics and Computing*, vol.17, pp.395-416, 2007.