Economic Modelling 81 (2019) 181–204

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Time-varying comovement and changes of comovement structure in the Chinese stock market: A causal network method

Hui Bu[a,1], Wenjin Tang[b,1], Junjie Wu[b,c,d,*,1]

[a] Department of Finance, School of Economics and Management, Beihang University (BUAA), Beijing, 100191, China
[b] Department of Information Systems, School of Economics and Management, Behang University (BUAA), Beijing, 100191, China
[c] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University (BUAA), Beijing, 100191, China
[d] Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beihang University (BUAA), Beijing, 100191, China

ARTICLE INFO

JEL classification:
G10
G12
G19

Keywords:
Comovement measure
Comovement structure
Topology of causal network
Common factors
Panel data model

ABSTRACT

The driving force for the comovement in stock returns is a long-standing debate between classical asset pricing theory and behavioral finance theory. It has become critically important recently for understanding systemic risk and risk contagion in the market. In this study, we propose complex networks enabled new methods to measure the causal comovement of individual stocks and the comovement structure of the market, which facilitate the examination of all kinds of hypotheses of comovement theories in a unified framework. Using a sample of the Chinese stock market from Jan. 1, 2006 to Dec. 31, 2016, we find that the degree of comovement generally intensifies over time, with a drastic increase from 2011 to 2015, while the comovement structure of the market changes with different market situations. Most importantly, our study reveals the driving force of causal comovement among individual stocks; that is, sentiment-based factors related to the market index indeed induce excess causal comovement in returns beyond that can be justified by fundamental factors including beta coefficient, book-to-market ratio, liquidity, profitability and volatility. Our study also reveals the determinants of comovement structure, which are attributable to the change of investors' behaviors in different periods. It turns out that investors in the Chinese stock market care about risk-return relationship in normal periods, while they seem to care only about risk in crisis periods.

1. Introduction

The comovement and spillover effect among markets, especially during the financial crises (1997 East Asian crisis, 1994 Mexican peso devaluation, 1987 U.S. market decline, and 2009 global financial crisis, among others), is often discussed in the literature. The studies found that many recent financial crises were initiated by episodes of local turmoil but ultimately spilled over to markets with seemingly little or even no economic linkages to those initial shocks. Financial contagion, the propagation of a shock to one security or market across fundamentally unrelated securities or markets, has become one of the most intriguing asset-pricing phenomena. After the 2009 global financial crisis, the International Monetary Fund (IMF), Banks for International Settlements (BIS), Financial Stability Board (FSB) and other regulatory authorities noticed that systemic risk was underestimated across the board before this crisis, and they advocated in-depth studies of this issue. All these are related to the study of the comovement of asset prices.

The comovement of security prices is one of the most fundamental aspects in asset pricing. One important issue that attracts much attention is whether the comovement is determined only by economic fundamentals or whether there is excess comovement. The traditional theories including the present discount models, derived from a frictionless economy with rational investors, state that comovement in prices reflects comovement in fundamental values. As Pindyck and Rotemberg (1993) summarized, "all present value models in which discount rates depend only on macroeconomic variables have a common implication: the prices of different stocks can move together only in response to common movements in earnings, or in response to common effects of changes in macroeconomic variables". However, many studies have found that the covariance between assets' prices observed in data is higher than that predicted by traditional asset pricing models, which is referred to as excess comovement. Much of the recent literature reveal both

* Corresponding author. Department of Information Systems, School of Economics and Management, Behang University (BUAA), Beijing, 100191, China.
E-mail address: wujj@buaa.edu.cn (J. Wu).
[1] The authors are ordered by the surname.

https://doi.org/10.1016/j.econmod.2019.03.002
Received 15 December 2017; Received in revised form 6 March 2019; Accepted 6 March 2019
Available online 12 March 2019
0264-9993/

cross-sectional and time-series patterns in excess comovement (Pindyck and Rotemberg, 1993; Kallberg and Pasquariello, 2008; Morck et al., 2000; among others), but the driving factors behind them, which go beyond economic fundamentals, remain a controversial issue calling for further investigation.

The extant studies try to explain excess comovement from a variety of perspectives especially the economies with frictions or with irrational investors, including: pure information transmission (King and Wadhwani, 1990), wealth effects (Kyle and Xiong, 2001), risk-averse agents' rebalancing activities (Fleming et al., 1998; Kodres and Pritsker, 2002), financial market fragility (Allen and Gale, 2000), financial constraints (Yuan, 2005), investors' trading patterns and sentiment (Barberis et al., 2005), cost of acquiring information (Veldkamp, 2006), strategic trading by heterogeneously informed speculators (Pasquariello, 2007), and relative real output shocks (Pavlova and Rigobon, 2007), as summarized by Kallberg and Pasquariello (2008). The friction- and sentiment-based theories of comovement proposed by Barberis, Shleifer and Wurgler (i.e., BSW, 2005) incorporate some above-mentioned perspectives and emphasize more on the effects of trading behavior and information diffusion, which are totally different from fundamentals-based theories. BSW (2005)'s friction- and sentiment-based theories of comovement include three specific views, namely, category view, habitat view, and information diffusion view. From the category view, they argue that many investors classify assets into categories first and then allocate funds at the level of these categories, rather than at the individual asset level. As BSW (2005) has noted, "if some of the investors using categories are noise traders with correlated sentiment, and if their trading affects prices, then as they move funds from one category to another, their coordinated demand induces common factors in the returns of assets that happen to be classified into the same category, even when these assets' cash flows are uncorrelated". The habitat view argues that many investors choose to trade only a subset of all available securities, because of transaction costs, international trading restrictions, or lack of information. Hence, this view of comovement predicts there is a common factor in assets returns that is attributable to the habitat of a specific subset of investors, when these investors' risk aversion, sentiment, or liquidity needs change. From the perspective of information diffusion, information is incorporated more quickly into the prices of some stocks than others. This implies that a common factor will be observed in the returns of stocks that incorporate information at similar rates.

This paper investigates the debate about comovement between traditional asset pricing theories and behavior finance theories. Specifically, we aim to answer the first question: How to *measure* and *explain* the comovement among individual stocks appropriately? This paper proposes new methods based on causal networks to measure the comovement among individual stocks and investigate the driving force behind it. The subsequent question is, can comovement structure of the market provide more information about the risk characteristics of the whole market and trading patterns of investors than the measure of degree of comovement? This is a question rarely investigated in the previous literature. We aim to fill this gap by proposing new measures to reveal market comovement structure and finding its determinants.

The first contribution of this paper is that we propose a causal network method to describe the comovement of each pair of stocks as well as the comovement structure of the whole market. The Granger causality test is employed to measure the comovement of each stock pair. A robust order identification method for the Granger causality tests is further employed to avoid biased results stemming from the sensitivity of test configurations. The causal network method is a more direct way for observing the cross-sectional and time-varying patterns of comovement among individual stocks, the degree of comovement of the market, as well as the market comovement structure. The networked representation of causal relationships is beneficial to studying the information diffusion view of the friction-based theory of comovement, which has been rarely discussed in the literature. Moreover, this new method is conducive to the discussion of the hypotheses of the sentiment-based theories and traditional theories of comovement in a unified methodology, rather than in two steps usually adopted in the literature, which makes the direct comparison of different effects possible.

Our study provides a detailed investigation of individual stocks, which could help to gain indepth insight into one of the most influential emerging markets, the Chinese market, as a complement to the literature. Using a sample in the Chinese stock market from Jan. 1, 2006 to Dec. 31, 2016, which covers the 2007–2009 global financial crisis and the Chinese stock crash around 2015, we obtain the time-varying directional causal networks. Three *topological properties* of complex networks are adopted to measure the degree of comovement. Through these topological properties, we find the degree of comovement in the Chinese stock market increases slightly from 2007 to 2009 and increases dramatically from 2011 to 2015, with a sudden rise in 2015. When the Chinese stock market crash occurred, stocks became more interconnected compared with other years during our sample period. The results are understandable because financial crisis or stock market crash can be regarded as the realization of systemic risk, and the high degree of connectedness of the causal network increases the channels through which shocks and risk can spread.

Furthermore, this paper proposes a new measure, namely the *improved PageRank score*, to investigate the comovement structure of all stocks in the market. Our study sample covers two crisis periods, one is the global financial crisis caused by subprime mortgage crisis around 2007 to 2009, and the other is the Chinese stock market crash happened around 2015. We find that while the degree of comovement in the Chinese stock market generally increases over the years, the comovement structure of the Chinese stock market in crisis periods is different from that in normal periods. The systemically important stocks measured by the improved PageRank are more influential during the periods of crisis. In all, the changes in the topological properties of the proposed causal network are consistent with the actual situation in the Chinese stock market, which justifies the usefulness, suitability and robustness of our method.

The second contribution of this paper is to provide new evidence about the reasons behind the observed causal comovement of stock returns. This paper investigates whether there is excess comovement of stock returns beyond that predicted by fundamentals-based factors, and whether the excess comovement is related to sentiment-based factors. Rather than setting up empirical analysis through bivariate or multivariate regressions on the return series in the first step and then setting up models on the residuals in the second step as in BSW (2005), we set up panel data logistic regression models directly on whether there is comovement between a pair of individual stocks. We bring different explanatory variables from the fundamentals-based view and the sentiment-based view into a unified model to facilitate the direct comparison of different factors. Three type of factors are examined. First, asset pricing factors suggested by Campbell et al. (2008) are considered, including the systematic risk measured by the beta coefficient, excess return measured by the alpha coefficient, total risk measured by daily return volatility, size factor measured by the log of market value, book-to-market ratio, and the profitability, leverage, and liquidity of the firm. Second, the dummy variables that represent whether the paired stocks belong to the same industry sector are adopted to test the industry effect revealed by King (1966) and Meyers (1973). Third, the dummy variables related to the inclusion or exclusion of individual stocks in the market index are used to investigate whether the excess comovement can be explained by sentiment-based factors after controlling for fundamental factors. No matter an individual stock is included in or excluded from a market index, the fundamentals of this individual stock do not change with the inclusion or exclusion. The market index is only a representative of the overall market; it does not signal an opinion about fundamental value. Under the traditional view of comovement, inclusion of an index should not change the price comovement of the incorporated stocks with other stocks. However, investors' behavior or sentiment may be affected by this inclusion or exclusion from the friction- or sentiment-based views. As BSW (2005) has noted, the category and habitat views suggest that the index is a preferred habitat for some investors and a natural category, which may lead to fund flows in and out of that category (habitat). If arbitrage is limited, these fund flows raise the comovement of the included stocks' return with the returns of other stocks in the index. The information diffusion view also predicts a rise in the comovement of the added stock with other stocks in the index. Under this view, stocks in the index are quick to incorporate news about aggregate cash flows, perhaps because they have particularly low trading costs or are held by investors with better access to news. When a stock enters the market index, it starts to incorporate market-wide news at the same time as other index component stocks. In our empirical study, we choose the CSI 300 as the market index for discussion.

Our empirical results reveal the excess comovement do exist in the Chinese stock market and is driven by sentiment-based factors. On one hand, our results provide new insights into pricing and the comovement relationship. We find stocks with a high systematic risk beta coefficient, book-to-market ratio, liquidity and profitability, but with low price volatility, tend to Granger cause other stocks and lead other stocks in price, while our results reject the hypothesis that the causal comovement is attributable to the industry effect. On the other hand, our study reveals that most of the causal comovement occurs in the situation that the component stocks of the market index lead non-component stocks, while non-component stocks seldom lead component stocks. This reflects that the component stocks in the market index can absorb information more quickly than non-component stocks. Our results support the category view of sentiment-based theories of comovement and provide direct evidence for the information diffusion view. As stated in the literature, if the component stocks of the market index are traded by investors, the correlation of these stocks should increase, while the causal comovement should decrease. Our results are consistent with the latter.

The third contribution of this paper is to discuss the determinants of comovement structure based on the newly proposed measure, the improved PageRank score. In normal periods, the most important stocks in the market are those with high excess return, high total risk volatility, high profitability of the firm, no matter whether they belong to the market index or some sectors. However, the most important stocks in the market during crisis periods are those with low total risk volatility and large size, especially belonging to the market index. These determinants uncover that the risk preference of investors as well as the utility function used by investors for stock selection change in different market situations to some degree. It seems that investors prefer the stocks with high volatility that can provide more opportunities to earn high returns in normal periods, while investors prefer low volatility stocks that can defense against risk in crisis periods. Furthermore, on one hand, it seems that investors make decisions based on the risk-return balance in normal periods, which implies the utility function used by investors in normal periods includes risk and expected return. On the other hand, it seems that investors seek safe-haven assets in crisis periods and the utility function used by investors in crisis periods may only contain risk. Indeed, investors in the Chinese stock market during crisis period select stocks with the consideration that the total risk should be low, and the component stocks of the market index or some sectors are better, because large size stocks and some sectors can defense risk better. Investors draw funds out of high total risk and cyclical-sector stocks, while make funds flow into the stocks with low total risk and belonging to the market index or some defensive sectors.

These results have some extended implications about rational investors and rational trading behaviors. Since risk-return balanced investors are assumed to be rational investors, our results to some extent reveal the existence of rational investors in the Chinese stock market, which was previously regarded as a retail investor-dominated emerging market. Furthermore, our empirical results provide new evidence for the "style investing" mentioned in Barberis and Shleifer (2003), which refers to the trading practice of investors in allocating funds across labels of "styles" rather than individual securities. Our results uncover that investors would like to invest in the component stocks in the market index and some particular sectors during crisis periods.

The remainder of this paper is organized as follows. Section 2 provides a brief literature review. Section 3 discusses the methodology for causal network construction. Section 4 illustrates the network topological properties and the comovement trends. Section 5 introduces the models to identify the driving sources of the comovement. Section 6 investigates the key determinants of comovement structure in the market, and we finally conclude our work in Section 7.

2. Literature review

Researchers have uncovered numerous patterns of comovement in asset returns, including cross-sectional and time series patterns. The cross-sectional evidence shows strong common factors in, for example, the returns of small-cap stocks, value stocks, and stocks in the same industry, whereas the time series evidence shows there is business cycle related time-varying comovement. Pindyck and Rotemberg (1993) test whether comovements of individual stock prices can be justified by economic fundamentals, and their results reject this hypothesis and report excess comovement of returns. They also show that excess comovement for firms can be partly explained by company size and the degree of institutional ownership, suggesting market segmentation. Kallberg and Pasquariello (2008) observe that excess comovement is uniformly significant across industries and over time, among 82 industry indices in the United States' (U.S.) stock market. Ribeiro and Veronesi (2002) illustrate patterns of comovement change over the business cycle, that is, comovement is higher in recessions than in booms. Morck et al. (2000) show the degree of stock prices moving together is higher in emerging markets, whereas asset price comovement has been steadily decreasing in the U.S. and emerging markets. They argue that the systematic component of returns variation is large in emerging markets and appears unrelated to fundamental comovements, which are consistent with noise trader risk.

A large body of literature has investigated international market interdependence. Most studies have discussed the comovement or connectedness of asset prices worldwide through using indices in different countries as proxies for their analysis, while some other studies have focused on different sectors in one market by using the indices in different sectors as proxies. Moreover, other studies have focused on the comovement of the economy or economic policies between countries. Li and Peng (2017) discuss China's economic policy. Pan and Mishra (2018) discuss the development of the Chinese stock market and economy. Xue and Zhang (2017) discuss the predictability of stock returns in the Chinese stock market. Notably, there is very limited work on the comovement of individual stocks in the stock market, especially in the emerging market. Our study complements the literature by extending the analysis to the comovement of individual stocks in the Chinese stock market, an emerging market, to provide additional empirical evidence of the patterns of stock returns' comovement.

The methods used to measure the comovements of asset prices include correlation (e.g., Forbes and Rigobon, 2002; Tse et al., 2010), partial correlation (Jung and Chang, 2016), cointegration relationships (Awokuse et al., 2009; Tu, 2014), variance decomposition (Diebold and Yılmaz, 2014), and causal linkages (Masih and Masih, 1999; Billio et al., 2012). Recently, some studies have used complex network topology to address this issue (Mantegna, 1999; Bonanno et al., 2001; Tse et al., 2010; Peralta and Zareei, 2016; Billio et al., 2012; Tu, 2014). Correlation has been the most used measure of comovement. Most of the complex networks used to discuss asset prices' connectedness in the literature have been constructed according to correlations, including the work of Mantegna (1999), Bonanno et al. (2001), Tse et al. (2010), and Peralta and Zareei (2016). Mantegna (1999) analyzes the clustering structure of stocks using complex networks. Tse et al. (2010) construct complex networks to study correlations between the closing prices for all U.S. stocks and propose that stocks corresponding to nodes of high degrees can be used to compose a new index to reflect market variations naturally and adequately. Some recent studies have adopted Granger causality networks, such as Billio et al. (2012). Billio et al. (2012) find that all four sectors of hedge funds, banks, broker/dealers, and insurance companies have become highly interrelated over the past decade in the U.S. market, which is likely to increase the level of systemic risk in the finance and insurance industries through a complex and time-varying network of relationships. Their results show an asymmetry in the degree of connectedness among the four sectors, with banks playing a much more prominent role in transmitting shocks than other financial institutions.

Granger causality tests have been used frequently to investigate the relationship among economic variables and financial variables, such as Hsiao (1981), Tsani (2010), Bu (2011), Dagher and Yacoubian (2012). However, the discussions regarding the comovement of individual asset prices by using complex networks via Granger causality tests remain very limited. Our study examines the topology properties of causal networks constructed according to the Granger causality tests between pairs of stocks, rather than by using correlation and cointegration, which dominants the literature. This is because the causal networks are more useful for observing the lead-lag relationships among stocks returns and examine the information diffusion effect. We discover and prove the proposed topology properties of causal networks can measure the comovement feature of stock prices. Moreover, many studies have used the comovement of asset prices to illustrate risk contagion, for example, using observed increases in correlation after a shock to one country (or group of countries) as a measure of financial contagion (e.g., Forbes and Rigobon, 2002). Forbes and Rigobon (2002) also show that correlation coefficients are conditional on market volatility, and there was no increase in unconditional correlation coefficients (that is, no contagion) during the 1997 Asian crisis, 1994 Mexican devaluation, and 1987 U.S. market crash after adjusting for heteroscedasticity biases. The time-varying complex networks based on Granger causality tests of individual asset prices in our paper has the advantage of illustrating risk contagion. In addition, some researches have related the comovement of stocks price based on a complex network to portfolio selection (e.g., Onnela et al., 2003; Peralta and Zareei, 2016). The results of our study have significant implications for portfolio selection.

3. Preliminaries

3.1. Comovement of stock returns measured by the Granger causality tests

This paper examines the lead-lag relationships between stocks in the Chinese stock market. In doing so, we adopt the Granger causality test (Granger, 1969, 1980) to measure the lead-lag relationship, rather than the methods of contemporaneous correlation or cointegration. We specify the Granger causality test as the simplest linear form on the returns for each pair of stocks as Eqs. (1) and (2).

$$R_{X,t} = \alpha + \sum_{i=1}^{m} \lambda_i R_{X,\,t-i} + \sum_{j=1}^{n} \theta_j R_{Y,t-j} + \varepsilon_t, \tag{1}$$

$$R_{Y,t} = \delta + \sum_{i=1}^{r} c_i R_{X,t-i} + \sum_{j=1}^{s} d_j R_{Y,t-j} + \eta_t, \tag{2}$$

where $R_{X,t}$ and $R_{Y,t}$ are the log return of the closing price of stock X and Y respectively, $R_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$, ε_t and η_t are uncorrelated white-noise series, and (m,n) and (r,s) are the lag structure of each equation. For equation (1), the null hypothesis is $H_0: \theta_j = 0$ for each j. If the null hypothesis of Eq. (1) is rejected, the returns of stock Y lead the returns of stock X. For equation (2), the null hypothesis is $H_0: c_i = 0$ for each i. If the null hypothesis of Eq. (2) is rejected, the returns of stock X is the Granger cause of the returns of stock Y. We use the Wald test to test the hypothesis. The stock return series are all stationary series based on the augmented Dickey-Fuller (ADF) unit root test. We use stationary series to set up the Granger causality test to avoid the problem, as stated in Granger (1980), that the simple Wald test for the Granger causality test may not be valid due to the absence of a standard distribution of the test when time series are not stationary.

Considering that the results of the Granger causality test may be sensitive to the selected lag order, we discuss how to select the orders in greater detail and compare its effects on the Granger causality test results. The most common method to determine lag orders is to minimize the Akaike information criterion (AIC). Hsiao (1981) proposes a stepwise procedure based on the Akaike final prediction error criterion to identify the lag order of each variable for the Granger causality test. We add the idea of the stepwise procedure in Hsiao (1981) and modify the procedure based on the AIC to identify the lag order for the Granger causality test. We explain the order identification procedure by taking Eq. (1) as an example. First, we identify the lag order m by treating $R_{X,t}$ as a one-dimensional autoregressive process and choosing the order that yields the minimum AIC with the maximum lag order of M. Second, identify the order of lags n by treating $R_{X,t}$ as a controlled variable with the lag order m prespecified in step 1, and determining the lag order n of $R_{Y,t}$ that yields the minimum AIC of Eq. (1) with the maximum lag order of M. In this paper, we set the maximum order of lags is 5 in our empirical study.

To examine the robustness of our lag order identification method for the Granger causality test, we compare the network topological properties of the network constructed by our method and other alternative order identification methods for the linear Granger causality test. The two methods used for comparison are: first, the simplest method in the literature that specifies $m = n$, $r = s$ for Eqs. (1) and (2), and chooses the minimum AIC to identify the order of lags for each equation; second, the stepwise procedure based on the Akaike final prediction error criterion used in Hsiao (1981). Section 4.1 reports these results.

In this study, we consider the simplest form of the Granger causality test rather than nonlinear causality, which has been discussed frequently in the recent literature, for example, Hiemstra and Jones (1994), Diks and Panchenko (2006), and Diks and Wolski (2016). Because the main focus of this study is the lead-lag relationship between stock returns, which can be easily measured by linear Granger causality. Moreover, the linear Granger causality test is easier to estimate, compared with the nonlinear Granger causality test. Therefore, we choose the most simple and effective method.

Using the results of the Granger causality test, we can construct a directed network based on the lead-lag relationship between stock returns. The stocks are regarded as the vertices in the causal network. If stock I Granger causes stock J, there is a directed edge from stock I to stock J (e.g. $I \to J$). Additionally, the p-value of the Granger causality test from stock I to stock J is recognized for the edge from stock I to stock J. To ensure that the network not only contains the most important information but also is as simply as possible, we choose the 5% significant level for the simple Wald test. For a pair of stocks, there are three types of relationships between stock I and stock J: no causal relationship, a unidirectional causal relationship ($I \to J$ or $J \to I$), and bidirectional causal relationships ($I \to J$ and $J \to I$), respectively. To better measure the causal comovement of individual stock returns, we adopt the results of the Granger causality tests directly for each ordered pair of stocks, that is $Y_{i \to j,t} = 1$, if stock i is a significant Granger cause of stock j in year t; otherwise, $Y_{i \to j,t} = 0$. This definition can illustrate all kinds of causal relationships between each pair of stocks.

3.2. Network topological properties of causal networks

3.2.1. Density, average distance, and clustering coefficient

After constructing the causal networks, we examine topological properties, such as density, average distance, global clustering coefficient, and the improved PageRank score. Using these topological properties, we can measure the degree to which the whole causal network tends to co-move together and comovement structure. The density of a directed network is equal to the proportion of the directed edges present in the directed network. It is calculated as the number of the directed edges existing in the directed network divided by the possible number of directed edges. Since a directed edge may exist in an ordered pair of nodes, there are *N(N-1)* possible edges in a directed network. The density of the causal network (*DCN*) equals the fraction of statistically significant Granger causality relationships among *N(N-1)* pairs of *N* stocks. The *DCN* of stock market is defined as

$$DCN = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} A_{i \to j}, \tag{3}$$

where A is the adjacency matrix of the causal network. $A_{i\to j} = 1$, if stock i is the significant Granger cause of stock j, that is, there is a causal directed edge from node i to node j; otherwise, $A_{i\to j} = 0$. Notably, A is an asymmetric matrix in a causal directed network. The higher the density, the higher degree of the lead-lag comovement of the overall stock market network. The density might more precisely be understood as a "1-step" lead-lag relationship, because the counted directed edges are a direct lead-lag relationship.

There are transmission relationships in a causal directed network. For example, if stock i is not causal directed to node j directly, but stock i is the Granger cause of stock k, while stock k is the Granger cause of stock j, then there is a "multi-step" deductive lead-lag relationship from stock i to stock j. The average distance of the causal network (*ADCN*) is defined as the mean distance between any pair of nodes,

$$ADCN = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} Dist_{i \to j}, \tag{4}$$

The distance from node i to node j ($Dist_{i\to j}$) is defined as the smallest number of edges that must be traversed to move from node i to node j. Notably, $Dist_{i\to j}$ may not be equal to $Dist_{j\to i}$ in a directed network. Closely related to the idea of distance is average distance. When there is no path from node i to node j, we set $Dist_{i\to j} = N-1$, which represents the maximum distance that may appear in a directed network with N nodes. The smaller the average distance, the higher the degree of "multi-step" lead-lag comovement in the causal network.

A global clustering coefficient (*GCC*) provides an overall indication of the extent to which stocks tend to cluster together and create tightly connected groups. The *GCC* is defined as the ratio of the closed triples and the connected triples in the network, that is,

$$GCC = \frac{Number\ of\ the\ closed\ triples}{Number\ of\ the\ connected\ triples}. \tag{5}$$

The connected triples consist of open triplets and closed triplets, which correspond to the structure of three connected nodes with two and three edges, respectively. The higher the *GCC* of the stock network, the higher the extent to which stocks form a tightly connected group and co-move. Notably, the direction of the edge is ignored in calculating *GCC*.

3.2.2. Improved node importance measure in causal directed networks

PageRank is first proposed by Page et al. (1999), a founder of Google Inc. and successfully used in the Google search engine, which evaluates the importance of web pages by their hyperlinks. The higher the PageRank value, the greater the importance of the web page in the search results. The basic idea of PageRank is based on two assumptions regarding quantity and quality. One assumption is that when a web page obtains a greater number of in-links, it is more important. The concept of PageRank can be also understood as a voting scheme, that is, whenever there is a hyperlink from web page i to web page j, a vote from web page i to web page j is produced; hence, the rank of web page j increases. Additionally, the strength of the vote from web page i to web page j also depends on the rank of web page i, that is, the greater the importance of web page i, the greater the strength of its votes. Another assumption is that the web page will be more important when it gains in-links from important web pages. The PageRank, PR_i of web page i is defined as Eq. (6).

$$PR_i = d \sum_{j \in T_i} \frac{PR_j}{c_j} + \frac{(1-d)}{N}, \tag{6}$$

where T_i is the set of pages that point to i, PR_j is the PageRank value of web page j, c_j is the number of out-links from web page j, N is the number of pages, and d is the damping factor that is a scalar value between 0 and 1. The first term of the sum in Eq. (6) models the voting scheme, and the second term represents the probability of a surfer randomly jumping to any web page, for example, without following any hyperlinks. The damping factor, usually set in the [0.85.0.95] range, models the manner in which these two terms are combined at each step. The PageRank score is computed by a power iteration method. First, we set the same initial value, $1/N$, for every vertex as its original PageRank score. Next, we apply Eq. (6) to iteratively update the PageRank until the algorithm converges with an iteration count of greater than 1000 or the error threshold less than 0.001.

The PageRank algorithm has been applied to the analysis of financial institutions and securities market recently. Cetorelli and Peristiani (2013) construct a network of 45 financial centers based on their ability to attract IPOs, and then use the PageRank score to assess their global relative importance. Yang et al. (2014) investigate the cointegration relationships among 26 global stock indices and rank their influence by applying the PageRank algorithm to the constructed cointegration network. Tu (2014) constructs a complex network of the Chinese stock market based on the Engle-Granger cointegration test and analyzes the network by calculating topological properties and a PageRank score to deepen the understanding of the Chinese stock market.

The mentioned studies above have used the PageRank definition shown in Eq. (6) directly to analyze financial markets, which obtains the relative importance of nodes with in-links. However, our study emphasizes the importance of stocks that can affect others in the market, or the information discovery power of stocks. Thus, we should consider the importance of nodes with out-links, in contrast with web page analysis. Therefore, we propose a new out-links PageRank measure, $PRnew_i$, to analyze the Granger causality networks, represented in Eq. (7).

$$PRnew_i = \mathrm{d} \sum_{j \in K_i} \frac{PR_j}{e_j} + \frac{(1-d)}{N}, \tag{7}$$

where K_i is the set of stocks that stock i points to and e_j is the number of in-links to stock j. In other words, K_i is the set of stocks affected by stock i through price, and e_j is the number of stocks that lead stock j in price. We set the damping coefficient at 0.85 as a matter of experience and obtain the new out-links PageRank value for each stock in the causal directed stock network to measure the information discovery power of individual stocks, emphasizing the importance of stocks that can affect and lead the price movement of other stocks.

4. Data and analysis of comovement and comovement structure

4.1. Data and robustness test results of the Granger causality tests

This study examines the comovement characteristics of the Chinese stock market. We construct a network of individual stocks in the market. To decrease computational complexity, we select a pool of stocks. We select component stocks in the CSI 300 Index of the Chinese stock market, which contains stocks from the Shanghai Stock Exchange and the Shenzhen Stock Exchange, and accounts for nearly 60% of the total market value. The constituent stocks of the CSI 300 Index are adjusted every six months. We choose all the constituent stocks of the CSI 300

Index in every adjustment during our sample period to make sure the size of nodes in the complex network is the same in our whole study sample. Our sample period is from Jan. 1, 2006 to Dec. 31, 2016. After excluding stocks with shares continually suspended over 100 days during our sample period, not listed at the beginning of our study period, or delisted by the end of our study sample, we select 377 stocks to construct causal networks. We collect the daily closing prices of stocks from the WIND database. To show the dynamics of stock price comovement, we adopt the rolling window estimation of the Granger causality tests for each pair of stocks. We examine the Granger causality tests for each year using the daily data for the whole year and correspond the test results to the last day of each year.

To make sure the constructed causal networks is suitable enough, we perform several checks. Before preceeding, we examine the order identification of the Granger causality test to make sure the results of Granger causality tests are robust. We compare the order identification method for the Granger causality test proposed by this paper with two alternative methods that are discussed in Section 3.1. The first method for comparison (recorded as method 1) is the simplest method in the literature that specifies the two variables' orders of lags the same, taking Eq. (1) as example, i.e., assumes $m = n$, and chooses the minimum AIC to identify the orders of lags. The second method for comparison (recorded as method 2) is the stepwise procedure based on Akaike final prediction error criterion in Hsiao (1981). We are concerned with the consistency of the Granger causality results under different order identification methods. We calculate the proportion of the pairs of stocks with the same results for the statistically significant Granger causality relationship in all *N(N-1)* ordered pairs of stocks between the different order identification methods, which we define as consistency. Fig. 1 illustrates the comparison results. The consistency of our order identification method compared with method 1 is approximately 88.00% to 95.74% in the yearly causal networks, with an average of 93.77%. The consistency of our order identification method compared with method 2 is approximately 81.31% to 85.41%, with an average of 83.40%. From these comparison results, we confirm that our order identification method for the Granger causality test is robust.

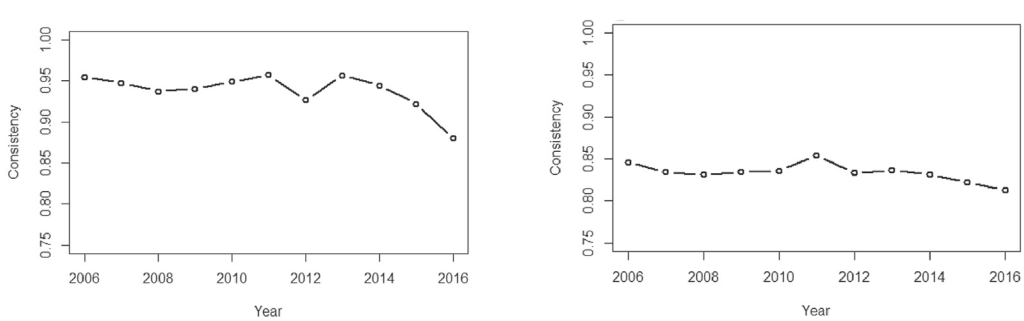
4.2. Illustration of comovement over time in the Chinese stock market

To make sure the causal networks reveal the actual situation of the Chinese stock market, we examine the comovement and the topological properties over time through the constructed causal networks in the Chinese stock market based on the constituent stocks of the CSI 300 Index in each year. As we have already known, our study sample covers two crisis periods. One is global financial crisis caused by subprime mortgage crisis around 2007 to 2009. Another one is Chinese stock market crash happened around 2015. A good measure constructed based on the causal networks should reflect the actual situation of the stock market. Therefore, we check the causal networks and measures constructed using our methods carefully.

First, we show the changes of comovement of individual stocks over time using the figures of the causal networks directly. To ensure the figures as clear as possible, we only report the 100 largest stocks in terms of their average market value during the whole sample period, illustrate them in four colors for the four quantiles, and only report the networks of these selected stocks in the selected two years 2011 and 2015 in Figs. 2 and 3. Red represents stocks with market value rankings from 1 to 25, black represents stocks with market value rankings from 26 to 50, yellow represents stocks with market value rankings from 51 to 75, and blue represents stocks with market value rankings from 76 to 100. From these two figures representing the two different years, we can easily observe the network changes. The reason why we choose to report the causal network in 2015 is that Chinese stock market crash happens in 2015, while the network in 2011 is used to comparison to show the time-varying characteristics of comovement in Chinese stock market. From these two figures, we can easily find the comovement among stocks increases a lot in 2015 compared to 2011.

Second, we examine topological properties, including density *(DCN)*, average distance *(ADCN)*, and global clustering coefficient *(GCC)* based on the constructed causal networks for the Chinese stock market in each year, to uncover the changes of comovement of individual stocks over time. As illustrated in Fig. 4, we observe similar changes for *DCN* and *GCC*, including a slight increasing trend from 2007 to 2009, a clear increasing trend from 2011 to 2015, and a sudden increase in 2015. The *ADCN* has decreasing trends from 2007 to 2009 and from 2011 to 2016, and a sudden decrease in 2015. Specifically, the *DCN* is 0.19 in 2014 and increases by nearly 35% to 0.2568 in 2015; the *GCC* is 0.4192 in 2014 and increases by nearly 35.69% to 0.5689 in 2015; the *ADCN* is 1.82 in 2014 and decreases by nearly 5.83% to 1.76 in 2015. From these figures and results, we can easily observe that the comovement of the Chinese stock increases a lot since 2011, and the degree of comovement has changed dramatically in 2015. Particularly, when the Chinese stock market crash occurred in 2015, stocks became more interconnected than that in other years during our sample period, which are illustrate by suddenly increase of *DCN* and *GCC* in 2015 and sudden decrease of *ADCN* in 2015. The changes in these topological properties of the Chinese stock causal network constructed by using our method are consistent with the actual situation in the market. The results are understandable because financial crisis or stock market crash can be regarded as the realization of systemic risk, and the high degree of connectedness of the stock network increases the channels through which shocks and risk can spread.

Finally, we examine whether a possible structural break existed in 2015 by applying the Z-A test (Zivot and Andrews, 1992), because the value of topological properties shows huge change in 2015. The Z-A test regards all the points of the time series data as possible structural breaks, conducts the ADF test step by step, and then determines the time point that minimizes the t-statistic as a structural break. Notably, we only allow for a break in intercept, considering that the time series of the topological properties do not show an obvious trend. Fig. 5 presents the Z-A test results of the structural breaks for the topological properties of the causal



(A) Consistency compared with method 1

(B) Consistency compared with method 2

Fig. 1. Consistency of Granger causality results under different order identification methods.
Note: This study uses the stepwise procedure based on the AIC as our order identification method. Method 1 is the simplest method in the literature that assumes $m = n$, $r = s$ for Eqs. (1) and (2), and Method 2 is the stepwise procedure based on the Akaike final prediction error criterion used in Hsiao (1981). The consistency here is the proportion of the pairs of stocks with the same results for the statistically significant Granger causality relationship between the different methods.

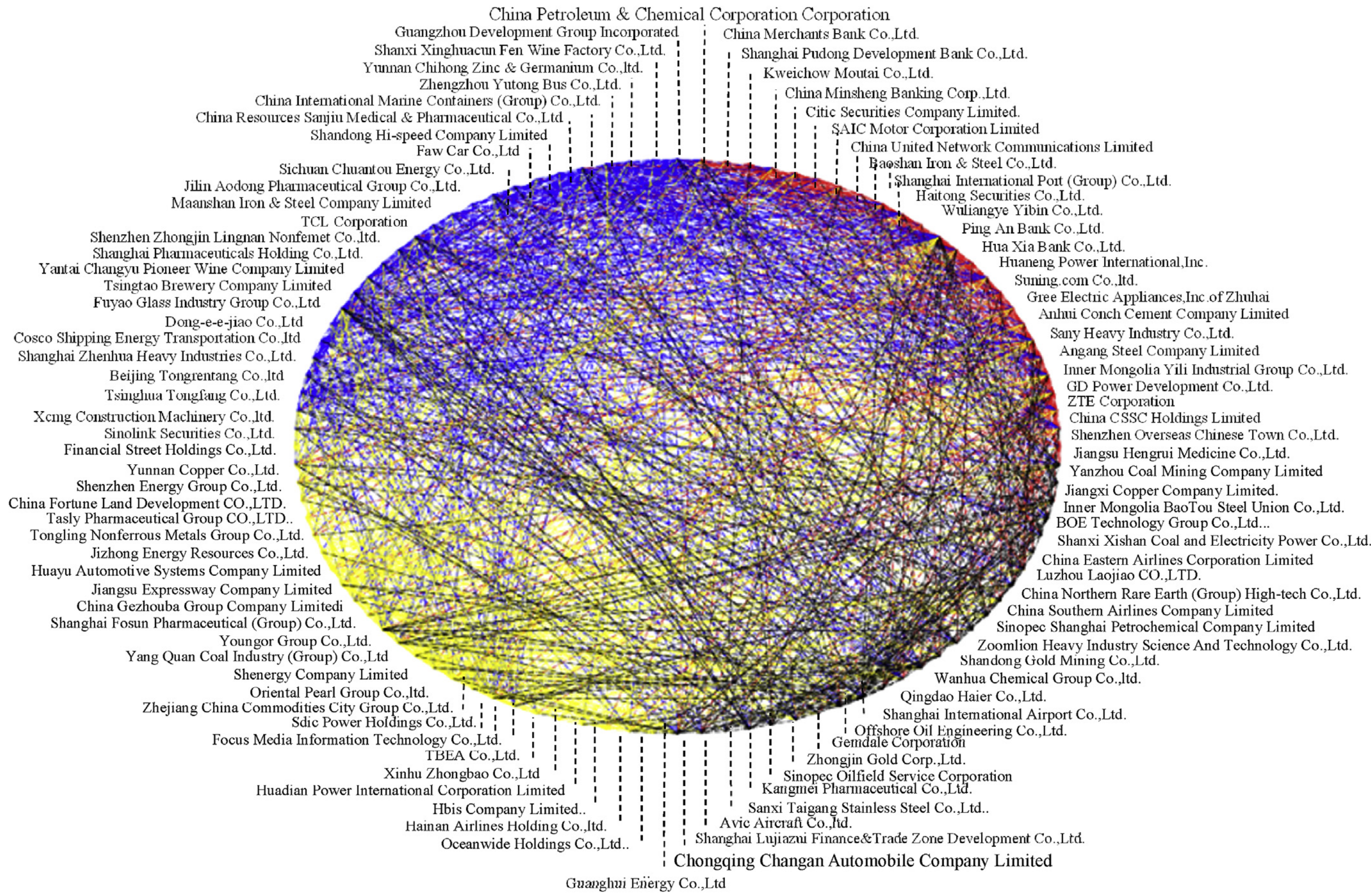


Fig. 2. Stock market network in 2011 of the 100 largest stocks in terms of average market value for the whole sample.
Note: The stocks are illustrated in four different colors representing the ranking quantiles according to their average market values for the whole sample. Red represents stocks with market value rankings from 1 to 25, black represents stocks with market value rankings from 26 to 50, yellow represents stocks with market value rankings from 51 to 75, and blue represents stocks with market value rankings from 76 to 100. The stock market network is characterized by lead-lag relationships calculated by Granger causality tests. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

network in the Chinese stock market. The Z-A test results show that structural breaks existed in 2015 according to the density and clustering coefficient, although the Z-A test based on the average distance shows a structural break in 2011. These results are consistent with the Chinese stock market crash in 2015.

Throughout the above analyses of the constructed causal networks, we confirm that the Granger causality test based on our proposed order identification method illustrates the causal comovement well. Therefore, we can adopt the results of Granger causality tests for each ordered pair of stocks as the comovement measure of individual stocks. Moreover, we prove the usefulness of topological properties as indicators of comovement and comovement structure in the stock market. The *DCN*, *ADCN*, and *GCC* can be used to reflect the degree of comovement in the market. In addition, *GCC* can provide the information of comovement structure. The higher of the value of *GCC*, the greater the extent to which stocks tend to cluster to together and create tightly connected groups.

4.3. Comovement structure revealed by the improved PageRank score

To investigate the comovement structure more deeply, we need good measures to address this issue. The newly proposed out-links PageRank score based on the causal networks, $PRnew_i$, can be used to measure the importance of stocks in the market by the definition of PageRank algorithm, and the information discovery power of individual stocks, because the based networks are constructed by the Granger causality tests. Here, we discuss more about the improved PageRank score to investigate its ability to uncover some characteristics related to comovement structure of networks. Fig. 6 shows the statistics of the improved PageRank score in each year. We find that the maximum value of $PRnew_i$ every year changes a lot, while the minimum value and mean of $PRnew_i$ every year are almost unchanging. Meanwhile, the standard deviation of $PRnew_i$ every year changes much. The deviation of the improved PageRank score remained at a high level from 2007 to 2009 and in 2015, corresponding with the two periods of the global financial crisis and Chinese stock market crash, respectively. The larger the deviation of the improved PageRank score, the more dispersed the distribution of the node influence. The high deviation of the improved PageRank score from 2007 to 2009 and in 2015 means that the most important leading stocks exist, which response to shocks more quickly than others in the market. Thus, we find that there are very important stocks with high information discovery power that can affect other stocks in the market during the periods with high maximum and high standard deviation of $PRnew_i$, that is around 2008 and in 2015.

To confirm the improved PageRank score can uncover the changes of comovement structure in the market, we adopt the quantile-quantile (Q-Q) plots of $PRnew_i$ in different years to compare the probability distributions. As shown in Fig. 7, we provide the Q-Q plots of $PRnew_i$ in different years against that in 2008, 2015 and 2011, corresponding to the global financial crisis, the Chinese market crash, and a normal year but with a structure break according to the degree of comovement, respectively. According to the panel (A) of Fig. 7, the underlying distribution of $PRnew_i$ in 2008 is quite similar to that in 2009, 2007, 2015 and 2016. According to the panel (B) of Fig. 7, the underlying distribution of $PRnew_i$ in 2015 is quite similar to that in 2016, 2007, 2009 and 2008. Although

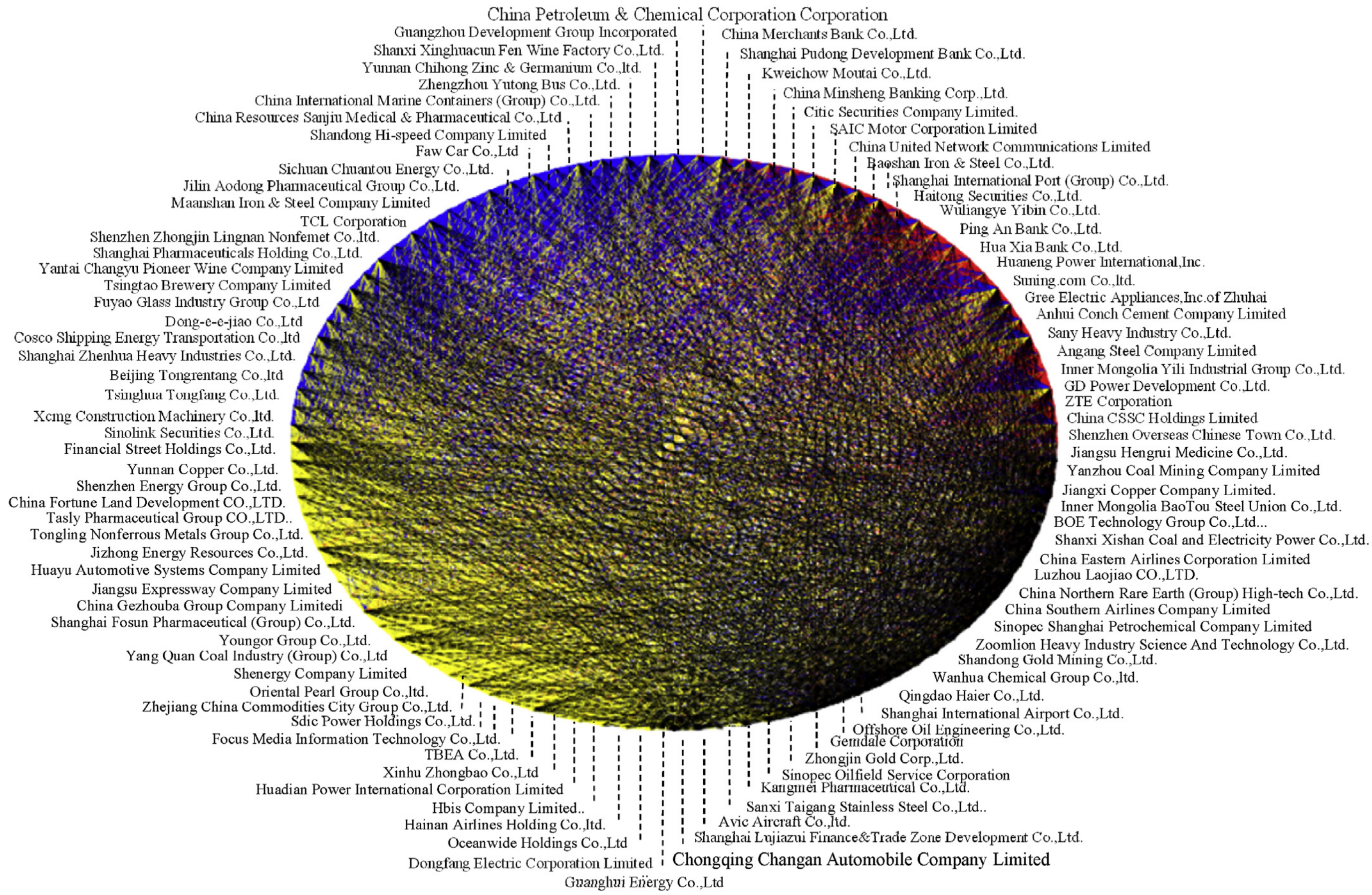
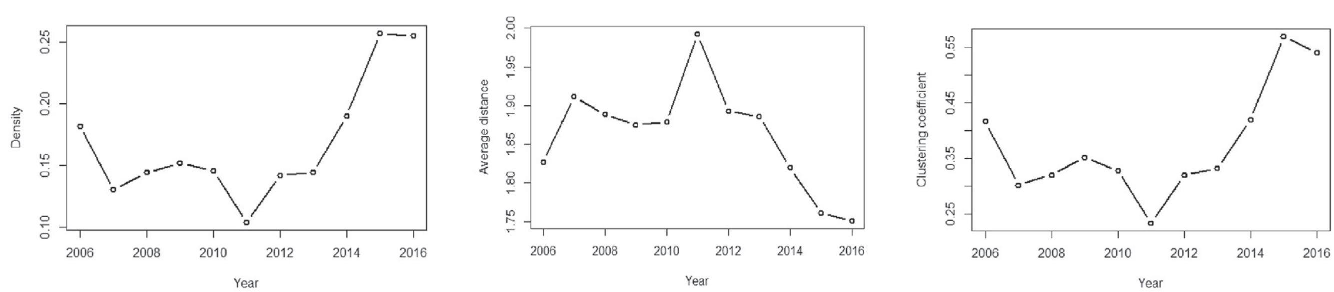


Fig. 3. Stock market network in 2015 of the 100 largest stocks in terms of average market value for the whole sample.
Note: The stocks are illustrated in four different colors representing the ranking quantiles according to their average market values for the whole sample. Red represents stocks with market value rankings from 1 to 25, black represents stocks with market value rankings from 26 to 50, yellow represents stocks with market value rankings from 51 to 75, and blue represents stocks with market value rankings from 76 to 100. The stock market network is characterized by lead-lag relationships calculated by Granger causality tests. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

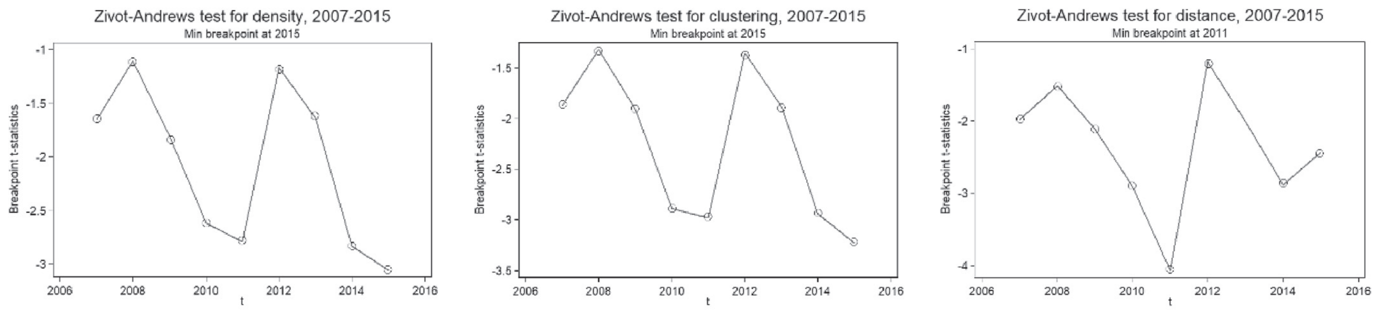


(A) Density (*DCN*) (B) Average distance (*ADCN*) (C) Global clustering coefficient (*GCC*)

Fig. 4. The topological properties of the causal networks over time.

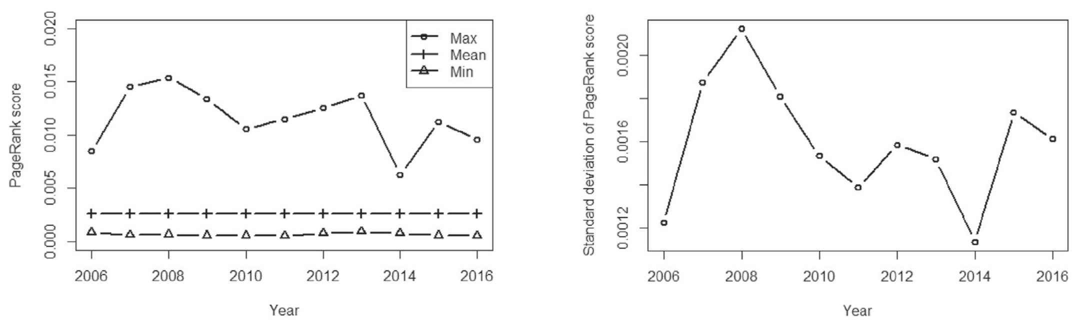
there is a structure break for the degree of comovement in 2011 according to the Z-A test for *ADCN*, 2011 is a normal period. According to the panel (C) of Fig. 7, we find the underlying distributions of $PRnew_i$ in 2011 is quite similar to that in 2010, 2012, 2013. The recent global financial crisis begins from the U.S. subprime mortgage crisis in 2007 and ends around 2009. The Chinese stock market crash happens in 2015 and the market still falls in 2016. Therefore, the underlying distributions of $PRnew_i$ in crisis periods are similar, while the underlying distributions of $PRnew_i$ in crisis periods are not similar to that in normal periods. This implies that the improved PageRank score can provide more information about comovement structure. Furthermore, the meanings of comovement structure and degree of comovement are not the same. This is because the degree of comovement in the Chinese stock market is increasing from 2011 to 2015, while the comovement structure is different during this period between 2011 and 2015.

To provide more evidence about the comovement structure and show the structure as clearly as possible, we use new ways to present the causal networks of the Chinese stock market in each year. We adopt the Kamada and Kawai (1989)'s algorithm (KK algorithm) to draw the graph of causal networks including all constituent stocks of the CSI 300 Index in our sample. In the KK algorithm, the concept of the ideal distance of non-adjacent nodes is proposed. The ideal distance between two nodes is proportional to the length of the shortest path between them. Thus, the nodes with many links with other nodes in a complex network would be



(A) Z-A test for *DCN* (B) Z-A test for *GCC* (C) Z-A test for *ADCN*

Fig. 5. Z-A test for the topological properties of the Chinese stock causal network.



(A) Max, min and mean of $PRnew_i$ (B) Standard deviation of $PRnew_i$

Fig. 6. The statistics of improved PageRank score of individual stocks over time.

placed in the middle of the graph according to KK algorithm. In addition, the nodes are illustrated in four different colors representing the ranking quartiles according to their improved PageRank scores in descending order. To be specific, according to the improved PageRank scores, red represents the nodes ranked in the first quartile, blue represents the nodes ranked in the second quartile, yellow represents the nodes ranked in the third quartile, and green represents the nodes ranked in the fourth quartile.[2] In crisis periods such as 2009, 2015 and 2016, the nodes of different improved PageRank levels appear to be circularly distributed, and the most important nodes are closely clustered in the middle. In normal periods such as the years from 2010 to 2013, the nodes of different improved PageRank levels appear to be disorderly distributed, and the nodes with different improved PageRank levels are scattered in the graph. To save the space, we only show several selected causal networks in Fig. 8.[3] These results show that the most important stocks can affect more stocks in the market in the crisis periods than that in normal periods.

Combining all results above related to comovement structure, we confirm the changes of comovement structure in the Chinese stock market. The comovement structure of stocks in crisis periods are different from that in normal periods. The systemic important stocks measured by the improved PageRank are more influential during times of crisis. It seems that the structure of causal networks is similar to star-shaped in the crisis periods.

5. Driving factors of causal comovement

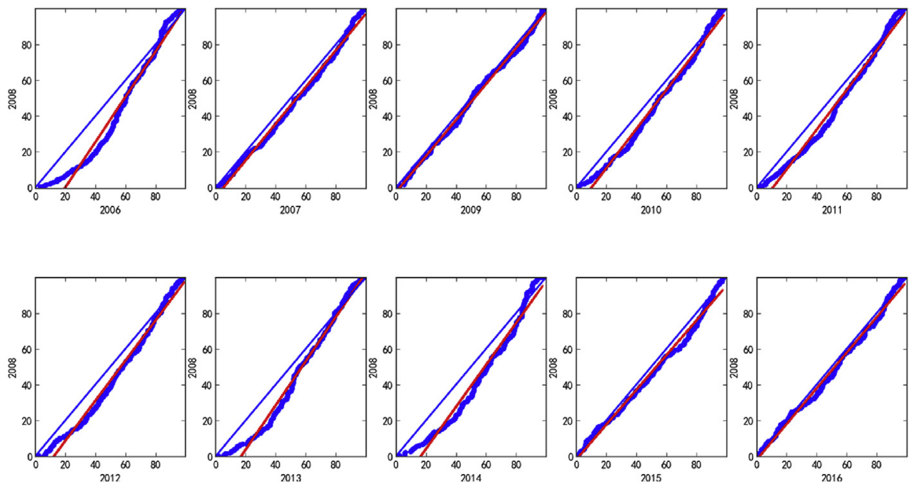
5.1. *Factors related to the fundamental sources of risk and firms' characteristics*

In this section, we investigate whether the causal comovement of individual stock returns can be justified by fundamental sources of risk, sector fundamentals, and fundamentals of firms. We first consider the fundamental sources of risk, which have been proven to be the pricing factors of stocks and some characteristics of financially distressed stocks discussed in the literature, such as Fama and French (1996) and Campbell et al. (2008) among others, including the systematic risk measured by the beta coefficient (*Beta*), size factor measured by the log of market capitalization (*Mvalue*), value factor measured by the book-to-market ratio (*BMratio*), excess return measured by the alpha coefficient (*Alpha*), total risk measured by daily stock return volatility (*Volatility*), the profitability measured by the ratio of net income to total assets (*Profitability*), leverage measured by the ratio of total liabilities to total assets (*Leverage*), and liquidity measured by the ratio of cash and short-term asset to total assets (*Liquidity*). Furthermore, we investigate whether industry effects exist in the causal comovement of stocks. The literature, for example King (1966), has argued that most of the correlation in returns is attributable to industry effects. We test this hypothesis by adding the variable representing whether the pair of stocks belongs to the same industry sector into the model. $INDUSTRY_{ij} = 1$, if stock i and stock j belong to the same industry sector; otherwise, $INDUSTRY_{ij} = 0$. Here, we adopt CITIC industry classification to study the industry effect.

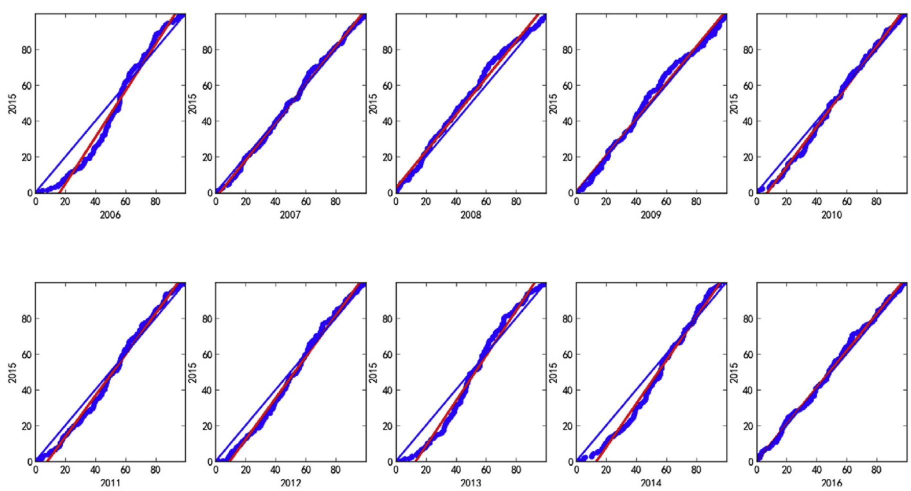
We estimate the *Beta* and *Alpha* coefficients for each stock through the index model of the CAPM by using the CSI 300 Index as the market index with an estimation window of the whole year of daily closing price log returns. Total risk is calculated by the standard deviation of each firm's daily stock return of each year. All variables related to the accounting data, including *BMratio*, *Profitability*, *Leverage* and *Liquidity*, are calculated based on the financial statements from listed companies' annual

[2] To be specific, red represents the nodes with the improved PageRank score ranked in the top 95, blue represents the nodes with the improved PageRank score ranked from 96th to 189th, yellow represents the nodes with the PageRank score ranked from 190th to 283rd, and green represents the nodes with the improved PageRank score ranked from 284th to 377th.

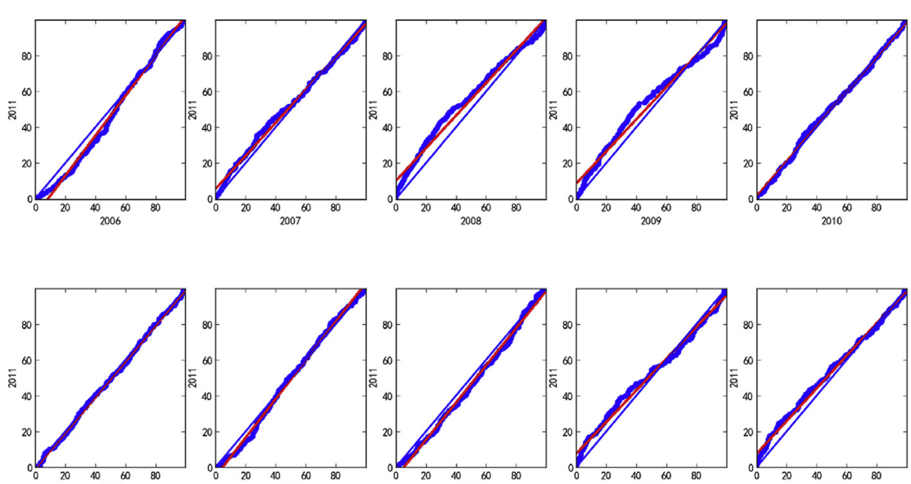
[3] To save space, the causal networks illustrated according to KK algorithm and different improved PageRank levels in each year are not all shown in Fig. 8, only selected graphs reported in this paper. The graphs are available upon request.



(A) The Q-Q plots of $PRnew_i$ in different years against that in 2008

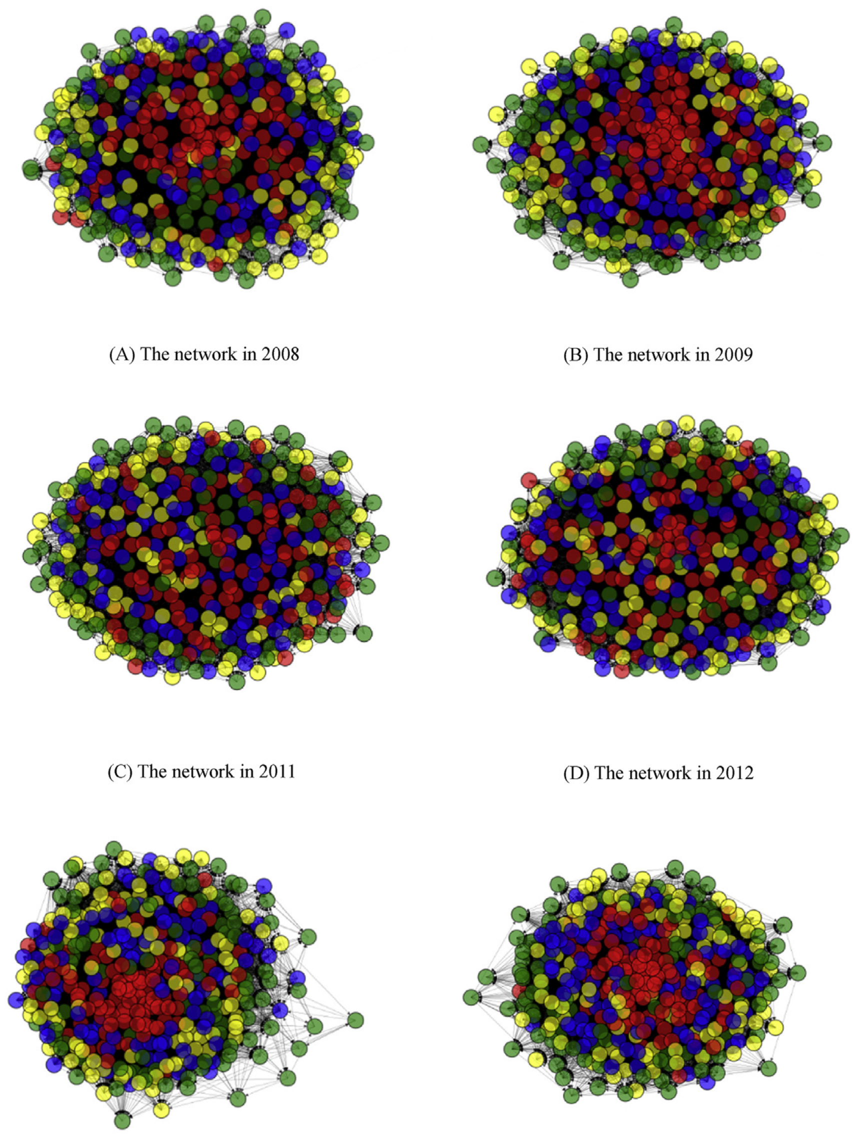


(B) The Q-Q plots of $PRnew_i$ in different years against that in 2015



(C) The Q-Q plots of $PRnew_i$ in different years against that in 2011

Fig. 7. The Q-Q plots of the improved PageRank score of stocks in different year.



(A) The network in 2008

(B) The network in 2009

(C) The network in 2011

(D) The network in 2012

(E) The network in 2015

(F) The network in 2016

Fig. 8. The causal networks illustrated according KK algorithm in different years.
Note: The causal networks are shown based on all the component stocks of CSI 300 Index in our sample. The nodes are illustrated in four different colors representing the ranking quartiles according to their improved PageRank scores in descending order. To be specific, according to the improved PageRank scores, red represents the nodes ranked in the first quartile, blue represents the nodes ranked in the second quartile, yellow represents the nodes ranked in the third quartile, and green represents the nodes ranked in the fourth quartile. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

reports. We adopt the book value of assets at the end of the fiscal year as the total assets used in the variables. The accounting data is sourced from CSMAR database. The annual reports of fiscal year *t* of listed companies are usually published in the first quarter of the next calendar year. We align each company's fiscal year appropriately with the converting fiscal year data to a calendar basis, and then lag annual accounting data by 1 year. We match the variables related to market in year *t* with the variable related to accounting data in year *t-1* to account for data availability.

Table 1 presents the descriptive statistics, the Jarque-Bera test statistic and the Augmented Dickey Fuller (ADF) test for the variables of fundamentals used to investigate the drivers of Chinese stock market comovements. We provide the statistics for all observations for the

Table 1
Descriptive statistics of the variables of risk factors and firms' characteristics.

Statistics	*Beta*	*Alpha*	*Volatility*	*Mvalue*	*BMratio*	*Profitablity*	*Leverage*	*Liquidility*
Mean	1.02	0.00	0.03	16.23	1.48	0.26	0.12	0.04
Max	2.06	0.01	0.08	21.22	21.92	8.26	0.99	0.44
Min	0.08	−0.01	0.01	12.55	0.03	0.01	0.00	−3.86
Std.Deviation	0.27	0.00	0.01	1.07	1.91	0.20	0.11	0.09
Skewness	−0.19	0.05	0.80	0.30	5.12	19.96	2.32	−19.29
Kurtosis	0.09	1.19	0.62	0.72	36.67	693.19	9.16	789.10
Jarque-Bera	26.02	243.09	502.13	148.36	2.46E+05	8.18E+07	1.79E+04	1.06E+08
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
ADF test	−8.25	−9.90	−4.50	−9.46	−11.19	−16.40	−15.32	−14.81
p-value	[-0.01]	[-0.01]	[-0.01]	[-0.01]	[-0.01]	[-0.01]	[-0.01]	[-0.01]

Note: The p-values of the above statistics are reported in square brackets.

sample period, rather than distinguish the cross-sectional and time series data for the individual stocks. The results show that the variables are not symmetric or Gaussian due to the nonzero skewness or excess kurtosis. The Jarque-Bera statistics reveal that all variables are not normally distributed. All variables are stationary, according to the results of the ADF test.

5.2. Panel data logistic regression models with fundamental factors

This study focuses on how the factors related to fundamentals affect the causal comovement between each pair of stocks firstly. To better specify the model, we adopt the dependent variable that describes whether there is a lead-lag relationship in a pair of stocks directly. $Y_{i\to j,t} = 1$, if stock i is a significant Granger cause of stock j in year t; otherwise, $Y_{i\to j,t} = 0$. Our sample contains 377 individual stocks and 11 years of data. We adopt the panel data logistic regression model, which is suitable to depict and model this problem more deeply. We define the independent variables as the comparison of the specific risk factor or firm characteristic of an ordered pair of stocks, i.e., $X_{i\to j,t} = x_{i,t} - x_{j,t}$, where $x_{i,t}$ is a risk factor or firm characteristic for stock i at time t, $x \in \{Beta, Alpha, Volatility, Mvalue, BMratio, Leverage, Liquidity, Profitability\}$For each risk factor or firm characteristic at each time, the independent variable is recorded as $X \in \{BETA, ALPHA, VOL, MV, BM, LEV, LIQ, PROFIT\}$.

When we investigate how the comovement structure of the stock market can be justified by the fundamental sources of risk and other firm characteristics, we adopt a panel data logistic regression model with the fixed effects of the individuals and time. One advantage of a panel data model is the possibility of consistent estimations that allow for unobserved individual heterogeneity and time heterogeneity. We consider individual heterogeneity and time heterogeneity simultaneously, because some unobservable factors determine the comovement structure of the stock market beyond the factors mentioned. We select fixed effect because the strong assumption of a random effect is difficult to satisfy. A random effect requires unobserved heterogeneity to be distributed independently of the explanatory variables. In our problem, heterogeneity and explanatory variables are correlated. To make the specification and estimation as simple as possible, we adopt a year's dummy to address the time effect. The model is specified as in Eq. (8).

$$\ln\left(\frac{p_{i\to j,t}}{1-p_{i\to j,t}}\right) = \beta_1 BETA_{i\to j,t} + \beta_2 ALPHA_{i\to j,t} + \beta_3 VOL_{i\to j,t} + \beta_4 MV_{i\to j,t} + \beta_5 BM_{i\to j,t} + \beta_6 LEV_{i\to j,t} + \beta_7 LIQ_{i\to j,t} + \beta_8 PROFIT_{i\to j,t} + \sum_{t=1}^{10} \rho_t D_{2006+t} + \lambda_{i\to j} + \varepsilon_{ij,t}, \tag{8}$$

where $p_{i\to j,t} = P(Y_{i\to j,t} = 1)$, $\lambda_{i\to j}$ represents the individual effect, and the dummy variables D_{2006+t} characterizes the time effect.

As has been argued in the literature, the maximum likelihood estimation (MLE) method fails to estimate coefficients consistently in nonlinear panel fixed effect models because of the incidental parameters problem, which occurs when the length of the panel is small and fixed. Specifically, in this case, the estimations of the heterogeneity parameters $\lambda_{i\to j}$ are inconsistent. The MLEs for $\lambda_{i\to j}$ and β are not independent of each other for the models with binary outcomes. When T is fixed, the inconsistency of $\lambda_{i\to j}$ is transmitted into the MLE for β. Hence, even if N tends to infinity, the MLE of β remains inconsistent. To obtain a consistent estimation of β, we must eliminate the individual effect $\lambda_{i\to j}$. Unfortunately, contrary to the linear panel data model in which the individual effects $\lambda_{i\to j}$ can be eliminated through a linear transformation such as the first difference, in general, no simple transformation exists to eliminate the incidental parameters from a nonlinear panel data model. To address the problem of the inconsistent estimation of β, we adopt the conditional MLE proposed by Chamberlain (1980). A sufficient statistic for the individual effect is the sum of the positive outcomes. Constructing a new set of probabilities conditional on this statistic forms the basis of a conditional MLE criterion. We maximize the conditional MLE iteratively to obtain consistent estimations of β.

We also investigate whether industry effects exist in the causal comovement of stocks by adding the variable representing whether the pair of stocks belongs to the same industry sector into the model. $INDUSTRY_{ij} = 1$, if stock i and stock j belong to the same industry sector; otherwise, $INDUSTRY_{ij} = 0$. Because the industry attribute is intrinsic special characteristic of a firm, it does not change over time. Thus, when we discuss industry effects, we adopt the logistic regression model with pool data specified in Eq. (9) as follows:

$$\ln\left(\frac{p_{i\to j,t}}{1-p_{i\to j,t}}\right) = \beta_0 + \beta_1 INDUSTRY_{ij} + \beta_2 BETA_{i\to j,t} + \beta_3 ALPHA_{i\to j,t} + \beta_4 VOL_{i\to j,t} + \beta_5 MV_{i\to j,t} + \beta_6 BM_{i\to j,t} + \beta_7 LEV_{i\to j,t} + \beta_8 LIQ_{i\to j,t} + \beta_9 PROFIT_{i\to j,t} + \varepsilon_{ij,t}, \tag{9}$$

where β_1 measures the industry effect on the comovement structure of the stock market. The logistic regression model by using the MLE method.

The estimated results of Eq. (8) under different settings of the independent variables are shown in Table 2. The first to the eighth columns of Table 2 show the results using different independent variables after controlling for the individual effects and time effect. The last column of Table 2 shows the estimation results of Eq. (8) with all independent variables after controlling for the individual effects and time effect. The model in last column of Table 2 are the best model according to the log likelihood and McFadden's pseudo R-squared. From the results, we confirm that the time effect exists in the comovement of the Chinese stock market, because all year's dummy variables are significant no matter how we change the independent variables in Eq. (8). We also find that the coefficients of the dummy variables representing the year effects are negative from 2007 to 2013 and are positive from 2014 to 2016. Compared with the coefficients of these years' dummy variables, the coefficients of D_{2015} and D_{2016} are nearly 0.45, larger than the coefficient of D_{2014}, which is nearly 0.07. And coefficients of year's dummy variable with a slight upward trend from 2007 to 2009 and an obvious upward trend from 2011 to 2015 with a sudden increase in 2015 reveal the degree of causal comovement in Chinese market increase during these two periods. The first period corresponds to the global financial crisis, while the period from 2015 to 2016 corresponds to the fact that the Chinese stock market suffered from a stock market crash. The stock market crash can be regarded as a realization of systemic risk. Shocks in one stock can be transmitted to other stocks quickly at this time, causing the prices of almost all stocks in the stock market to fall simultaneously. The increasing degree of comovement illustrated by the coefficients of year's dummy variables is consistent with the topological properties of the causal networks, such as *DCN*, *GCC*, and *ADCN*. These results confirm the usefulness of the topological properties of the causal networks.

After considering the individual effects and time effect, we can analyze the effect of risk factors and firm characteristics on the comovement of the stock market. First, we find that the estimation results when we use the explanatory variables one by one and when we use all the variables are similar. The original effect of the independent variables on the dependent variable remains almost unchanged after adding the other independent variables, and this result illustrates the low degree of collinearity between the variables to some extent. The variables $BETA_{i\to j,t}$, $VOL_{i\to j,t}$, $BM_{i\to j,t}$, $LIQ_{i\to j,t}$ and $PROFIT_{i\to j,t}$ are significant at the 1% significance levels, no matter whether adding other explanatory variables. The coefficients of $BETA_{i\to j,t}$, $BM_{i\to j,t}$, $LIQ_{i\to j,t}$ and $PROFIT_{i\to j,t}$ are positive, while the coefficients of $VOL_{i\to j,t}$ are negative. Thus, we conclude that stocks with a high systematic risk beta coefficient, book-to-market ratio, liquidity and profitability, but low volatility of prices, tend to Granger cause other stocks' price changes.

Table 2
Estimation results of the panel data logistic regression model shown as Eq. (8).

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$BETA_{i\to j,t}$	0.03								0.04
	(3.68)***								(4.65)***
$ALPHA_{i\to j,t}$		0.13							−0.89
		(0.11)							(−0.73)
$VOL_{i\to j,t}$			−1.39						−1.86
			(−5.94)***						(−7.31)***
$MV_{i\to j,t}$				2.33E-03					3.98E-03
				(0.67)					(1.03)
$BM_{i\to j,t}$					0.02				0.02
					(10.10)***				(11.6)***
$LEV_{i\to j,t}$						−0.02			0.02
						(−1.81)*			(1.48)
$LIQ_{i\to j,t}$							0.13		0.09
							(5.38)***		(3.87)***
$PROFIT_{i\to j,t}$								0.28	0.30
								(14.1)***	(13.5)***
D_{2007}	−0.34	−0.34	−0.34	−0.34	−0.34	−0.34	−0.34	−0.34	−0.34
	(−30.2)***	(−30.2)***	(−30.2)***	(−30.2)***	(−30.2)***	(−30.2)***	(−30.2)***	(−30.4)***	(−30.5)***
D_{2008}	−0.24	−0.24	−0.24	−0.24	−0.24	−0.24	−0.24	−0.24	−0.24
	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.8)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.8)***
D_{2009}	−0.18	−0.18	−0.18	−0.18	−0.18	−0.18	−0.18	−0.18	−0.18
	(−17.0)***	(−17.1)***	(−17.1)***	(−17.1)***	(−17.0)***	(−17.1)***	(−17.1)***	(−17.1)***	(−17.0)***
D_{2010}	−0.23	−0.23	−0.23	−0.23	−0.24	−0.23	−0.23	−0.23	−0.24
	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.7)***	(−21.8)***
D_{2011}	−0.63	−0.64	−0.64	−0.64	−0.63	−0.64	−0.64	−0.64	−0.64
	(−54.1)***	(−54.1)***	(−54.2)***	(−54.1)***	(−54.1)***	(−54.1)***	(−54.2)***	(−54.2)***	(−54.2)***
D_{2012}	−0.27	−0.27	−0.27	−0.27	−0.27	−0.27	−0.27	−0.27	−0.27
	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***	(−25.1)***
D_{2013}	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25	−0.25
	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***	(−23.0)***
D_{2014}	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
	(7.19)***	(7.19)***	(7.16)***	(7.18)***	(7.24)***	(7.19)***	(7.18)***	(7.19)***	(7.20)***
D_{2015}	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	(46.3)***	(46.4)***	(46.4)***	(46.4)***	(46.4)***	(46.4)***	(46.4)***	(46.4)***	(46.3)***
D_{2016}	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
	(45.7)***	(45.7)***	(45.7)***	(45.7)***	(45.7)***	(45.7)***	(45.7)***	(45.7)***	(45.7)***
Log-likelihood	−438270	−438277	−438256	−438277	−438226	−438276	−438263	−438177	−438069
LR χ^2 test	21667	21653	21688	21654	21755	21656	21682	21853	22290
p-value	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.00]***
Pseudo R^2	0.0241	0.0241	0.0241	0.0241	0.0242	0.0241	0.0241	0.0243	0.0246

Note: The dependent variable, whether there is a lead-lag relationship in a pair of stocks, has a binary outcome taking only the values 0 and 1. $Y_{i\to j,t} = 1$, if stock i is a significant Granger cause of stock j in year t; otherwise, $Y_{i\to j,t} = 0$. $p_{i\to j,t} = P(Y_{i\to j,t} = 1)$. The independent variable $X_{i\to j,t} = x_{i,t} - x_{j,t}$, where $x_{i,t}$ is the specific risk factor or firm characteristic of stock i at time t, $x \in \{Beta, \ Alpha, \ Volatility, \ Mvalue, \ BMratio, \ Leverage, \ Liquidity, \ Profitability\}$. $X \in \{BETA, \ ALPHA, \ VOL, \ MV, \ BM, \ LEV, \ LIQ, \ PROFIT\}$. The dummy variables D_{2006+t} characterize the time effect. To save space, we do not provide the estimation of $\lambda_{i\to j}$. The t-statistics are reported in parentheses under the estimated parameters. ***, **, and * indicate that the t-statistics are significant at the 1%, 5%, and 10% levels, respectively. The p-values of LR Chi-square test are shown in square brackets under the Chi-square statistics. Pseudo R^2 refers to McFadden's pseudo R-squared.

Next, we estimate Eq. (9) by using different control variables to investigate in more depth. Table 3 presents the estimation results of the logistic regression using all observations as pool data. Each column presents a specific model's results. The first column in Table 3 shows the result of the model that includes only $INDUSTRY_{ij}$ as the explanatory variable. Columns 2 to 9 in Table 3 correspond to the results of models including $INDUSTRY_{ij}$ and one variable of a risk factor or firm characteristic one by one as control variables. Then, we provide the estimated model results including all risk factors and firm characteristics in Eq. (9) in Column 10. We find the model shown in the last column is the best model according to the log-likelihood and McFadden's pseudo R-squared. We also add year's dummy variables into Eq. (9) and get similar results as Table 2, which are not reported here to save space. From the estimation results, we find that no matter how we change the control variables in Eq. (9), the coefficient of the variable $INDUSTRY_{ij}$ is insignificant. We reject the hypothesis that the causal comovement is attributable to the industry effect. Our results further support the literature that has argued against the industry effect hypothesis, such as Pindyck and Rotemberg (1993), and Jung and Chang (2016). Pindyck and Rotemberg (1993) reject the hypothesis that comovements in the individual stock prices are justified by economic fundamentals and explain the results by the market segmentation caused by the institutional investors. Jung and Chang (2016) study the market structure by using the partial correlation analysis and compare the members of the clusters with the Global Industry Classification Standard (GICS) and find that most clusters are a mix of multiple sectors, which is inconsistent with the hypothesis.

Most of estimation results of other independent variables in Eq. (8) and Eq. (9) are similar, except MV, LEV becomes significant and the coefficient of BM becomes negative in Eq. (9). And we confirm the robustness of variables $BETA_{i\to j,t}$, $VOL_{i\to j,t}$, $LIQ_{i\to j,t}$ and $PROFIT_{i\to j,t}$ on causal comovement of individual stocks. The difference of two type of models may come from the individual effect specification and estimation methods or may come from the changes of comovement structure. To address the issue about comovement structure, we re-estimate Eq. (9) using observations in each year and investigate the effects of industry, risk factors and firm characteristics on the causal comovement among stocks. Table 4 shows the results. The coefficients of the industry variable $INDUSTRY_{ij}$ are insignificant in most years and only significant in 2008, 2014 and 2016, at 5%, 1% and 5% significance level, respectively. We conclude that the comovement is not attributable to industry effects, in general, based the results above. However, in certain years around financial crises and stock market crashes, such as financial crisis around

Table 3
Estimation results of the logistic regression model shown as Eq. (9) by using all observations as pool data.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
INDUSTRY	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
	(1.12)	(1.12)	(1.12)	(1.14)	(1.12)	(1.42)	(1.22)	(1.12)	(1.21)	(1.51)
BETA		0.03								0.04
		(4.31)***								(6.21)***
ALPHA			−3.81							0.36
			(−3.58)***							(0.32)
VOL				−1.07						−1.93
				(−5.45)***						(−8.85)***
MV					0.01					0.01
					(7.53)***					(6.47)***
BM						−0.01				−0.01
						(−15.5)***				(−10.5)***
LEV							−0.14			−0.06
							(−18.5)***			(−6.57)***
LIQ								0.16		0.10
								(11.3)***		(6.79)***
PROFIT									0.37	0.22
									(22.1)***	(11.1)***
β_0	−1.63	−1.63	−1.63	−1.63	−1.63	−1.63	−1.63	−1.63	−1.63	−1.63
	(−736)***	(−736)***	(−736)***	(−736)***	(−736)***	(−727)***	(−730)***	(−736)***	(−735)***	(−722)***
Log-likelihood	−694545	−694536	−694539	−694530	−694517	−682958	−694376	−679932	−693121	−669635
LR χ^2 test	1.24	9.92	7.03	15.45	29.01	119.94	170.06	64.51	244.33	87.04
p-value	[0.2654]	[0.0070]***	[0.0297]**	[0.0004]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.0000]***
Pseudo R^2	8.9289E-07	1.4288E-05	1.0126E-05	2.2248E-05	4.1763E-05	0.0167	0.0002	0.0210	0.0021	0.0359

Note: Similar to the note in Table 2. $INDUSTRY_{ij}=1$, if stock i and stock j belong to the same industry sector; otherwise, $INDUSTRY_{ij}=0$.

2008, and Chinese market crash around 2015 and 2016, the pair of stocks belonging to the same industry tends to co-move in returns. Moreover, besides the changes of industry effect on comovement, we find the driving factors of comovement in each year change a little. This may imply the changes of comovement structure.

Thus, we conclude that stocks with a high systematic risk beta coefficient, book-to-market ratio, liquidity and profitability, but low volatility of prices, tend to Granger cause other stocks' price changes. And the causal comovement of stock returns is not attributable to industry effects, in general, but can be attributable to industry effect around crisis period. We also reveal the driving factors of comovement change a little over time, which may imply the changes of comovement structure. These results support the fundamentals-based view of comovement, because the fundamental risk factors and firm characteristics rather than industry property can affect the causal comovement of stock returns. Our study supports many existing studies on pricing factors, such as Fama and French (1996) and Campbell et al. (2008) among others.

5.3. Logistic regression models with factors of both fundamentals and sentiment-based

In this section, we explore whether the excess causal comovement of stock returns that are not totally explained by fundamentals can be explained by the variables related to sentiment-based view. Similar to the ideas in BSW (2005) and Boyer (2011), we bring several variables related to the inclusion and exclusion from the CSI 300 Index into the panel data logistic regression model, meanwhile we maintain the significant fundamental factors of Eq. (8) as shown in Table 2 in the model. CSI 300 Index consists of the 300 largest and most liquid Chinese A-share stocks. This index is only one of the market index labels in Chinese stock market, the constituents of which are determined by the committee. Therefore, there are other information in excess of fundamentals in the index labels.

Define a dummy variable $D_{i,t}^{CSI}$ to represent the stock i as the component stock of CSI 300 Index at year t. $D_{i,t}^{CSI}=1$, if stock i in year t belongs to the component stocks of CSI 300 Index; otherwise, $D_{i,t}^{CSI}=0$. Based on this dummy variable, we define four types of stocks: continuously included stocks in the index, newly included stocks, newly eliminated stocks, and continuously excluded stocks. Define continuously included stock as $CSI_{i,t}$. $CSI_{i,t}=1$, if $D_{i,t-1}^{CSI}=1$ and $D_{i,t}^{CSI}=1$; otherwise, $CSI_{i,t}=0$. Define newly included stocks as $Enter_{i,t}^{CSI}$. $Enter_{i,t}^{CSI}=1$, if $D_{i,t-1}^{CSI}=0$ and $D_{i,t}^{CSI}=1$; otherwise $Enter_{i,t}^{CSI}=0$. Define newly eliminated stocks as $Eliminate_{i,t}^{CSI}$. $Eliminate_{i,t}^{CSI}=1$, if $D_{i,t-1}^{CSI}=1$ and $D_{i,t}^{CSI}=0$; otherwise, $Eliminate_{i,t}^{CSI}=0$. The stock is continuously excluded from the index, if $D_{i,t-1}^{CSI}=0$ and $D_{i,t}^{CSI}=0$. The CSI 300 Index constituents are adjusted twice a year, at the beginning and middle of the year. No matter the stock is a constituent of the CSI 300 Index at the beginning of the year or at the middle of the year, the stock is considered as a constituent of CSI 300 Index during the year in our empirical study. We can discuss the inclusion in the market index from another way. For the directed stock pair $i\rightarrow j$, according to included or nonincluded in the index of each stock, $D_{i,t}^{CSI}$ and $D_{j,t}^{CSI}$, there are four types. Then, we define three dummy variables to indicate the above four categories of the directed stock pairs, $D_{i\rightarrow j,t}^{CSI\rightarrow CSI}$, $D_{i\rightarrow j,t}^{nonCSI\rightarrow CSI}$ and $D_{i\rightarrow j,t}^{CSI\rightarrow nonCSI}$, representing the stock i and the stock j both are included, only the stock j is included, or only stock i is included in the CSI 300 Index in year t. $D_{i\rightarrow j,t}^{CSI\rightarrow CSI}=1$, if $D_{i,t}^{CSI}=1$ and $D_{j,t}^{CSI}=1$; otherwise $D_{i\rightarrow j,t}^{CSI\rightarrow CSI}=0$. $D_{i\rightarrow j,t}^{nonCSI\rightarrow CSI}=1$, if $D_{i,t}^{CSI}=0$ and $D_{j,t}^{CSI}=1$; otherwise $D_{i\rightarrow j,t}^{nonCSI\rightarrow CSI}=0$. $D_{i\rightarrow j,t}^{CSI\rightarrow nonCSI}=1$, if $D_{i,t}^{CSI}=1$ and $D_{j,t}^{CSI}=0$; otherwise $D_{i\rightarrow j,t}^{CSI\rightarrow nonCSI}=0$.

After controlling for the significant fundamental factors that have been shown in Table 2, we bring in the inclusion or exclusion properties of stock i, stock j or both stocks for the directed stock pair $i\rightarrow j$, and the model can be written as Eqs. (10)–(12), respectively. The empirical results are shown in Table 5.

$$\ln\left(\frac{p_{i\rightarrow j,t}}{1-p_{i\rightarrow j,t}}\right)=\beta_1 Enter_{i,t}^{CSI}+\beta_2 Eliminate_{i,t}^{CSI}+\beta_3 CSI_{i,t}+\beta_4 BETA_{i\rightarrow j,t}+\beta_5 VOL_{i\rightarrow j,t}+\beta_6 BM_{i\rightarrow j,t}+\beta_7 LIQ_{i\rightarrow j,t}+\beta_8 PROFIT_{i\rightarrow j,t}+\sum_{t=1}^{9}\rho_t D_{2007+t}+\lambda_{i\rightarrow j}+\varepsilon_{ij,t}, \tag{10}$$

Table 4
The estimation results of logistic regression model shown as Eq. (9) using observations in each year.

Variable	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
INDUSTRY	−0.02	0.05	−0.09	−3.90E-04	0.05	−0.05	0.03	0.02	0.21	0.04	−0.07
	(−0.57)	(−1.38)	(−2.31)**	(−0.01)	(−1.47)	(−1.08)	(−0.73)	(−0.47)	(6.98)***	(−1.24)	(−2.25)**
BETA	0.04	0.01	−0.15	−0.48	−0.07	0.14	−0.14	0.45	0.16	0.11	−0.07
	(1.77)*	(−0.28)	(−4.27)***	(−18.27)***	(−2.84)***	(4.63)***	(−6.35)***	(19.40)***	(7.17)***	(5.89)***	(−3.72)***
ALPHA	−1.89	11.59	43.86	12.74	−18.04	−11.84	−30.47	−31.6	4.33	−1.02	−58.7
	(−0.55)	(3.00)***	(10.59)***	(3.18)***	(−4.94)***	(−1.84)*	(−5.54)***	(−9.09)***	(−1.00)	(−0.34)	(−12.16)***
VOL	0.35	7.24	13.66	−3.53	9.10	−1.26	4.83	5.81	0.27	−17.36	−16.18
	(−0.47)	(7.48)***	(15.90)***	(−4.14)***	(12.05)***	(−1.34)	(5.39)***	(7.65)***	(−0.40)	(−32.13)***	(−23.44)***
MV	0.02	0.11	−0.10	0.03	0.04	−0.01	−0.11	0.07	3.05E-03	0.04	0.04
	(−3.56)	(17.21)***	(−18.08)***	(5.03)***	(6.82)***	(−1.48)	(−17.61)***	(10.82)***	(−0.50)	(7.49)***	(6.19)***
BM	−0.03	−0.02	−0.22	−1.92E-03	−0.06	−0.07	−0.04	−0.01	0.03	4.30E-03	−0.01
	(−7.59)***	(−4.02)***	(−20.22)***	(−0.56)	(−11.10)***	(−17.30)***	(−11.89)***	(−3.37)***	(10.33)***	(−1.31)	(−4.18)***
LEV	0.16	−0.26	−0.36	−0.42	−0.07	0.07	0.06	0.08	−0.34	−0.30	0.19
	(2.53)**	(−3.91)***	(−9.71)***	(−11.60)***	(−2.43)**	(−1.18)	(1.81)*	(4.65)***	(−5.27)***	(−5.51)***	(3.41)***
LIQ	0.27	0.17	0.04	0.48	0.40	−0.35	0.08	−0.02	−0.64	−0.12	−0.42
	(4.66)***	(2.86)***	(−0.86)	(9.40)***	(8.36)***	(−6.61)***	(−1.44)	(−0.39)	(−11.04)***	(−2.26)**	(−8.51)***
PROFIT	−0.09	0.02	−0.74	0.37	−0.08	0.38	0.4	1.63E-03	−0.68	2.88	−0.58
	(−1.23)	(−0.28)	(−7.62)***	(4.85)***	(−1.08)	(3.11)***	(3.42)***	(−0.01)	(−8.14)***	(28.33)***	(−5.95)***
β_0	−1.56	−1.91	−1.83	−1.76	−1.81	−2.20	−1.85	−1.82	−1.50	−1.12	−1.11
	(−200.24)***	(−224.44)***	(−222.25)***	(−225.18)***	(−228.43)***	(−239.36)***	(−229.53)***	(−229.04)***	(−211.46)***	(−171.21)***	(−173.04)***
Log-likelihood	−55714	−50292	−54063	−58660	−57229	−45569	−55966	−56991	−67255	−75044	−76923
Pseudo R^2	0.1417	0.0768	0.0661	0.0186	0.0153	0.0185	0.0231	0.0128	0.0095	0.0572	0.0311

Note: Similar to the note in Table 3.

Table 5
The results of the panel data logistic regression models shown as Eqs. (10)–(12).

Variable	(1)	(2)	(3)	(4)
$BETA_{i\to j,t}$	0.03 (3.95)***	0.03 (3.88)***	0.03 (3.16)***	0.02 (2.80)***
$VOL_{i\to j,t}$	−1.86 (−7.29) ***	−1.91 (−7.48) ***	−1.72 (−6.72) ***	−1.70 (−6.65) ***
$BM_{i\to j,t}$	0.03 (13.12) ***	0.03 (13.58) ***	0.03 (13.82)***	0.03 (14.29) ***
$LIQ_{i\to j,t}$	0.08 (2.97)***	0.07 (2.84)***	0.06 (2.53)**	0.06 (2.43)**
$PROFIT_{i\to j,t}$	0.30 (14.34) ***	0.30 (14.08) ***	0.29 (13.57)***	0.28 (13.39) ***
$CSI_{i,t}$		0.01 (−0.70)		
$Eliminate^{CSI}_{i,t}$		−0.04 (−4.04) ***		
$Enter^{CSI}_{i,t}$		0.04 (3.26)***		
$CSI_{j,t}$			−0.07 (−9.12) ***	
$Eliminate^{CSI}_{j,t}$			−0.05 (−4.86) ***	
$Enter^{CSI}_{j,t}$			−0.13 (−10.17) ***	
$D^{CSI\to CSI}_{i\to j,t}$				−0.05 (−4.58) ***
$D^{nonCSI\to CSI}_{i\to j,t}$				−0.07 (−7.49) ***
$D^{CSI\to nonCSI}_{i\to j,t}$				0.03 (2.89)***
Controlled by individual effect and time effect	Yes	Yes	Yes	Yes
Log-likelihood	−389208	−389190	−389135	−389153
LR χ^2 test	21998	22033	22142	22108
p-value	[0.00]***	[0.00]***	[0.00]***	[0.00]***
Pseudo R^2	0.02748	0.02752	0.02766	0.02762

Note: Similar to the note of Table 2.

$$\ln\left(\frac{p_{i\to j,t}}{1-p_{i\to j,t}}\right) = \beta_1 Enter^{CSI}_{j,t} + \beta_2 Eliminate^{CSI}_{j,t} + \beta_3 CSI_{j,t} + \beta_4 BETA_{i\to j,t} + \beta_5 VOL_{i\to j,t} + \beta_6 BM_{i\to j,t} + \beta_7 LIQ_{i\to j,t} + \beta_8 PROFIT_{i\to j,t} + \sum_{t=1}^{9}\rho_t D_{2007+t} + \lambda_{i\to j} + \varepsilon_{ij,t}, \quad (11)$$

$$\ln\left(\frac{p_{i\to j,t}}{1-p_{i\to j,t}}\right) = \beta_1 D^{CSI\to CSI}_{i\to j,t} + \beta_2 D^{nonCSI\to CSI}_{i\to j,t} + \beta_3 D^{CSI\to nonCSI}_{i\to j,t} + \beta_4 BETA_{i\to j,t} + \beta_5 VOL_{i\to j,t} + \beta_6 BM_{i\to j,t} + \beta_7 LIQ_{i\to j,t} + \beta_8 PROFIT_{i\to j,t} + \sum_{t=1}^{9}\rho_t D_{2007+t} + \lambda_{i\to j} + \varepsilon_{ij,t}, \quad (12)$$

Throughout the results shown in column (2) of Table 5, we can find that leading ability of newly included stocks in the market index is improved, while the ability of newly excluded stocks is weakened. The results shown in column (3) of Table 5 also provide similar results from the other side. Compared to non-component stocks, the component stocks of market index cannot be affected and led by other stocks, and stocks that have been newly included in the index have changed the most in the causal comovement with others. The results shown in column (4) of Table 5 provide new insights about the causal comovement between individual stocks. Most of the causal comovement occurs in the situation that the component stocks lead non-component stocks, while non-component stocks seldom lead component stocks. Even in the group of component stocks, the causal comovement also rarely appears. This reflects that the component stocks in the market index can reveal the information more quickly than non-component stocks, which lead to the causal comovement between component stocks and non-component stocks. These results provide direct evidence about the information diffusion view of sentiment-based theories of comovement. Moreover, if the component stocks of the market index are traded by investors, the correlation of these stocks should increase, while the causal comovement should decrease. Our results are consistent with the latter. That means our results support the category (habitat) view of sentiment-based theories of comovement.

6. Determinants of comovement structure of stock market

6.1. Discussion about the fundamental factors

We investigate the determinants of the comovement structure by using the improved PageRank score of individual stocks, $PRnew_i$. The improved PageRank algorithm can also measure information discovery power of stocks in the stock causal networks, through which we can find the systemically important stocks that can lead other stocks in the stock market network. For a given network, the stocks with greater importance have a higher improved PageRank score. The greater importance of the stocks is, the stronger is the ability of discovering information. We note that the improved PageRank score can be compared only within the same stock market network. Specifically, if the improved PageRank score of stock A in the causal network in year t_1 is larger than the PageRank score of stock B in the causal network in year t_2, we cannot conclude that stock A in year t_1 is more important than stock B in year t_2. To make the importance of nodes in different yearly stock market network comparable, we rank all stocks from 1 to 377 according to their improved PageRank score in the corresponding causal network in each year. We regard the rankings by the improved PageRank score as the dependent variable, $RPRnew_{i,t}$. To investigate the determinants of the importance of individual stocks, we construct a panel data model based on $RPRnew_{i,t}$.

First, we examine the effect of fundamentals. To determine whether the model should include the fixed individual effect and time effect, we examine a panel data model with the fixed individual effect and the time effect and observe the significance of these effects. To make the specification and estimation as simple as possible, we adopt the year dummies to address the time effect. There are two ways to specify the model. One is using the original values of the factors as the explanatory variables in the model, as shown in the following Eq. (13). Another one is using the rankings of each stock according to different factors as the explanatory variables in the model, as shown in the following Eq. (14).

$$RPRnew_{i,t} = c_0 + c_1 Beta_{i,t} + c_2 Alpha_{i,t} + c_3 Volatility_{i,t} + c_4 Mvalue_{i,t} + c_5 BMratio_{i,t} + c_6 Leverage_{i,t} + c_7 Liquidlity_{i,t} + c_8 Profitability_{i,t} + \sum_{t=1}^{10}\rho_t D_{2006+t} + d_i + \varepsilon_{i,t}, \quad (13)$$

$$RPRnew_{i,t} = c_0 + c_1 RBeta_{i,t} + c_2 RAlpha_{i,t} + c_3 RVolatility_{i,t} + c_4 RMvalue_{i,t} + c_5 RBMratio_{i,t} + c_6 RLeverage_{i,t} + c_7 RLiquidlity_{i,t} + c_8 RProfitability_{i,t} + \sum_{t=1}^{10}\rho_t D_{2006+t} + d_i + \varepsilon_{i,t}, \quad (14)$$

where $RPRnew_{i,t}$ is the ranking of stock i according to the improve PageRank score at time t, $RBeta_{i,t}$, $RAlpha_{i,t}$, $RVolatility_{i,t}$, $RMvalue_{i,t}$, $RBMratio_{i,t}$, $RLeverage_{i,t}$, $RLiquidlity_{i,t}$, and $RProfitability_{i,t}$ represent the rankings of stock i at time t according to *Beta, Alpha, Volatility, Mvalue, BMratio, Leverage, Liquidity,* and *Profitability* of stock i at time t, respectively. In Eq. (13) and Eq. (14), d_i characterizes the individual effect, and the dummy variables D_{2006+t} characterize the time effect. The estimations of Eqs. (13) and (14) are based on a least-square dummy variable estimator, which is obtained by an ordinary least-squares analysis with regarding heterogeneity as dummy variables. Add each of the explanatory variables into the model, controlling for individual fixed effects and time fixed effect. Then, include all the explanatory variables after controlling for individual fixed effects and time fixed effect. In addition to the coefficient estimation results and t-statistics of the coefficient, we also examine the F-test results for the null hypothesis that the individual effect does not exist, i.e., $d_i = 0, \quad \forall i$. According to the estimation results, we find there is no individual effect or time effect in the data. To save space, we do not report the detailed results for Eq. (13) and Eq. (14).[4] For the estimation results of Eq. (13), we observe that the p-values of F-tests are greater than 0.10, and all the dummy variables are insignificant, no matter how we change the explanatory variables in the regression. When we change the format of the independent variables from the original values to the rankings, we observe that the individual and time effects do not exist either for Eq. (14). There is no individual and time effect for the improved PageRank measures of individual stocks, which is totally different from the results of the model for the comovement between individual stocks in Section 5. This is because the improved PageRank measure provide information about comovement structure.

Throughout the above analysis, we exclude the individual fixed effects and time fixed effect in Eqs. (13) and (14) and obtain the more suitable specification for the analysis of the importance of individual stocks in the market, as shown in Eqs. (15) and (16) as follows.

$$\begin{aligned} RPRnew_{i,t} &= c_0 + c_1 Beta_{i,t} + c_2 Alpha_{i,t} + c_3 Volatility_{i,t} + c_4 Mvalue_{i,t} \\ &\quad + c_5 BMratio_{i,t} + c_6 Leverage_{i,t} + c_7 Liquidlity_{i,t} + c_8 Profitability_{i,t} \\ &\quad + \varepsilon_{i,t}\ . \end{aligned} \tag{15}$$

$$\begin{aligned} RPRnew_{i,t} &= c_0 + c_1 RBeta_{i,t} + c_2 RAlpha_{i,t} + c_3 RVolatility_{i,t} + c_4 RMvalue_{i,t} \\ &\quad + c_5 RBMratio_{i,t} + c_6 RLeverage_{i,t} + c_7 RLiquidlity_{i,t} \\ &\quad + c_8 RProfitability_{i,t} + \varepsilon_{it}\ . \end{aligned} \tag{16}$$

We show the estimation results of Eqs. (15) and (16) in Tables 6 and 7, respectively.

The estimation results of the first 9 columns of Eq. (15) are obtained based on the whole sample. When we examine the variables separately, we find that only excess return (*Alpha*) and market value (*Mvalue*) are significant at the 1% significance level, while profitability (*Profitability*) is significant at the 10% level. When we add all the variables into the model, we still find that the excess return alpha is significant at the 1% level, market value is significant at the 5% level, and profitability is significant at the 10% level. The estimation results of Eq. (16) based on the whole sample is quite similar to that of Eq. (15). When we consider the explanatory variable separately, we find that the excess return alpha and market value are significant at the 1% level, while profitability is significant at the 5% level. When considering all the variables, the significant variables in Eq. (16) are excess return alpha and market value. The sign of the coefficients of the significant variables in the two models are opposite. This result is understandable because of the difference in the format of the explanatory variables. Taking the market value as an example, the higher the *Mvalue* is, the lower is the *RMvalue,* because the high values of the independent variable are accompanied by low rankings.

[4] The estimation results of Eq. (13) and Eq. (14) are not reported in the paper to save space. The results are available upon request.

Table 6
The results of determinants of the importance of stocks using the model in Eq. (15).

Variable	Estimations using the whole sample									Normal periods	Crisis periods
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Beta	−4.59								−10.25	−10.54	−1.40
	(−0.72)								(−1.54)	(−1.17)	(−0.14)
Alpha		−4285							−4346	−7063.94	−2133.49
		(−3.83)***							(−3.79)***	(−4.50)***	(−1.19)
Volatility			171.54						238.62	−856.09	956.04
			(1.10)						(1.45)	(−2.92)***	(3.76)***
Mvalue				−5.32					−4.04	−0.74	−5.37
				(−3.34)***					(−2.37)**	(−0.30)	(−2.14)**
BMratio					0.06				0.70	−0.86	2.92
					(0.06)				(0.71)	(−0.71)	(1.59)
Leverage						0.86			−5.64	−4.71	−10.49
						(0.10)			(−0.57)	(−0.41)	(−0.50)
Liquidity							−3.37		6.21	−1.62	21.17
							(−0.21)		(0.38)	(−0.07)	(0.85)
Profitability								−35.93	−35.52	−80.63	−23.09
								(−1.94)*	(−1.66)*	(−2.14)**	(−0.81)
c_0	193.51	188.22	183.41	275.15	188.74	188.61	189.24	190.39	257.87	237.40	239.70
	(28.92)***	(110)***	(35.19)***	(10.62)***	(87.60)***	(67.78)***	(73.1)***	(101)***	(9.26)***	(5.75)***	(5.45)***
Adjust R^2	−0.0001	0.0034	5.08E-05	0.0025	−0.0002	−0.0002	−0.0002	0.0007	0.0054	0.0122	0.0091
F-test	0.5238	14.6896	1.2068	11.1415	0.0041	0.01	0.0449	3.7804	3.7416	4.4547	3.1382
p-value	[0.4693]	[0.0001]***	[0.2720]	[0.0009]***	[0.9490]	[0.9204]	[0.8323]	[0.0519]*	[0.0002]***	[2.26E-05]***	[0.0026]***

Note: The estimations of the normal periods are based on the observations in 2006 and that from 2010 to 2014. The estimations of crisis periods are based on the observations from 2007 to 2009, and that from 2015 to 2016. The t-statistics are reported in parentheses under the estimated parameters. ***, **, and * indicate that the t-statistics are significant at the 1%, 5%, and 10% level, respectively. The p-values of the above statistics are reported in squared brackets.

Table 7
The results of determinants of the importance of stocks using the model in Eq. (16).

Variable	Estimations using the whole sample									Normal periods	Crisis periods
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
RBeta	−0.01								0.02	0.03	0.01
	(−0.40)								(−0.83)	(1.12)	(0.49)
RAlpha		0.05							0.06	0.09	0.01
		(3.54)***							(3.62)***	(4.36)***	(0.57)
RVolatility			−0.02						−0.02	0.06	−0.12
			(−0.99)						(−1.02)	(2.29)**	(−4.29)***
RMvalue				0.05					0.05	0.03	0.05
				(3.33)***					(2.74)***	(1.40)	(1.93)*
RBMratio					−0.01				−0.01	−0.03	0.03
					(−0.77)				(−0.42)	(−1.17)	(1.02)
RLeverage						−0.02			−0.01	−0.01	−0.03
						(−1.13)			(−0.70)	(−0.24)	(−0.91)
RLiquidity							−0.02		−0.03	−0.03	−0.03
							(−1.17)		(−1.88)*	(−1.38)	(−1.39)
RProfitability								0.03	0.02	0.01	0.01
								(2.06)**	(−0.79)	(−0.48)	(0.40)
c_0	190.18	178.61	191.92	179.24	191.25	192.32	192.43	182.97	177.03	160.15	200.47
	(56.14)***	(52.81)***	(56.66)***	(52.98)***	(56.46)***	(56.78)***	(56.81)***	(54.04)***	(19.84)***	(13.43)***	(14.93)***
Adjust R^2	0.0002	0.0028	−2.46E-06	0.0024	−9.95E-05	6.80E-05	0.0003	0.0008	0.0053	0.0145	0.0157
F test	0.1619	12.5635	0.9898	11.0798	0.5876	1.2820	1.3627	4.2259	3.7880	5.1518	4.7580
p-value	[0.6874]	[0.0004]***	[0.3199]	[0.0009]***	[0.4434]	[0.2576]	[0.2431]	[0.0399]**	[0.0002]***	[2.18E-06]***	[8.40E-06]***

Note: The estimations of the normal periods are based on the observations in 2006 and that from 2010 to 2014. The estimations of crisis periods are based on the observations from 2007 to 2009, and that from 2015 to 2016. The t-statistics are reported in parentheses under the estimated parameters. ***, **, and * indicate that the t-statistics are significant at the 1%, 5%, and 10% level, respectively. The p-values of the above statistics are reported in squared brackets.

To test the determinants of comovement structure, we separate the whole sample into two parts: the crisis periods sample (including global crisis periods from 2007 to 2009, and the Chinese stock market crash periods from 2015 to 2016) and the normal periods sample. We re-estimate the models shown in Eqs. (15) and (16) and report the results in the column (10)–(11) of Tables 6 and 7, respectively. According to the results of Eq. (15), the determinants of comovement structure in normal periods is excess return (*Alpha*), total risk volatility, and profitability of the firm, while the determinants of comovement structure in crisis periods is total risk volatility and size factor. They are different from the results obtained from the whole sample reported in column (9) in Table 6. When we change the specification of the variables and re-estimated the model in Eq. (16), we get similar results for different periods. As reported in Table 7, the determinants of comovement structure in normal periods is excess return (*Alpha*), total risk volatility, while the determinants of comovement structure in crisis periods is total risk volatility at 1% significant level and size factor at 10% significant level. In normal periods, the most important stocks in the market are that with high excess return, high total risk volatility, high profitability of the firm. However, the most important stocks in the market during crisis periods are that with low total risk volatility and large size. The most interesting result is the opposite of total risk volatility coefficients in different periods.

These results are consistent with much of the literature related to factor models of asset pricing. Moreover, these results above reveal the risk preference of investors changes in different market situation. In normal periods, investors prefer the stocks with high volatility that can provide more opportunities to high returns, while investors prefer low volatility stocks that can defense against risk in crisis periods.

6.2. Discussion about the sentiment-based factors

Next, we investigate whether sentiment-based factors affect the importance of stocks and comovement structure in addition to fundamental factors. Based on the results of Eqs. (15) and (16), we maintain the significant variables of fundamental factors as the control variables and bring the sentiment-based factors into the models. Here, we adopt the dummy variable that represents the component stock of CSI 300 Index, $D_{i,t}^{CSI}$, as a sentiment-based factor. Moreover, we bring in the CITIC industry index to investigate more about the industry effect and the category view of comovement. There are 29 subindustries in CITIC index. We choose the bank sector as the benchmark and define 28 dummy variables, I_i^1 to I_i^{28} to represent the other 28 industries, including electric utilities and public utilities (I^1), petroleum and petrochemicals (I^2), coal mining (I^3), nonferrous metals (I^4), steel (I^5), construction (I^6), construction materials (I^7), real estate (I^8), chemical (I^9), machinery (I^{10}), defense (I^{11}), automobiles (I^{12}), electric power equipment (I^{13}), electronic components (I^{14}), computers (I^{15}), communications (I^{16}), household appliances (I^{17}), light manufacturing (I^{18}), textiles and garments (I^{19}), food and beverages (I^{20}), catering and tourism (I^{21}), retail and trade (I^{22}), health care (I^{23}), transportation (I^{24}), medias (I^{25}), composite (I^{26}), nonbanking financials (I^{27}), agriculture, forestry, animal husbandry & fishery (I^{28}), respectively. The model to study the importance of stocks can be modified as follows:

$$RPRnew_{i,t} = c_0 + \gamma_1 D_{i,t}^{CSI} + \sum_{k=1}^{28} \rho_k I_i^k + \sum_h c_h Fundamentals_{i,t}^h + \varepsilon_{i,t}, \tag{17}$$

$$RPRnew_{i,t} = c_0 + \gamma_1 D_{i,t}^{CSI} + \sum_{k=1}^{28} \rho_k I_i^k + + \sum_h c_h Fundamentals_{i,t}^h + \varepsilon_{i,t}, \tag{18}$$

where *Fundamentals* is the significant fundamental factors obtained from Eqs. (15) and (16) as controlled variables, respectively. The CSI 300 Index comprises the 300 largest and most liquid Chinese A-share stocks,

Table 8
The results of determinants of the importance of stocks using the model in Eq. (17).

Variable	Estimations using the whole sample				
	(1)	(2)	(3)	(4)	(5)
c_0	263.69 (10.11)***	191.09 (84.76)***	191.69 (85.03)***	210.50 (13.92)***	287.74 (8.04)***
Alpha	−3889.50 (−3.46)***		−4828.90 (−4.23)***	−4814.40 (−4.21)***	−3850.12 (−3.41)***
Mvalue	−4.65 (−2.90)***				−4.55 (−2.62)***
D^{CSI}		−2.51 (1.60)	−3.94 (−2.36)**	−4.12 (−2.38)**	
I^1				−34.52 (−2.15)**	−40.76 (−2.47)**
I^2				−43.49 (−2.31)**	−47.95 (−2.52)**
I^3				−28.10 (−1.64)	−35.01 (−1.99)**
I^4				2.32 (0.14)	−5.45 (−0.32)
I^5				−13.10 (−0.79)	−19.42 (−1.14)
I^6				−32.45 (−1.56)	−39.79 (−1.88)*
I^7				−25.06 (−1.30)	−32.03 (−1.63)
I^8				−6.20 (−0.38)	−13.03 (−0.78)
I^9				0.03 (2.1E-03)	−8.54 (−0.49)
I^{10}				−22.52 (−1.34)	−29.41 (−1.70)*
I^{11}				−14.41 (−0.78)	−22.10 (−1.17)
I^{12}				−12.28 (−0.74)	−19.67 (−1.16)
I^{13}				1.33 (0.07)	−5.05 (−0.26)
I^{14}				−9.23 (−0.54)	−16.99 (−0.97)
I^{15}				−24.73 (−1.34)	−32.42 (−1.71)*
I^{16}				−14.36 (−0.76)	−20.86 (−1.09)
I^{17}				−27.72 (−1.50)	−35.16 (−1.86)*
I^{18}				−19.82 (−0.94)	−28.22 (−1.31)
I^{19}				−35.86 (−1.72)*	−44.03 (−2.06)**
I^{20}				−26.11 (−1.52)	−32.51 (−1.86)*
I^{21}				−41.364 (−1.48)	−47.02 (−1.66)*
I^{22}				−17.96 (−1.01)	−24.88 (−1.36)
I^{23}				−28.59 (−1.78)*	−35.58 (−2.16)**
I^{24}				−24.19 (−1.52)	−30.53 (−1.86)*
I^{25}				−1.64 (−0.08)	−8.87 (−0.44)
I^{26}				−25.75 (−1.22)	−34.57 (−1.61)
I^{27}				−23.39 (−1.15)	−28.63 (−1.39)
I^{28}				−18.79 (−0.94)	−26.18 (−1.28)
Adjust R^2	0.0052	0.0003	0.0045	0.0096	0.0099
F test	11.5590	2.3484	10.1358	2.3103	2.3506
p-value	[0.0000]***	[0.1255]	[0.0000]***	[0.0000]***	[0.0000]***

Variable	Normal periods		Crisis periods		
	(6)	(7)	(8)	(9)	(10)
c_0	212.89	230.14	158.48	247.47	187.25
	(27.11)***	(11.11)***	(16.03)***	(4.56)***	(7.89)***
Alpha	−6708.41	−7179.43			
	(−4.41)***	(−4.73)***			
Volatility	−927.93	−910.66	988.08	744.49	793.32
	(−3.34)***	(−3.13)***	(4.04)***	(2.91)***	(3.14)***
Mvalue				−4.00	
				(−1.57)	
Profitability	−82.87	−67.23			
	(−2.53)**	(−1.96)*			
D^{CSI}	1.92		−7.75		−8.26
	(0.83)		(−3.21)***		(−3.33)***
I^1		−27.73		−43.96	−43.99
		(−1.30)		(−1.79)*	(−1.84)*
I^2		−41.57		−43.56	−44.70
		(−1.65)*		(−1.54)	(−1.60)
I^3		−33.15		−24.71	−25.36
		(−1.43)		(−0.94)	(−0.99)
I^4		−0.50		0.50	4.07
		(−0.02)		(0.02)	(0.16)
I^5		−8.18		−25.16	−23.86
		(−0.37)		(−1.00)	(−0.97)
I^6		−42.16		−25.16	−22.44
		(−1.49)		(−0.80)	(−0.73)
I^7		4.55		−62.49	−62.96
		(0.18)		(−2.14)**	(−2.19)**
I^8		−4.83		−8.62	−8.12
		(−0.22)		(−0.35)	(−0.34)
I^9		−12.73		11.36	11.42
		(−0.58)		(0.44)	(0.46)
I^{10}		−20.68		−21.89	−21.00
		(−0.91)		(−0.85)	(−0.84)
I^{11}		−18.91		−7.42	−5.72
		(−0.76)		(−0.26)	(−0.21)
I^{12}		−2.37		−25.41	−23.94
		(−0.11)		(−1.01)	(−0.97)
I^{13}		−0.64		5.16	3.99
		(−0.02)		(0.18)	(0.14)
I^{14}		−8.82		−9.58	−10.78
		(−0.39)		(−0.37)	(−0.43)
I^{15}		−29.86		−14.22	−13.06
		(−1.20)		(−0.50)	(−0.47)
I^{16}		−21.82		−8.35	−9.76
		(−0.87)		(−0.29)	(−0.35)
I^{17}		−33.03		−25.33	−23.95
		(−1.34)		(−0.90)	(−0.87)
I^{18}		−6.01		−17.51	−16.13
		(−0.93)		(−0.55)	(−0.52)
I^{19}		−20.52		−57.87	−57.35
		(−0.74)		(−1.83)*	(−1.85)*
I^{20}		−16.59		−32.7	−30.79
		(−0.71)		(−1.26)	(−1.21)
I^{21}		−13.78		−72.27	−77
		(−0.37)		(−1.69)*	(−1.82)*
I^{22}		−22.95		−9.83	−9.22
		(−0.96)		(−0.36)	(−0.35)
I^{23}		−27.12		−27.99	−28.74
		(−1.26)		(−1.14)	(−1.20)
I^{24}		−16.23		−31.82	−32.07
		(−0.77)		(−1.31)	(−1.36)
I^{25}		3.34		−3.68	−3.44
		(0.13)		(−0.12)	(−0.12)
I^{26}		−46.77		−4.60	−1.26
		(−1.67)*		(−0.14)	(−0.04)
I^{27}		−29.4		−13.95	−14.82
		(−1.08)		(−0.45)	(−0.49)
I^{28}		−1.45		−43.50	−44.07
		(−0.05)		(−1.43)	(−1.48)
Adjust R^2	0,0132	0.0135	0.0134	0.0186	0.0232
F-test	8.4168	1.9788	13.5240	2.1652	2.4626
p-value	[9.75E-07]***	[0.0011]***	[1.48E-06]***	[0.0003]***	[1.92E-05]***

Note: The note similar to Table 6. We choose to banks sector as the benchmark, and define 28 dummy variables, I_i^1 to I_i^{28} to represent 28 other industries, including electric utilities and public utilities (I^1) , petroleum and petrochemicals (I^2), coal mining (I^3), nonferrous metals (I^4), steel (I^5), construction (I^6), construction materials (I^7), real estate (I^8), chemical (I^9), machinery (I^{10}), defense (I^{11}), automobiles (I^{12}), electric power equipment (I^{13}), electronic components (I^{14}), computers (I^{15}), communications (I^{16}), household appliances (I^{17}), light manufacturing (I^{18}), textiles and garments (I^{19}), food and beverages (I^{20}), catering and tourism (I^{21}), retail and trade (I^{22}), health care (I^{23}), transportation (I^{24}), medias (I^{25}), composite (I^{26}), nonbanking financials (I^{27}), agriculture, forestry, animal husbandry & fishery (I^{28}), respectively.

Table 9
The results of determinants of the importance of stocks using the model in Eq. (18).

Variable	Estimations using the whole sample				
	(1)	(2)	(3)	(4)	(5)
c_0	169.34	191.47	180.83	201.73	194.7
	(38.28)***	(85.84)***	(51.61)***	(13.30)***	(13.06)***
RAlpha	0.05		0.06	0.06	0.06
	(3.47)***		(3.93)***	(3.94)***	(3.94)***
RMvalue	0.05				0.05
	(3.25)***				(2.89)***
D^{CSI}		−2.75	−3.97	−4.12	
		(−1.70)*	(−2.40)**	(−2.41)**	
I^1				−36.46	−39.85
				(−2.29)**	(−2.48)**
I^2				−45.75	−48.18
				(−2.45)**	(−2.57)**
I^3				−31.97	−36.23
				(−1.89)*	(−2.12)*
I^4				0.83	−3.34
				(0.05)	(−0.20)
I^5				−16.28	−20.10
				(−0.99)	(−1.21)
I^6				−33.79	−37.48
				(−1.63)	(−1.80)*
I^7				−27.25	−30.98
				(−1.42)	(−1.61)
I^8				−8.48	−12.30
				(−0.53)	(−0.76)
I^9				−1.23	−7.60
				(−0.07)	(−0.45)
I^{10}				−25.93	−29.68
				(−1.55)	(−1.76)*
I^{11}				−15.75	−20.43
				(−0.86)	(−1.11)
I^{12}				−14.06	−18.58
				(−0.86)	(−1.12)
I^{13}				−1.05	−4.03
				(−0.05)	(−0.21)
I^{14}				−11.78	−17.60
				(−0.70)	(−1.03)
I^{15}				−28.70	−33.28
				(−1.57)	(−1.80)*
I^{16}				−16.89	−20.36
				(−0.90)	(−1.08)
I^{17}				−28.49	−33.14
				(−1.56)	(−1.80)*
I^{18}				−18.47	−25.42
				(−0.89)	(−1.20)
I^{19}				−38.62	−44.50
				(−1.86)*	(−2.12)*
I^{20}				−28.23	−31.87
				(−1.66)*	(−1.86)*
I^{21}				−39.09	−42.09
				(−1.42)	(−1.53)*
I^{22}				−21.35	−25.68
				(−1.20)	(−1.43)
I^{23}				−31.71	−35.18
				(−1.99)**	(−2.19)**
I^{24}				−25.75	−29.50
				(−1.63)	(−1.85)*
I^{25}				−4.57	−8.82
				(−0.24)	(−0.45)
I^{26}				−26.07	−32.24
				(−1.26)	(−1.54)
I^{27}				−24.34	−27.67
				(−1.23)	(−1.39)
I^{28}				−20.96	−25.72
				(−1.05)	(−1.28)
Adjust R^2	0.0051	0.0004	0.0039	0.0094	0.0101
F test	11.5648	2.87305	9.18059	2.31752	2.40369
p-value	[0.0000]***	[0.0901]*	[0.0001]***	[0.0000]***	[0.0000]***

Variable	Normal periods		Crisis periods		
	(6)	(7)	(8)	(9)	(10)
c_0	149.31	163.76	217.09	230.5	243.25
	(19.59)***	(7.39)***	(41.52)***	(10.09)***	(10.52)***
RAlpha	0.09	0.10			
	(4.35)***	(4.68)***			
RVolatility	0.07	0.08	−0.12	−0.10	−0.09
	(3.43)***	(3.51)***	(−5.08)***	(−4.08)***	(−3.73)***
RMvalue				0.04	
				(1.62)	
RProfitability	0.04	0.03			
	(1.90)*	(1.16)			
D^{CSI}	1.04		−6.61		−7.43
	(0.45)		(−2.75)***		(−3.01)***
I^1		−24.81		−50.17	−50.98
		(−1.16)		(−2.12)**	(−2.17)**
I^2		−39.76		−51.56	−52.84
		(−1.57)		(−1.86)*	(−1.92)*
I^3		−33.40		−36.10	−36.49
		(−1.43)		(−1.41)	(−1.44)
I^4		3.56		−10.68	−7.73
		(0.16)		(−0.43)	(−0.31)
I^5		−9.13		−33.18	−32.46
		(−0.41)		(−1.35)	(−1.34)
I^6		−37.95		−32.60	−31.47
		(−1.35)		(−1.06)	(−1.03)
I^7		8.16		−72.34	−73.17
		(0.31)		(−2.53)**	(−2.57)**
I^8		0.55		−20.89	−20.57
		(0.03)		(−0.87)	(−0.86)
I^9		−6.19		−0.50	−0.03
		(−0.28)		(−0.02)	(0.00)
I^{10}		−18.60		−29.77	−29.57
		(−0.82)		(−1.19)	(−1.20)
I^{11}		−10.46		−19.17	−17.39
		(−0.42)		(−0.70)	(−0.64)
I^{12}		1.82		−33.80	−32.98
		(0.08)		(−1.38)	(−1.36)
I^{13}		3.44		−2.67	−4.84
		(0.13)		(−0.09)	(−0.17)
I^{14}		−2.06		−22.95	−23.14
		(−0.09)		(−0.90)	(−0.93)
I^{15}		−26.31		−28.36	−27.25
		(−1.05)		(−1.03)	(−1.00)
I^{16}		−17.64		−15.41	−17.28
		(−0.70)		(−0.55)	(−0.62)
I^{17}		−30.54		−29.44	−28.66
		(−1.23)		(−1.08)	(−1.06)
I^{18}		−25.36		−19.29	−17.90
		(−0.91)		(−0.62)	(−0.58)
I^{19}		−17.31		−65.85	−65.38
		(−0.62)		(−2.12)**	(−2.13)**
I^{20}		−17.48		−37.50	−36.85
		(−0.74)		(−1.48)	(−1.47)
I^{21}		−9.85		−65.50	−71.09
		(−0.27)		(−1.61)	(−1.76)*
I^{22}		−20.20		−21.01	−20.60
		(−0.84)		(−0.79)	(−0.79)
I^{23}		−26.36		−34.96	−36.93
		(−1.20)		(−1.47)	(−1.57)
I^{24}		−14.00		−36.07	−37.24
		(−0.65)		(−1.54)	(−1.61)
I^{25}		9.61		−13.45	−13.66
		(0.37)		(−0.47)	(−0.48)
I^{26}		−39.61		−13.17	−10.49
		(−1.41)		(−0.42)	(−0.34)
I^{27}		−21.47		−26.73	−27.29
		(−0.80)		(−0.90)	(−0.93)
I^{28}		3.99		−54.30	−54.61
		(0.15)		(−1.82)*	(−1.85)*
Adjust R^2	0.0139	0.0162	0.0186	0.0217	0.0250
F-test	8.9866	2.2018	18.8634	2.3912	2.6124
p-value	[3.37E-07]***	[0.0002]***	[7.74E-09]***	[3.66E-05]***	[4.74E-05]***

Note: Similar to Table 8.

which is intended to reflect the overall performance of China A-share market. There is high correlation between $D_{i,t}^{CSI}$ and the market value $Mvalue_{i,t}$ of stocks. Therefore, we choose either $D_{i,t}^{CSI}$ or the market value $Mvalue_{i,t}$ for the model, rather than bring them both into the models. We present the estimation results based on the whole sample, normal periods sample and crisis periods sample to discuss the changes of determinants of importance of stocks in Tables 8 and 9.

According to the whole sample estimation results in the fist five columns of Tables 8 and 9, when we substitute the market value variable into the market index dummy variable $D_{i,t}^{CSI}$, we find that being a component stock of CSI 300 Index can significantly decrease the ranking of improved PageRank, that is, increase the importance of the stock in the market. However, compared to $D_{i,t}^{CSI}$, the market value $Mvalue_{i,t}$ is more powerful in explaining the importance of the stocks. When we bring the industry index into the model, we find that industry effect exists. The industry property of each stock provides explanatory power, by improving the adjusted R^2 of the models. Moreover, we find the most important industry in the Chinese stock market that is always significant in different specifications of the models is the electric and public utilities sector, I^1, and the petroleum and petrochemical sector, I^2. The stocks in these two sectors tend to have high node influence in the stock market, that is, they discover information more quickly than others. Those two industries are mainly the energy sector, which are the upstream industries of the entire economy. The stocks in these two sectors have more information discovery power, which may be consistent with the economic structure in China.

According to the results of the separated samples of normal periods and crisis periods, we confirm that the component stocks of the CSI 300 Index can significantly increase the importance of the stocks in the market in the crisis periods as reported in column (8) and (10) in Tables 8 and 9, while there is no such effect in normal periods as reported in column (6) in Tables 8 and 9. For the industry indices, we find different results in different periods. In the normal periods, there is no industry effect on the importance of stocks as shown in column (7) in Tables 8 and 9, while there do exist industry index effect in crisis periods as shown in column (9) and (10) in Tables 8 and 9. After controlling different fundamental factors, we find the most important industry during the crisis periods in the Chinese stock market is construction materials, I^7, as reported in column (10) of Table 8, while that is the electric and public utilities sector, I^1, and construction materials, I^7, textiles and garments, I^{19}, as reported in column (10) of Table 9. Some of these industries belongs to cyclical industry, such as construction materials and textiles and garments. Some of them belongs to non-cyclical industry and defensive industry, for example, the electric and public utilities sector.

There results reveal that the comovement structure of the Chinese stock market changes as illustrated by the improved PageRank, and the changes of comovement structure is due to the changes of determinants in crisis periods and normal periods. In normal periods, investors in the Chinese stock market only care about the total risk, excess return and profitability of individual stocks, no matter whether the stocks belong to the market index or some sectors. The stocks with larger excess return alpha, profitability and volatility is more important and have better information discovery power in the market. It looks like that the investors in the Chinese stock market are more rational in normal periods, because sentiment-based factors are insignificant, and the significant fundamental factors covers the variables related to risk-return relationship. On the other side, the stocks with low volatility and belonging to the market index is more important in the market in the crisis periods. It seems that investors seek safe-haven assets in the crisis periods in the Chinese stock market. They care about risk rather than returns, draw funds out of cyclical-sector stocks, and make funds flow into the stocks belonging to the market index and some defensive sectors.

In all, our results reveal the risk preference of investors as well as the utility function used by investors change in different market situations in the Chinese stock market. It seems that investors prefer the stocks with high volatility that can provide more opportunities to high returns in normal periods, while investors prefer low volatility stocks that can defense against risk in crisis periods. Moreover, it seems that the utility function used by investors in normal periods is related to the risk-return balance, which is always the assumption of rational investors. However, the utility function used by investors in crisis periods seems only contain the risk and some sentiment variables. Thus, our results reveal to some extent the existence of rational investors in the Chinese stock market. Because the Chinese stock market is usually regarded as a retail investor-dominated emerging market and institutions are viewed as better-informed, rational, and prudential investors relative to retail investors, our results uncover some evidence about the rational behavior of institutional investors in the Chinese stock market.

7. Conclusion

This study proposes new methods to measure the causal comovement among individual stocks and the comovement structure of the market, based on the complex network constructed via the Granger causality tests. These measures are very convenient to illustrate the time-varying characteristics of causal comovement and the changes of comovement structure in the market. Furthermore, these new measures facilitate the examination of all kinds of hypotheses of comovement theories in a unified framework of methodology, which enables the comparison of the effects of different factors. In addition, these measures can be easily extended for investigating some important topics like risk contagion and systemic risk.

Our empirical results provide new evidence for the theories of comovement and asset pricing, and provide new insight into trading behaviors of investors in the Chinese stock market. We find that sentiment-based factors related to including stocks in or excluding stocks from the market index induce excess causal comovement in returns beyond that can be justified by fundamentals. From a micro perspective, we reveal that the information diffusion directions between pairs of stocks in the Chinese market are from the stock with high systematic risk beta coefficient, book-to-market ratio, liquidity, profitability, but low volatility of returns, and included in the market index, to the stock with the opposite characteristics. Most of these fundamental factors are pricing factors in multi-factor models.

From a macro perspective, our results show the comovement structure, measured by the newly proposed improved PageRank score, changes with the market situations, which is caused by the changes of investors' behaviors in different market situations in China. In normal periods, investors care more about the total risk, excess return and profitability of individual stocks, which are related to risk-return relationship, while they do not care whether they belong to the market index or some sectors. On the other side, the stocks with low volatility and belonging to the market index are more important in the crisis periods. It seems that investors only care about risk in crisis periods. They seek safe-haven assets. Investors in the Chinese stock market draw funds out of the high total risk and cyclical-sector stocks, while make funds flowing into the stocks with low total risk and belonging to the market index or some defensive sectors.

These results have some extended implications about rational investors and rational trading behaviors. Our empirical results reveal to some extent the existence of rational investors in the Chinese stock market, which was traditionally regarded as a retail investor-dominated emerging market. Furthermore, our empirical results provide new evidence for the "style investing" mentioned in Barberis and Shleifer (2003). Our results uncover that investors would like to invest in the component stocks in the market index and particular sectors during crisis periods. Therefore, the results of this paper have significant implications for the risk management and portfolio selection.

Acknowledgement

This work was supported by Prof. Bu's National Natural Science Foundation of China (NSFC) [grant numbers 91846108, 71671012, 71373001], and by Prof. Wu's National Natural Science Foundation of China (NSFC) [grant numbers 71725002, 71531001, U1636210, 71471009]. We thank the Handling Editor Prof. Sushanta Mallick, the Guest Editor Prof. Roman Matousek, and the two anonymous referees for their very useful comments and suggestions. We thank two students Fengzhi Li and Xiaojian Wang for their help in drawing Figs. 2 and 3. This paper was presented and discussed at the 14th International Conference on Services Systems and Services Management (ICSSSM 2017) and 15th International Conference on Financial System Engineering and Risk Management (FSERM 2017). We thank the participants of these conferences for their comments on this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.econmod.2019.03.002.

References

Allen, F., Gale, D., 2000. Financial contagion. J. Political Econ. 108 (1), 1–33.
Awokuse, T.O., Chopra, A., Bessler, D.A., 2009. Structural change and international stock market interdependence: evidence from Asian emerging markets. Econ. Modell. 26 (3), 549–559.
Barberis, N., Shleifer, A., 2003. Style investing. J. Financ. Econ. 68 (2), 161–199.
Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. J. Financ. Econ. 75 (2), 283–317.
Billio, M., Getmansky, M., Lo, A.W., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J. Financ. Econ. 104 (3), 535–559.
Bonanno, G., Lillo, F., Mantegna, R., 2001. High-frequency cross-correlation in a set of stocks. Quant. Finance 1 (1), 96–104.
Boyer, B.H., 2011. Style-related comovement: fundamentals or labels? J. Financ. 66 (1), 307–332.
Bu, H., 2011. Price dynamics and speculators in crude oil futures market. Syst. Eng. Procedia 2, 114–121.
Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. J. Financ. 63 (6), 2899–2939.
Cetorelli, N., Peristiani, S., 2013. Prestigious stock exchanges: a network analysis of international financial centers. J. Bank. Financ. 37 (5), 1543–1551.
Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econ. Stud. 47 (1), 225–238.
Dagher, L., Yacoubian, T., 2012. The causal relationship between energy consumption and economic growth in Lebanon. Energy Policy 50 (6), 795–801.
Diebold, F.X., Yılmaz, K., 2014. On the network topology of variance decompositions: measuring the connectedness of financial firms. J. Econom. 182 (1), 119–134.
Diks, C., Panchenko, V., 2006. A new statistic and practical guidelines for nonparametric Granger causality testing. J. Econ. Dyn. Control 30 (9–10), 1647–1669.
Diks, C., Wolski, M., 2016. Nonlinear Granger causality: guidelines for multivariate analysis. J. Appl. Econom. 31 (7), 1333–1351.
Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. J. Financ. 51 (1), 55–84.
Fleming, J., Kirby, C., Ostdiek, B., 1998. Information and volatility linkages in the stock, bond, and money markets. J. Financ. Econ. 49, 111–137.
Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: measuring stock market comovements. J. Financ. 57 (5), 2223–2261.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.
Granger, C.W.J., 1980. Testing for causality: a personal view. J. Econ. Dyn. Control (2), 329–352.
Hiemstra, C., Jones, J.D., 1994. Testing for linear and nonlinear Granger causality in the stock price-volume relation. J. Financ. 49 (5), 1639–1664.
Hsiao, C., 1981. Autoregressive modelling and money-income causality detection. J. Monet. Econ. 7 (1), 85–106.
Jung, S.S., Chang, W., 2016. Clustering stocks using partial correlation coefficients. Phys. A Stat. Mech. Appl. Stat. Mech. Appl. 462, 410–420.
Kallberg, J., Pasquariello, P., 2008. Time-series and cross-sectional excess comovement in stock indexes. J. Empir. Financ. 15 (3), 481–502.
Kamada, T., Kawai, S., 1989. An algorithm for drawing general undirected graphs. Inf. Process. Lett. 31 (1), 7–15.
King, B.F., 1966. Market and industry factors in stock price behavior. J. Bus. 39 (1), 139–190.
King, M.A., Wadhwani, S., 1990. Transmission of volatility between stock markets. Rev. Financ. Stud. 3 (1), 5–33.
Kodres, L.E., Pritsker, M., 2002. A rational expectations model of financial contagion. J. Financ. 57 (2), 769–799.
Kyle, A.S., Xiong, W., 2001. Contagion as a wealth effect. J. Financ. 56 (4), 1401–1440.
Li, X.M., Peng, L., 2017. US economic policy uncertainty and comovements between Chinese and US stock markets. Econ. Modell. 61, 27–39.
Masih, A.M., Masih, R., 1999. Are Asian stock market fluctuations due mainly to intra-regional contagion effects? Evidence based on Asian emerging stock markets. Pac. Basin Finance J. 7 (3), 251–282.
Mantegna, R.N., 1999. Hierarchical structure in financial markets. Eur. Phys. J. B 11 (1), 193–197.
Meyers, S.L., 1973. A re-examination of market and industry factors in stock price behavior. J. Financ. 28 (3), 695–705.
Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? J. Financ. Econ. 58 (1), 215–260.
Onnela, J.P., Chakraborti, A., Kaski, K., Kertesz, J., Kanto, A., 2003. Dynamics of market correlations: taxonomy and portfolio analysis. Phys. Rev. E 68 (5), 056110.
Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab.
Pan, L., Mishra, V., 2018. Stock market development and economic growth: empirical evidence from China. Econ. Modell. 68, 661–673.
Pasquariello, P., 2007. Imperfect competition, information heterogeneity, and financial contagion. Rev. Financ. Stud. 20, 391–426.
Pavlova, A., Rigobon, R., 2007. Asset prices and exchange rates. Rev. Financ. Stud. 20 (4), 1139–1180.
Peralta, G., Zareei, A., 2016. A network approach to portfolio selection. J. Empir. Financ. 38, 157–180.
Pindyck, R.S., Rotemberg, J.J., 1993. The comovement of stock prices. Q. J. Econ. 108 (4), 1073–1104.
Ribeiro, R., Veronesi, P., 2002. The Excess Comovement of International Stock Markets in Bad Times: a Rational Expectations Approach. Unpublished working paper. University of Chicago, Chicago, IL.
Tsani, S.Z., 2010. Energy consumption and economic growth: a causality analysis for Greece. Energy Econ. 32 (3), 582–590.
Tse, C.K., Liu, J., Lau, F.C., 2010. A network perspective of the stock market. J. Empir. Financ. 17 (4), 659–667.
Tu, C., 2014. Cointegration-based financial networks study in Chinese stock market. Phys. A Stat. Mech. Appl. Stat. Mech. Appl. 402, 245–254.
Veldkamp, L.L., 2006. Information markets and the comovement of asset prices. Rev. Econ. Stud. 73 (3), 823–845.
Xue, W.J., Zhang, L.W., 2017. Stock return autocorrelations and predictability in the Chinese stock market-Evidence from threshold quantile autoregressive models. Econ. Modell. 60, 391–401.
Yang, C., Chen, Y., Niu, L., Li, Q., 2014. Cointegration analysis and influence rank—a network approach to global stock markets. Phys. A Stat. Mech. Appl. Stat. Mech. Appl. 400, 168–185.
Yuan, K., 2005. Asymmetric price movements and borrowing constraints: a REE model of crisis, contagion, and confusion. J. Financ. 60, 379–411.
Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. J. Bus. Econ. Stat. 20 (1), 25–44.