No. 2505 [EN]

IMI Working Paper

Measuring systemic risk from textual Analysis: Evidence from Chinese Banks

Yi Fang, Hao Lin and Liping Lu

INTERNATIONAL MONETARY INSTITUTE

IMI Working Paper No. 2505 [EN]

Measuring systemic risk from textual Analysis: Evidence from Chinese Banks*

By* Yi Fang, Hao Lin, Liping Lu

October 2025

Abstract

Systemic risk is often measured with the interconnection among listed banks. However, the systemic risk of small and medium-sized banks is rarely addressed due to a lack data. Thus, we build a network of 711 banks in China using the co-occurrence analysis with media reports data, and construct an index based on the negative news to measure the systemic risk. The interconnection among large banks is relatively stable in the context of market turmoil, while the one between small and medium-sized banks is characterized by a transition from centralization to decentralization. In contrast with large banks, small and medium-sized banks become the main driver of systemic risk in the banking sector after 2013. It is mainly due to a hike of interbank business of small and medium-sized banks and cross-region operations, which have strengthened the interconnections among small and medium-sized banks, and their interconnections with large banks.

JEL Classification: G21, G33, N25
Keywords: Co-occurrence analysis, Text analysis, Systemic risk

---

*Published in International Review of Economics and Finance, Volume 103, October 2025.
* Yi Fang, National Academy of Development and Strategy, Renmin University of China; Hao Lin, Society Hub, The Hong Kong University of Science and Technology (Guangzhou); Liping Lu, China Banking Research Center and School of Finance, Renmin University of China

1. Introduction

Systemic risk of the banking sector is a core regulatory issue in China, especially the small and medium-sized banks since the recent bankruptcy of *Baoshang Bank*, i.e., a city commercial bank. According to the *National Financial Regulatory Administration* (NFRA), there are more than four thousand small and medium-sized banks, which account for about 40 % of total banking assets. In contrast to large state-owned banks, small and medium-sized banks are more vulnerable with lower capital adequacy, more aggressive business strategy and less government preferential treatment, which have become a key driver for the systemic risk in the banking sector.

Systemic risk is defined as the risk that negative shocks to certain financial institutions spread to other financial institutions through their interconnections, which results in the collapse of the whole financial system and further spillovers to the real economy (Billio et al., 2012; Benoit et al., 2017). Negative shocks are the fuse of the systemic risk (Dicks & Fulghieri, 2019), which include internal shocks such as bank runs or defaults of counterparties (Elliott et al., 2014), external shocks such as a downturn in the real economy (Bernanke et al., 1999). Interconnection is an important factor in constructing the indicators of systemic risk (Allen & Babus, 2009; Patro et al., 2013). The higher the interconnectedness among financial institutions, the more likely that the risk is amplified in thework topology and heat maps can displa network[1] (Battiston et al., 2012; Glasserman & Young, 2015).

To capture the interconnections among banks, existing studies often utilize the direct or indirect business connection methods, or the co-movement method. The direct business connection method employs the data on inter-bank asset and liability, while bank risks can be contagious through derivatives trading and inter-bank lending. If a financial institution fails, it will lead to losses of its counterparties, and may even bring domino-like failures. As most data on bilateral exposures is proprietary, which is even partially accessible for financial authorities (Anand et al., 2015; Bisias, Flood, Lo, & Valavanis, 2012; Upper, 2011). Therefore, the maximum entropy method and the minimum density method are often employed to estimate the bilateral exposures (see, e.g., Anand et al., 2013; Bargigli et al., 2015). However, these estimations do not fully capture the features of financial network due to the strong underlying assumptions. Mistrulli (2011) shows that the maximum entropy approach based on bilateral exposures overestimates the contagion under certain conditions. Anand et al. (2015) employ the minimum-density method to construct a multiplex network with a core-periphery structure for more than 1700 banks in Germany, while it tends to overestimate the contagion in stress testing. In contrast, the indirect business connection method captures the network due to common assets holdings, which employs balance sheet data with higher accessibility. It focuses on asset liquidity risk, i.e., price discount of asset sales by the bank will end up in losses suffered by banks with common asset holdings, which is an indirect risk contagion (Greenwood, 2015). Duarte and Eisenbach (2021) extend the model by Greenwood (2015) and introduce time-varying haircuts to track bank risks over

[1] Some research argues that a complete network allows for risk diversification, which can enhance systemic stability under small shocks (Acemoglu et al., 2015; Allen & Gale, 2000). Nevertheless, the network constructed in this paper is incomplete and some of the shocks in this paper have wide-spread impacts, such as the global financial crisis in 2008 and stock market crash in China in 2015. Thus, we only consider the positive correlation between interconnectedness and systemic risk.

time. Using balance sheet data for large US banks during 1996–2016, Duarte and Eisenbach (2021) show that a concentration of illiquid assets, i.e., indirect interconnectedness due to fire sales, plays an important role for banks' aggregate vulnerability, which can serve as a measure of systemic risk. Di Gangi et al. (2018) employ the maximum entropy approach in the framework of Greenwood (2015), which is a more generalized method based on bank assets and capitalization. Nevertheless, the financial data is lagged with a low frequency, and banks often have incentives to manipulate the data, which undermines the data reliability (Cao et al., 2022; Karpoff, 2021). In addition, many unlisted banks do not disclose their financial data, which attract less attention from the academia and regulators.

The co-movement method is based on the data from the financial market, which uses the correlation of stock volatility, stock returns, credit default swap (CDS) and other market data of financial institutions to measure the systemic risk. Diebold and Yilmaz (2009) measure a dynamic correlation of stock return volatility of financial institutions in 19 stock markets during 1992–2007 with generalized forecast error variance decomposition, and show that the peak of volatility overflow index corresponds well to the breakout of major risk events. Ando et al. (2022) modify the Diebold-Yilmaz method by estimating quantile dependencies in a quantile-VaR model which allows for residuals driven by a series of latent factors, which is particularly suitable for studying large and rare idiosyncratic shocks. Patro et al. (2013) examine 22 bank holding companies and investment banks during 1988–2008, and find that the correlation of daily stock returns is a simple and robust indicator for the systemic risk. Tobias and Brunnermeier (2016) propose a measure of systemic risk, ΔCoVaR, which is defined as the difference between the CoVaR of financial system in the context of distressed financial institutions and that in the median state. Markose et al. (2012) construct the network of CDS market in the US and find that the risk of highly inter-connected financial institutions will cause connected institutions to fail. Oh and Patton (2018) employ copula-based models to estimate time-varying high-dimensional distributions of CDS spreads on U.S. firms, which evaluates the expected distress proportion of firms conditional on another firm's distress. Their findings suggest that risk spillovers from the real sector to the financial sector are greater than those in the reverse direction. Tian et al. (2022) develop a GARCH copula quantile regression model to measure dynamic risk spillovers between oil markets and stock markets. However, some empirical studies identify the in efficiency of financial markets in China (Chen et al., 2017; Kristoufek & Vosvrda, 2013), while a violation of efficient market hypothesis may lead to a bias in the estimation of systemic risk (Cerchiello et al., 2017; Malkiel, 2003). Besides, a limited availability of market data for unlisted financial institutions renders it impossible to fully capture the risk of the whole financial system.

Generally speaking, these standard methods have some limitations. On the one hand, they cannot monitor a large number of unlisted banks due to limited data availability. On the other hand, they only capture certain aspects of the inter-connectedness. Interconnectedness includes not only the direct connections due to inter-bank transactions and co-movements in the market data, but also indirect connections from inter-bank competition, management profiles, customer and investor characteristics, reputational linkages, and competitive dynamics (Acemoglu et al., 2015; Acharya & Thakor, 2016; Allen et al., 2012; Capponi et al., 2016; Dicks & Fulghieri, 2019; Eisenberg & NOE, 2001). All these linkages are potential channels for the risk contagion (Aharony & Swary, 1983; Račickas & Vasiliauskaitė, 2012; Sarlin, 2016), while we can employ measures with alternative datasets to overcome the weakness of standard methods.

Due to the limitations of standard datasets, researchers have explored textual data to measure systemic risk. An emerging strand of literature uses investor sentiment similarity to estimate inter-institutional connections. For instance, Cerchiello et al. (2017) employ Twitter sentiment indices to estimate systemic risk for Italian banks, while Nyman et al. (2021) analyze narrative topic consensus in UK news to provide early warning indicators of systemic risk events. Similarly, Andrieş¸ et al. (2022) find that covariates of news sentiment convey information on risk spillovers from global systemically important banks to other systemically important institutions. In addition, some literature explores inter-institutional connections from 10-K discussions. For example, Hoberg and Phillips (2016) use similarities of the description of products in firms' 10-Ks to measure firms' interconnectedness. Similarly, Bushman et al. (2016) assess banks' interconnectedness based on the similarities of topics of 10-Ks discussions.

While these studies have made significant contribution in the text-based systemic risk measurement, they primarily focus on interconnection, which represents only one element of systemic risk. Few studies attempt to capture both interconnections and negative shocks, the two core elements of systemic risk. We combine co-occurrence analysis and sentiment analysis techniques to capture both interconnections and negative shocks. By precisely identifying bank co-occurrences in negative news, we construct a novel risk indicator, NCOI, which can capture the evolution of systemic risk across thirteen major risk events from 2000 to 2021.

Specifically, we estimate inter-bank connectedness through news co-occurrence, which can measure direct and indirect connections. Co-occurrence analysis captures the connection tightness among these keywords by counting the co-occurrence frequency of keywords in the text. Firms reported in a piece of news often have certain connections, such as business contact, management mode similarity, customer or investor interactions, reputational linkages, geographical proximity and competitive dynamics (Bernstein et al., 2002; Kazinnik et al., 2024; Yaros & Imieliński, 2014). By counting the co-occurrence frequency of firms in the text, co-occurrence analysis captures the relative strength of these connections. An emerging strand of literature supports the validity of this approach. For instance, Rönnqvist and Sarlin (2015) employ pairwise co-occurrence of banks in Reuters online news as the links' weights of bank network, and use the information centrality in network analysis to measure the systemic importance of banks. Wan et al. (2021) construct a co-occurrence network with 87 firms that show up most frequently in Reuters news, and find that clustering firms by co-occurrence tightness is consistent with traditional industry classification. Similarly, Zheng and Schwenkler (2020) find that dis tressed firms' linkages implied in news act as channels for risk contagion. Yang et al. (2023) identify communities in the co-occurrence network of Chinese financial firms using headline news from Sina Finance, while Kazinnik et al. (2024) examine the dynamic co-occurrence network of DFAST banks in response to COVID-19. These studies provide empirical evidence that co-occurrence in news articles reflects the strength of interconnections among financial institutions.

We also employ the text sentiment analysis to capture negative shocks, another core element of systemic risk. Bank-related news contain comments on the movement of financial markets or banks from banks' insiders, analysts, institutional investors, and regulators, etc. (Engelberg et al., 2012; Kearney & Liu, 2014; Thompson et al., 1987). News with positive sentiment does not necessarily mean positive signals for banks due to the existence of media corruption or control, while news with negative sentiment often corresponds to negative shocks suffered by banks (Li, 2013; Wang & Ye, 2015; Zou et al., 2019), e.g., credit defaults, devaluation of asset holdings, stock price plummet, regulatory pressure, and geopolitical crisis, etc. Recall that bank co-occurrence reveals inter-dependencies between banks, which are potential links for risk contagion (Azizpour et al., 2018; Herskovic et al., 2020; Zheng & Schwenkler, 2020). Therefore, the co-occurrence of two banks reported in the same negative news implies that the risk derives from the shock to a bank spillovers to the other through the interconnection. Fig. 1 shows that we identify the inter-bank risk contagion precisely through combining sentiment analysis and co-occurrence analysis, which distinguishes our work from Rönnqvist and Sarlin (2015).

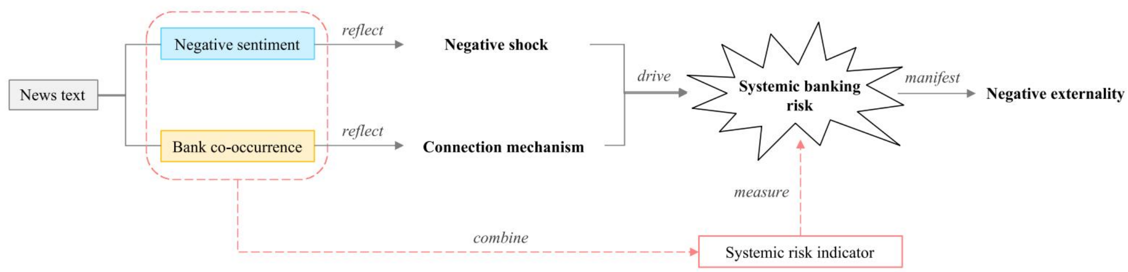


Fig. 1. News sentiment, co-occurrence and systemic risk.

We employ news data to measure the systemic risk due to the following advantages. First, news data offers more comprehensive information on direct and indirect inter-bank connections, such as business activities, investment strategies, stock performance, management, customer and investor characteristics, etc. Second, news data has a higher data availability and frequency, which can cover both listed and unlisted banks. Also, real-time news data is conducive to track the risk of financial institutions in a timely way. Besides, a huge volume of news data can alleviate the concern on the noise, which can reveal inter-dependencies between banks more precisely (Wan et al., 2021).

We use co-occurrence analysis to calculate news co-occurrence frequency of bank pairs and store them in co-occurrence matrix, and construct co-occurrence network and heat map of large banks versus small and medium-sized banks. We show that news co-occurrence can reveal inter-bank connections in a proper way. We further introduce the sentiment index in order to obtain the co-occurrence index based on negative news (NCOI). We examine the validity of co-occurrence index through the comparison with traditional risk in dicators. We also examine the capability of co-occurrence index to capture individual bank risk and explore the factors that can affect the systemic risk.

We employ news co-occurrence analysis and sentiment analysis to construct a systemic risk indicator from the news reports. Our indicator serves for large banks as well as small and medium-sized banks, which captures the two drivers of systemic risk, i.e., negative shock and interconnection. A vast amount of real-time news can capture the inter-bank connections in various dimensions, which renders our measurement with substantial practical value. Besides, we present the banking network using data from about 950,000 news articles in the INFOBANK economic news database[2], which enables us to track more than 700 banks over 20 years. The network topology and heat maps can display the hierarchical structure of the banking sector and the interconnections among banks reported in the news in China. We show that inter-bank connectedness of large banks is relatively stable during the financial crisis, while that of small and medium-sized banks is characterized by a transition to decentralization. The significant risk fluctuation highlights the importance of studying the risks of small and medium-sized banks in China.

This paper contributes to a growing literature on systemic risk measurement in the banking sector, which predominantly relies on market and balance sheet data of listed banks. A lot of research has applied systemic risk measures, such as ΔCoVaR and MES, to evaluate the systemic risk contributions of Chinese listed banks, i.e., Duan et al. (2021), Lee et al. (2023), and Yan et al. (2023). Wang et al. (2014) employ extremal quantile regression with market data to measure systemic risk, who show that listed banks contribute more to systemic risk than other financial institutions. Using 16 Chinese listed banks during 2010–2016, Fang, Xiao, Yu, and You (2018) find that the PCA model provides more reliable systemic risk rankings than five popular risk measures, while fundamentals-based measures perform better than price-based measures in terms of risk ranking. Gong et al. (2019) construct a causal network model to measure systemic risk across Chinese financial institutions, and find intensified network connections during crisis periods. Shi et al. (2024) employ bank balance sheet data in a bipartite network model, and find that indirect losses from common exposures often exceed direct losses, while mortgage exposure is the primary vulnerability.

In contrast, this paper adopts the most comprehensive sample with over 700 banks, which covers various types of banks, such as large state-owned banks, joint-stock banks, city/rural commercial banks, rural credit cooperatives and new types of rural financial institutions, foreign banks, and private banks, etc. Our study captures systemic risk across a spectrum of institutions that are often excluded in the existing literature, which often focuses on listed banks. Furthermore, we employ textual analysis on millions of news reports to assess systemic risk, which offers a novel perspective that complements the existing research. This comprehensive approach enables us to provide a more holistic view of systemic risk in the Chinese banking sector, and addresses a critical issue by covering less-studied institutions that play a significant role in the financial system. This paper makes a distinctive contribution to the literature on systemic risk of Chinese banks.

The rest of the paper is organized as follows. Section 2 introduces the methodology and data. Section 3 shows the static and dynamic characteristics of bank network and constructs a measurement for the systemic risk. Section 4 substantiates the validity of the

---

[2] INFOBANK is a popular news database in China. Wang and Ye (2015) collect news reports on 366 Chinese listed firms from INFOBANK during 2003–2006 and examine the impact of media reports on corporate valuation. Zou et al. (2019) show that media reports on a listed firm increase investors' recognition of the firm, which is reflected in a higher stock return. Huang (2018) finds that news reports on listed firms have a positive impact on the stock return in the short term, but there is a reversal effect in the long term. INFOBANK has collected daily news data since 1992, which covers professional news reports from more than 1200 mainstream media in mainland China, such as general newspapers, financial and economic newspaper, and financial and economic websites. Therefore, we select "China Economic News Database" in INFOBANK as the news data source considering the coverage, information density and crawl difficulty.

measurement in various dimensions. Section 5 examines the robustness of the indicators to various types of news media. Section 6 concludes the paper.

2. Institutional background

According to the *China Banking and Insurance Regulatory Commission* (CBIRC), the number of banking financial institutions has reached 4604 in 2021, which includes 6 large state-owned banks, 12 joint-stock banks, 133 city commercial banks, 1539 rural commercial banks, and 1637 village and town banks. While the total assets of the big six state-owned banks account for 40.1 % of total banking assets in 2021, city commercial banks and rural financial institutions account for 13.1 % and 13.3 % respectively. The total number of small and medium-sized banks and their total assets have been growing rapidly, which are essential players in the banking market.

The banking sector has a tight inter-bank connectedness in China, which forms a highly dependent and complex network. The risk of a bank can often spread rapidly within the network through inter-bank connections, which ends up in risk contagion effects. Thus, the supervision of systemic risk involves identifying fragile and unstable bank nodes, monitoring their risk status, and curbing risk spillovers in a timely way.

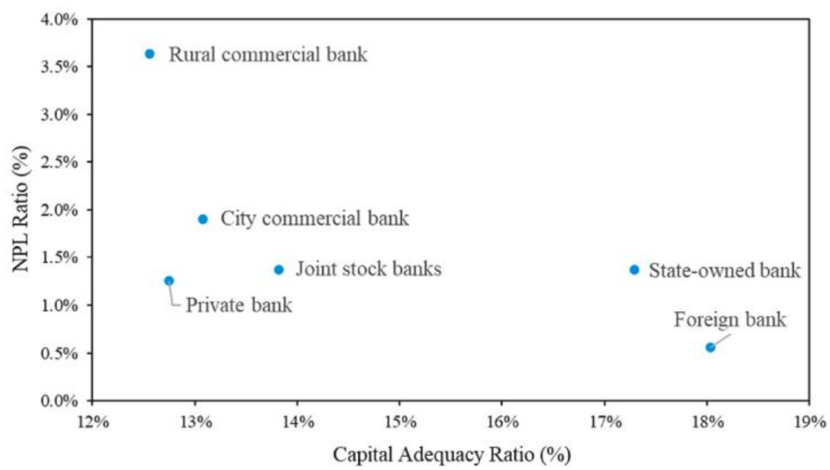


Fig. 2. Non-performing loan ratio and capital adequacy ratio of banks in 2021.

Non-performing loan ratio can measure the loan risk of banks, and capital adequacy ratio can measure the capability to resist risks. We compare non-performing loan ratios and capital adequacy ratios of various types of banks at the end of 2021 in Fig. 2. Large state-owned banks and foreign banks have higher capital adequacy ratios and lower non-performing loan ratios. In contrast, rural commercial banks have the highest non-performing loan ratios and the lowest capital adequacy ratios among all types of banks. In addition, although non-performing loan ratios of private banks and city commercial banks are lower than that of rural commercial banks, their capital adequacy ratios are lower than that of large state-owned banks.

The *People's Bank of China* disclosed the stress tests results of 4015 domestic banks in the China Financial Stability Report (2021). Capital adequacy ratios of 30 large and medium-sized banks met the regulatory requirement of 10.5 % under various shocks in the credit risk stress test. However, 1390 (among 3985 in total) small and medium-sized banks could not pass the stress test under the mildest overall credit risk shock, which ends up in a failure rate of 34.88 %. In addition, all 30 large and medium-sized banks passed the mild stress test in the liquidity risk stress test, while nearly 140 small and medium-sized banks failed to pass the test. Moreover, small and medium-sized banks in China are subject to less stringent regulation than state-owned banks, which enables them to engage in riskier activities and accumulate risks in less transparent ways. For instance, state-owned banks must meet additional leverage ratio requirements on top of the standard requirements for commercial banks. The additional leverage ratio requirement is 50 % of the additional capital requirement for systemically important banks and should be met by Tier 1 capital. These stricter regulatory measures ensure that state-owned banks maintain robust buffers against financial shocks. In contrast, small and medium-sized banks with less regulatory constraints can operate with thinner capital margins, leaving them more vulnerable to risk accumulation (Shi et al., 2022).

The risks of small and medium-sized banks are often contagious, which can be amplified rapidly through the intricate connections in the banking network. The frequent inter-bank trading is a direct connection channel. Banks that fail the liquidity stress test often have high dependence on inter-bank market and liquidity stress at the same time. Thus, a small negative shock could transmit to their counterparties, which creates a collective liquidity crunch and endangers the entire banking system. Besides, the herding behaviors of small and medium-sized banks lead to similar risk exposures, which forms an indirect connection channel.

Banks follow herding behaviors as they want to be rescued first. Acharya and Yorulmazer (2007) show that the optimal regulatory choice is to bail out failed banks in order to avoid continuous losses in the context of a large number of bank failures. The implicit government guarantee stimulates banks towards herding behaviors, such as lending to similar industries and bearing similar interest rate risks in order to increase the possibility of being rescued, i.e., “too many to fail”. Thus, it will increase the common risk exposure of banks, which elevates the probability of a crisis (Benoit et al., 2017). Banks adopt herding behaviors in order to cut their borrowing costs. Acharya and Yorulmazer (2008) and Silva-Buston (2019) find that the borrowing cost of a bank will increase in the context of negative news about other banks when bank lending rates have common systemic factors, which may convey unfavorable information about the common factors. Banks have a more pronounced increase in the borrowing costs when there is less commonality in bank lending rates. Thus, banks engage in herding behaviors in the investment to alleviate the impact of information contagion on the borrowing costs.

We emphasize the importance of small and medium-sized banks in curtailing systemic risks due to the concern on “too many to fail” instead of “too big to fail”.[3] A small shock can spread to multiple banks in the context of a large number of vulnerable banks, which can trigger contagion effects in a similar way with banks at the center of the network. Therefore, these small banks can also become a vital part of the network (Varotto & Zhao, 2018). The cost of bailouts will also be quite high as the total assets of these banks are

[3] "Too big to fail" refers to that certain corporations or financial institutions are so large and complex that their failure would have disastrous effect on the economy. It is often related with government intervention to prevent such failures, as the collapse of a "too big to fail" entity could bring severe consequences, such as financial instability and economic recession. The term gained prominence during the global financial crisis, when several large banks and financial institutions in the United States were treated as "too big to fail" and received government bailouts.

sizeable, and the government will face the dilemma between financial crisis and fiscal collapse (Brown & Dinç, 2011; Morrison, 2011).

Thus, the regulatory authorities should strengthen the supervision of small and medium-sized banks due to a high failure risk, a high rescue cost in case of collective failures of these banks, and the moral hazard problem under the implicit government guarantee. However, the standard methods of systemic risk measurement are limited by the data of listed banks, which cannot measure the risks of small and medium-sized banks. We employ innovative data source and risk measurement to bring all banks into a unified framework, which can achieve a comprehensive real-time risk monitoring.

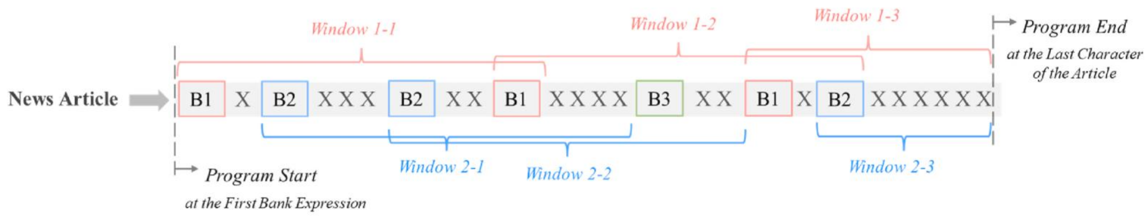


Fig. 3. News co-occurrence frequency. "B1" (or "B2", "B3") represents phrases that match regular expressions of bank "B1" (or "B2", "B3"). The symbol "X" represents other Chinese characters.

3. Methodology and data

3.1. Text sentiment analysis

We employ the Chinese financial text sentiment dictionary[4] constructed by Jiang et al. (2019, 2021[5]), which is designed specifically for the analysis of Chinese financial news. The dictionary integrates the *word2vec* algorithm and manual filtering methods, which expands a couple of Chinese and English financial sentiment dictionaries such as the dictionary by Loughran and McDonald (2011), and NTUSD Simplified Chinese Sentiment Dictionary, etc. The dictionary includes a total of 9228 emotional words, i.e., 5890 negative words and 3338 positive words.

For a news article A, the corresponding sentiment index is calculated with the "Relative sentiment shifts" formula in Nyman et al. (2021):

$$Senti_A = \frac{|n(positive)_A| - |n(negative)_A|}{|n(positive)_A| + |n(negative)_A|} \tag{1}$$

We can obtain the sentiment index of each news through capturing the relative sentiment of news text. As the overall sentiment of a news article may not necessarily represent the sentiment of the bank in the report, we narrow the text sentiment analysis to the sentence where the bank keyword is located. We analyze the text sentiment in the

[4] The dictionary is here: https://github.com/MengLingchao/Chinese_financial_sentiment_dictionary.

[5] Their article was published in the Chinese-language journal China Economic Quarterly. China Economic Quarterly was founded in October 2001. It is a comprehensive economics journal supervised by Peking University, sponsored by the China Center for Economic Research of Peking University, and published by Peking University Press. The journal publishes original theories, experiences, Review and commentary Chinese eco nomics papers, and it is a source journal of Chinese Social Science Citation Index (2021–2022). According to CNKI, as of 2022, China Economic Quarterly has been downloaded 4,706,031 times and cited 136,252 times, with a composite impact factor of 8.173 and a comprehensive impact factor of 6.013 for the 2019 edition.

units of sentences and take the average value as the sentiment index of the news article, which is simple, easy to interpret, and informative. According to Eq. (1), if the value of $Senti_A$ of a news article A is larger, the overall mood of the article is more likely to be positive rather than negative. Thus, if $Senti_A > 0$, A is considered as positive news, while if $Senti_A \leq 0$, A is considered as negative news.

3.2. Co-occurrence analysis and systemic risk

We use regular expressions in *Python* to match bank names from the text. A regular expression describes a pattern of string matching, which can check whether a sub-string is contained in a piece of text and extract sub-strings that meet a certain condition from the text. There might be various names for the same bank in the news reports, such as full name, abbreviations, synonymous names, and historical names, etc. Therefore, we should account for these name variations when counting the co-occurrence frequency. We specify each sample bank as a set of regular expressions that represent its full name and name variations.

Rönnqvist and Sarlin (2015) show that too large a text scope may result in a statistical "co-occurrence relationship" with no practical significance. For example, two banks reported in different events may also be treated as co-occurrence. In contrast, too small a text scope may render a long sentence truncated artificially, which may lose some co-occurrence information. Following Rönnqvist and Sarlin (2015), we set the text scope as 400 Chinese characters, and check whether there are bank pairs in each sliding window. Specifically, the program will scan the text, look for sub-strings that match predefined regular expressions, identify all banks reported in the text, and calculate the co-occurrence frequency of banks through pairwise combination. In order to avoid repeated counting, we look backward to see if regular expressions of other banks exist within a 400-character window. Furthermore, the co-occurrence frequency of the same bank pair that occur multiple times in the same text will be accumulated to reflect the importance of the bank pair. The co-occurrence of regular expressions of the same bank will not be counted. The window scope is not consistently equal to 400 characters, as the last window is limited to the scope of the last character of the news article. Fig. 3 shows that the co-occurrence frequency of bank B1 and bank B2 in 3 sliding windows starting at B1 and in 3 windows starting at B2 in the news article equals 2, 1, 1, 1, 1, 0 respectively. Thus, the co-occurrence frequency of the bank pair (B1, B2) equals 6.

We store the co-occurrence frequency of bank pair (i, j) in the co-occurrence matrix CM in the corresponding period. We standardize it by the number of sample banks in the current period to eliminate the impact of the number of sample banks on the co-occurrence frequency. Each element of the matrix is taken as the weights of the pair-wise links between each bank node in the network, which can form a series of cross-sectional networks. We can obtain the co-occurrence matrix of the current year by adding the co-occurrence matrix of 12 months in the same year; and we can obtain the co-occurrence matrix of a certain period by taking the average of cooccurrence matrix of several adjacent years.

We use the co-occurrence index COI_t to describe its connection strength for the entire banking system:

$$COI_t = \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} CM_t(i,j) \tag{2}$$

where t is the time (observation month); $CM_t(i, j)$ is the element of the co-occurrence matrix CM_t at the position (i, j) of the observation month$t$, which equals the co-occurrence frequency of bank i and j in the observation month$t$ divided by $N_t \times (N_t - 1)$[6]; N_t is the number of sample banks. Therefore, if the connection strength among banks is higher, the co-occurrence index of the banking system will be higher.

As the negative shock is one of the driving factors of systemic risk, we calculate the co-occurrence index of negative news as a risk indicator, i.e., $NCOI_t$, at time t. We use the text sentiment analysis to obtain its sentiment index $Senti_t$ according to Eq. (1) for each cooccurrence news. The co-occurrence index of bank pair (i, j) in a news article is defined as $R_{News}(i,j)$. If all news where co-occurrence bank pair (i, j) reported in observation month t constitute the news set $NewsSet_t(i,j)$, we can obtain the co-occurrence matrix based on negative news:

$$CM_t^{neg}(i,j) = \sum_{News \in NewsSet_t(i,j)} R_{News}(i,j) \times \mathbf{1}_{[Senti(News) \leq 0]} \tag{3}$$

where $\mathbf{1}_{[Senti(News) \leq 0]}$ is an integer function, which indicates whether the news corresponding to the bank is negative. Finally, we can obtain the co-occurrence index based on negative news:

$$NCOI_t = \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} CM_t^{neg}(i,j) \tag{4}$$

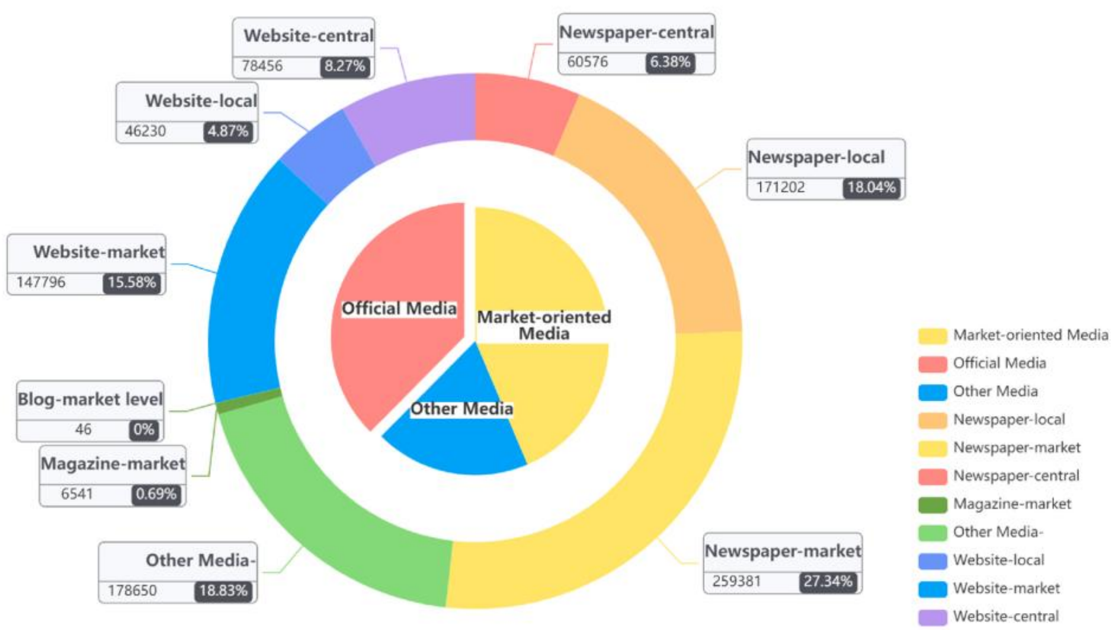


Fig. 4. Distribution of news articles from various types of media.

[6] For a network of N banks, the maximum number of internal links is $C_N^2 = N \times (N - 1)$.

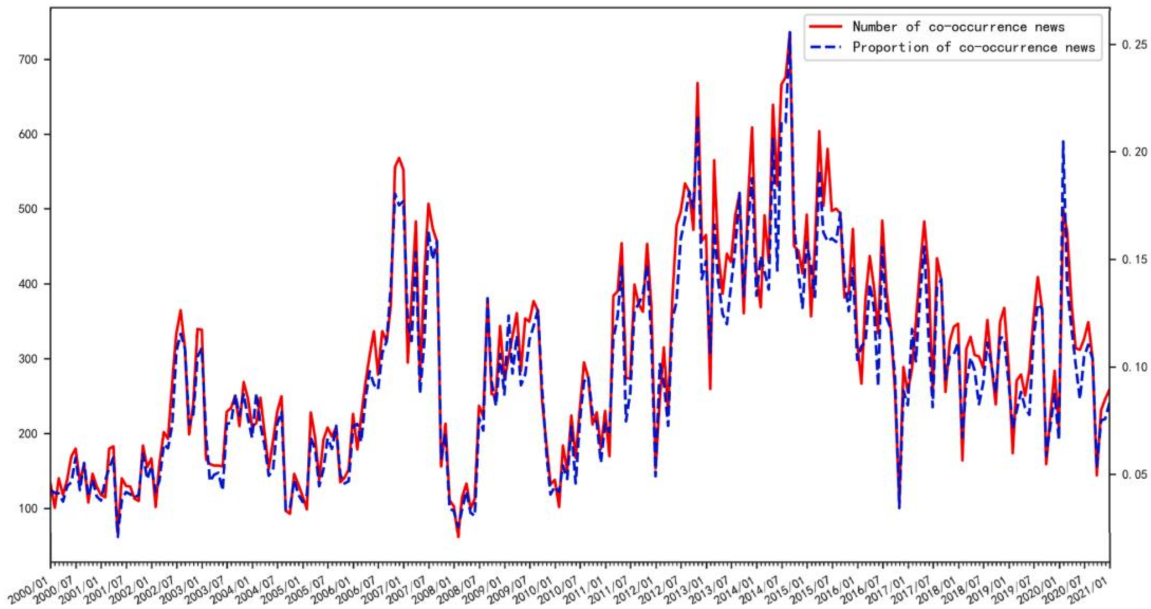


Fig. 5. Monthly changes in the number and proportion of co-occurring news.

Table 1
Sample coverage of banks in China.

Bank type		Sample #	Total #	Coverage
State-owned bank		6	6	100 %
Joint stock banks		12	12	100 %
City commercial bank		130	130	100 %
Rural small and medium-sized bank	Village Bank	197	1642	12 %
	Rural cooperative Bank	9	26	34.62 %
	Rural commercial Bank	200	1569	12.75 %
	Rural credit cooperative	36	609	5.91 %
	Rural fund mutual cooperative	4	41	9.76 %
Foreign bank		36	41	87.80 %
Private bank		19	19	100 %

3.3. Sample description

The time span is from Jan 1, 2000 to Dec 31, 2021. The total number of news reports in the sample period is about 6.38 million. We set the fuzzy query condition with “bank” as the key word, and finally screens out 948,878 news articles to construct a news text database at the monthly frequency. Fig. 4 shows the distribution of media sources for the news article database. News reports from official media (including central and local media) account for about 40 %, which can provide a decent dataset to construct news cooccurrence and sentiment indicators.[7]

Fig. 5 shows the monthly change in the total number of co-occurrence news and its proportion in total news from Jan 2000 to Dec 2021. Both the number and proportion of co-occurrence news reach their peaks in 2007–2008, 2013 and 2015, which corresponds to the global financial crisis, the credit crunch in the inter-bank market, and the stock

[7] For the media classification, we rely on the results by INFOBANK, which is also the database of news articles analyzed in our study. Our dataset comprises 39 central-level newspaper, 132 local newspaper, 308 market-oriented newspaper, 21 market-oriented blogs, 39 central-level websites, 240 local websites, 333 market-oriented websites, and 136 market-oriented magazines. A complete breakdown of all media outlets is in the online appendix: https://lin1103.github.io/Measure-systemic-risk-using-text-analysis/.

market crash in China respectively. The inter-bank connectedness goes up gradually before the shock, while the transactions between banks are frequent and the bank performance is rising. When the shock comes, the interconnectedness reaches a peak with a continuous decline afterwards. The decline of connectedness reflects the reduction of inter-bank lending, the number of common customers, and the divergence of performance of banks, etc. Therefore, news co-occurrence can effectively capture the changes in the connections among banks before and after the shock. In addition, the trend of the number and the proportion of co-occurrence news vary with time similarly, which indicates that the number of co-occurrence news is not affected by the total number of news texts in different periods. Thus, similar with existing studies based on news co-occurrence analysis (Rönnqvist & Sarlin, 2015), we take the co-occurrence frequency as a measure to display the change more intuitively, although the sample size of the bank has been considered into the construction of COI_t and $NCOI_t$.

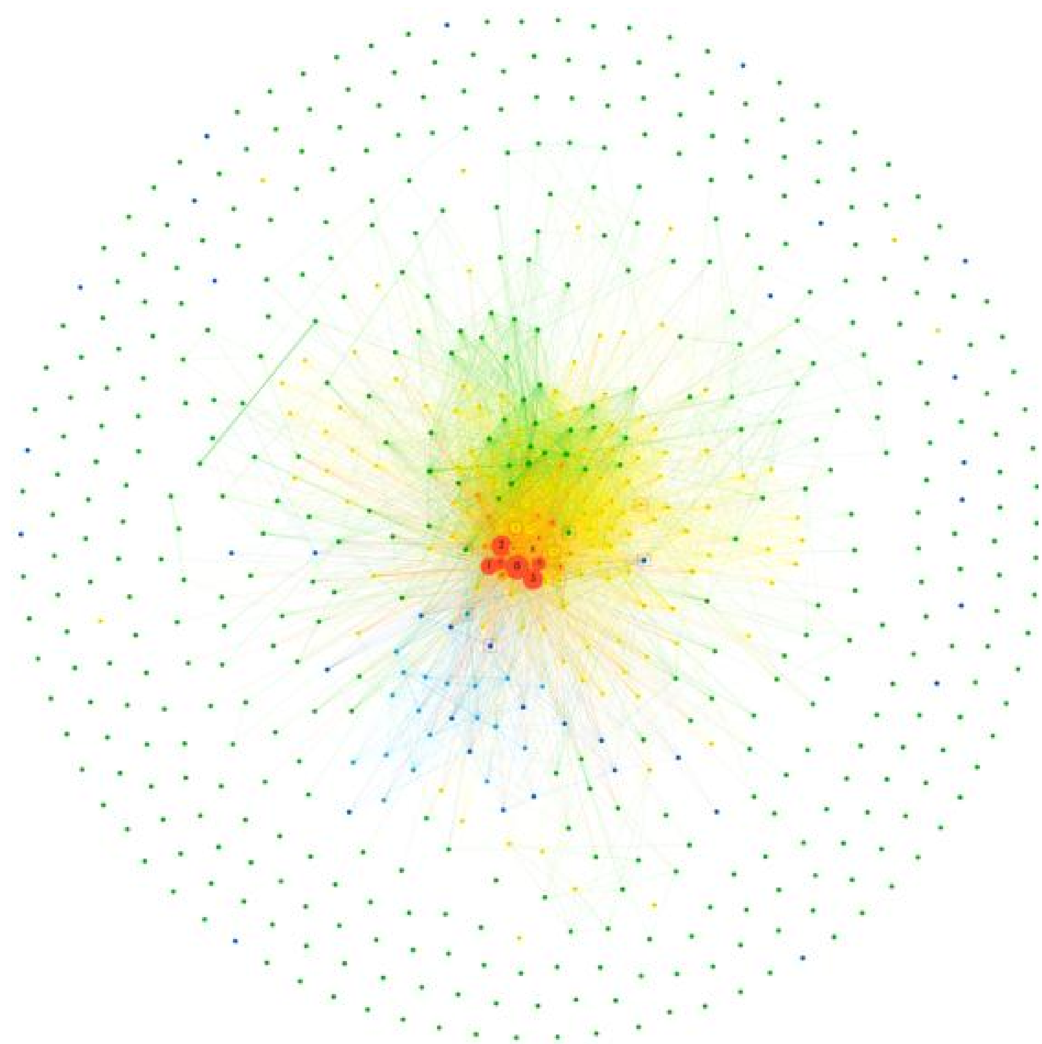


Fig. 6. Bank network in China in 2021.[8]

[8] For high-resolution pictures, please refer to the online appendix: https://lin1103.github.io/Measure-systemic-risk-using-text-analysis/.

We keep banks reported by at least three news articles.[9] Table 1 shows that the sample covers all state-owned banks, joint-stock banks, city commercial banks and private banks that exist by 2021, and the coverage ratio of foreign banks is 87.8 %. Although the sample has a relatively low coverage for rural small and medium-sized banks, the size of these banks is much smaller than other types of banks. In addition, the banking sector has witnessed a wave of bank name changes, restructuring, and M&As, so the list of banks changes over the years. We collect the historical information of each bank, which incorporates its old name,[10] name before restructuring, original bank names before M&As, so as to build a dynamic sample with a total of 711 banks.[11] Our sample covers banks covering a majority of the total banking assets, which can reflect the systemic risk of the banking sector. Appendix 1 shows a complete list of sample banks.

4. Co-occurrence network analysis and risk indicator

4.1. Static network of the sample bank

The co-occurrence among banks can reflect the relationship between bank pairs. A complex relationship network can be formed by examining pair-wise co-occurrence of sample banks. We can draw the inter-bank connection network in 2021 by the Fruchterman Reingold layout algorithm (Fruchterman & Reingold, 1991) in Fig. 6. The numbered nodes represent banks, and the node color represent bank types. Red nodes represent state-owned banks; orange nodes represent joint stock banks; yellow nodes represent city commercial banks; green nodes represent rural small and medium-sized banks; blue nodes represent foreign banks; and dark blue nodes represent private banks. Bank name and type corresponding to each number are listed in Appendix 1. If there is a relationship between two banks, they are connected by an edge. The tighter the inter-bank connectedness, the thicker the edge. Node size is proportional to the degree of the node, which equals the number of nodes associated with the node. The greater the degree of the node, the higher the centrality, and the more important it is in the network.

[9] Some banks occur less in the news, probably as they receive less media attention or they have few connections with other banks. To mitigate the bias due to the differences in media attention, we try to find a threshold to filter out banks that have received sufficient attention. We find that keeping banks that have been reported by at least three news articles in the sample can well balance the number of sample banks, the represen tativeness of banks, and potential differences in media attention.

[10] Some banks have different names in the history without type changes. We put their historical names into their respective regular expressions rather than treating them as different banks. For example, *Shengjing Bank* used to be named *Shenyang Commercial Bank*. Moreover, *Mengshang Bank* and *Baoshang Bank* share the same bank label in our sample, as the former is established based on the latter after bankruptcy.

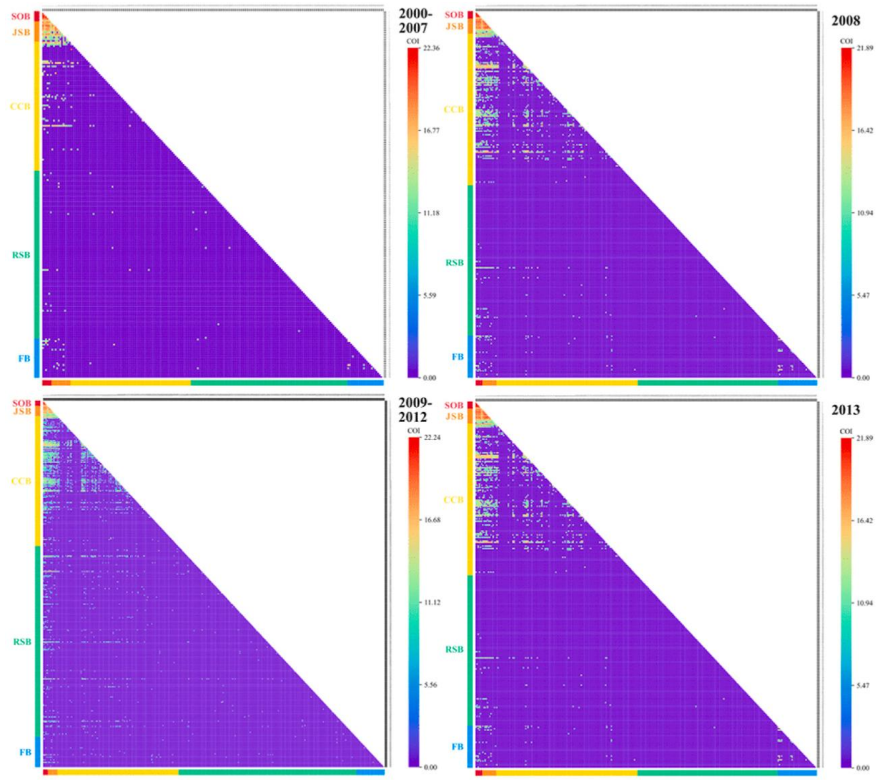
[11] We dynamically adjust the sample year by year, and take the middle of the year as the threshold. Banks that exist between Jan–June of the year will be included in the sample of the current year, while banks that exist between July–Dec of the year will be included in the sample of the next year. Some city commercial banks and small and medium-sized rural banks are restructured from urban/rural credit cooperatives, which are not included in the sample due to the data availability.

Fig. 6 shows that the network has a core-periphery structure,[12] which has an agglomeration and stratification pattern. State-owned banks are in the center, surrounded by joint stock banks, and small and medium-sized banks on the outside, such as city commercial banks and rural small and medium-sized banks. In terms of the node centrality, state-owned banks have the largest degree and the strongest centrality, and they are the core of the network. The degrees are relatively smaller for joint stock banks and city commercial banks, rural small and medium-sized banks, foreign banks and private banks. A large number of rural small and medium-sized banks (mainly village bank and rural credit cooperative) are even isolated in the network. The business complexity and the importance decrease sequentially from state-owned banks to small and medium-sized banks.

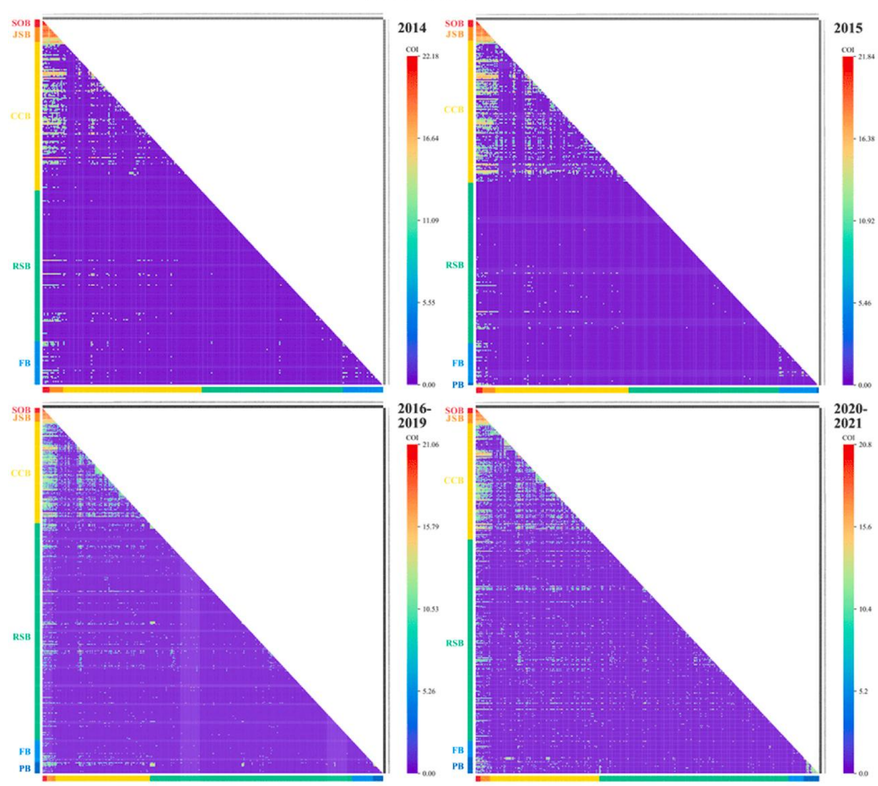
Banks close to the center of the network tend to be larger with stronger business connection among all types of banks. We have circled several representative banks with red boxes. In yellow nodes, we see *Bank of Beijing* (No. 18), *Bank of Shanghai* (No. 52) and *Bank of Jiangsu* (No. 53), which are the top three in total assets among city commercial banks. *Bank of Jinzhou* (No. 36), which is partially owned by *Industrial and Commercial Bank of China* (No. 0), is also located not far from the center of the network. In green nodes, though *Shanghai Rural Commercial Bank* (No. 251), *Shenzhen Rural Commercial Bank* (No. 608) and *Chengdu Rural Commercial Bank* (No. 525) all rank top 10 in total assets among rural small and medium-sized banks, their interconnection characteristics are different. *Shanghai Rural Commercial Bank* shows more connection to city and rural commercial banks, while *Shenzhen Rural Commercial Bank* and *Chengdu Rural Commercial Bank* are closer to state-owned banks. In dark blue nodes, both *HSBC Bank* (China) (No. 613) and *Standard Chartered Bank* (China) (No. 639), top two of foreign banks in terms of branch number, are closely related to state-owned banks. In blue nodes, *Zhejiang E-Commerce Bank* (No. 150), the largest private banks in China, exhibits stronger connection with other types of banks than with private banks.

---

[12] This structure also exists in interbank networks in many other countries, such as Italy (Bargigli et al., 2015), Germany (Craig & Von Peter, 2014), and the Netherlands (van Lelyveld, 2014). It is interesting and perplexing that financial institutions form a core-periphery network structure, which is widely believed to increase financial system instability rather than promote adequate risk sharing. Theoretical studies have proposed multiple mechanisms for such network structure. Lux (2015) show that the core-periphery structure is naturally derived from a banking system with het erogeneous balance sheet size. Altinoglu and Stiglitz (2023) suggest that the implicit government guarantee renders small banks with tighter interconnectedness to large banks through inter-bank lending, which leads to the core-periphery structure and excessive risk-taking.



a. 2000-2013



b. 2014-2021

(caption on next page)

Fig. 7. Annual inter-bank connection heat maps. As the co-occurrence matrix is symmetric, the heat map shows only the lower triangular region. Banks which do not co-occur with other banks in one period are not in the heat map for that period. The bank number for each block is marked at the top and right, and the bank type for each block is marked at the left and bottom. SOB (red-colored), JSB (orange-colored), CCB (yellow-colored), RSB (green-colored), FB (blue-colored) and PB (dark-blue-colored) refer to state-owned banks, joint stock banks, city commercial banks, small and medium-sized rural banks, foreign banks and private banks respectively.

4.2. Dynamic network of various types of banks

In order to visualize the dynamic relationship among various banks, we draw heat maps based on co-occurrence matrix. Fig. 7 displays the interconnections among various types of banks, such as state-owned banks (SOB), joint stock banks (JSB), city commercial banks (CCB), rural small and medium-sized banks (RSB), foreign banks (FB) and private banks (PB).

The heat maps follow the rainbow color system from red to blue, which represents the gradual decrease in the co-occurrence index.[13] Areas with brighter colors in the heat map indicate tighter inter-bank connections. Only banks connected to at least one of other banks are displayed in heat maps due to too large a sample size, and the maximum value represented by red is not always consistent across time as the range of co-occurrence frequency varies over the time.

Fig. 7 shows that the internal connectedness characteristics of large banks and small and medium-sized banks have similar trends from 2000 to 2021. The inter-bank connectedness reaches the peak during four shocks: global financial crisis in 2008, credit crunch in the inter-bank market in 2013, the stock market crash in 2015, and COVID-19 in 2020, which are manifested in the expansion of bright-colored area (especially the red area) in the heat maps. The tightness of inter-bank connection tends to decrease after these events, which is manifested in the expansion of cool-colored areas.

There is some heterogeneity in the connectedness across bank types. Although the internal connections of state-owned banks and joint stock banks change dynamically, the extent of the change is not significant. The heat map of each period is mainly composed of bright-colored blocks, which indicates that the two types of banks are in high connectedness in general, and the connectedness difference between bank pairs is relatively small. However, the proportion of dark-colored blocks in the heat map is significantly higher than that of the first two types of large banks for many small and medium-sized banks, and the maximum connection tightness of bank pairs is also smaller than that of large banks. Most city commercial banks do not have connections with each other before 2009, which is due to geographical restrictions by the regulation. However, the regulatory authorities relax the restrictions on the branch footprint of city commercial banks since 2009. Thus, city commercial banks start inter-city operations and their internal connections increasingly tighten. This trend was further accelerated by a rapid growth of interbank business after 2013, i.e., see the expansion of bright-colored areas in

[13] To be precise, it is equal to multiplying the co-occurrence index by 1,000,000 and taking the logarithm of 2 to balance the large denominator of COI formula.

the upper-left corner of the heat map.[14] Additionally, the increase in inter-connectedness of city commercial banks is also partly attributed to their utilization of shadow banking products. Hachem and Song (2021) show that small banks are more constrained by liquidity requirements than large banks. Therefore, city commercial banks have increasingly utilized shadow banking products with stricter liquidity regulations after 2009, such as non-guaranteed wealth management products (WMPs) and trust loans in order to bypass regulatory restrictions (Hachem, 2018; He & Wei, 2023). Furthermore, city commercial banks tend to connect more frequently with state-owned banks and joint-stock banks up until 2012, i.e., Appendix 2 shows a detailed visualization. After 2013, however, they rapidly increase their connections with other small and medium-sized banks, which is consistent with their expansion.

In contrast with city commercial banks, the internal connection network of rural small and medium-sized banks is sparse. It is mainly due to severe restrictions on the inter-city operation of rural small and medium-sized banks. In addition, the inter-bank and off-balance sheet business has encountered strict regulation since 2016, and it is more difficult for rural small and medium-sized banks to form a close internal network, which is reflected by few bright-colored blocks in the central part of the heat map. In addition, bright-colored area in the lower right corner of the heat map expands rapidly, i.e., the connection tightness among foreign banks and private banks rises drastically. It corresponds to the acceleration of financial opening-up, liberalizing restrictions on foreign banks, supportive policies on consumer finance and inclusive finance, and the surge of private banks in China since 2018.

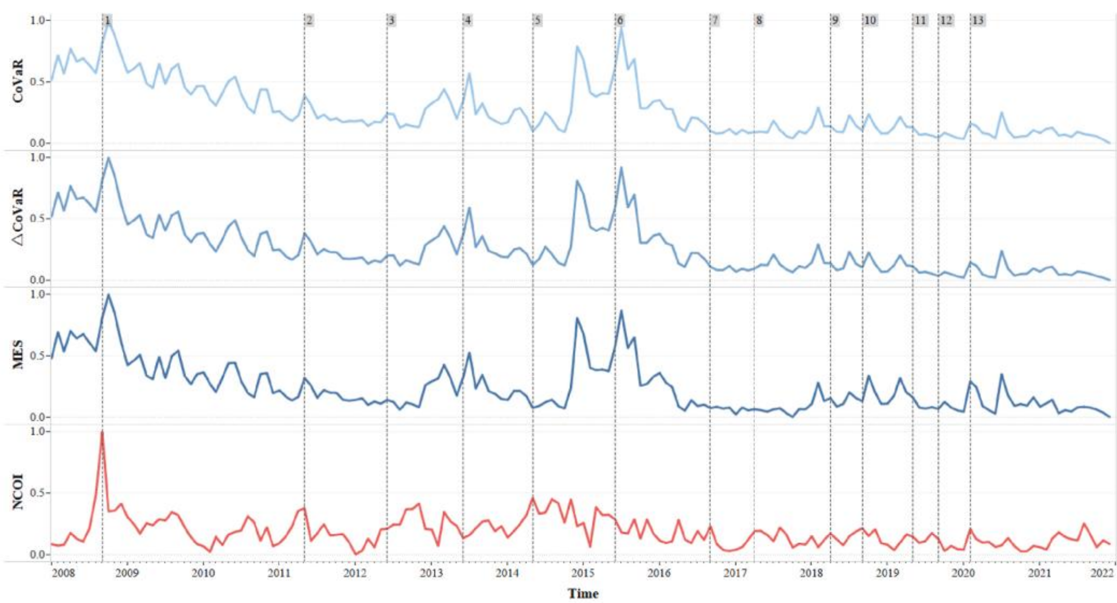


Fig. 8. Comparison of risk indicators.

[14] Unlike state-owned banks, which benefit from stable household deposits and implicit government support, small and medium-sized banks rely heavily on interbank funding, often sourced from large banks such as state-owned banks and some joint stock banks (Acharya et al., 2019; Song & Xiong, 2018). For instance, Industrial Bank's "Yinyin Platform" provides financing and settlement services to hundreds of small and medium-sized banks (Wang et al., 2018).

The internal connection structure of small and medium-sized banks changes from centralization to decentralization around the shocks. For example, only a small number of small and medium-sized banks are highly correlated with each other before the four shocks mentioned above, while other small and medium-sized banks have weak interconnections, i.e., bright-colored blocks are concentrated in some specific regions of the heat map and the value of COI_t on the right side of each heat map is relatively high. However, the difference in the number of bank co-occurrence decreases after these shocks. The connection tightness of strongly correlated banks decreases, while that of weakly correlated banks increases, and the COI_t drops. The number of bright-colored blocks in the heat map is in a significant boost and more evenly distributed in each region, while the inter-bank connection structure has switched from centralization to decentralization.

4.3. Time series analysis of risk indicator

Sections 3.1 and 3.2 show that the co-occurrence index COI_t reflects the actual connectedness tightness between banks. Therefore, $NCOI_t$, which combines sentiment index $Senti_A$ with COI_t, should be able to capture the two driving factors of systemic risk, i.e., negative shock and connection mechanism. We will examine its capability in identifying the risk from the time dimension.

We compare $NCOI$ with traditional risk contribution indexes ($CoVaR_t$; $\Delta CoVaR_t$; MES_t) to investigate whether $NCOI_t$ can capture systemic risk dynamics around shocks.[15] We set the time range between 2008 and 2020, and all indicators have been standardized. Fig. 8 shows that $NCOI_t$ exhibits large fluctuations during the global financial crisis (Event 1), credit crunch in the interbank market (Event 4), the stock market crash (Event 6), COVID-19 (Event 13) and other shocks in Table 2, which is consistent with the traditional indicators of systemic risk around shocks.

However, the peaks of various indicators in different periods are not fully consistent with each other. The peaks that traditional risk indicators reach during the stock market crash equal to that during the global financial crisis, while the peak of $NCOI_t$ in the global financial crisis is significantly higher than that in stock market crash. Traditional risk indicators mostly use stock market data, which often reflects the linkage among listed banks and is sensitive to banks' stock market performance. Traditional risk indicators may overestimate the risk contagion of the stock market crash as unlisted banks' risk are not represented. Instead, $NCOI_t$ has reflected multilevel inter-connectedness which accounts for many unlisted banks. Thus, the far-reaching effect of the global financial crisis are by no means the same as those of the stock market crash in China.

---

[15] $\Delta CoVaR$ captures risk contribution of institution i relative to institution j or the entire system (Here we omit the subscript of t). The former refers to pairwise spillover effect, while the latter refers to overall systemic impact of institution i. Specifically, $CoVaR_q^{j|i}$ is the value at risk (VaR) of institution j in context of some event $S(X^i)$ of institution i at the level of q, that is, $\mathbb{P}(X^j \leq -CoVaR_q^{j|i} | S(X^i)) = q$. Tobias and Brunnermeier (2016) define institution i's contribution to j at the level of q as $\Delta CoVaR_q(i, j) = CoVaR_q^{j|X^i=VaR_q^i} - CoVaR_q^{j|X^i=Median^i}$, where $CoVaR_q^{j|X^i=VaR_q^i}$ is for institution j given the distress of institution i, while $CoVaR_q^{j|X^i=Median^i}$ is for institution j given institution i in its normal conditions. In practice, q = 5%. The formulation highlights the impact of i's extreme events on j. MES measures the marginal contribution of institution i to systemic risk, as measured by the expected shortfall (ES) of the system (Acharya et al., 2017).

Table 2
Summary of shock events from 2008 to 2021.

Index	Time	Description
1	Sep-08	Global financial crisis
2	May-11	Inflationary pressure increases, and the central bank raises the reserve ratio five times in a row, which increases the pressure on bank credit
3	Jun-12	"Commercial Bank Capital Management Measures (Trial)"
4	Jun-13	Credit crunch in the inter-bank market
5	May-14	"Notice on Regulating Inter-bank Business Governance of Commercial Banks" issued by the CBRC to restrain the maturity mismatch of banking institutions and strengthen liquidity management
6	Jun-15	Stock market crash
7	Sep-16	Shanghai Stock Exchange tightens bond issuance by real estate companies, and banks tighten project loans to real estate companies
8	Apr-17	"Notice on Effectively Making Up for Regulatory Shortcomings and Improving Regulatory Efficiency"
9	Apr-18	"New Asset Management Regulations"
10	Sep-18	"Measures for the Supervision and Administration of the Wealth Management Business of Commercial Banks" issued by the CBIRC, which is a supporting rule for the "New Asset Management Regulations" to refine the regulatory requirements for wealth management products
11	May-19	*Baoshang Bank* is taken over by the central bank and the CBIRC, which entrust *China Construction Bank* to manage its business; *China Merchants Bank*'s "Qianduan" APP exploded.
12	Sep-19	"Administrative Measures for the Net Capital of Wealth Management Subsidiaries of Commercial Banks (Trial) (Draft for Comment)" issued by the CBIRC, which proposes the net capital requirements of wealth management subsidiaries and rectifies the shadow banking business.
13	Feb-20	COVID-20 (The peak of the epidemic in China)

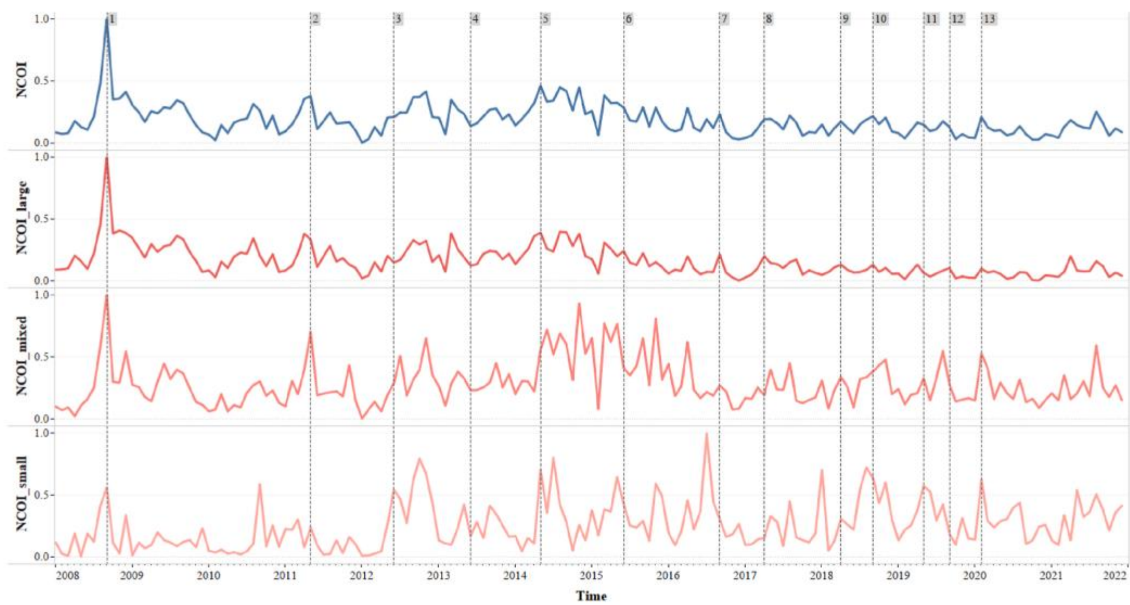


Fig. 9. Decomposition of co-occurrence index *NCOI* by bank types.

In addition, $NCOI_t$ captures the impact of policy events on banks' business activities and operation strategies, while it rises in periods of the event 2, 5, 8, 9, 10 and 12. For example, the "*New Asset Management Regulations*" (i.e., new AMP Rules) are aimed to curtail the shadow banking sector, which has been expanding rapidly since 2010,[16] and created huge risk spillovers in the banking system. Thus, when the regulatory authorities tighten the supervision rules, the risks erupt quickly and form a contagion effect in the

[16] The economic growth has slowed down as the regulatory authorities turned to tightening policies since 2010. In particular, the real estate industry, local government financing platforms and SMEs have a severe shortage of funds, which has stimulated the development of the shadow banking sector. In addition, when the deposit reserve ratio is high, banks have great incentives to use financial innovation to make use of regulatory arbitrage, which further promotes the shadow banking sector. According to the CBIRC, the total size of the shadow banking sector exceeded 90 trillion RMB up to 2016 in China, which endangers the stability of the banking sector in China.

bank network. Such connectedness originated from similar business mode is difficult to be revealed directly with market data, and the fluctuations of traditional risk indicators are not significant over the same period. Thus, news-based *NCOIt* can indeed capture more comprehensive inter-bank connections that include information on stock market, operation performance, and business activities, etc.

We decompose *NCOI* by bank types, i.e., $NCOI_{large}$ considers the connections among state-owned and joint stock banks; $NCOI_{small}$ considers the connections among small and medium-sized banks; $NCOI_{mixed}$ considers both connections among small and medium-sized banks, and pair-wise connections between state-owned (joint stock banks) and small and medium-sized banks.

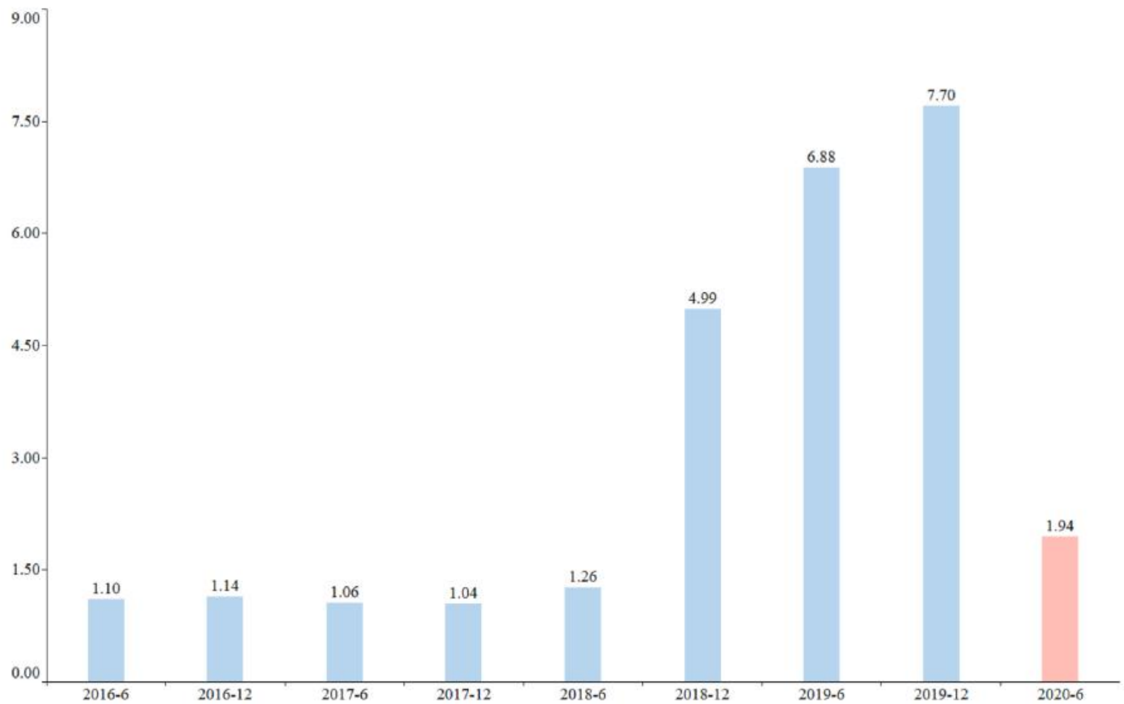


Fig. 10. Non-performing loan ratio of *Bank of Jinzhou* from 2016 to 2020 (%).

Fig. 9 shows the trend of *NCOI* and its decomposition during 2008–2021. $NCOI_{large}$ captures the risk diffusion by the connection among large banks; $NCOI_{small}$ captures the risk diffusion by the connection among small and medium-sized banks; $NCOI_{mixed}$ captures the risk spillovers by the connection between small and medium-sized banks and large banks. As small and medium-sized banks face strict restrictions on international business before the global financial crisis, the shock matter less for them than large banks, i.e., $NCOI_{large}$ has a peak in this period. With the expansion of inter-city operation, interbank businesses, and shadow banking activities of small and medium-sized banks, both $NCOI_{mixed}$ and $NCOI_{small}$ rise dramatically over the time. The impact of the shock on small and medium-sized banks become more pronounced, such as the "Commercial Bank Capital Management Measures (Trial)" released in June 2012 (event 3), which raises regulatory capital standards and renders small and medium-sized banks with low capital adequacy ratios face higher risks. Regulatory pressure makes $NCOI_{small}$ rise slightly although $NCOI_{large}$ has little change in the same period. During the stock market crash, small and medium-sized banks become a main source of risk contributor. The relative fluctuation of $NCOI_{mixed}$ and $NCOI_{small}$ is larger than that of $NCOI_{large}$ in the same period. The tightened regulations such as restrictions on related-party transactions and WMPs, contribute to the fluctuation of $NCOI_{small}$ after 2016. While these regulations temporarily reduce risk-taking, our findings suggest that small and medium-sized banks continue to face challenges due to their reliance on interbank funding and shadow banking, which is supported by persistently high magnitudes of $NCOI_{mixed}$ and $NCOI_{small}$. In contrast, large banks exhibit a low and stable magnitude of $NCOI_{large}$ over the past decade, underscoring their relatively conservative risk profiles. This divergence highlights the systemic risk contributed by business structures and regulatory arbitrage of small and medium-sized banks. When the credit risk of *Baoshang Bank* bursts in May 2019 (event 11), which has more than 60 % of inter-bank counterparties as small and medium-sized financial institutions, $NCOI_{small}$ also witness a rapid growth in the same period. It suggests that *NCOI* has capability in identifying systemic risk, which reveals the instability of small and medium-sized banks versus large banks.[17]

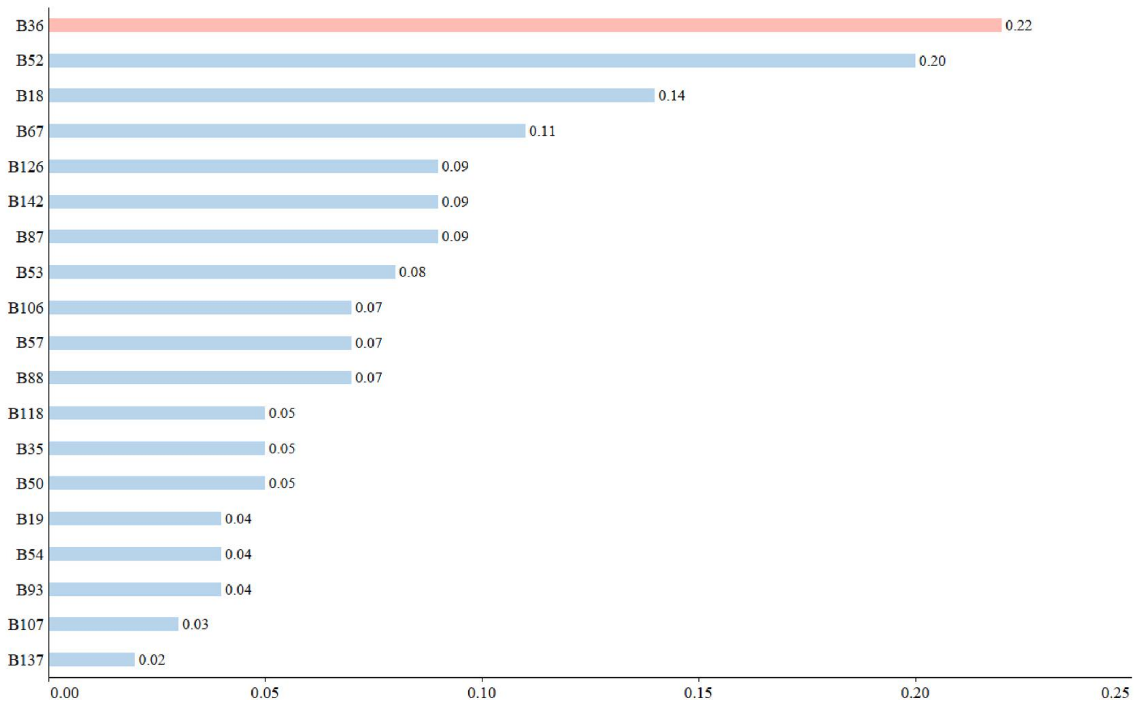


Fig. 11. Inter-bank liabilities/interest-bearing liabilities of listed city commercial banks in June 2018. B36 is the *Bank of Jinzhou*.

[17] Recent studies also suggest an increasingly important role of small and medium-sized banks in systemic risk. For example, Wang et al. (2018) explore a risk contagion network of 24 financial institutions and find that small banks such as Bank of Beijing and Bank of Ningbo are detected as systemic risk emitters due to their huge risk spillovers, particularly from their rapidly developing shadow businesses. Huang et al. (2019) show that the average MES of three city commercial banks (Bank of Beijing, Bank of Nanjing, and Bank of Ningbo) even exceeds that of large state-owned banks during 2008–2014. Shi et al. (2022) find that most of top 10 vulnerable banks in 2019 and 2020 are city commercial banks and rural commercial banks, including Bank of Handan, and Bank of Jiangsu, and Qingdao Rural Commercial Bank, primarily due to their common exposures to specific asset types. Notably, banks highlighted in these studies are also included in our sample. However, our study offers a more comprehensive examination of small and medium-sized banks, utilizing a much larger sample of 711 banks over 2000 to 2021.

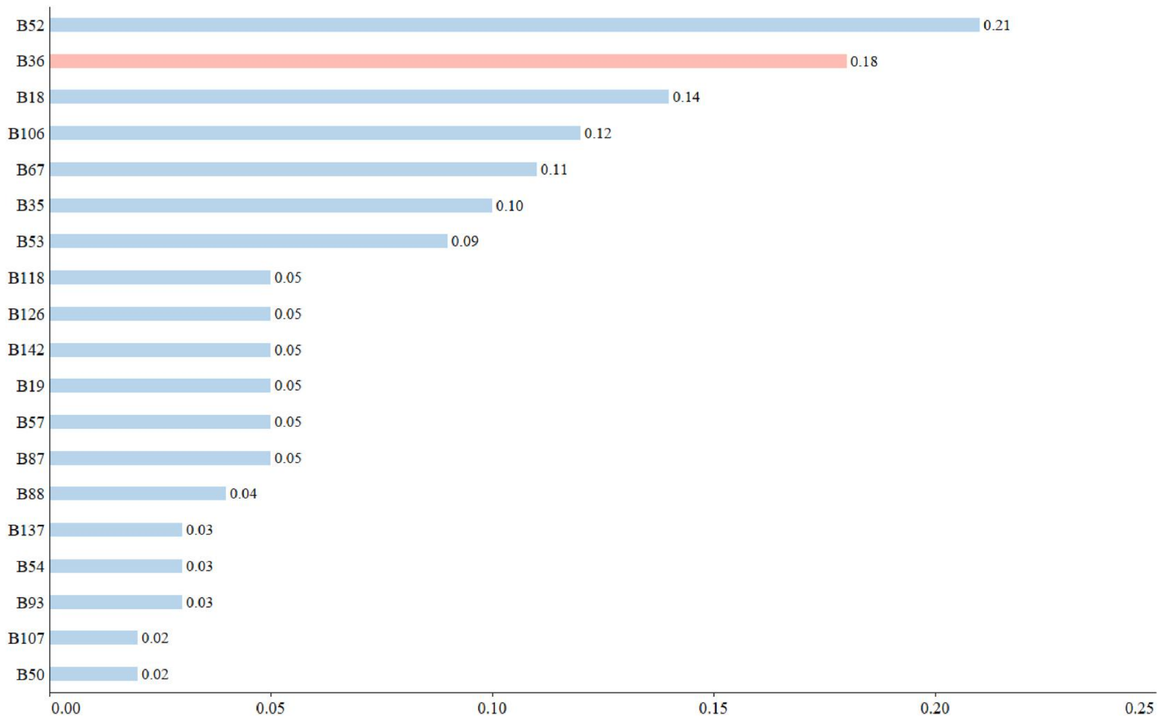


Fig. 12. Inter-bank liabilities/interest-bearing liabilities of listed city commercial banks in June 2019. B36 is the *Bank of Jinzhou*.

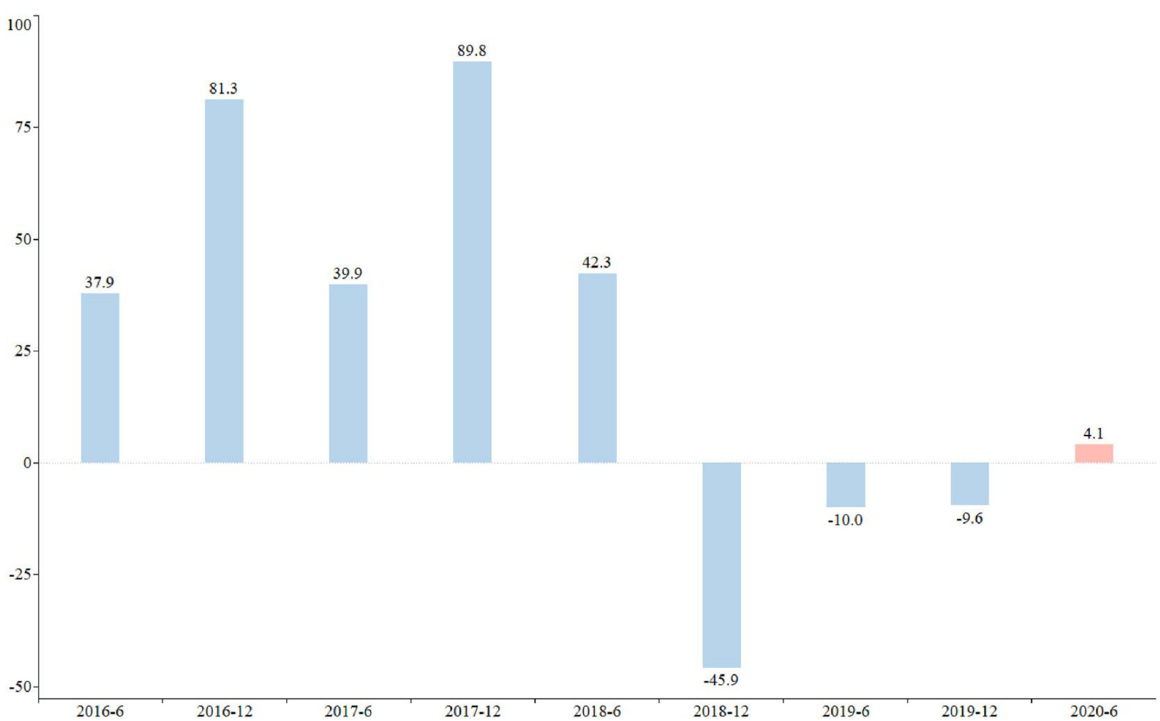


Fig. 13. Net profit of *Bank of Jinzhou* from 2016 to 2020 (Unit: 100 million RMB).

5. Further analysis of the measurement

News co-occurrence can capture inter-bank connection from both static and dynamic perspectives. News sentiment index *Senti* is introduced to construct the risk indicator $NCOI_t$ that considers the two systemic risk drivers of negative shock and interconnection. By comparing $NCOI_t$ with traditional risk indicators and decomposing $NCOI_t$ by bank types, we find that $NCOI_t$ has capability in identifying the systemic risk. Thus, we will conduct an in-depth analysis of the effectiveness of $NCOI_t$ from the time and space dimensions.

5.1. Time dimension: a case study at the individual level

The capability of $NCOI_t$ in identifying systemic risk is in the banking sector. However, the precise handling of high-risk banks is an arduous task in preventing and resolving financial risks. Thus, it is necessary to test the validity of $NCOI_t$ at the individual bank level. If $NCOI_t$ changes in accordance with financial indicators, it indicates that $NCOI_t$ can capture the risks of individual banks timely in a precise way.

We focus on the *Bank of Jinzhou*, which has witnessed a risk event in the past three years. *Bank of Jinzhou* is the second largest city commercial bank in *Liaoning* province while *Shengjing Bank* has the largest assets. In June 2019, a serious inter-bank squeeze broke out in *Bank of Jinzhou*, while the *People's Bank of China* pushed forward the restructuring of the bank, such as the introduction of strategic investors, management turnover, and asset disposal, etc. The asset quality and profitability of the bank recovered soon, and the "squeeze" crisis and risk contagion were alleviated in time as well. With the capital increase in Oct 2020, *Bank of Jinzhou* achieved the intended goal in the restructuring.

The risk event of *Bank of Jinzhou* originated from the default of *Baota Petrochemical Group* in July 2018. As a key customer of *Bank of Jinzhou*, the default of *Baota Petrochemical Group* has a great negative impact on the capital and asset quality of *Bank of Jinzhou*. The deployment of the restructure work in August 2019 is the threshold for the two periods: before the restructure (2018.7–2019.8) and restructure (2019.9–2020.10):

Before the restructure is the period of risk generation and diffusion of *Bank of Jinzhou*, which can be divided into two stages. One stage is the initial shock stage (2018.7–2019.4). In July 2018, the default of *Baota Petrochemical Group* serves as the risk fuse, which exerts a negative impact on the capital and asset quality of the bank, and leads to a sharp rise in non-performing loan ratio. Fig. 10 shows that non-performing loan ratio of *Bank of Jinzhou* is low during Jan 2016 and June 2018, while it skyrocketed to 4.99 % at the end of 2018 due to the default of *Pagoda Petrochemical* notes. The other stage is the risk overflow stage (2019.5–2019.8). Ernst & Young resigned from its audit business for the bank at the end of May 2019, and the non-performing loans are gradually revealed. As *Bank of Jinzhou* relies heavily on the inter-bank funds, its risk exposure triggers the advance withdrawal of loans by many banks, which leads to a liquidity crisis. Figs. 11 and 12 show the inter-bank capital dependence of all listed city commercial banks in China including *Bank of Jinzhou* in 2018 and 2019,[18] i.e., inter-bank capital dependence is the inter-bank liabilities over interest-bearing liabilities. Before the liquidity crisis, the inter-bank capital dependence of *Bank of Jinzhou* is as high as 0.22, ranking the first

[18] The interbank liability data of Harbin Bank's semi-annual report is missing and is replaced by the annual report data.

among listed city commercial banks. Due to the withdrawal of loans by many banks, its inter-bank funds holdings fall sharply afterwards, which drops to 18 % of the inter-bank capital dependence.

Restructure is the period of risk resolution of *Bank of Jinzhou* (2019.9–2020.10). The bank has adopted various measures to address non-performing loans, such as introduction of strategic investors, replacement of management, disposal of non-performing loans, and private placement to supplement capital, which improves the asset quality and profitability. Fig. 13 shows that *Bank of Jinzhou* achieves a net profit of approximately 410 million RMB in the first half of 2020 after two consecutive years of losses. At the same time, Fig. 10 shows that non-performing loan ratio drops to 1.94 % in the same period, and its asset quality has risen as well. In short, bank restructure has contained the risks accordingly.

If the risk indicator $NCOI_t$ constructed by news co-occurrence and sentiment analysis can identify risks, we should observe the risk and spillover in various stages. In the initial shock stage, as *Bank of Jinzhou* is affected by the default of its key client, the news attention (i.e., the number of times reported in news, OI_t) should increase. However, the risk indicator $NCOI_t$ may only increase mildly as the risk is contained within the bank. In the risk spillover stage, OI_t may still increase over time as banks suffer fund shortage; the risk may spread through the inter-bank connection due to fund withdrawal by related banks, and thus $NCOI_t$ will reach a peak. In the restructure stage, OI_t drops to a normal level with the risk disposal and the improvement of asset quality, and $NCOI_t$ may decline as well.

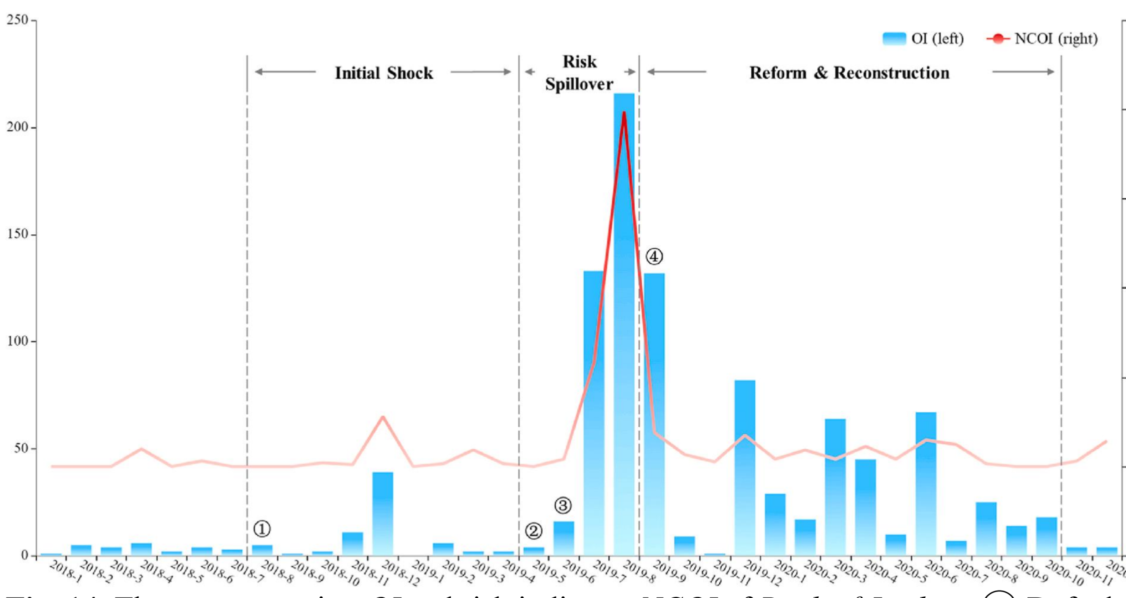


Fig. 14. The news attention *OI* and risk indicator *NCOI* of *Bank of Jinzhou*. ① Default of *Baota Petrochemical Group*; ② Resignation of *Ernst & Young* for the audit business; ③ Withdrawal of loans by related banks; ④ Full deployment of the restructure.

Fig. 14 shows the pattern of OI_t and $NCOI_t$ for *Bank of Jinzhou* from 2018 to 2020. In the initial shock stage, the news attention OI_t rises after the default of *Baota Petrochemical Group* in July 2018 while $NCOI_t$ rises mildly, which substantiates Hypothesis 1. In the risk spillover stage, the news attention OI_t increases for four consecutive months due to the resignation of *Ernst & Young*. At the same time, the risk indicator $NCOI_t$ shows an upward trend and reaches a local peak in August 2019, which is consistent with a risk spread across the bank network and substantiates Hypothesis 2. During the restructure stage, the news attention OI_t and risk indicator $NCOI_t$ show a declining trend, which suggests that the restructure has improved asset quality and alleviated bank risk and spillover effect. However, the market remains concerned about the bank risk, and OI_t and $NCOI_t$ do not revert to their *ex-ante* level in the short term.[19]

In addition, it reveals the necessity of real-time and high-frequency risk monitoring indicators. If the risk is only examined from the financial indicators such as non-performing loan ratio, it is often subject to limited frequency and disclosure time of financial statements. In contrast, risk indicators based on high frequency news can help monitor risk in the supervision.

5.2. Time dimension: news-based risk indicator NCOI

Existing research has highlighted the relationship between bank characteristics and systemic risk measured by traditional indicators such as $\Delta CoVaR$, *MES* and *SRISK*. Anginer et al. (2018) disclose potential ways that bank size affects systemic risk. De Jonghe et al. (2015) explore the effect of annual growth in bank assets and the ratio of deposits to assets on the systemic risk. Ellul and Yerramilli (2013) show a positive connection between bank risks and capital ratios. Zhang et al. (2021) show that capital adequacy ratio and return on assets mitigate the systemic risk contribution of banks. van Oordt & Zhou (2019) find no strong relationship between the proportion of non-interest income and tail risks of banks. Brunnermeier et al. (2020) examine the relationship between the proportion of non-interest income, liquidity ratio, non-performing loan ratio of banks and the systemic risk.

Table 3

Two-way fixed-effect estimates on the determinants of $NCOI_t$. $NCOI_t$ is multiplied by 100. Log assets is the natural logarithm of total assets. Growth in assets is the annual growth in total assets. Capital adequacy ratio is the total capital adequacy ratio. NPL ratio is the non-performance loan ratio. Liquidity ratio is the ratio of liquid assets over total assets. ROAA is return on average assets. Non-interest income share is the ratio of non-interest income over total assets. The inflation rate is calculated based on CPI. Significance ***, 0.01; **, 0.05; *, 0.10.

[19] According to Alibaba's judicial auction platform, on Feb 6, 2021, 90 million shares of Bank of Jinzhou held by Jincheng International Logistics Group and 6 million shares held by Dalian Changxingdao Greentown Development were put on auction, but ultimately failed due to a lack of bid.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log assets$_{t-1}$	30.2**	34.6***	29.1**	58.6**	30.5***	32.7***	30.4***	65.6***
	(2.57)	(2.73)	(2.47)	(2.48)	(2.63)	(2.84)	(2.66)	(2.83)
Growth in assets$_{t-1}$		0.029						0.077
		(0.53)						(0.72)
Capital adequacy ratio$_{t-1}$			0.20***					−2.56
			(2.69)					(-1.62)
NPL ratio$_{t-1}$				7.05**				0.99
				(2.11)				(0.28)
Liquidity ratio$_{t-1}$					0.11			−0.84
					(0.41)			(-1.61)
ROAA$_{t-1}$						−8.55**		−42.7**
						(-2.33)		(-2.28)
Non − interest income share$_{t-1}$							−3.83	7.96
							(-0.70)	(0.96)
Constant	−801.6***	−901.8***	−610.2**	−1187.5**	−813.5***	−856.8***	−805.9***	−1141.0**
	(-2.88)	(-2.97)	(-2.17)	(-2.08)	(-2.96)	(-3.16)	(-2.98)	(-1.99)
Bank fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
N	1431	1267	1324	825	1430	1429	1431	742
R-sq	0.12	0.15	0.12	0.19	0.12	0.13	0.12	0.25
adj. R-sq	0.11	0.14	0.11	0.18	0.11	0.12	0.11	0.23

We include these common factors to conduct an analysis on the determinants of *NCOI*, although the impact of some factors on systemic risk are still controversial. We obtain annual financial data of 223 banks from *Bank Focus* over 2005 and 2021,[20] and employ two-way fixed-effect models for the determinants of $NCOI_t$ in Table 3.[21] We find that bank size is significant at conventional levels in all columns, which suggests that it is an important risk factor (Laeven et al., 2016). We add other variables one-by-one to determine the risk factors that can explain $NCOI_t$. Capital adequacy ratio, NPL ratio and ROAA are all significant although the annual growth in assets, liquidity ratio and non-interest income share show limited effect on $NCOI_t$. In order to alleviate the bias due to missing variables, we add these factors into the regression and find that the signs of the coefficients of most factors are as expected. Liquidity ratio and ROAA are all negatively correlated with $NCOI_t$, and NPL ratio is positively correlated with $NCOI_t$. Capital adequacy ratio is positively correlated with $NCOI_t$, which is consistent with Ellul and Yerramilli (2013). However, some factors are no longer significant when we put all variables together in a single regression for $NCOI_t$. Finally, we do not find any significant impact of non-interest income share on the systemic risk, which supports the finding of van Oordt & Zhou (2019). In short, our indicator $NCOI_t$ is driven by similar determinants at the bank level as other traditional risk indicators (see Table 3).

In addition, we repeat the above analysis on two sub-periods following the global financial crisis, i.e., 2009–2017 and 2018–2021, to investigate the changes of the risk drivers after 2018 due to the introduction of the "New Asset Management Regulations". These regulations aimed at addressing key risks in shadow banking, reducing regulatory arbitrage, and enhancing overall financial stability. The results from the two sub-periods highlight a significant evolution in the key drivers of systemic risk over the time. Before 2018, systemic risk is predominantly driven by bank size, capital buffer, and profitability, with larger banks playing a more crucial role in banking stability. However, after 2018, marked by the introduction of the "New Asset Management Regulations", asset quality emerges as a more critical factor in driving systemic risk. The diminished significance of bank size, capital buffer, and profitability reveals a structural change in the banking sector following stricter regulatory measures, which suggests the increasingly importance

[20] See Appendix 1 for the list of banks in the regression.

[21] To deal with outliers, we eliminate top (bottom) 1 % of observations for *NCOI* in each regression.

of small and medium-sized banks for the stability of the banking system. Appendix 3 provides more details.

5.3. Cross-sectional dimension: systemic important banks

Systemically important banks are more linked with other banks, which are at the center of the systemic risk contagion network. In short, the systemic importance of a bank depends on the number of banks with connection and the connectedness tightness. Therefore, the systemic importance of bank *i* is defined as:

$$NCOI_t^i = \sum_{j=1}^{N_t} CM_t^{neg}(i,j) \tag{5}$$

where *t* is the time; N_t is the number of sample banks at time *t*; $CM_t^{neg}(i,j)$ is the element in (*i, j*) of the co-occurrence matrix CM_t^{neg} at time *t*, which only considers negative news; $NCOI_t^i$ is the systemic importance of bank *i* at time *t*. We use $NCOI_t^i$ to distinguish $NCOI_t$ of a single bank from the whole banking system.

Table 4

Ranking of systemic importance of banks during 2000–2021. The color of the cell corresponds to bank types: red for state-owned banks, orange for joint stock banks, yellow for city commercial banks, green for rural small and medium-sized banks, blue for foreign banks, and dark blue for private banks.

Rank	2000–2007	2008	2009–2012	2013	2014	2015	2016–2019	2020–2021
1	B2	B0	B0	B2	B2	B0	B0	B0
2	B0	B2	B2	B0	B0	B2	B2	B3
3	B3	B3	B3	B3	B1	B1	B3	B2
4	B1	B8	B1	B1	B3	B3	B1	B1
5	B8	B1	B4	B8	B8	B4	B8	B8
6	B4	B4	B8	B10	B4	B8	B4	B4
7	B10	B13	B10	B4	B13	B13	B6	B5
8	B9	B9	B13	B13	B10	B6	B9	B13
9	B11	B6	B7	B12	B12	B10	B10	B12
10	B13	B10	B9	B9	B9	B9	B13	B6
11	B7	B18	B6	B7	B6	B12	B12	B9
12	B52	B52	B18	B6	B7	B7	B7	B7
13	B6	B11	B11	B18	B18	B18	B53	B10
14	B18	B54	B137	B11	B137	B54	B52	B137
15	B650	B7	B54	B106	B11	B11	B11	B53
16	B14	B137	B52	B54	B54	B137	B137	B52
17	B54	B650	B12	B52	B53	B5	B5	B144
18	B137	B26	B14	B137	B52	B53	B54	B11
19	B649	B106	B57	B14	B5	B142	B18	B14
20	B12	B5	B5	B67	B14	B14	B16	B142
21	B106	B14	B106	B53	B17	B16	B57	B67
22	B19	B78	B53	B5	B57	B52	B14	B57
23	B124	B16	B50	B97	B106	B67	B142	B16
24	B17	B57	B136	B57	B67	B17	B118	B251
25	B16	B101	B67	B136	B97	B57	B88	B18
26	B94	B19	B650	B63	B50	B50	B93	B54
27	B96	B15	B63	B107	B35	B106	B67	B74
28	B74	B67	B107	B516	B15	B35	B107	B107
29	B67	B12	B94	B525	B167	B150	B17	B140
30	B107	B251	B516	B35	B16	B74	B15	B15
31	B620	B72	B35	B142	B36	B36	B516	B93
32	B141	B49	B15	B118	B144	B97	B35	B106
33	B666	B107	B74	B144	B118	B15	B19	B464
34	B53	B50	B19	B31	B107	B151	B36	B58
35	B63	B127	B144	B464	B142	B20	B106	B516
36	B105	B167	B613	B36	B283	B149	B70	B17
37	B251	B613	B251	B15	B19	B144	B124	B127
38	B659	B614	B96	B50	B80	B152	B464	B55
39	B5	B62	B97	B19	B516	B88	B50	B87
40	B93	B63	B16	B96	B74	B251	B140	B124

Table 5

Structural parameters of regressions in each period. This table presents the estimates of structural parameters in Eq. (8) and Eq. (9). For computing the Diebold-Yilmaz spillover matrix, we utilize a 12-week-ahead forecast error variance decomposition. Significance: ***, 0.01; **, 0.05; *, 0.10.

Time	β_{dy}	β_{dcovar}
2017 Q4	0.24***	−0.08**
2018 Q1	−0.08	0.53***
2018 Q2	0.01	0.13***
2018 Q3	0.15**	1.05***
2018 Q4	−0.01	0.67***
2019 Q1	0.12	1.01***
2019 Q2	0.03	0.10
2019 Q3	0.04	0.71***
2019 Q4	0.11**	−0.09
2020 Q1	0.18***	0.58***
2020 Q2	0.59***	0.67***

Following Brownlees and Engle (2017) on the effectiveness of risk measurement, we verify whether $NCOI_t^i$ are effective in the spatial dimension in terms of identifying systemically important banks. The Financial Stability Board (FSB) identifies systemically important banks on five dimensions, such as cross-border activities, scale, relevance, substitutability, and complexity. Therefore, an ideal systemic risk indicator should capture the above five factors in the identification of the systemic importance of banks. However, we only discuss the effectiveness of risk indicator from the dimensions of scale and relevance[22] due to limited data availability. For example, if the systemic importance of a bank with a large asset is higher than that of a small bank, it suggests that the risk indicator is valid in the spatial dimension.

$NCOI_t^i$ has included the systemic importance of interconnection. Although the construction of *NCOI* does not contain the size of banks, the identification of systemically important banks based on $NCOI_t^i$ can reflect the factor of scale. We integrate the ranking of the systemic importance of banks in each year and divides the top 40 banks into 3 groups. Table 4 shows that $NCOI_t^i$ have a clear pattern. Although the scale factor is not included in the indicator construction, $NCOI_t^i$ is still able to capture the scale feature, which shows that *Industrial and Commercial Bank of China*, *Agricultural Bank of China*, *Bank of China*, *China Construction Bank*, *Bank of Communications*, and *China Merchants Bank* are ranked in the top of systemic importance. The second tier of banks ranked in top 7–20 are joint stock banks and a few large city commercial banks. The third tier of banks ranked 21 to 40 mainly consist of small and medium-sized banks such as city commercial banks, rural small and medium-sized banks, foreign banks and private banks.

In contrast with state-owned and joint stock banks, the rankings of small and medium-sized banks are quite unstable. *Bank of Jinzhou* (B36) and *Baoshang Bank* (B144,

[22] Actually, several theories support the view that large and complex banks contribute to systemic risk. One explanation is that large banks tend to take wide-range risks in business activities and rely more on short-term debt, resulting in more vulnerabilities to generalized liquidity shocks and market failures such as liquidity shortages and fire sales (Boot & Ratnovski, 2016; Greenwood et al., 2015; Shleifer & Vishny, 2010). Another explanation is that large banks' moral hazard behavior making them engage more in high-risky activities because of their expectation on gov ernment bailouts (Farhi & Tirole, 2012). Moreover, Laeven and Levine (2007) point out that large banks' risks also derive from awful corporate governance faced with their complex business activities.

reorganized into *Mengshang Bank* since 2020), which shows up as high-risk banks with an increasing ranking of systemic importance as early as 2013, and continues to climb up the rankings before their risk events during 2016 and 2019. It is also true for *Evergrowing Bank* (B17, with chaotic operations and tight liquidity from 2016 to 2019) and *Bohai Bank* (B15, violations of regulatory rules in 2010 and the associated sanctions). It substantiates the varying systemic importance of small and medium-sized banks, which suggests the necessity of covering small and medium banks in the supervision of systemic risk in the country.

5.4. Cross-sectional dimension: systemic risk spillovers

In order to substantiate the validity of risk indicator *NCOI* in the spatial dimension, we use panel regression to investigate whether the co-occurrence matrix can explain the risk spillover matrix of traditional risk measurement, such as Diebold-Yilmaz (DY) pairwise spillover index, denoted by $S_{ij,t}^{g}(H)$ with a forecast horizon of H steps, and $\Delta CoVaR$, which reflects the connection bank risks with the stock price data of banks.[23]

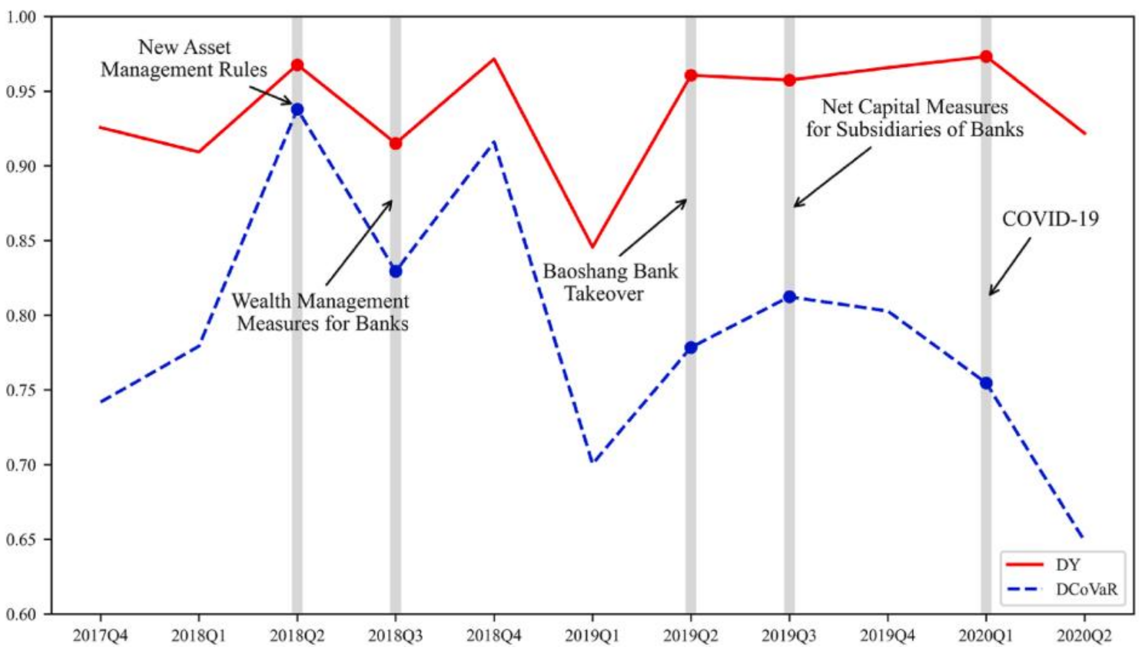


Fig. 15. Adjusted R-squared of regressions in each period.

[23] Diebold-Yilmaz spillover index quantifies the interconnectedness among financial institutions. It is derived from a forecast error variance decomposition of a vector autoregressive (VAR) model: $X_t = v + \sum_{i=1}^{P} \Phi_i X_{t-i} + \varepsilon_t$. Using H-step-ahead forecasts, the i-th variable's contribution to the j-th variable's forecast error variance is $\theta_{ij}^{g}(H) = \frac{\sigma_{ii}^{-1} \sum_{h=0}^{H-1} (e_i A_h \sum e_j)^2}{\sum_{h=0}^{H-1} (e_j A_h \sum e_i)}$. For more details about these notations, please see Diebold and Yilmaz (2009). The pairwise directional connectedness is defined as $S_{ij}^{g}(H) = \theta_{ji}^{\sim g}(H) = \frac{\theta_{ij}^{g}(H)}{\sum_{j=1}^{N} \theta_{ij}^{g}(H)}$, which measures bank i's spillover effect to bank j.

Both $S^{g}_{ij,t}(H)$ and $\Delta CoVaR_t(i, j)$ are directional, i.e., the spillover from bank i to bank$j$ is does not equal the one from bank j to bank$i$, and both matrixes are asymmetric. The risk spillover based on the co-occurrence in negative news, $CM^{neg}_{t}(i, j)$, is directionless, i.e., cooccurrence matrix CM^{neg}_{t} is a symmetric matrix. Therefore, we should transform DY matrix and $\Delta CoVaR$ matrix into symmetric ones. The elements of DY matrix and $\Delta CoVaR$ matrix at (i, j) are defined as:

$$DY_t(i,j) = 0.5 \times \left(S^{g}_{ij,t}(H) + S^{g}_{ji,t}(H)\right) \tag{6}$$

$$DCoVaR_t(i,j) = 0.5 \times (\Delta CoVaR_t(i,j) + \Delta CoVaR_t(j,i)) \tag{7}$$

where $S^{g}_{ij,t}(H)$ and $\Delta CoVaR_t(i, j)$ are the risk spillover of bank i to bank$j$ at time t based on Diebold-Yilmaz variance decomposition, and $\Delta CoVaR$ method.

Thus, we have the following panel regression equations:

$$DY_t(i,j) = \alpha_{1,q} + \beta_{dy,q} \times CM^{neg}_{t}(i,j) + \beta_{con1,q} \times DY_{t-1}(i,j) + \varepsilon_1 \tag{8}$$

$$DCoVaR_t(i,j) = \alpha_{2,q} + \beta_{dcovar,q} \times CM^{neg}_{t}(i,j) + \beta_{con2,q} \times DCoVaR_{t-1}(i,j) + \varepsilon_2 \tag{9}$$

where $CM^{neg}_{t}(i, j)$ is the co-occurrence index of bank i and bank j in negative news; t is the observation month; q is the quarter that month t belongs to, and $\beta_{con,q}$ controls the historical trend of traditional indexes. We estimate the above equations for the three months of each quarter, which yields time-varying parameters in terms of quarter. We choose 19 listed banks in the sample due to data availability[24] over Oct 2017–Jun 2020.[25] For the risk spillover matrices CM^{neg}_{t}, DY_t and $DCoVaR_t$ of monthly frequency in each quarter, we estimate the regression parameters one by one in Eq. (8) and Eq. (9).

[24] The sample banks include B0, B1, B2, B3, B4, B6, B7, B8, B9, B10, B11, B12, B13, B18, B53, B54, B118, B137 and B259.

[25] After 2020, the impact of the pandemic in China has been greatly alleviated without other severe shocks, so June 2020 is set as the end of our time frame here.

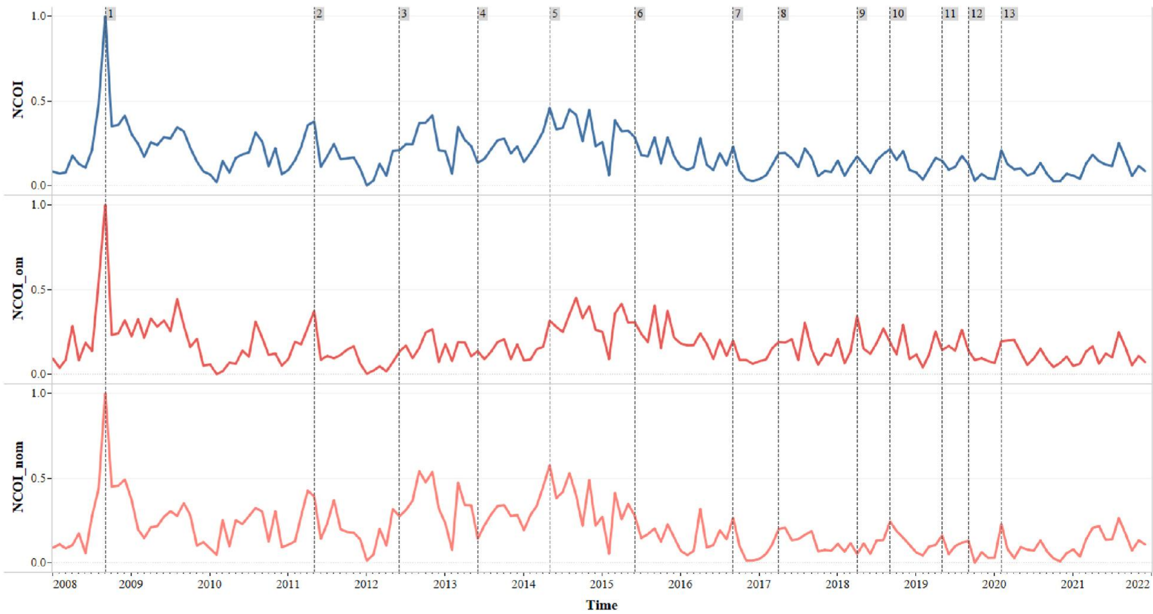


Fig. 16. *$NCOI_t$* based on news reports from various types of media.

Table 5 shows regression results when we employ a 12-week-ahead forecast error variance decomposition for computing the Diebold-Yilmaz spillover index.[26] The structural parameters (β) are predominantly positive, with half of them significant at the 1 % level. The adjusted R-squared ($\tilde{R}^2$) for each period exceeds 0.65, indicating that there is no discernible deviation in trends between traditional systemic risk indicators and the proposed indicator, *NCOI*. Notably, $\tilde{R}^2$ reaches local peaks during five key events high lighted in Fig. 15, which further substantiates the consistency between *NCOI* and traditional indicators. Interestingly, $\tilde{R}^2$ in the regression of *DY* spillover matrix is consistently higher than that of $\Delta CoVaR$. It can be attributed to a distinct conceptual focus of two measures: *DY* spillover index evaluates the overall degree of network interdependence among financial institutions, while $\Delta CoVaR$ captures pairwise spillovers under tail-dependent stress scenarios relying on an exogenous quantile threshold. Unlike $\Delta CoVaR$, *DY* spillover index does not depend on predefined thresholds and is more suitable to capture broader and systemic interdependencies during both normal and stressed periods. This characteristic aligns closely with *NCOI*'s design, which accesses systemic risk by analyzing co-occurrence patterns in negative news, encompassing both direct and indirect contagion pathways. *NCOI* is relatively effective in detecting indirect interconnectedness that arises from external common factors, such as policy changes and operational linkages that are not restricted to extreme tail events. Fig. 15 shows that *NCOI* aligns more closely with *DY* spillover index during the stress periods of 2017–2020 with heightened regulatory interventions and broader financial and operational disruptions, which similarly captures system-wide interdependence in contrast to $\Delta CoVaR$. This distinction underscores the broader applicability of *NCOI* in assessing systemic risk under various market conditions.

---

[26] The result is robust the selection of the forecast horizon. In Appendix 4, we repeat the spillover index computation and the corresponding regression analysis for 2-week and 4-week ahead forecasts and find similar results.

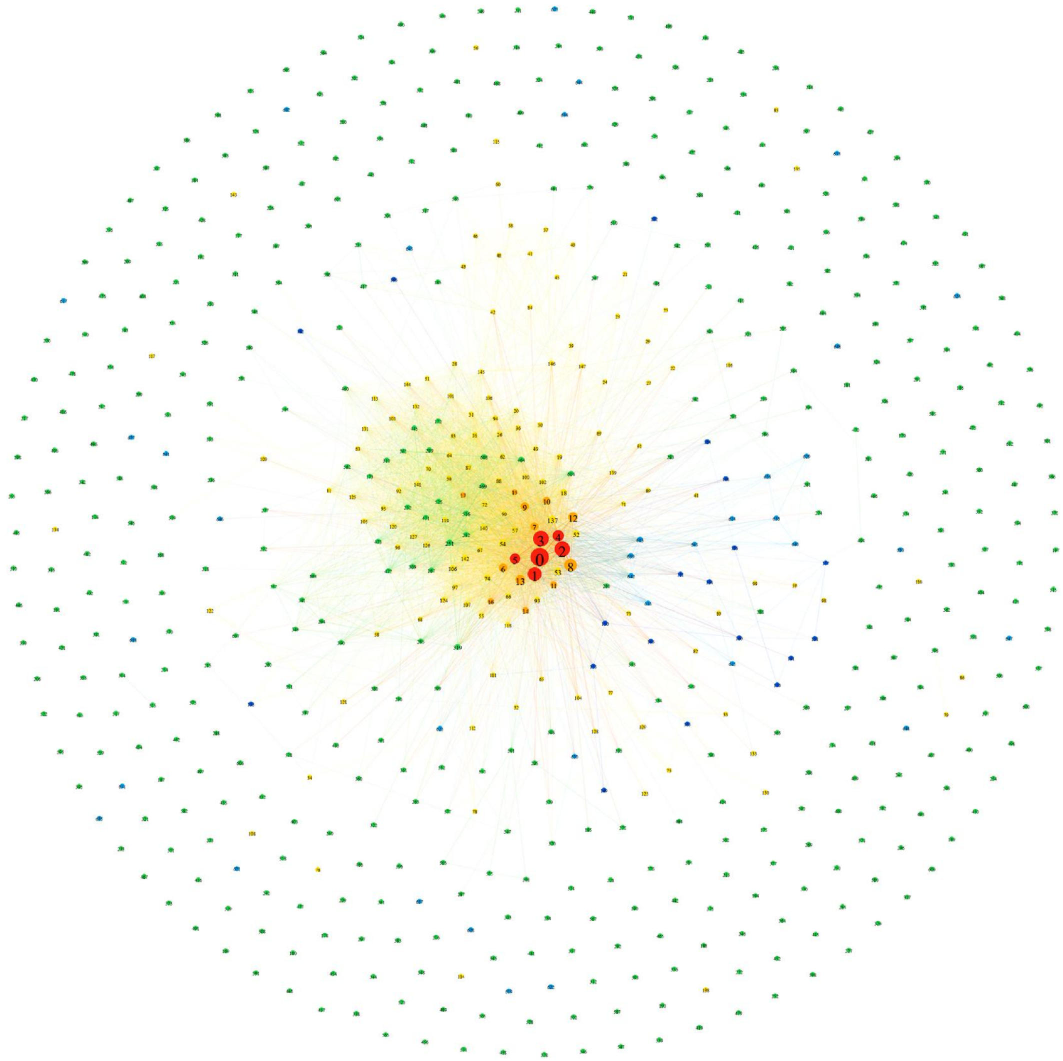


Fig. 17. Bank network in China in 2021 (constructed using official media reports).

6. Discussions on media bias

Text analysis of media reports may suffer from the media bias, especially the independence of official media reporting. Most of these criticisms focus on the lag of information release and the authenticity of the content. You et al. (2018) show that the government may deliberately delay the release of information on negative events for political purposes. Piotroski et al. (2017) and Borochin & Cu (2018) show that the official media disclose less information about firms' operation and management, and report more positive news than negative news. For example, the *Yinguangxia* accounting fraud event was first exposed by market-oriented media rather than the official media.

As nearly 40 percent of the news data comes from the official media, it is necessary to examine whether the media bias affect the risk identification of $NCOI_t$. Fig. 16 shows that $NCOI_t$ is based on news reports from all media, official media and non-official media. Although the peak of official media is not as significant as that of non-official media during the two risk events of credit crunch in the inter-bank market and the stock market crash, the fluctuation of official media is similar with non-official media during the global financial crisis. It indicates that the reports of official media also have certain information value. In addition, $NCOI_t$ of official media, i. e., $NCOI_{om,t}$ in Fig. 16, shows higher volatility than non-official media, i.e., $NCOI_{nom,t}$, when the "New Asset Management Regulations" was introduced in April 2018, and around the takeover of *Baoshang Bank* by the *People's Bank of China* in May 2019. It suggests that official media may have advantage in the collection and dissemination of government related information, which coincides with Piotroski et al. (2020).

It is not easy to judge the merits of official media versus non-official media. Even though the official media may report negative events with bias (such as events 4 and 5), $NCOI_t$ based on the reports of multi-source news media can still provide a decent measurement of the systemic risk.

To further substantiate the robustness of our results on the media category, we extract news data exclusively from official media sources and construct the bank co-occurrence network with the methodology in Section 4.1. Comparing Fig. 17 with Fig. 6, we find that the basic structure of bank co-occurrence network does not exhibit significant differences due to different categories of news media, which still presents a "core-periphery" topology. Specifically, the center of the network is predominantly composed of large state-owned banks such as Industrial and Commercial Bank of China (B0), Agricultural Bank of China (B1), and Bank of China (B2). Banks surrounding the center are joint-stock banks like China Merchants Bank (B8), Ping An Bank (B12), and Industrial Bank (B13), while the outer layer consists of city commercial banks, rural small and medium-sized banks, and private banks. In terms of node centrality, there is a gradual decline from large state-owned banks to joint-stock banks, and then to small and medium-sized banks, with many rural small and medium-sized banks even remaining isolated within the network. The co-occurrence network serves as a visual representation of co-occurrence matrix, which forms a foundational basis for the risk indicator proposed in this paper. Thus, the high similarity between Figs. 17 and 6 further illustrates the robustness of our findings.

Fig. 6 displays the results by applying equal weighting to news information from all media sources. In contrast, Fig. 17 reflects the results by assigning a weight of 1 to information from official media and a weight of 0 to information from non-official media. It suggests that even if we assume the existence of media bias and apply specific weighting schemes to the information from official and non-official media, the outcomes can still be derived from linear operations on co-occurrence matrices associated with Figs. 17 and 6. Given the structural consistency between these figures, the application of any relative weighting scheme across different media types is unlikely to impact our main findings.

7. Conclusion

The regulation of systemic risk is a priority for the financial stability in China. We show that a text analysis of news can serve as a tool to monitor the systemic risk of banks in the country. In contrast to traditional risk indicators based on market and transaction data, we establish a more comprehensive bank network to capture the systemic risk. We use co-occurrence analysis to capture inter-bank relation from news texts, and apply the sentiment analysis to identify negative shocks, while we combine both to construct a cooccurrence index based on negative news. We substantiate the effectiveness of the risk indicator from the time and spatial dimensions.

The co-occurrence index based on negative news is a real-time risk indicator, which is built on the sentiment index and cooccurrence index. News text sentiment can capture the occurrence of negative shocks, which can reveal the multi-level connection between banks. Therefore, co-occurrence index based on negative news can capture the two driving factors of systemic risk, i.e., negative shock and amplification. News co-occurrence can capture the inter-bank connection better than the traditional risk measures based on the market data and business transactions.

In the static connection network, state-owned banks are located at the center of the network and have the most tight connectedness with other banks in the network. Joint stock banks are tightly connected with state-owned banks and some city commercial banks. A large number of small and medium-sized banks are distributed in the periphery of the network. We find that before the occurrence of risk events, interconnections among small and medium-sized banks are mainly concentrated among a few banks, and most banks have few connections. However, most banks will have certain connections with some other banks after the occurrence of risk events.

The risk spillover of small and medium-sized banks is rising recently in China, which necessitates a timely risk monitoring system for these banks that are not covered by traditional systemic risk indicators. Banks exhibit large fluctuations in the ranking of systemic importance around risk events of *Baoshang Bank*, *Bank of Jinzhou* and *Bohai Bank*. The systemic risk of the Chinese banking sector is mainly dominated by state-owned banks and joint stock banks before 2013, while small and medium-sized banks have become main drivers afterwards. The systemic risk indicator based on news sentiment and co-occurrence analysis provides a simple and feasible tool for regulatory authorities, especially for the small and medium-sized banks that receive less attention from the regulatory authorities.

CRediT authorship contribution statement

Yi Fang: Conceptualization, Methodology, Funding acquisition. **Hao Lin:** Data, Software, Writing. **Liping Lu:** Re-writing, Reviewing, Supervision.

Declarations of competing interest

None.

Acknowledgments

This work was supported by Major Project of National Social Science Fund of China in 2023 under Grant No.23&ZD058 (named "Research on Statistical Monitoring, Early Warning and policies of China's Financial Security"), the National Natural Science Foun dation of China under Grant No [72173144&71973162] (named "Financial Textual Big Data and Banking's Systemic Risk: Indicators' Construction, and Their Application, Evaluation, and Integration" and "A Study on the Prevention and Resolution of Systemic Risk in China's Banking from the Perspective of Financial Cycle"). Lu acknowledges the funding by the MOE (Ministry of Education in China) Liberal arts and Social Sciences Foundation (No. 24YJA790039) and Asian Research Center at Renmin University of China (23YYA005). The order of authors is alphabetic. All authors contribute equally.

Appendix 1. Lists of sample banks and media outlets

The list of sample banks in Section 3.3 and the list of banks in regression in Section 5.2 is as follows: https://lin1103.github.io/ Measure-systemic-risk-using-text-analysis/.

Appendix 2. Detailed visualizations for city commercial banks

We add more detailed visualizations on CCBs (city commercial banks). Specifically, Figures A1 and A2 present annual heat maps to illustrate the interconnections between CCBs and LBs (large banks) and within CCBs, respectively. Figure A3 displays the monthly dynamics of the co-occurrence frequency between CCBs and LBs, as well as the co-occurrence frequency between CCBs and SMBs (small and medium-sized banks). Additionally, it shows the percentage of co-occurrence between CCBs and LBs relative to the total cooccurrence frequency between CCBs and all bank types.

Figures A1 and A2 show that the interconnectedness of CCBs with both LBs and other CCBs follows a similar trend over time. Before 2009, only a small number of banks are interconnected. However, when regulatory authorities relaxed restrictions on the branch footprint of city commercial banks after 2009, we find a significant increase in interconnections. This trend accelerates further after 2013, which coincides with the rapid growth of inter-bank business. The growth in interconnections slows temporarily between 2016 and 2019, which is likely due to strict regulations targeting inter-bank and off-balance sheet business introduced since 2016. After 2020, however, a large proportion of CCBs establish interconnections with either LBs or other CCBs. Comparing Figure A1 with Figure A2, we can observe that CCBs tend to connect more frequently with LBs than with other CCBs, i.e., the larger share of bright-colored areas in Figure A1 than Figure A2.

Figure A3 supports these observations in monthly frequency. The co-occurrence frequencies between CCBs and LBs, as well as between CCBs and SMBs, exhibit a similar trend: a rapid increase from 2009 to 2016, some fluctuations between 2016 and 2020, and another increase after 2020. Additionally, the percentage of co-occurrence frequency between CCBs and LBs relative to the cooccurrence frequency between CCBs and all bank types is initially very high, reaching 90 % in 2009, and remains above 50 % for an extended period before 2012. However, it gradually declines below 50 % after 2012 and falls to less than 40 % by the end of 2021. It reflects a rapid growth of CCBs in recent decades and highlights their increasing influence in the banking sector.

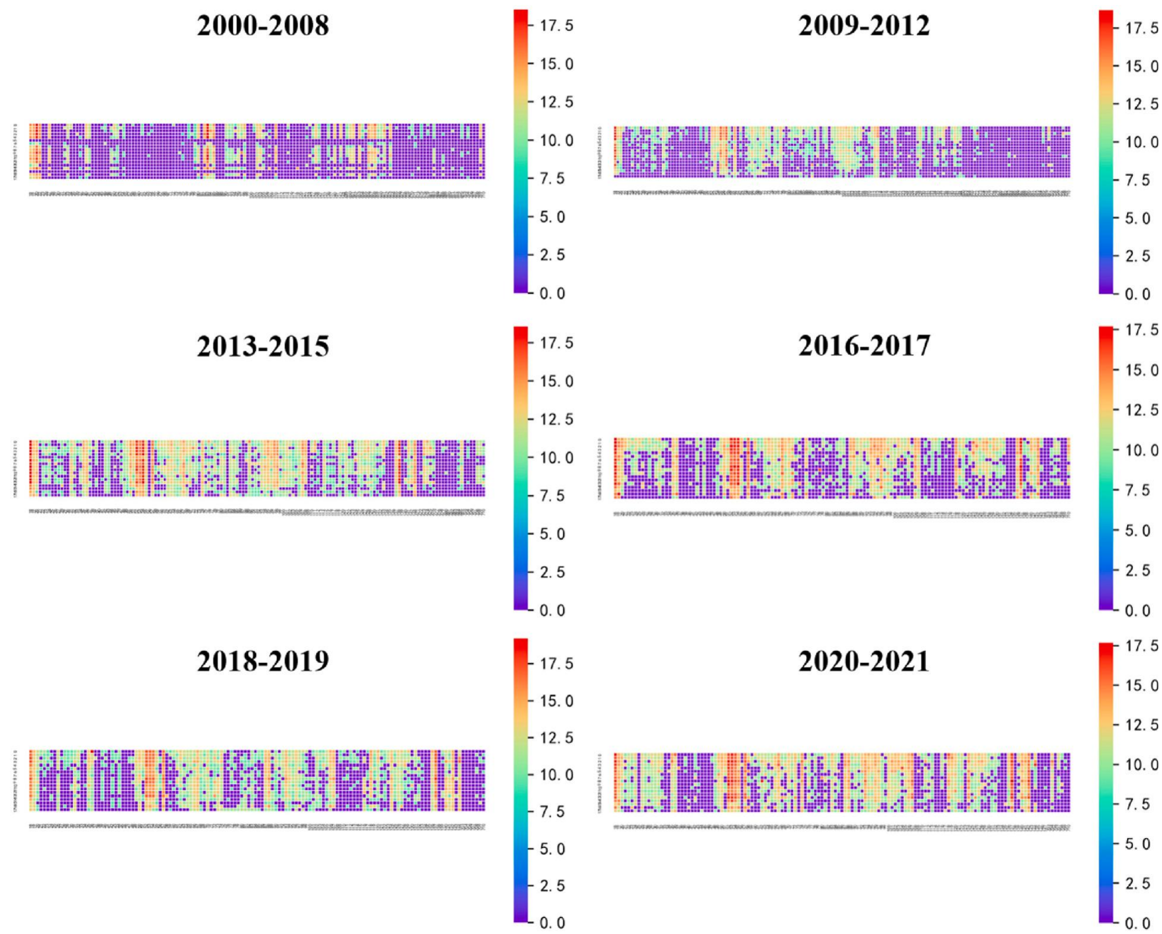


Fig. A1. Inter-bank connection between CCBs and LBs.

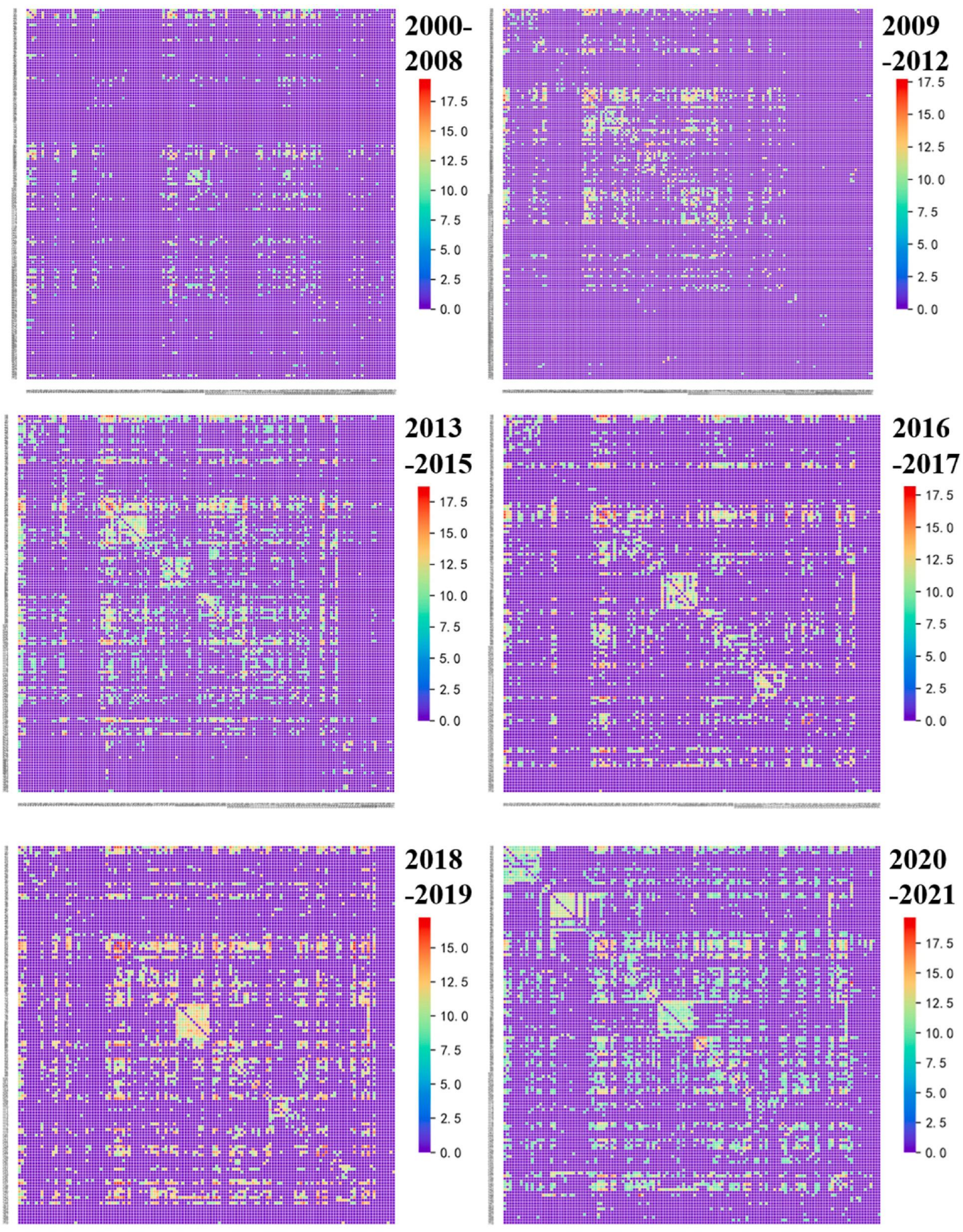


Fig. A2. Inter-bank connection within CCBs.

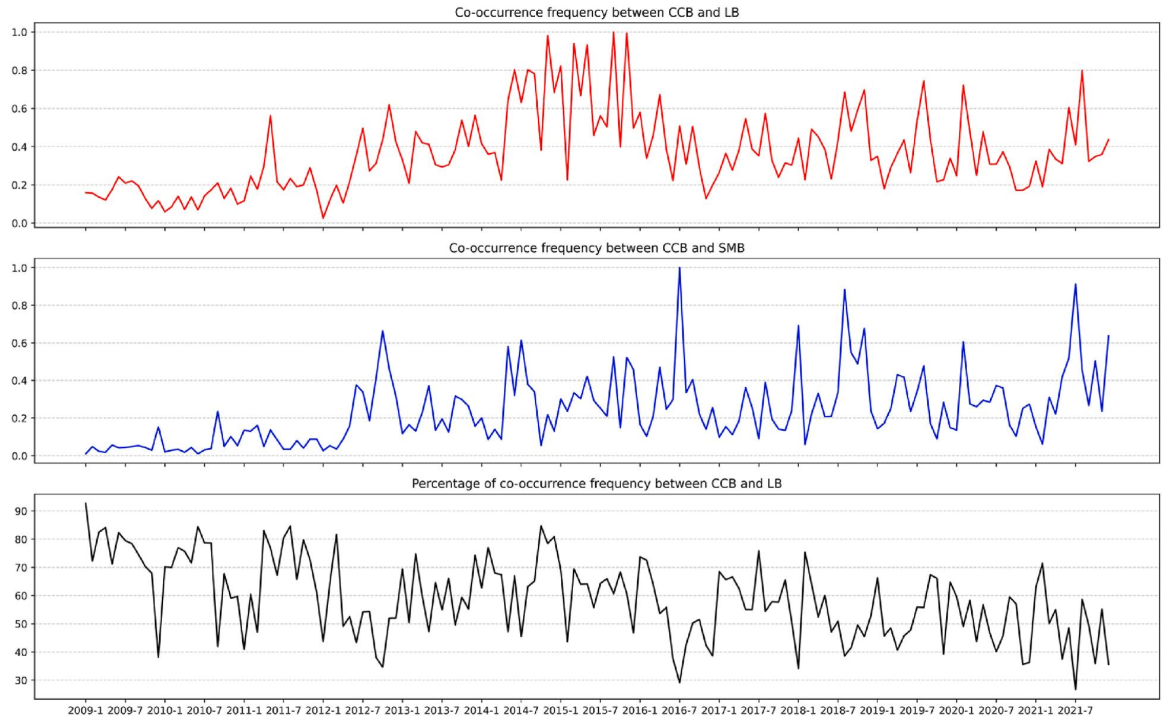


Fig. A3. Monthly co-occurrence frequency between CCBs and LBs, between CCBs and SMBs, and the relative percentage of co-occurrence frequency between CCBs and LBs to all bank types. Note that co-occurrence frequencies have been normalized for a comparison of dynamics.

Appendix 3. Analysis of the driving factors of systemic risk in different periods

To understand how the driving factors of systemic risk evolve over time, we repeat our analysis on the determinants of $NCOI_t$, our systemic risk indicator in Section 5.2. We examine two distinct sub-periods following the global financial crisis, i.e., 2009–2017 and 2018–2021, to investigate the changes of risk drivers after 2018 after introduction of the "New Asset Management Regulations". These regulations represent a landmark reform aimed at addressing key risks in shadow banking, which reduces regulatory arbitrage, and enhances overall financial stability.

Table A1 presents the results for 2009–2017, which suggests that three factors, i.e., bank size, capital adequacy ratio, and ROAA, contribute significantly to the systemic risk. First, the positive coefficients for bank size in all columns suggest that larger banks contribute more to systemic risk during this period. This indicates that the size of banks is a critical factor in driving systemic risk, as larger institutions have greater exposure to interconnected financial activities, which aligns with Anginer et al. (2018) and Laeven et al. (2016). Second, although the coefficient for capital adequacy ratio is positive and significant in column (4), it becomes negative in column (8) after controlling for other variables. A potential explanation is that banks with higher capital buffers are more confident in engaging in riskier activities, while capital adequacy ratio is negatively associated with systemic risk after including additional controls to alleviate omitted variable bias. Third, ROAA exhibits a negative coefficient in columns (6) and (7), which suggests that more profitable banks are associated with lower systemic risk during this period. This result aligns with the notion that profitability reflects better financial health and risk management, which is consistent with the findings by Zhang et al. (2021).

However, Table A2 presents the results for the period 2018–2021, which suggests a shift in the driving factors of systemic risk. Unlike the earlier period, the coefficients for bank size become mostly insignificant across all columns, which suggests that the size effect diminish as a systemic risk driver during this period. This change likely reflects the impact of the "New Asset Management Regulations," which impose stricter regulatory oversight on larger banks, reducing their contribution to systemic risk. Similarly, neither the capital adequacy ratio nor ROAA remains significant in this period. This decline in significance can be attributed to several factors. First, the "New Asset Management Regulations" aims to curb regulatory arbitrage and enhance transparency in shadow banking activities, which likely diminish the direct impact of capital adequacy on systemic risk. Second, the post-2018 period is characterized by intensified competition and narrowing profit margins for banks, partly due to regulatory reforms and economic uncertainties. These factors weaken the importance of profitability as an indicator of financial health and risk management. A notable difference in this period is the emergence of the NPL ratio as a driver of systemic risk. The NPL ratio exhibits a positive and significant coefficient in column (8), indicating that credit quality issues contribute significantly to systemic risk after 2018. This result underscores the rising importance of asset quality in the regulatory environment after 2018.

The results from the two sub-periods highlight a significant evolution in the key drivers of systemic risk over time. Before 2018, systemic risk is predominantly driven by bank size, capital buffer, and profitability, with larger banks playing a more crucial role in banking stability. However, after 2018, marked by the introduction of the “New Asset Management Regulations”, asset quality emerges as a more critical factor driving systemic risk. The diminished significance of bank size, capital buffer, and profitability reveals a structural change in the banking sector following the introduction of stricter regulatory measures, which suggests an increasingly importance of small and medium-sized banks in the overall stability of the banking system.

Table A1 and Table A2 provide a deeper understanding of the findings in Table 3 in Section 5.2. Specifically, Table 3, which covers the entire period from 2005 to 2021, identifies bank size, capital adequacy ratio, profitability, and credit quality as critical drivers of systemic risk. The additional analysis of the two sub-periods in Tables A1 and A2 show that the relative importance of these factors changes over time. These changes are closely linked to regulations such as the “New Asset Management Regulations” introduced in 2018. It highlights the connection between regulatory interventions and the evolution of systemic risk drivers, which underscores the dynamics of systemic risk.

Table A1

Two-way fixed-effect estimates on the determinants of $NCOI_t$ during 2009 and 2017. $NCOI_t$ is multiplied by 100. Log assets is the natural logarithm of total assets. Growth in assets is annual growth in total assets. Capital adequacy ratio is total capital adequacy ratio. NPL ratio is non-performance loan ratio. Liquidity ratio is the ratio of liquid assets over total assets. ROAA is return on average assets. Non-interest income share is the ratio of noninterest income over total assets. The inflation rate is calculated based on CPI. Significance ***, 0.01; **, 0.05; *, 0.10.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log assets$_{t-1}$	47.2**	62.7**	59.1**	177.4**	47.2**	50.4**	46.9**	192.6***
	(2.06)	(2.24)	(2.13)	(2.57)	(2.12)	(2.16)	(2.04)	(3.26)
Growth in assets$_{t-1}$		−0.030						0.34
		(-0.58)						(1.23)
Capital adequacy ratio$_{t-1}$			0.22**					−8.16**
			(2.27)					(-2.29)
NPL ratio$_{t-1}$				−3.10				8.93
				(-0.38)				(0.35)
Liquidity ratio$_{t-1}$					0.017			−2.33
					(0.04)			(-1.62)
ROAA$_{t-1}$						−9.76*		−149.9***
						(-1.91)		(-2.82)
Non − interest income share$_{t-1}$							6.61	58.8*
							(0.73)	(1.88)
Constant	−1141.9**	−1535.1**	−1442.4**	−4364.7**	−1144.4**	−1215.4**	−1139.1*	−4450.8***
	(-1.99)	(-2.18)	(-2.07)	(-2.51)	(-2.05)	(-2.09)	(-1.96)	(-3.12)
Bank fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
N	921	798	848	418	921	920	921	358
R-sq	0.13	0.16	0.14	0.24	0.13	0.13	0.13	0.36
adj. R-sq	0.12	0.15	0.13	0.22	0.12	0.12	0.12	0.33

Table A2

Two-way fixed-effect estimates on the determinants of $NCOI_t$ during 2018 and 2021. $NCOI_t$ is multiplied by 100. Log assets is the natural logarithm of total assets. Growth in assets is annual growth in total assets. Capital adequacy ratio is total capital adequacy ratio. NPL ratio is non-performance loan ratio. Liquidity ratio is the ratio of liquid assets over total assets. ROAA is return on average assets. Non-interest income share is the ratio of noninterest income over total assets. The inflation rate is calculated based on CPI. Significance ***, 0.01; **, 0.05; *, 0.10.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log assets$_{t-1}$	−4.58	−3.27	−4.59	−3.76	−7.58	−4.38	−4.9	−4.58
	(-0.47)	(-0.35)	(-0.44)	(-0.35)	(-0.73)	(-0.47)	(-0.48)	(-0.41)
Growth in assets$_{t-1}$		−0.021						0.013
		(-0.39)						−0.21
Capital adequacy ratio$_{t-1}$			1.11					2.26
			−0.95					−1.44
NPL ratio$_{t-1}$				1.22				2.45**
				−1.55				−2.03
Liquidity ratio$_{t-1}$					0.21			0.19
					−1.42			−1.06
ROAA$_{t-1}$						1.95		−2.28
						−0.23		(-0.23)
Non − interest income share$_{t-1}$							−0.64	1.22
							(-0.18)	−0.26
Constant	171.5	136.8	157.6	146.6	245.7	164.8	180.4	131.8
	−0.66	−0.55	−0.58	−0.52	−0.9	−0.67	−0.67	−0.45
Bank fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
N	437	435	429	386	437	437	437	381
R-sq	0.07	0.07	0.07	0.06	0.07	0.07	0.07	0.07
adj. R-sq	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.04

Appendix 4. Robustness to the selection of forecast horizons

To compare with the results in Section 5.4, which utilizes a 12-week-ahead forecast error variance decomposition for the Diebold-Yilmaz spillover index, we repeat the spillover index computation and the regression analysis for 2-week and 4-week ahead forecasts.

First, Table A3 reports the estimates of structural parameters across different forecast horizons. The results indicate that both the values and significance levels of the structural parameters remain quite stable, which suggests the consistency across various time frames in the variance decomposition.

Second, Figure A4 illustrates the adjusted R-squared of the regressions for each forecast horizon. Our analysis reveals that the dynamics of adjusted R-squared follow similar trends across three forecast horizons, with only negligible differences in magnitude during each period. Notably, we observe that adjusted R-squared reaches local peaks during five risk events in Figure A4, which confirms the robustness of our regression results regardless of the forecast horizon selected.

Table A3

Structural parameters of regressions with different forecast horizons. This table compares the estimates of structural parameters in cases of 2, 4, 12-step ahead forecast error variance decomposition, denoted by $\beta_{dy}(2)$, $\beta_{dy}(4)$, and $\beta_{dy}(12)$, respectively. Significance: ***, 0.01; **, 0.05; *, 0.10.

Time	$\beta_{dy}(2)$	$\beta_{dy}(4)$	$\beta_{dy}(12)$	β_{dcovar}
2017 Q4	0.23***	0.23***	0.24***	−0.08**
2018 Q1	0.01	−0.04	−0.08	0.53***
2018 Q2	−0.02	0.02	0.01	0.13***
2018 Q3	0.24***	0.15**	0.15**	1.05***
2018 Q4	0.06	0.01	−0.01	0.67***
2019 Q1	−0.02	0.08	0.12	1.01***
2019 Q2	0.09*	−0.01	0.03	0.10
2019 Q3	0.18***	0.09*	0.04	0.71***
2019 Q4	0.06	0.1**	0.11**	−0.09
2020 Q1	0.06	0.17***	0.18***	0.58***
2020 Q2	0.54***	0.57***	0.59***	0.67***

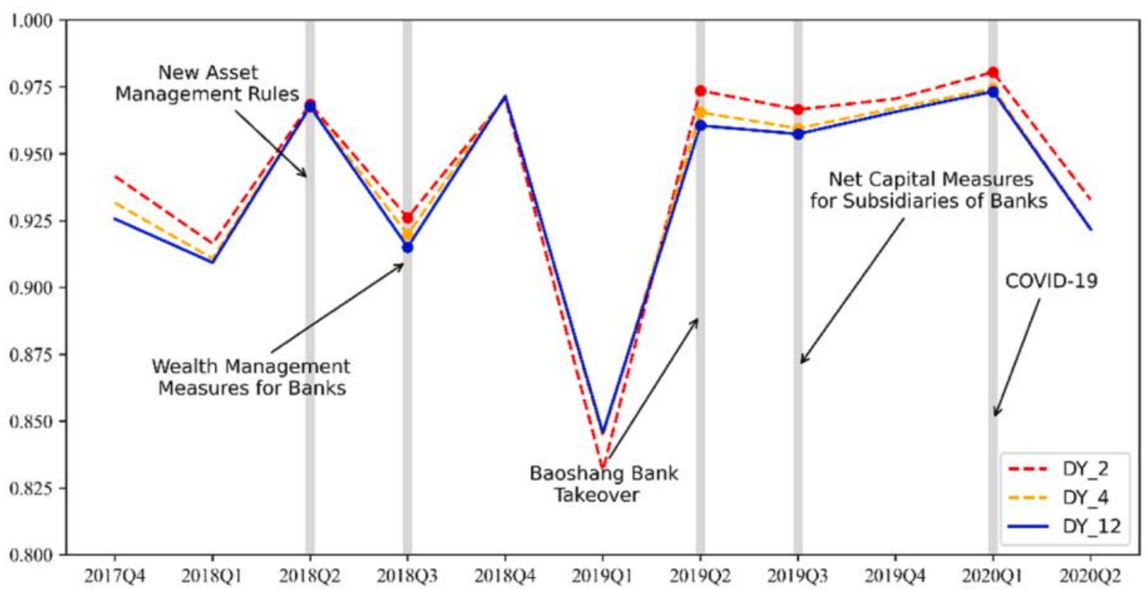


Fig. A4. Adjusted R-squared of regressions with different forecast horizons. This figure shows the values of adjusted R-squared of regressions with 2- , 4-, 12-step-ahead forecast error variance decomposition in each quarter over Oct 2017–Jun 2020.

Data availability

Data will be made available on request.

References

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *The American Economic Review, 105*(2), 564–608.

Acharya, V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies, 30*(1), 2–47.

Acharya, V., Qian, J., Su, Y., & Yang, Z. (2019). In the shadow of banks: Wealth management products and issuing banks' risks in China. *NYU stern working paper*. New York: New York University.

Acharya, V., & Thakor, A. V. (2016). The dark side of liquidity creation: Leverage and systemic risk. *Journal of Financial Intermediation, 28*, 4–21.

Acharya, V., & Yorulmazer, T. (2007). Too many to fail: An analysis of time-inconsistency in bank closure policies. *Journal of Financial Intermediation, 16*(1), 1–31.

Acharya, V., & Yorulmazer, T. (2008). Information contagion and bank herding. *Journal of Money, Credit, and Banking, 40*(1), 215–231.

Aharony, J., & Swary, I. (1983). Contagion effects of bank failures: Evidence from capital markets. *Journal of Business*, 305–322.

Allen, F., & Babus, A. (2009). *Networks in finance* (Vol. 367). The Network Challenge: Strategy, Profit, and Risk in An Interlinked World.

Allen, F., Babus, A., & Carletti, E. (2012). Asset commonality, debt maturity and systemic risk. *Journal of Financial Economics, 104*(3), 519–534.

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of Political Economy, 108*(1), 1–33.

Altinoglu, L., & Stiglitz, J. E. (2023). Collective moral hazard and the interbank market. *American Economic Journal: Macroeconomics, 15*(2), 35–64.

Anand, K., Craig, B., & Von Peter, G. (2015). Filling in the blanks: Network structure and interbank contagion. *Quantitative Finance, 15*(4), 625–636.

Anand, K., Gai, P., Kapadia, S., Brennan, S., & Willison, M. (2013). A network model of financial system resilience. *Journal of Economic Behavior & Organization, 85*, 219–235.

Ando, T., Greenwood-Nimmo, M., & Shin, Y. (2022). Quantile connectedness: Modeling tail behavior in the topology of financial networks. *Management Science, 68*(4), 2401–2431.

Andries¸, A. M., Ongena, S., Sprincean, N., & Tunaru, R. (2022). Risk spillovers and interconnectedness between systemically important institutions. *Journal of Financial Stability, 58*, Article 100963.

Anginer, D., Demirguc-Kunt, A., Huizinga, H., & Ma, K. (2018). Corporate governance of banks and financial stability. *Journal of Financial Economics, 130*(2), 327–346.

Azizpour, S., Giesecke, K., & Schwenkler, G. (2018). Exploring the sources of default clustering. *Journal of Financial Economics, 129*(1), 154–183.

Bargigli, L., Di Iasio, G., Infante, L., Lillo, F., & Pierobon, F. (2015). The multiplex structure of interbank networks. *Quantitative Finance, 15*(4), 673–691.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control, 36*(8), 1121–1141.

Benoit, S., Colliard, J. E., Hurlin, C., & Perignon, C. (2017). Where the risks lie: A survey on systemic risk. *Review of Finance, 21*(1), 109–152.
Bernanke, B., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics, 1*, 1341–1393.
Bernstein, A., Clearwater, S., Hill, S., Perlich, C., & Provost, F. (2002). *Discovering knowledge from relational data extracted from business news*. *Working Paper*.
Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics, 104*(3), 535–559.
Bisias, D., Flood, M., Lo, A. W., & Valavanis, S. (2012). A survey of systemic risk analytics. *Annual Review Financial Economics, 4*(1), 255–296.
Boot, A., & Ratnovski, L. (2016). Banking and trading. *Review of Finance, 20*(6), 2219–2246.
Brown, C., & Dinç, I. S. (2011). Too many to fail? Evidence of regulatory forbearance when the banking sector is weak. *Review of Financial Studies, 24*(4), 1378–1405.
Brownlees, C., & Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies, 30*(1), 48–79.
Brunnermeier, M., Dong, G., & Palia, D. (2020). Banks' noninterest income and systemic risk. *Review of Corporate Finance Studies, 9*(2), 229–255.
Bushman, R., Chen, J., & Williams, C. D. (2016). *Bank connectedness: Qualitative and quantitative disclosure similarity and future tail comovement*. *Unpublished Working Paper*.
Cao, Q., Yang, F., & Liu, M. (2022). Impact of managerial power on regulatory inquiries from stock exchanges: Evidence from the text tone of Chinese listed companies' annual reports. *Pacific-Basin Finance Journal, 71*, Article 101646.
Capponi, A., Chen, P., & Yao, D. (2016). Liability concentration and systemic losses in financial networks. *Operations Research, 64*(5), 1121–1134.
Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion. *Neurocomputing, 264*, 50–56.
Chen, C. D., Cheng, C. M., & Demirer, R. (2017). Oil and stock market momentum. *Energy Economics, 68*, 151–159.
Craig, B., & Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation, 23*(3), 322–347.
De Jonghe, O., Diepstraten, M., & Schepens, G. (2015). Banks' size, scope and systemic risk: What role for conflicts of interest? *Journal of Banking & Finance, 61*, S3–S13.
Di Gangi, D., Lillo, F., & Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic Dynamics and Control, 94*, 117–141.
Dicks, D. L., & Fulghieri, P. (2019). Uncertainty aversion and systemic risk. *Journal of Political Economy, 127*(3), 1118–1155.
Diebold, F. X., & Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal, 119*(534), 158–171.
Duan, Y., El Ghoul, S., Guedhami, O., Li, H., & Li, X. (2021). Bank systemic risk around COVID-19: A cross-country analysis. *Journal of Banking & Finance, 133*, Article

106299.
Duarte, F., & Eisenbach, T. M. (2021). Fire-sale spillovers and systemic risk. *The Journal of Finance, 76*(3), 1251–1294.
Eisenberg, L., & Noe, T. (2001). Systemic risk in financial systems. *Management Science, 47*(2), 236–249.
Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial networks and contagion. *The American Economic Review, 104*(10), 3115–3153.
Ellul, A., & Yerramilli, V. (2013). Stronger risk controls, lower risk: Evidence from US bank holding companies. *The Journal of Finance, 68*(5), 1757–1803.
Engelberg, J. E., Reed, A. V., & Ringgenberg, M. C. (2012). How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics, 105*(2), 260–278.
Fang, L., Xiao, B., Yu, H., & You, Q. (2018). A stable systemic risk ranking in China's banking sector: Based on principal component analysis. *Physica A: Statistical Mechanics and Its Applications, 492*, 1997–2009.
Farhi, E., & Tirole, J. (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *The American Economic Review, 102*(1), 60–93.
Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164.
Glasserman, P., & Young, H. P. (2015). How likely is contagion in financial networks? *Journal of Banking & Finance, 50*, 383–399.
Gong, X. L., Liu, X. H., Xiong, X., & Zhang, W. (2019). Financial systemic risk measurement based on causal network connectedness analysis. *International Review of Economics & Finance, 64*, 290–307.
Greenwood, R., Landier, A., & Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics, 115*(3), 471–485.
Hachem, K. (2018). Shadow banking in China. *Annual Review of Financial Economics, 10*(1), 287–308.
Hachem, K., & Song, Z. (2021). Liquidity rules and credit booms. *Journal of Political Economy, 129*(10), 2721–2765.
He, Z., & Wei, W. (2023). China's financial system and economy: A review. *Annual Review of Economics, 15*(1), 451–483.
Herskovic, B., Kelly, B., Lustig, H., & Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. *Journal of Political Economy, 128*(11), 4097–4162.
Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy, 124*(5), 1423–1465.
Huang, T. L. (2018). The puzzling media effect in the Chinese stock market. *Pacific-Basin Finance Journal, 49*, 129–146.
Huang, Q., De Haan, J., & Scholtens, B. (2019). Analysing systemic risk in the Chinese banking system. *Pacific Economic Review, 24*(2), 348–372.
Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics, 132*(1), 126–149.
Jiang, F., Meng, L., & Tang, G. (2021). Media text sentiment and stock return prediction. *China Economic Quarterly, 21*(4), 1323–1344 (in Chinese).
Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance, 66*, Article 101694.

Kazinnik, S., Scida, D., Wu, J., & Killen, C. (2024). *News and networks: Using financial news coverage to measure bank interconnectedness*.
Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis, 33*, 171–185.
Kristoufek, L., & Vosvrda, M. (2013). Measuring capital market efficiency: Global and local correlations structure. *Physica A: Statistical Mechanics and Its Applications, 392*(1), 184–193.
Laeven, L., & Levine, R. (2007). Is there a diversification discount in financial conglomerates? *Journal of Financial Economics, 85*(2), 331–367.
Laeven, L., Ratnovski, L., & Tong, H. (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking & Finance, 69*, S25–S34.
Lee, C. C., Wang, Y., & Zhang, X. (2023). Corporate governance and systemic risk: Evidence from Chinese-listed banks. *International Review of Economics & Finance, 87*, 180–202.
Li, R. (2013). Media corruption: A Chinese characteristic. *Journal of Business Ethics, 116*(2), 297–310.
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65.
Lux, T. (2015). Emergence of a core-periphery structure in a simple dynamic model of the interbank market. *Journal of Economic Dynamics and Control, 52*, A11–A23.
Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives, 17*(1), 59–82.
Markose, S., Giansante, S., & Shaghaghi, A. R. (2012). 'Too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk. *Journal of Economic Behavior & Organization, 83*(3), 627–646.
Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking & Finance, 35*(5), 1114–1127.
Morrison, A. D. (2011). Systemic risks and the 'too-big-to-fail'problem'. *Oxford Review of Economic Policy, 27*(3), 498–516.
Nyman, R., Kapadia, S., & Tuckett, D. (2021). News and narratives in financial systems: Exploiting big data for systemic risk assessment. *Journal of Economic Dynamics and Control, 127*, Article 104119.
Oh, D., & Patton, A. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics, 36*(2), 181–195.
Patro, D. K., Qi, M., & Sun, X. (2013). A simple indicator of systemic risk. *Journal of Financial Stability, 9*(1), 105–116.
Piotroski, J. D., Wong, T. J., & Zhang, T. (2017). Political bias in corporate news: The role of conglomeration reform in China. *The Journal of Law and Economics, 60*(1), 173–207.
Piotroski, J. D., Wong, T. J., Zhang, S., & Zhang, T. (2020). *Firm news and market views: The informational role of official newspapers in China*. *Working Paper*.
Račickas, E., & Vasiliauskaitė, A. (2012). Model of financial risk contagion in the global financial markets. *Economics and Management, 17*(1), 91–102.

Rönnqvist, S., & Sarlin, P. (2015). Bank networks from text: Interrelations, centrality and determinants. *Quantitative Finance, 15*(10), 1619–1635.
Sarlin, P. (2016). Macroprudential oversight, risk communication and visualization. *Journal of Financial Stability, 27*, 160–179.
Shi, Q., Sun, X., & Jiang, Y. (2022). Concentrated commonalities and systemic risk in China's banking system: A contagion network approach. *International Review of Financial Analysis, 83*, Article 102253.
Shleifer, A., & Vishny, R. W. (2010). Unstable banking. *Journal of Financial Economics, 97*(3), 306–318.
Silva-Buston, C. (2019). Systemic risk and competition revisited. *Journal of Banking & Finance, 101*, 188–205.
Song, Z., & Xiong, W. (2018). Risks in China's financial system. *Annual Review of Financial Economics, 10*(1), 261–286.
Thompson, R. B., Olsen, C., & Dietrich, J. R. (1987). Attributes of news about firms: An analysis of firm-specific news reported in the wall street journal index. *Journal of Accounting Research*, 245–274.
Tian, M., Alshater, M., & Yoon, S. (2022). Dynamic risk spillovers from oil to stock markets: Fresh evidence from GARCH copula quantile regression-based CoVaR model. *Energy Economics, 115*, Article 106341.
Tobias, A., & Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review, 106*(7), 1705.
Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability, 7*(3), 111–125.
van Lelyveld, I. (2014). Finding the core: Network structure in interbank markets. *Journal of Banking & Finance, 49*, 27–40.
van Oordt, M., & Zhou, C. (2019). Systemic risk and bank business models. *Journal of Applied Econometrics, 34*(3), 365–384.
Varotto, S., & Zhao, L. (2018). Systemic risk and bank size. *Journal of International Money and Finance, 82*, 45–70.
Wan, X., Yang, J., Marinov, S., Calliess, J. P., Zohren, S., & Dong, X. (2021). Sentiment correlation in financial news networks and associated market movements. *Scientific Reports, 11*(1), 1–12.
Wang, Y., Chen, S., & Zhang, X. (2014). Measuring systemic financial risk and analyzing influential factors: An extreme value approach. *China Finance Review International, 4*(4), 385–398.
Wang, G. J., Jiang, Z. Q., Lin, M., Xie, C., & Stanley, H. E. (2018). Interconnectedness and systemic risk of China's financial institutions. *Emerging Markets Review, 35*, 1–18.
Wang, J., & Ye, K. (2015). Media coverage and firm valuation: Evidence from China. *Journal of Business Ethics, 127*(3), 501–511.
Yan, Y., Jeon, B. N., & Wu, J. (2023). The impact of the COVID-19 pandemic on bank systemic risk: Some cross-country evidence. *China Finance Review International, 13*(3), 388–409.
Yang, X., Liu, Z., Li, J., & Xie, Q. (2023). Communities of co-occurrence network of financial firms in news. *Procedia Computer Science, 221*, 821–825.

Yaros, J. R., & Imieliński, T. (2014). Diversification improvements through news article co-occurrences. In *2014 IEEE conference on computational intelligence for financial engineering & economics (CIFEr)* (pp. 130–137). IEEE.
You, J., Zhang, B., & Zhang, L. (2018). Who captures the power of the pen? *Review of Financial Studies, 31*(1), 43–96.
Zhang, X., Fu, Q., Lu, L., Wang, Q., & Zhang, S. (2021). Bank liquidity creation, network contagion and systemic risk: Evidence from Chinese listed banks. *Journal of Financial Stability, 53*, Article 100844.
Zheng, H., & Schwenkler, G. (2020). *The network of firms implied by the news (no. 108)*. European Systemic Risk Board.
Zou, L., Cao, K. D., & Wang, Y. (2019). Media coverage and the cross-section of stock returns: The Chinese evidence. *International Review of Finance, 19*(4), 707–729.