

Al-Haschimi, Alexander; Apostolou, Apostolos; Azqueta-Gavaldon, Andres;  
Ricci, Martino

**Working Paper**

## Using machine learning to measure financial risk in China

ECB Working Paper, No. 2767

**Provided in Cooperation with:**

European Central Bank (ECB)

*Suggested Citation:* Al-Haschimi, Alexander; Apostolou, Apostolos; Azqueta-Gavaldon, Andres; Ricci, Martino (2023) : Using machine learning to measure financial risk in China, ECB Working Paper, No. 2767, ISBN 978-92-899-5509-6, European Central Bank (ECB), Frankfurt a. M., <https://doi.org/10.2866/565243>

This Version is available at:

<https://hdl.handle.net/10419/278306>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



EUROPEAN CENTRAL BANK

EUROSYSTEM

## Working Paper Series

Alexander Al-Haschimi, Apostolos Apostolou,  
Andres Azqueta-Gavaldon, Martino Ricci

### Using machine learning to measure financial risk in China

No 2767 / January 2023

**Disclaimer:** This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

### **Abstract**

We develop a measure of overall financial risk in China by applying machine learning techniques to textual data. A pre-defined set of relevant newspaper articles is first selected using a specific constellation of risk-related keywords. Then, we employ topical modelling based on an unsupervised machine learning algorithm to decompose financial risk into its thematic drivers. The resulting aggregated indicator can identify major episodes of overall heightened financial risks in China, which cannot be consistently captured using financial data. Finally, a structural VAR framework is employed to show that shocks to the financial risk measure have a significant impact on macroeconomic and financial variables in China and abroad.

*Keywords:* China; financial risk; textual analysis; machine learning; topic modelling; LDA.

*JEL-Classification:* C32, C65, E32, F44, G15.

## **Non-technical summary**

Since the Global Financial Crisis (GFC), financial risks in China have been accumulating. Credit to the nonfinancial sector, as a percentage of GDP, has increased at a more rapid pace than in any other major emerging market, and it soared above that of major advanced economies. Rising debt levels have been observed both in the household and in the corporate sector with leverage being particularly high in the real estate sector. Risks also arise from over-reliance on short-term funding in some sectors, the opacity of the shadow banking sector, and hidden risks in traditional banking operations. Moreover, additional headwinds such as international trade disputes, the COVID-19 pandemic, geopolitical tensions, policy uncertainty regarding the future growth model for China, and regulatory changes in key sectors further complicate the landscape for financial market participants.

In this environment, materialisation of risks in China's financial system recurs periodically. Recent examples are the equity price collapse experienced in 2015, the insolvency problems and subsequent default of Baoshang Bank in 2019-2020, the defaults of several major real estate developers in 2021, and a stock market correction coinciding with unusually high capital outflows from China in early 2022. This last episode followed an increase in regulation of the fast-growing technology sector and rising geopolitical risks associated with the war in Ukraine. These episodes exemplify that financial markets in China remain subject to volatility, while the potential triggers of risk are diverse.

However, monitoring financial risk in the world's second-largest economy remains a challenging task. The rise of the shadow banking sector and the entry of technology firms into financial intermediation involving payments and in credit services implies that the source of risk is rapidly changing, as are financial regulations and financial reporting requirements. The availability of consistent financial data that has sufficient history to allow for standard time series analysis is often lacking, thus posing challenges to quantifying the various aspects of financial risks in the system using traditional measures.

In this paper, we therefore augment the set of available risk indicators by developing a measure of financial risk by applying machine learning techniques to a large number of newspaper articles. Specifically, we rely on text-based analysis to identify major episodes of financial risks in China, quantify them and disentangle the different sources of financial risks. We do this using the Latent Dirichlet Allocation (LDA) algorithm, a machine learning technique which allows for topic modelling. Subsequently, we use a structural vector autoregressive (SVAR) model, as common in the literature on risk and uncertainty, to quantify the impact of rising financial risk on the Chinese and the global economy. We find that an increase in the financial risk index has a statistically significant negative impact on both Chinese and global macro

and financial variables.

## 1 Introduction

Financial risks in China are increasing. The real estate sector is under severe stress following a string of defaults by major developers, including Evergrande. The tensions in the real estate sector had been rising over time as leverage reached levels which pressured regulators to significantly tighten access to credit, thereby prompting a liquidity squeeze among developers. More broadly, financial risks in China have been accumulating since the Global Financial Crisis (GFC). Credit to the nonfinancial sector, as a percentage of GDP, has increased at a more rapid pace than in any other major emerging market, and it soared above that of major advanced economies (Figure 1a). Rising debt levels have been observed both in the household and in the corporate sector with leverage being particularly high in the real estate sector (Figure 1b). Risks also arise from over-reliance on short-term funding in some sectors, the opacity of the shadow banking sector, and hidden risks in traditional banking operations.<sup>1</sup> Moreover, additional headwinds such as international trade disputes, the COVID-19 pandemic, geopolitical tensions, policy uncertainty regarding the future growth model for China, and regulatory changes in key sectors further complicate the landscape for financial market participants.

In this environment, materialisation of risks in China's financial system recurs periodically. In 2015, equity prices collapsed and erased USD 3 trillion in shareholder value (Song and Xiong, 2018). In 2019, the insolvency problems and subsequent default of Baoshang Bank triggered liquidity disruptions in the interbank market. In 2021, excessive leverage in the real estate sector, in combination with significantly tightened regulation by the government, led to defaults of several major real estate developers, causing significant volatility and uncertainty in Chinese financial markets. In 2022, a Chinese stock market correction coincided with unusually high capital outflows from China, following increased regulation of the fast-growing technology sector and rising geopolitical risks associated with the war in Ukraine. These episodes exemplify that financial markets in China remain subject to volatility, while the potential triggers of risk are diverse.

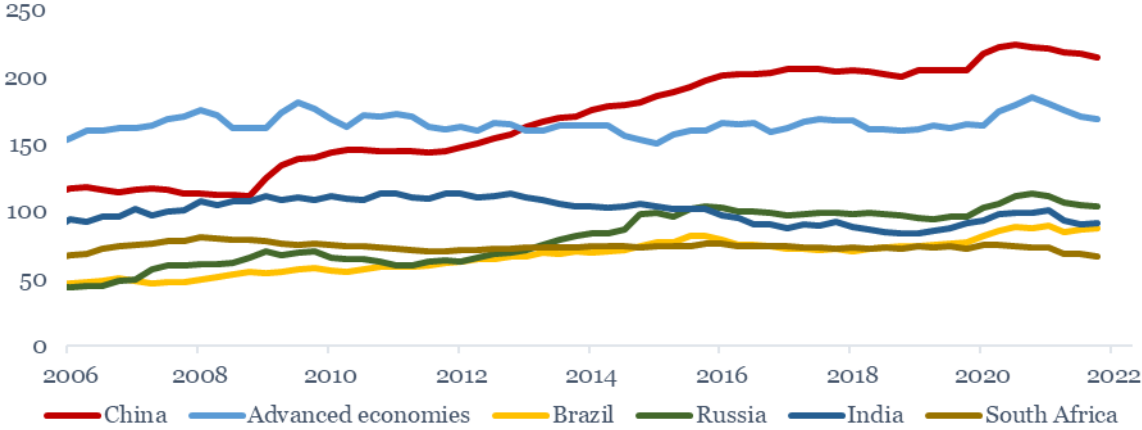
At the same time, monitoring financial risk in the world's second-largest economy remains a challenging task. China's financial sector is changing rapidly: While it is largely under the purview of the public sector and relatively closed to international private capital, it is undergoing many structural changes, from the partial liberalisation of the capital account to evolving monetary and regulatory regimes,

<sup>1</sup> Shadow banking refers to non-bank financial institutions providing bank-like financial services and in particular credit instruments.

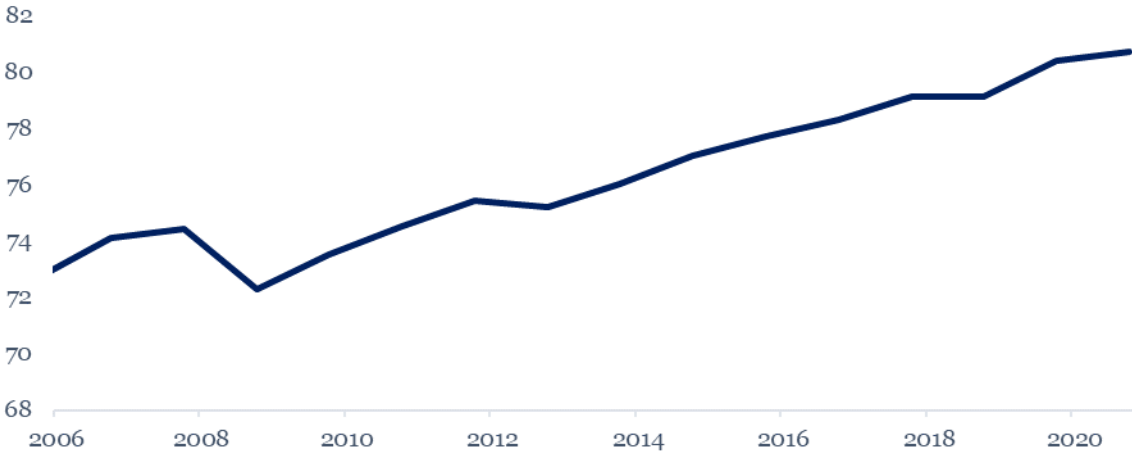
and the rise of new financial market participants (Song and Xiong, 2018). Technology companies, in particular, have assumed a significantly greater role in financial intermediation and credit provision, with regulatory oversight often playing catch up to mitigate new risks created by new market participants.

**Figure 1**

**Figure 1a - Credit to the private nonfinancial sector**  
(% of GDP)



**Figure 1b - China: Real Estate Investment: Ratio of Liabilities to Assets**  
(percent)



**Source:** BIS and China Statistics Yearbook, National Bureau of Statistics (NBS) of China via Haver Analytics.  
**Notes:** Credit to nonfinancial corporation is from all sectors and at market value, namely the price at which the asset would change hands if sold on the open market. It includes both credit to households and to nonfinancial corporations.

In this paper, we augment the often sparse available conventional data on financial risks in China with a measure of financial risk obtained by applying machine learning

techniques to a large number of newspaper articles. Specifically, we rely on text-based analysis to identify major episodes of financial risks in China, quantify them and disentangle the different sources of financial risks. We do this using the Latent Dirichlet Allocation (LDA) algorithm, a machine learning technique which allows for topic modelling. Subsequently, we use a structural vector autoregressive (SVAR) model, as common in the literature on risk and uncertainty, to quantify the impact of rising financial risk on the Chinese and the global economy. We find that an increase in the financial risk index has a statistically significant negative impact on both Chinese and global macro and financial variables.

The rest of the paper is structured as follows. Section 2 provides a literature review and discusses the identification of financial risks using text-based analysis; Section 3 describes the data and methodology used; Section 4 presents the empirical findings and validation of the financial risk indices; Section 5 discusses the SVAR analyses of the link between financial risk and economic activity; and Section 6 concludes.

## **2 Literature review**

### **2.1 Financial risks in China**

China's growth slowdown after the GFC and the subsequent surge in credit levels have induced conditions under which the financial system is showing increasing strains. To revitalise the economy, the government implemented a major stimulus plan focusing on large investment projects financed in significant part by credit. To support this effort, the large Chinese state-owned banks increased substantially their provision of loans. By 2010, a cap on the ceiling of the loan-to-deposit ratio forced these banks to increase deposits. Facing increased competition for deposits, smaller and medium-size banks, started to issue so-called Wealth Management Products (WMP), which are off-balance sheet items with short maturities that are substitutes for deposits (Acharya et al., 2020).

Acharya et al. (2020) find that WMPs increased risk in the Chinese banking system, and present empirical evidence that stock market investors price in rollover risks from WMPs during times of stressed funding conditions. Sun (2019) finds that shadow banking activities impair the effectiveness of banking regulation and contribute to the accumulation of systemic risk in China's financial system. Non-bank financial actors increased in prominence over time, in part by managing to circumvent regulations and specifically limits on leverage. For instance, the Fintech sector created shadow-financed accounts enabling retail investors to trade in the Chinese stock market with higher leverage than allowed using regulated brokerage accounts. In 2015, regulations began to tighten on shadow-financed equity trading, and excess leverage induced fire-sales that contributed to the major stock market correction in mid-2015 (Bian et al., 2018).

More recently, China experienced a string of bank failures, with the People's Bank of China (PBC) and the banking regulator taking control of Baoshang Bank in May 2019, citing severe credit risks. Baoshang Bank represented the most high-profile bank failure in 20 years. In July and August of that year, two more banks failed, prompting the PBC to conduct liquidity injections to avoid an escalation of systemic risk (Lo, 2019). The incident highlighted a source of risk in the form of implicit guarantees. Many investors do not adequately account for risks in their investment decisions, believing that the state will bail out banks, even in the absence of formal guarantee agreements. A similar concern regarding implicit guarantees affects WMPs used to invest in the real estate sector. Over 40% of outstanding WMPs have maturities of three months or less, while being used as a source for longer-term lending. This type of short-term funding for longer-term investments exposes these products to liquidity and rollover risks (Apostolou et al., 2021). However, investors are not adequately internalising these risks as investors perceive WMPs as being implicitly covered by guarantees from the issuing bank or the government, even though 70% of WMPs issued since 2007 are not covered by explicit guarantees (Dang et al., 2019).

In 2021, excessive leverage led to stress in the real estate sector. Preceding the real estate turmoil, government regulatory efforts had imposed limits on leverage for property developers. In August 2020, new regulations introduced a set of thresholds for financial ratios, the so-called three red lines, which when exceeded, would limit the ability of developers to raise debt. Subsequently, liquidity in the sector dried up, and several developers, and most notably Evergrande, defaulted on their debt in 2021-22. Again, the PBC intervened with liquidity injections to prevent the spillover of risks to other sectors. Nevertheless, the volatility in the real estate sector, and resulting slowdown in the housing sector, weighed on investor sentiment and economic growth. Moreover, the stress in the real estate sector highlighted a number of sources of underlying financial risks in China's economy, relating to excessive leverage, the circumvention of leverage limits via off-balance sheet funding sources, and the mispricing of risk due to the assumption of implicit guarantees, among others. Overall, the nature of a rapidly developing financial sector in China presents challenges for macro- and micro-prudential policies to keep pace with the ongoing innovation in financial markets, most recently by large emerging fintech companies. As such, developing broad measures that aim to quantify a wide variety of risks in the financial system can be a useful additional monitoring tool.

## **2.2 Textual analysis in economics**

One approach that can capture a broad measure of risk is based on textual analysis. This approach has been used in economics in a number of contexts and Gentzkow et al. (2019) provide a rich review of the literature. Textual analysis can be conducted



through a variety of methodologies. The dictionary method, which has garnered wide attention with the Economic Policy Uncertainty index by Baker et al. (2016), entails counting the number of newspaper articles that contain a set of words linked to a specific topic. Results have shown that the counting-method produces a fairly accurate index that can identify episodes of increased uncertainty about economic policies, such as for example the start of severe trade tensions between China and the US.<sup>2</sup> A similar approach has recently been employed to construct a risk indicator relating to geopolitical developments. The geopolitical risk index by Caldara and Iacoviello (2022) counts newspaper articles citing words in different categories describing specific types of geopolitical risks, including those relating to threats from war and terrorist organisations, among others. The authors show that their index appears to capture well global episodes of elevated geopolitical risks.

Other methods employ statistical analysis from the big data literature. Among them, machine learning tools are used to analyse text with supervised and unsupervised machine learning algorithms and neural networks. In supervised machine learning, a so-called training sample of text data is first classified by researchers who would, for example, note whether the sentiment of a given article is positive or negative. Subsequently, the algorithm learns to distinguish between articles with positive and negative sentiment using this training sample of classified articles as a guide.

Unsupervised learning include algorithms that evaluates text , without prior training samples. One well-known application of unsupervised learning is topic modelling, based for instance on the Latent Dirichlet Allocation (LDA) algorithm as in Blei et al. (2003). The method allows researchers to identify the main topics discussed in a body of text and track how the frequency of these topics changes over time. For example, a study of Norwegian news articles found that the frequency of certain topics discussing financial market developments, and in particular credit and borrowing, can have predictive power for the evolution of key quarterly economic variables and asset prices (Larsen and Thorsrud, 2019).

While applications of textual analysis to measure financial risk are rather limited, there are a few papers that specifically link text-based indicators to financial variables. Piškorec et al. (2014) developed a news cohesiveness index, which reflects the similarity of word frequencies across different articles and find that their measure correlates well with US and EU stock market indices, as well as stock market volatility indices such as the VIX. Ormerod et al. (2015) develops a text-based index of market sentiment by computing the frequency of words indicating anxiety-related sentiment expressed in the Thompson-Reuters news feed. The resulting indicator is shown to have predictive power for one-quarter ahead US GDP growth. The authors also find that this sentiment index Granger causes the financial stress indices of the

<sup>2</sup> The measurement and impact of economic policy uncertainty in China has been studied in some detail; see, e.g., Davis et al. (2019), He et al. (2020), Huang and Luk (2020), Li and Wu (2020), Liu and Zhang (2020), and Sha et al. (2020), among others.

Cleveland and St Louis Federal Reserve Banks.

### **3 Data and methodology**

The following sections describe the data and sources used in the analysis, as well as the topic modelling approach employed to identify the topics that drive the aggregate risk measure in our collection of newspaper articles.

#### **3.1 Textual data**

In order to obtain a data source that is consistent over an extended time period, the text data was sourced from the US print edition of the Wall Street Journal (WSJ) and the South China Morning Post (SCMP), a Hong Kong-based newspaper that covers mainland China extensively, via the Dow Jones Factiva database. The two newspapers provide authoritative coverage of developments in China and have sufficient history in the database used to access the articles. There are two main concerns around the construction of a text-based indicator on the Chinese economy using newspaper articles. The first is related to the potential manipulation of information by authorities at the source. This problem, which abstracts from the degree of independence of a given media outlet, limits the scope of our analysis as important news could be censored or distorted before reaching the public. The second relates to the direct influence Chinese authorities exert on some media. Relying on foreign rather than on domestic, or mainland, newspapers aims to reduce the impact of censorship of mainland Chinese news media. At the same time, in particular the Hong Kong-based SMCP may increasingly face editorial pressures. To the extent that this impacts the news coverage, it may also affect the news-based risk indicator. Upon closer inspection, we believe that our indicator is, by construction, relatively robust to official or self-censorship. First, there are studies which constructed risk or uncertainty measures for China based on textual analysis (e.g Davis et al., 2019; Huang and Luk, 2020) using mainland newspapers. This, presumably, exposes them to issues concerning editorial independence, reliability and censorship given the tight grip of the Chinese Communist Party on media.<sup>3</sup> However, those authors argue that censorship has no qualitative impact on their index (Huang and Luk, 2020). Second, we verified that comparing the frequency of the risk incidence in the news reporting is qualitatively the same whether the SMCP is included in the sample or not. One reason for this is that our analytical approach does not rely on sentiment analysis, whereby the index reacts to the intensity of positive or negative reporting, which may be affected by overly optimistic coverage

<sup>3</sup> According to Reporters Without Borders, an international non-profit organization which seeks to defend and promote freedom of information, China is ranked 175<sup>th</sup> in over 180 countries in the world in 2022 for the freedom of its press.

resulting from censorship. Instead, the risk index reflects the frequency with which certain subjects are reported, which is a metric that is less sensitive to the way specific news are covered. Our approach therefore follows that of related studies, including Chen and Tillmann (2021), who conduct textual analysis on China using foreign media coverage.

The choice of print over online newspaper editions is designed instead to guard against upward trends in the amount of news coverage available through online media, while the editorial process of print newspapers also reduces duplication or updates of existing articles. The use of print media has the cost of potentially reducing the sample size of available news articles given the proliferation of online media.

The text data we analyse cover the period 1 January 2005 to January 2022. From the Dow Jones database, all articles relating to financial risks in China were downloaded. For the regional restrictions, we employed Factiva's geographical filters, selecting mainland China (i.e., excluding Hong Kong). The articles filtered geographically are around 10,000 per year over an average of around 68,000 articles published in both newspapers yearly in the sample period considered. Given its proximity to China, the share of articles discussing China-related issues over total articles published is larger for the SCMP than for the WSJ: of the 10,000 articles on China, around 80% are published in the SCMP.

To identify articles relating to financial risks, we filtered for a set of words contained in the article. To obtain this list of words, we started with the words '*risk*' and '*financial*' as well as permutations of these terms such as risks, riskiness, etc. We then generated a list of words that are semantically similar using the word2vec algorithm created by Mikolov et al. (2013).<sup>4</sup> The word2vec algorithm was applied to a training sample of 1,000 articles for 2015, which was a year in which China experienced elevated volatility in its exchange rate, capital flows as well as in economic growth rates. The algorithm generated 100 words that were most similar to '*risk*' and '*financial*', and judgment was used to reduce this list to those words that were relevant both in meaning and sentiment. For the term finance, the related words identified as relevant comprise banking, lending and commerce. We also included "shadow banking" given its unique importance in China's financial system. For the term '*risk*', 29 words were identified to be relevant. The resulting set of search terms is listed in Table 1.<sup>5</sup> This refined search produces around 2,000 China risk-related articles per year. Interestingly, in this refined sample, the share of articles identified in the WSJ increases to around 35% (from 20% if we consider the geographical filtering only). This is explained by the fact that the WSJ focuses more on financial matters than the more general topics covered by the SCMP.

<sup>4</sup> We use the Python implementation of the word2vec algorithm.

<sup>5</sup> Table A.1. in the Appendix shows the words obtained with the world2vec algorithm.

**Table 1: Final search query**

Text	(china or chinese) and (financ\$ or bank\$ or lend\$ or commerc\$ or shadow bank\$) and (risk\$ or lever\$ or problem\$ or opaqu\$ or pain\$ or unstable or danger\$ or default\$ or unwind\$ or deterioat\$ or vulnerab\$ or volatil\$ or negative\$ or contagio\$ or imbalance\$ or hazard\$ or trigger\$ or delever\$ or bad or uncertain\$ or jeopard\$ or system\$ or serious\$ or consequ\$ or shock\$ or bubbl\$ or stress\$ or threat\$ or burden\$)
Date	01/01/2005 to 31/01/2022
Source	The Wall Street Journal (print) Or South China Morning Post (print)
Author	All Authors
Company	All Companies
Subject	All Subjects
Industry	All Industries
Region	China Not Hong Kong
Language	English

Source: Dow Jones Factiva database.

### 3.2 Topic modelling

In the next step, we apply an unsupervised machine learning algorithm to the selected number of news articles. Specifically, we use the Latent Dirichlet Allocation (LDA), as developed by Blei et al. (2003). Intuitively, the algorithm studies the co-occurrences of words across articles to frame each topic as a distribution of words with a specific probability of belonging to a given topic. Each article, in turn, is represented by a distribution of topics. The classification of topics is conducted using an unsupervised learning approach in that the algorithm computes the two latent distributions without any prior labelling of the articles. The only parameter input to the algorithm is the given number of topics,  $K$ , that exist within a document.

The data generating process is therefore described by two probability distributions: topics as a distribution of words, and articles as a distribution of topics. The model recovers these two distributions by obtaining the parameters that maximise the probability of each word appearing in each article, given the total number of topics  $K$ . In this respect, the probability of a word  $\omega_i$  occurring in an article is:

$$P(\omega_i) = \sum_{j=1}^K P(\omega_i | z_i = j) P(z_i = j) \quad (1)$$

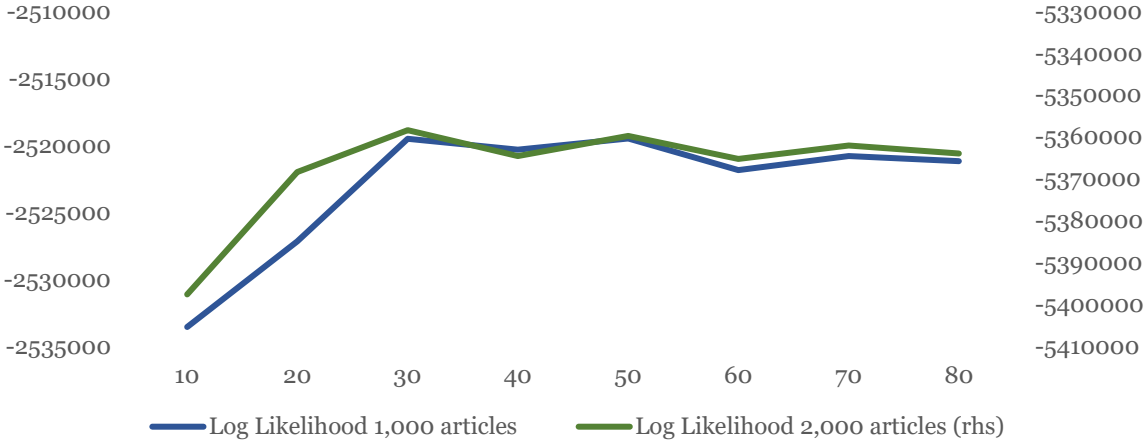
where  $z_i$  is a latent variable indicating the topic from which the  $i$ th word was drawn and  $P(\omega_i | z_i = j)$  is the probability of word  $\omega_i$  being drawn from topic  $j$ .  $P(z_i = j)$  is

the probability of drawing a word from topic  $j$  in the current article, which will vary across different articles. Intuitively,  $P(w|z)$  indicates which words are important to a topic, whereas  $P(z)$  is the prevalence of those topics within an article.

In its complete format, the posterior distribution can be represented by all its components:  $p(z, K, \theta | w, a, \eta)$ , where  $z$  is the topic assignment (the probability of choosing a given topic across the set of articles),  $K$  is the number of topics, and  $\theta$  is the article-topic distribution. Moreover, the list of words is given by  $w$ , and the hyper-parameters of the model are  $a$  and  $\eta$ . High levels of  $\eta$  represent the probability distribution of words to topics being more even, while a low level of  $\eta$  represents fewer words having a much higher probability of defining that topic than the rest. Similarly, high levels of  $a$  indicate articles containing a similar topic distribution per article while low levels of  $a$  indicate a more disperse distribution.

The choice of the optimal number of topics  $K$  is a model selection problem, which we solve relying on a likelihood maximisation method. This method involves estimating empirically the likelihood  $P(w|K)$  which is the likelihood of the data (words) for any given model specified by different values of  $K$ . We then select the model, and thus a value for  $K$ , which best fits the data. The likelihood cannot be directly estimated since it requires summing over all possible assignments of words to topics but can be approximated using the harmonic mean of a set of values of  $P(w|z, K)$ , when  $z$  is sampled from the posterior distribution (Griffiths and Steyvers, 2004). We sampled 1,000 and 2,000 random articles (without replacement) across the whole text corpus and computed the log-likelihood score groups of topics ranging from  $N = 10, 20, 30, \dots, 80$ . Figure 2 shows the resulting scores and that according to this procedure 30 is the optimal number of topics for both sample subsets. Specifically, increasing  $K$  above 30 does not increase the likelihood that a given document contains more than 30 distinct topics.

**Figure 2:** Number of topics and log-likelihood scores



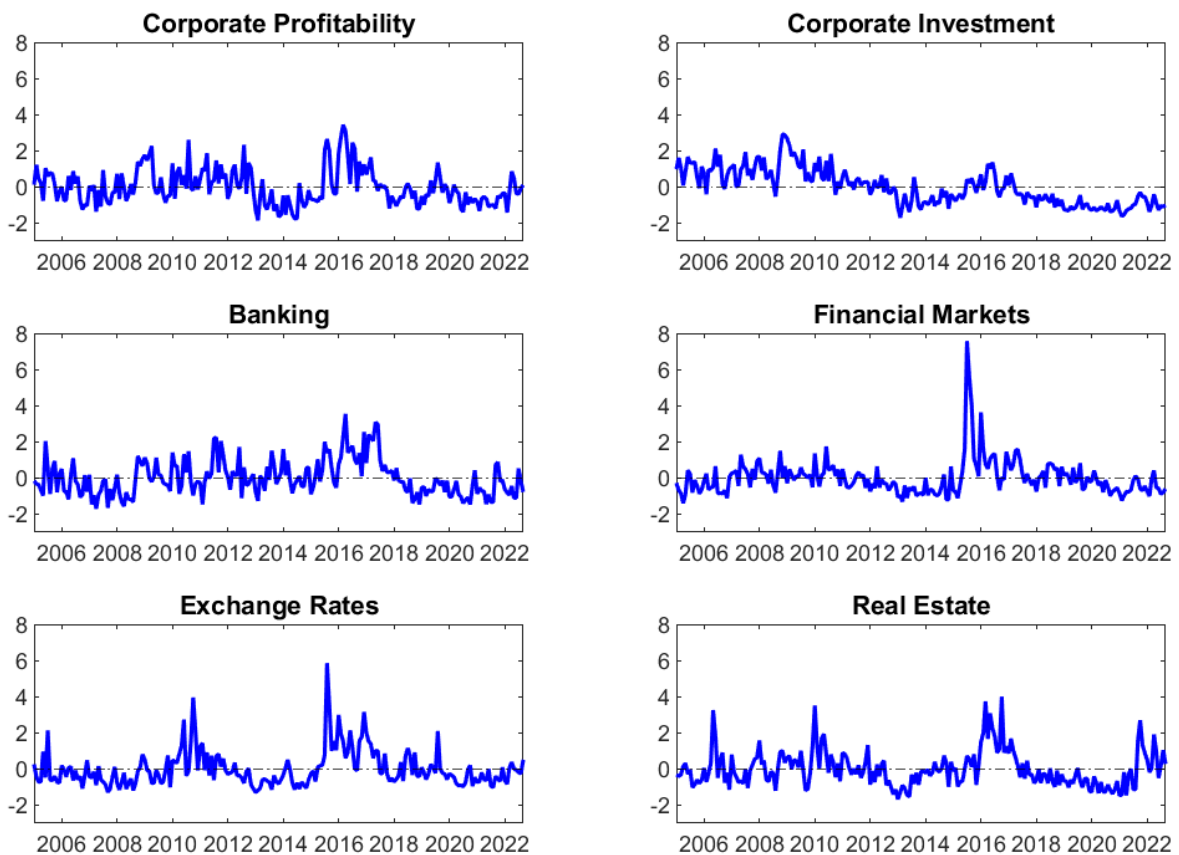
**Source:** Authors' calculations.



financial risk index. Third, we scale the index by the total number of articles about China published by both newspapers over the period considered. To select the final set of topics, we employ judgment by (i) considering the main focus of the topic, (ii) reading the articles that are most representative of the topics – i.e. have the largest probabilities of belonging to a given topic – and by (iii) observing the proximity in the topic mixture which informs our judgment on how well identified the topics are and how they interact with one another. This leads us to identify 6 topics: (1) corporate profitability, (2) corporate investment, (3) banking, (4) financial markets, (5) exchange rates and (6) real estate. Of the topics selected, 4 are very well identified according to the proximity map and consistently represented in the corpus of articles considered (topics iii to vi).<sup>6</sup> The two topics related to the corporate sector appear to be less relevant following those criteria, but we include them in order to capture potential materialisation of risks in the corporate sector, given the risk posed by high corporate leverage and its relevance for financial stability concerns (see, e.g., Apostolou et al. 2022).

Figure 4 shows the time series of those six indicators, scaled by the total number of articles on China.

**Figure 4:** Components of financial risk in China



**Source:** Authors' calculations.

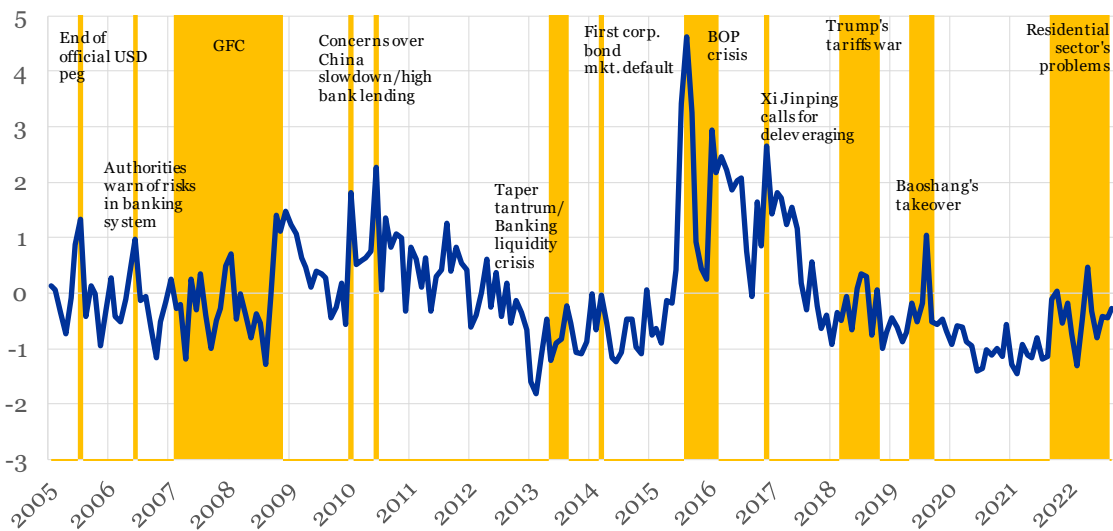
**Notes:** Time coverage 2005M01-2022M09. Series are standardised to have zero mean and standard deviation

<sup>6</sup> See Figure A.4 in Appendix A for the proximity map.

equal to one over the sample considered. The series are scaled monthly by the total number of articles related to China published by the newspapers considered.

Figure 5 shows the resulting aggregation of the indices presented in Figure 4. Plotting the index against some important global and China specific events, we can see that it spikes in correspondence to major episodes in which the Chinese and the global economy have experienced turbulences in the financial sector. Most notably, it picks up the period of a major stock market correction and the RMB sell-off in 2015-2016 when the financial risk indicator reaches its all-time high and more recently during the rising concerns over financial stability and growth caused by the turmoil in the residential sector. The episode of the global financial crisis in 2008-2009 also coincides with a sharp rise in the risk index initially, which however declined shortly thereafter. This corresponds well with the massive fiscal stimulus programme by the Chinese government that was implemented from the fourth quarter in 2008 through 2009 and 2010. The fiscal support was the largest in the world and about three times the size of the U.S. stimulus programme (Wong, 2011). In addition to the government support, credit issued by state-owned banks rose significantly after the start of the crisis and total credit increased by more than one third in 2009, compared to the previous year. Loose financial conditions led to an asset bubble that induced steep rises in land and house prices. As such, the financial risk index corresponds closely with the financial conditions over the period of the Global Financial Crisis.

**Figure 5:** Financial risk index



**Source:** Authors' calculations.

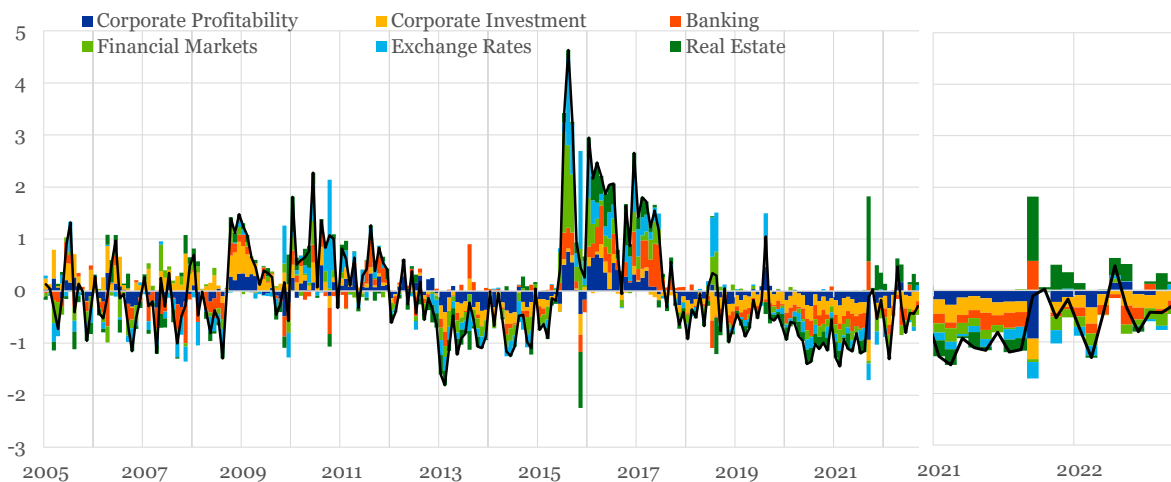
**Notes:** Time coverage 2005M01-2022M09. The series is standardised to have zero mean and standard deviation equal to one over the sample considered. The series is scaled monthly by the total number of articles related to China published by the newspapers considered.

Having constructed the financial risk indicator as the sum of sector specific indices, our approach allows us to decompose the index into the contribution of each single topic.



This alludes to the main reasons contributing to an increase in financial risk in the Chinese economy, enabling the tracking of specific sources of risk over time. Figure 6 shows that given the limited exposures of the Chinese financial sector to the subprime mortgage market and the domestic credit focus of China’s financial system, the stress experienced during the GFC is relatively contained and broadly balanced across sectors. However, during 2015-6, China’s financial sector experienced higher stress with corporate bond defaults, a sharp drop in the stock market, and a RMB sell-off. Consistently, our indicator main drivers in this period are “financial markets” and “exchange rates”. Interestingly, we observe other increases in financial risks picked up by our measure such as the collapse of Baoshang in 2019. Finally, zooming in on the last two years, the increase in the index appears to be substantially due to an increase in risk in the residential sector.

**Figure 6: Financial risk index: Decomposition**



**Source:** Authors’ calculations.

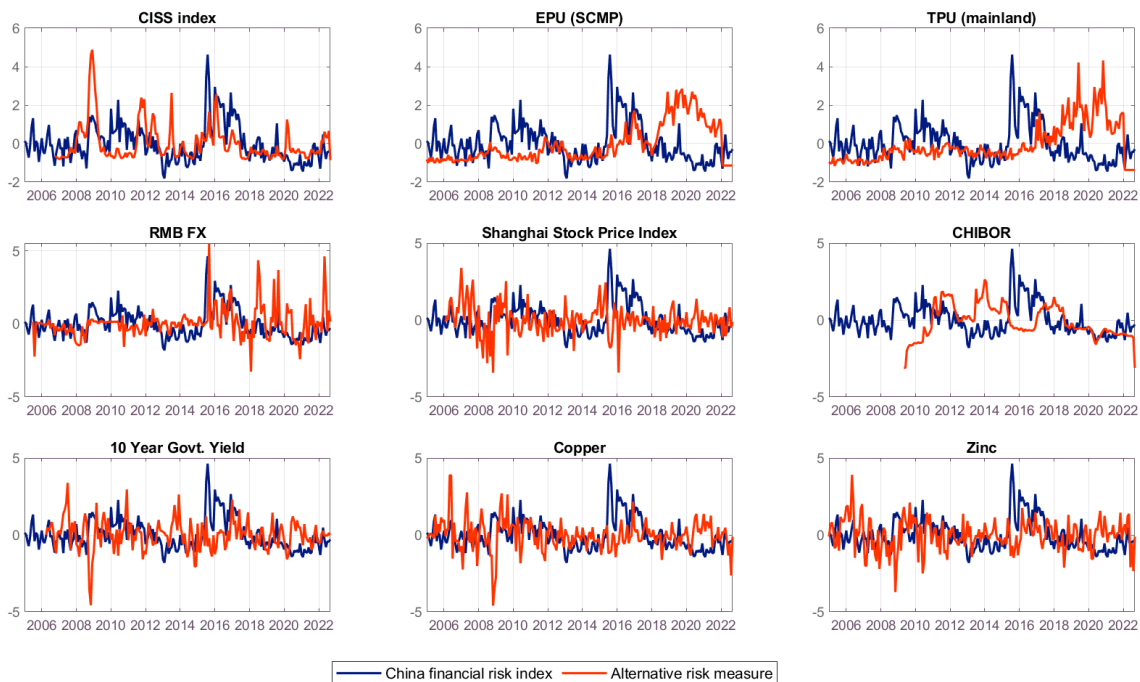
**Notes:** Time coverage 2005M01-2022M09. The series are scaled monthly by the total number of articles related to China published by the newspapers considered.

## 4.2 Validation of the financial risk index

As financial risk is a latent variable, our indicator can only be considered a proxy which aims at capturing the *true* level of financial risk in the Chinese economy. Thus, the issue of validating the indicator is non-trivial. In order to provide some additional evidence of the usefulness of our index, we benchmark it against a range of variables that are broadly used as measures of risk, uncertainty or stress in China’s economy and financial sector. More precisely, we rely on a set of measures that are either developed by the literature or are used by analysts to monitor risk and stress in specific key segments of the Chinese financial sector. Specifically, we explore whether our overall indicator is related to five indicators from the literature, namely the China-CISS index (Ma et al.,

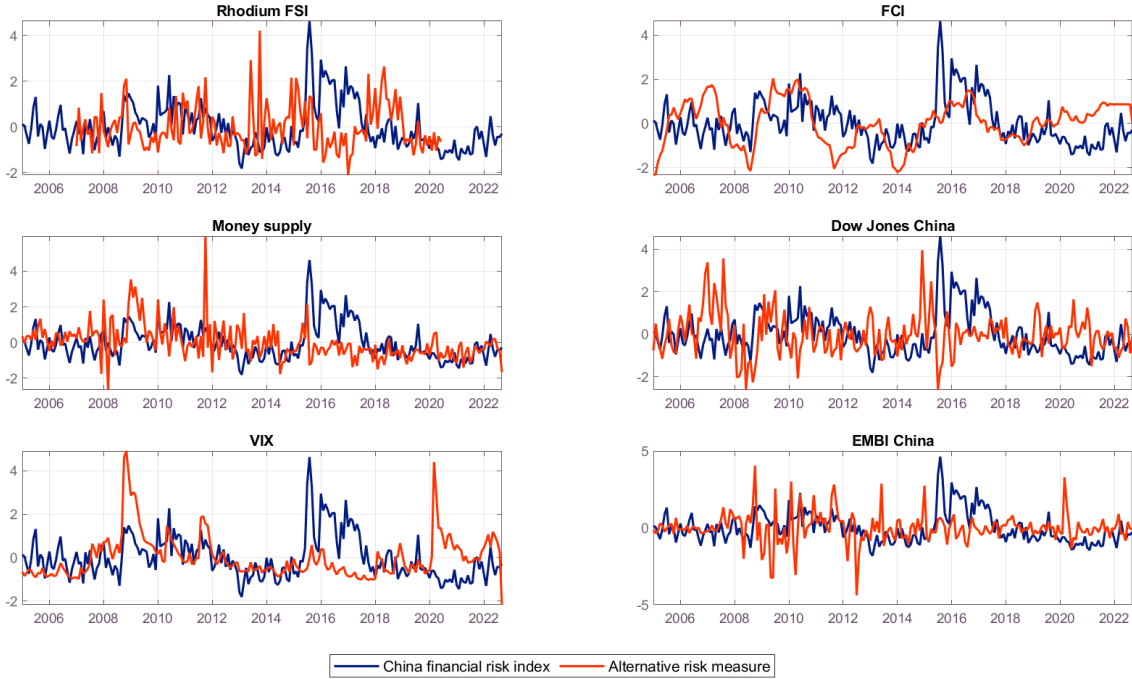
2019),<sup>78</sup> the economic policy uncertainty index developed by Baker et al. (2013), both based either on the SCMP or on mainland newspapers, the Rhodium Group’s China Risk Matrix Index and a financial conditions index (FCI) from Arrigoni et al. (2020). Figure 7 plots our financial risk index (in blue) against each alternative measure considered (in red). Although our indicator is not constructed to specifically capture *systemic* risk, among the indicators used to benchmark our index the CISS China is the one which shows the highest correlation with it (almost 0.4). Indeed, both indices spike around similar events, although with different magnitudes. Interestingly, the CISS index increased substantially in 2020Q1, when the COVID-19 outbreak erupted, while our indicator does not show a spike in the same period. One reason behind this difference is that the event was well captured by a specific topic produced by the LDA procedure (see Appendix A Figure A.3 third subplot) labelled *pandemic and health*, which we however decided to exclude from our core index. Our financial risk index is poorly correlated with the economic policy uncertainty indices, reassuringly suggesting that we are capturing something different from uncertainty about economic policies. The correlation with financial variables such as the equity prices and commodity prices, is mainly the result of the correlation with our financial markets sub-index, while unsurprisingly, the bilateral exchange rate between the renminbi and the US dollar is also correlated with the measure mainly through the exchange rate sub-index.

**Figure 7:** Comparison of the financial risk index with alternative risk measures



<sup>7</sup> See also Hollo et al. (2012) for the methodology used to construct the index.

<sup>8</sup> The China-CISS defines systemic risk events as “episodes when both the covariance and co-extremeness across markets are jointly high as “systemic stress events” and is “based on 13 financial indicators of the equity market, the bond market, financial institutions and the foreign exchange market” (Ma et al., 2019).



**Source:** Authors’ calculations, ECB, China’s National Bureau of Statistics, People’s Bank of China, JP Morgan, Wall Street Journal, State Administration of Foreign Exchange, Shanghai Stock Exchange, Rhodium Group and <https://www.policyuncertainty.com/>. EPU (SCMP) is the economic policy uncertainty index (Baker et al., 2013) based on the South China Morning Post. TPU is economic uncertainty index based on mainland newspapers. FCI index from Arrigoni et al. (2020).

**Notes:** Time coverage 2005M01-2022M09. Series are standardised to have zero mean and standard deviation equal to one over the sample considered. The financial risk index series is scaled monthly by the total number of articles related to China published by the newspapers considered. Correlation of the index with the measures presented in the figure are (from top left to bottom right): 0.39; -0.12; -0.22; 0.25; -0.16; -0.10; -0.05; -0.09; -0.05; -0.07; 0.21; 0.16; -0.18; 0.14; 0.0.

## 5 Empirical analysis

As standard in the literature on risk and uncertainty (e.g., Baker et al., 2016), we explore the relationship between our overall financial risk indicator and various macro and financial variables using a structural vector autoregression (SVAR) framework. More precisely, we estimate the model as the  $p$ th-order VAR:

$$y_t = B_1 y_{t-1} + \dots + B_p y_{t-p} + u_t$$

$$u_t \sim N(0, \Sigma),$$

where  $y_t$  denotes a  $q \times 1$  vector of endogenous variables,  $u_t$  a  $q \times 1$  vector of errors and  $B_1, \dots, B_p$  and  $\Sigma$  represents matrices of suitable dimensions containing the unknown parameters of the model. More precisely, the coefficients of lagged endogenous variables are in the matrices  $B_1, \dots, B_p$  while  $\Sigma$  is the covariance matrix. To overcome possible “overfitting” issues we employ Bayesian estimation techniques using normal-inverse Wishart priors and

standard values for the hyperparameters.<sup>9</sup>

## 5.1 Baseline results

For the baseline specification we include the following set of variables ordered from the most exogenous to the most endogenous: the natural logarithm of global industrial production exports to China, the natural logarithm of global oil prices, the emerging market sovereign spread (EMBI Global spread), the natural logarithm of the Chinese equity price index (CN equity index), our financial risk measure, the natural logarithm of the Chinese industrial production index, the natural logarithm of the Chinese consumer price index (CPI), and the Chinese 7-day repo rate. Most of the data are collected from national sources and other data providers via Haver Analytics, the consumer price index and the Chinese 7-day repo rate are taken from the Federal Bank Reserve of Atlanta China's Macroeconomy: Time Series Data database (see Table 2 and its footnote for additional details).

**Table 2**

<b>VAR setup</b>	<b>Source</b>
ln(global industrial production ex China)	Netherlands Bureau for Economic Policy Analysis/Haver Analytics
ln(oil prices)	Financial Times/Haver Analytics
EMBI Global spread	JP Morgan/Haver Analytics
ln(CN equity price index)	Dow Jones/Haver Analytics
Financial risk measure	Authors' calculations
ln(Chinese industrial production)	China National Bureau of Statistics/Haver Analytics
ln(CN CPI)	China National Bureau of Statistics/Federal Reserve Bank of Atlanta
Chinese 7-day repo rate	People's Bank of China/Federal Reserve Bank of Atlanta

**Notes:** Oil prices refer to brent crude oil (\$/bbl), Brent crude oil prices starting in 2007 are Brent blended prices from the Financial Times. Prices prior to 2007 are European Free Market Prices from the Wall Street Journal. The Chinese equity price index is the Dow Jones China 88. Industrial production measures and the consumer price index are seasonally adjusted by sources. The Federal Reserve Bank of Atlanta provides seasonally adjusted monthly [data series for the Chinese economy](#) (for additional details see also Higgins et al., 2016).

A shock to financial risk is identified with a recursive identification procedure obtained by using a Cholesky decomposition of the covariance matrix of the VAR reduced-form residuals. In doing so, we place the global variables at the beginning of the VAR given their more exogenous nature. Global industrial production is a global business cycle indicator which measures monthly developments in price-adjusted global output. Global oil prices

<sup>9</sup> To estimate the model, we use version 5.0 of the BEAR Toolbox (see Dieppe et al. (2016) for additional information on the estimation procedure).

are included to account for the important role played by China in the market of this commodity. The EMBI Global sovereign spread measures the spread of government bonds issued by several emerging markets against a US benchmark. Incorporating such indices before our financial risk measure guarantees that any global exogenous shocks that could be latent in our financial risk index is being adjusted for. Moreover, including the stock market index mitigates concerns of endogeneity because stock markets are forward-looking and stock prices react to all sources of information (Baker et al., 2016). The remaining domestic variables are ordered after the faster-moving financial market data. Specifically, we include the industrial production index as a proxy for China's activity, the consumer price index and the 7-day repo rate which is a widely used measure to proxy China's monetary policy stance.<sup>10</sup> Finally, we estimate the model using monthly data over the period 2005M1- 2019M12 using 12 lags.<sup>11</sup> We explicitly exclude the period covering the COVID-19 outbreak to avoid large outliers to affect our estimation results. Figure 8 displays the impulse responses of the variables included in the model to a one standard deviation increase in the financial risk indicator.

Overall, we observe a negative and statistically significant impact on global industrial production excluding China, which contracts by around 0.2pp while the response of China's activity is qualitatively similar but not significant.<sup>12</sup> A possible explanation for this inconsistency might be given by the constant effort exerted by authorities in stabilising growth in China through active fiscal policy and direct support to SOEs in the period considered during the empirical analysis. This might have in turn limited the responsiveness of real activity proxies to financial shocks. Alternatively, those results underline the difficulty of using indicators which are capable of properly tracking China's activity due to a potential interference of Chinese authorities with official statistics (see Fernald et al. (2021) among others for a discussion on the issue).

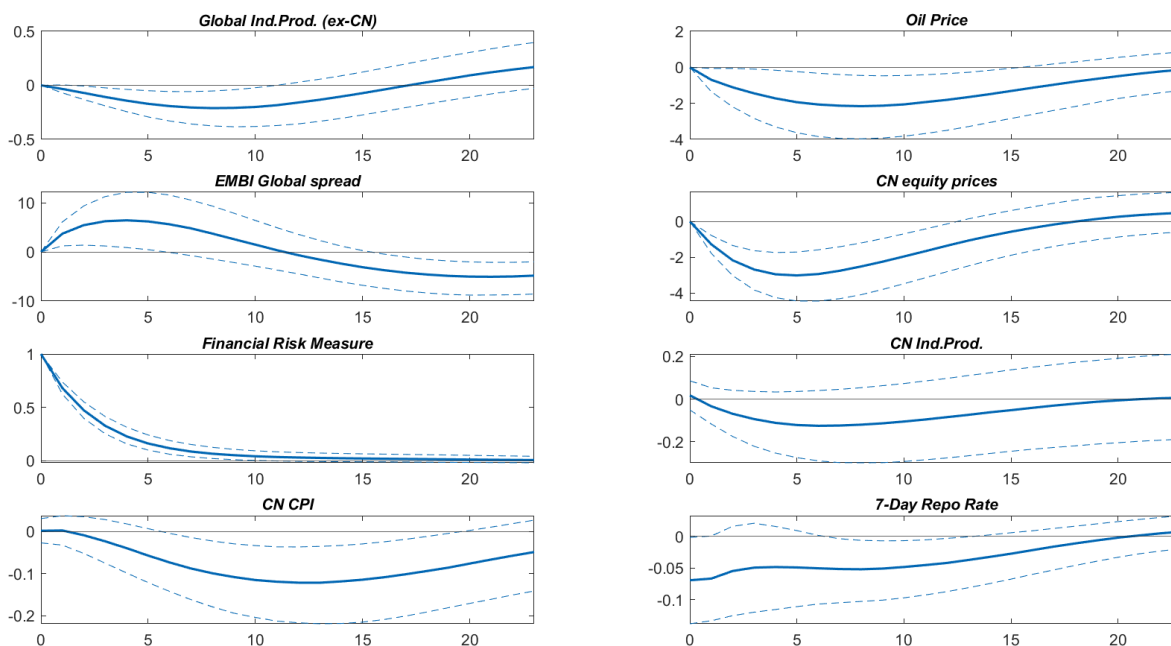
In line with the contraction in global activity, and potentially lower demand from China, which represents the world's second largest consumer of oil, oil prices decline by around 2pp in response to the shock to financial risk. The EMBI spread instead, increases by around 7bps, suggesting a tightening in financial conditions in emerging markets, while China's equity prices decline by around 3pp after around 4 to 6 months before returning to previous levels after around a year. The decline in the consumer price index in China is consistent with an effect of heightened financial risk akin to that of a negative demand shock. Consistently, we observe a downward movement in China's repo rate in response to the shock, pointing to a loosening of monetary policy to counteract the increase in financial risks.

<sup>10</sup> On the informativeness of the 7-day repo rate on China's monetary policy stance see Kamber and Mohanty (2018) among others.

<sup>11</sup> The choice of the number of lags is rather standard considering the model frequency. Results are qualitatively and quantitatively similar if we change the number of lags to 6 or 8 (see Appendix B, Figure B1 and B2).

<sup>12</sup> This result seems to be confirmed also if we use alternative proxies of activity such as retail sales as shown in Appendix B Figure B3.

**Figure 8:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator.



**Source:** Authors' calculations

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval.

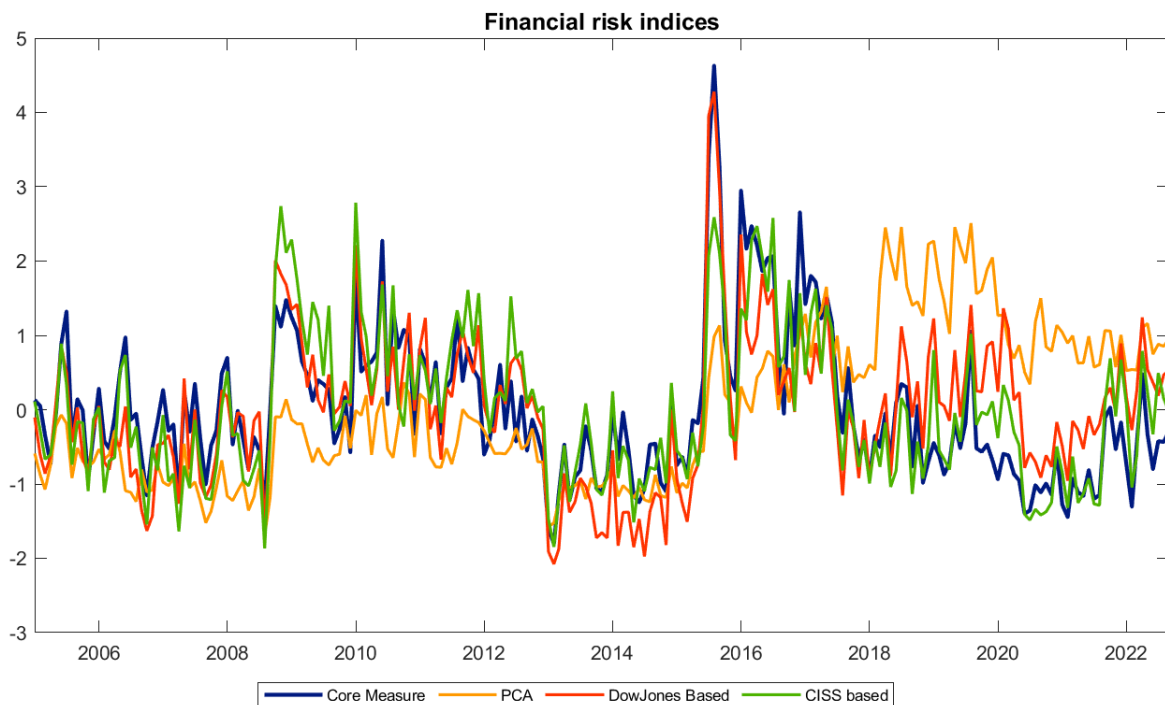
## 5.2 Robustness checks

In this section, we present some robustness exercises concerning both the construction of the financial risk index and the empirical analysis.

**Construction of the financial risk index:** Lacking an objective reliable strategy to construct a financial risk index, in this section we present the results from adopting alternative approaches. Specifically, we construct three alternative indicators. The first one is based on principal component analysis (PCA), while the second and the third make use of regression analysis. The PCA index is constructed by taking the first principal component of all 30 topics identified with the LDA procedure in Section 3.2. This approach has the advantage of being more objective than the one proposed in Section 4.1, but given that it considers all topics identified, a disadvantage might be that the variance captured by the first principal component could reflect some commonality in the variations of the indices that is not strictly due to financial risk. The PCA based index (yellow line, Figure 9) is poorly correlated to our core indicator (blue line) and it does not spike substantially during the stock market crash in 2015, while it is consistently higher since 2018, when US-China trade war began. The second and third indicators have been constructed by selecting a subset of the 30 topics identified with the LDA procedure on the basis of their explicatory power with reference to two measures which either track directly financial

developments in China (i.e. the China Dow Jones index) or are related to financial stress in the Chinese economy (CISS index). To do so, we run two regression models, one with the China Dow Jones index and one with the CISS index as dependent variable. Then, starting with the sub-set of indicators used for our core index presented in Section 4.1, we run different specifications of those models changing in turn the explanatory variables by drawing them from the 30 identified topics. Tables B.1. and B.2. in the Appendix B present in details the results of this exercise. In both cases, we selected the model with the highest adjusted R-squared (Model 6) and aggregated the topics used as regressors in those models to construct alternative measures of financial risk. Those two indicators (red and green lines in Figure 9) correlate well with our core index (0.78 for the Dow Jones based index and 0.85 for the CISS based index) and display very similar dynamics, both qualitatively and quantitatively.<sup>13</sup>

**Figure 9:** Alternative financial risk indices



**Source:** Authors' calculations.

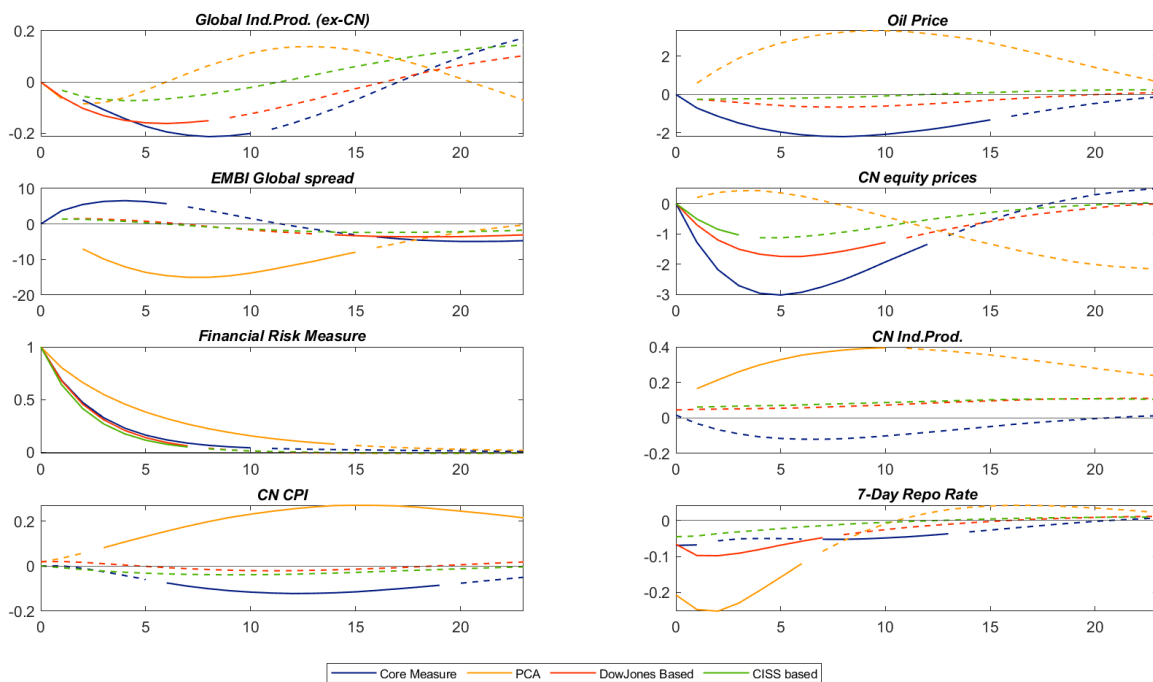
**Notes:** Time coverage 2005M01-2022M09. The series are scaled monthly by the total number of articles related to China published by the newspapers considered. The 'core measure' is the index presented in Section 4.1, the 'PCA' index has been constructed taking the first principal component of all topics identified with the LDA procedure while the 'Dow Jones Based' and the 'CISS based' indices have been constructed considering only the topics that have the highest explicatory power with reference to those variables respectively.

Figure 10 shows the impulse response functions of macro-financial variables to

<sup>13</sup> Table B.3. in Appendix B presents a correlation matrix of the four indicators.

shocks to the four different financial risk measures considered using the same empirical model presented in Section 5.1. The alternative indices constructed using regression analysis produce qualitative similar IRFs compared to shocks to our core index, although with a rather smaller magnitude, while the PCA indicator responses are in general less significant and in some instances (EMBI Global spread, oil prices, CPI) qualitatively different from the other indicators.

**Figure 10:** Impulse-response functions of macro-financial variables to shocks in four different financial risk indicators.



**Source:** Authors' calculations

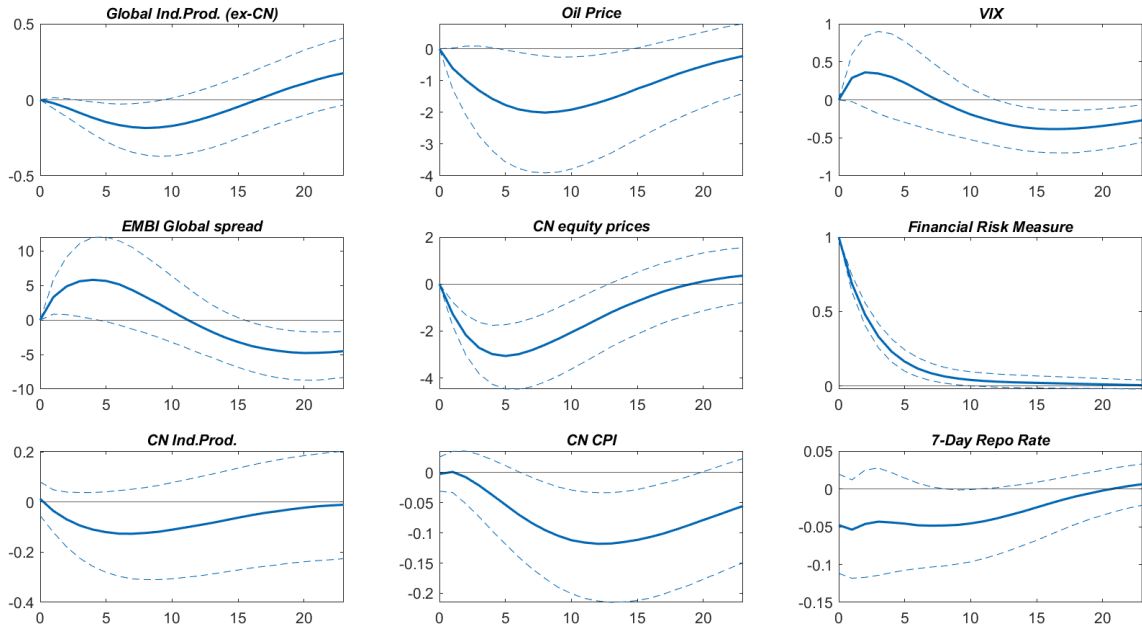
**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Continuous lines report median IRFs which lay inside the 68% credibility interval, while dotted lines refer to median IRFs laying outside the 68% credibility interval.

**Empirical analysis:** To run some robustness checks around the baseline model described in Section 5.1, we extend the SVAR model to control for other widely used measures of risk and uncertainty. To this end, we first add to the model the implied volatility index (VIX). This index is based on market data, is forward-looking and is often referred to as the investor fear gauge: The higher the index, the greater the fear (Whaley, 2000). Significantly, implied volatility indices are often used as a proxy for overall uncertainty (see for example Baker et al. (2016) and Gulen and Ion, 2016). In addition, we consider the Chinese economic policy uncertainty (EPU) index developed by Baker et al. (2016). Although fundamentally different, in the sense that this latter indicator captures events related to government policy uncertainty while our deals with purely financial risk, there is one component in the EPU index related to uncertainty regarding financial regulation. For this reason, it might be possible



that our financial risk indicator is embedded in the overall EPU. Following a conservative approach, we include these two variables (one at a time) before our financial risk measure. As can be seen in Figures 11 and 12, shocks to our financial risk measure have almost identical effects on both foreign and domestic variables as those described in the previous section. Additional robustness checks are shown in Appendix B.

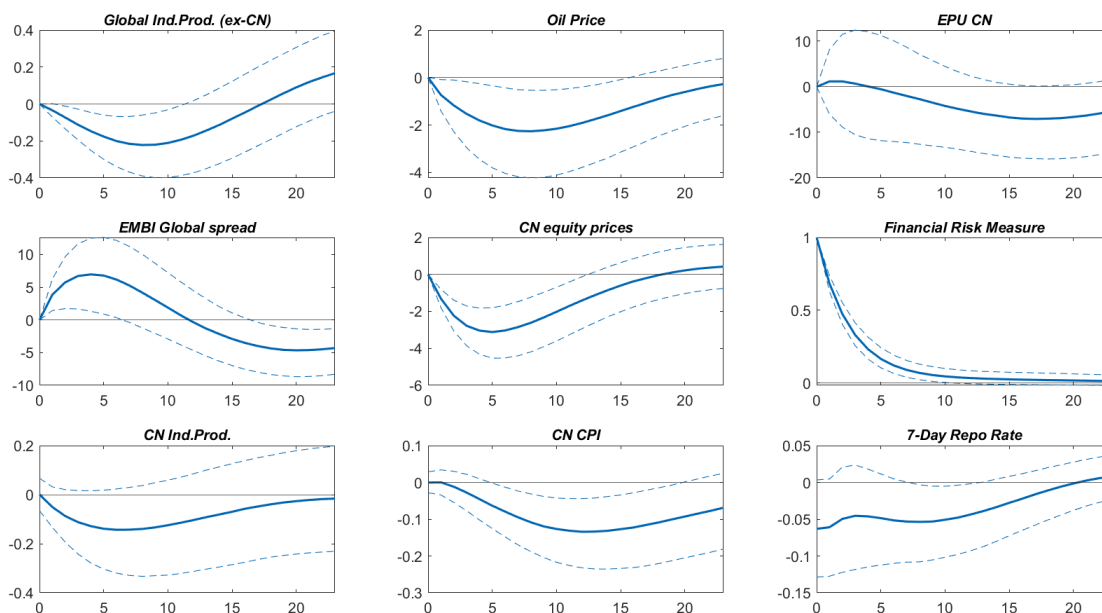
**Figure 11:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator when controlling for the US VIX index.



**Source:** Authors' calculations

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval.

**Figure 12:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator when controlling for the Chinese Economic Policy Uncertainty Index.



**Source:** Authors' calculations

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval.

## 6 Conclusion

In this paper, we present a model-based bottom-up approach to estimate an indicator of overall financial risk in China together with its constituent sub-components. To construct the indicator, we apply an unsupervised machine learning algorithm on a large number of newspaper articles reporting on financial risk in China published since 2005. This strategy has the benefit of endogenously extracting individual financial risk components while at the same time assessing their weight and impact on the overall financial risk indicator. We find that our core measure of financial risk correlates with other indicators of risk used for the Chinese economy and that the indicator captures financial risk episodes in China and does so in a timely fashion.

Using an SVAR framework to measure the possible interactions between our financial risk indicator and global and domestic macroeconomic and financial variables, we find that increases in overall financial risk foreshadow a decline in global oil prices and equity prices in China. In addition, we find that shocks to financial risk in China spillover globally as world industrial production contracts in response to an increase in financial risk, while emerging markets government bond spreads increase. Furthermore, while the decline in

industrial production in China in response to the shock is not statistically significant, the decline in the consumer price index coupled with a loosening of monetary policy suggest that the shock affects the economy similarly to a negative demand shock.

Overall, these results suggest that a textual analysis of financial risks in China can represent a valid alternative to conventional quantitative measures to track financial risk and that this methodology could be in principle extended to other countries. Moreover, by constructing sub-indicators, our methodology allows to identify and categorise the sources of financial risk, which might be useful to track risk in specific sectors of the economy. A potential limitation of our analysis relates to the extent to which newspaper publications are affected by censorship and manipulation by Chinese authorities. Our data coverage and approach seek to avoid this problem, but it cannot be entirely ruled out that our indicator is somehow distorted by authorities' information and editorial control.

Future research could seek to incorporate an additional metric of the risk embedded in the media in the form of sentiment analysis. Doing so would likely increase the precision of the estimate and provide a timelier account of the financial risks faced by the Chinese economy. Furthermore, the literature on uncertainty has underlined the challenges associated with correctly assessing whether risk and uncertainty are endogenous to the business cycle or rather an exogenous source of fluctuations (Ludvigson et al., 2021), paving the way to new identification solutions. An extension of our work along those lines would also provide more clarity on the potential implications of rising financial risk on both real and nominal variables.

## References

- Acharya, V.V., Qian, J., Su, Y., and Yang, Z. (2020), “In the shadow of banks: Wealth management products and issuing banks’ risk in China”, CEPR Discussion Paper No. DP14957.
- Apostolou, A., Al-Haschimi, A., and Ricci, M. (2022), “Financial risks in China’s corporate sector: real estate and beyond”, Economic Bulletin Article, European Central Bank, Vol. 2.
- Arrigoni, S., Bobasu, A., and Venditti, F. (2020), “The simpler the better: measuring financial conditions for monetary policy and financial stability”, Working Paper Series No. 2451, European Central Bank.
- Baker, S., Bloom, N., Davis, S.J., and Wang, X. (2013), “Economic Policy Uncertainty in China,” mimeo, University of Chicago.
- Baker, S., Bloom, N., and Davis, S.J. (2016), “Measuring economic policy uncertainty”, *The Quarterly Journal of Economics* Vol. 131, pp. 1593–1636.
- Bian, J., He, Z., Shue, K., and Zhou, H. (2018), “Leverage-induced fire sales and stock market crashes”, NBER Working Paper 26040, National Bureau of Economic Research, September.
- Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003), “Latent dirichlet allocation”, *Journal of Machine Learning Research* Vol. 3, pp. 993–1022.
- Caldara, D., and Iacoviello, M. (2022), “Measuring geopolitical risks”, *American Economic Review*, Vol. 112(4), pp. 1194-1225.
- Chen, H. and Tillmann, P. (2021), “Monetary policy uncertainty in China”, *Journal of International Money and Finance*, Vol. 110, 102309.
- Dang, T.V., Liu, L., Wang, H., and Yao, A. (2019), “Shadow banking modes: The Chinese versus US system”, mimeo, Columbia University.
- Davis, S.J., Liu, D., and Sheng, X.S. (2019), “Economic policy uncertainty in China since 1949: The view from mainland newspapers”, Chicago Booth Research Paper 19-04.
- Dieppe, A., van Roye, B., and Legrand, R. (2016), “The BEAR toolbox”, Working Paper Series No. 1934, European Central Bank.
- Fernald, J. G., Hsu, E. Spiegel, M.M. (2021), “Is China fudging its GDP figures? Evidence from trading partner data”, *Journal of International Money and Finance*, Vol. 110, 102262.
- Gentzkow, M., Kelly, B., and Taddy, M. (2019), “Text as data”, *Journal of Economic Literature* Vol. 57(3), pp. 535–74.
- Griffiths, T.L. and Steyvers, M. (2004), “Finding scientific topics”, *Proceedings of the National Academy of Sciences* Vol. 101, pp. 5228-5235.
- Gulen, H. and Ion, M. (2016), “Policy Uncertainty and Corporate Investment”, *The*

- Review of Financial Studies*, Vol. 29(3), pp. 523–564.
- He, F., Ma, Y., and Zhang, X. (2020), “How does economic policy uncertainty affect corporate innovation? Evidence from China listed companies”, *International Review of Economics & Finance* Vol. 67, pp. 225–239.
- Higgins, P., Zha, T., and Zhong, W. (2016), “Forecasting China’s economic growth and inflation”, *China Economic Review* Vol. 41, pp. 46-61.
- Hollo, D., Kremer, M., and Lo Duca, M. (2012), “CISS - a composite indicator of systemic stress in the financial system”, Working Paper Series No. 1426, European Central Bank.
- Huang, Y. and Luk, P. (2020), “Measuring economic policy uncertainty in China”, *China Economic Review* Vol. 59, 101367.
- Kamber, G. and Mohanty, M. (2018), “Do interest rates play a major role in monetary policy transmission in China?” Working Paper No. 714, Bank for International Settlements.
- Larsen, V.H. and Thorsrud, L.A. (2019), “The value of news for economic developments”, *Journal of Econometrics* Vol. 210, pp. 203–218.
- Li, M. and Wu, G. (2020), “The impact of economic policy uncertainty on real estate development in China”, *Journal of Applied Finance and Banking* Vol. 10, pp. 25–42.
- Liu, G. and Zhang, C. (2020), “Economic policy uncertainty and firms’ investment and financing decisions in China”, *China Economic Review* Vol. 63, 101279.
- Lo, C. (2019), “China’s bank failures: a turning point for the system”, BNP Paribas Asset Management.
- Ludvigson, S.C., Ma, S., and Ng, S. (2021), “Uncertainty and business cycles: Exogenous impulse or endogenous response?” *American Economic Journal: Macroeconomics*, Vol.13(4), pp. 369-410.
- Ma, J., Kremer, M., He, X., Liu, S., and Wang, L. (2019), “Monitoring of China’s systemic financial stress”, Finance and Development Research Center, National Institute of Finance, Tsinghua University. URL: <https://www.pbcfsf.tsinghua.edu.cn/portal/article/index/id/3929.html>. Original document in Chinese.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), “Efficient estimation of word representations in vector space”, mimeo.
- Ormerod, P., Nyman, R., Tuckett, D. (2015), “Measuring financial sentiment to predict financial instability: A new approach based on text analysis”, mimeo.
- Piškorec, M., Antulov-Fantulin, N., Novak, P.K., Mozetič, I., Grčar, M., Vodenska, I., and Šmuc, T. (2014), “Cohesiveness in financial news and its relation to market volatility”, *Scientific Reports* Vol. 4, 5038.
- Sha, Y., Kang, C., and Wang, Z. (2020), “Economic policy uncertainty and mergers and

- acquisitions: Evidence from China”, *Economic Modelling* Vol.89, pp. 590-600.
- Song, Z. and Xiong, W. (2018), “Risks in China’s financial system”, *Annual Review of Financial Economics* Vol. 10, pp. 261–286.
- Sun, G. (2019), “China’s shadow banking: Banks’ shadow and traditional shadow banking”, Working Paper No. 822, Bank for International Settlements.
- Whaley, R. (2000), “The Investor Fear Gauge,” *The Journal of Portfolio Management* Vol. 26(3), pp. 12-17.
- Wong, C. (2011), “The fiscal stimulus programme and public governance issues in China”, *OECD Journal on Budgeting* Vol. 11/3.

# Appendices

## Appendix A – Additional figures and tables LDA and search

The below table shows the first 30 words that were identified as being most similar to risk and finance by applying the word2vec algorithm. Judgement has been then applied to select the most relevant words for our final search query.

**Table A.1 Expanded search term list**

<b>Term</b>	<b>Similarity with finance</b>	<b>Term</b>	<b>Similarity with risk</b>
financing	0.43	risks	0.64
proposed	0.39	leverage	0.51
govern	0.38	problem	0.47
jiwei	0.38	conditions	0.43
prime	0.38	standards	0.43
banking	0.37	nature	0.43
lending	0.36	opaque	0.43
commerce	0.36	given	0.42
securitiesfinance	0.35	painful	0.41
tsipras	0.35	backdrop	0.41
mckinsey	0.35	unstable	0.40
jointly	0.34	extremely	0.40
tighten	0.34	technical	0.40
drafting	0.34	stable	0.40
regulate	0.34	situation	0.39
trustee	0.33	danger	0.39
chiefly	0.32	vasudevan	0.39
factual	0.32	solvency	0.39
nagano	0.32	proponent	0.39
broaden	0.32	regulation	0.39
jams	0.32	defaults	0.38
padoan	0.31	governance	0.38
amplified	0.31	quality	0.38
guidelines	0.31	unwinding	0.38
poorest	0.31	deterioration	0.38
benefitting	0.31	causes	0.38
peer	0.31	avoid	0.38
designed	0.30	vulnerabilities	0.38
shinzo	0.30	volatility	0.38
calls	0.30	risky	0.37

**Figure A.1: Other identified topics**

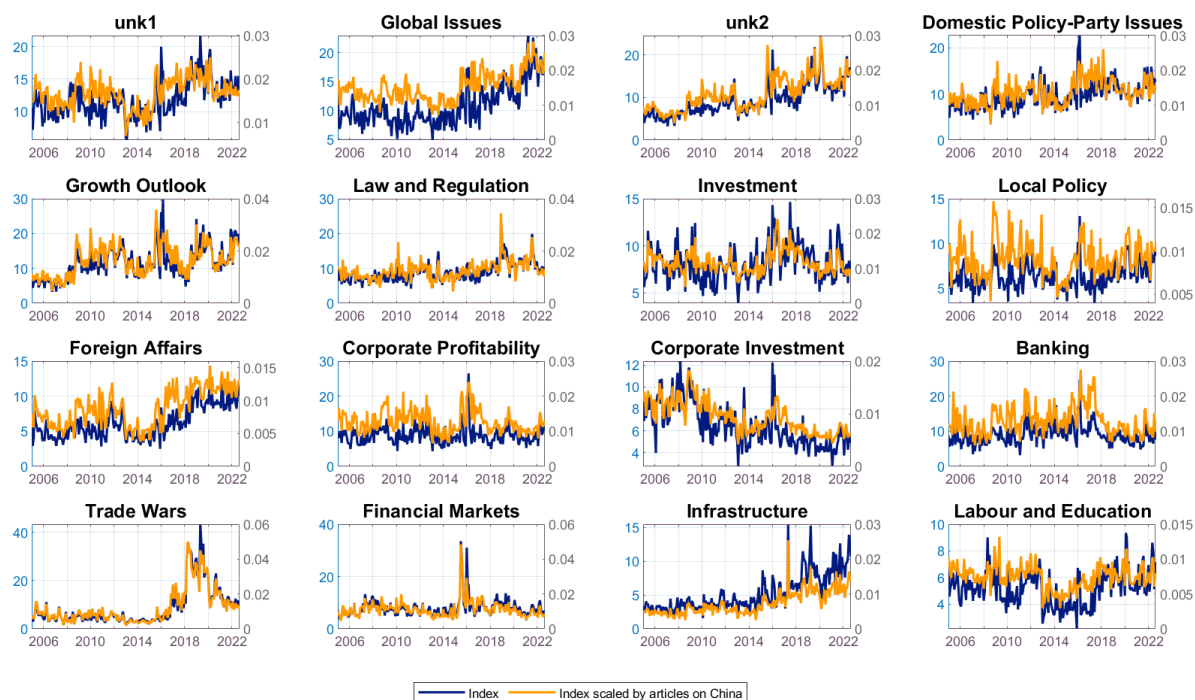


**Source:** Authors' calculations.

**Notes:** The size of the words in the word clouds reflect the number of occurrences of that word in the topic considered. As common with this procedure, some topics end up being a collection of common words that do not form a logically coherent group. These are indicated with “unk. 1” and “unk. 2” which is short for unknown/unidentified.

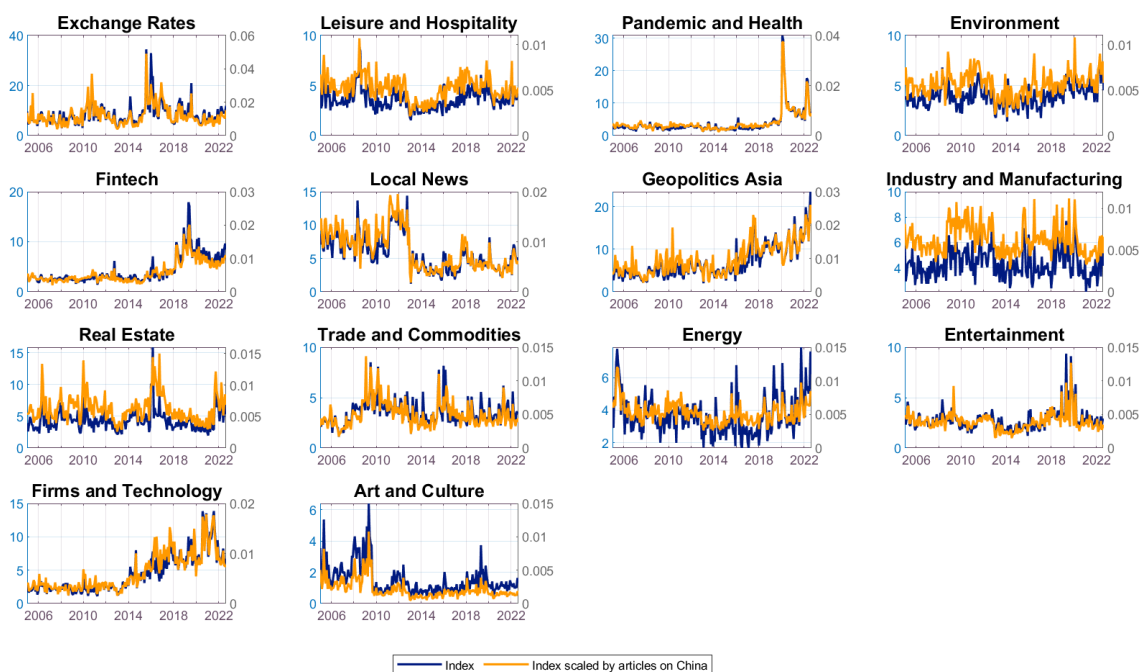


**Figure A.2:** Indices obtained with the LDA procedure.



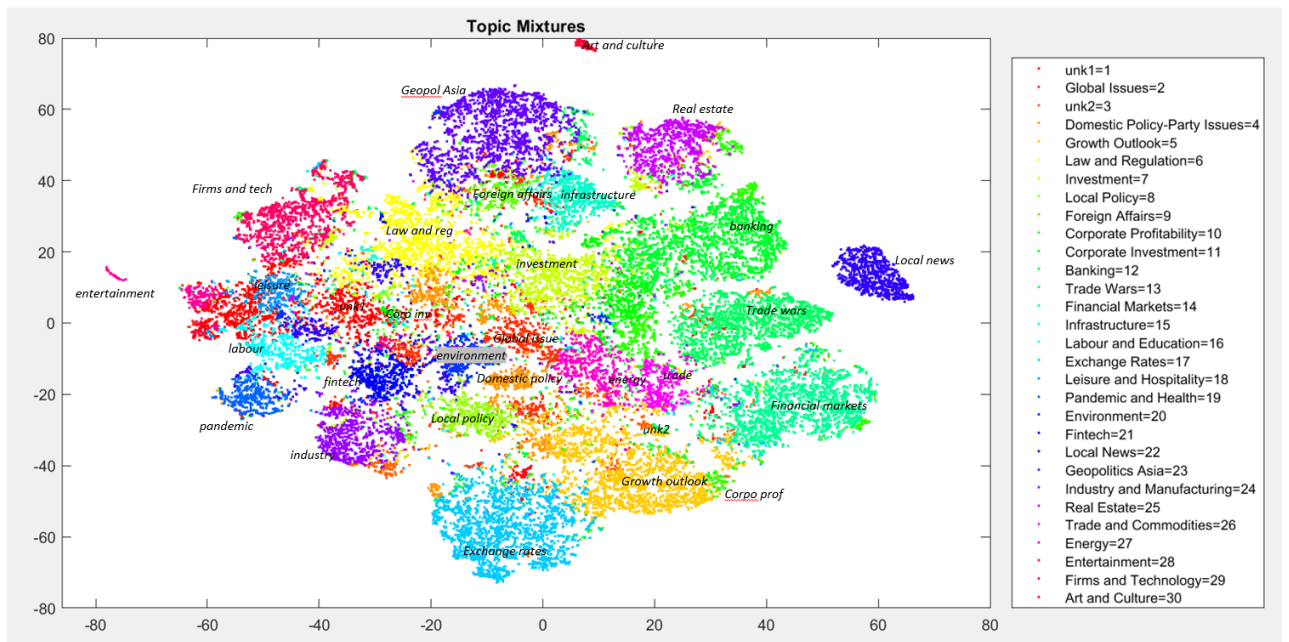
**Source:** Authors' calculations.

**Figure A.3:** Indices obtained with the LDA procedure.



**Source:** Authors' calculations.

**Figure A.4:** Proximity of the topics identified with the LDA procedure.



**Source:** Authors' calculations.

## Appendix B – Additional results: Robustness

Table B.1. and B.2. present results from regression analysis run to identify the topics used to construct two additional financial risk indicators. More precisely, both tables present 6 different models which differ for the indicators used as regressors. Table B.1 presents regression analysis where the dependent variable is the China Dow Jones index, while Table B.2 presents regression analysis where the dependent variable is the CISS index. Starting from Model 1, where the regressors are the topics used to construct our core index in Section 4.1., we try in turn different model specifications by changing the regressors – guided by their pairwise correlation with the dependent variables - with the aim of finding the model that maximizes the adjusted R-squared of the regressions. In both cases the model that delivers the highest adjusted R-squared is Model 6. In the case of the Dow Jones China the model explains 0.14% of the variation of the dependent variable, in the case of the CISS the explained variance is almost double 0.35.

**Table B.1.:** Regression analysis results with China Dow Jones index as dependent variable.

VARIABLES	(1) DowJones	(2) DowJones	(3) DowJones	(4) DowJones	(5) DowJones	(6) DowJones
FinancialMarkets	-40.92** (17.94)	-32.05** (15.57)	-34.45** (15.03)	-34.83** (15.49)	-36.80** (15.03)	-33.35** (15.14)
Banking	25.54 (19.74)					
FX	6.898 (14.18)					
RealEstate	-41.02 (37.23)					
CorpProf	-62.07** (29.60)	-27.53 (31.31)	-38.46 (25.51)	-39.03 (26.11)	-46.84* (25.90)	
CorpInv	57.27** (26.81)	72.40** (30.68)	82.40*** (25.79)	82.19*** (25.92)	112.1*** (31.28)	86.47*** (27.97)
GrowthOutlook		-9.293 (15.39)				-31.31** (13.99)
LeisANDHosp		-162.5*** (51.32)	-163.9*** (51.18)	-165.0*** (52.29)	-205.0*** (56.65)	-222.1*** (56.93)
Energy		-92.48** (42.13)	-95.62** (41.74)	-96.18** (42.17)	-113.3*** (42.90)	-112.7*** (42.72)
IndANdManu				4.600 (42.69)		
ForeignAffairs					48.96* (29.47)	71.78** (32.38)
Constant	0.544* (0.290)	1.507*** (0.344)	1.452*** (0.331)	1.445*** (0.337)	1.183*** (0.367)	1.185*** (0.363)
Observations	213	213	213	213	213	213
Adjusted R-squared	0.056	0.126	0.129	0.124	0.136	0.143

Standard errors in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.2.:** Regression analysis results with China CISS index as dependent variable.

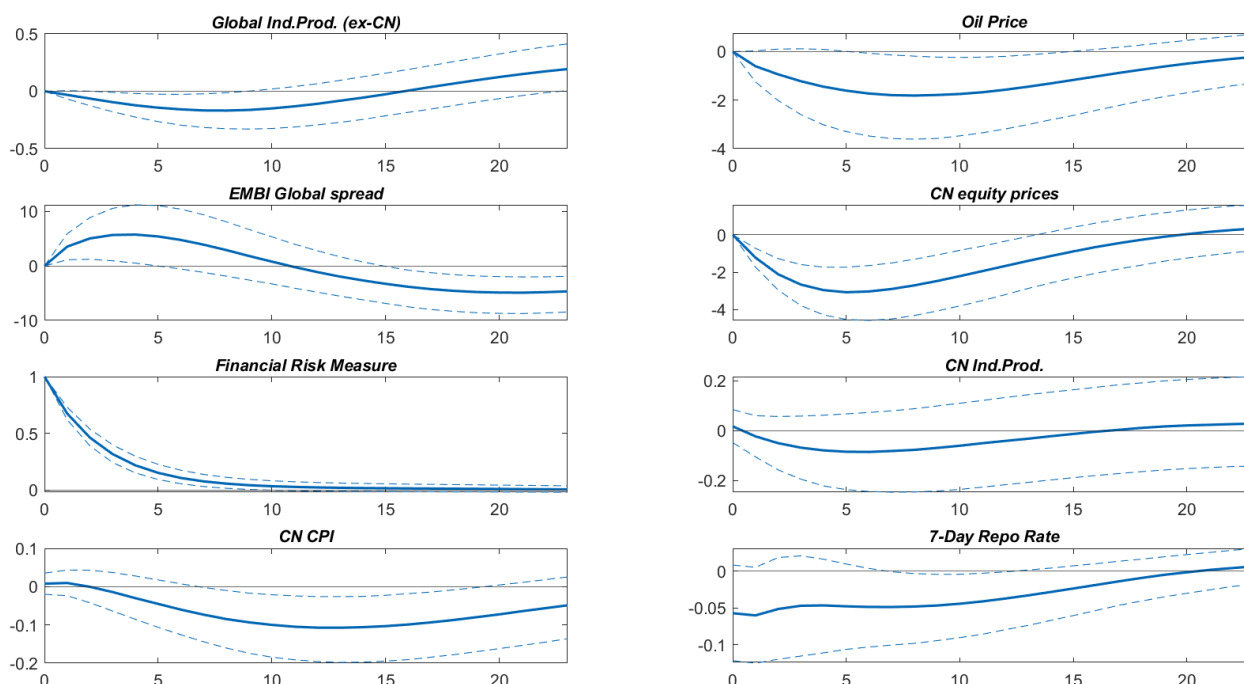
VARIABLES	(1) CISS	(2) CISS	(3) CISS	(4) CISS	(5) CISS	(6) CISS
FinancialMarkets	-11.22 (16.64)	-13.57 (15.90)	-12.07 (15.78)	-9.304 (15.69)	-10.82 (15.50)	
Banking	32.50* (18.20)	29.32* (17.40)	32.17* (17.02)	29.46* (16.92)	38.05** (17.07)	35.47** (16.79)
FX	-2.065 (13.18)	-14.92 (12.93)	-13.88 (12.85)	-10.32 (12.85)	-3.117 (13.03)	
RealEstate	-44.55 (35.79)	-61.15* (34.38)	-58.36* (34.18)	-78.32** (35.18)	-72.54** (34.81)	-75.74** (34.10)
CorpProf	89.78*** (27.17)	24.31 (29.99)				
CorpInv	96.02*** (26.10)	148.3*** (27.66)	159.4*** (23.99)	185.0*** (26.74)	138.1*** (32.74)	132.3*** (32.05)
GrowthOutlook		65.46*** (15.04)	71.58*** (13.00)	85.63*** (14.52)	80.98*** (14.46)	77.79*** (13.73)
IndANdManu				-88.13** (42.03)	-87.14** (41.49)	-94.28** (40.51)
ArtandCulture					134.8** (55.76)	141.3*** (53.97)
Constant	-1.986*** (0.263)	-2.419*** (0.270)	-2.401*** (0.269)	-2.236*** (0.278)	-2.234*** (0.274)	-2.206*** (0.268)
Observations	192	192	192	192	192	192
Adjusted R-squared	0.249	0.316	0.317	0.329	0.346	0.350

Standard errors in parentheses, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.3.:** Correlation across different financial risk indices and selected variables

	Core index	PCA based	Dow Jones Based	CISS based	CISS	Dow Jones China
Core index	1	0.15	0.78	0.85	0.39	-0.17
PCA based index	0.15	1	0.49	0.18	-0.09	-0.12
Dow Jones based index	0.78	0.49	1	0.83	0.40	-0.28
CISS based index	0.85	0.18	0.83	1	0.51	-0.18
CISS	0.39	-0.09	0.4	0.51	1	-0.24
Dow Jones China	-0.17	-0.12	-0.28	-0.18	-0.24	1

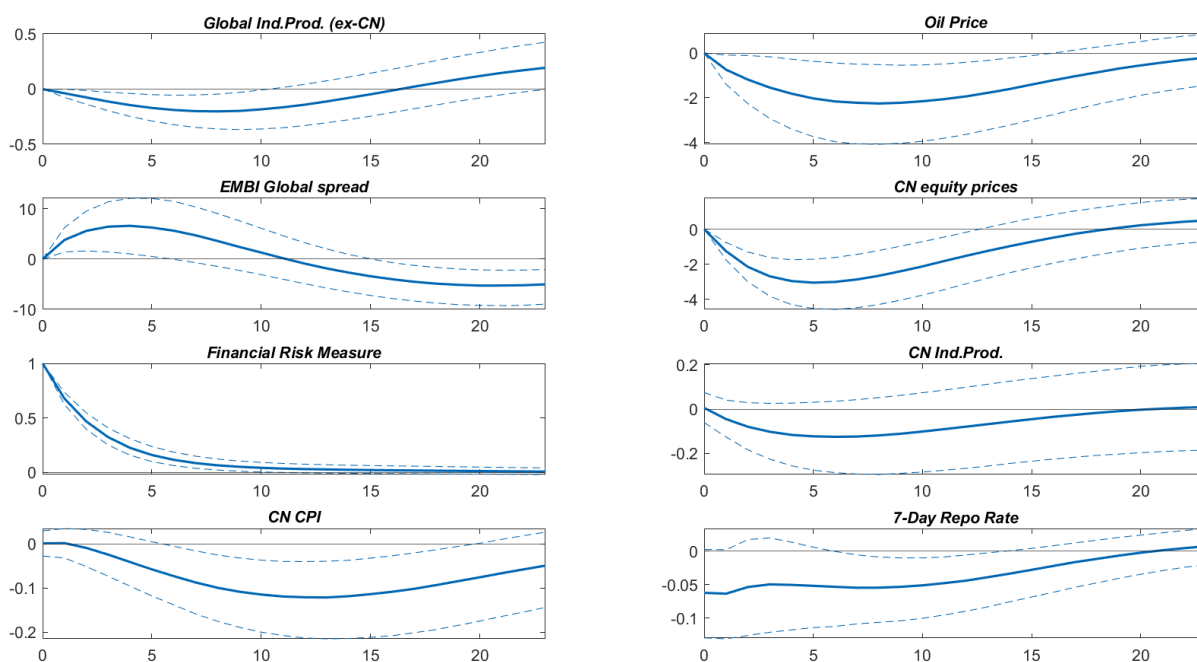
**Figure B.1:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator. 6 lags instead of 12



**Source:** Authors' calculations.

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval. Retail sales is used instead of industrial production for China as an alternative activity variable.

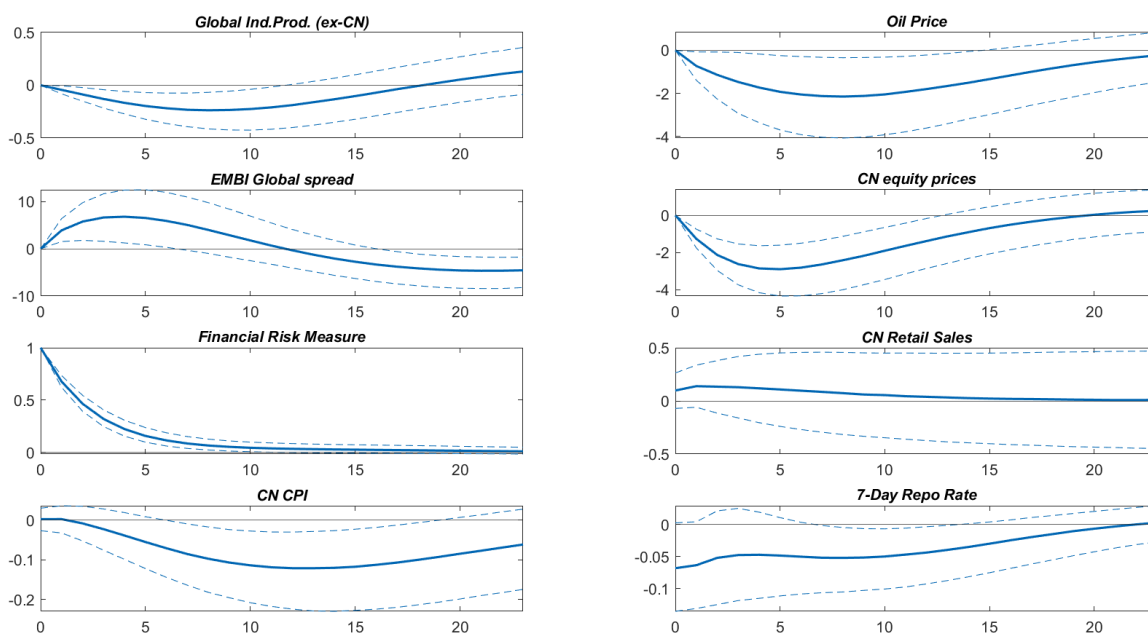
**Figure B.2:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator. 8 lags instead of 12



**Source:** Authors' calculations.

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval. Retail sales is used instead of industrial production for China as an alternative activity variable.

**Figure B3:** Impulse-response functions of macro-financial variables to shocks in the overall financial risk indicator. Alternative activity variable for China.



**Source:** Authors' calculations.

**Notes:** IRFs report percentage changes for all variables excluding EMBI Global spreads and the 7-day repo rate which are reported in bps. Dotted lines report the 68% credibility interval. Retail sales is used instead of industrial production for China as an alternative activity variable.

### Acknowledgements

The authors would like to thank an anonymous referee and participants at the “12th Annual International Conference on the Chinese Economy” organised by the Hong Kong Institute for Monetary and Financial Research of the Hong Kong Monetary Authority as well as participants at a seminar at the European Central Bank (ECB) for useful comments and suggestions. We are grateful to Martin Rasmussen for his excellent research assistance at an early stage of this project.

The paper was written while the author Apostolos Apostolou was at the European Central Bank.

The analysis and conclusions expressed herein are those of the authors and should not be interpreted as those of the ECB or the IMF.

### Alexander Al-Haschimi

European Central Bank, Frankfurt am Main, Germany; email: [Alexander.Al-Haschimi@ecb.europa.eu](mailto:Alexander.Al-Haschimi@ecb.europa.eu)

### Apostolos Apostolou

International Monetary Fund, Washington D.C., United States; email: [aapostolou@imf.org](mailto:aapostolou@imf.org)

### Andres Azqueta-Gavaldon

Sensyne Health, Oxford, United Kingdom; email: [Andres.Azqueta@gmail.com](mailto:Andres.Azqueta@gmail.com)

### Martino Ricci

European Central Bank, Frankfurt am Main, Germany; email: [martino.ricci@ecb.europa.eu](mailto:martino.ricci@ecb.europa.eu)

### © European Central Bank, 2023

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website [www.ecb.europa.eu](http://www.ecb.europa.eu)

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from [www.ecb.europa.eu](http://www.ecb.europa.eu), from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

PDF

ISBN 978-92-899-5509-6

ISSN 1725-2806

doi:10.2866/565243

QB-AR-23-004-EN-N