EUROPEAN CENTRAL BANK

EUROSYSTEM

Measuring Financial Risk in China using Textual Analysis

The 12th annual international conference on the Chinese economy

Hong Kong Monetary Authority

23/11/2022

Martino Ricci

European Central Bank

Joint with A. Al-Haschimi, A. Apostolou, A. Azqueta

Overview of the presentation

News-based financial risk measure in China:

- Motivation
- Literature

Analytical approach

- Data collection
- Standard textual analysis (count of articles)
- Topic Modelling - unsupervised machine learning for textual data (*Latent Dirichlet Allocation*: LDA)
- Construction of the index and decomposition or risk

Empirical analysis

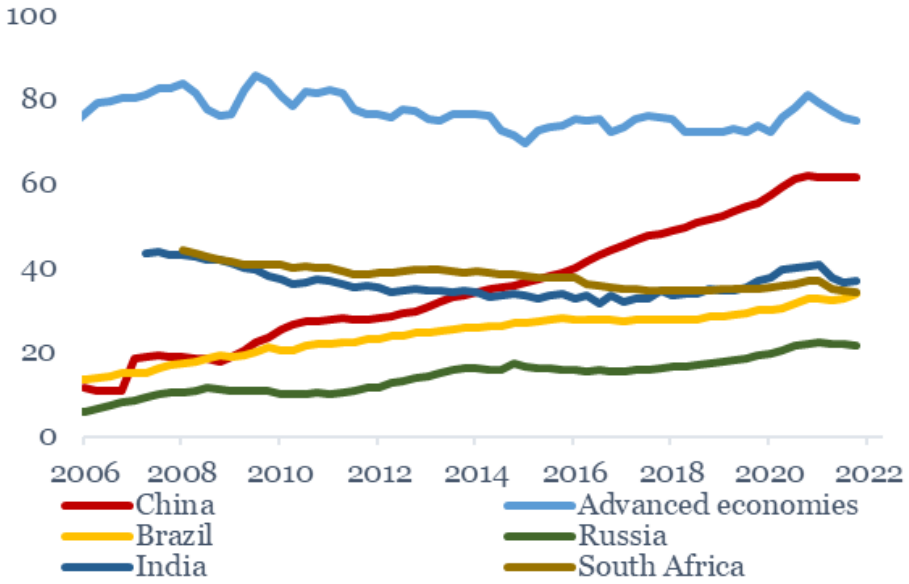
- Impact on macro-financial variables of increase in China's financial risk

1.

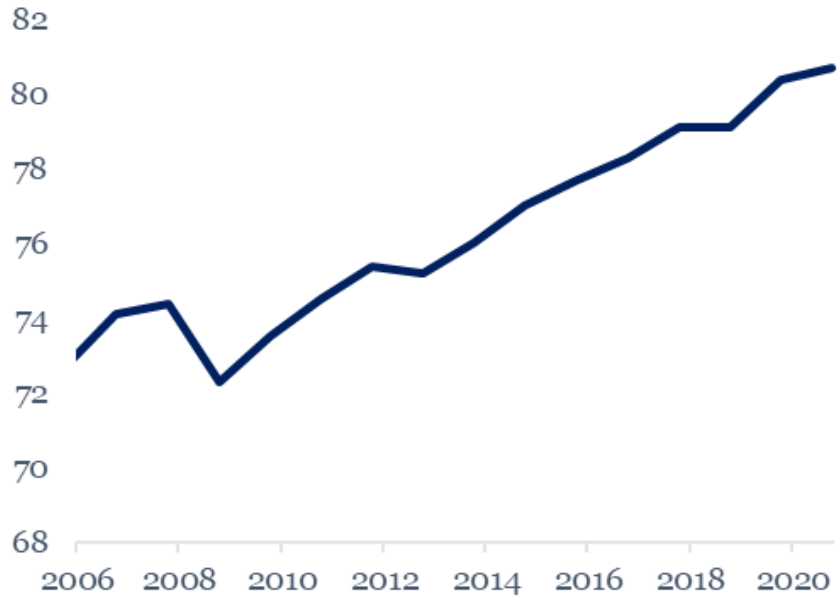
Objectives and Motivation

Motivation: Financial risk in China on the rise

Credit to the private non-financial sector (% of GDP)



Real Estate Investment: Ratio of Liabilities to Assets (%)



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.

Motivation: Monitoring China using hard data is becoming increasingly difficult

Literature of financial risk in China and text based approaches

Text-based approaches in Economics:

Baker et al. (2016): “Economic policy uncertainty”.

Caldara et al. (2019) “The economic effects of trade policy uncertainty”.

Caldara and Iacoviello (2019): “Measuring geopolitical risk”.

Angelico et al. (2020) “Can We Measure Inflation Expectations Using Twitter?”.

Financial risk/text based work on China:

Song et al. (2018) “Risks in China’s financial system”.

Ma et al. (2019) “Monitoring of china’s systemic financial stress”.

Wright and Rosen (2018) “Credit and Credibility: Risks to China’s Economic Resilience”.

Davis et al. (2019) “Economic policy uncertainty in China since 1949: The view from mainland newspapers.”

Huang and Luk (2020) “Measuring economic policy uncertainty in China.”

Chen and Tillmann (2021) “Monetary policy uncertainty in China”.

2.

The construction of the Financial Risk Index

Using textual analysis and topic modelling to gauge China’s financial risk and sources

Overview of the Analytical Strategy

- **Step 1:** We collect data (newspaper articles) from two authoritative sources on China and financial markets
- **Step 2:** We perform standard textual analysis -> filter articles relevant to increasing *financial risk*; data and pre-processing
- **Step 3:** Topic modelling -> unsupervised machine learning: assign a probabilistic distribution to identified articles of talking about specific topics.
- **Step 4:** Construct time-series based on topics; select most relevant topics aggregate in a single indicator and validate the measure.

Step 1: Data and pre-processing

THE WALL STREET JOURNAL.

South China Morning Post

Data, sources and timespan

- We collect articles from the **SCMP** and **the Wall Street Journal.**
- English version of the newspaper.
- Printed editions only (more constant reporting).
- January 2005 to September 2022. Total of around 36.000 articles.

Step 2: Filtering articles discussing increasing financial risk in China

Filtering articles discussing **financial risk** in **China** over time using **keywords:**

I. Initial terms: start search with all articles containing words *China/Chinese, risk* and *finance*.

II. Expand initial search using *word2vec* algorithm.

III. Select most relevant words associated with negative sentiment (i.e. danger/opaque/defaults).

IV. Repeat the search.

Final search:

- {(**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$)}

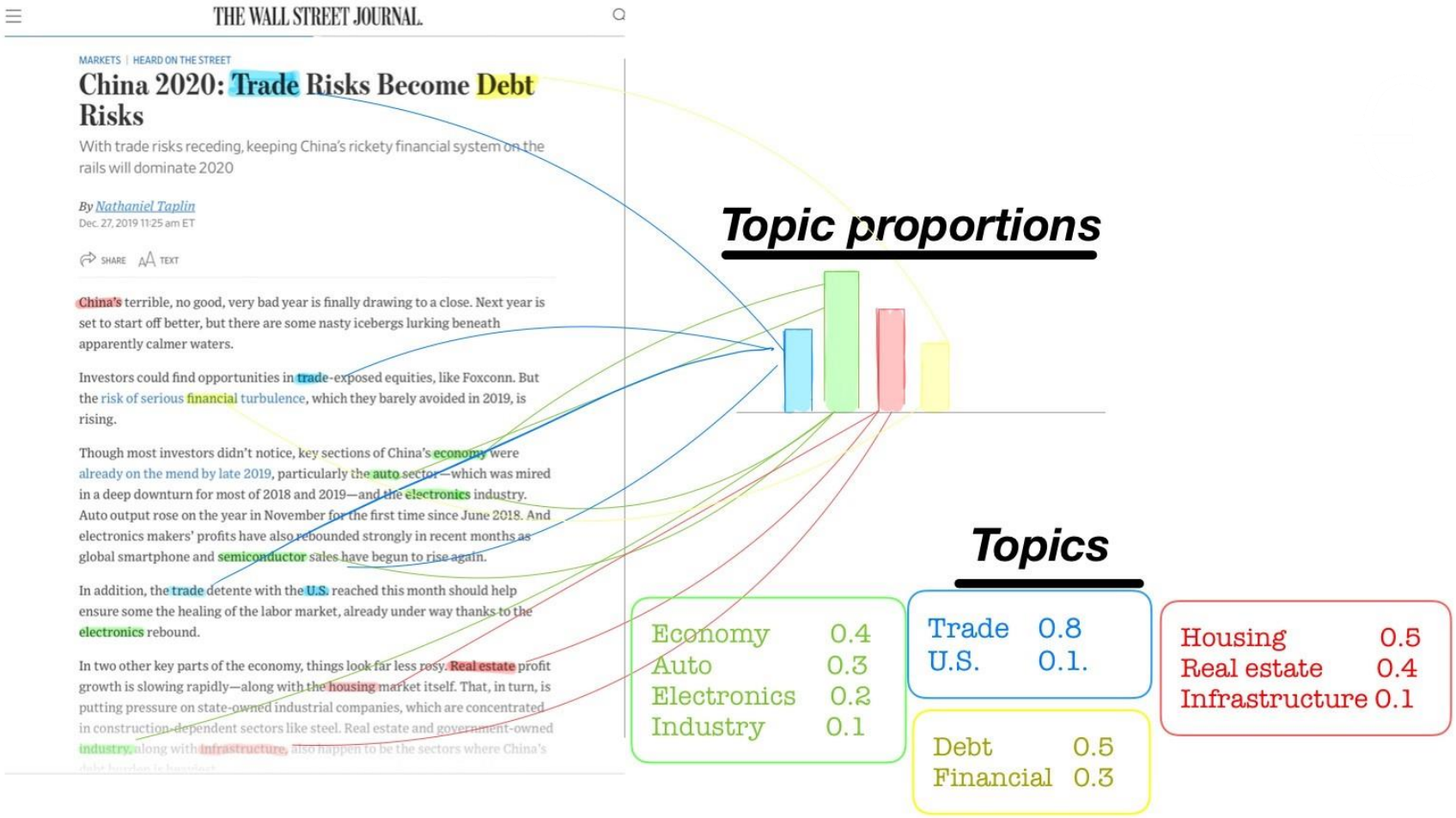
Outcome: We obtain around 2000 articles per year related to financial risk in China.

Step 3: Topic Modelling and the LDA algorithm

Topic modelling *is a type of statistical modelling for discovering the abstract* ***"topics"*** *that occur in a* ***collection of documents****. Latent Dirichlet Allocation (****LDA****) is an example of topic model and is used to classify text in a document to a particular topic.*

- It uses **unsupervised machine learning** to unveil the distinctive **topics** of overall financial risk.
- This method **does not require** selecting additional specific **keywords** (unsupervised).
- It is **not language specific**, and can be easily adapted to multi-language setups.
- We can construct a **wide range** of **sub-indicators** related to financial risk: *Banking*, *financial markets, trade and commodities, exchange rates….-> filter topics / decompose sources of risk*
- *We find an* ***optimal*** *number of topics equal to* ***30***

Step 3: Topic Modelling, graphical representation



Step 3: Topics need to be labelled. We do this looking at most frequent words and reading most representative articles

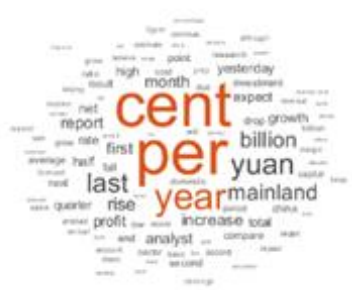
Banking

Financial Markets

Exchange Rates

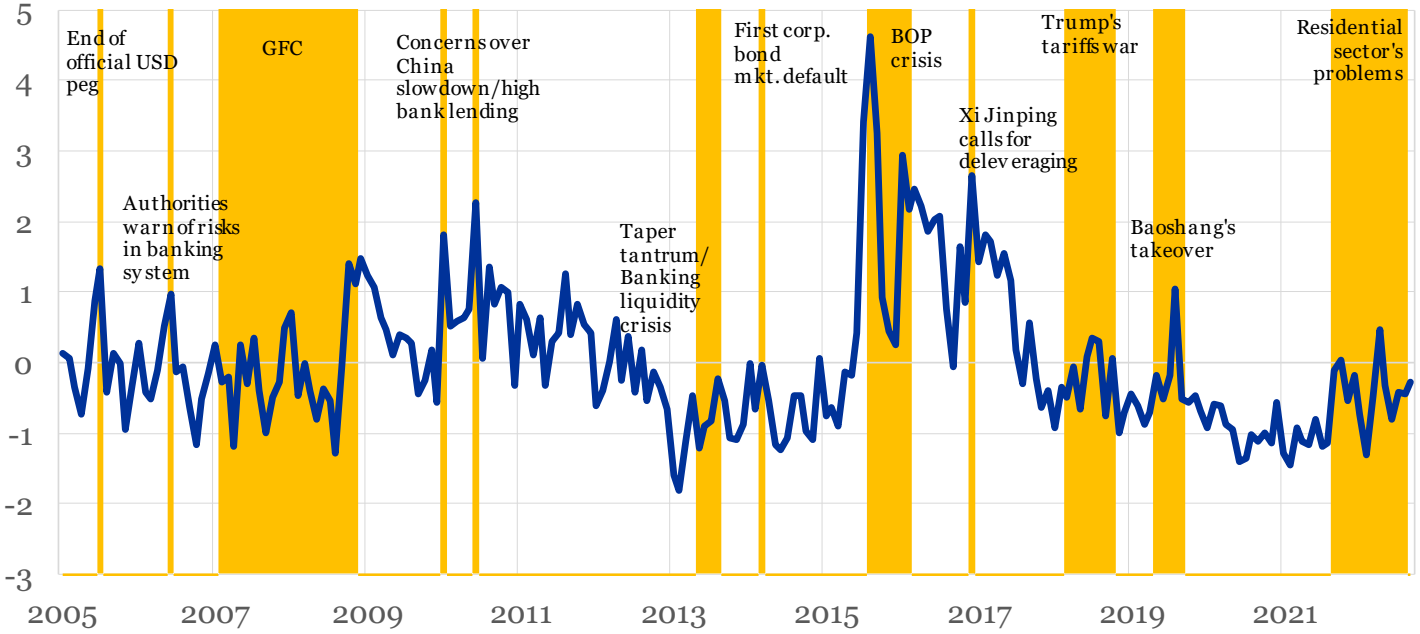
Real Estate

Corporate Profitability



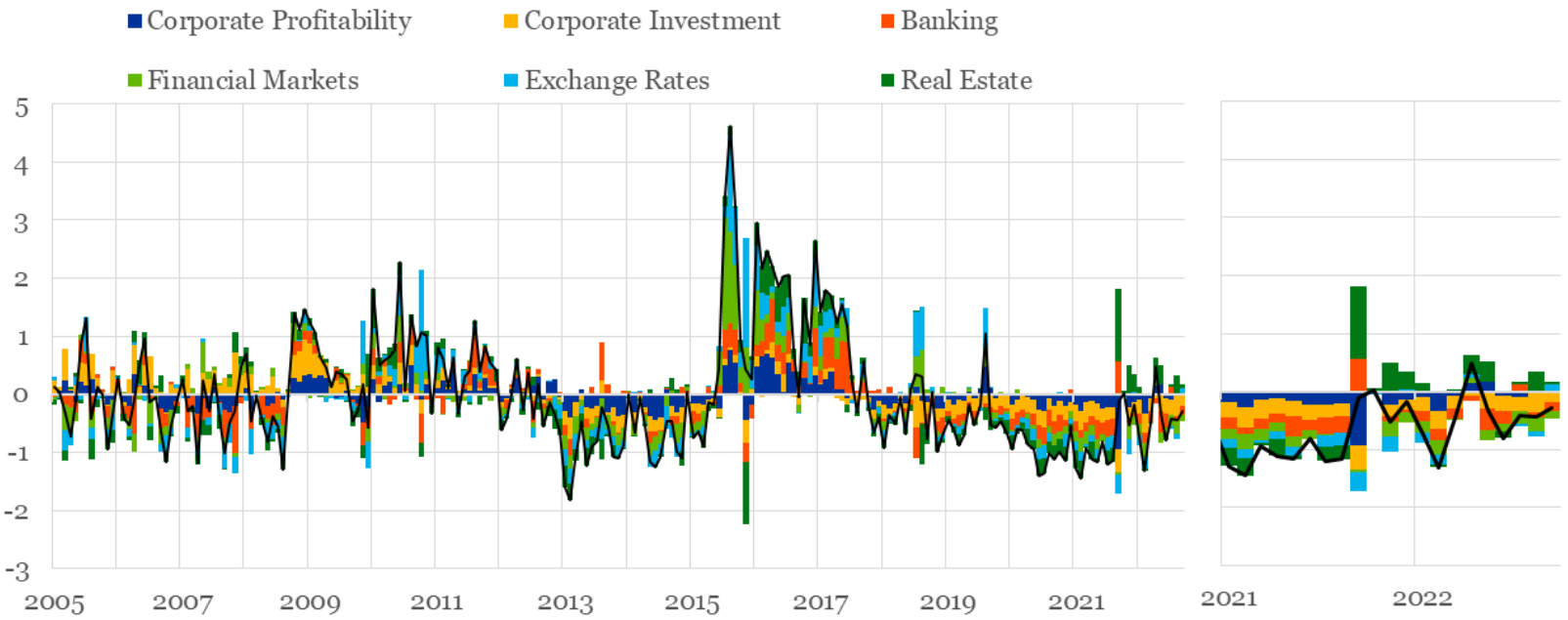
Corporate Investment

Step 4: Financial risk in China – aggregated index



Notes: ECB staff calculations. Time coverage 2005/01-2022/09. Y-axis in standard deviation with mean zero over the whole horizon.

Step 4: Financial risk in China - decomposition of risk



Notes: ECB staff calculations. Time coverage 2005/01-2020/11. Y-axis in standard deviation with mean zero over the whole horizon.

3.

Empirical analysis

How heightened financial risk affects China and the rest of the world

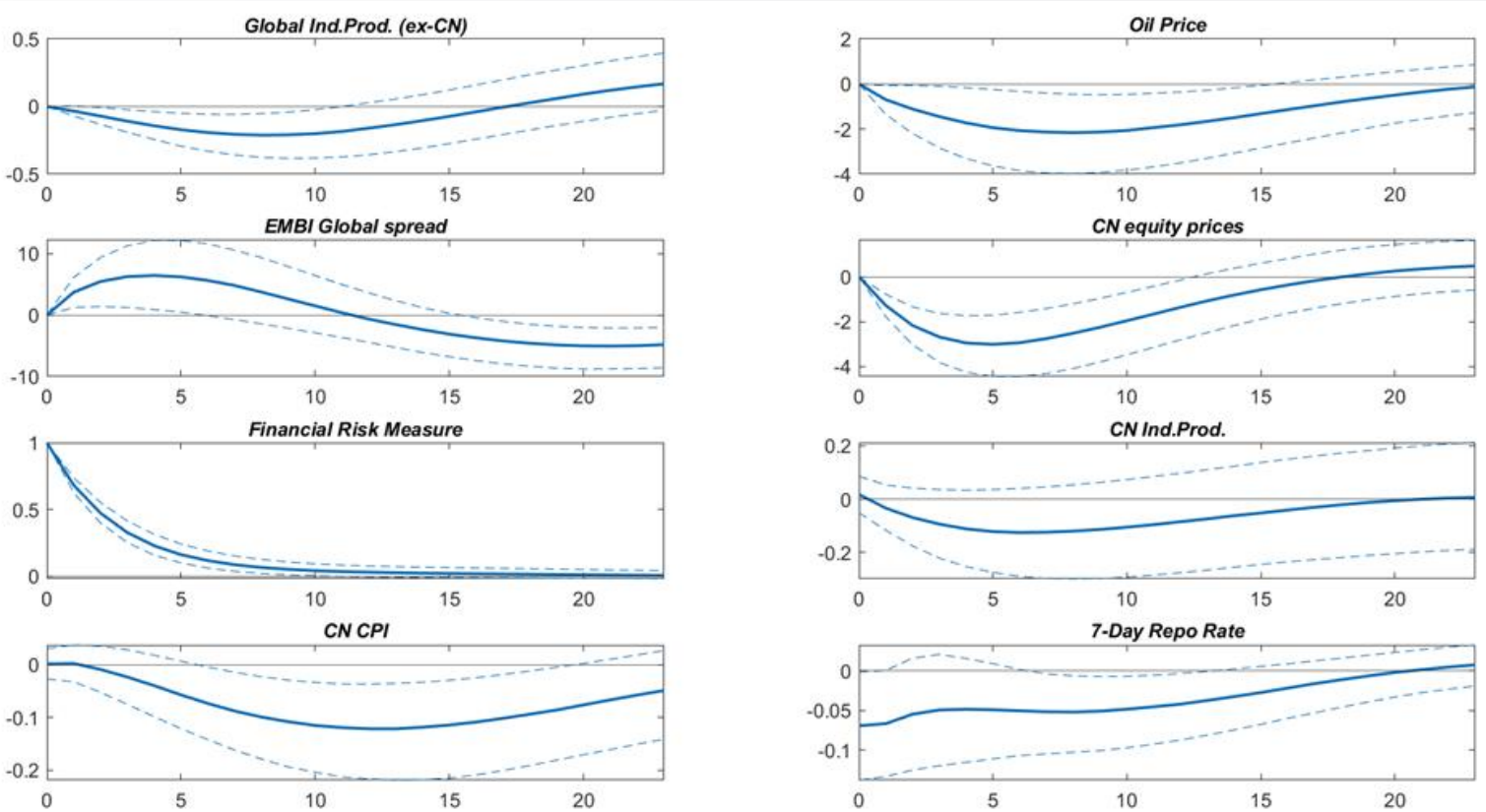
Empirical analysis: The Model

Table 2

VAR setup	Source
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 Chines economy (for additional details see also Higgins et al., 2016).

Empirical analysis: Impact of financial risk shocks to macro-financial variables



Conclusions

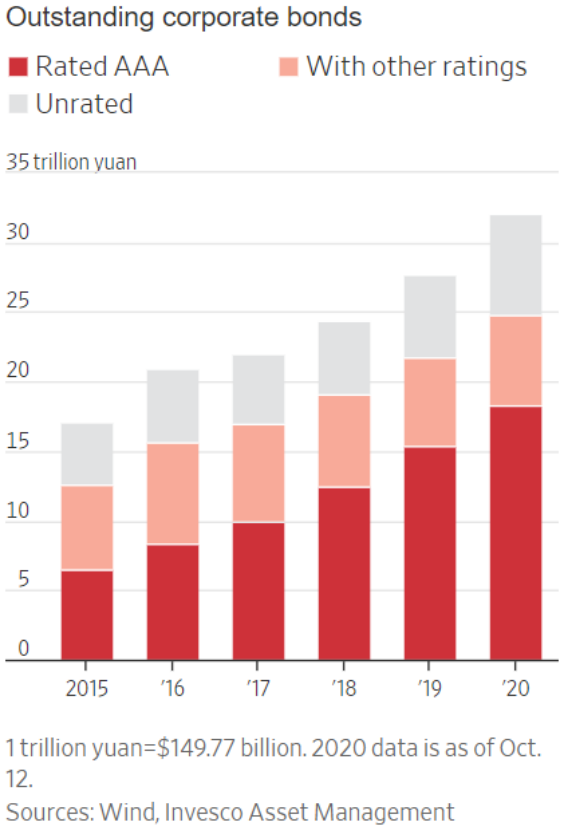
- We construct a text-based index of financial risk in China.
- The index seems to correctly accounting for past instances of increasing risk in China.
- The index has the advantage of being timely and can be decomposed in further sub-indices.
- Although being low compared to previous events the index shows some recent spikes due to problems in real estate.
- An empirical exercise shows that an increase in risk tightens financial conditions in China and has negative repercussion on (global) macro financial variables.

Thank you

Background slides

Financial risks in China: Difficult to extract signals from data

"In China, not all triple-A bonds are created equal" (WSJ)



Alternative estimates of non-performing loans

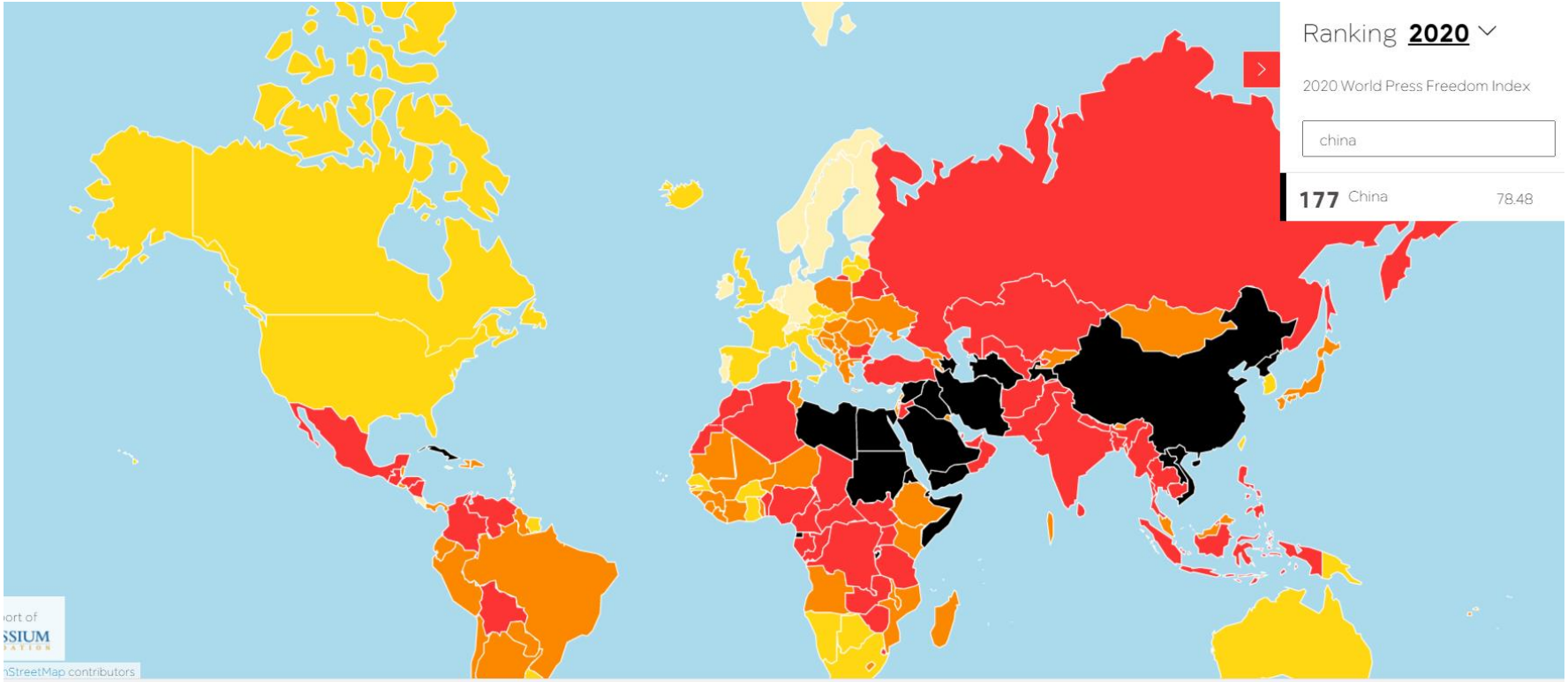
WSJ (Oct. 2020): "In China, not all triple-A rated bonds are created equal"
More than half of corporate debt in China rated AAA
In the US, bonds of only two companies are rated AAA by S&P
PBoC identified systemic risks from hidden links between non-financial holding companies and banks, how to measure this?

News database used for the analysis: Dow Jones Factiva

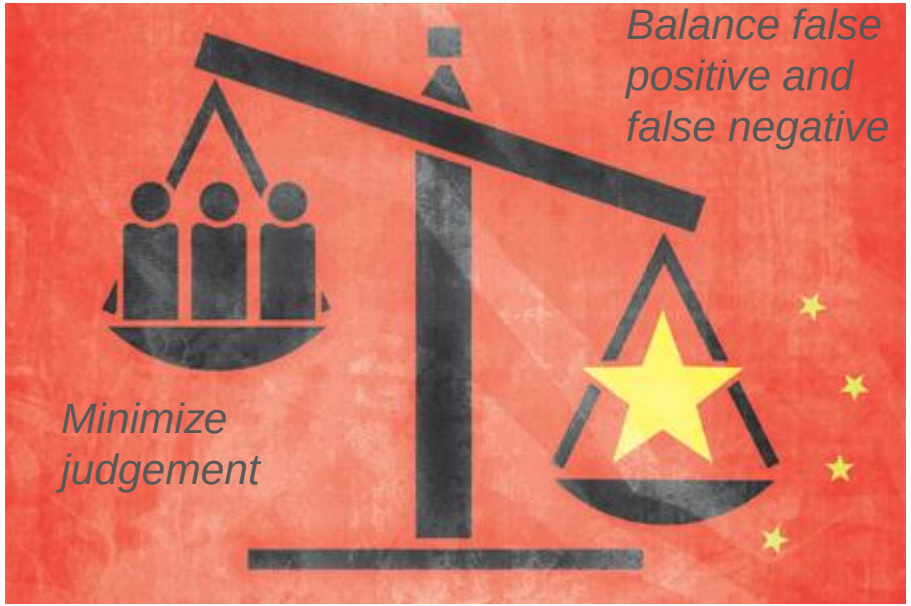
- More than **35,000 sources** from more than 200 countries in 26 languages.
- Nearly **400 continuously updating newswires.**
- **Newspapers**: Same-day and archival coverage of several newspapers (e.g. The Wall Street Journal, The New York Times, Washington Post..).
- **Factiva Search:** Gives full control to build precise searches using a combination of keywords and Dow Jones Intelligent Indexing.

Link to Darwin document Factiva Description

Freedom of the press: China ranked 177 out of 180 nations



Step 2: Filtering articles discussing increasing financial risk in China



Count the **occurrence** of **news articles** describing **financial risk** in **China** over time using **keywords:**

I. Initial terms: start search with all articles containing {China or Chinese AND risk$ AND finance$}.

II. Expand initial search using *word2vec* algorithm.

III. Select most relevant words associated with negative sentiment (i.e. danger/opaque/defaults).

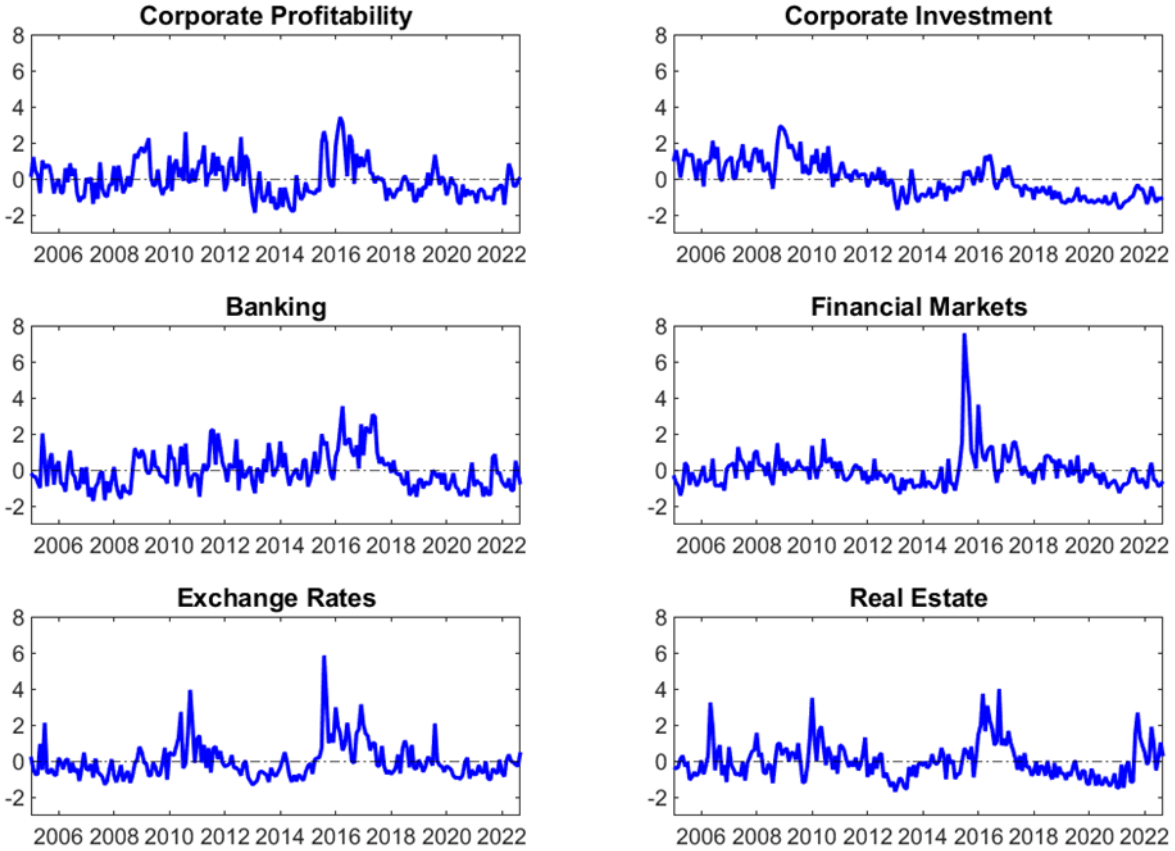
IV. Repeat the search.

Step 1: Final search

- {(**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$)}

Outcome: We obtain around 2000 articles per year related to financial risk in China.

Step 4: Topics probabilities are then aggregated to form time-series

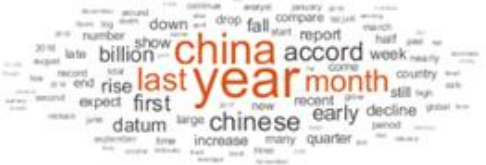


Topic Modelling: other topics

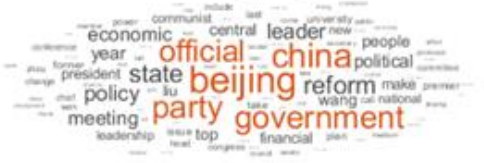
unk1

Global Issues

unk2



Domestic Policy-Party Issues



Growth Outlook

Law and Regulation

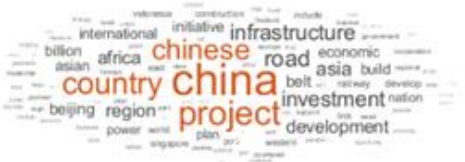
Investment

Local Policy

Foreign Affairs

Trade Wars

Infrastructure



Labour and Education

Topic Modelling: other topics

Leisure and Hospitality

Pandemic and Health

Environment

Fintech

Local News

Geopolitics Asia

Industry and Manufacturing

Trade and Commodities

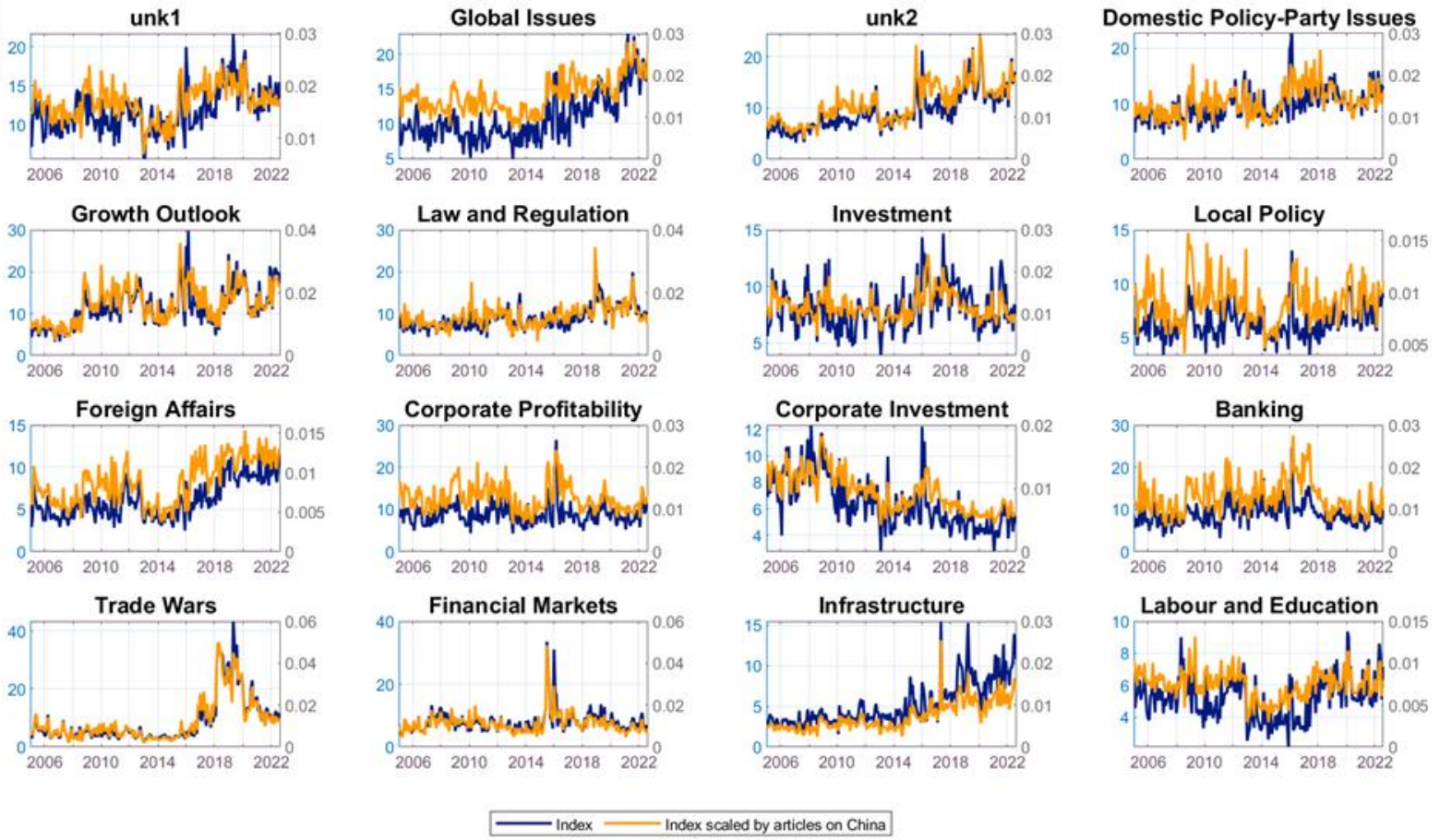
Energy

Entertainment

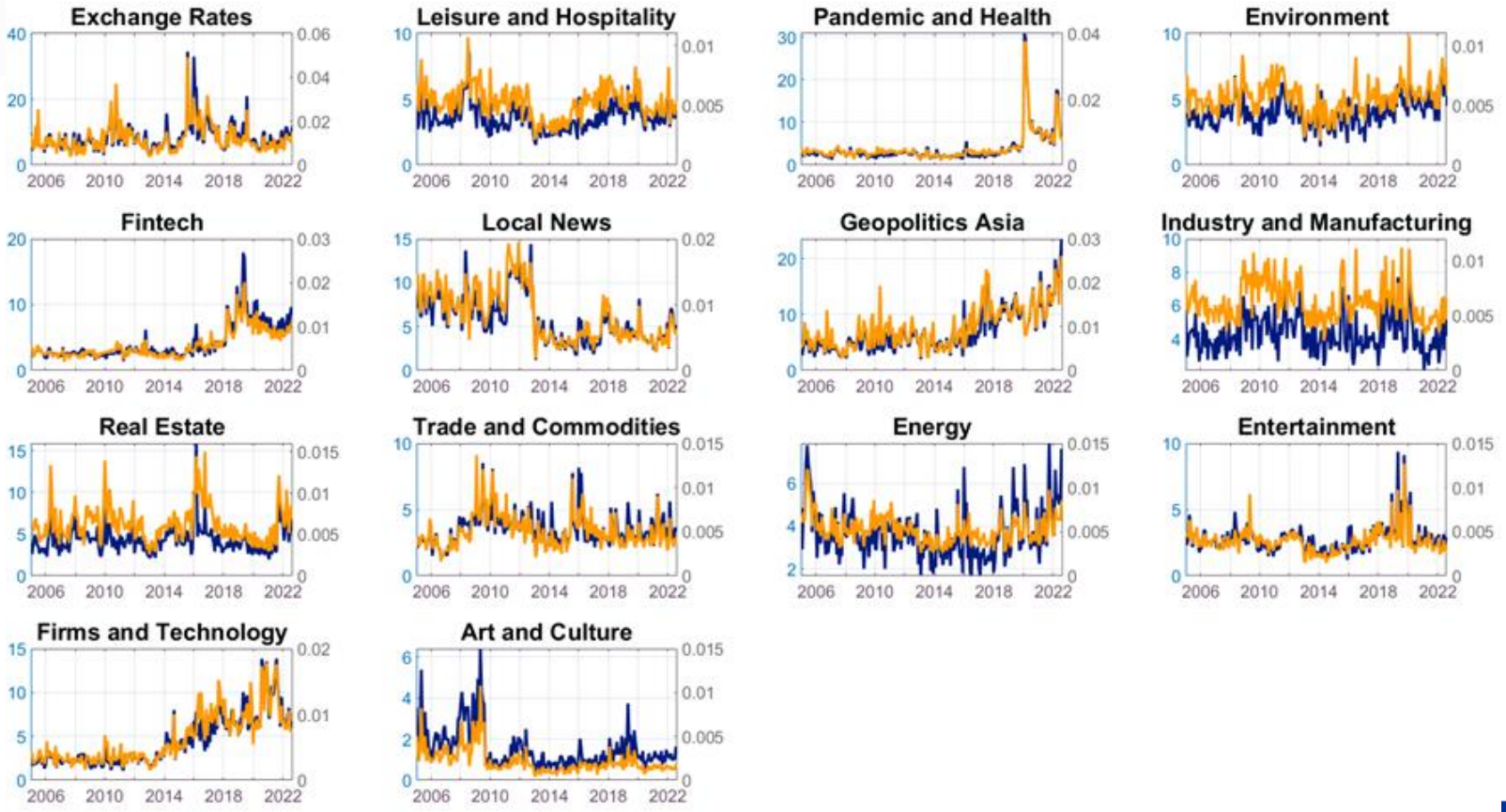
Firms and Technology

Art and Culture

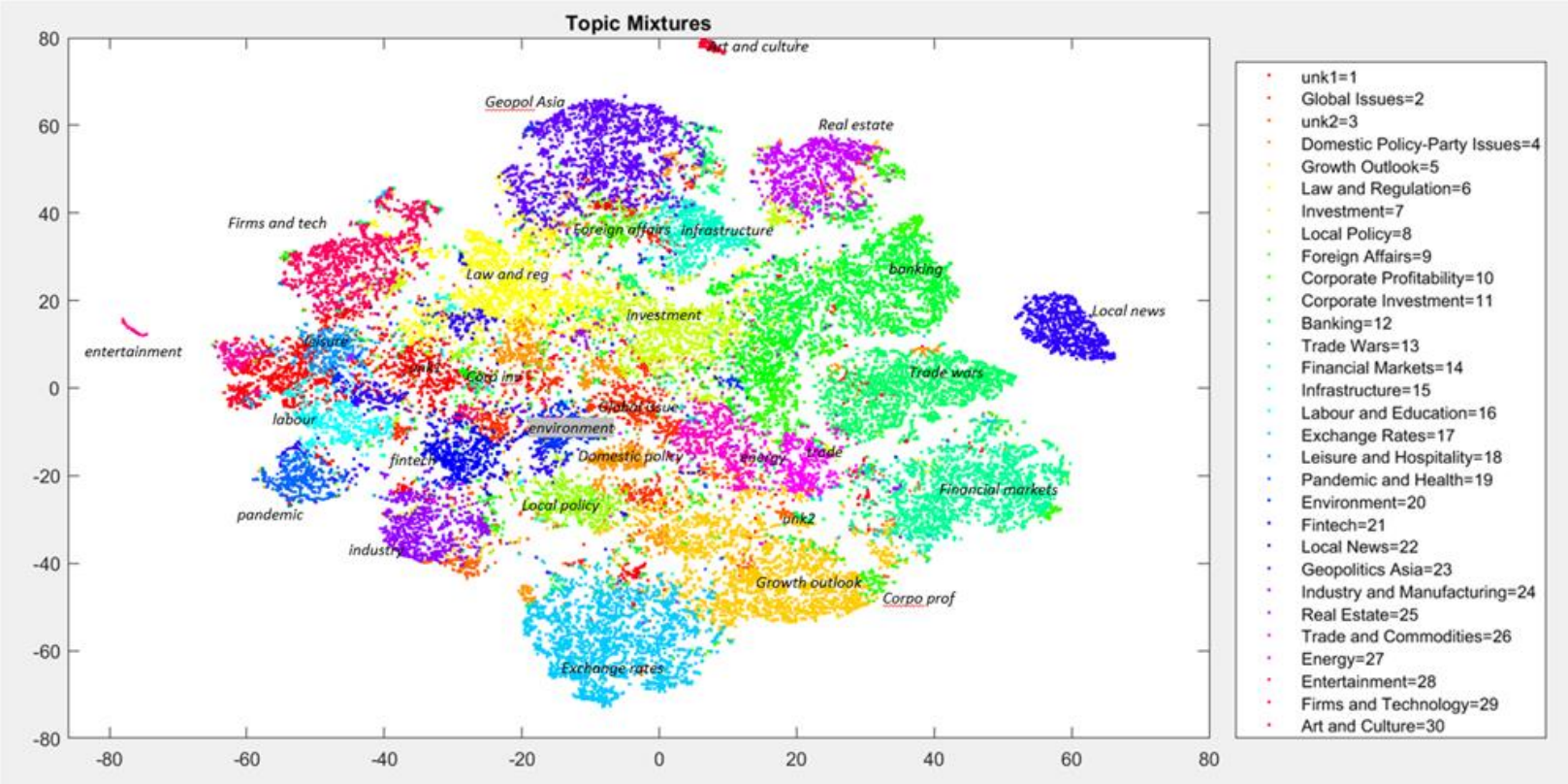
Topic Modelling: identified topics



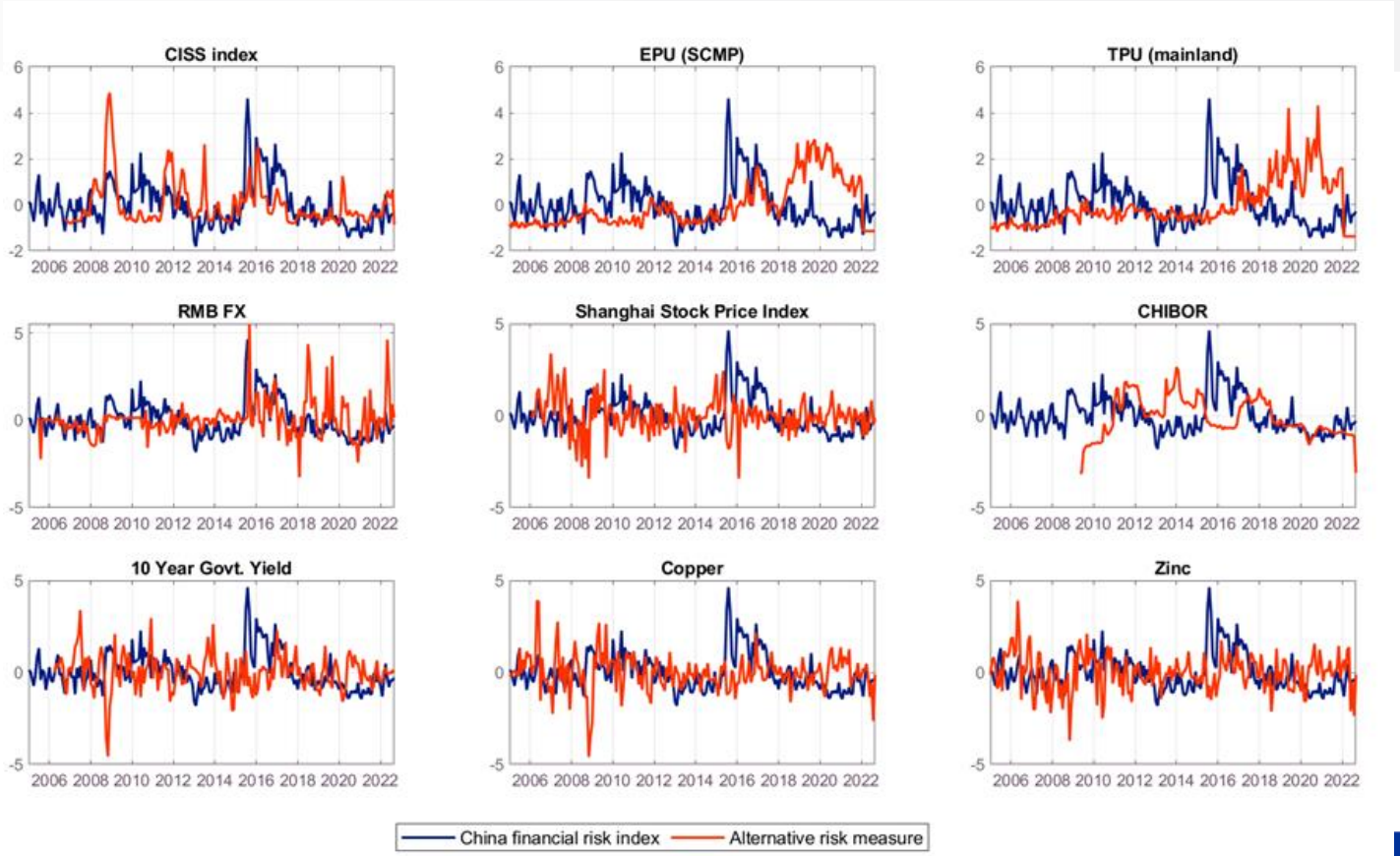
Topic Modelling: identified topics



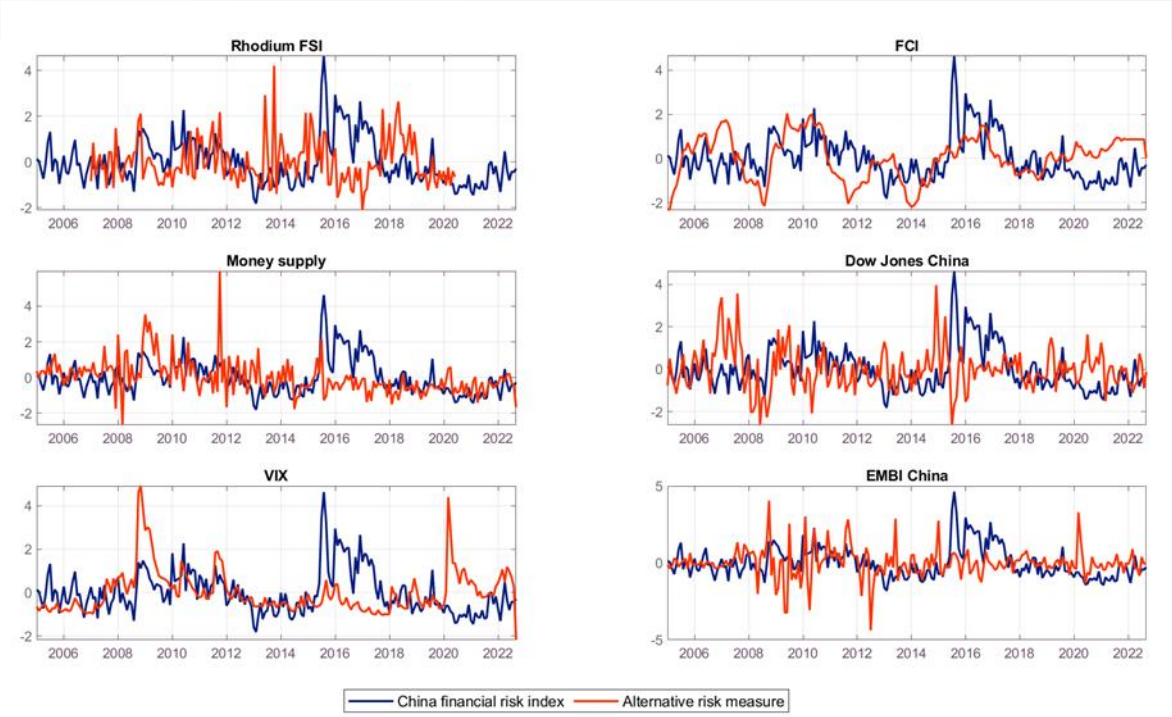
Topic Modelling: Proximity analysis



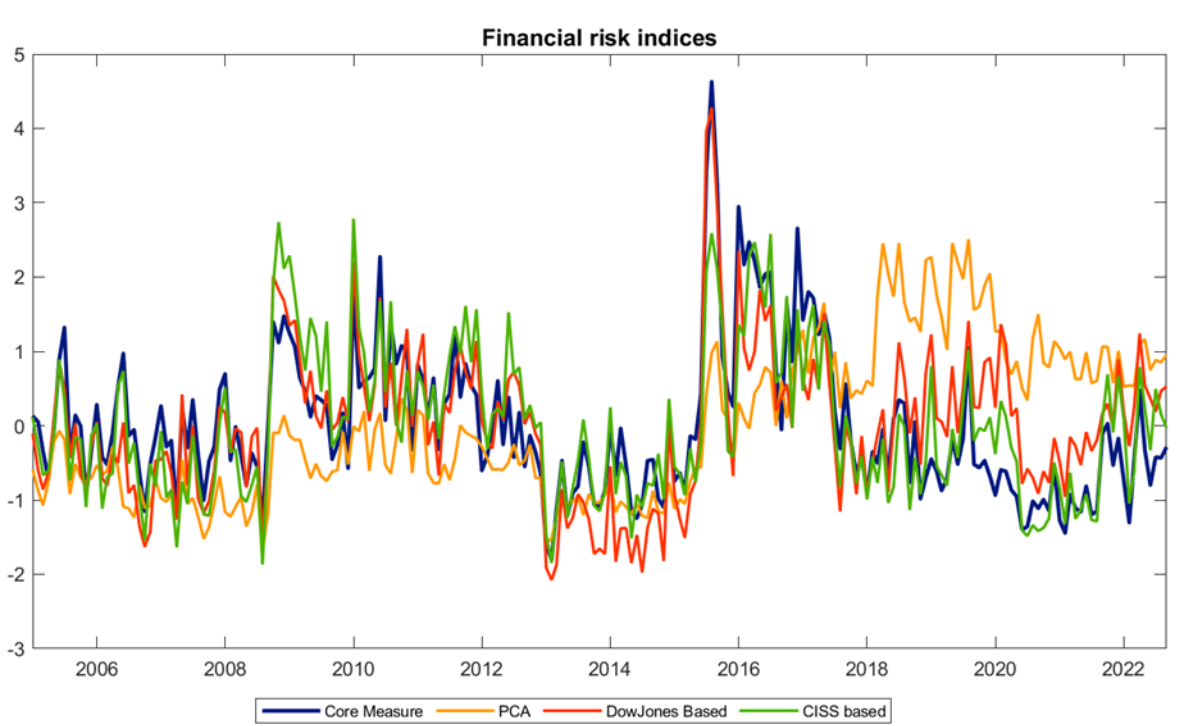
Topic Modelling: Comparison with other indicators



Topic Modelling: Comparison with other indicators



Topic Modelling: unveiling China's sources of financial risk



Topic Modelling: unveiling China's sources of financial risk

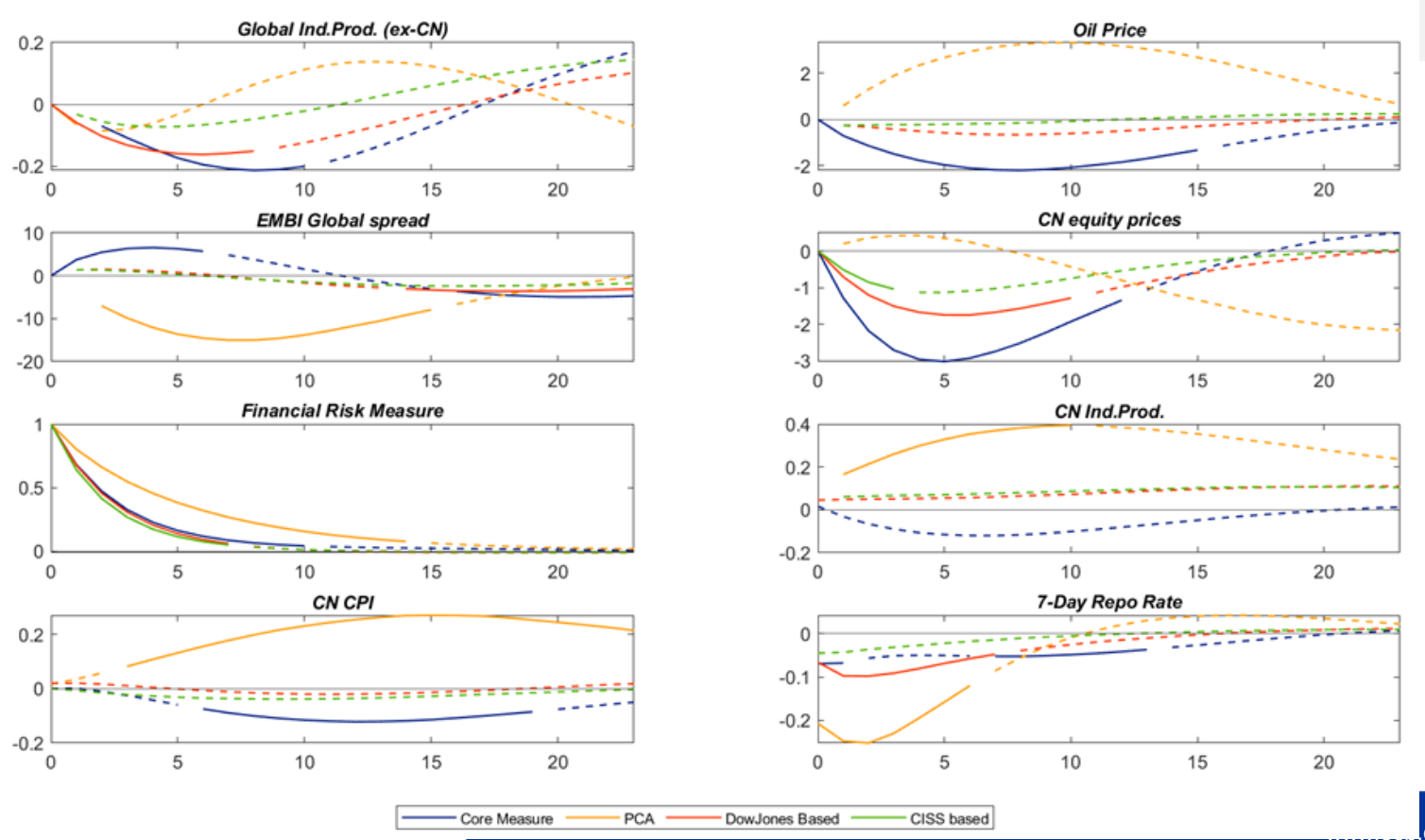


Table B.1.: Regression analysis results with China Dow Jones index as dependent variable.

VARIABLES	(1) DowJones	(2) DowJones	(3) DowJones	(4) DowJones	(5) DowJones	(6) DowJones
FinancialMarkets	-40.92**	-32.05**	-34.45**	-34.83**	-36.80**	-33.35**
	(17.94)	(15.57)	(15.03)	(15.49)	(15.03)	(15.14)
Banking	25.54					
	(19.74)					
FX	6.898					
	(14.18)					
RealEstate	-41.02					
	(37.23)					
CorpProf	-62.07**	-27.53	-38.46	-39.03	-46.84*	
	(29.60)	(31.31)	(25.51)	(26.11)	(25.90)	
CorpInv	57.27**	72.40**	82.40***	82.19***	112.1***	86.47***
	(26.81)	(30.68)	(25.79)	(25.92)	(31.28)	(27.97)
GrowthOutlook		-9.293				-31.31**
		(15.39)				(13.99)
LeisANDHosp		-162.5***	-163.9***	-165.0***	-205.0***	-222.1***
		(51.32)	(51.18)	(52.29)	(56.65)	(56.93)
Energy		-92.48**	-95.62**	-96.18**	-113.3***	-112.7***
		(42.13)	(41.74)	(42.17)	(42.90)	(42.72)
IndANdManu				4.600		
				(42.69)		
ForeignAffairs					48.96*	71.78**
					(29.47)	(32.38)
Constant	0.544*	1.507***	1.452***	1.445***	1.183***	1.185***
	(0.290)	(0.344)	(0.331)	(0.337)	(0.367)	(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	-13.57	-12.07	-9.304	-10.82	
	(16.64)	(15.90)	(15.78)	(15.69)	(15.50)	
Banking	32.50*	29.32*	32.17*	29.46*	38.05**	35.47**
	(18.20)	(17.40)	(17.02)	(16.92)	(17.07)	(16.79)
FX	-2.065	-14.92	-13.88	-10.32	-3.117	
	(13.18)	(12.93)	(12.85)	(12.85)	(13.03)	
RealEstate	-44.55	-61.15*	-58.36*	-78.32**	-72.54**	-75.74**
	(35.79)	(34.38)	(34.18)	(35.18)	(34.81)	(34.10)
CorpProf	89.78***	24.31				
	(27.17)	(29.99)				
CorpInv	96.02***	148.3***	159.4***	185.0***	138.1***	132.3***
	(26.10)	(27.66)	(23.99)	(26.74)	(32.74)	(32.05)
GrowthOutlook		65.46***	71.58***	85.63***	80.98***	77.79***
		(15.04)	(13.00)	(14.52)	(14.46)	(13.73)
IndANdManu				-88.13**	-87.14**	-94.28**
				(42.03)	(41.49)	(40.51)
ArtandCulture					134.8**	141.3***
					(55.76)	(53.97)
Constant	-1.986***	-2.419***	-2.401***	-2.236***	-2.234***	-2.206***
	(0.263)	(0.270)	(0.269)	(0.278)	(0.274)	(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