BBVA Research

N°. 17/13

WORKING PAPER

Tracking Chinese Vulnerability in Real Time Using Big Data

Álvaro Ortiz / Tomasa Rodrigo / Le Xia / Joaquín Iglesias / Carlos Casanova

Tracking Chinese Vulnerability in Real Time Using Big Data

Alvaro Ortiz / Tomasa Rodrigo / Le Xia / Joaquin Iglesias / Carlos Casanova

Abstract[1]

In this paper we develop an indicator to track the vulnerability sentiment in China in real time using Big Data. The China Vulnerability Sentiment Index (CVSI) is combination of traditional hard data, financial time series and sentiment indicators parsed from digital news through sentiment analysis using Data Science techniques. The index is composed by the key four different vulnerability dimensions such as the highly leveraged state owned enterprises (SOEs), the fast expansion of shadow banking activities, the risks associated to a potential correction in the housing market and the external vulnerability of the Chinese economy. The results show that CVSI index constitutes a good indicator to assess the vulnerability sentiment of China and it can be used as a real time indicator of the vulnerability perception by agents and the evolution of the risk narrative of the country. The index shows that the evolution of the vulnerability of China has not been uniform, with the shadow banking and the SOEs risk dimensions improving faster than the rest of the components. As Natural Language Processing techniques allows us to distinguish between different languages media, we also test for the existence of a systematic bias in information stemming from local (Mandarin written media) and foreign media (English written media). We found that there is no systematic bias between Mandarin and English written news. However, while the local news should have more information on local vulnerability, it is the English news which has a higher correlation with Chinese market risk measures, supporting the information asymmetries hypothesis in the emerging markets.

Key words: China, Vulnerability, Big Data, Sentiment Analysis, Information asymmetries, Emerging Markets, NPL

1: We are grateful to the participantsof the presentation of this paper at HKMA. We also thank JMs Sally Chen (IMF office at Hong Kong) for helpful comments.

1. Introduction

Nobel Prize winner Robert Shiller (2017) has recently claimed that the use of new empirical research through sentiment analysis and Big Data should be extended to analyze the role of narratives in business cycle. In this paper, we build on this block of empirical research to analyze the role of narratives in the context of Chinese vulnerability sentiment. Thus, rather than focusing on ways to measure economic activity sentiment (i.e. traditional confidence index, animal spirits); we develop an index to track Chinese vulnerability sentiment in real time using Big Data. This could help us to identify the key narratives of the vulnerabilities in China, which have the potential to become self-fulfilling or self-defeated.

As Chinese economy vulnerability has several dimensions, the China Vulnerability Sentiment Index (CVSI) is composed by four different vulnerability components. These include indebtedness concerns among highly leveraged state owned enterprises (SOEs); the rapid expansion of shadow banking activities, the risks of a potential correction in the housing market bubble and the speculative pressures in the exchange rate market. The rationale behind the inclusion of these four dimensions of vulnerability is explained in Section 2.2.

On the empirical strategy, the CVSI combines daily time series on sentiment from a Big Data database known as Global Database on Events Location and Tone (GDELT)[2] with traditional official statistics (hard data) and financial indicators in daily, weekly, monthly and quarterly frequencies, making it unique in its robustness and depth. The index is also unique in the sense that it enables us to track vulnerability and sentiment along a number of risk dimensions in real time. To construct the overall index as well as the indices of each component, we use principal component analysis to reduce complex data sets to a lower dimension so as to reveal the most relevant underlying trends (Jolliffe, I.T., 2002).

The empirical results show that the information provided by the textual analysis can be considered an efficient indicator to assess the Chinese vulnerability sentiment in real time. Beyond this, it contains some explanatory power on market measures of risk as the risk premiums proxied by the Sovereign Credit Default Swap. Lastly, the differentiation of vulnerability dimensions can help investors and policymakers to formulate more nuanced views on the economy.

Finally, the Big Data analysis allows us to check for the existence of a systematic bias in the narrative of different languages. As Natural Language Processing and the GDELT database allow us to distinguish between alternative sources of news (i.e those coming from Mandarin Chinese language sources and English language sources), we can trace the difference in the vulnerability sentiment of local (proxied by Mandarin) and global (proxied by English) news. Our findings reveal that there is no systematic bias between English and Chinese media sentiment when we investigate vulnerability in China related news. However, the results show that while local media sources have some leading-indicator properties over the foreign ones (english written news), the correlation with market risk measures (i.e Chinese Sovereign CDS) is higher with the global sentiment (English written news). This could be the result of

2: https://www.gdeltproject.org/

information asymmetries affecting to the pricing mechanism of liquid financial instrument in Emerging Markets stemming from structural liquidity problems in local markets and/or the existence of margin calls in the financial centres. If this argument holds, the market indicators for risks could be affected by the language used in the major financial centers (Eyssell et al., 2012).

The paper is organized as follows: Section 2 explains the motivation of this study and defines the rationale for the components of the index. In Section 3, we describe the data, the sentiment algorithm as well as the used methology for contructing the indices. Section 4 presents the results as well as a robusteness check of the analysis using Bayesian Model Averaging techniques. We check for the existence of language bias in the news according to the media source in the index. Finally, section 5 concludes.

BBVA | Research

2. Assessing China's vulnerability

2.1. Motivation

China has become increasingly integrated in the global economy. A growing volume of literature has already examined implications of the moderation of Chinese growth on the global economy, including spillovers to the rest of the world through a number of channels such as global trade (Drummond et al., 2013; Casanova et al., 2016), financial markets (Zhou et al., 2012) and investments (Ahua and Nabar, 2012; Lee et al. 2016). Inmediately after the 2009 crisis, the Chinese authorities implemented a RMB 4 trillion (USD 586 billion) fiscal stimulus package that helped to buffer Chinese domestic demand and balance global growth. However, the rapid pace of economic growth after the stimulus has ben accompanied by growing Chinese macroeconomic imbalances.

2.2. The different Chinese vulnerability dimensions

While the importance of Chinese vulnerability is straightforward, the nature of Chinese vulnerability is more difficult to stablish. Despite it, most analysts and research have defined four main dimensions in the Chinese vulnerability:

1. **The Stated owned enterprises (SOE):** This is basically linked to overcapacity and highly leveraged indebted SOEs, which is behind the rapid increase of indebtness of the economy, now equivalent to approximately 250% of GDP. Most of this debt is held by corporates, not household or the government. The surge in corporate indebtedness in China can be traced back to an increase in leveraging by SOEs, coinciding with the implementation of China's huge post-financial crisis stimulus package (Huang 2014). Despite the "strategic" considerations, the truth is that leverage is substantially higher than private companies and that SOEs debt is concentrated in some specific sectors (Shipbuilding,Mining,…) rather than being broad-based vulnerability. In any case, excessively high debt levels could result in moderation of growth as the ability of these companies to continue to invest will be somehow constrained. Moreover, inefficiencies associated with less than optimal allocation of resources mean that every new unit of new credit has a harder time finding a productive project to invest in. Over time, the consecuences of the econom's 2016) this could put strain on bank balance sheets, requiring huge liquidity injection. The government is trying to steer the economy away from its dependency on debt by gradually removing its implicit guarantees on SOEs. This means that for smaller or regional SOEs could find themselves in a vulnerable position in case the cost of servicing their debt increases, as we have observed by the unusually large number of corporate defaults in the first quarter of 2017.

2. **The Housing bubble vulnerability:** This includes risks of a potential correction in the housing market bubble. As China doesn't have mature stock markets nor developed pension systems, many households have opted for housing and land has an investment vehicle. Thus, aside from fundamental factors (increasing incomes, household formation, urbanization and agglomeration) the house price trend has been exacerbated by a number of financial factors, including the collapse of the stock market in 2015, and excessive liquidity following from many years of loose monetary policy in China (Xu and Chen, 2012). Moreover, much of the economy itself is based on construction and real-estate development which directly and indirectly could account for near 30% of GDP. Not

surprisingly, mortages have been one of the main drivers behind the increase in total credit. This has driven housing prices in China up, increasing the risk of a hard landing of the residential sector. A downturn could wreck havoc to the economy, leading to liquity issues in smaller banks which are more exposed to loans and leveraged buying (Koss, 2015).

3. **Shadow banking:** The fast expansion of shadow banking activities constitutes an important risk for the economy. An expansion of shadow banking could pose threats to the stability of China's financial institutions, leading to an overall deterioration of credit conditions (Li et al., 2014). Broader credit has soared, as measured by Total Social Financing (TSF), and is now equivalent to roughly 220% of GDP, up from just 206% in 2015. Higher levels of economy-wide leveraging have been fueled by bank lending to corporates, mortgages and corporate bond issuance. By contrast, the growth in "core" shadow banking activities included in TSF remains subdued. However, this moderation does not take into account "non-core" components such as the issuance of Wealth Management Products (WMPs), which have soared in recent months. China's shadow banking activities now account for more than 80% of GDP or RMB58 trillion according to estimates by Moody's (2016), having grown by 19% in the first half of 2016. If this situation continues indefinitely, it could lead to a worsening of credit conditions, leading to dire consequences in case of any shocks to onshore liquidity conditions.

4. **Exchange rate (FX) speculative pressure:** The Chinese currency has been subject to depreciatory pressures since late 2014. This has led the authorities to expend a vast amount of sovereign resources to prop up the value of the currency. FX reserves dropped by USD1 trillion as a consequence, causing great pain to the authorities. As in the rest of the big Emerging Markets this component can be driven by idiosyncratic or global fianancial forces.

These four risks are highly specific to the situation in China; which makes the CVSI unique to assess more accuratelyy and timely the evolution of the Chinese risks. Besides, it also allows us to assess vulnerability and sentiment in a way which offers more granularity. China is a complex market and it will not be enough to talk about vulnerability in general terms. Some parts may be booming while risks mount in other areas, this is something which has significant implications for policymakers and investors alike.

3. A vulnerability sentiment index for the Chinese economy

3.1. The Data: Hard and Sentiment indicators

The database that we have used to develop the index is based in both hard traditional risk indicators and real time sentiment indicators. In order to build these high frequency vulnerability components, we mix hard data (e.g. macroeconomic and financial vulnerability indicators), with high frequency soft data from the "*Global Database of Events, Language and Tone* (GDELT)", a database which collects information from the media created by Leetaru and Scrodt (2013)[3]. Thus, we complement data from official statistics and market information with data sentiment in real time coming from news, which allows us to nowcast the official data as well as to draw information from a new dimension, the emotional one. Thus, we incorporate sentiment variables on daily basis to the traditional analysis using hard and market data.

Figure 3.1.1 Hard data and sources

Indicator	Frequency	Source	Justification
Total Profits (SOEs)	Monthly	Ministry of Finance of the People's Republic of China	Vulnerability sentiment increases in case SOE profits decrease
Total Liabilities (SOEs)	Monthly	Ministry of Finance of the People's Republic of China	Vulnerability sentimented increases un case SOE liabilities increase
Mortgages as a percentage of total loans	Quarterly	The People's Bank of China	Vulnerability sentiment increases in case mortgages as a percentage of total loans increase
GICS housing index	Daily	Wind Information Service	Vulnerability increases if the price increases
Housing price index	Monthly	National Bureau of Statistics	Vulnerability increases if the price increases
New construction	Monthly	National Bureau of Statistics of the People's Republic of China	Vulnerability increases in new construction decreases
Real estate investment	Monthly	National Bureau of Statistics of the People's Republic of China	Vulnerability increases if real estate investments decrease
NPL ratio	Quarterly	China Banking Regulatory Commission	Vulnerability increases id NPL rations increase
Total social financing	Monthly	The People's Bank of China	Vulnerability increases if TSF increases
Entrusted loans	Monthly	The People's Bank of China	Vulnerability increases if the use of entrusted loand increases
Wenzhou Index	Daily	City of Wenzhou, Finance Bureau	Vulnerability increases as the Wenzhou Index increases
WMP yields	Weekly	Wind Information Service	Vulnerability increases with higher yields
Bank acceptance bill yields (6M Pearl River Delta)	Daily	Wind Information Service	Vulnerability increases withi higher yields
Foreign Exchange reserves	Monthly	The People's Bank of China	Vulnerability increases when FX reserves fall
USDCNY	Daily	Bloomberg	Vulnerability increases if the exchange rate depreciates
USDCNH	Daily	Bloomberg	Vulnerability increases if the exchange rate depreciates
O/N Hibor	Daily	Bloomberg	Vulnerability increases if O/N hibor rates spikes, pointing to tighter liquidity on intervention

Source: BBVA Research

3: Leetaru, Kalev and Schrodt, Phillip3 2013."Global Database of Events, Language, and Tone (GDELT)

The hard data variables were selected to represent the four dimensions of risk outlined previously. We rely on high frequency series (daily) wherever possible, in order to combine these with the soft data sentiment data from GDELT. There are several reasons for mixing different sources of information. First, there are no many hard indicators to capture Chinese vulnerability. In order to deal with the lack of risk indicators, we use information from the media reflecting sentiment on Chinese vulnerability related terms. However, relying exclusively on sentiment indicators could provide and excess of noise rather than signal. Thus, including factor analysis between hard and sentiment data will reassure us that the information steming from sentiment is associated to any of these vulnerabilities. Finally, we mixed frequencies with all the sources of information, including quarterly, monthly and weekly data converted into daily series and normalized for inclusion in the index. A list of the official statistics and their sources can be found in Figure 3.1.1.

To compute soft sentiment data, we rely on GDELT[4], an open access database on international news which pins down and processes news in broadcast, print and web media globally in over 100 languages on daily basis. In GDELT, the information is extracted from the media and systematized using PETRACH algorithm. As news articles are being parsed to build the GDELT database, different algorithms are run in order to assess the themes that a given piece of news was dealing with. Key words from different taxonomies and dictionaries are identified in the articles and thus a theme classification of the article is provided in the database. GDELT also identifies thousands of emotions, organizations, locations, counts, news sources and events across the world as well as the average tone of the analyzed news articles, this is, how positive or negative the wording of the article is (more details in Section 3.2).

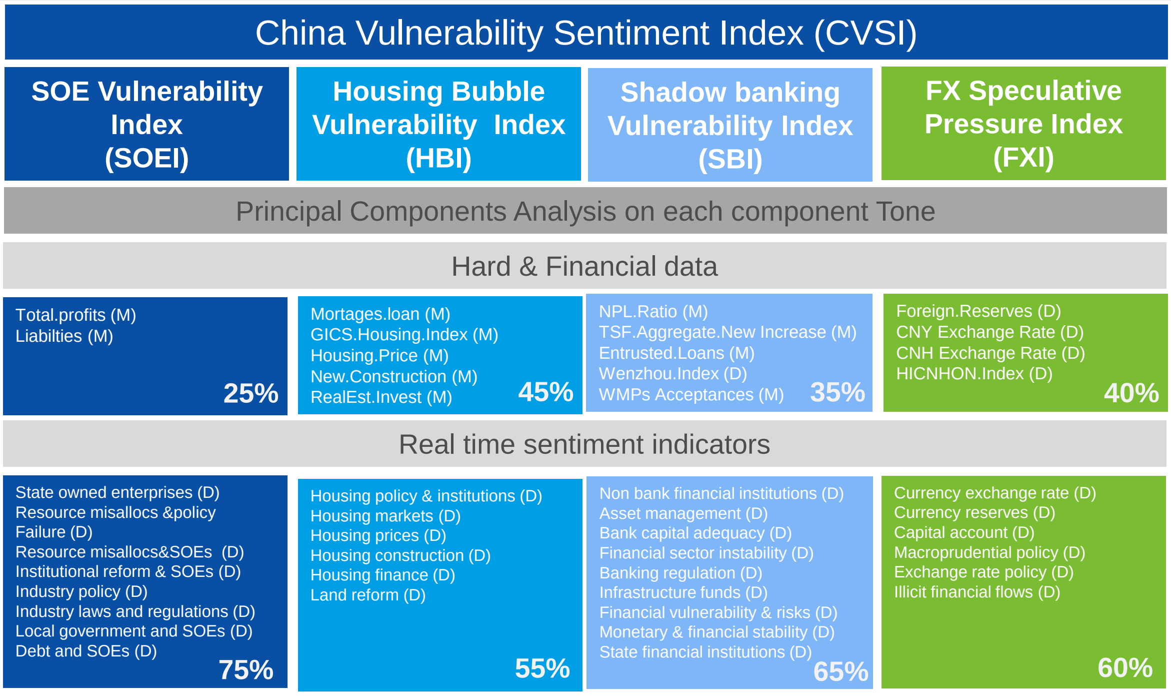
Sentiment variables from GDELT are defined as the daily average tone or sentiment of the news containing a particular theme. To delimit the scope of the news into the space desired for our analysis, some interaction series were also built, this is, the daily average tone of the news dealing with multiple of these themes. Specifically, the tone series regarding *Debt*, *Resource Misallocations,* and *Local Government* have been interacted with the theme on *State Owned Enterprises* to make sure that the pieces of news included in the analysis are indeed related to the State Owned Enterprises component of the index.

The selected series for each component of the index are shown in Figure 3.1.2. As can be observed the data are balanced, in the sense that in most of the components the amount of soft sentiment data is higher than the 50%. The only exception is the SOE component, for which the lack of real indicators led us to overweigth sentiment indicators.

To extract, manipulate and analyze this huge information set, we use Google Cloud with BigQuery, which allows us to efficiently interact with the database using Structure Query Language (SQL) processing terabytes of data in just seconds. In addition, BigQuery's scalability and flexibility lets to have a fluid schema where information is constantly growing, combining historial data with real time data.

4: Further information can be found in the following link.

Figure 3.1.2. Included variables in the Index.



Source: BBVA Research

3.2. Measuring "Sentiment"

Both theoretical and empirical economic works have stressed that sentiment or narratives can influence agents' decisions. The theoretical importance of sentiment in economics is far from new. Pigou (1927) believed that business cycle fluctuations are driven by expectations and entrepreneurs' errors of optimism and pessimism are crucial determinants of these fluctuations. Later, the seminal work by Keynes (1936) highlighted the importance of changes in expectations that are not necessarily driven by rational probabilistic calculations, but by what he labeled "animal spirits". More recently, Keynes's original hypothesis has been one of the salient features of the 1990-91 recession (Blanchard, 1993) while Angeletos and La'O (2013) develop a unique-equilibrium, rational-expectations, macroeconomic model which features "animal spirits", labeled sentiments. Shiller (2017) shows "narratives" can explain aggregate fluctuations though epidemic models.

There is also empirical literature which has stressed the importance of sentiment in economics. Angeletos, Collard and Dellas (2015) have quantified the importance of the variations in sentiment (or confidence) in macroeconomic DSGE models. They find that sentiment shocks lead to strong co-movement between employment, output, consumption and investment and that they account for around one half of GDP variance and one third of the nominal interest rate variance at business-cycle frequencies. Barsky and Sims (2012) found that this informational component forms the main link between sentiment and future activity in international business cycles. More recently, Recently Shapiro, Sudhoff, and Wilson (2017) show how the news sentiment measures outperform the University of Michigan and Conference board measures in predicting the federal funds rate, consumption, employment, inflation, industrial production, and the S&P500.

In this paper, we measure sentiment vulnerability in the media through textual search using natural language processing or text mining and sentiment analysis techniques. Sentiment analysis, also known as opinion mining, extracts meaning from the text processing words with emotional connotation in order to get insights about the perception of a given topic in the text. Due to development of digital news and its tremendous value for practical applications, there has been an explosive growth of both research in academia and applications in the industry (Liu, 2010). Shiller (2017) observes that textual search is a small but expanding area in economic research. Textual analysis has been used by economists, for example, to document changes in party affiliation (Kuziemko and Washington, 2015); political polarization (Gentzkow et al. 2016); and news and speculative price movements (Roll 1988; Boudoukh et al. 2013). Tetlock et al (2008) uses textual analysis to extract sentiment from corporate reports while Tetlock (2007) analyzes the sentiment in media. In the spirit of our work, Wang et al (2013) shows the existence of strong correlations between financial sentiment words in financial reports and the risk of companies. More recently, Shapiro, Sudhof and Wilson (2017) construct indexes measuring the economic sentiment embodied in newspaper articles.

There are two general approaches to quantify sentiment in text. The first is known as the Lexicon approach. It relies on a pre-defined dictionary of words that are associated with an emotion (commonly positive or negative). Thus, the lexicon-based techniques work with individual words missing context and words co-occurrence. The second approach uses Statistical Natural Language Processing (NLP) tools, a subfield of computational linguistics, which relies on machine learning techniques. Sentiment analysis attempts to extract emotional content from a set of text based both on word choice (lexicon) and the context (combinations and structure) of words.

To measure the tone or sentiment in the media we use the GDELT dataset and tools. As explained in Section 3.1, GDELT relies on hundreds of thousands of broadcast, print and online news sources from every corner of the globe in more than 100 languages. GDELT uses “directional” dictionary measuring words associated with positive and negative connotation based on more than 40 sentiment dictionaries included in Wordnet[5] (and translating each article into English from more than 65 languages).

GDELT uses a refined lexicon (with a special emphasis on classifying words differently based on the tenses of the phrase), to build a score ranging from -100 (extremely negative) to +100 (extremely positive) for each pice of news, with common values ranging between -10 (negative) and +10 (positive), with 0 indicating neutral tone.

A neutral sentiment can be the result of a neutral language or a balancing of some extreme positive sentiments compensated by negative ones. Once negative and positive words are identified, we construct a tone variable based on the balance between the percentage of all words in the article having a positive and negative emotional connotation within an article divided by the total number of words included the article:

$$\text{Average tone} = 100 * \frac{\sum Positive\ words\ - \sum Negative\ words}{\sum Total\ words}$$

5: See http://wordnet.princeton.edu/ : Some of the most used sentiment dictionaries included in economic analysis as the Harvard IV and Loughran and MacDonald are also included in GDELT.

3.2. Construction of the index: methodology

Vulnerability is a non-observable variable, so it is difficult to measure it quantitatively. Still, there are some proxies for general vulnerability or risk as market proxies (Risk premiums, CDs…) and rating agencies scores. In this work, we assume that vulnerability in China is a function of the underlying trend of a set of vulnerability dimensions described in Section 3.1. Furthermore, we can approximate a measure of vulnerability sentiment by looking at the underlying pattern governing these correlated observable variables and their variances. To find this commont trends in vulnerability we use Principal Component Analysis (PCA) which enables us to reduce complex data sets to a lower dimension, thus revealing underlying trends by examining the variances associated to the principal component (Jolliffe, I.T., 2002).

To identify the alterative sources of vulnerability we proceed with PCA in two steps. First, we apply PCA to the individual set of vulnerability groups as State Owned Enterprises (SOE), Housing Bubble (HB), Shwdow Banking (SB) and external vulnerability (FX). .Simultaneously, the four subcomonents summarize the main underlying trend related to state-owned enterprises, the housing sector, the shadow banking and FX pressures respectively in a single index. They are defined as follows:

$$(1) \qquad \mathrm{SOE} = \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_{10} x_{10} + \epsilon_1$$

$$(2) \qquad \mathrm{HB} = \delta_1 y_1 + \delta_2 y_2 + \ldots + \delta_{11} y_{11} + \epsilon_2$$

$$(3) \qquad SB = \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{15} z_{15} + \epsilon_3$$

$$(4) \qquad FX = \rho_1 v_1 + \rho_2 v_2 + \ldots + \rho_{10} v_{10} + \epsilon_4$$

Where $(x_1, \ldots, x_{10})$ represent the variables included in the SOE component (see Figure 3.1.2); $(y_1, \ldots, y_{11})$ represent those for the housing bubble component (see Figure 3.1.2); $(z_1, \ldots, z_{15})$ shadow banking component (see Figure 3.1.2), and $(v_1, \ldots, v_{10})$ are the variables included in the FX speculative pressure component (see Figure 3.1.2). $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ are the errors of each equation respectively. Thus, in each case, we can define the weights of each variable in the components and in the overall index as following:

$$(5) \qquad \omega_i = \frac{\sum_{j=1} \lambda_j X_{ji}}{\sum_{j=1} \lambda_j}$$

Where j is the first principal component, i is the number of variables in the index or sub-index, λ_j is the variance of the first principal component, X_{ji} is the eigenvector of the correlation matrix. Results are shown in Figure 3.2.2. Weights are balanced between components and also between variables in each component in most of the cases[6]. The first component of the PCA explains between 69.5% in the case of Shadow Banking to 80% in the case of the external component. The SOE and Housing Bubble account for 63.2% and 65.8% respectively

6: For robustness check, we have created a reduced version of the index incluiding just the most correlated variables in each component with the component itself. Results do not change significantly and conclusions remain.

Figure 3.2.1 Relative weights of each variable and component in the overall index

Variable	Weights
Total profits	19.63
Institutional_reform_and_SOEs	12.37
Debt_and_SOEs	11.92
Local_government_and_SOEs	9.75
Industry_policy	9.50
Resource_misallocations_and_policy_failure	8.18
SOEs	7.15
Liabilities	10.62
Industry_laws_and_regulations	5.28
Resource_misallocations_and_SOEs	5.61
SOE Vulnerability Index	**29.18**

Variable	Weights
New Construction	16.37
mortages loan	14.57
Land_reform	12.62
Housing Price	11.60
Housing_construction	10.59
Housing_prices	10.05
Housing_policy_and_institutions	8.94
Housing_finance	7.83
Housing_markets	6.71
GICS Housing Index	0.36
Real Est. Investment	0.35
Housing Bubble Vulnerability Index	**26.05**

Variable	Weights
Wenzhou Index	16.58
WMP Yields	13.63
Infrastructure_funds	10.92
NPL Ratio	9.46
State_financial_institutions	8.95
Banking_regulation	7.28
Financial_vulnerability_and_risks	6.82
Asset_management	5.62
Financial_sector_instability	5.35
Bank_capital_adequacy	4.60
Non_bank_financial_institutions	4.35
Monetary_and_financial_stability	3.54
Acceptances	2.22
TSF Aggregate NewIncreased	0.57
Entrusted Loans	0.13
Shadow banking Vulnerability Index	**23.12**

Variable	Weights
Currency_exchange_rate	19.94
Exchange_rate_policy	17.95
Macroprudential_policy	15.33
HICNHON Index	13.84
CNY Curncy	11.92
Capital_account	10.05
CNH Curncy	8.73
Illicit_financial_flows	1.58
Foreign Reserves	0.60
Currency_reserves	0.06
FX Speculative Pressure Index	**21.64**

China Vulnerability Sentiment Index

Source: BBVA Research

In the second step, once we have the individual PCA we apply PCA to the four components in order to obtain a weighted index of independent vulnerability components as in (6), where μ_1 (29%), μ_2 (26%), μ_3 (23%) and μ_4 (22%) stand for the weights of every component in the total index. The percentage of the variance explained by the first component of the total index account for 61%.

$$(6) \qquad CSVI = \mu_1 \mathrm{SOE} + \mu_2 \mathrm{HB} + \mu_3 \mathrm{SB} + \mu_4 \mathrm{FX} + \varepsilon$$

Figure 3.2.2 Percentage of explained variance by the first component in each case

Indices	% of explained variance by the first component
SOE Vulnerability Index	63.17
Housing Bubble Vulnerability Index	65.76
Shadow banking Vulnerability Index	59.53
FX Speculative Pressure Index	78.99
China Vulnerability Sentiment Index	**61.16**

Source: BBVA Research

4. Results

4.1. The Evolution of China Vulnerability Sentiment Index (CVSI)

The CVSI index can be observed in Figure 4.1.1. The index captures the key risk events driving vulnerability sentiment in China. The CVSI shows a gradually improving vulnerability sentiment since the second half of 2016, just once the authorities started to react. The reaction was triggered by the the sharp decline of the index observed in the second half of 2015 and the first half of 2016 when two stock market correction and a sharp devaluation of the currency raised all the alarms. During this period, the CVSI dipped quite steeply in mid-2015, refecting deteriorating vulnerability sentiment following from two stock market crashes and an ill-timed devaluation of the RMB. [7]

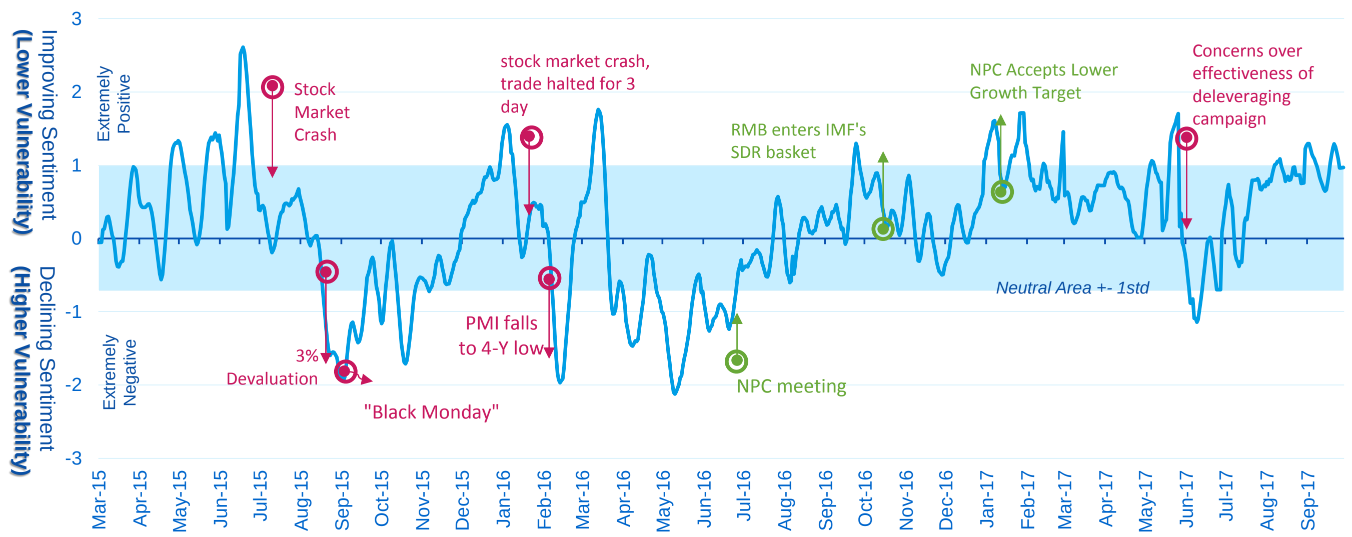
As this risky period wouldhave potential implications on the global financial system, it's not surprisingly that the risk narratives in global media deteriorated rapidly. Nevertheless, we see an inflexion point nearing the second half of 2016 and coinciding with the conclusion of the Chinese National Policy Committee meeting. During this event, authorities made clear their commitment to growth targets maintaining the compromise of reforms, adjusting the policy mix to more fiscal support combined with a tighter monetary policy stance and a stricter macro-prudential framework.

The general CVSI started to improve after the second half of 2016 despite a small correction in mid-2017 when the authorities' monetary prudence and regulatory tightenings aiming at shadow banking activities had pushed up interbank interest rates and added investors' concern over financial stability. It quickly rebounded after the authorities use more flexible and sophisticated approaches to pacify the market while maintaining the direction of financial deleveraging unchanged.

7: The series of stock market crashes ultimately resulted in a third of the value of the Shanghai Stock Exchange being wiped out in what is now referred to as "Black Monday". Together with concernes surrounding high corporate debt levels, the bulk of which has been accrued by SOEs (see Section 2.2) and one-sided expectations of Yuan devaluation, the events of the summer of 2015 fueled fears of a potential collapse of the Chinese economy.

BBVA Research

Figure 4.1.1. Chinese Vulnerability Sentiment Index (CVSI)
(Evolution of the "Tone" of main followed themes about vulnerability in China. Lower values indicate a deterioration of sentiment and higher vulnerability)



Source: www.gdelt.org and BBVA Research

Our strategy of differentiate in alternative sets of vulnerability provided some important insights as the performance of the components has not been homogeneous (see figure 4.1.2):

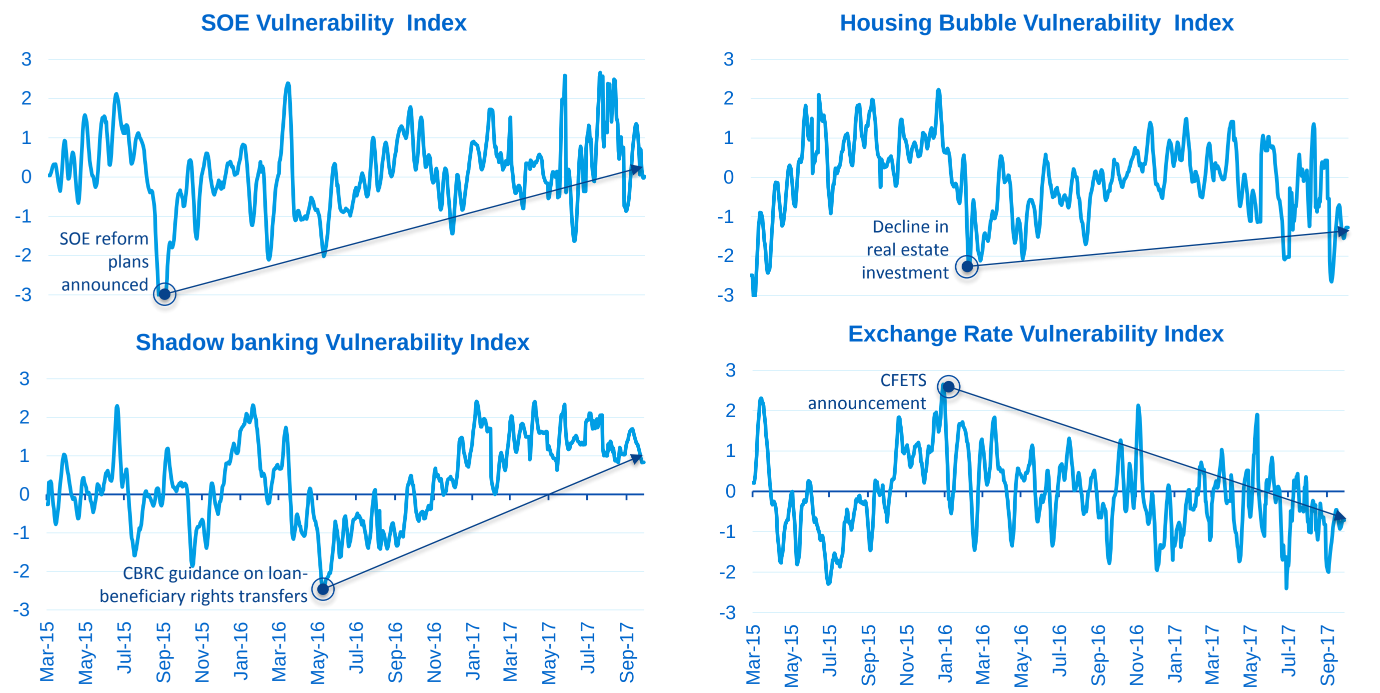
- **The SOE component has improved since the announcement of "supply side" reforms in the third quarter if 2015.** On September 2015, China's State Council issued a series of guidelines aimed to reform the SOEs and reduce overcapacity through closures and mergers by 2020. The plans also included were annual targets for overcapacity reduction in the bloated steel and coal sectors. Through a combination of policy measures and external factors, targets for 2016 were exceeded with state media announcing that steel and coal was cut by 45 million tons and 250 million tons respectively contributive to a more positive narrative overall. According to latest news reports, it is likely that the authorities can achieve their 2017 targets of eliminating overcapacity before the fourth quarter.

- **The shadow banking component reflects the most significant improvement in vulnerability sentiment.** This is the result of the combination of tighter liquidity conditions and stricter regulatory scrutiny on the banks' off-balance-sheet activities will has restricted bank's incentive to engage in regulatory arbitrage, gradually dampening the fast-growing pace of shadow banking activities (Moody's, 2017). Since the second half of 2016, the authorities have been persistently and painstakingly tightening regulations in a bid to clamp down shadow banking activities. More importantly, the authorities concluded its once-in-five-year Central Finanical Working Meeting in July 2017, at which the leadership announced that maintain financial stability will be their long-term target and they will establish a high-level financial stability committee to coordinate different regulators.

- **The housing bubble index improvement in vulnerability sentiment has been milder**. Since the first quarter of 2016, the authorities have stepped up their effors to limit price increases in first-tier cities such as Shanghai, Shenzhen and Beijing, limiting the risks of rapid increasing prices and a posterior stronger correction. As prices moderated the index started a recovery from negative values to near neutral. Although prices in tier-1 cities have moderate since the start of 2017, the housing market still have some overheating signals in smaller cities. As a result, the narratives of housing market have been relatively volatile. This is also a reflection of the fact that the Chinese authorities have allowed prices to increase in order to support GDP while simultaneously intervening to contain asset price risks in major cities.

- **The only component where we do not observe an improvement is the external one or Foreign exchange vulnerability index.** This is due to the fact that the RMB saw significant depreciatory pressures periodically in the aftermath of the one-off devaluation in August 2015, which led to outflows and a depletion of FX reserves during the period of August 2015 and January 2017. The depreciation once accelerated in the last quarter of 2016 after Donald Trump won the US presidential election and initiate a round of strong USD appreciation. Coming into 2017, the weakness of the USD has helped the Yuan to maintain its stability of exchange rate. The FX reserves have registered a streak of positive growth since February 2017. However, the narratives of exchange rate vulnerability haven't improved given that investors are still concerned with the Yuan's performance if a strong USD comes back. Above all, China's FX reserves, the most important cushion against external shocks, are now way below its level two years ago.

Figure 4.1.2. Chinese Vulnerability Sentiment Index: Improving in 3/4 components
(Evolution of the "Tone" of main followed themes about vulnerability in China. Lower values indicate a deterioration of sentiment and higher vulnerability)

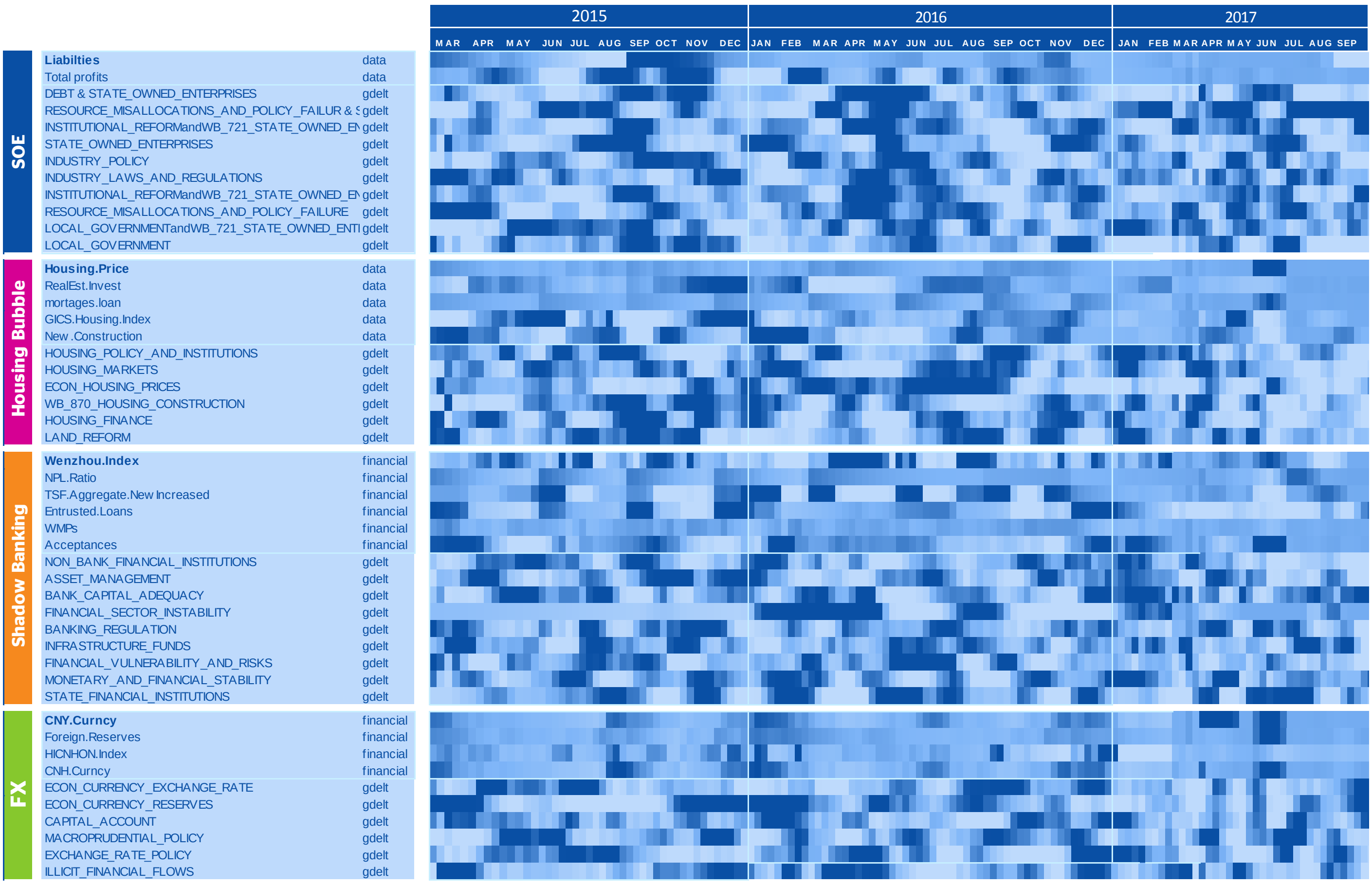


Source: www.gdelt.org and BBVA Research

4.2. Properties of the index

We find that high frecuency sentiment indicators can be used to nowcast vulnerability. In Figure 4.2.1, we show a heat map of the weekly evolution of each included variable in the index from March 2015 to September 2017. The map is divided by the subcomponents, with the hard data at the top of any of the group. The included variables in the heatmap are scaled, centered and de-trended when necessary. We can observe that there are co-movement patterns in the data. This is especially noticeable in the GDELT sentiment series for the SOE and FX components where pockets of vulnerabilities in darker blue colours are more apparent. In the remaining series we can see more coincident and gradual sinusoid patterns. Furthermore, there seems that at worst the sentiment indicators are coincident with some of them (SOE component) showing some nowcasting properties.

Figure 4.2.1. Chinese Vulnerability Sentiment Index Color Map: Patterns and Co-movements



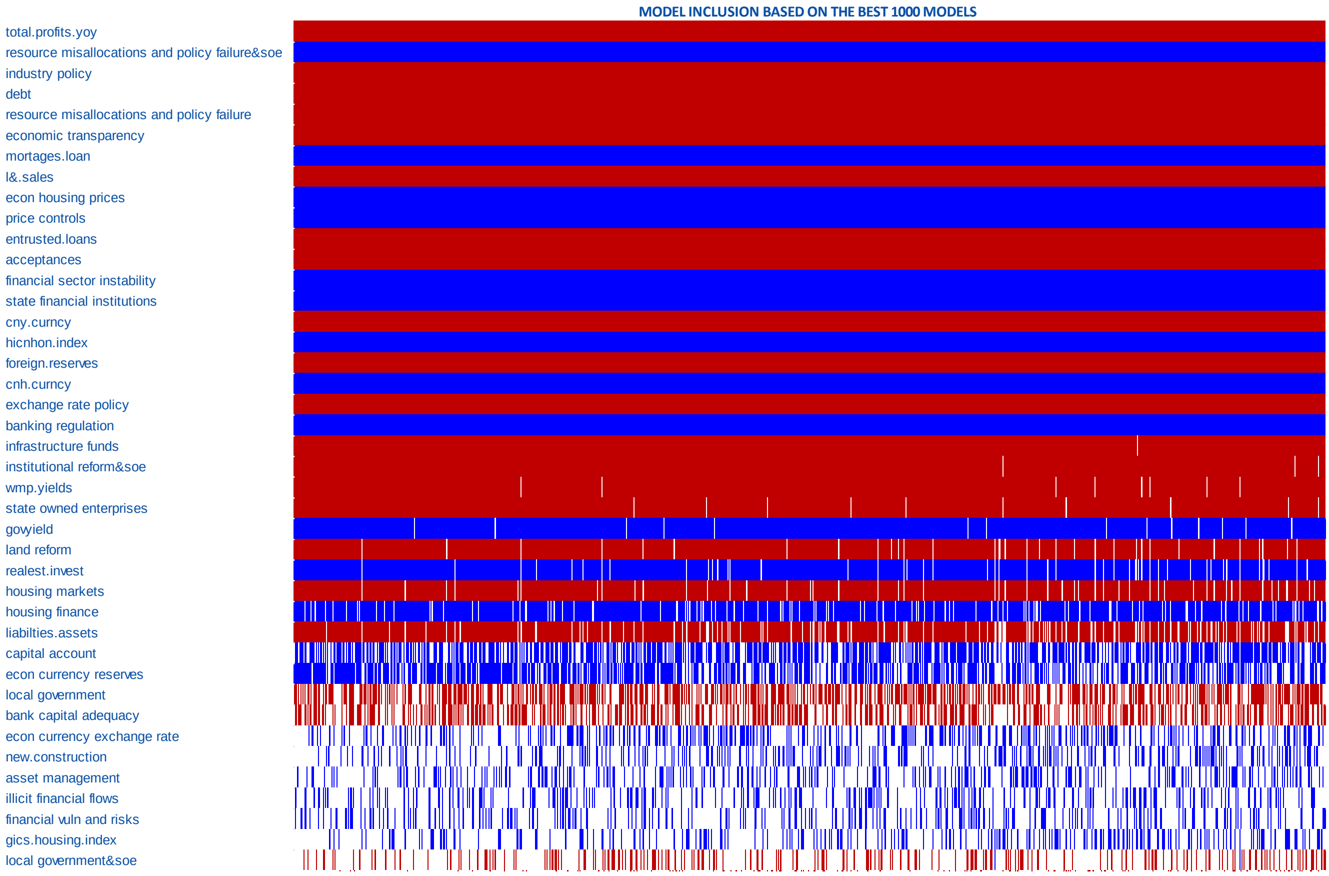
Source: BBVA Research

4.3. Robustness check

To check the robustness of the analysis we run a Bayesian Model Averaging analysis (Hoeting et al., 1999), using as dependent variable the CDS spread of the five year Chinese bond a market proxy for vulnerability). We test wether the set of our “a priori” selected variables will be including in a model explaining market risk.

The results of this analysis are presented in Figure 4.3.1. Here we can see variables in rows and 1000 of the estimated models by columns. A cell is colored if the Posterior Inclusion Probability or PIP is siginificant while the color indicates the sign (red for negative and blue for positive). Variables with a PIP smaller than 15% have been excluded from Figure 4.3.1. As we can see, more than half of the variables we selected have a very high Posterior Inclusion Probability (PIP), and more than half the variables (26 out of 54) have a probability of inclusion greater than 50%, and 87% of these most likely to be included in the CDS spread linear model are actually included in the index, providing solid evidence that the selected variables are indeed relevant in order to build a measure of risk and vulnerability. A 62% (16 out of the 26) of the variables with higher PIP than 50% are sentiment variables.

Figure 4.3.1. BMA Results



Source: BBVA Research

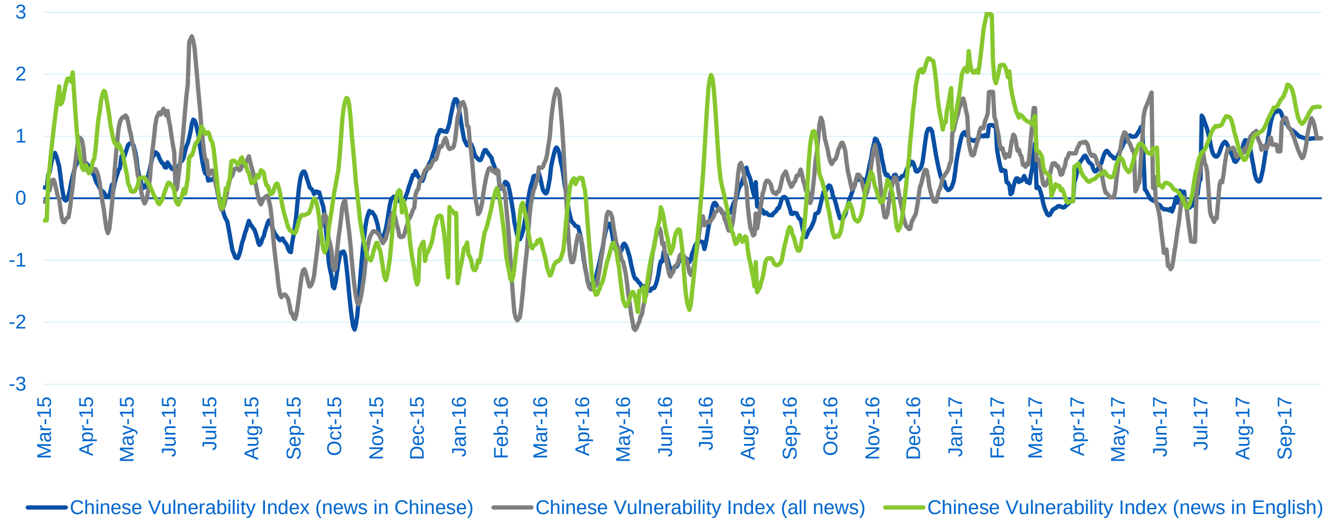
4.4. Checking for language bias

In this section we check whether there is a systematic language bias in the vulnerability sentiment about China. The GDELT database includes several languages around the world including the Chinese one. We build up two new indexes according to the languages in which the news are written. In our case we choose English (as representative media of worldwide financial centers) and Chinese written media.

Figure 4.4.1 shows some interesting patterns. First, there is no systematic bias in sentiment in any of the analyzed languages. Sentiment in English media was more negative at the end of 2015 and the beginning of 2016, but it has become more positive thereafter. Second, the Chinese index looks more stable and its standard deviation is in fact lower (0.66 the Chinese and 0.99 the English one), thus global narratives could amplify vulnerability sentiment. Third, according to the correlogram, the Chinese language sentiment seems to have leading indicator properties over the English one. This is not surprisingly at all if we assume more complete information by the Chinese media.

However, an important result is that the English media seems to have a higher correlation with market risk measures like the Chinese CDS than the local one. A reason of it is that this instrument will be priced in more liquid markets and therefore more influenced by the language of big financial centres. In fact, some previous works show that global factors have increased the role in explaining the Chinese CDS (Eyssell, Fung and Zhang, 2013). Moreover, the fact that there seems to be no inherent bias in English versus Chinese media also confirms theories by other authors. Pan, King and Roberts (2017) argue that the Chinese government fabricates 448 million social media comments a year with the objective of distracting public opinion with themes that are favorable for China. This is in contrast to previously held assumptions that online censors intervene to oppose dissenting views or change narratives.

Figure 4.4.1 Indices for different languages.



Source: BBVA Research

5. Conclusions

We develop an index to track vulnerability sentiment in China in real time using Big Data techinques which has shown to be robust and coherent to reflect the developments on the ground in China. The included variables in the analysis were carefully selected in order to capture the main vulnerabilities in China. The Bayesian Model Average analysis confirms that the posterior inclusion probability of most of the sentiment variables contribute to the market price of risk. Thus, our empirical analysis confirms that “sentiment” and narratives matter and can also explain vulnerabilities.

The results show that most of the analyzed components were deteriorating since mid 2015 and bottoming out at the first quarter of 2016. However, the performance of the vulnerabilities has not been symmetric and policy measures results have being more obvious. This is the case of the Shadow Banking and State owned Enterprises vulnerabilities, which according to our results improved more markedly than the rest.

The analysis shows that there’s no any systematic language bias in the vulnerability sentiment about China and, as expected, the local Chinese language sentiment seems to be a leading indicator over the English one. However, the latter has a higher correlation with market risk measures like the Chinese CDS. This is a key result as it confirms the existence of information asymmetries in emerging markets that could arise in low liquid markets. In this sense, communications policies could be enhanced if they are directed to these centers.

Further research will focus on analyzing the effects of CSVI using VAR analysis including policy variables. This would help us to design optimal policies to cope with global vulnerability in the four different dimensions of vulnerability. The effects of the language narratives have been tested with market measures of risk (CDS), but the analysis could be extended to a more analytical risk premium measures.

References

Ahuja, A. and Nabar, M. (2012): “Investment-Led Growth in China: Global Spillovers”, IMF Working Paper No. 12/267

Akerloff, G. and Shiller, R. (2009): *Animal Spirits :How human psychology drives the economy and why it matters for global capitalism*. Princeton Unversity Press 2009.

Angeletos, G.-M., and J. La'O (2013): “Sentiments", Econometrica, 81(2), 739-779.

Angeletos, GM, Collard, F. and Dellas, H. (2015). “Quantifying Confidence”, mimeo, Massachusetts Institute of Technology.

Barsky, R. B. and E. R. Sims (2012): “Information, animal spirits, and the meaning of innovations in consumer confidence," *The American Economic Review*, 102(4), 1343-1377.

Blanchard, O., 1993. "Consumption and the Recession of 1990-1991," Working papers 93-5, Massachusetts Institute of Technology (MIT), Department of Economics.

Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2013). *Which news moves stock prices? a textual analysis* (No. w18725). National Bureau of Economic Research.

Casanova, Carlos, Xia, Le and Ferreira, Romina (2016): "Measuring Latin America’s export dependency on China", *Journal of Chinese Economic and Foreign Trade Studies*, Vol. 9 Iss 3 pp. 213 – 233

Drummond, P. and Liu, E. (2013): “Africa’s Rising Exposure to China: How Large are Spillovers Through Trade?”, IMF Working Paper, Washington D.C.

Eyssell, T., Fung, HG, Zhang, G (2013) Determinants and price discovery of China sovereign credit default swaps. *China Economic Review*, Volume 24, Issue null, Pages 1-15

Huang, Yukon and Bosler, Canyon (2014): “China’s Debt Dilemma: Deleveraging While Generating Growth”, Carnegie Endowment for International Peace

Hoeting Jennifer A, Madigan David, Raftery Adrian E. and Volinsky Chris T. “Bayesian Model Averaging: A Tutorial”. *Statistical Science*, 1999, Vol. 14, No. 4, 382–417.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2016). *Measuring polarization in high-dimensional data: Method and application to congressional speech* (No. w22423). National Bureau of Economic Research.

Jolliffe, I.T. (2002) *Principal Component Analysis*, New York: Springer

Keynes, John Maynard (1936). *General Theory of Employment, Interest and Money*. London: Palgrave Macmillan.

Koss, R. (2015): “How Significant is Housing Risk in China?” IMF Global Housing Watch, Comments at the International Symposium on Housing and Financial Stability in China, Shenzhen

Kuziemko, I., & Washington, E. (2015). *Why Did the Democrats Lose the South? Bringing New Data to an Old Debate* (No. w21703). National Bureau of Economic Research.

Lee, M. Park, D. and Ramayandi, A. (2016):"How Growth Deceleration in the People's Republic of China Affects Other Asian Economies: An Empirical Analysis", Asian Development Bank Economics Working Paper Series No. 484

Leetaru, Kalev and Schrodt, Phillip3 2013."Global Database of Events, Language, and Tone (GDELT)

Li, J. Hsu, S. and Qin, Y. (2014): "Shadow banking in China: Institutional risks", *China Economic Review*, Vol. 31, pp. 119-129

Liu, B. (2010): "Sentiment Analysis and Subjectivity" *Handbook of natural language processing*, 2, 627-666.

Loughran,T. and McDonald,B (2011), "When is a Liability not a Liability," Journal of Finance, V66, pp. 35-65.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

Moody's Investor Services (2016): "Government of China: Sovereign Exposed to Sizeable, Rising Contingent Liabilities"

Moody's Investor Services (2017): "Quarterly China Shadow Banking Monitor"

Pan, J. King, G. and Roberts, M. (2017): "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument", American Political Science Review

Pigou, Arthur Cecil (1927). *Industrial Fluctuations*, London: Palgrave MacMillan.

Roll, R. (1988). The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis. *The Journal of Finance*, *43*(3), 541-566.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel Wilson. 2017. "Measuring News Sentiment." FRB San Francisco Working Paper 2017-01.

Shiller, R. (2017), "Narrative Economics". Cowles Foundation Discussion Paper N 2069.

Socher, R., Perelygin, A., Wu, J., Chuang, J. Manning. C. Ng A., Potts, C. "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank". In EMNLP 2013, pages 1631–1642

Stone,P, Bales,R. Namenwirth,Z & Ogilvie D (1962). The General Inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. Behavioral Science, 7(4), 484-498

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139-1168.

Tetlock, P. C., Saar‐Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437-1467.

Xu, E. and Chen, T. (2012): "The effect of monetary policy on real estate price growth in China*", Basin Pacific Finance Journal* Vol. 20, Iss. 1, pp. 62-77

Zhou, X., Zhang, W., and Zhang, J. (2012): "Volatility spillovers between the Chinese and world equity markets", *Pacific Basin Finance Journal*, Vol. 20, Iss. 2, pp. 247-270

Working Papers

2017

17/13 **Álvaro Ortiz, Tomasa Rodrigo, Le Xia, Joaquín Iglesias y Carlos Casanova:** Tracking Chinese Vulnerability in Real Time Using Big Data.

17/12 **José E. Boscá, Rafael Doménech, Javier Ferri y José R. García:** Los Desplazamientos de la Curva de Beveridge en España y sus Efectos Macroeconómicos.

17/11 **Rafael Doménech y José Manuel González-Páramo:** Budgetary stability and structural reforms in Spain: lessons from the recession and options for the future.

17/10 **Ángel de la Fuente:** Series enlazadas de algunos agregados económicos regionales, 1955-2014. Parte I: Metodología, VAB, PIB y puestos de trabajo.

17/09 **José Félix Izquierdo:** Modelos para los flujos de nuevo crédito en España.

17/08 **José María Álvarez, Cristina Deblas, José Félix Izquierdo, Ana Rubio y Jaime Zurita:** The impact of European banking consolidation on credit prices.

17/07 **Víctor Adame García, Javier Alonso Meseguer, Luisa Pérez Ortiz, David Tuesta:** Infrastructure and economic growth from a meta-analysis approach: do all roads lead to Rome?

17/06 **Víctor Adame García, Javier Alonso Meseguer, Luisa Pérez Ortiz, David Tuesta:** Infraestructuras y crecimiento: un ejercicio de meta-análisis.

17/05 **Olga Cerqueira Gouveia, Enestor Dos Santos, Santiago Fernández de Lis, Alejandro Neut y Javier Sebastián:** Monedas digitales emitidas por los bancos centrales: adopción y repercusiones.

17/04 **Olga Cerqueira Gouveia, Enestor Dos Santos, Santiago Fernández de Lis, Alejandro Neut and Javier Sebastián:** Central Bank Digital Currencies: assessing implementation possibilities and impacts.

17/03 **Juan Antolín Díaz and Juan F. Rubio-Ramírez:** Narrative Sign Restrictions for SVARs.

17/02 **Luis Fernández Lafuerza and Gonzalo de Cadenas:** The Network View: applications to international trade and bank exposures.

17/01 **José Félix Izquierdo, Santiago Muñoz, Ana Rubio and Camilo Ulloa:** Impact of capital regulation on SMEs credit.

2016

16/21 **Javier Sebastián Cermeño:** Blockchain in financial services: Regulatory landscape and future challenges for its commercial application

16/20 **Máximo Camacho and Matías Pacce:** Forecasting travelers in Spain with Google queries.

16/19 **Javier Alonso, Alfonso Arellano, David Tuesta:** Factors that impact on pension fund investments in infrastructure under the current global financial regulation.

16/18 **Ángel de la Fuente:** La financiación regional en Alemania y en España: una perspectiva comparada.

16/17 **R. Doménech, J.R. García and C. Ulloa:** The Effects of Wage Flexibility on Activity and Employment in the Spanish Economy.

16/16 **Ángel de la Fuente:** La evolución de la financiación de las comunidades autónomas de régimen común, 2002-2014.

16/15 **Ángel de la Fuente:** La liquidación de 2014 del sistema de financiación de las comunidades autónomas de régimen común: Adenda.

16/14 **Alicia García-Herrero, Eric Girardin and Hermann González:** Analyzing the impact of monetary policy on financial markets in Chile.

16/13 **Ángel de la Fuente:** La liquidación de 2014 del sistema de financiación de las comunidades autónomas de régimen común.

16/12 **Kan Chen, Mario Crucini:** Trends and Cycles in Small Open Economies: Making The Case For A General Equilibrium Approach.

16/11 **José Félix Izquierdo de la Cruz:** Determinantes de los tipos de interés de las carteras de crédito en la Eurozona.

16/10 **Alfonso Ugarte Ruiz:** Long run and short run components in explanatory variables and differences in Panel Data estimators.

16/09 **Carlos Casanova, Alicia García-Herrero:** Africa's rising commodity export dependency on China.

16/08 **Ángel de la Fuente:** Las finanzas autonómicas en 2015 y entre 2003 y 2015.

16/07 **Ángel de la Fuente:** Series largas de algunos agregados demográficos regionales, 1950-2015.

16/06 **Ángel de la Fuente:** Series enlazadas de Contabilidad Regional para España, 1980-2014.

16/05 **Rafael Doménech, Juan Ramón García, Camilo Ulloa:** Los efectos de la flexibilidad salarial sobre el crecimiento y el empleo.

16/04 **Angel de la Fuente, Michael Thöne, Christian Kastrop:** Regional Financing in Germany and Spain: Comparative Reform Perspectives.

16/03 **Antonio Cortina, Santiago Fernández de Lis:** El modelo de negocio de los bancos españoles en América Latina.

16/02 **Javier Andrés, Ángel de la Fuente, Rafael Doménech:** Notas para una política fiscal en la salida de la crisis.

16/01 **Ángel de la Fuente:** Series enlazadas de PIB y otros agregados de Contabilidad Nacional para España, 1955-2014.

Click here to Access the Working Paper published

Spanish

and English

Contact details:
BBVA Research
Azul Street, 4
La Vela Building - 4th and 5th floors
28050 Madrid (Spain)
Tel.: +34 91 374 60 00 and +34 91 537 70 00
Fax: +34 91 374 30 25
bbvaresearch@bbva.com
www.bbvaresearch.com