Tracking China Vulnerability in Real Time Using Big Data: The CVSI Index

Big Data Empirics and Policy Analysis Conference
Bank of England

Alvaro Ortiz., PhD*
BBVA Research

Summary

We have developed a China Vulnerability Sentiment Index using Big Data. The CVSI index allows us to track the 4 different vulnerability components: SOEs, Shadow Banking, Housing Bubble & FX speculative.

Robustness Check: “prior” inclusion of Sentiment Indicators is “Posterior” confirmed through Bayesian Model Averaging (BMA).

No Systematic Bias between English & Chinese Sentiment in Local and foreign media but transitory deviations can affect market sentiment.

Additional tools to track Risk in Real Time, with High Granularity and Geo-referenced Information.
Index

01 Measuring Sentiment using GDELT
02 Estimation & Robustness Check
03 Results: The CSVI Index
04 Additional Risk Analysis Tools
01
Measuring Sentiment Using GDELT
We introduce the Big Data (GDELT) analysis to track different events, themes,
We can extract more than 2300 emotions of more than 30000 topics in thousands of Newspapers of the world (Chinese & English) in real time.
Sentiment & Economics: References

Theoretical Literature

- **Pigou (1927):** business cycle fluctuations are driven by expectations and entrepreneurs’ errors of optimism and pessimism are crucial determinants of these fluctuations.
- **Keynes (1936)** highlighted the importance of changes in expectations that are not necessarily driven by rational probabilistic calculations, but by “animal spirits.”
- **Shiller (2017)** shows “narratives” can explain aggregate fluctuations through epidemic models.

Empirical Literature

- **Angeletos, Collard and Dallas (2015)** find that sentiment shocks account for around $1/2$ of GDP variance and $1/3$ 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.
- **Shapiro, Sudhoff, and Wilson (2017)** show how the news sentiment measures outperform the University of Michigan confidence index.
Measuring Sentiment

**Text Input**

- Print and web news media in over 100 languages from across every country
- Translated in real time to English and parsed

**Shallow Parsing**

- "Tokenization" or sentences
- Cleaning texts "The", "a" "," "," ... urls...
- Roots Econom* verb*
  ... 

**Deep Parsing**

- Sophisticated natural language processing algorithms

**Sentiment Aggregation**

\[
\text{Sentiment} = 100 \times \frac{\sum \text{Positive } W}{\sum \text{Total words}} - \frac{\sum \text{Negative } W}{\sum \text{Total words}}
\]

- **Positive W:** This is the percentage of all words in the article that were found to have a positive emotional connotation. Ranges are normally from -10 to +10.

- **Negative:** This is the percentage of all words in the article that were found to have a positive emotional connotation. Ranges from 0 to +100


Estimation and Robustness Check
Tracking China Vulnerability in Real Time: Value Added from Big Data

- **Hard Data Indicators**
  - ... provide accurate information but at lower frequencies and with delays...

- **Market Indicators**
  - ... Real Time but limited Information also influenced by global factors...

- **Sentiment Indicators**
  - ... complementing Hard-Soft-Markets in Real Time “sentiment” on Special topics not quotes.
A Balanced set of Information in the Database: Hard Data, Markets and Sentiment

**China Vulnerability Sentiment Index (CVSI)**

<table>
<thead>
<tr>
<th>SOE Vulnerability Index (SOEI)</th>
<th>Housing Bubble Vulnerability Index (HBI)</th>
<th>Shadow banking Vulnerability Index (SBI)</th>
<th>FX Speculative Pressure Index (FXI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgages.loan (M)</td>
<td>NPL.Ratio (M)</td>
<td>Foreign.Reserves (D)</td>
<td></td>
</tr>
<tr>
<td>GICS.Housing.Index (M)</td>
<td>TSF.Aggregate.New Increase (M)</td>
<td>CNV Exchange Rate (D)</td>
<td></td>
</tr>
<tr>
<td>Housing.Price (M)</td>
<td>Entrusted.Loans (M)</td>
<td>CNY Exchange Rate (D)</td>
<td></td>
</tr>
<tr>
<td>New.Construction (M)</td>
<td>Wenzhou.Index (D)</td>
<td>CNH Exchange Rate (D)</td>
<td></td>
</tr>
<tr>
<td>RealEst.Invest (M)</td>
<td>WMPs Acceptances (M)</td>
<td>HICNHON.Index (D)</td>
<td></td>
</tr>
</tbody>
</table>

**Principal Components Analysis on each component Tone**

- **Hard & Financial data**
  - Total.profits (M) Liabilities (M) 25%
  - Mortgages.loan (M) 45%
  - NPL.Ratio (M) 35%
  - Foreign.Reserves (D) 40%

- **Real time sentiment indicators**
  - State owned enterprises (D) 75%
  - Housing policy & institutions (D) 55%
  - Non bank financial institutions (D) 65%
  - Currency exchange rate (D) 60%
  - Resource misallocs & policy Failure (D) 45%
  - Housing prices (D) 55%
  - Bank capital adequacy (D) 65%
  - Institutional reform & SOEs (D) 60%
  - Housing construction (D) 55%
  - Financial sector instability (D) 65%
  - Industry policy (D) 60%
  - Housing finance (D) 55%
  - Banking regulation (D) 65%
  - Industry laws and regulations (D) 65%
  - Land reform (D) 55%
  - Infrastructure funds (D) 65%
  - Local government and SOEs (D) 60%
  - Debt and SOEs (D) 60%
  - Non bank financial institutions (D) 65%
  - Financial vulnerability & risks (D) 65%
  - State financial institutions (D) 60%
  - Monetary & financial stability (D) 65%
  - Macroprudential policy (D) 60%
  - Exchange rate policy (D) 65%
  - Implicit financial flows (D) 60%
A two step procedure to extract common vulnerability factor from Hard data, Markets and News Sentiment

1st Step Estimation: Components

(1) \( \text{SOEI} = \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_{10} x_{10} + \epsilon_1 \)
(2) \( \text{HBI} = \delta_1 y_1 + \delta_2 y_2 + \ldots + \delta_{11} y_{11} + \epsilon_2 \)
(3) \( \text{SBI} = \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{15} z_{15} + \epsilon_3 \)
(4) \( \text{FXI} = \rho_1 v_1 + \rho_2 v_2 + \ldots + \rho_{10} v_{10} + \epsilon_4 \)

with \( \gamma_i, \delta_i, \beta_i, \rho_i \) being the weight of every variable in the first principal component

2nd Step Estimation: Index

(6) \( \text{CSI} = \mu_1 \text{SOE} + \mu_2 \text{HB} + \mu_3 \text{SBI} + \mu_4 \text{FXI} + \epsilon \)

with \( \mu_1, \mu_2, \mu_3, \mu_4 \) being the weight of every component in the first principal component of the four components
### Results show the Importance of Sentiment

**Chinese Vulnerability Sentiment Index (CVSI): Weights**

<table>
<thead>
<tr>
<th>SOE Vulnerability</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Profits</td>
<td>19.63</td>
</tr>
<tr>
<td>Institutional Reform &amp; SOEs</td>
<td>12.37</td>
</tr>
<tr>
<td>Debt &amp; SOE</td>
<td>11.92</td>
</tr>
<tr>
<td>Liabilities</td>
<td>10.62</td>
</tr>
<tr>
<td>Local Government &amp; SOE</td>
<td>9.75</td>
</tr>
<tr>
<td>Industry Policy</td>
<td>9.5</td>
</tr>
<tr>
<td>Resource Mis &amp; P. Failure</td>
<td>8.18</td>
</tr>
<tr>
<td>SOE</td>
<td>7.15</td>
</tr>
<tr>
<td>Industry Laws &amp; Regulation</td>
<td>5.28</td>
</tr>
<tr>
<td>Resource Misalloc. &amp; SOE</td>
<td>5.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Housing Bubble</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Construction</td>
<td>16.37</td>
</tr>
<tr>
<td>Mortgages Loans</td>
<td>14.57</td>
</tr>
<tr>
<td>Land Reform</td>
<td>12.62</td>
</tr>
<tr>
<td>Housing Price</td>
<td>11.6</td>
</tr>
<tr>
<td>Housing Construction</td>
<td>10.59</td>
</tr>
<tr>
<td>Housing Prices</td>
<td>10.05</td>
</tr>
<tr>
<td>Housing Policy &amp; Institutions</td>
<td>8.94</td>
</tr>
<tr>
<td>Housing Finance</td>
<td>7.83</td>
</tr>
<tr>
<td>Housing Markets</td>
<td>6.71</td>
</tr>
<tr>
<td>GICS Housing Index</td>
<td>0.36</td>
</tr>
<tr>
<td>Real State Investment</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shadow Banking</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wenzhou Index</td>
<td>16.58</td>
</tr>
<tr>
<td>WMP Yields</td>
<td>13.63</td>
</tr>
<tr>
<td>Infrastructure_funds</td>
<td>10.92</td>
</tr>
<tr>
<td>NPL ratio</td>
<td>9.46</td>
</tr>
<tr>
<td>State &amp; Financial Inst</td>
<td>8.95</td>
</tr>
<tr>
<td>Banking_Regulation</td>
<td>7.28</td>
</tr>
<tr>
<td>Financial Vulnerability</td>
<td>6.82</td>
</tr>
<tr>
<td>Asset_Management</td>
<td>5.62</td>
</tr>
<tr>
<td>Financial Sector Instability</td>
<td>5.35</td>
</tr>
<tr>
<td>Bank Capital Adequacy</td>
<td>4.6</td>
</tr>
<tr>
<td>Non Bank Financial Inst</td>
<td>4.35</td>
</tr>
<tr>
<td>Monetary &amp; F.Stability</td>
<td>3.54</td>
</tr>
<tr>
<td>Acceptances</td>
<td>2.22</td>
</tr>
<tr>
<td>TSF Aggregate New</td>
<td>0.57</td>
</tr>
<tr>
<td>Entrusted Loans</td>
<td>0.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FX Speculative Pressure</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currency Exchange Rate</td>
<td>19.94</td>
</tr>
<tr>
<td>Exchange Rate Policy</td>
<td>17.95</td>
</tr>
<tr>
<td>Macroprudential_Policy</td>
<td>15.33</td>
</tr>
<tr>
<td>Hinchon Index</td>
<td>13.84</td>
</tr>
<tr>
<td>CNY Currency</td>
<td>11.92</td>
</tr>
<tr>
<td>Capital Account</td>
<td>10.05</td>
</tr>
<tr>
<td>CNH Currency</td>
<td>8.73</td>
</tr>
<tr>
<td>Illicit Financial Flows</td>
<td>1.58</td>
</tr>
<tr>
<td>Foreign Reserves</td>
<td>0.6</td>
</tr>
<tr>
<td>Currency Reserves</td>
<td>0.06</td>
</tr>
</tbody>
</table>

- % of Sentiment in Component (S) | 69.76
- % of Variance by 1st PC in the CSV | 63.2
- % of Variance by 1st PC in the CSV | 29.18
- % of Sentiment in Component (S) | 56.74
- % of Variance by 1st PC in the CSV | 65.8
- % of Variance by 1st PC in the CSV | 26.05
- % of Sentiment in Component (S) | 57.43
- % of Variance by 1st PC in the CSV | 59.5
- % of Variance by 1st PC in the CSV | 23.12
- % of Sentiment in Component (S) | 48.27
- % of Variance by 1st PC in the CSV | 78.99
- % of Variance by 1st PC in the CSV | 21.64
Robustness Check: Co-movement (Hard-Market-Sentiment)

Chinese Vulnerability Sentiment Index Colour Map: Patterns and Comovements
(Standard values. Light Blue values indicate a improvement of sentiment while dark blue stands for higher vulnerability sentiment)
Source: own calculations
Robustness Check: Do our selected variables account for “underlying” vulnerability

Posterior Inclusion Probability of a set of “X” candidates variables in explaining “Risk” (“Y”) (Risk = y = proxied by Chinese CD Swap)

\[ p(y|M_\gamma, X, g) \propto (y - \bar{y})' (y - \bar{y})^{-\frac{N-1}{2}} (1 + g)^{-\frac{k_\gamma}{2}} \left( 1 - \frac{g}{1 + g} \right)^{-\frac{N-1}{2}} \]

\[ X_{\gamma} \in \{ X \} \]

Zellner’s Uniform Prior (Non informative Prior, \( \beta = 0 \))

\[ \beta_\gamma | g \sim N \left( 0, \sigma^2 \left( \frac{1}{g} X_\gamma'X_\gamma \right)^{-1} \right) \]

Hyperparameter “g” shows certain is the researcher fo B to be zero
Bayesian Model Averaging Robustness check confirms the relevance of Sentiment in explaining Market Risk proxies
Results: The CVSI index
The index shows that Vulnerability sentiment has been improved since the authorities implemented some policies.

Chinese Vulnerability Sentiment Index (CVSI)
(Evolution of the “Tone” or “Sentiment”. Lower values indicate a deterioration of sentiment and higher vulnerability)

Source: www.gdelt.org & Own Calculations
Some components have some an important response to policy changes (i.e. Shadow Banking)

CVSI: SOEs Component
(Evolution of the "Tone" or "Sentiment").

CVSI: Shadow Banking Component
(Evolution of the "Tone" or "Sentiment").

- NDRC and State Council promulgated the "Ten rules" of SOE reform
- Moody's Downgrade
- S&P Downgrade
- SoE Mixed ownership Reforms Plan
- The start of Mixed ownership reform of China Unicom
- Stock Market Crash
- Stock Market Crash and RMB depreciation
- CBRC Guidance on Loan Beneficiaries rights Transfer
- Moody's Downgrade
- S&P Downgrade
Other components present more random performance with some dependence of global affairs (External)

**CVSI: Real State Component**
(Evolution of the "Tone" or "Sentiment").

- Sharp decline in Real State Investment
- A series of tightening measures were promulgated
- Large Chinese real estate developers sold their land

**CVSI: External**
(Evolution of the "Tone" or "Sentiment").

- FED announced the interest rate hike
- FED Doubts On Hawkish Stance
- DXY depreciated sharply
- FED Confirms Exit path
- CFTS Announcement

A series of tightening measures were promulgated.
04

Additional Risk Analysis Tools
The CSVI is associated to other High Frequency Indicators as Growth and Market Risk

**CVSI Index and Economic Activity (PMI)**
(index and net balance)

**CVSI Index and Risk Premia (CDS)**
(index and basis points Inverted)
Chinese Vulnerability Sentiment Index: total news VS Chinese news

Chinese Vulnerability Sentiment Index: total, Chinese and English
(Evolution of the “Tone” of main followed themes about vulnerability in China in Mandarin and Non-Translated-English)

No Systematic Bias in English & Chinese Language Media

Transitory divergences Can affect markets
It can be used to analyze vulnerability in a very granular way…
Including Geo referenced risks

Geographical Analysis Housing Prices
(sentiment on Housing Prices)
Tracking China Vulnerability in Real Time Using Big Data: The CVSI Index

Big Data Empirics and Policy Analysis Conference
Bank of England

Alvaro Ortiz., PhD*
BBVA Research