

The use of Big Data at BBVA Research

to better understand the society,
the economy and the world

Structured Products & Digitalization in
Private Banking

Geneva, Bloomberg Office

January 2020

Index

- 01 Opportunities in the digital era. Big Data at BBVA Research
- 02 Geopolitics, Trade and Spill overs
- 03 Economic & Risk indicators in Real Time
- 04 Text Mining and Sentiment analysis

01

Opportunities
in the digital era.
Big Data at
BBVA Research

New sources of data provide a better, "Real Time" and "High Definition" analysis for our clients...



Economic indicators in Real Time



Economic Analysis in High Definition



Social & Economic Networks



The use of **Big Data** and **Data science** techniques allows us to quantify these trends.

... and also exploiting non traditional data to better understand political, social and geopolitical challenges



Geopolitical events



Political events and social reaction



Natural disasters and epidemics



New data may end up changing the way in which we approach empirical questions and the tools they use to answer them.

Text as Numbers: a key source of information to enrich analysis



of the total amount of web pages on the internet is given **textual or unstructured data**



Text mining to extract meaning from strings of letters



The potential use of textual information and text sources **improves the understanding of the social, economic and financial systems**

It helps us to comprehend what drives **agent's economic expectations, policy decisions, measures of uncertainty, ...**

New framework in the digital era



> Availability of new types of datasets (including text)



> Combination of historical data with real time data



> Advanced data science techniques and algorithms



> New answers to old questions



> Better, faster infrastructure higher computational abilities



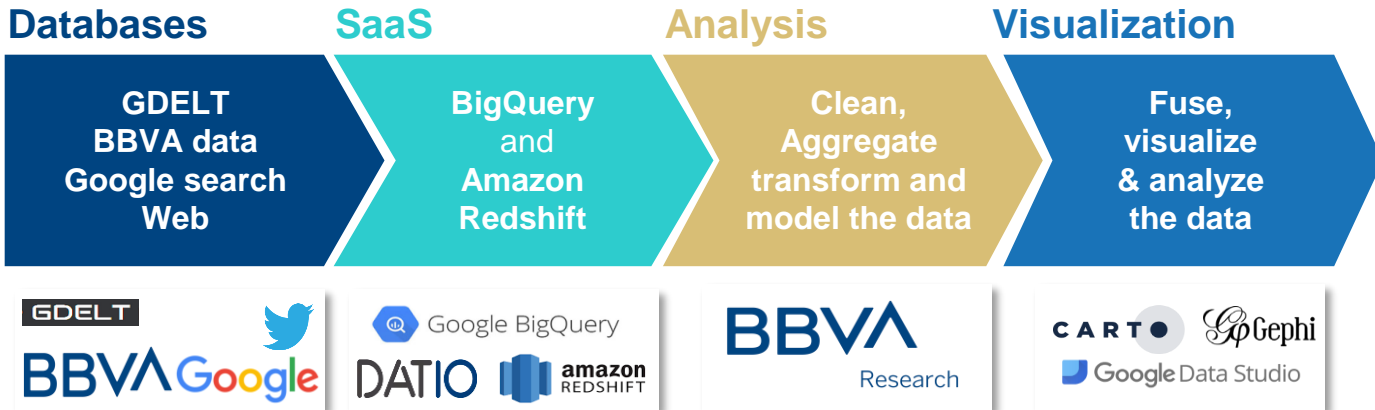
> Especially useful in countries where lower data availability

20%
of used data

80%
of available data



Our working process



Technology provides us new opportunities to analyze massive data on real time and high definition coming from new sources

Massive Flows of New Data

On Real Time

High Definition

New Sources: Texts & Images

Political, Geopolitical Social Indexes (Political Index)



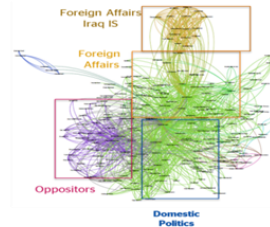
To elaborate Political, Social & geopolitical Indexes...

Geographical Analysis Housing Prices (sentiment on Housing Prices)



... analyze with high granularity the transmission of shocks...

Politics & Financial Networks (Political Networks)



...Build Networks to understand the interaction between agents...

Monetary & Stability tones by Central Banks (sentiment Analysis on Central Bank texts)



... Use Text as Numbers to interpret sentiment in policy docs...

Measuring Sentiments, Narratives on News (sentiment Analysis on Economy and Society News)



... disentangle "Information Bias" between what we read and reality

Mix Hard data & Sentiment & VAR models (Vulnerability and Risks Index Models)



... build hybrid models including hard and sentiment data to ...

02

News & Narratives: News & Networks for Economics, Political and Geopolitical analysis

News as a new source of information.

What is GDELT?

Global Database on Events Location and Tone

- Open database of human society from every corner of the globe dating back to 1979
- ... georeferenced across the entire planet...
- ... including over 300 events around the world and more than 30000 themes...
- ...and collecting emotions using some of the most sophisticated algorithms

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



Global Content Analysis Measures (GCAM)

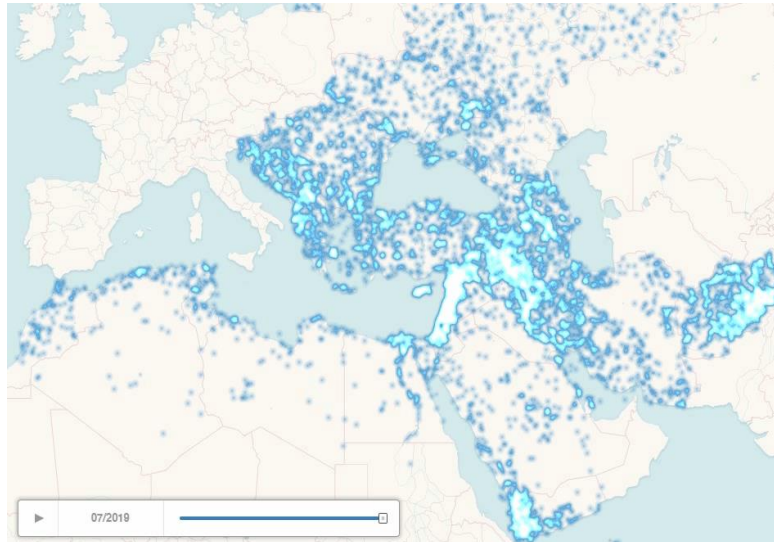
40 different **sentiment dictionaries** to capture more than **2300 emotions and dimensions** with a value score

GDELT allow us Tracking Geopolitics on real time...

Is useful to identify the main hot spots and potential spillovers

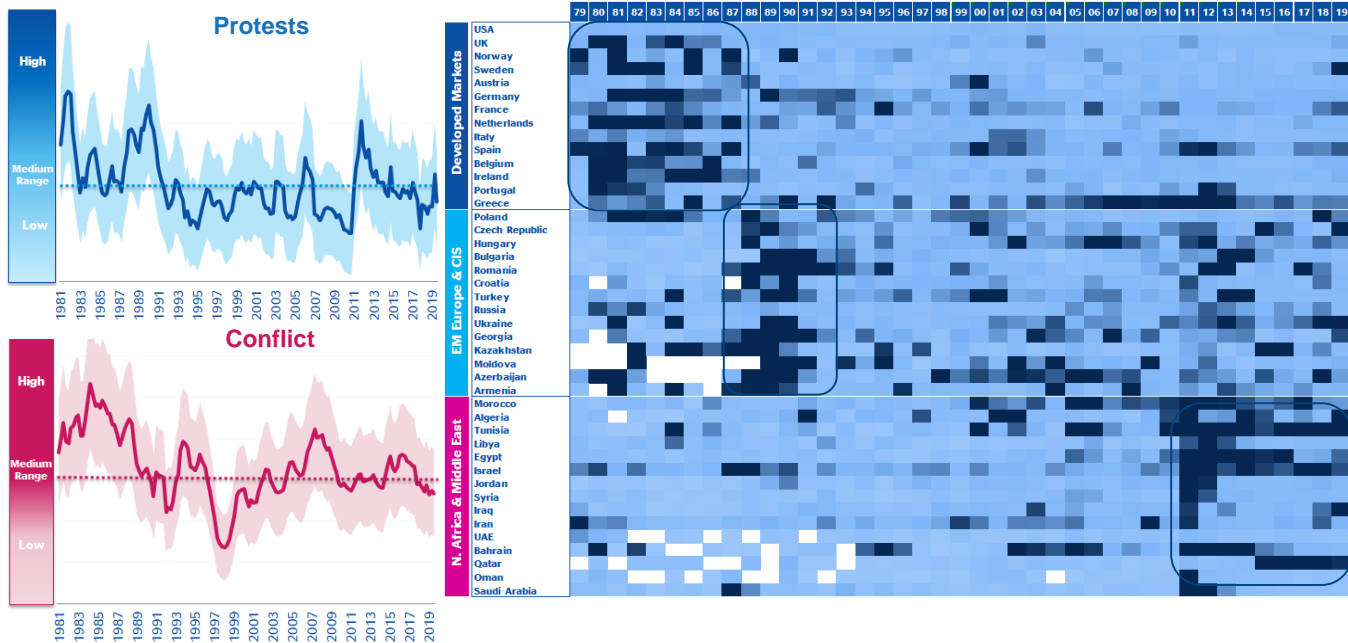
CONFLICT INTENSITY MAP 2017-19

(NUMBER OF CONFLICTS/ TOTAL EVENTS)



...and also from an historical perspective...

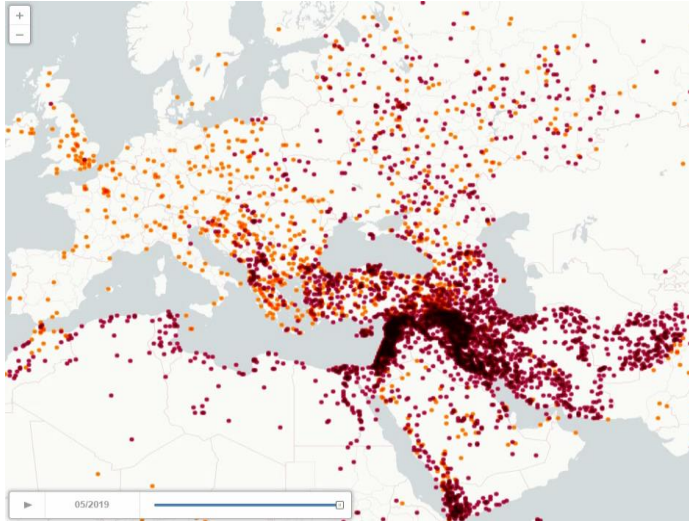
BBVA RESEARCH WORLD CONFLICT AND PROTEST INTENSITY INDEX 1979-2019



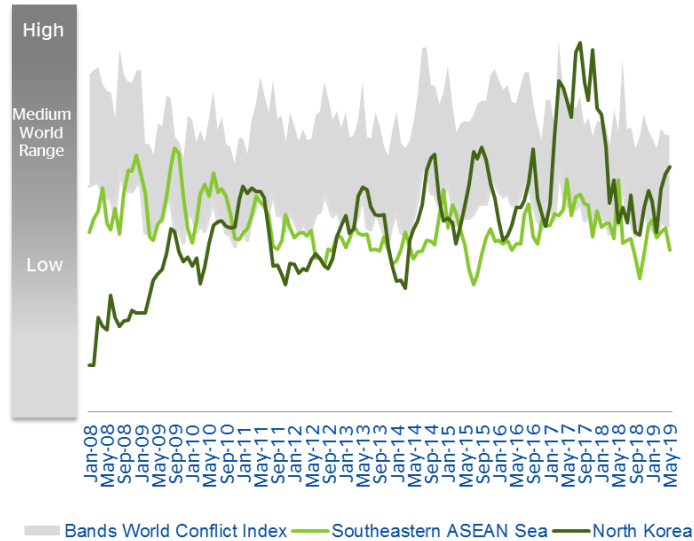
...to the current “hot” spots...

BBVA RESEARCH REFUGEES FLOWS MAP IN 2015-19

MEDIA CITATIONS ABOUT REFUGEES' INFLOWS
AND OUTFLOWS



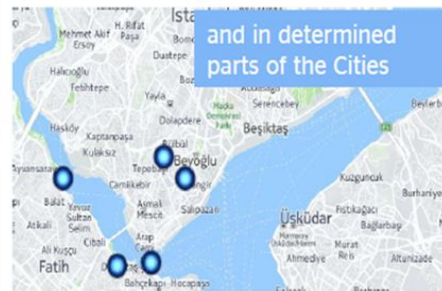
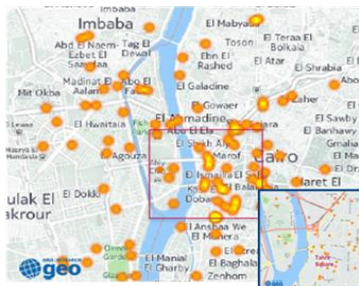
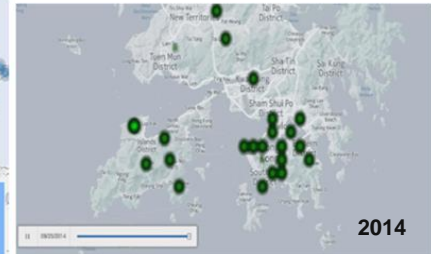
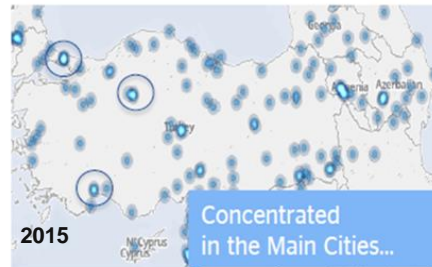
BBVA RESEARCH ASIA CONFLICT INTENSITY INDEX 2008-19



... in “high definition”

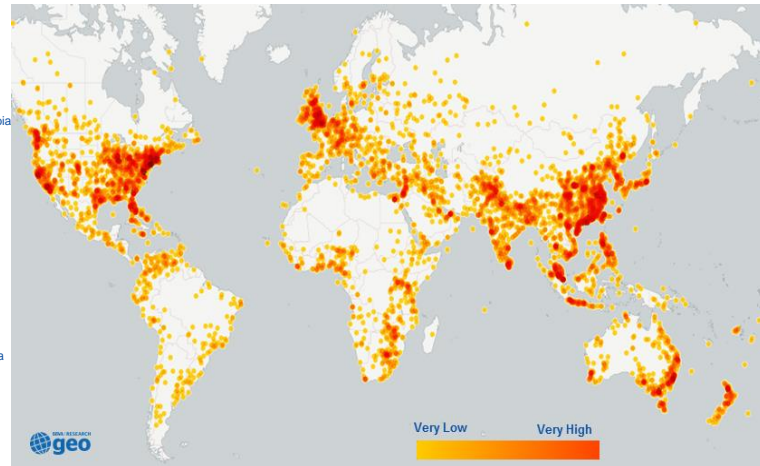
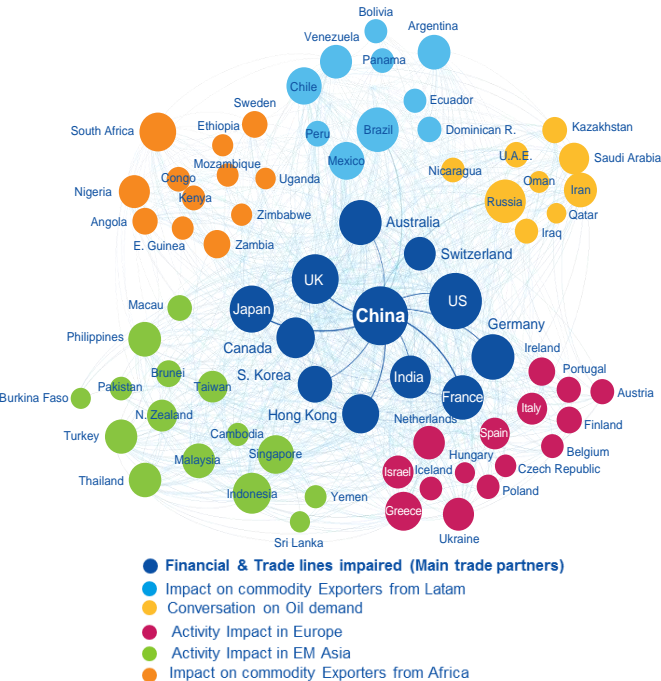
SOCIAL UNREST EVENTS ACROSS THE WORLD: CAIRO, ISTANBUL AND HONG KONG CASES

PROTEST EVENTS



Exploring spillover effects: from the China's links with the rest of the world...

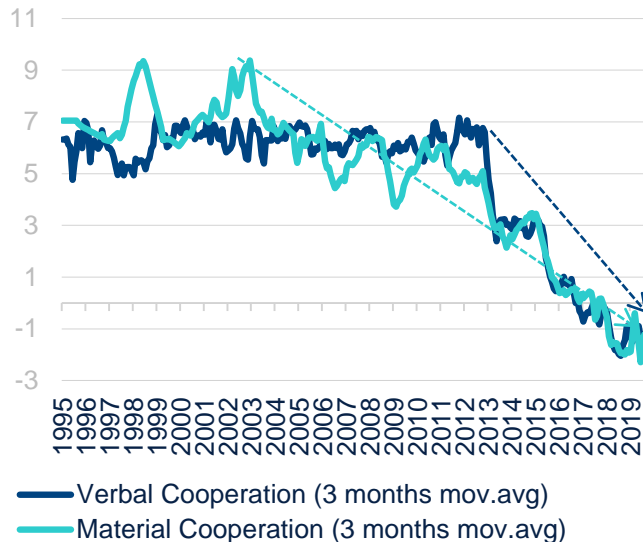
CHINESE SLOWDOWN: THE COUNTRY NETWORK AND THE MEDIA PERCEPTION



“Words” can be different from “Deeds” some times... like the world trade material evolution and its narrative

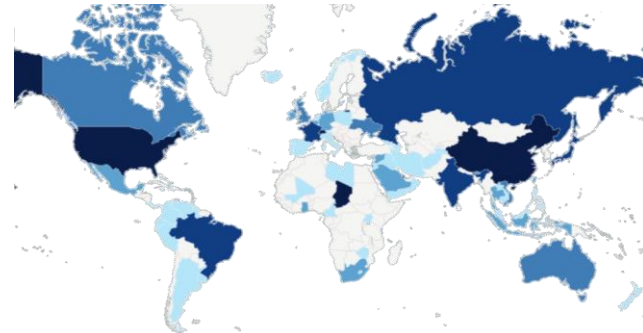
BBVA RESEARCH WORLD TRADE SUPPORT INDEX

(TONE & COVERAGE VERBAL COOPERATION AT WTO)



BBVA RESEARCH TRADE SUPPORT INDEX 2008-17

(VERBAL COOPERATION INDEX, IN DIFF 2017-2008)

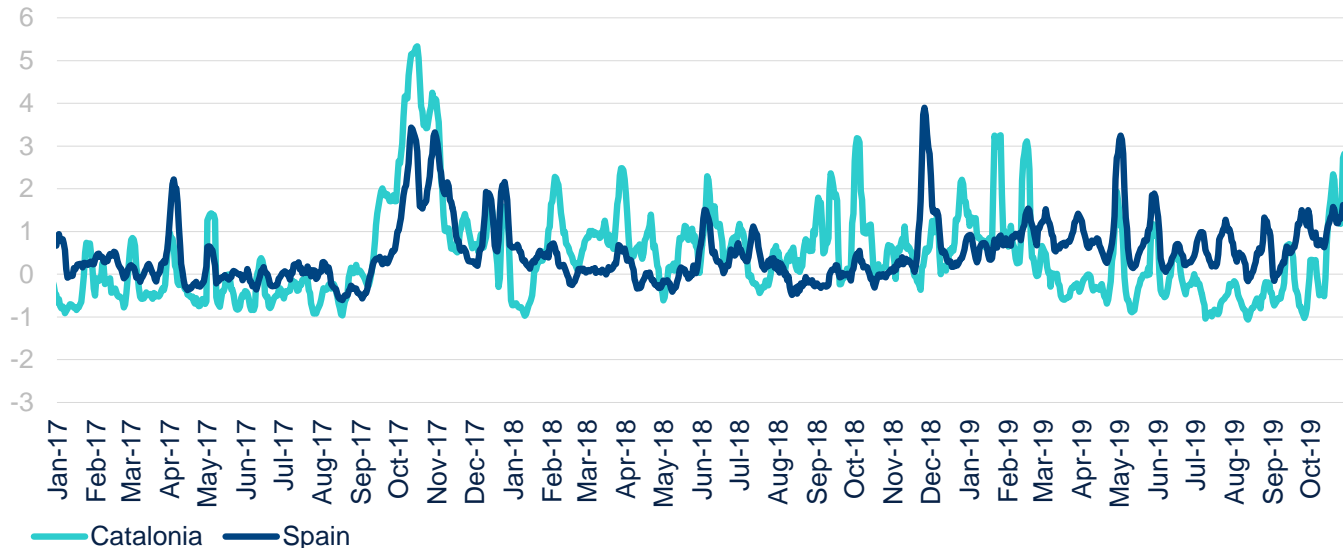


THE GLOBAL INCIDENCE OF PROTECTIONISM 2008-2015 (GLOBAL TRADE ALERT)



We can quantify events such as the Economic Policy Uncertainty Index in Spain based on news

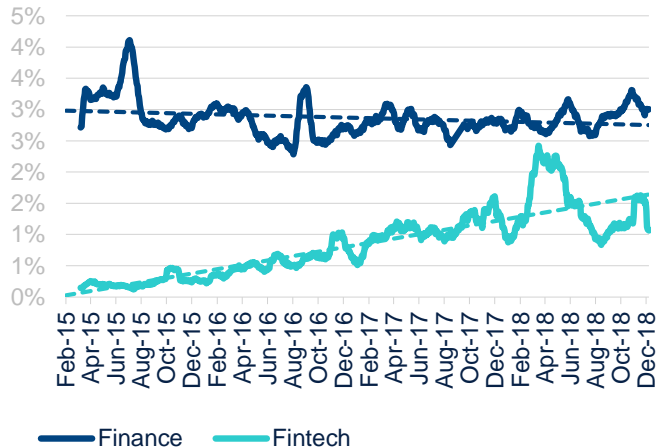
BBVA RESEARCH ECONOMIC POLICY UNCERTAINTY INDEX: SPAIN AND CATALONIA (NORMALIZED DATA, MOVING AVERAGE 7 DAYS)



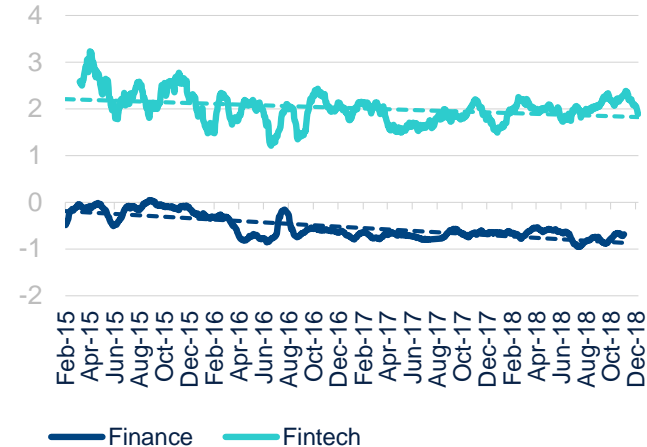
Economic Policy Uncertainty increased in Spain and Catalonia since 2015. General and Catalonian elections, Brexit, 1-O, ... are moving the series upwards.

The Fintech Universe through the Eyes of the Media and Social Networks

FINTECH AND FINANCE MEDIA COVERAGE OVER TIME 2015-2018 (MOV AVG 30 DAYS. RELATIVE RATIO WITH RESPECT TO TOTAL NEWS)



FINTECH AND FINANCE MEDIA SENTIMENT OVER TIME 2015-2018 (MOV AVG 30 DAYS)



Media coverage is calculated as the ratio between news related with Fintech/ Fintech&Data / Fintech&Digital Payments with respect to the total number of news in the same day. Media sentiment goes normally from -10 (negative perception) to 10 (positive perception), with 0 indicating neutral.

Media coverage of FinTech increased over time, catching up with that of Finance and with a persistent gap between the perception of both topics.

03

Economics in Real Time and High Definition: Media and Google Trends

External databases: Google searches database



- Google Search provides several features beyond searching for words and are available since July 2007.
- The analysis of the frequency of search terms may indicate the evolution of economic, social and health trends.

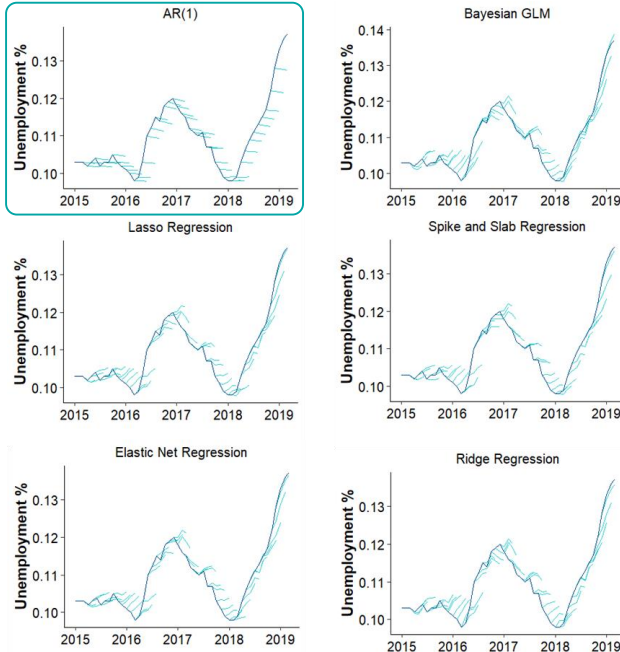
Example: a database with aggregate information about Google queries related to unemployment in Turkey has been developed



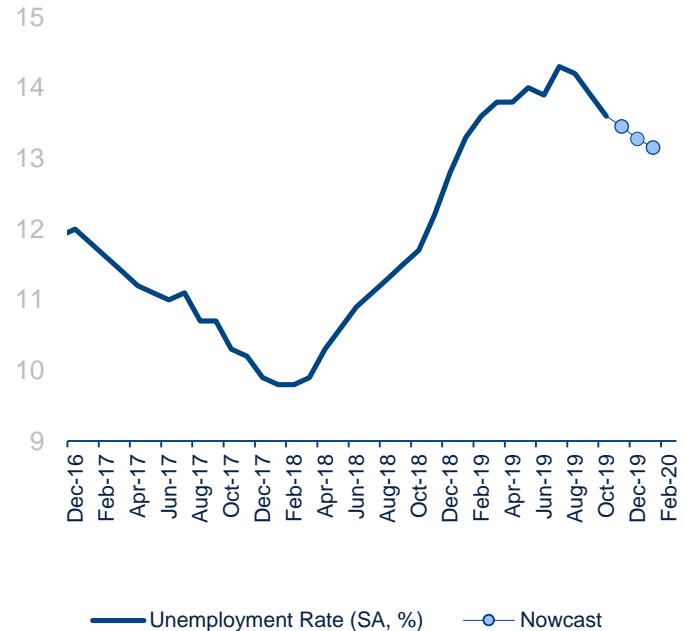
Providing extra information to nowcast unemployment. In the case of Turkey providing an advantage of 3 months

UNEMPLOYMENT OUT-OF-SAMPLE FORECASTS (3M RECURSIVE OUT OF SAMPLE FORECASTS)

Benchmark to beat



TURKEY: UNEMPLOYMENT RATE (SA) NOWCAST

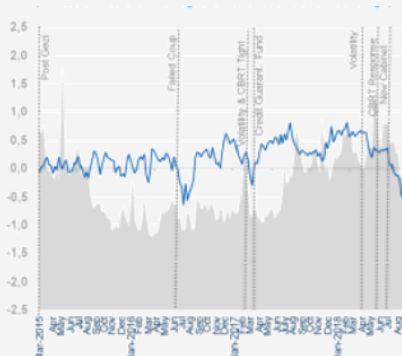


Corporate news Sentiment can provide us “Early Warning Signals” of corporate balance sheet health...

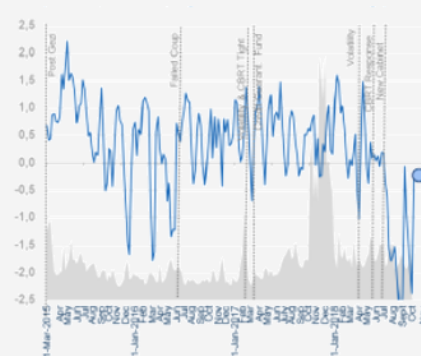
TURKEY: BBVA RESEARCH MEDIA SENTIMENT AND COVERAGE ON TURKISH CORPORATES



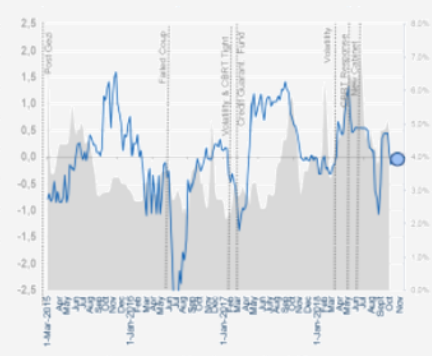
TOTAL CORPORATES



BANKS

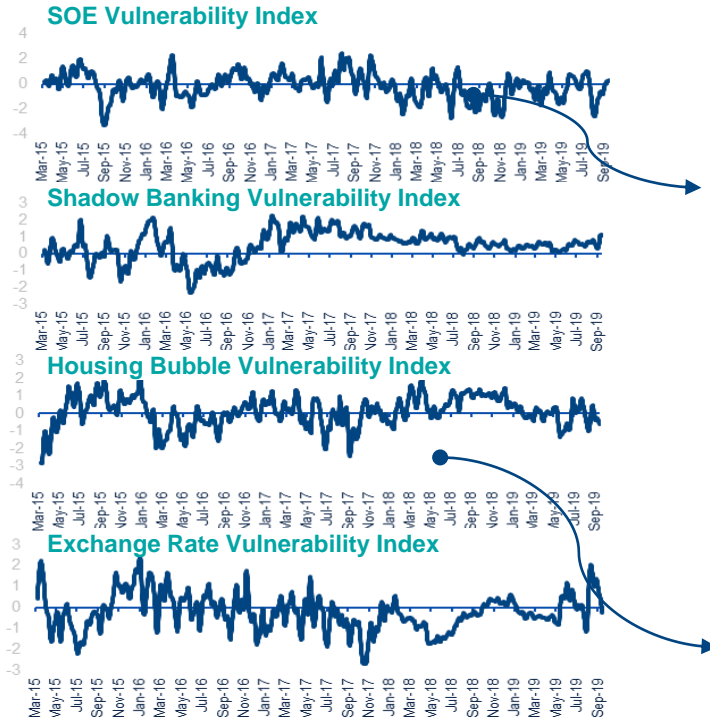


MANUFACTURERS

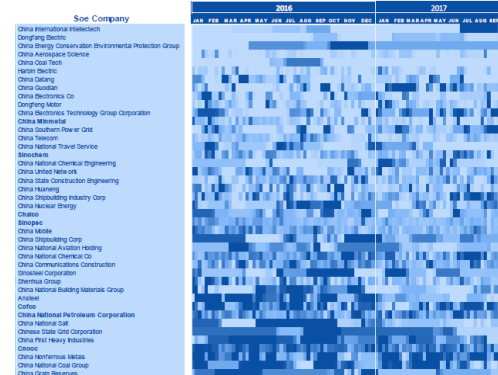


And we can develop hybrid Indicators (Hard Data & Sentiment Data) to disentangle risks in China

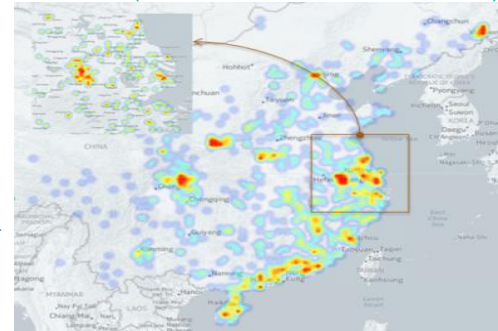
CHINESE VULNERABILITY SENTIMENT INDEX (CVSI)



CHINA SOE MAP (SENTIMENT ON SOE)



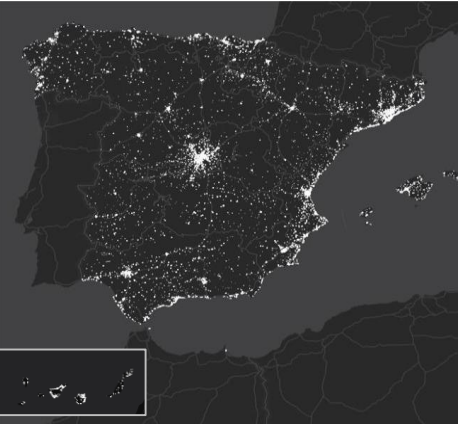
GEOGRAPHICAL ANALYSIS HOUSING PRICES (SENTIMENT ON HOUSING PRICES)



03

Economic & Risk indicators through Transactions, Google Searches & International News

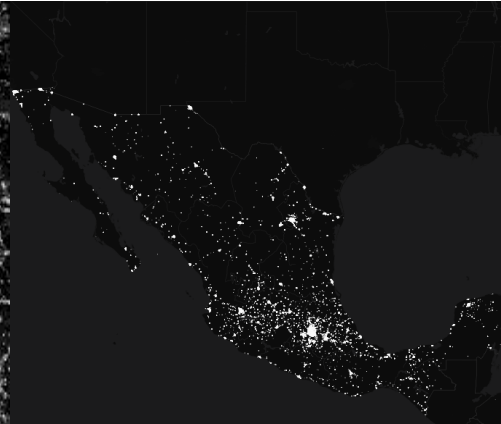
Internal databases: working with aggregated and anonymized BBVA Data



710M card transactions from 1M PoS, made by 53M people, representing €43.000M



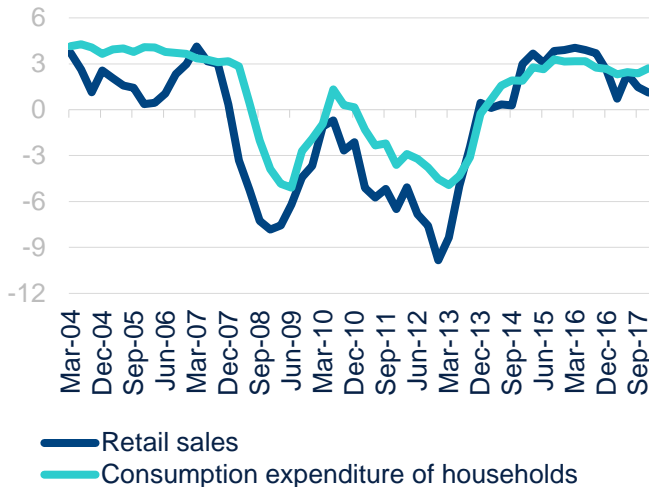
4 billion card transactions made by 14M people



1.500M card transactions from 1,1M PoS, made by 88M people, representing €41.000M

Retail trade sector dynamic leads the evolution of the aggregate consumption, which represents a high share of the GDP

SPANISH CASE: RETAIL SALES VS. CONSUMPTION EXPENDITURE OF HOUSEHOLDS (% , YOY)



Source: BBVA Research and BBVA Data & Analytics from INE data

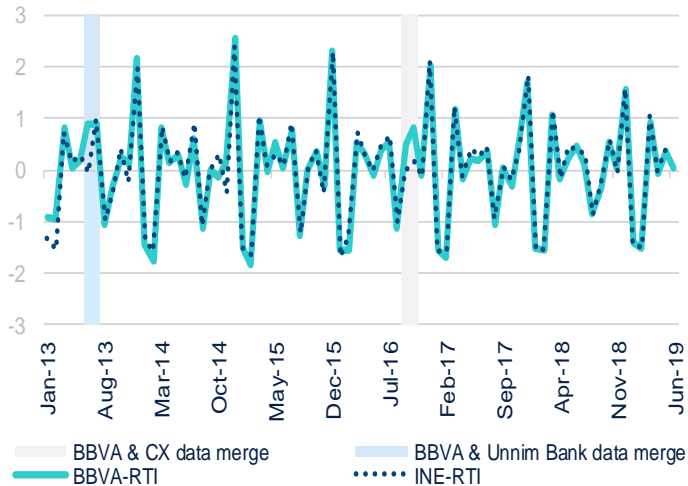
REPLICATING INE DATA TREATMENT AND METHODOLOGY USING TRANSACTIONAL DATA



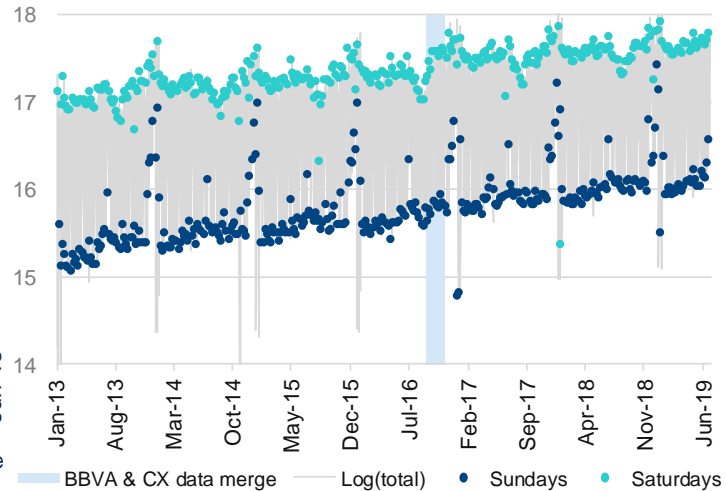
Source: BBVA Research and BBVA Data & Analytics

Using BBVA data, we replicate national figures...

RETAIL TRADE INDICES: BBVA VS INE (STANDARDIZED MONTHLY GROWTH RATE)



AGGREGATE RETAIL TRADE - DAILY FREQUENCY (LOGARITHMS)



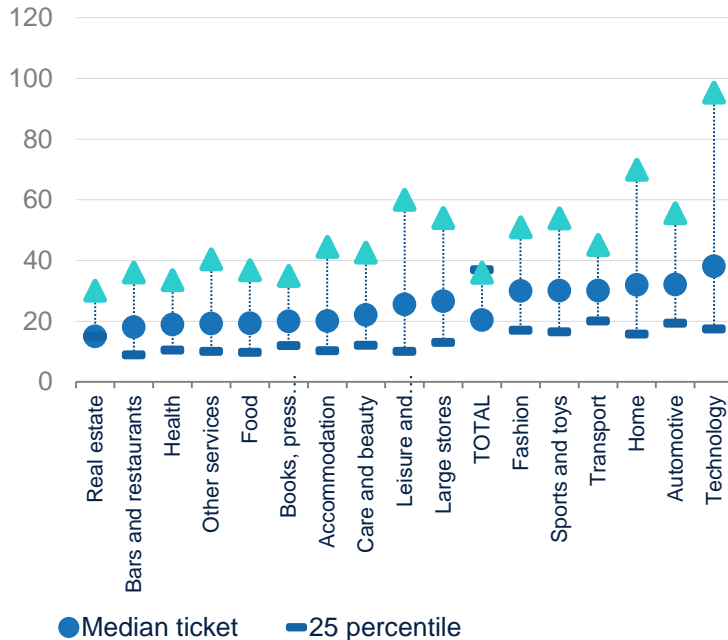
High granularity:
Dynamics down to
subnational level

Multi Dimensional:
More detailed
socioeconomic features

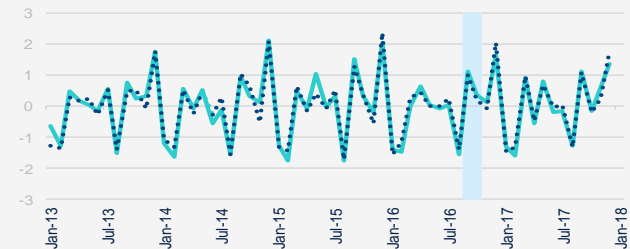
Ultra High Frequency:
Dynamics up to
sub-monthly frequency

... going to further . The retail sales by sector and provinces

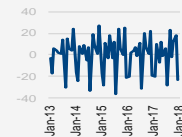
BBVA RTI BY MERCHANT (MEDIAN TICKET IN DEC-17, €)



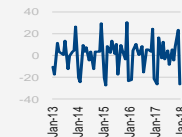
BASQUE COUNTRY



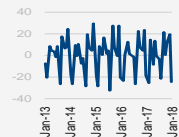
ÁLAVA



GUIPÚZCOA



VIZCAYA



BBVA & CX data merge

BBVA-RTI

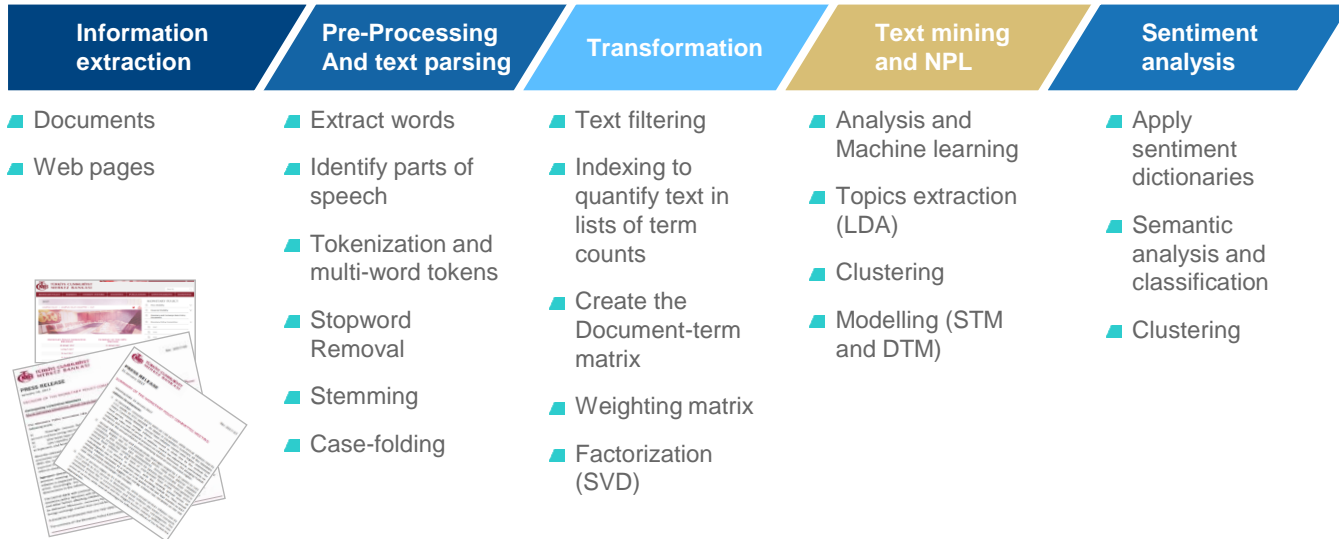
INE-RTI

04

Text Mining and Sentiment analysis

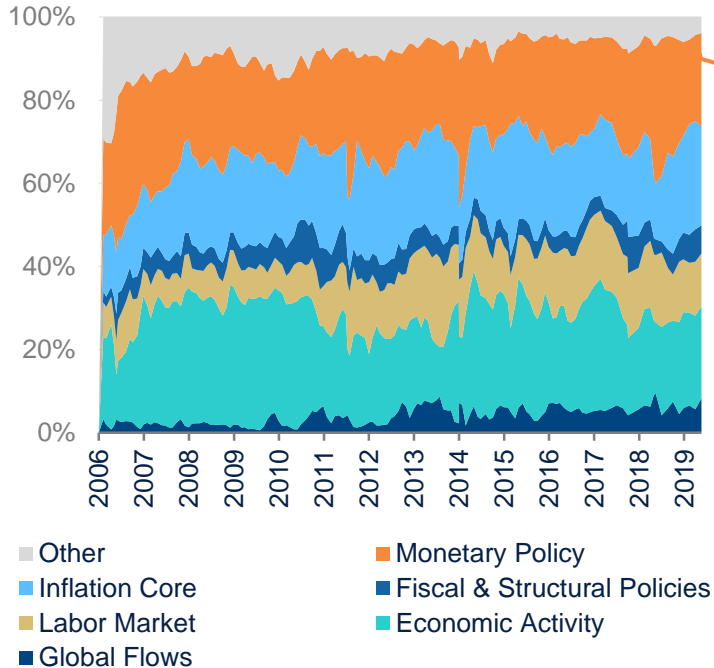
External databases: web scrapping and NPL techniques

Text mining makes information extraction from huge volumes of data easier and structures the information as important facts, key terms or persons.

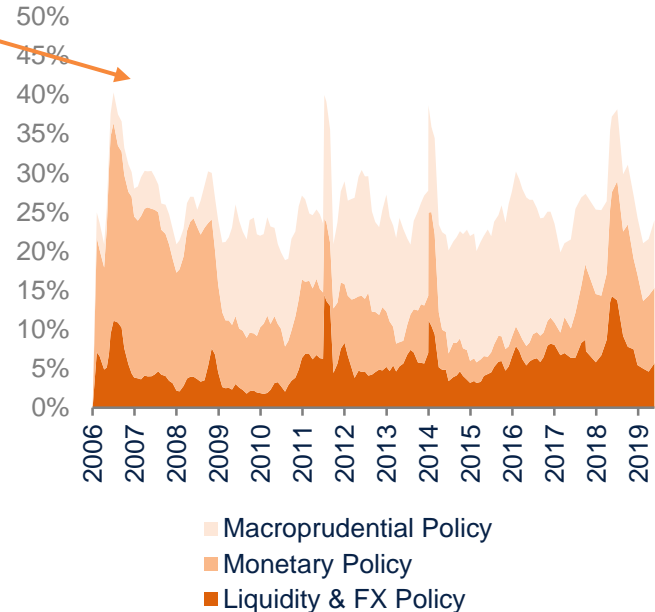


Disentangling communication policy by topic

CENTRAL BANK OF TURKEY: EVOLUTION OF TOPICS



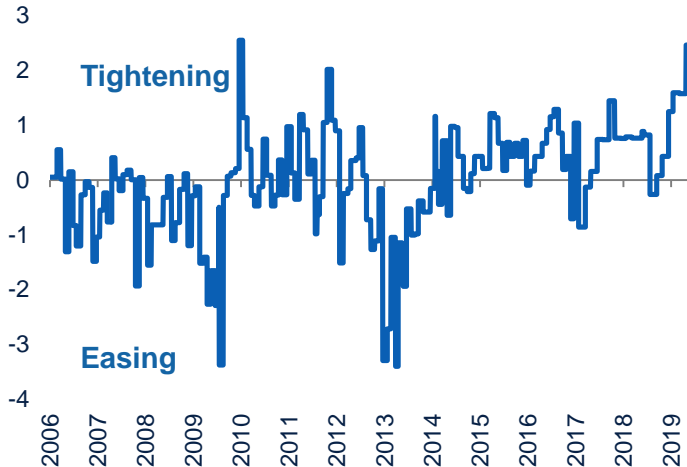
MONETARY POLICY TOPICS DISTRIBUTION (% OF TOTAL)



Through Sentiment Analysis we can check “how” the central bank is talking and assessing the monetary policy stance...

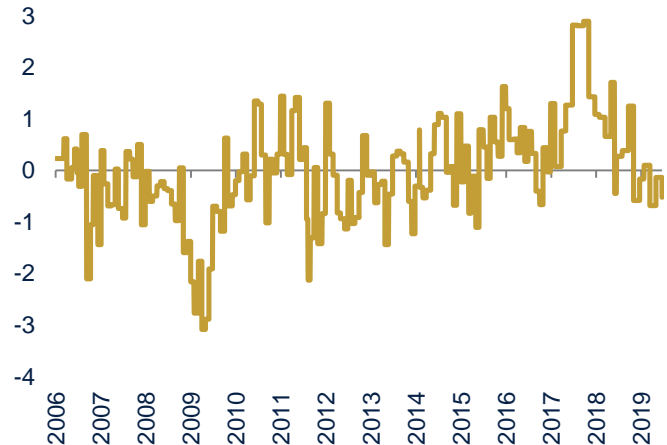
CENTRAL BANK OF TURKEY: MONETARY POLICY SENTIMENT (STANDARDIZED, ESTIMATED THROUGH BIG DATA LDA AND STM TECHNIQUES FROM MINUTES & STATEMENTS)

MONETARY POLICY “STATEMENTS”



A more formal Statement...

MONETARY POLICY “MINUTES”



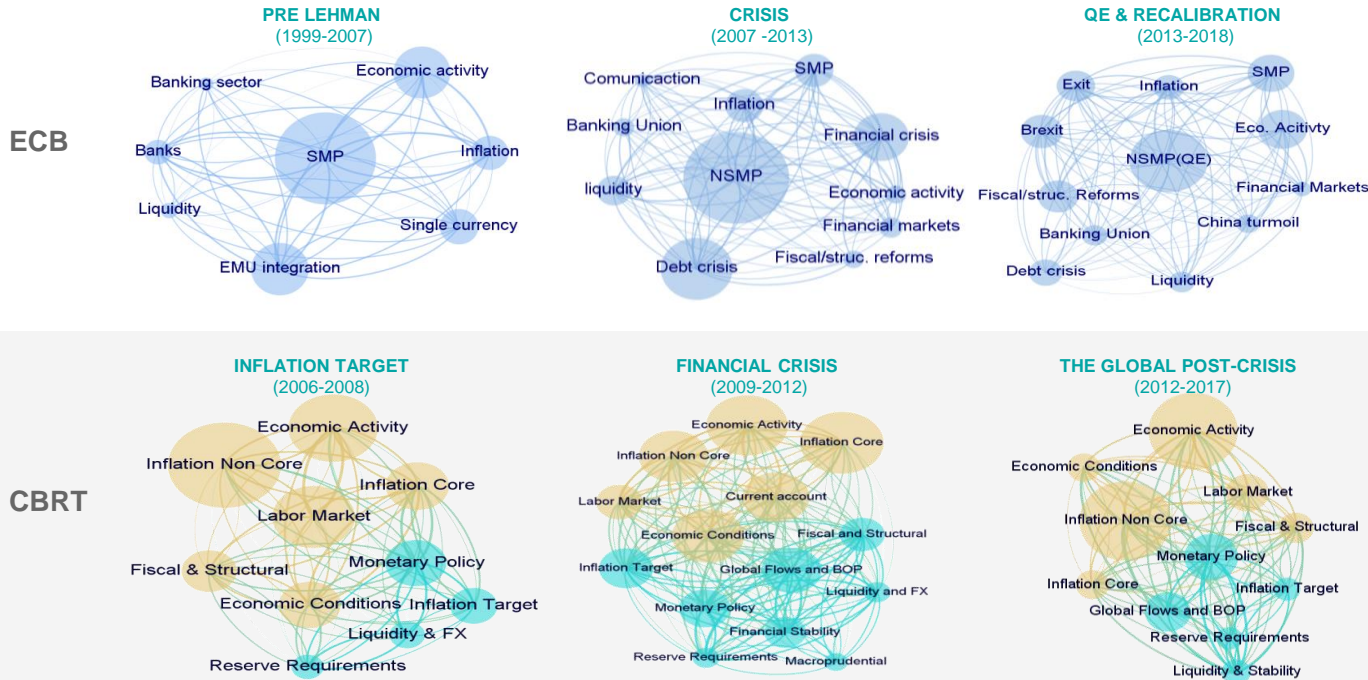
More extensive and analytical...

Source: Iglesias, J, Ortiz, A & Rodrigo, T (2017)

Source: BBVA Research

Understanding the Interconnectedness between Developed and Emerging Central Banks

MONETARY POLICY IN DEVELOPED ECONOMIES AND RESPONSE IN THE EMERGING MARKETS (NETWORKS)



Going one step further: we create our own dictionaries according to the Central Bank idiosyncratic narrative

Word2Vec methodology

Neural networks to learn word embedding and reconstruct linguistic contexts of words

Encodes semantic relationship among words by mapping vocabulary to vectors of numbers

Word vectors will be closed if they have similar meaning
i.e. increase and rise will be closer

Allows to add and subtract words as though they are capturing the meaning of a word

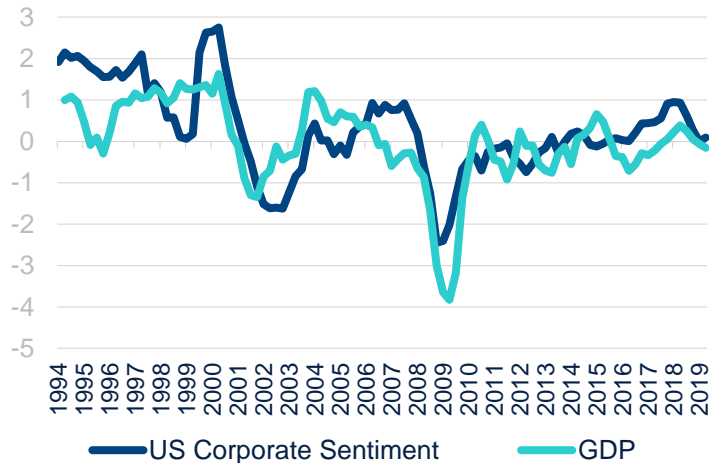
i.e. King—Man + Woman = Queen

Exploring the main filling reports by the US companies to the SEC

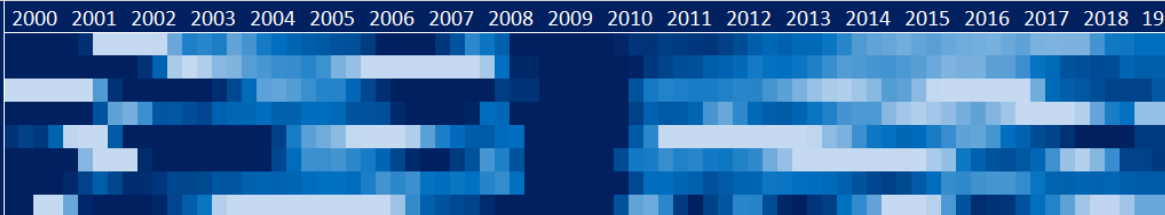
QUARTERLY AND ANNUAL REPORTS



BBVA RESEARCH US SENTIMENT INDEX BASED ON US CORPORATE REPORTS AND US GDP (STANDARDIZED, ESTIMATED THROUGH BIG DATA TECHNIQUES FROM 10Q & 10K REPORTS)



Finance
Insurance
Manufacturing
Retail Trade
Services
Transportation
Wholesale Trade
US Corporate Sentiment



You can find us at:



www.bbvaresearch.com



BBVA Research / Tomasa Rodrigo



@BBVAResearch / @Tomasarodrigo



Thank you!

The use of Big Data at BBVA Research

to better understand the society,
the economy and the world

Structured Products & Digitalization in
Private Banking

Geneva, Bloomberg Office

January 2020