Twitter Sentiment and Banks' Financial Ratios: Is There Any Causal Link?

Giuseppe Bruno
Bank of Italy

Paola Cerchiello
University of Pavia

Juri Marcucci[1]
Bank of Italy

Giancarlo Nicola
University of Pavia

Workshop on
"Harnessing Big Data & Machine Learning Technologies for Central Banks"
Bank of Italy
Rome, March 26–27, 2018

[1]The opinions expressed are those of the authors and do not reflect the views of the Bank of Italy, the ECB or the Eurosystem.

Market sentiment and asset prices

- There is a wide literature studying **how changes in investors' sentiment affect stock prices**
  - Fisher and Statman (2000), Baker and Wurgler (2006, 2007), Kumar and Lee (JF 2006), Tetlock (JF 2007), Huang et al (RFS 2015)
- **Market sentiment** or **investor attention** $\Rightarrow$ general prevailing attitude of investors
- This attitude is the accumulation of a variety of fundamental and technical factors
- Sentiment can be defined as
  - **optimism** vs **pessimism**
  - **bullish** vs **bear**
  - **animal spirits?**
- Market sentiment is usually considered as a *contrarian indicator*
- Market sentiment is used because it is believed to be a good predictor of market moves

Market sentiment and asset prices (Cont.)

- The literature connects results from behavioural finance, changes of investor attention on financial markets, and asset pricing
  - Barberis et al. (1998), Barberis & Thaler (2003), and Baker & Wurgler (2006, 2007).
- Two types of investors
  1. **rational arbitrageurs** who are sentiment-free
  2. **irrational traders** who are prone to exogenous sentiment.
- Behavioral patterns of retail investors have a significant impact on market returns.
- At least 5 approaches to measure investor attention:
  1. financial **market-based** measures (volume, VIX, TED spread, etc)
  2. **survey-based** sentiment indexes (e.g. Consumer Confidence index)
  3. **Internet search** behavior (Da et al., 2011)
  4. **non-economic factors** (news, weather, health conditions)
  5. textual sentiment data from **on-line resources** (like **Twitter**)

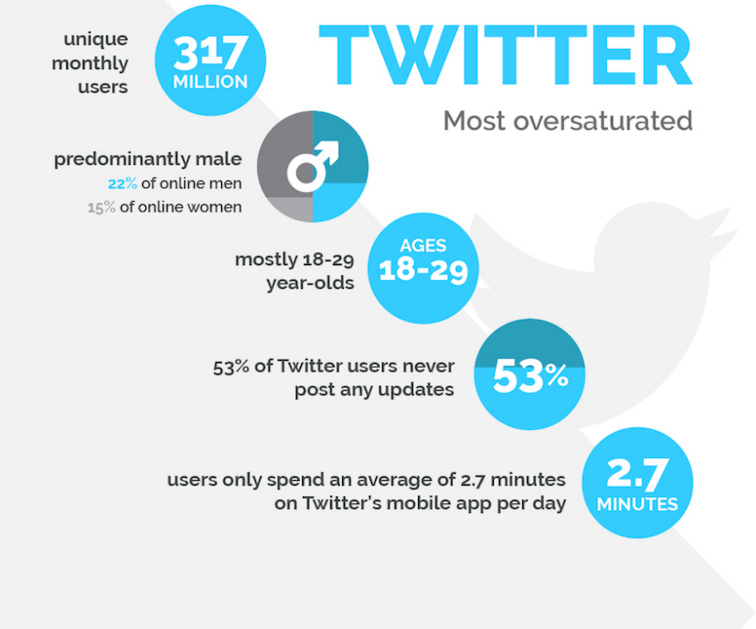
Related literature: Sentiment from social media and asset prices

- Bukovina (2016) reviews the literature on the **link between social media and capital markets**. Investors' sentiment or public mood is influential for asset pricing and capital market volatility
- Antweiler & Frank (2004), Das & Chen (2007), Tumarking & Whitelaw (2001): data from **internet message boards** (**Yahoo!Finance**) → mixed evidence.
- Azar and Lo (2016) show that **tweets** mentioning the FOMC around FOMC meetings do contain information to predict future returns
- Liew and Budavari (2017) show that features derived from social media (**StockTwits**) have power in explaining time-series variation of daily returns
- Liew and Wang (2016) investigate the relationship between the IPO's first-day returns and the sentiment extracted from **tweets** (iSENTIUM LLC)
- Souza and Aste (2016) study the DJIA components suggesting that social media (**tweets**) (PsychSignal) and stock markets have a nonlinear causal relationship.
- Plakandaras et al. (2015) show that investors' sentiment (**StockTwits**) has valuable information for future movements of 4 exchange rates
- Chen et al. (2014) show that peer-based advice extracted from user-generated opinions (**Seeking Alpha**) predict future stock returns and earning surprises.
- Sprenger et al. (2014a, 2014b), Sul, Dennis & Yuan (2014), Bollen et al. (2011), Karagozoglu & Fabozzi (2017), etc ... find **significant association** between **Twitter message features and market features**

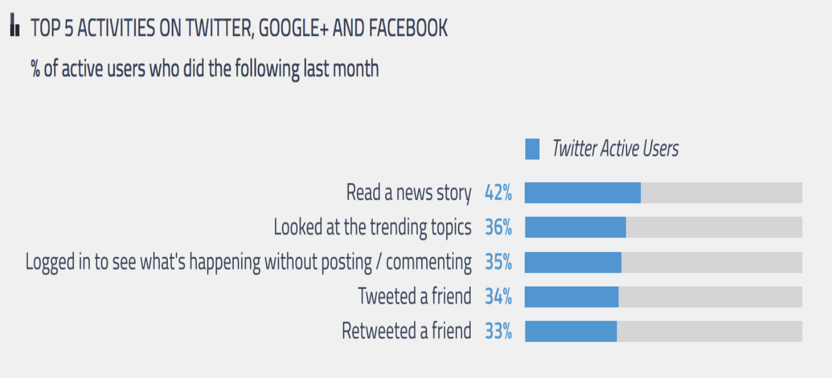
Our contribution and main findings

- We use **tweets in Italian** (from public and private APIs) on **four major Italian banks** (BMPS, UCG, ISP, UBI) and **Deutsche Bank** to extract **sentiment indicators** using a combination of **unsupervised techniques**
- We compute a simple and weighted average of sentiment on these banks and relate them to some banks' financial variables (**returns, volatility, volume, CDS** and **bond spreads**)
- We do find that **sentiment does Granger cause** some of the financial variables for some banks even after 5 business days
- We also find that the **volume of tweets** is another important indicator
- We find that sentiment is **helpful in predicting** financial variables
- In particular, these results are even stronger for the **most buzzed banks** (BMPS or DBK)

Why Twitter?


unique
monthly
users
317
MILLION
TWITTER
Most oversaturated
predominantly male
22% of online men
15% of online women
mostly 18-29
year-olds
AGES
18-29
53% of Twitter users never
post any updates
53%
users only spend an average of 2.7 minutes
on Twitter's mobile app per day
2.7
MINUTES

Why Twitter?



- 42% of Twitter users use Twitter to read news (news aggregator)

Is Twitter representative?

Data on Italy

- In **2016** in Italy there were **39 millions internet users** (**66%** of population, www.internetlivestats.com)
- In **2014** the number of **social network users** in Italy was **21.6 millions** (www.statista.com)
- In 2015 the popularity of Twitter and Google+ is marginal in Italy if compared to Facebook (www.digitalnewsreport.org).
- Around **22%** of internet users use Twitter in Italy;
- **10%** of Italians use Twitter weekly for searching news (similarly in the US and UK)
- Furthermore, Twitter is more often used by professionals and **highly-educated people** (e.g. bloggers, journalists, economists, etc.)

Twitter data

- We obtained tweets from the **public API** (Application Programming Interface) provided by Twitter/GNIP
- We collect all **tweets and retweets in Italian** from **all active Twitter accounts** which contain the following keywords:
-

Bank	**Ticker**	**Keywords**
Banca MPS	BMPS	'MPS', 'Banca Monte dei Paschi di Siena', 'Monte dei Paschi',
Unicredit	UCG	'Unicredit'
Intesa S.Paolo	ISP	'Intesa San Paolo', 'Banca Intesa'
UBI Banca	UBI	'UBI', 'UBI Banca', 'UBIBanca'
Deutsche Bank	DBK	'Deutsche Bank'

- 28 months of data: from **August 2015 to January 2018**

Descriptive statistics of tweets

Bank	**Ticker**	**# of total tweets**	**# of Retweets**	**# of tweets in Italian**	**# of tweets used**	**Average daily # of tweets used**
Banca MPS	BMPS	783,150	606,006	345,510	260,496	341
Unicredit	UCG	27,435	23,400	4,407	18,993	25
Intesa S.Paolo	ISP	14,249	12,708	807	11,901	16
UBI	UBI	7,696	3,541	1,689	1,852	2
Deutsche Bank	DBK	2,422,559	79,593	45,394	34,199	45

- For BMPS 77% of ReTweets (RT), UCG 85%, ISP 85%, UBI 42%, DBK 32%
- Pitfall in text analysis: "**MPS**" vs "**MPs**". We looked for "**Monte dei Paschi di Siena**" and we found UK "**Members of Parliament**" talking about **Brexit**!
- Same at the Bank of England: "**RBS**" vs "**RBs**" i.e. "**Royal Bank of Scotland**" vs "**Running Backs**"

Possible Pitfalls in Text Analysis?

While looking for Monte dei Paschi di Siena (MPS), we found Members or Parliament (MPs) discussing on Brexit

Possible Pitfalls in Text Analysis?

It also happened to the Bank of England during the Scottish referendum

Trying to predict a bank run, the Bank of England discovers the Minnesota Vikings

Minnesota Vikings running back Matt Asiata runs over New England Patriots cornerback Kyle Arrington during the second quarter at TCF Bank Stadium. Tweets about Vikings running backs got caught up in an analysis project by the Bank of England.

Example of a tweet and caveat

- Twitter is a web site where you can broadcast very short messages (max 280 characters) to anyone who is signed up to receive them. It's like a cross between a blog and a chat room. In Twitter you can find the wit and wisdom of millions of people. Twitter is just the first communications channel in history.

Example of tweet and caveat: Max 280 characters!

- Twitter is a web site where you can broadcast very short messages (max 280 characters) to anyone who is signed up to receive them. It's like a cross between a blog and a chat room. In Twitter you can find the wit and wisdom of millions of people. Twitter is just the first communications channel in history.

Example of tweet and caveat: Max 280 characters!

- *Twitter is a web site where you can broadcast very short messages (max 280 characters) to anyone who is signed up to receive them. It's like a cross between a blog and a chat room. In Twitter you can find the wit and wisdom* ~~of millions of people. Twitteris just the first communications channel in history.~~
- Therefore, it is difficult to analyze such a **short piece of text** which usually contains also
  - more than one hashtag (e.g. #MPS)
  - tiny urls (https://bloom.bg/2rDZVFy or https://wp.me/pMm6L-Dl2)
  - emoticons (e.g. :-) and similar)
  - etc.
- This makes the **extraction of sentiment** more difficult also because natural language processing (NLP) algorithms are very **well trained for English** but **not for Italian...**

Sentiment analysis with unsupervised methods

- Our sentiment analysis is based on a **dictionary-based approach** that **maps pre-assigned lists of positive and negative words to tweets**
- The final score is given by a function of positive and negative counts. (We used the *R* library **TextWiller**).
- *TextWiller* is based on a list of specific words in Italian with both **positive** and **negative** **polarities**.
- Current **limitations**: **neither negatives nor superlatives**. However, it is the best we can have for the Italian language.
- Our classifier is based on words and accounts for the relative quotas of positive and negative words in each tweet.

Sentiment analysis with unsupervised methods

- Only a fraction of tweets was selected for our sentiment analysis.
- First only tweets **in Italian** were retained.
- For the selection we employed an **unsupervised clustering** procedure based on **two steps**:
  1. 1st step:
     - **text vectorization**
     - **Latent Semantic Analysis (LSA)** (with dimension reduction)
     - *k-means* **clustering**
  2. 2nd step:
     - **Latent Dirichelet Allocation** (LDA) to investigate the main topics for each discussion and see how they change over time

Text Vectorization

- With **text vectorization** we represent documents in a vector space, creating a mapping from terms to term ids.
- We call them *terms* instead of *words* because they can be arbitrary n-grams not just single words.
- We represent a set of documents as a sparse matrix, where each row corresponds to a document and each column corresponds to a term.
- This is done using a vocabulary.
- We applied the **Bag of Words** (BoW) approach: a text is represented as an unordered collection of words, in which only counts for each tweet matter.
- We collect all word frequencies in a **Term Document Matrix (TDM)**
- Apply weights with TF-IDF (Term Frequency Inverse Document Frequency) algorithm.

Bag of Words example

- Suppose you have the **vocabulary**:
- $\{brown, dog, fox, jumped, lazy, over, quick, the, zebra\}$
- The sentence "*the quick brown fox jumped over the lazy dog*" could be encoded as:
- $< 1, 1, 1, 1, 1, 1, 1, 2, 0 >$
- The sentence "*the zebra jumped*" - even though it is shorter in length - would then be encoded as:
- $< 0, 0, 0, 1, 0, 0, 0, 1, 1 >$

Latent Semantic Analysis

- Latent semantic analysis (LSA) is a technique in natural language processing of analyzing relationships between a set of documents and the terms they contain by producing a set of **concepts**.
- LSA assumes that words that are close in meaning will occur in similar pieces of text (the distributional hypothesis).
- Then singular value decomposition (SVD) on TDM is used to reduce the dimensionality
- Words are then compared by taking the cosine of the angle between the two vectors (or the dot product between the normalizations of the two vectors) formed by any two rows.
- Values close to 1 represent very similar words while values close to 0 represent very dissimilar words.
- Then only singular values above a certain threshold are retained (**dimensionality reduction**) and the rest are set to 0 (like a factor model!)

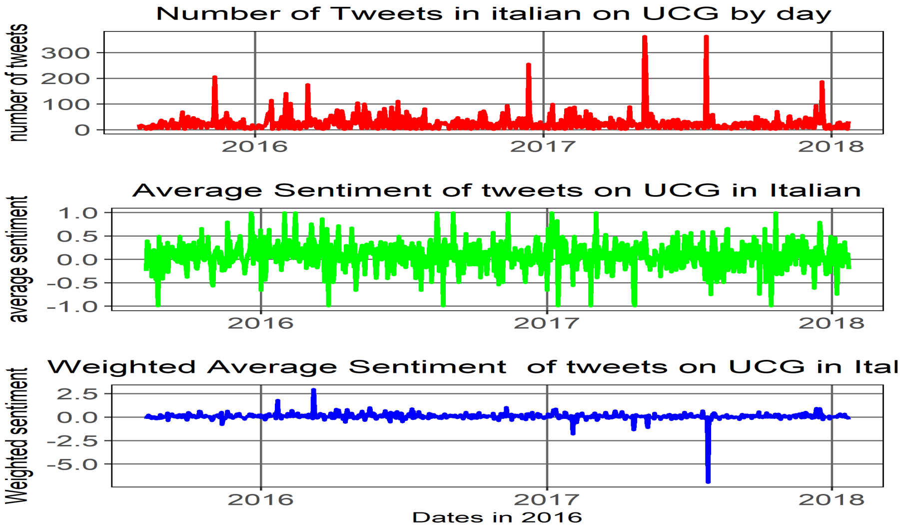
k-*means* clustering

- The reduced space is equipped with a norm which allows to evaluate the distance among documents
- To **group together similar documents**, we applied k-*means* clustering on the lower dimensional space
- k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- This results in a partitioning of the data space into Voronoi cells.
- The problem is computationally difficult (NP-hard); however, there are efficient algorithms (e.g. EM) that are commonly employed and converge quickly to a local optimum.

Latent Dirichelet Allocation (LDA)

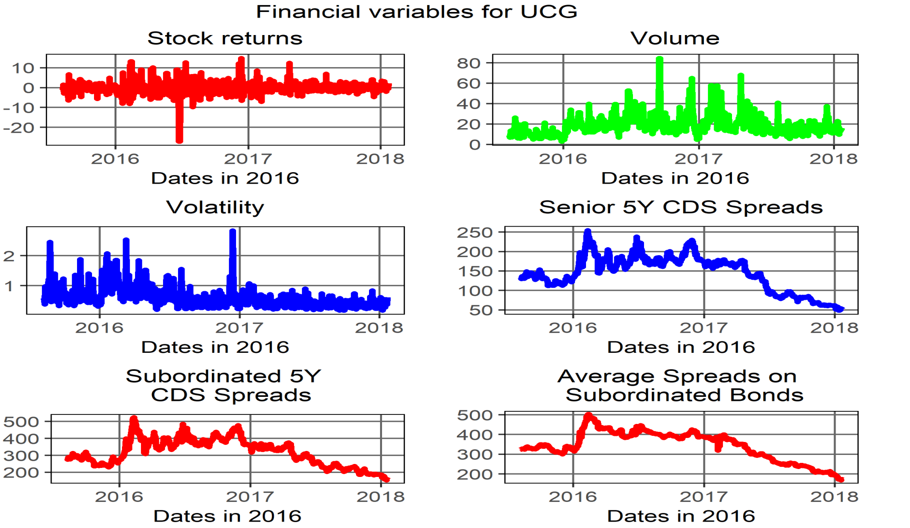
- In the second step Latent Dirichelet Allocation has been used to investigate the **main topics**.
- LDA is a generative statistical model that allows sets of observations to be **explained by unobserved groups** that explain why some parts of the data are similar.
- For example, if observations are words collected into documents, it posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics.
- LDA is an example of a topic model, i.e. a type of statistical model for discovering the abstract "topics" that occur in a collection of documents
- We can see how the main topics evolve over time for MPS

UCG: Twitter data

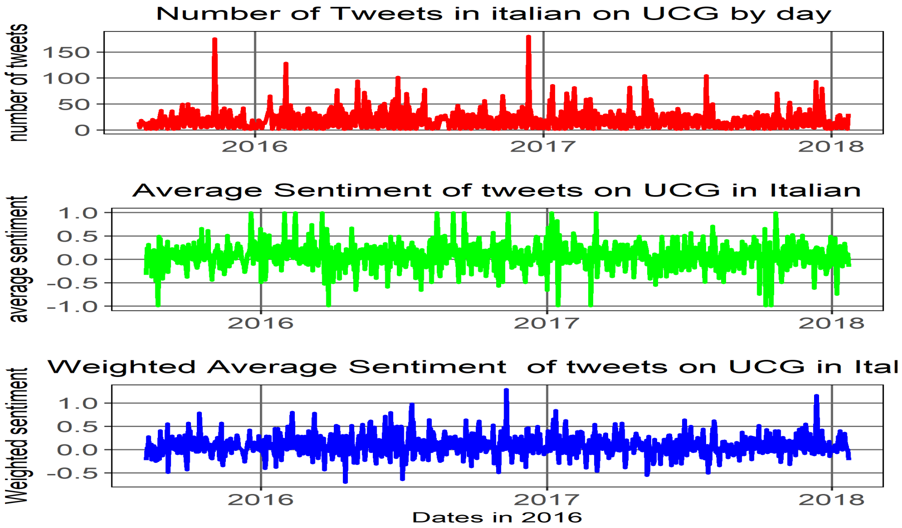


- Sentiment: -1=negative, +1=positive, 0=neutral
- simple average and weighted average (weights=ratio of tweets on each day and average number of tweets)

UCG: financial variables



UCG: Twitter data (no retweets)



- Sentiment: -1=negative, +1=positive, 0=neutral
- simple average and weighted average (weights=ratio of tweets on each day and average number of tweets)

Bank UCG: Correlations

Sample: 2015-08-07 - 2018-01-10, T = 840

	1	2	3	4	5	6	7	8	9
1 tweet_number	1	**-0.01**	**-0.13**	-0.02	0.17	-0.20	0.11	0.10	0.10
2 tweet_sent_mean		1	0.49	**0.19**	**0.02**	**0.02**	**0.03**	**0.03**	**0.05**
3 Weighted_sentiment			1	0.16	0.03	0.08	0.08	0.08	0.09
4 return				1	0.02	-0.04	-0.03	-0.04	0.00
5 Volume					1	0.36	0.36	0.38	0.31
6 Volatility						1	0.36	0.33	0.43
7 CDS Sen							1	0.98	0.95
8 CDS Sub								1	0.93
9 Sub Bond									1

Relationship between financial ratios and sentiment

Stationarity tests and Granger Causality tests

- Time series of financial ratios and sentiment (simple and weighted) are stationary
- Check if sentiment does Granger cause Italian banks' financial ratios
- The Granger causality principle is straightforward: if lagged values of X_t contribute to forecast current values of Y_t in a forecast achieved with lagged values of both X_t and Y_t then we say X *Granger causes* Y_t.
- $$y_t = \mu + \sum_{i=1}^{L} \alpha_i \cdot y_{t-i} + \sum_{i=1}^{L} \beta_i \cdot x_{t-i} + \varepsilon_t \qquad (1)$$
- The null hypothesis is: $H_0 : \beta_1 = \beta_2 = \cdots = \beta_L = 0$.
- Up to 5 lags (daily data)

Granger causality tests with const. and trend - $sent_{ts}$

Bank		Variable → Twitter Sentiment					Variable ← Twitter Sentiment				
	Lags	1	2	3	4	5	1	2	3	4	5
BMPS											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	**	**	**
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
ISP											
	r_t	*	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	*	-	-	-	-	-	-	-
UCG											
	r_t	**	**	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	**	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
DBK											
	r_t	-	-	*	***	**	-	-	-	-	**
	VV_t	-	-	-	-	-	*	-	**	**	***
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	*	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	*	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-

Granger causality tests with const. and trend - $sent_{we}$

Bank		Variable → Twitter Sentiment					Variable ← Twitter Sentiment				
	Lags	1	2		4	5	1	2	3	4	5
BMPS											
	r_t	-	-	-	-	-	**	-	**	**	*
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	**	**	**	-	**
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	**
	CDS_t^{Sub}	**	-	-	-	-	*	*	**	***	**
	$Bond_t^{Sub}$	-	*	*	**	**	-	-	-	-	-
ISP											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
UCG											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	**	**	**	*	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	*	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
DBK											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	***	**	***	**	*
	Vol_t	-	-	-	-	-	***	***	**	**	*
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	**
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	**	***

Granger tests with const. and trend - $sent_{ts}$ (No retweets)

Bank		Variable → Twitter Sentiment					Variable ← Twitter Sentiment				
	Lags	1	2	3	4	5	1	2	3	4	5
* BMPS											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	*	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
ISP											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	*	*	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	-
UCG											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	*	*	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	-	-	-	-	-	*	*
	$Bond_t^{Sub}$	-	*	-	-	-	-	-	-	-	-
DBK											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	*	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	*	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	*	-	-	-	-	-	-	-	-	-
	$Bond_t^{Sub}$	*	*	-	-	-	-	-	-	-	-
UBI											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	*									

Granger tests with const. and trend - $sent_{we}$ (No retweets)

Bank		Variable → Twitter Sentiment					Variable ← Twitter Sentiment				
	Lags	1	2	3	4	5	1	2	3	4	5
* BMPS											
	r_t	-	-	-	-	-	***	***	***	***	***
	VV_t	-	-	-	-	-	***	***	***	***	***
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	*	-	-	*	**
	CDS_t^{Sub}	-	-	-	-	-	***	***	***	***	***
	$Bond_t^{Sub}$	-	-	*	**	**	-	-	-	-	-
ISP											
	r_t	-	-	-	*	*	-	-	-	*	*
	VV_t	*	*	*	*	*	-	-	-	-	-
	Vol_t	*	**	**	*	*	-	-	-	-	-
	CDS_t^{Sen}	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sub}	-	-	-	*	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	**	**	**	**	*
UCG											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	-	-	-	-	-	-	-	*	*	*
	CDS_t^{Sen}	*	**	**	**	**	-	**	**	*	*
	CDS_t^{Sub}	**	**	**	**	**	-	*	**	***	***
	$Bond_t^{Sub}$	*	**	**	**	**	-	-	-	-	-
DBK											
	r_t	-	-	-	-	-	*	**	**	***	***
	VV_t	-	-	-	-	-	-	-	***	**	***
	Vol_t	-	-	-	-	-	-	-	-	-	-
	CDS_t^{Sen}	*	*	-	*	*	-	-	-	-	-
	CDS_t^{Sub}	*	*	-	*	-	-	-	-	-	-
	$Bond_t^{Sub}$	-	-	-	-	-	-	-	-	-	***
UBI											
	r_t	-	-	-	-	-	-	-	-	-	-
	VV_t	-	-	-	-	-	-	-	-	-	-
	Vol_t	*	*	**	*	*	-	-	-	-	-
	CDS_t^{Sen}										

Regression with Twitter Sentiment (simple) and Volume

BMPS	Intercept	Sentiment	F-stat	Intercept	Sentiment	TW_vol	F-stat
Stock Return	-0.004	0.033*	6.01	0.001	0.029*	-0.000*	5.45
Volume	-0.002	0.041***	17.71	-0.001	0.038***	0.000	8.96
Volatility	7.097E5***	3.574E4	0.08	6.124E5***	8.733E4	4.030E2***	10.67
CDS_spread_sen	2.422***	-0.578	0.89	2.433***	-0.584	-0.000	0.45
CDS_spread_sub	4.557E2***	4.179E1	2.25	4.637E2***	3.753E1	-0.033	2.51
Bond_spread_sub	1.338E3***	-5.646E1	0.83	1.285E3***	-2.861E1	0.217***	13.89
BNL/BNP Paribas	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	1.100E-3	-0.000	0.01	0.000	-0.000	1.193E-5	0.06
Volume	4.547E6***	2.484E4	0.00	4.591E6***	3.407E4	-2.093E3	0.11
Volatility	1.285***	-0.011	0.02	1.279***	-0.012	0.000	0.043
CDS_spread_sen	7.793E1***	1.065	0.63	7.780E1***	1.039	0.006	0.41
CDS_spread_sub	1.676E2***	1.487	0.37	1.677E2***	1.495	-0.002	0.19
Bond_spread_sub	1.654E2***	9.309	1.27	1.637E2***	8.953	0.080	1.09
ISP	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.003	0.009	2.38	-0.001	0.009	-0.000	2.50
Volume	1.446E8***	1.254E7	0.47	1.231E8***	-1.502E7	6.412E5**	4.675
Volatility	-0.074***	0.001	0.02	0.068***	0.000	0.000	2.353
CDS_spread_sen	1.365E2***	-5.769*	4.44	1.327E2***	-6.210*	0.114***	8.798
CDS_spread_sub	2.818***	-1.174E1*	4.00	2.746E2***	-1.257E1*	0.214**	6.93
Bond_spread_sub	3.751E2***	-7.234	2.69	3.680E2***	-8.056	0.213***	10.27
DBK	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.003	-0.007	2.13	-0.002	-0.007	-0.000	1.16
Volume	1.436E7***	-9.160E5	0.21	1.160E7***	-8.154E5	5.851E4***	43.5
Volatility	0.524***	0.008	0.02	0.469***	0.010	0.001***	19.22
CDS_spread_sen	2.043E2***	-4.018	0.96	2.010E2***	-3.900	0.071***	11.77
CDS_spread_sub	4.117E2***	-9.816	1.69	4.055E2***	-9.589	0.132***	12.54
Bond_spread_sub	4.197E2***	-1.431E1	2.62	4.151E2***	-1.414E1	0.098**	5.695
UCG	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.008*	0.029*	4.95	-0.010*	0.030*	0.000	2.69
Volume	1.158E8***	-1.760E7	1.15	1.048E8***	-1.288E7	1.057E5**	4.09
Volatility	0.122***	-0.015	0.80	0.105***	-0.008	0.000***	8.66
CDS_spread_sen	1.809E2***	-3.355	0.57	1.802E2***	3.057	0.007	0.47
CDS_spread_sub	3.894E2***	-4.491	0.28	3.902E2***	-4.821	-0.007	0.20
Bond_spread_sub	4.824E2***	-3.319E1	2.77	4.815E2***	-3.282E1	0.008	1.39

Main findings

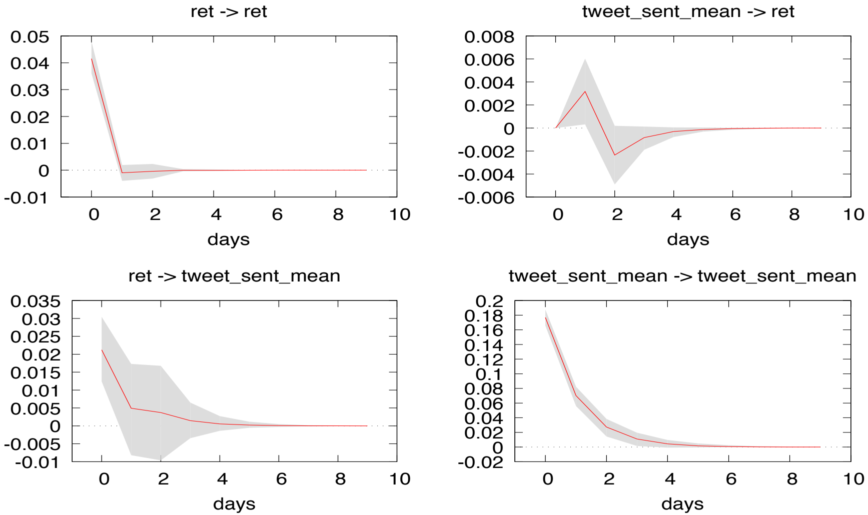
- Using both a simple average of Twitter sentiment or a weighted average, we find that **Twitter sentiment does Granger cause** some Italian banks' financial ratios even at longer lags (up to 5)
- In line with expectations, we find that financial variables do Granger cause Twitter sentiment (particular news generate buzz on social media)
- Results are robust across different specifications of the test with higher significance for the more buzzed banks (e.g. BMPS and DBK).
- In our regression analysis we notice that Twitter sentiment positively affects the stock returns and volume of traded stocks, while it seems to negatively affect the CDS spreads
- Twitter volume instead seems to negatively affect returns, and positively volume, volatility and CDS and bond spreads.
- We also find that sentiment has some predictive power for banks' financial variables

Conclusion and next steps (work in progress)

- We have confirmed the importance of social media sentiment for the financial variables of some Italian banks and DBK
- We have suggested how to extract sentiment indicators with unsupervised methods from tweets written in Italian
- We have shown that Twitter sentiment and volume are important to determine some banks' financial variables
- With respect to the previous literature, we have extended the link between asset pricing and sentiment to bond and CDS spreads
- We plan to extend the analysis to other major European and US banks for which we will use more standard techniques developed for English texts.
- We will also examine tweets in English related to Italian banks, because traders and investors 'talk' in English!
- We will also extend the sample possibly to start in 2009/09

Thank you for your attention!

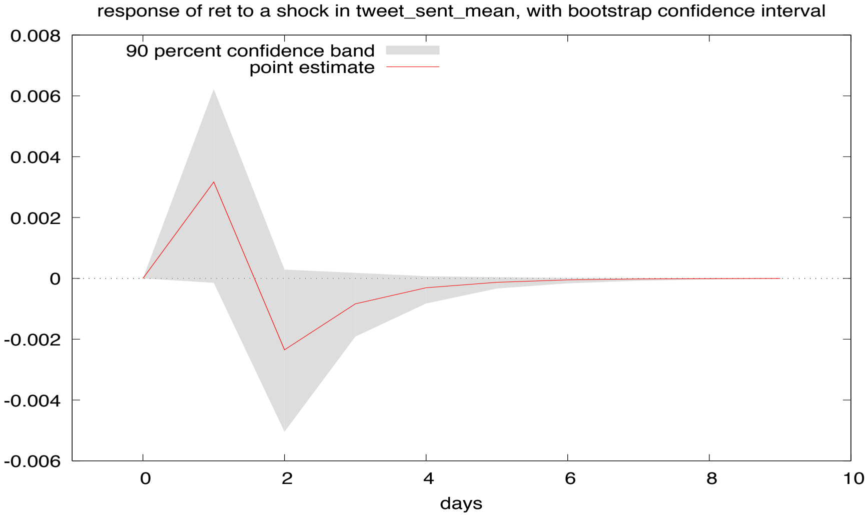
BMPS: IRFs for VAR: return and sentiment with 2 lags



- Estimation sample: 4/1/2015-10/18/2016.

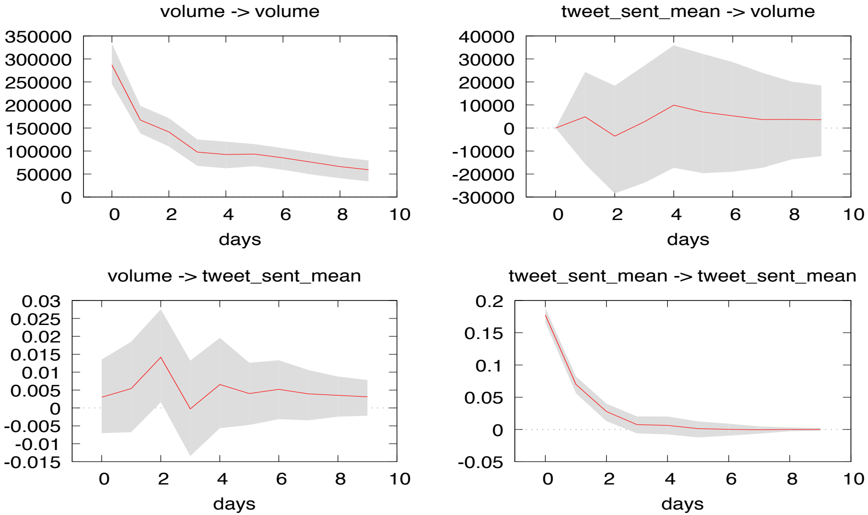
BMPS: IRFs for VAR: return and sentiment with 2 lags

Response of return on sentiment



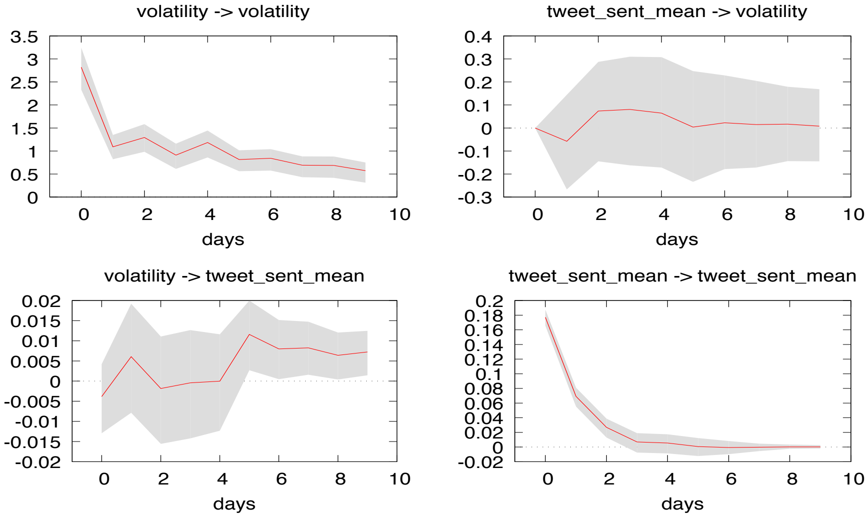
- Estimation sample: 4/1/2015-10/18/2016.

BMPS: IRFs for VAR: volume and sentiment



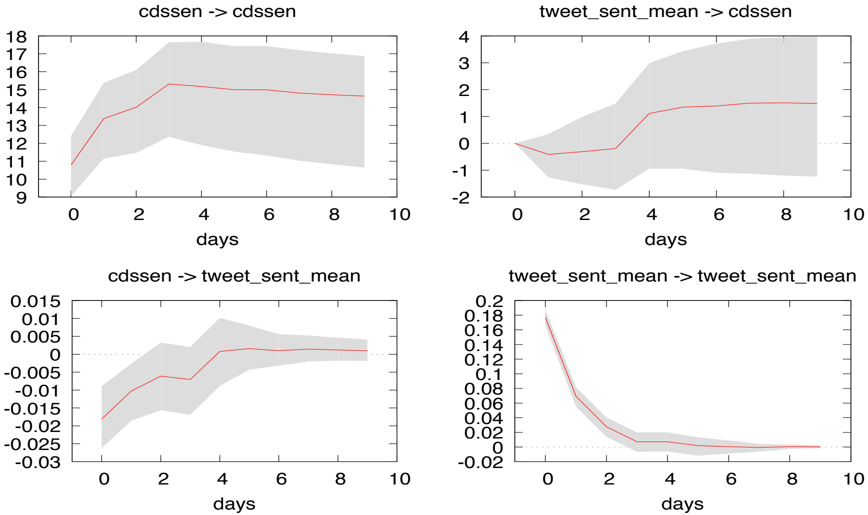
- Estimation sample: 4/1/2015-10/18/2016.

BMPS: IRFs for VAR: volatility and sentiment



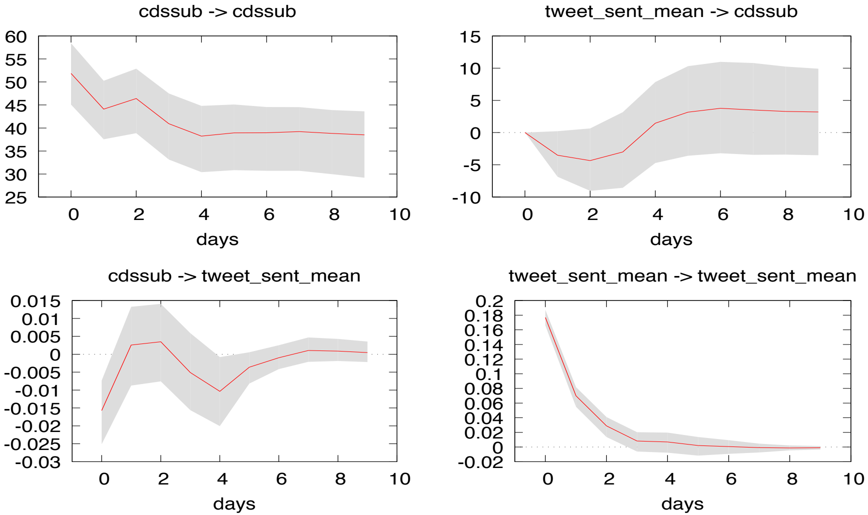
- Estimation sample: 4/1/2015-10/18/2016.

BMPS: IRFs for VAR: CDS senior and sentiment



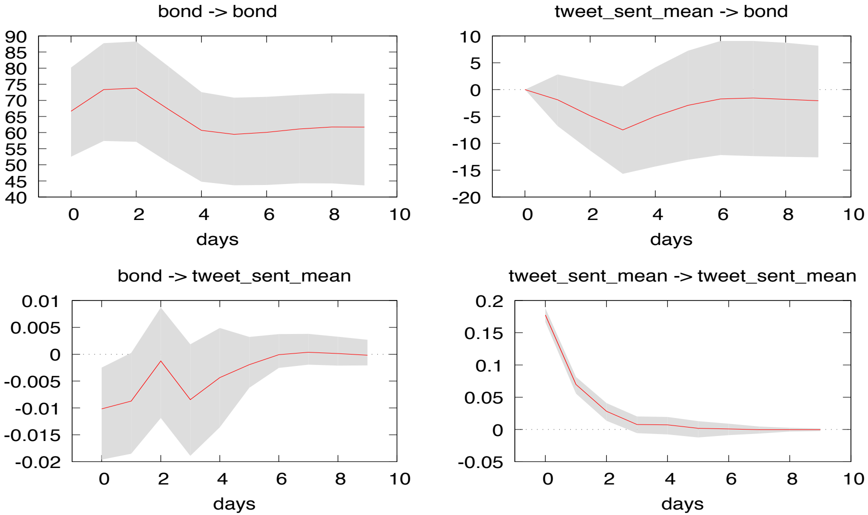
- Estimation sample: 4/1/2015-10/18/2016.

BMPS: IRFs for VAR: CDS sub. and sentiment



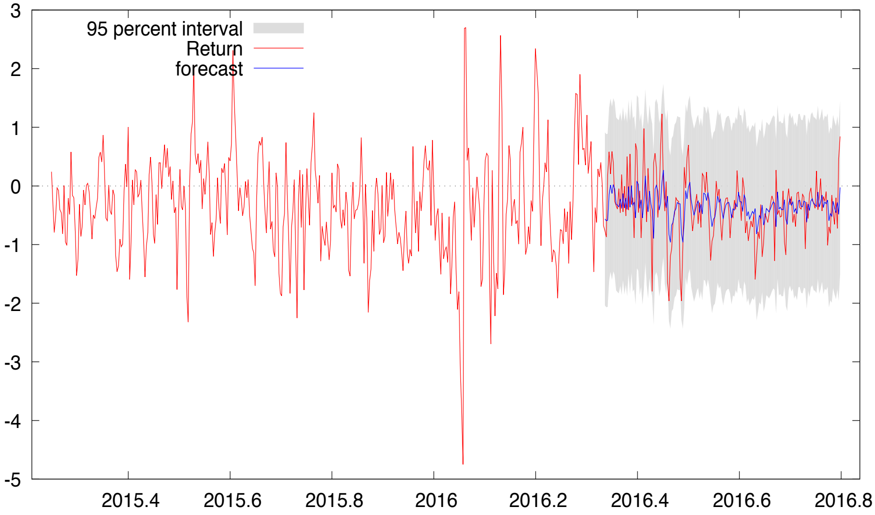
- Estimation sample: 4/1/2015-10/18/2016.

BMPS: IRFs for VAR: bond and sentiment



- Estimation sample: 4/1/2015-10/18/2016.

BMPS: Forecasting with a VAR: return and sentiment



- In-sample: 4/1/2015-5/1/2016. Out-of-sample: 5/2/2016-10/18/2016
- RMSE: 0.49498

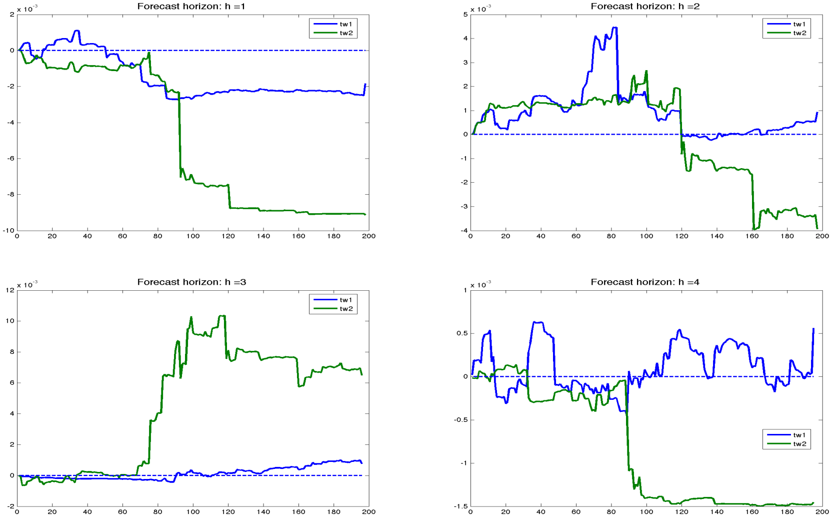
BMPS: Forecasting with AR-X: return and sentiment

Variable	1	2	3	4	5	6	7	8	9	10	11	12
$AR(p)$ *(RMSE)*	0.042	0.043	0.043	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.041	0.041
$TW1_t$	1.003	0.999	0.999	0.999	1.001	1.003	0.986	0.978	0.994	1.001	1.004	0.998
$TW2_t$	1.013+	1.005	0.991	1.002	1.001	1.001	1.023	1.000	1.000	1.005++	1.000	1.001

- In-sample: 4/1/2015-5/1/2016. Out-of-sample: 5/2/2016-10/18/2016

BMPS: OOS Evaluation - CSSED

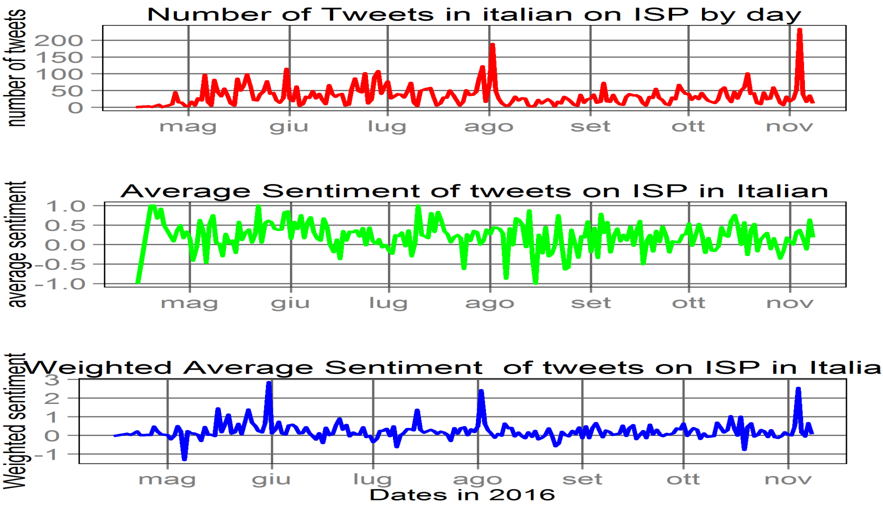
$$CSSED_{m,\tau} = \sum_{\tau=R}^{T}(\hat{e}^2_{bm,\tau} - \hat{e}^2_{m,\tau})$$



Regression with Twitter Sentiment (weighted) and Volume

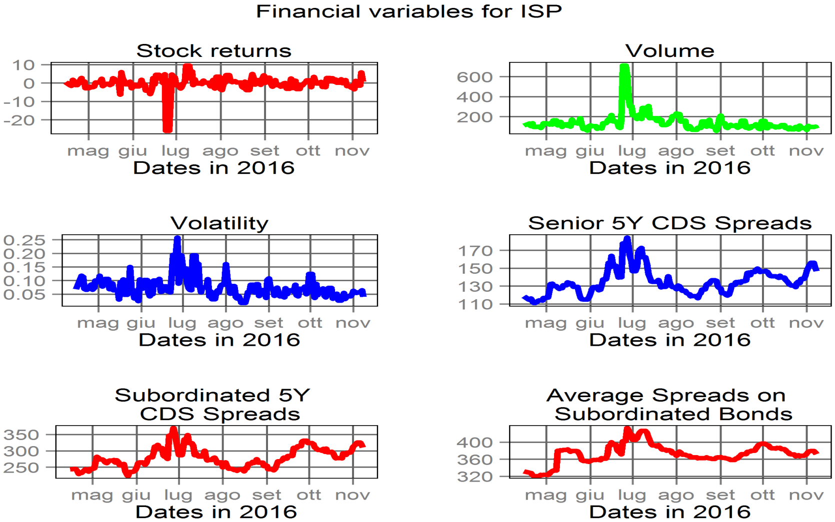
BMPS	Intercept	Sentiment	F-stat	Intercept	Sentiment	TW_vol	F-stat
Stock Return	-0.002	0.041***	17.71	-0.001	0.039***	-0.000	8.96
Volume	6.095E5	-1.328E5***	1.89	6.066E5***	8.255E4	4.338E2***	10.73
Volatility	2.434***	-0.264	0.32	2.466***	-0.341	-0.000	0.21
CDS_spread_sen	4.561E2***	3.347E1	2.49	4.616E2***	2.012E1	-0.027	1.95
CDS_spread_sub	1.337E3***	-4.968E1	1.11	1.285E3***	7.598E1	-0.025***	15.09
Bond_spread_sub	1.683E3***	-1.897E2	1.13	1.517E3***	2.106E2	0.807***	10.45
BNL/BNP Paribas	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	0.001	0.001	0.07	0.001	0.000	0.000	0.06
Volume	4.609E6***	-2.023E5	0.57	4.606E6***	2.147E5	3.095E2	0.28
Volatility	1.272***	0.035	0.41	1.275***	0.046	-0.000	0.23
CDS_spread_sen	7.806E1***	0.450	0.30	7.803E1***	0.343	0.003	0.16
CDS_spread_sub	1.682E2***	-0.925	0.37	1.681E2***	-1.583	0.017	0.31
Bond_spread_sub	1.666E2***	3.510	0.48	1.657E2***	0.231	0.082	0.50
ISP	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.001	0.002	0.14	0.003	0.008	-0.000	2.21
Volume	1.450E8***	-1.329E7	0.81	1.176E8***	-5.128E7**	1.045E6***	9.15
Volatility	0.076***	-0.006	1.01	0.067***	-0.019	0.000*	5.94
CDS_spread_sen	1.359E2***	-2.459	1.23	1.311E2***	-9.164***	0.185***	13.09
CDS_spread_sub	2.806E2***	-5.289*	1.24	2.713E2***	-1.810E1*	0.352**	10.29
Bond_spread_sub	3.734E2***	0.744	0.04	3.659E2***	-9.657*	0.286***	11.44
DBK	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.000	0.003	1.70	-0.001	0.008	0.000	1.50
Volume	1.288E7***	-7.315E5***	61.64	1.185E7***	-1.467E5	4.997E4***	43.95
Volatility	0.492***	-0.129***	22.12	0.466***	0.021	0.001***	19.32
CDS_spread_sen	2.043E2***	-7.052**	10.55	2.017E2***	3.872	0.093***	11.81
CDS_spread_sub	4.108E2***	-13.591***	11.57	4.073E2***	5.925	0.168***	11.94
Bond_spread_sub	4.214E2***	-6.755	1.98	4.176E2***	-1.489E1	0.185**	5.971
UCG	**Intercept**	**Sentiment**	**F-stat**	**Intercept**	**Sentiment**	**TW_vol**	**F-stat**
Stock Return	-0.007*	0.035**	10.50	-0.007	0.035**	0.000*	5.22
Volume	1.123E8***	1.506E7	0.01	1.023E8***	1.084E7	1.053E5**	4.10
Volatility	0.120***	0.001	0.01	0.104***	-0.006	0.000***	8.62
CDS_spread_sen	1.801E2***	4.325	1.39	1.795E2***	4.082	6.063	0.85
CDS_spread_sub	3.880E2***	9.252	1.77	3.889E2***	9.633	-0.010	0.99
Bond_spread_sub	4.782E2***	1.932	0.01	4.766E2***	1.259	0.017	0.06

ISP: Twitter data

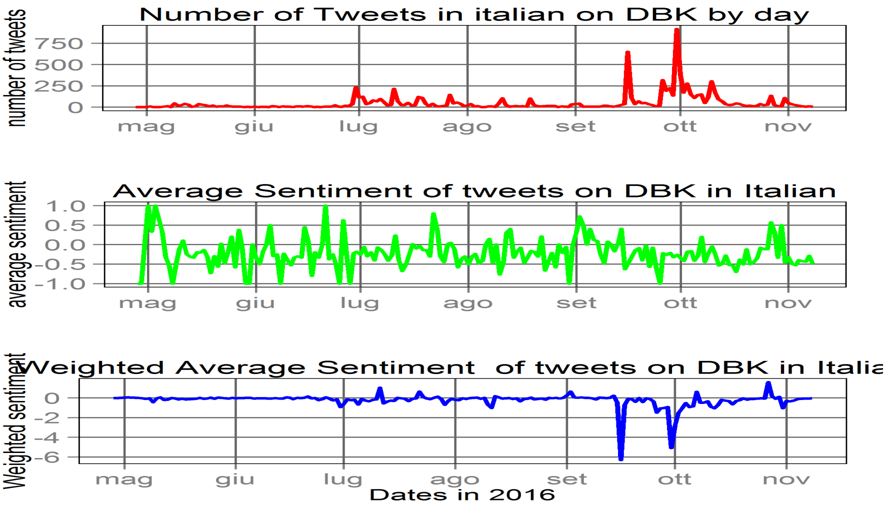


- Sentiment: -1=negative, +1=positive, 0=neutral
- simple average and weighted average (weights=ratio of tweets on each day and average number of tweets)

ISP: financial variables



DBK: Twitter data



- Sentiment: -1=negative, +1=positive, 0=neutral
- simple average and weighted average (weights=ratio of tweets on each day and average number of tweets)

DBK: financial variables