



LISTENING TO THE BUZZ: SOCIAL MEDIA SENTIMENT AND RETAIL DEPOSITORS' TRUST

Matteo Accornero, Mirko Moscatelli



MOTIVATION

1. Exploit the power of big data, providing both opportunities and challenges:
 - Heterogeneity of agents
 - Timeliness
 - Large amount of structured and unstructured data
 - Large amount of noise
2. Analyze the predictive power of measures based on social networking services in nowcasting financial indicators.
3. Create an early-warning indicator targeted at evaluating retail depositors' level of trust.



1. Can big data enhance forecasts of in-outflows of retail deposits?
2. Can big data help in detecting contagion dynamics at work in the retail deposits market?

1. Contributes to the empirical literature concerning:
 - i. the usefulness of social networking services data for the nowcasting of bank distress and financial indicators
 - ii. identification of informational contagion dynamics



DATA

TWITTER



- Twitter is a social networking service, an on line service enabling users to publish short messages and read other users' messages.
- Every second on average around 6000 tweets are tweeted, corresponding to over 500 million tweets per day.

TWITTER: OUR DATA



- Data provided by GNIP Inc.
- Period: 1st April 2015 – 30th April 2016
- First 100 Italian banks in terms of retail deposits
- More than 500.000 tweets



- Aggregation at banking group level
- Dropped foreign banks
- Dropped banks having on average less than 10 tweets per month

SENTIMENT ANALYSIS

- *Dictionary based:*
 - *Top-Down approach: pre-defined dictionary*
 - *Bottom-Up approach: machine learning-based dictionary definition*
- *Our choice: Top-Down approach with customized dictionary tailored on financial topics*

SENTIMENT ANALYSIS

- "**Sit-in** davanti Banca Etruria Civitavecchia: **Protesta** organizzata dal Codacons, gente grida '**vergogna**' "
- " Ferrara: Banca Carife, **indagato** per **truffa** Grassano, ex dg. Perquisita la sede della Cassa #crisitaly "
- " UBI Banca Lavora con noi: posizioni aperte - "
- " Invece si occupa dell'azienda del Papà di renzie, si occupa della banca etruria del papà del ministro boschi "

SENTIMENT INDICATOR

$$S_{b,t} = \frac{S'_{b,t} - E[S'_b]}{STD[S'_b]}$$

- $S'_{b,t}$ = Number of negative words in tweets regarding bank b in month t
- $E[S'_b]$ = Mean of $\{S'_{b,1}, \dots, S'_{b,n}\}$
- $STD[S'_b]$ = Standard deviation of $\{S'_{b,1}, \dots, S'_{b,n}\}$

SENTIMENT INDICATOR (EXTRA)

- $S'_{b,t}$ = Number of negative words in tweets regarding bank b in month t
- In other words, $S'_{b,t} = T'_{b,t} \cdot Q'_{t,b}$ where:
 - $T'_{b,t}$ = Number of tweets regarding bank b in month t
 - $Q'_{b,t}$ = Number of negative words per tweet on tweets regarding bank b in month t

INTERCONNECTION INDICATOR

$$I_{b,t} = \sum_{i \neq b} (P_{i,b} \cdot S_{i,t})$$

- $S_{i,t}$ = Sentiment indicator for bank i in month t .
- $P_{i,b}$ = Ratio of tweets regarding bank b that regards also bank i , i.e. the percentage of co-occurrences of the two banks.

FULL SET OF VARIABLES

Acronym	Definition	Frequency
ret_dep_gro_tot	Monthly growth rate of retail deposits	M
ret_dep_gro_ins	Monthly growth rate of insured retail deposits	M
ret_dep_gro_unins	Monthly growth rate of uninsured retail deposits	M
int_rat_spre	Spread interest rates on dep. and gov. bonds (1-3 y)	M
log_tot_asset	Logarithm of total assets (in millions)	Q
t1ratio	Tier 1 capital on risk weighted assets	Q
bad_loan_rat	Rate of new quarterly bad loans on total loans	Q
roa	Operating profits on total assets	Q
ci_rat	Operating costs on operating profits	Q
liq_asset	Liquid funds (cash, ST treasury bonds, demand and overnight bank deposits) on total assets	M
whs_fun	Wholesale funding on total funding	Q
sen_sco	Standardized sentiment score	M
inter_ind	Interconnection indicator	M
tweet_std	Standardized number of monthly tweets	M
neg_ratio	Negative terms divided by number of terms	M



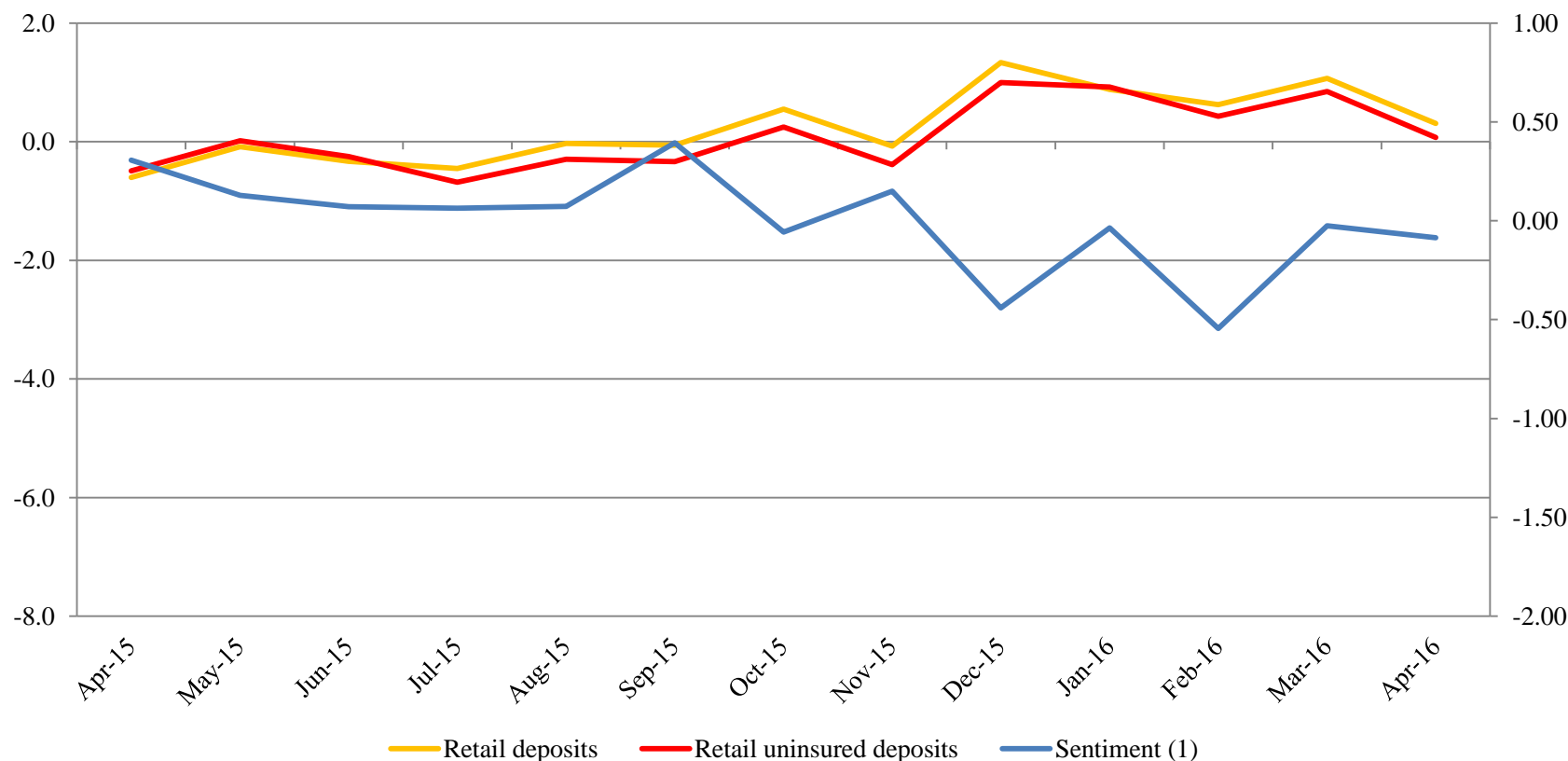
DESCRIPTIVE STATISTICS

ITALIAN BANKING SYSTEM FUNDING

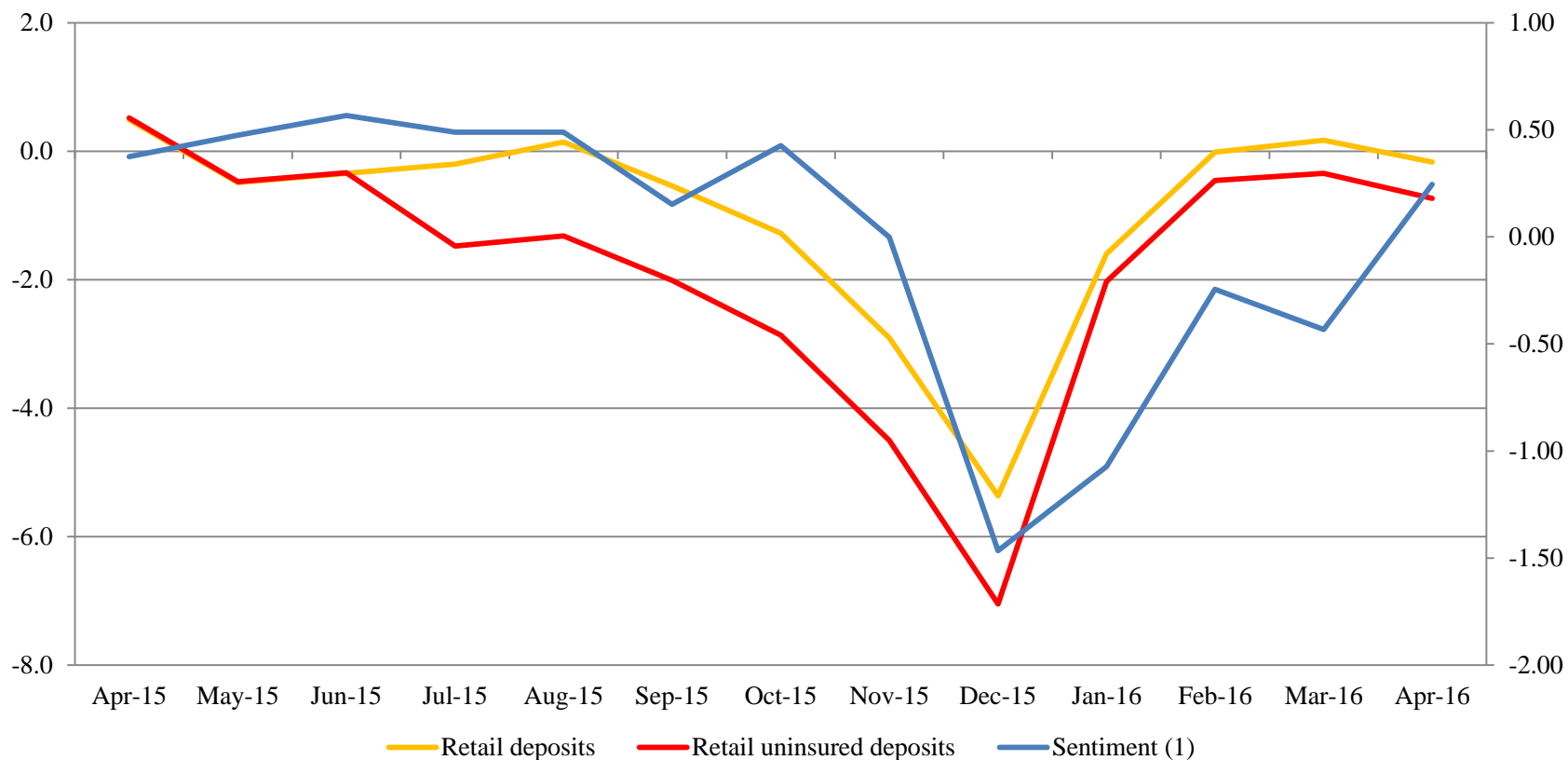
(PERCENTAGE VALUES)

Instrument	2008	2011	2015Q3
Deposits from residents (excluding banks)	49.8	47.5	59.2
<i>of which: insured deposits</i>	35.5	33.9	40.7
<i>of which: uninsured deposits</i>	14.3	13.6	18.5
Bonds held by retail investors	15.1	15.9	9.5
Bonds held by wholesale investors	11.3	9.7	8.4
Other deposits	21.1	16.0	13.4
Liabilities against CCPs	0.4	2.2	2.5
Eurosystem refinancing	2.3	8.7	7.1

THE DYNAMICS OF SENTIMENT AND FUNDING FOR 'NON-DISTRESSED' BANKS

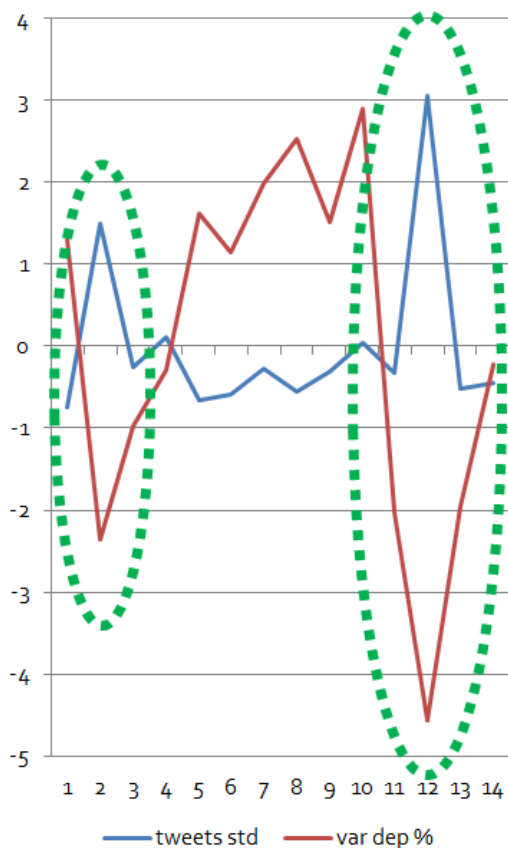


THE DYNAMICS OF SENTIMENT AND FUNDING FOR 'DISTRESSED' BANKS

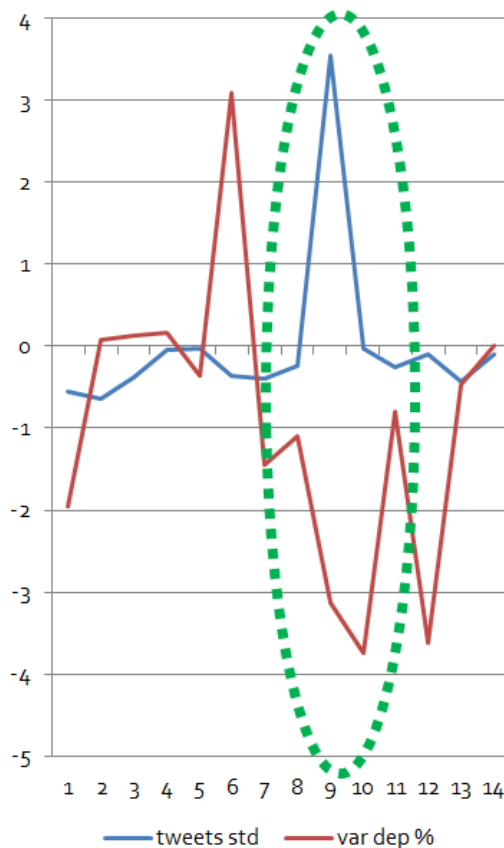


AN OLD ANALYSIS

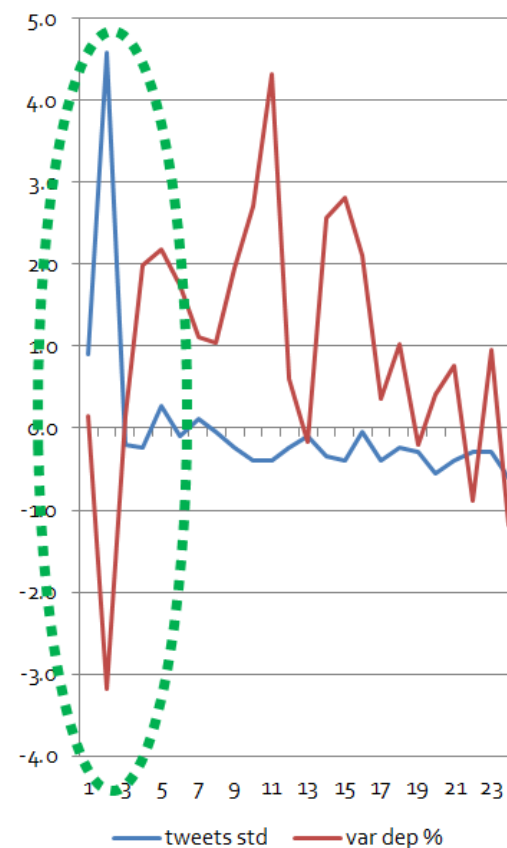
Empirical case 1



Empirical case 2



Empirical case 3





RESULTS

THE MODEL

$$D_{i,t} = \alpha_i + \beta S_{i,t} + \gamma I_{i,t} + \delta S_{i,t} * T1R_{i,t-1} \\ + \sum_k \zeta_k C_{k,i,t-1} + \eta Y_{i,t-1} + \theta D_{i,t-1} + \varepsilon_{i,t}$$

$D_{i,t}$	Monthly percentage variation of households deposits
$S_{i,t}$	Standardized sentiment score
$I_{i,t}$	Interconnection indicator
$Y_{i,t}$	Spread between interest rate on deposits and benchmark
$C_{k,i,t}$	Set of financial control variables

IMPACT OF TWITTER BASED VARIABLES

Variables	(1) Benchmark	(2) Bench. with sentiment	(3) Bench. with sent. and interconnect.	(4) Main model
L.ret_dep_gro_tot	0.141*** (0.0516)	0.110** (0.0556)	0.0984** (0.0488)	0.0425 (0.0544)
L.int_rat_spre	1.151* (0.645)	1.719** (0.797)	1.918** (0.768)	2.204*** (0.839)
L.log_tot_asset	-0.00506 (0.159)	0.0472 (0.180)	0.0660 (0.187)	0.0770 (0.205)
L.t1ratio	0.190*** (0.0353)	0.201*** (0.0434)	0.194*** (0.0463)	0.183*** (0.0492)
L.roa	0.0909 (0.0708)	0.0972 (0.0756)	0.0954 (0.0792)	0.0777 (0.0856)
L.ci_rat	-0.00145 (0.00316)	-0.00241 (0.00362)	-0.000839 (0.00373)	-0.00225 (0.00417)
L.bad_loan_rat	-0.0626 (0.178)	-0.124 (0.176)	-0.244 (0.197)	-0.358 (0.223)
L.liq_asset	0.801 (1.721)	1.193 (1.970)	1.048 (1.984)	1.803 (2.008)
L.whs_fun	-0.620 (0.939)	-0.794 (0.954)	-0.801 (0.929)	-0.895 (0.922)
sen_sco		0.317* (0.169)	0.245 (0.150)	2.003*** (0.493)
inter_ind			1.419*** (0.473)	1.047*** (0.308)
sen_sco*L.t1ratio				-0.163*** (0.0435)
Constant	-3.315 (2.153)	-4.431* (2.475)	-4.622* (2.517)	-4.741* (2.790)

RESIDUAL MEAN SQUARE ERROR

	Distressed banks	Other banks	Total
Benchmark model	6.6	4.0	4.7
Model augmented with sentiment score	6.2	4.3	4.8
Model augmented with sentiment score and interconnection index	5.9	4.3	4.7
Main model	4.6	4.3	4.4

INSURED AND UNINSURED DEPOSITORS

Variables	(1) Growth rate of retail deposits	(2) Growth rate of uninsured retail deposits	(3) Growth rate of insured retail deposits
L.ret_dep_gro_tot	0.0425 (0.0544)		
L. ret_det_gro_unins		0.0846 (0.0645)	
L. ret_det_gro_ins			0.0498 (0.0510)
L.int_rat_spre	2.204*** (0.839)	2.269** (1.051)	2.214*** (0.776)
L.log_tot_asset	0.0770 (0.205)	0.0750 (0.252)	0.0945 (0.190)
L.tlratio	0.183*** (0.0492)	0.210*** (0.0650)	0.168*** (0.0464)
L.roa	0.0777 (0.0856)	0.0810 (0.109)	0.0669 (0.0752)
L.ci_rat	-0.00225 (0.00417)	0.00406 (0.00452)	-0.00533 (0.00411)
L.bad_loan_rat	-0.358 (0.223)	-0.749*** (0.236)	-0.200 (0.224)
L.liq_asset	1.803 (2.008)	4.064* (2.155)	0.487 (2.132)
L.whs_fun	-0.895 (0.922)	-1.423 (1.175)	-0.878 (0.897)
sen_sco	2.003*** (0.493)	2.200*** (0.540)	1.922*** (0.471)
inter_ind	1.047*** (0.308)	1.285*** (0.362)	0.973*** (0.298)
sen_sco*L.tlratio	-0.163*** (0.0435)	-0.176*** (0.0483)	-0.157*** (0.0415)
Constant	-4.741* (2.790)	-5.596 (3.492)	-4.391* (2.562)



CONCLUSIONS

1. There is a significant correlation between social media sentiment and the variation of retail deposits, mainly driven by weaker banks.
2. Twitter data can help improving forecasts concerning deposit withdrawals and banks liquidity distress.
3. Twitter can be used to capture informational contagion effects across banks.



NEXT STEPS

1. Extend the indicators to a longer time period using the new data provided by GNIP, and implement them to receive semi-real time warnings.
2. Bottom-Up sentiment analysis using a machine learning approach.



THANK YOU

For any questions or suggestions:
mirko.moscatelli@bancaditalia.it
matteo.accornero@bancaditalia.it