

RESEARCH

Open Access



Big data analysis for financial risk management

Paola Cerchiello^{*†} and Paolo Giudici[†]

*Correspondence:

paola.cerchiello@unipv.it

[†]Paola Cerchiello and Paolo Giudici contributed equally to this work

Department of Economics and Management Science, Università degli Studi di Pavia, Via San Felice, 5, Pavia, Italy

Abstract

A very important area of financial risk management is systemic risk modelling, which concerns the estimation of the interrelationships between financial institutions, with the aim of establishing which of them are more central and, therefore, more contagious/subject to contagion. The aim of this paper is to develop a novel systemic risk model. A model that, differently from existing ones, employs not only the information contained in financial market prices, but also big data coming from financial tweets. From a methodological viewpoint, the novelty of our paper is the estimation of systemic risk models using two different data sources: financial markets and financial tweets, and a proposal to combine them, using a Bayesian approach. From an applied viewpoint, we present the first systemic risk model based on big data, and show that such a model can shed further light on the interrelationships between financial institutions.

Keywords: Twitter data analysis, Graphical Gaussian models, Graphical model selection, Banking and finance applications, Risk management

Introduction

Systemic risk models address the issue of interdependence between financial institutions and, specifically, consider how bank default risks are transmitted among banks.

The study of bank defaults is important for two reasons. First, an understanding of the factors related to bank failure enables regulatory authorities to supervise banks more efficiently. In other words, if supervisors can detect problems early enough, regulatory actions can be taken, to prevent a bank from failing and, therefore, to reduce the costs of its bail-in, faced by shareholders, bondholders and depositors; as well as those of its bail-out, faced by the governments and, ultimately, by the taxpayers. Second, the failure of a bank very likely induces failures of other banks or of parts of the financial system. Understanding the determinants of a single bank failure may thus help to understand the determinants of financial systemic risks, were they due to microeconomic, idiosyncratic factors or to macroeconomic imbalances. When problems are detected, their causes can be removed or isolated, to limit “contagion effects”.

Background and literature review

The literature on predictive models for single bank failures is relatively recent: until the 1990s most authors emphasize the absence of default risk of a bank (see [1, 2]), in the presence of a generalised expectation of state interventions. However, in the last years we have witnessed the emergence of financial crisis in different areas of the world, and a correlated emphasis on systemic financial risks. Related to this, there have been many developments of the international financial regulation, aimed at mitigating such risks (see, for example, the Bank of International Settlements regulations). In addition, governments themselves are less willing than before to save banks, partly for their financial shortages and partly for a growing negative sentiment from the public opinion. As a consequence of all of these aspects, the very recent years are seeing a growing body of literature on bank failures, and on systemic risks originated from such them.

Most research papers on bank failures are based on financial market models, that originate from the seminal paper of [3], in which the market value of bank assets, typically modelled as a diffusion process, is matched against bank liabilities. Due to its practical limitations, Merton's model has been evolved into a reduced form (see e.g. [4]), leading to a widespread diffusion of the resulting model, and the related implementation in regulatory models.

The literature on systemic risk is very recent and follows closely the developments of the financial crisis, started in 2007. A comprehensive review is provided in [5]. Specific measures of systemic risk have been proposed, in particular, by [5] and [6]. All of these approaches are built on financial market data, on the basis of which they lead to the estimation of appropriate quantiles of the estimated loss probability distribution of a financial institution, conditional on a crash event on the financial market. A different approach, explicitly geared towards estimation of the interrelationships among all institutions, is based on network models, and has been proposed in [7].

Here we shall follow this latter approach, and add a stochastic framework, based on graphical Gaussian models. We will thus be able to derive, on the basis of market price data on a number of financial institutions, the network model that best describes their interrelationships and, therefore, explains how systemic risk is transmitted among them.

All models described so far, both in failure estimation as well in systemic risk modelling, are based on financial market data. Market data are relatively easy to collect, are public, and are quite objective. On the other hand, they may not reflect the true fundamentals of the underlying financial institutions, and may lead to a biased estimation of the probability of failure. This bias may be stronger when the probability of multiple failures are to be estimated, as it occurs in systemic risk. Indeed, the recent paper by Hirsch [8] shows that market models are not much reliable in predictive terms.

More generally, it is well known that market prices are formed in complex interaction mechanisms that, often, reflect speculative behaviours rather than the fundamentals of the companies to which they refer. Market models and, specifically, financial network models based on market data may, therefore, reflect "spurious" components that could bias systemic risk estimation. This weakness of the market suggests to enrich financial market data with data coming from other, complementary, sources. Indeed, market prices are only one of the evaluations that are carried out on financial institutions: other

relevant ones include ratings issued by rating agencies, reports of qualified financial analysts, and opinions of influential media.

Most of the previous sources are private, not available for data analysis. However, summary reports from them are now typically reported, almost in real time, in social networks and, in particular, in financial tweets. Therefore, big data offers the opportunity to extract from them very useful evaluation data that can complement market prices and that can, in addition, “replace” market information when not available (as it occurs for banks that are not listed). To extract from tweets data that can be assimilated to market prices, their text has to be preprocessed using semantic analysis techniques. In our context, if financial tweets on a number of banks are collected daily, it becomes possible to express, using semantic analysis, a daily “sentiment” towards them that expresses, for each day, how each considered bank is, on average, being evaluated by tweeterers.

In this paper we propose how to select and model semantic based tweet data, so to compare and integrate them with market data, within the framework of graphical network models. We thus propose a novel usage of twitter data, aimed at assessing systemic risk with a graphical model built on daily variation of bank “sentiment”. We also introduce a criteria, the T-index, aimed at selecting in advance the most relevant twitter sources, to avoid using non-informative data that may distort the results.

The novelty of this paper is twofold. From a methodological viewpoint, we propose a framework that can estimate systemic risks with models based on two different sources: financial markets and financial tweets, and suggest a way to combine them, using a Bayesian approach.

From an applied viewpoint, we present the first systemic risk model based on big data, and show that such a model can shed further light on the interrelationships between financial institutions.

The rest of the paper is organised as follows: in “[Background and literature review](#)” section we introduce our proposal; in “[Methods](#)” section we apply our proposal to financial and tweet data on the Italian banking market and, finally, in “[Application and results](#)” section we present some concluding remarks.

Methods

In this section we introduce our proposal. First we describe a methodology able to select in advance tweets, based on the H-index proposed by Hirsch [8], employed to measure research impact, for which a stochastic version has been proposed by Cerchiello [9].

The h-index is employed in the bibliometric literature as a merely descriptive measure, that can be used to rank scientists or institutions where scientists work. A similar ranking can be achieved for tweeterers; however the stochastic variability surrounding tweet citations (retweets) is greater than that of paper citations. suggest to formalise the *h*-index for tweet data, named T index, in a proper statistical framework. Here we briefly recall such methodology that we will use in the following.

Let X_1, \dots, X_n be n random variables representing the number of retweets of the N_p tweets (henceforth for simplicity n) of a given twitterer. In the context of research impact measurement, the n random variables are the citations of the n papers of a given scientist. We assume that X_1, \dots, X_n are independent with a common retweet distribution function F . Beirlant [10] and Pratelli [11], among other contributions, assume that F is

continuous, at least asymptotically, even if retweet counts have support on the integer set.

According to this assumption, the T index can be defined in a formal statistical way as in Glänzel [12] and Beirlant [10]:

$$t : 1 - F(t) = \frac{t}{n} \tag{1}$$

The definition should be as much as possible coherent with the nature of the data and, therefore, assuming that F is discrete and, in order to define the T index, order statistics can be profitably employed.

In this paper, given a set of n tweets of a tweeterer to which a count vector of the retweets of each tweet is associated, we consider the ordered sample of retweets $\{X_{(i)}\}$, that is $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$, from which obviously $X_{(1)}$ ($X_{(n)}$) denotes the most (the least) cited tweet. Consequently the T index can be defined as follows:

$$T = \max\{t : X_{(t)} \geq t\}. \tag{2}$$

In our context, only the twitterers with highest values of the T-index will be included in the tweet data source that will be employed to estimate systemic risk. This on the basis of an implicit assumption that the most cited twitterers, being the most influential, are also the most reliable.

Having introduced a method aimed at selecting financial tweets, we now introduce the graphical network models that will be used to estimate relationships between N banks, both with market and tweet data.

Relationships between banks can be measured by their partial correlation, that expresses the direct influence of a bank on another. Partial correlations can be estimated assuming that the observations follow a graphical Gaussian model, in which Σ is constrained by the conditional independence described by a graph (see e.g. [13]).

More formally, let $x = (x_1, \dots, x_N) \in R^N$ be a N - dimensional random vector distributed according to a multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$. Without loss of generality, we will assume that the data are generated by a stationary process, and, therefore, $\mu = 0$. In addition, we will assume throughout that the covariance matrix Σ is not singular.

Let $G = (V, E)$ be an undirected graph, with vertex set $V = \{1, \dots, N\}$, and edge set $E = V \times V$, a binary matrix, with elements e_{ij} , that describe whether pairs of vertices are (symmetrically) linked between each other ($e_{ij} = 1$), or not ($e_{ij} = 0$). If the vertices V of this graph are put in correspondence with the random variables X_1, \dots, X_N , the edge set E induces conditional independence on X via the so-called Markov properties (see e.g. [13]).

In particular, the pairwise Markov property determined by G states that, for all $1 \leq i < j \leq N$:

$$e_{ij} = 0 \iff X_i \perp X_j | X_{V \setminus \{i,j\}}; \tag{3}$$

that is, the absence of an edge between vertices i and j is equivalent to independence between the random variables X_i and X_j , conditionally on all other variables $x_{V \setminus \{i,j\}}$.

Let the elements of Σ^{-1} , the inverse of the variance-covariance matrix, be indicated as $\{\sigma^{ij}\}$ Whittaker [14] proved that the following equivalence also holds:

$$X_i \perp X_j | X_{V \setminus \{i,j\}} \iff \rho_{ijV} = 0 \tag{4}$$

where

$$\rho_{ijV} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \tag{5}$$

denotes the ij th partial correlation, that is, the correlation between X_i and X_j , conditionally on the remaining variables $X_{V \setminus \{i,j\}}$.

Therefore, by means of the pairwise Markov property, and given an undirected graph $G = (V, E)$, a graphical Gaussian model can be defined as the family of all N -variate normal distributions that satisfy the constraints induced by the graph on the partial correlations, as follows:

$$e_{ij} = 0 \iff \rho_{ijV} = 0 \tag{6}$$

for all $1 \leq i < j \leq N$.

Stochastic inference in graphical models may lead to two different types of learning: structural learning, which leads to the estimation of the graphical structure G that best describes the data and quantitative learning, that aims at estimating the parameters of a graphical model, for a given graph. In the systemic risk framework, we are mainly interested in structural learning. Structural learning can be achieved by choosing the graphical structure with maximal likelihood, or its penalised versions, such as AIC and BIC. Here we follow the backward selection procedure implemented in the software R and, specifically, in the function `glasso` from package `glasso`.

For the aim of structural learning, we now recall the expression of the likelihood of a graphical Gaussian model.

For a given graph G , consider a sample X of size n from $P = \mathcal{N}_N(0, \Sigma)$, and let S_n be the corresponding observed variance-covariance matrix. For a subset of vertices $A \subset N$, let Σ_A denote the variance-covariance matrix of the variables in X_A , and define with S_A the corresponding observed variance-covariance submatrix.

When the graph G is decomposable (and we will assume so) the likelihood of the data, under the graphical Gaussian model specified by P , nicely decomposes as follows (see e.g. [15]):

$$p(x|\Sigma, G) = \frac{\prod_{C \in \mathcal{C}} p(x_C|\Sigma_C)}{\prod_{S \in \mathcal{S}} p(x_S|\Sigma_S)}, \tag{7}$$

where \mathcal{C} and \mathcal{S} respectively denote the set of cliques and separators of the graph G , and:

$$P(x_C|\Sigma_C) = (2\pi)^{-\frac{n|C|}{2}} |\Sigma_C|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}\left(S_C(\Sigma_C)^{-1}\right)\right] \tag{8}$$

and similarly for $P(x_S|\Sigma_S)$.

Operationally, a model selection procedure compares different G structures by calculating the previous likelihood substituting for Σ its maximum likelihood estimator under G . For a complete (fully connected) graph such an estimator is simply the observed

variance-covariance matrix. For a general (decomposable) incomplete graph, an iterative procedure, based on the clique and separators of a graph, must be undertaken (see e.g. [13]).

Through model selection, we obtain a graphical model that can be used to describe relationships between banks and, specifically, to understand how risks propagate in a systemic risk perspective. More precisely, in our context, we select one graphical model for each given data source: one from market data and one from tweet data.

Besides comparing the two models, it is quite natural to aim at integrating them into a single model. This task can be achieved within a Bayesian framework, as follows.

We first specify a prior distribution for the parameter Σ . Dawid [15] propose a convenient prior for Σ , the hyper inverse Wishart distribution. It can be obtained from a collection of clique specific marginal inverse Wisharts as follows:

$$l(\Sigma) = \frac{\prod_{C \in \mathcal{C}} l(\Sigma_C)}{\prod_{S \in \mathcal{S}} l(\Sigma_S)} \tag{9}$$

where $l(\Sigma_C)$ is the density of an inverse Wishart distribution, with hyperparameters T_C and α , and similarly for $l(\Sigma_S)$. For the definition of the hyperparameters here we follow [16] and let T_C and T_S be the submatrices of a larger “scale” matrix T_0 of dimension $N \times N$, and choose $\alpha > N$.

Dawid [15] and Giudici [16] show that, under the previous assumptions, the posterior distribution of the variance-covariance matrix Σ is a hyper Wishart distribution with $\alpha + n$ degrees of freedom and a scale matrix given by:

$$T_n = T_0 + S_n \tag{10}$$

where S_n is the sample variance-covariance matrix.

The previous result can be used to combine market data with tweet data, assuming that the former represent “data” and the latter “prior information” in a Bayesian prior-to-posterior analysis.

To achieve this task we recall that, under a complete, fully connected graph, the expected value of the previous inverse Wishart is:

$$E(\Sigma|X) = T_n = (T_0 + S_n)/(\alpha + n) \tag{11}$$

and, therefore, the Bayesian estimator of the unknown variance covariance matrix, the a posteriori mean, is a linear combination between the prior (tweet) mean and the observed (market) mean.

When the graph G is not complete, a similar result holds locally, at the level of each clique and separator.

The previous results suggest to use the above posterior mean as the variance-covariance matrix of a complete graph on which to base (backward) model selection, thereby leading to a new selected graphical model, based on a “mixed” data source, that contains both financial and tweet data, in proportions determined by the quantities α and n .

Application and results

In this section we consider the application of our proposed methodology. For reasons of information homogeneity we concentrate on a single banking market: the Italian banking system, a very interesting context, characterised by a large number of important banks, dominating the economy of the country, in a rapidly changing environment. We focus on large banks that are listed, for which there exist daily financial market data, that we would like to compare and integrate with tweet data.

The list of banks that we consider, along with their total assets at the end of the last quarter of 2013 (in Euro), a measure of bank size, is contained in Table 1. Banks are described by their stock market code (ticker).

For each bank we consider the daily return, obtained from the closing price of financial markets, for a period of 148 consecutive days in the year 2013, as follows:

$$R_t = \log(P_t/P_{t-1}), \quad (12)$$

where t is a day, $t - 1$ the day that precedes it and P_t (P_{t-1}) the corresponding closing price of that bank in that day.

For the same period, we have crawled Twitter, using the software TwitteR, available open source within the R project environment, and chosen all tweets that contain, besides one of the banks in Table 1, a keyword belonging to a financial taxonomy, that we have built, based on our knowledge of which balance sheet information may affect systemic risk. Each obtained tweet has then been elaborated by a commercial partner of ours, Expert System, that has transformed each tweet into a sentiment class, with categories ranging from 1 to 5. Such categories are associated to tweets on the basis of a semantic analysis that allows a text to be automatically processed on the basis of codified rules based the experience of our partner company in business textual analysis. The higher the category, the more positive the sentiment (or value) that the tweet assigns to the bank under analysis.

Table 2 describes our proposed taxonomy, along with the frequency and average sentiment associated to each keyword in our considered big database.

Table 1 List of considered listed Italian banks

Bank Name	Code	Total assets
UniCredit	UCG	926827
Intesa Sanpaolo	ISP	673472
Banca Monte dei Paschi di Siena	BMPS	218882
Unione di Banche Italiane	UBI	132433
Banco Popolare	BP	131921
Mediobanca	MB	72841
Banca Popolare Emilia Romagna	BPE	61637
Banca Popolare di Milano	PMI	52475
Banca Carige	CRG	49325
Banca Popolare di Sondrio	BPSO	32349
Credito Emiliano	CE	30748
Credito Valtellinese	CVAL	29896

Table 2 Taxonomy proposed and descriptive sentiment analysis

Item	Frequency × 100	Average sentiment
Commissions	0.03	2.67
Labour costs	1.49	3.21
Deposits	0.08	2.83
Interbank	0.14	2.19
Management	28.58	3.01
Interest margin	4.91	2.79
Subsidiaries	0.99	3.02
Capital	35.67	3.07
Loan losses	0.73	2.90
Loans	10.11	2.93

The problem with the above data is that information sources (tweets in our case) are not selected in advance. All tweets that contain the considered keywords are crawled. A tweet on a bank may thus be given a high sentiment because it comes from a twitterer that is very favorable with the bank, for example from a bank’s stakeholder. For this reason we made two choices: first we have considered only tweets coming from twitterers specialised in economics and finance. Second we have calculated, for each twitterer, its T-index, and selected only tweets coming from the twitterers with the highest values of such index (see [17] for more details). The selected sources, along with their T-indexes, are reported in Table 3.

We have then focused our analysis only on tweets coming from the above sources. For each bank we have calculated a sentiment daily variation, that mimicks market returns, as follows:

$$S_t = \log(T_t/T_{t-1}) \tag{13}$$

where t is a day, $t - 1$ the day that precedes it, and T_t is the corresponding average daily sentiment on that bank for that day.

From a descriptive viewpoint, we expect the market and the tweet “returns” to show some degree of correlation although, given their different informational content, we do not expect such correlation to be very high. Table 4 below reports, for each bank, the correlation between financial returns and sentiment returns.

Table 3 List of selected twitterers ordered according to their T-index

Source	T-index
La Voce	24
Sole 24 ore	18
Dagospia	14
Linkiesta	12
Reuters Italia	8
Milano Finanza	7
Italia Oggi	6
Ansa Economia	4

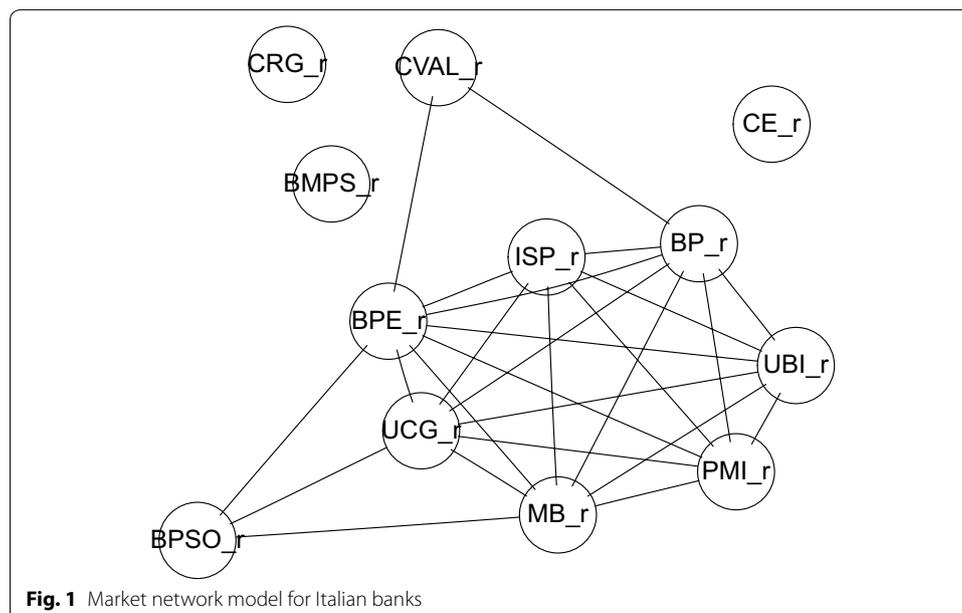
From Table 4 note that correlations are low, especially for smaller banks, that have less tweet information, and this was expected. A detailed inspection of each bank tweet data reveals that banks of similar size show a higher correlation when more information is disclosed. This explains, for example, the difference in the market-sentiment correlation of UCG and ISP, that between CRG and PMI as well as that between BPSO and CVAL.

Our main interest is, however, not in predicting market returns using tweets but, rather, to model systemic risk with both sources of data. In this respect, Fig. 1 reports the selected graphical model obtained with market data. In the graph, each bank is indicated with its ticker code, with the subscript *r* indicating that market returns are being considered.

The graph in Fig. 1 shows a core network of banks, highly correlated with each other, that comprises the largest banks: UCG, ISP, UBI, MB, BP, BPE, PMI. Smaller, more regional banks such as CRG, BPSO, CE, CVAL and, in addition, BMPS, which has gone

Table 4 Correlation between financial and sentiment returns

Bank code	Correlation
UCG	0.31
ISP	0.23
BMPS	0.16
UBI	0.33
BP	0.16
MB	0.27
BPE	0.24
PMI	0.26
CRG	0.09
BPSO	0.08
CE	0.14
CVAL	0.22



through a period of severe crisis, are independent or less connected. In terms of a systemic risk framework, the first group is more central than the second and, in particular, BPE is the most central, followed by UCG and MB.

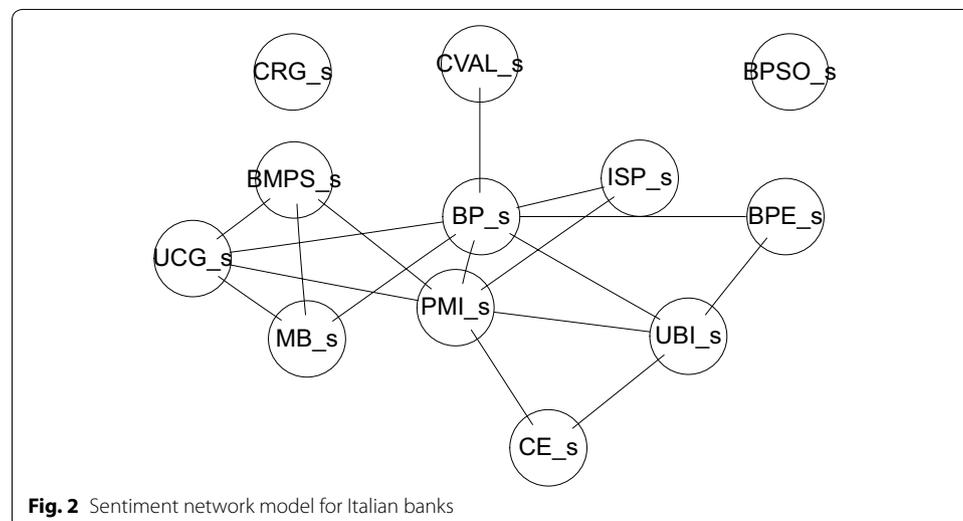
Figure 2 reports the selected graphical model obtained with tweet data. In the graph, each bank is indicated with its ticker code, with the subscript s that indicates that tweet Sentiment returns are being considered.

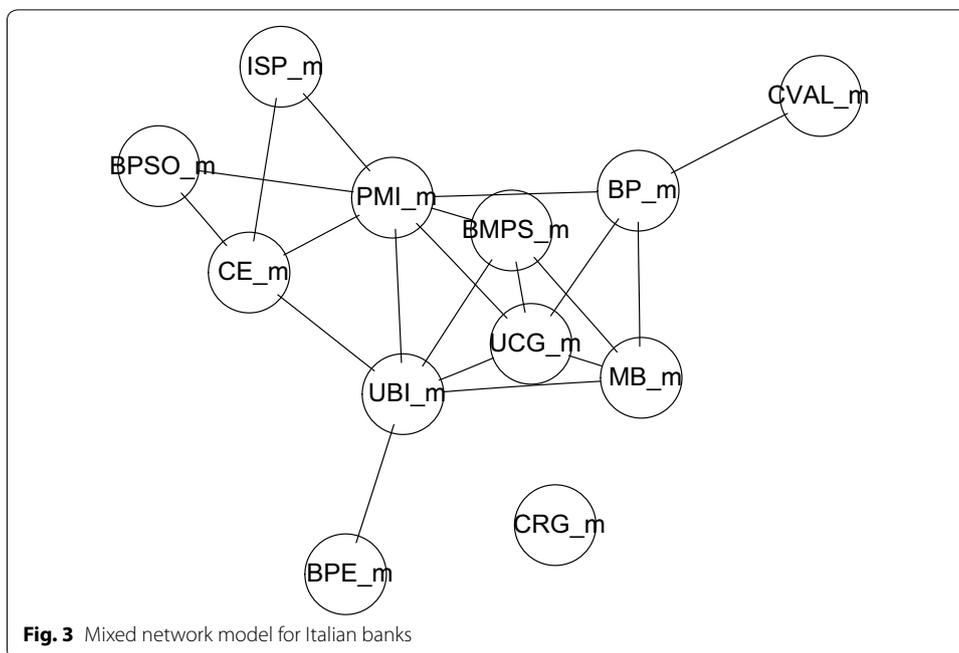
From Fig. 2 note that the network is a little more sparse than that in Fig. 1. This reflects the lower variability of tweet data, with respect to market data. Again there is a central hub of bigger banks, with smaller ones more isolated (CRG, BPSO, CE, CVAL). BMPS is now linked with bigger banks: this is a negative news for regulators, given the critical situation of BMPS, which has the highest probability of failure of all considered banks. In terms of systemic risk, note that the most central banks are BP and PMI, that have, in the considered period, often appeared in Twitter because of frequent news on possible and actual changes in governance and management.

We now consider the selected graphical model, obtained by means of (backward) model selection from the mixed data source, obtained by averaging the complete variance-covariance matrices of financial and tweet data, as shown in the “Application and results” section.

Figure 3 reports the selected model. For the sake of simplicity, and without loss of generality, we have taken $\alpha = n$ so that the market and the tweet data component have equal weights. In the graph, each bank is indicated with its ticker code, with the subscript m that indicates that mixed data are being considered.

Figure 3 emphasises again the distinction between “large” and “small” banks, that especially comes from market data. In addition, it puts in the core of the system banks that are more cited in twitter in the period, such as PMI. Coherently with this “mixing” behaviour the most central banks appear to be PMI, UBI and UCG. The first is the most cited, the third is the largest in size, and the second is highly positioned in both terms. In addition, Fig. 3 further emphasises the relevant systemic risks associated with BMPS, which is now more connected than before.





Conclusions

In this paper we have shown how big data and, specifically, tweet data, can be usefully employed in the field of financial systemic risk modelling.

By means of an appropriate selection of tweets, and of the employment of graphical Gaussian models to estimate relationships between bank tweet sentiment variations, the paper shows how tweet data can be used to estimate systemic risk networks.

Furthermore, the paper shows how to combine tweet based systemic risk networks with those obtained from financial market data, using the a posteriori Bayesian mean of the complete variance-covariance matrix.

We believe that our proposal can be very useful to estimate systemic risk and, therefore, to individuate the most contagious/subject to contagion financial institutions. This because it can compare and integrate two different, albeit complementary, sources of information: market prices and twitter information.

Another important value of the model is its capability of including in systemic risk networks institutions that are not publicly listed, using the tweet component alone: a relevant advantage for banking systems as the Eurozone one, where only 45 out of 131 of the largest banks, subject to the European Central Bank assessment of 2014, are listed.

Authors' contributions

The article is the result of the close collaboration between the authors however PC produced the development and implementation of the methodology and the production of the results. PG guided the initial research idea, and played a pivotal role in editing the article and in interpreting the results. Both authors read and approved the final manuscript.

Acknowledgements

The authors acknowledge financial support from the PRIN project MISURA: multivariate models for risk assessment. They also acknowledge a very useful discussion at the European Central Bank workshop on "Using Big data for forecasting and statistics", on 7/8 april 2014, where a preliminary version of this paper has been presented.

Competing interests

The authors declare that they have no competing interests.

Received: 12 March 2015 Accepted: 16 June 2015
Published online: 01 October 2016

References

1. Gup BE. Bank failures in the major trading countries of the world: causes and remedies. Santa Barbara: Greenwood Publishing Group; 1998.
2. Roth M. "Too-big-to-fail" and the stability of the banking system: some insights from foreign countries. Berlin: Business Economics; 1994. p. 43–9.
3. Merton RC. On the pricing of corporate debt: the risk structure of interest rates. *J Financ*. 1974;29(2):449–70.
4. Vasicek OA. Credit valuation. San Francisco: KMV Corporation; 1984.
5. Acharya VV, Pedersen LH, Philippon T, Richardson MP. Measuring systemic risk. NYU Working Paper. 2010.
6. Huang X, Zhou H, Zhu H. Systemic risk contributions. *J Financ Serv Res*. 2012;42(1–2):55–83.
7. Billio M, Getmansky M, Lo AW, Pelizzon L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Financ Econ*. 2012;104(3):535–59.
8. Hirsch JE. An index to quantify an individual's scientific research output. In: Proceedings of the National academy of sciences of the United States of America; 2005. p. 16569–16572.
9. Cerchiello P, Giudici P. On a statistical h index. *Scientometrics*. 2014;99(2):299–312.
10. Beirlant J, Einmahl JH. Asymptotics for the hirsch index. *Scand J Stat*. 2010;37(3):355–64.
11. Pratelli L, Baccini A, Barabesi L, Marcheselli M. Statistical analysis of the hirsch index. *Scand J Stat*. 2012;39(4):681–94.
12. Glänzel W. On the h-index—a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*. 2006;67(2):315–21.
13. Lauritzen SL. Graphical models, vol. 17. Oxford: Clarendon Press; 1996.
14. Whittaker J. Graphical models in applied multivariate analysis. Chichester, New York: Wiley; 1990.
15. Dawid AP, Lauritzen SL. Hyper markov laws in the statistical analysis of decomposable graphical models. *Ann Stat*. 1993;21(3):1272–317.
16. Giudici P, Green PJ. Decomposable graphical Gaussian model determination. *Biometrika*. 1999;86(4):785–801.
17. Cerchiello P, Paolo G. How to measure the quality of financial tweets. *Qual Quant*. 2016;50:1695–713.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
