BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Press news and social media in credit risk assessment: the experience of Banca d'Italia's In-house Credit Assessment System

by Giulio Gariano and Gianluca Viggiano

July 2022

Number 24

BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Approfondimenti

(Research Papers)

Press news and social media in credit risk assessment: the experience of Banca d'Italia's In-house Credit Assessment System

by Giulio Gariano and Gianluca Viggiano

Number 24 – July 2022

The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of the Bank of Italy in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.

The views expressed in the papers are those of the authors and do not necessarily reflect those of the Bank of Italy.

The series is available online at www.bancaditalia.it.

Printed copies can be requested from the Paolo Baffi Library: richieste.pubblicazioni@bancaditalia.it.

ISSN 2724-6418 (online)
ISSN 2724-640X (print)

Banca d'Italia
Via Nazionale, 91 - 00184 Rome - Italy
+39 06 47921

Designed and printing by the Printing and Publishing Division of the Bank of Italy

Press news and social media in credit risk assessment: the experience of Banca d'Italia's In-house Credit Assessment System

by Giulio Gariano* and Gianluca Viggiano**

Abstract

This article uses press news and Twitter messages to improve the predictive power of the ICAS rating model of Banca d'Italia. We construct two credit sentiment indicators that display a good discriminating power and marginally increase the discriminating power of the ICAS. Scores based on press news prove more effective than those based on Twitter messages.

JEL Classification: G30, C58.

Keywords: credit sentiment, corporate default, social networks, natural language processing.

Sintesi

Il lavoro sfrutta articoli di stampa e annunci Twitter per migliorare il potere predittivo del modello di rating ICAS della Banca d'Italia. Si costruiscono due indicatori di sentimento creditizio (credit sentiment) che dimostrano un buon potere discriminante e accrescono al margine quello del modello ICAS. I punteggi basati su articoli di stampa si dimostrano più efficaci rispetto a quelli basati su annunci Twitter.

* Bank of Italy, Financial Risk Management Directorate.

** Bank of Italy, Economic Research Division, Milan.

CONTENTS

1. Introduction[1]

The acceptance of bank loans to non-financial companies has long been one of the cornerstones of the collateral framework adopted by the European Central Bank (ECB) for the implementation of monetary policy (Auria *et al.*, 2021). In this context, in order to correctly assess the value of loans various tools are used by central banks, inter alia agency ratings and *internal rating based systems* (IRB) validated for this purpose. In addition, in order to support smaller banks, some national central banks, including Banca d'Italia, have developed internal models that are referred to as *In-house credit assessment systems* (ICAS).

The ICAS of Banca d'Italia consists of two building blocks: a purely statistical rating model, which exploits information from financial statements and the Credit Register in order to estimate a *probability of default* (PD) with a one-year horizon[2]; a subsequent process of "expert judgement" by financial analysts (Giovannelli *et al.*, 2020). Combining the two blocks one derives the "complete" rating for the company examined.

The "expert" analysis includes, among other things, an assessment of the qualitative information available on the web about the company, such as press news, interviews, reviews, the company website, etc. This analysis is heterogeneous and difficult to summarize using quantitative indicators that can be used in a statistical model. However, some of the more mechanical steps, such as the analysis of press news, can be automated through text analysis techniques.

The purpose of this note is to analyse the contribution which two data sources may provide to improve Banca d'Italia's ICAS performance: press news available in the database of the Dow Jones Factiva service and messages from the Twitter social network. The relevance of this exercise is twofold:

(i) In general, the statistical model of the ICAS of Banca d'Italia represents a rather difficult benchmark to beat, as it incorporates not only public balance sheet information, but also full information from the Credit Register for all Italian joint-stock companies. In the estimation phase, therefore, a snapshot of the banking relationships of the entire non-financial sector is available, rather than a limited subset, as is the case, for example, in the IRBs of commercial banks. Improving on the performance of ICAS would therefore represent a significant success for text analysis.

(ii) From a practical point of view, only a minority (about 10 per cent) of companies assessed by ICAS are subject to the "expert analysis" phase, mainly due to a cost-benefit analysis that is unfavourable for smaller companies. Therefore, having an automatic and "scalable" integration of qualitative information in the statistical ICAS model would allow to improve

[1] The authors would like to thank Filippo Giovannelli, Aviram Levy, Antonio Scalia and Stefano Siviero for their useful comments, and express special thanks to Juri Marcucci for his valuable technical contribution.

[2] The definition of default used by national ICASs is quite specific. Overlooking some details, a company is considered in "ICAS default" when non-performing exposures (bad debts, unlikely to pay and past-due loans) exceed 5% of the total exposure for three consecutive months, when considering exposures to the entire banking system. For more details, see Giovannelli *et al.* (2020).

the quality of ratings and consequently the allocative efficiency of monetary policy instruments.

In this note we construct two credit sentiment indicators, one based on news articles from the Dow Jones Factiva database and one based on tweets extracted from the social networking service Twitter. We then check whether these indicators discriminate the creditworthiness of companies both on a stand-alone basis and combined with the ICAS statistical model. It turns out that scoring based on press news significantly improves the discriminating power of the ICAS model; the Twitter-based analogue, on the other hand, does not show significant improvements.

The remainder of the note is organized as follows: section 2 presents a review of the most recent literature; sections 3-5 describe the data and the methodology adopted; section 6 presents the results; section 7 concludes.

2. Literature review

Traditional credit risk models are based on quantitative indicators derived from company balance sheets and market data (Altman, 1983). These indicators can be easily included in multivariate regressions or, more recently, in machine learning models (Altman, Barboza and Kimura 2017; Ciampi and Gordini, 2013). In contrast, a wide range of qualitative information contained, for example, in notes and comments on company financial statements, in company press releases and in press news is not exploited by these kind of models.

The most recent literature shows that the exploitation of textual information can increase the accuracy of the credit risk models: Lu, Hung and Tsai (2016) construct a sentiment score based on quarterly reports of US companies with listed CDSs as well as related press reports and show that this score improves the forecast of CDS spreads. Similarly Cathcart *et al.* (2020) show that a news sentiment score obtained via the Thomson Reuters News Analytics service is a significant variable in forecasting sovereign CDS spreads.

The exploitation of alternative textual bases such as social networks and search engines is the object of a rich literature as well. González-Fernández and González-Velasco (2020) use Google Trends service to build a sentiment index for bank credit risk, in alternative to classic CDS-based indices. Lau *et al.* (2018) extract an “emotion indicator” from Twitter messages that can integrate the usual financial indicators in order to forecast company ratings. Bozzon *et al.* (2020) extract information from the Facebook pages of Dutch SMEs to build indicators that complement the traditional balance sheet ones in a credit risk model.

3. Data

For the construction of the dataset used in this study, information from four databases was cross-referenced: ICAS, InfoCamere, Factiva and Twitter. The period considered for the analysis runs from January 2016 to December 2020, for a total of 60 monthly reference dates.

3.1 ICAS

The companies included in the sample are those that received a complete ICAS rating in the survey period: about 6,000 companies, of which 500 were insolvent, according to the ICAS definition, at least once in the 5 years of observation (Table 1). On average, the one-year default rate of the sample is around 1.6 percent (roughly 100 companies defaulting each year). The number of records (i.e., the number of the company-date pairs) is more than 350,000.

Table 1. Companies dataset

Number of companies	6,004
Number of companies defaulting at least once in 5 years	500
Number of records	351,755
Number of records of defaulting companies	5,490

The dataset contains, for each company and for each date: (i) the statistical ICAS score[3] and (ii) the default "flag" of the company, i.e. a binary variable equal to 1 if the company would be in default in the following 12 months and 0 otherwise.

The default "flag", which refers to the following 12 months and corresponds to the dependent variable of our regressions (see section 5), should not be confused with the default "status" of a company at the current date. Figure 1 illustrates the difference using the example of a company which is in default from April 2017 to September 2017. For such a company, the default flag is equal to 1 from April 2016 to March 2017 (and 0 otherwise) since, for each of these dates, a default is observed in the following 12 months.

In addition, records relating to companies already in default at the beginning of the period were removed from the sample, as the ICAS only assigns a PD to companies which are not already in default. Figure 1 shows this filter as well (records from April 2017 to September 2017 are removed from the sample).

---

[3] The statistical ICAS PD is obtained with a logistic model that processes financial statements and the Credit Register data. The relationship between PD and score is: $PD = \frac{1}{1+e^{-score}}$.

Figure 1. Default status and default flag (an example)

Date	Default status	Default flag
31/01/2016	bonis	0
29/02/2016	bonis	0
31/03/2016	bonis	0
30/04/2016	bonis	1
31/05/2016	bonis	1
30/06/2016	bonis	1
31/07/2016	bonis	1
31/08/2016	bonis	1
30/09/2016	bonis	1
31/10/2016	bonis	1
30/11/2016	bonis	1
31/12/2016	bonis	1
31/01/2017	bonis	1
28/02/2017	bonis	1
31/03/2017	bonis	1
30/04/2017	default	Removed from sample
31/05/2017	default	Removed from sample
30/06/2017	default	Removed from sample
31/07/2017	default	Removed from sample
31/08/2017	default	Removed from sample
30/09/2017	default	Removed from sample
31/10/2017	bonis	0
30/11/2017	bonis	0
31/12/2017	bonis	0

The choice of the companies to be included in the sample (i.e. those that have received a complete ICAS rating) takes into account the expected use of the models to be developed in order to assist the financial analysts in the "expert analysis" phase which follows the statistical rating. The selection of a different sample, for example one with a lower concentration of large companies, would have been less representative of the set of companies on which the models would probably be applied[4]. For the same reason, the statistical ICAS score was used in the sample, instead of the subsequent final rating[5].

About 80 percent of the companies considered are medium-large in size and about 75 percent are located in Northern Italy (Table 2).

[4] The models that we are going to develop are difficult to include directly into the ICAS statistical model for several reasons, including: (i) technical difficulties due to the different and non-communicating platforms on which data are hosted and (ii) large initial investment required for generating an appropriate name of the companies (see also section 4).

[5] A further reason for selecting the statistical score is its availability on all dates. The final rating, on the other hand, would be available only on certain dates as the list of companies subject to the "expert judgement" by financial analysts generally changes from year to year.

Table 2. Companies distribution by size and geographical area
(percentage values)

	North-West	North-East	Centre	South	Total
Micro	3	2	1	1	7
Small	6	4	2	1	13
Medium	19	16	7	5	48
Large	13	11	5	3	32
Total	41	34	15	10	

3.2 InfoCamere

The list of managers of each ICAS company was determined using the InfoCamere database. On average, approximately 11 managers per company were identified in the sample period, for a total of approximately 66,000 managers (Table 3).

Table 3. Managers dataset

Number of managers	66,415
Minimum number of managers per company	1
Maximum number of managers per company	62
Average number of managers per company	11

3.3 Factiva

The database we exploited contains on average 600,000 press articles per year, extracted from more than 60 newspapers. The list of newspapers and the number of articles per newspaper are shown in the appendix. This database was obtained from the Dow Jones Factiva service applying a preliminary filter for articles related to economics and finance in order to render its size more manageable (for a similar approach see Aprigliano *et al.*, 2021).

For each article, title, summary, text, date and source are available, as well as a series of metadata that can be used to facilitate the analysis, including: (i) a list of topics covered; (ii) a list of companies mentioned in the article.

In order to analyze only the relevant articles, the following were excluded: (i) articles classified by Factiva in clearly irrelevant topics (e.g. sport, weather, etc.); (ii) articles containing many numerical values; (iii) articles concerning a large number of companies. These last two filters aim at eliminating generic articles on the closing of the stock markets, frequent in the database but not informative for our purposes. Overall, these cleansing operations reduce the sample by around 24 percent.

Table 4 reports, for each year, the total number of available articles, the number of relevant articles (i.e. the number of articles about at least one company), the number of distinct company-date pairs. The number of relevant articles is about 8 percent of the total number of articles. The number of distinct company-date (monthly) pairs is approximately 10 percent of the number of relevant articles, as companies, if mentioned, are mentioned on average in 10 different articles in the same month.

Table 4. Factiva dataset

Year	Number of articles	Number of relevant articles	Number of company-date pairs
2016	581,908	39,043	3,458
2017	603,423	48,872	4,497
2018	641,913	52,123	4,829
2019	651,222	55,716	4,936
2020	397,466	26,185	3,066
Total	2,875,932	221,939	20,786

Since the default "flag" is defined at the level of company-date (monthly) pair, the credit sentiment indicators must be constructed at this same level, aggregating at least all the articles of the same month mentioning the same company.

3.4 Twitter

We leveraged a proprietary Banca d'Italia database containing approximately 30 million tweets per year. This database contains tweets related to general economic and financial topics such as inflation, prices, economic uncertainty and the like (see Angelico *et al.*, 2022). This represents a sensible preliminary filter for the purposes of our research that might affect final results, as some potentially relevant tweets are not considered (i.e., those related to a specific company but not to general economic and financial topics).

The Twitter database does not contain some of the useful metadata available in the Factiva database, such as the list of topics covered[6] and the list of companies mentioned in the article (see previous section), which makes the analysis more time consuming.

Table 5 reports the number of tweets in the database, the number of relevant tweets (i.e. the number of tweets mentioning at least one company), the number of distinct company-date (monthly) pairs. While the number of total tweets is very high (around 50 times the number of articles on Factiva), the number of relevant tweets is only 0.5 percent of the total number of tweets (around 3 times the number of relevant articles on Factiva). The number of distinct company-date pairs is approximately 1.4 percent of the

[6] A list of topic from tweets can be computed with the Latent Dirichlet Allocation approach, see for example Angelico et al. (2022).

number of relevant tweets, since companies, if mentioned, are mentioned on average in 70 different tweets in a month. This is partly due to the interactive nature of Twitter, where tweets trigger a number of other tweets on the same theme (retweets).

Table 5. Twitter dataset

Year	Number of tweets	Number of relevant tweets	Number of company-date pairs
2016	21,134,968	122,832	1,452
2017	23,912,375	110,880	1,697
2018	31,737,472	147,875	2,006
2019	30,187,455	127,761	1,994
2020	33,780,905	131,890	1,946
Total	140,753,175	641,238	9,095

4. Entity recognition

The processing of our text data requires, first of all, to identify whether an article (tweet) mentions one of the companies in our sample. This type of analysis is typically based on an entity recognition algorithm (*Named Entity Recognition*, hereinafter NER) or on a raw search for the company name (denomination) in the text.

There are many algorithms used for NER, almost always developed for the English language, and occasionally extended to other languages[7]. For the Italian language the options are few[8] and, at least on our sample of articles and tweets, not very effective.

In this work, the recognition of companies was therefore carried out through the analysis of metadata, when available (see section 3.3), or through a raw search of the denomination in the text, in the remaining cases. In this second case, the definition of denominations is of fundamental importance, since denominations must be short enough to be contained in an article (tweet) but also sufficiently specific to avoid ambiguity.

The denominations used in our analysis are the result of a first automatic step, in which, starting from those available in our internal databases, some key words were removed (e.g. “Briefly”, “in abbreviation”, “limited liability company“) and names in abbreviation are selected when available, and a second step in which denominations are manually edited (Table 6).

[7] A rather used (Java) package is the Stanford Named Entity Recognizer, available also in German, Spanish and Chinese. In the Python environment, the spaCy library provides several functions for the entity recognition.

[8] An example is TINT (The Italian Nlp Tool).

Table 6. Examples of re-denomination of companies

Original denomination	Denomination used
ASK INDUSTRIES SOCIETA' PER AZIONI	ASK INDUSTRIES
SOCIETA' ITALIANA ACETILENE E DERIVATI S.I.A.D. S.P.A. IN BREVE S.I.A.D. S.P.A.	SIAD
CONSORZIO AGRARIO DEL TIRRENO SOCIETA' COOPERATIVA	CONSORZIO AGRARIO DEL TIRRENO

The companies were then manually divided into two groups: those with sufficiently "specific" names to avoid ambiguity (e.g. Antica Distilleria Domenico Sibona, Distribuzione Elettrica Adriatica), and those with more "generic" names (eg. San Carlo, Il Raccolto)[9]. Generic names, if present in an article (tweet), may not necessarily refer to the company in our sample. In order to select only pertinent data, articles and tweets containing "generic" names are classified as relevant only if they also mention one or more of the managers of the same company. Articles and tweets containing specific names, on the other hand, are automatically classified as relevant, regardless of whether the managers are mentioned. Companies with specific names represent about 80 percent of companies (Table 7).

With reference to Factiva, denominations are searched for only in the *title* or in the summary (*snippet*) of the article[10], while the managers are also searched in the *body* of the article. In the case of Twitter, however, the entire tweet is considered in all cases.

As mentioned in paragraph 3.3, the Factiva metadata includes, for each item, the list of codes of the companies covered by the article (if any and if identified by Factiva). This metadata can be used to increase the ability to identify companies. However, to exploit this information, it is necessary to associate the codes used by Factiva with the various ICAS companies.

In order to match the two samples of companies, an algorithm based on the similarity of the names was used, associating each ICAS company with the Factiva code corresponding to the company with the most similar name. The similarity indicator, which is based on the number and length of the words appearing in both denominations, is computed as:

$$I = \frac{Number\ of\ characters\ of\ words\ present\ in\ both\ denominations}{Number\ of\ characters\ in\ the\ shortest\ denomination}$$

Firstly, for each ICAS company, the Factiva company with the highest similarity indicator was selected. Secondly, pairs with a too low similarity index are removed. Finally, the matches were checked manually.

[9] This is the case, for example, of companies whose names coincide with the names of famous people or commonly used words.

[10] If a company is mentioned in the body of an article but not in its title/snippet then it is likely that the article is not about such company.

In this way it was possible to determine the Factiva code for 659 ICAS companies (Table 7). In any case, these companies have been assigned a specific name (or, in rare cases, a *hashtag*), to be used in the Twitter database[11].

Table 7. Classification of companies by type of denomination

Type of denomination	Number of companies
Factiva code	659
"Specific" name	4,813
"Generic" name	532
Total	6,004

5. Methodology

This paragraph describes the way we construct credit sentiment indicators derived from articles and tweets. The calculation of the indicators is carried out in two stages: (i) first, a credit sentiment score is calculated at the level of a single article (tweet); (ii) subsequently, an aggregate credit sentiment indicator is calculated by combining the scores calculated on all the articles (tweets) available in the last period.

5.1 Credit sentiment score of a single article (tweet)

The calculation of a credit sentiment score at the level of a single article (tweet) is based on the use of a dictionary of keywords. There are many dictionaries already developed for the *sentiment analysis*[12], but almost all of them are in English.

Also taking into account the limited availability of dictionaries in Italian and the fact that they are typically "all-purpose"[13], a custom dictionary has been built that specializes in the semantic area of a corporate crisis.

The dictionary was built by means of an automatic procedure followed by a quality control performed by an analyst:

(i) first of all, a set of words associated with the highest default rates was identified, i.e. words contained in articles (tweets) relating to companies that would have defaulted in the following 12 months;

[11] Factiva codes can only be used in the processing of news articles stored in the Factiva database. When processing Twitters messages, a company denomination is required.

[12] For example the Loghran and McDonald Master Dictionary, or the Bing Liu Sentiment Lexicon.

[13] An example is Sentita, trained on generic texts extracted from Twitter and Wikipedia.

(ii) words that are rarely used or lacking an evident negative connotation have been removed from the previous set;

(iii) about 200 words, some positive, some negative, manually identified as very relevant, were finally added to the dictionary.

The articles used for step (i) relate exclusively to the years 2016-2018, which can be considered *in sample*. The years 2019-2020 can therefore be considered *out of sample*.

The resulting dictionary contains more than 400 words, with a focus on legal, employment and corporate issues (Table 8).

Table 8. Examples of negative words included in the dictionary, by topic

Topic	Word (Italian)	Word (English)
Legal	abusivo	abusive
	arrestato	arrested
	collusione	collusion
	condanna	conviction
	costituirsi	turning oneself
	denunciare	press charges
	domiciliari	house arrests
	vertenza	dispute
Employment	ammortizzatori	welfare
	cassa integrazione	furlough
	esuberi	redundancy
	licenziati	laid off
	proteste	protests
	ricollocazioni	restructuring
	sindacati	labor union
	slitta	fall
Corporate	cessione	divestiture
	delocalizzare	delocalize
	dismessi	write off
	indebitato	indebted
	liquidare	liquidate
	perdita	loss
	ricapitalizzare	recapitalize
	voragine	hole
Default	bancarotta	bankruptcy
	cessare	cease
	chiusura	closing
	default	default
	ristrutturazione	restructuring
	salvataggio	rescue
	scaduto	past-due
	tracollo	collapse
Other	bassissimo	very low
	contrazione	contraction
	crollo	collapse
	debolezze	weaknesses
	disastro	disaste
	disattesi	disregarded
	duramente	harshly
	faticosamente	with difficulty

Each word was assigned a value ranging from -3 (very positive) to +3 (very negative). This choice is motivated by the fact that a very negative score corresponds to a PD of approximately 0 (good company) and a very positive score corresponds to a PD of almost 1 (bad company).

This dictionary was then used to assign a score to each article, calculated as the sum of the values of each keyword found in the text (counted only once). Formally, if *t* is the text of the analyzed article (tweet) and *D* the constructed dictionary, the score of the article is calculated as:

$$s(t) = \sum\nolimits_{w \in D} I(w \in t) \cdot v(w)$$

where $v(\cdot)$ is the value of each word in the dictionary and $I(\cdot)$ the event indicator function in brackets, which is 1 if and only if the word is present in the text.

5.2 Aggregate credit sentiment indicator

The aggregate credit sentiment indicator was obtained by combining the scores calculated on all the articles (tweets) available in the last period. In particular, the following two indicators were calculated for each company and for each date:

(i) average credit sentiment score in the articles (tweets) of the last 6 months (*avg_sent*);
(ii) maximum credit sentiment score in the articles (tweets) of the last 6 months (*max_sent*).

The first indicator provides an indication of the average sentiment about a company, while the second indicator is meant to highlight any particularly negative articles.

Some variants of indicators (i) and (ii) were considered and then discarded. Increasing the time window beyond 6 months, for example, does not add information.

Figures 2 and 3 illustrate the relationship between the indicators constructed as shown above and the observed default rates.

Figure 2. Default rates per buckets of credit sentiment indicators (Factiva)

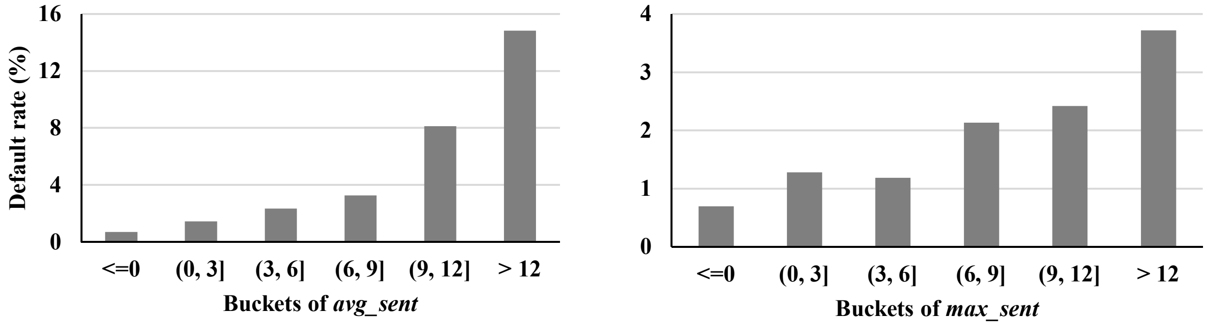
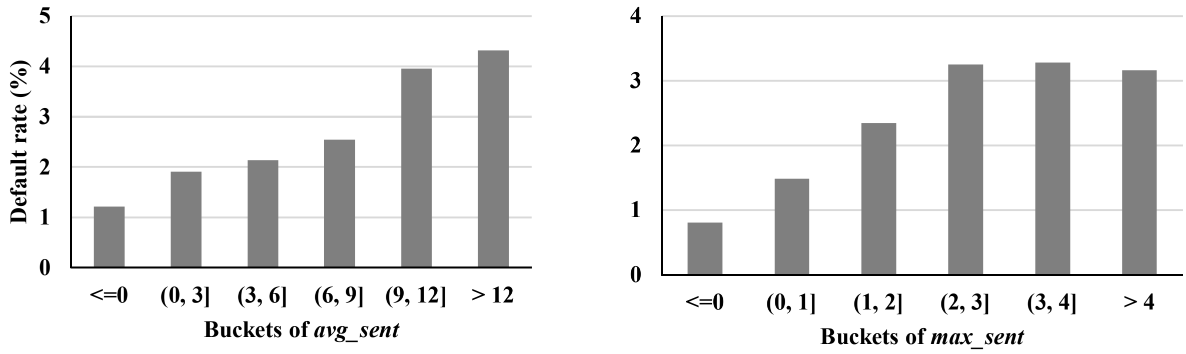


Figure 3. Default rates per buckets of credit sentiment indicators (Twitter)



6. Results

This paragraph analyzes the discriminatory power of the credit sentiment indicators described in paragraph 5.2. First, the following logistic model is estimated:

$$g(P[def\_icas_i = 1]) = \beta_0 + \beta_1 \cdot avg\_sent_i + \beta_2 \cdot max\_sent_i \qquad (1)$$

where def_icas_i is the default flag binary variable, equal to 1 if and only if an ICAS default has occurred in the 12 months following the date of publication of the article (tweet), $g(\pi) = log\big(\pi/(1-\pi)\big)$ is the logistic function, avg_sent and max_sent are the two credit sentiment indicators considered.

As reported in the following paragraphs, the indicators considered have good predictive power, especially those built on Factiva news articles.

Subsequently, a second logistic model was estimated:

$$g(P[def\_icas_i = 1]) = \beta_0 + \beta_1 \cdot avg\_sent_i + \beta_2 \cdot max\_sent_i + \beta_3 \cdot score\_icas_i \tag{2}$$

where $score\_icas$ is the statistical ICAS score.

The purpose of this second regression is to verify whether the two credit sentiment indicators, in addition to having good predictive power, are also able to add information to the statistical ICAS model. The latter includes balance sheet data (available with a delay of between 12 and 18 months) and Credit Register data (available with a delay of two months).

The following paragraphs report the results of the logistic regressions, with the *p-values* of the estimated parameters and the area under the ROC curve (AUROC)[14]. The appendix contains further information, including results disaggregated by year and by size class.

6.1 Factiva

Table 9 presents some information on the dataset used for the logistic regressions with Factiva data, which includes 20,786 observations (see also Table 4).

Each record relates to a company-date (monthly) pair, and contains the default flag and the two credit sentiment indicators calculated as described in the section 5.2.

Table 9. Dataset for the logistic regressions (Factiva)

Year	Number of records	of which: in bonis	of which: defaults
2016	3,458	3,370	88
2017	4,497	4,393	104
2018	4,829	4,741	88
2019	4,936	4,846	90
2020	3,066	3,012	54
Total	20,786	20,362	424

[14] The ROC curve (receiver operating characteristic curve) is a graphical representation of model performance. It plots the true positive rate versus the false positive rate at various thresholds. The area under the ROC curve (AUROC) summarizes the performance with a number between 0 and 1, where 0 represents a model whose predictions are 100% wrong and 1 a model whose predictions are 100% correct. A random classifier corresponds to an AUROC of 0.5, which is considered a lower bound for any classification model. Credit scoring models based on financial statement data typically reach AUROC values around 0.7; the ICAS statistical model, using both financial statement and Credit Register data, reaches a value around 0.85.

It is important to highlight that the dataset used for logistic regression includes only the company-date pairs for which it was possible to find at least one article in the Factiva database; in addition, these pairs represent a small fraction (about 6 percent) of all the company-date pairs (which are 351.755, see Table 1). This is mainly due to two reasons:

(i) some companies, about 60 per cent, are never mentioned in the news articles available in the database (at least using the procedure described in the paragraph 4);
(ii) the remaining companies are mentioned, but only in a limited time interval (9 months over 5 years, on average).

Table 10 reports the details by type of denomination of the companies that are mentioned in the news articles. About 40 percent of companies with a Factiva code or with a specific name are found in at least one article, while the coverage of companies with generic name is lower, around 30 percent, as it is required that at least one of the company managers is also mentioned (see section 4), the latter being an event which is not very common in articles.

Table 10. Companies mentioned in news articles by type of denomination

Type of denomination	**Number of companies**
Factiva code	296
“Specific” name	1982
“Generic” name	147
Total	2,425

Compared to the original dataset, the companies mentioned in the articles are more concentrated in the medium/large size classes, which represent 87 percent of the companies found (it was 80 percent in the original dataset, see Table 2). This result was expected, as medium-large companies are of more general interest. The distribution by geographical area is, on the other hand, substantially in line with that of the original dataset. For more details, see the appendix.

In Table 11, column (b) reports the results of regression (1) with only the credit sentiment indicators. Both indicators are statistically significant. The AUROC of the overall model, equal to 66.8 percent, is in line with those reported in Bozzon *et al.* (2020) and slightly lower than that generally found for statistical rating models based on financial statements data (typically around 70 per cent).

Column (c) shows the regression results and the AUROCs of the logistic model (2), which also includes the statistical ICAS score. The credit sentiment indicators continue to remain statistically significant, and marginally increase the AUROC of the ICAS model, which goes from 85.2 to 86.3 percent.

Table 11. Regression results (Factiva)

	def_icas		
	(a)	(b)	(c)
score_icas	0.967*** (0.015)		1.178*** (0.025)
avg_sent		0,111*** (0.014)	0.092*** (0.017)
max_sent		0.053*** (0.008)	0.047*** (0.008)
Number of records	20,786	20,786	20,786
AUROC	0.852	0.668	0.863

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.001$. Standard errors in parentheses.

Further details on the performance of the estimated logistics models are provided in the appendix. Firstly, the "in sample" (years 2016-2018) performance of the model (1) is in line with the one "out of sample" (years 2019-2020): the AUROC is equal to 66.6 and 67.3 per cent, respectively. Secondly, model (1) is particularly effective in the case of large companies (AUROC equal to 0.719), whereas for the remaining companies the discriminating power is lower (AUROC from 0.584 to 0.671). This is probably due to the fact that the estimation sample is very concentrated towards large companies. The statistical ICAS model, on the other hand, performs better with small companies (AUROC of 0.873) than with others (AUROC from 0.786 to 0.828) and this generates an integrated model with a good discriminating power on all size classes.

6.2 Twitter

Table 12 reports some information on the dataset used for the logistic regressions with Twitter data, which includes 9,095 observations (see also Table 5).

Table 12. Dataset for the logistic regressions (Twitter)

Year	Number of records	of which: in bonis	of which: defaults
2016	1,452	1,414	38
2017	1,697	1,658	39
2018	2,006	1,965	41
2019	1,994	1,930	64
2020	1,946	1,911	35
Total	9,095	8,878	217

Table 13 presents the distribution by type of denomination of the companies that are mentioned in the tweets. Only 19 percent of companies with specific names are mentioned in at least one tweet, while the coverage of companies with generic names is close to zero, as it is extremely rare for managers of any company to be mentioned in a tweet. Similarly to what happened with the Factiva database, when companies are mentioned this happens only in a limited time interval (again, on average 9 months over 5 years).

Table 13. Companies mentioned in tweets by type of denomination

Type of denomination	Number of companies
"Specific" name	1,058
"Generic" name	7
Total	1,065

Compared to the original dataset, the companies mentioned in the tweets are more concentrated in the medium/large size classes, which represent 90 percent of the companies (this share was 80 percent in the original sample, see Table 2). The distribution by geographical area is substantially in line with that of the original sample. For more details, see the appendix.

Column (b) of Table 14 reports the results of the first regression, which includes only the credit sentiment indicators. The AUROC of the overall model, equal to 61.2 percent, is lower than that of the model based on Factiva indicators (66.8 percent, see paragraph 6.1). Column (c) shows the results of the regression and the AUROCs of the second logistic model, which also includes the statistical ICAS score. Although the credit sentiment indicators remain statistically significant, the AUROC of the ICAS model does not increase[15].

[15] In column (a), the AUROC of the ICAS model (84.9%) changes from the previous paragraph (where it was 85.2%) because the dataset on which it is calculated (i.e. the company-date pairs considered) has changed.

Table 14. Regression results (Twitter)

	def_icas		
	(a)	(b)	(c)
score_icas	0.906*** (0.020)		1.014*** (0.031)
avg_sent		0.287*** (0.072)	0.244*** (0.079)
max_sent		0.067** (0.029)	0.059* (0.031)
Number of records	9,095	9,095	9,095
AUROC	0.849	0.612	0.847

Note: * p < 0.1; ** p < 0.05; *** p < 0.001. Standard errors in parentheses.

The lower performance of Twitter-based indicators is partly explainable in light of the different nature of the information source, since social networks are more general purpose and less specialized than press articles.[16]

Further details on the performance of the estimated logistics models are provided in the appendix.

7. Conclusions

In this paper we used the information available in the Factiva and Twitter databases in order to predict the default of Italian companies and increase, where possible, the already high performance of the ICAS model.

Credit sentiment indicators built from press articles show good discriminating power and can help improve the predictive capacity of the ICAS model. The indicators built on tweets, on the other hand, show a more modest predictive power, insufficient to improve that of the ICAS. The lower performance of Twitter-based indicators is partly explainable in light of the different nature of the information source, since social networks are more general purpose and less specialized than press articles.

Ultimately, both from a theoretical and an empirical viewpoint the exercise gave positive results. In particular, the constructed credit sentiment indicators could be made available to ICAS analysts in the "expert" phase, through a routine that queries the Factiva/Twitter databases in real time after having received the name (or a series of names) of the company as input.

[16] The lower performance could also reflect a lower effectiveness of the company recognition algorithm on Twitter, which is characterized by a more concise language and where denominations are often abbreviated or replaced by hashtags.

References

Altman, E.I. (1983), A complete guide to predicting, avoiding, and dealing with bankruptcy. *Corporate Financial Distress, New York*.

Altman, E., Barboza, F. and Kimura, H. (2017), Machine learning models and bankruptcy prediction, *Expert Systems with Applications*, *83*, 405-417.

Angelico, C., Marcucci, J., Miccoli, M., and Quarta, F. (2022), Can we measure inflation expectations using Twitter?, *Journal of Econometrics*, *228*(2), 259-277.

Aprigliano, V., Emiliozzi, S., Guaitoli, G., Luciani, A., Marcucci, J., and Monteforte, L. (2021), The power of text-based indicators in forecasting the Italian economic activity, *Banca d'Italia Working Papers,* No. 1321.

Auria, L., Bingmer, M., Charavel, C., Gavilá, S., Graciano, C., Iannamorelli, A., Levy, A., Maldonado, A., Mateo, C., Resch, F., Rossi, A.M. and Sauer, S. (2021), Overview of Central Banks' In-House Credit Assessment Systems in the Euro Area, *European Central Bank Occasional Paper Series*, No. 284.

Bozzon, A., Putra, S.G.P., Joshi, B. and Redi, J. (2020), A Credit Scoring Model for SMEs Based on Social Media Data, in *International Conference on Web Engineering* (pp. 113-129), Springer, Cham.

Cathcart, L., Gotthelf, N.M., Uhl, M., and Shi, Y. (2020), News sentiment and sovereign credit risk, *European Financial Management*, *26*(2), 261-287.

Ciampi, F. and Gordini, N. (2013), Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises, *Journal of Small Business Management*, *51*(1), 23-45.

Giovannelli, F., Iannamorelli, A. Levy, A. and Orlandi, M. (2020), The In-house credit assessment system of Banca d'Italia, *Banca d'Italia Occasional Paper*, No. 586.

González-Fernández, M. and González-Velasco, C. (2020), An alternative approach to predicting bank credit risk in Europe with Google data, *Finance Research Letters*, *35*, 101281.

Lau, R.Y., Li, C., Yuan, H. and Wong, M.C. (2018), Mining Emotions of the Public from Social Media for Enhancing Corporate Credit Rating, in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)* (pp. 25-30), IEEE.

Lu, H.M., Hung, M.W. and Tsai, F.T. (2016), The impact of news articles and corporate disclosure on credit risk valuation, *Journal of Banking e Finance*, *68*, 100-116.

Appendix

A.1 Publications included in Factiva

Table 15 shows the list of the top 10 newspapers by number of articles available in the Factiva database and the number of articles present in each year. Overall, the database includes more than 60 newspapers, but 62 percent of the articles come from the top 10.

Table 15. Number of articles by newspapers and year

	2016	2017	2018	2019	2020
La Repubblica	62,222	59,328	65,408	63,687	59,411
Il Sole 24 Ore	58,731	54,938	49,047	46,953	44,287
Corriere della Sera	24,907	31,100	46,858	51,161	69,901
Il Gazzettino	78,506	38,322	30,081	33,261	10,184
Il Messaggero	86,825	37,491	26,062	29,147	8,493
Il Mattino	72,701	32,896	22,908	32,321	24,484
Il Tirreno	6,799	38,431	35,368	31,775	13,112
Corriere Adriatico	59,554	23,679	12,812	12,914	4,575
La Gazzetta del Mezzogiorno		12,695	42,268	40,205	12,378
La Stampa	14,317	13,889	12,253	15,862	23,947
Other (53)	117,346	260,654	298,848	293,936	126,694
Total	581,908	603,423	641,913	651,222	397,466

A.2 Factiva: other statistics

Table 16 shows the distribution by size class and by geographical area of the 2,425 companies mentioned in the articles (see also Table 10).

Table 16. Distribution of companies mentioned in news articles
(percentage values)

	North-West	North-East	Centre	South	Total
Micro	2	1	1	0	4
Small	4	3	2	1	9
Medium	14	16	7	5	42
Large	17	17	7	4	45
Total	36	37	17	10	

Table 17 reports, for each size class and for each geographical area, the percentage of company mentioned in news articles (with respect to all companies in the initial dataset).

Table 17. Company mentioned in news articles

(percentage values)

	North-West	**North-East**	**Centre**	**South**	Total
Micro	25	17	20	25	22
Small	27	29	30	27	28
Medium	29	40	41	37	35
Large	54	60	60	56	57
Total	36	44	43	41	40

Table 18 reports the AUROC of the logistic models on portions of the original dataset consisting of only a few years (periods 2016-2018 and 2019-2020) or some size classes[17]. Column (a) refers to the statistical ICAS, column (b) refers to the model consisting of the two credit sentiment indicators calculated on the Factiva articles, column (c) refers to the integrated ICAS plus Factiva model.

Table 18. AUROC of the logistic models by years and size classes

		Number of records	**(a)**	**(b)**	**(c)**
Years	2016-2018	12,784	0.859	0.666	0.877
	2019-2020	8,002	0.837	0.673	0.834
Size	Micro	733	0.840	0.671	0.861
	Small	1.816	0.873	0.601	0.814
	Medium	6.550	0.786	0.584	0.808
	Large	11.623	0.828	0.719	0.837

A.3 Twitter: other statistics

Table 19 shows the distribution by size class and geographical area of the 1065 companies (Table 13) mentioned in the tweets.

[17] AUROC by size class does not take into account 64 observations (out of the total 20,786) which refer to a dozen companies for which the size class is not available.

Table 19. Distribution of companies mentioned in tweets

(percentage values)

	North-West	North-East	Centre	South	Total
Micro	1	1	1	0	3
Small	3	2	2	1	8
Medium	14	11	6	3	35
Large	22	19	9	5	55
Total	40	33	17	9	

Table 20 reports, for each size class and for each geographical area, the percentage of company mentioned in tweets (with respect to all companies in the initial dataset).

Table 20. Company mentioned in tweets

(percentage values)

	North-West	North-East	Centre	South	Total
Micro	6	4	14	3	7
Small	9	10	13	10	10
Medium	13	12	15	11	13
Large	31	30	32	28	30
Total	17	17	20	16	18

Table 21 reports the AUROCs of the logistic models on portions of the original sample consisting of only a few years (2016-2018 vs 2019-2020) or some size classes[18]. Column (a) refers to the statistical ICAS, column (b) refers to the model consisting of the two credit sentiment indicators calculated from the tweets, column (c) refers to the integrated ICAS plus Twitter model.

Table 21. AUROC of the logistic models by years and size classes

		Number of records	(a)	(b)	(c)
Years	2016-2018	5,155	0.898	0.631	0.898
	2019-2020	3,940	0.791	0.584	0.786
Size	Micro	421	0.817	0.659	0.814
	Samll	692	0.917	0.483	0.922
	Medium	2,204	0.913	0.545	0.899
	Large	5,701	0.734	0.597	0.732

[18] AUROC by size class does not take into account 77 observations (out of the 9,095 total) which refer to a dozen companies for which the size class is not available.

Papers published in the 'Markets, Infrastructures, Payment Systems' series

n. 1 TIPS - TARGET Instant Payment Settlement – The Pan-European Infrastructure for the Settlement of Instant Paymentsi, *by Massimiliano Renzetti, Serena Bernardini, Giuseppe Marino, Luca Mibelli, Laura Ricciardi and Giovanni M. Sabelli* (Institutional Issues)

n. 2 Real-Time Gross Settlement systems: breaking the wall of scalability and high availability, *by Mauro Arcese, Domenico Di Giulio and Vitangelo Lasorella* (Research Papers)

n. 3 Green Bonds: the Sovereign Issuers' Perspective, *by Raffaele Doronzo, Vittorio Siracusa and Stefano Antonelli* (Research Papers)

n. 4 T2S - TARGET2-Securities – The pan-European platform for the settlement of securities in central bank money, *by Cristina Mastropasqua, Alessandro Intonti, Michael Jennings, Clara Mandolini, Massimo Maniero, Stefano Vespucci and Diego Toma* (Institutional Issues)

n. 5 The carbon footprint of the Target Instant Payment Settlement (TIPS) system: a comparative analysis with Bitcoin and other infrastructures, *by Pietro Tiberi* (Research Papers)

n. 6 Proposal for a common categorisation of IT incidents, *by Autorité de Contrôle Prudentiel et de Résolution, Banca d'Italia, Commissione Nazionale per le Società e la Borsa, Deutsche Bundesbank, European Central Bank, Federal Reserve Board, Financial Conduct Authority, Ministero dell'Economia e delle Finanze, Prudential Regulation Authority, U.S. Treasury* (Institutional Issues)

n. 7 Inside the black box: tools for understanding cash circulation, *by Luca Baldo, Elisa Bonifacio, Marco Brandi, Michelina Lo Russo, Gianluca Maddaloni, Andrea Nobili, Giorgia Rocco, Gabriele Sene and Massimo Valentini* (Research Papers)

n. 8 The impact of the pandemic on the use of payment instruments in Italy, *by Guerino Ardizzi, Alessandro Gambini, Andrea Nobili, Emanuele Pimpini and Giorgia Rocco* (Research Papers) (in Italian)

n. 9 TARGET2 – The European system for large-value payments settlement, *by Paolo Bramini, Matteo Coletti, Francesco Di Stasio, Pierfrancesco Molina, Vittorio Schina and Massimo Valentini* (Institutional Issues) (in Italian)

n. 10 A digital euro: a contribution to the discussion on technical design choices, *by Emanuele Urbinati, Alessia Belsito, Daniele Cani, Angela Caporrini, Marco Capotosto, Simone Folino, Giuseppe Galano, Giancarlo Goretti, Gabriele Marcelli, Pietro Tiberi and Alessia Vita* (Institutional Issues)

n. 11 From SMP to PEPP: a further look at the risk endogeneity of the Central Bank, *by Marco Fruzzetti, Giulio Gariano, Gerardo Palazzo and Antonio Scalia* (Research Papers)

n. 12 TLTROs and collateral availability in Italy, *by Annino Agnes, Paola Antilici and Gianluca Mosconi* (Research Papers) (in Italian)

n. 13 Overview of central banks' in-house credit assessment systems in the euro area, *by* Laura *Auria, Markus Bingmer, Carlos Mateo Caicedo Graciano, Clémence Charavel, Sergio Gavilá, Alessandra Iannamorelli, Aviram Levy, Alfredo Maldonado, Florian Resch, Anna Maria Rossi and Stephan Sauer* (Institutional Issues)

n. 14 The strategic allocation and sustainability of central banks' investment, *by Davide Di Zio, Marco Fanari, Simone Letta, Tommaso Perez and Giovanni Secondin* (Research Papers) (in Italian)

n. 15 Climate and environmental risks: measuring the exposure of investments, *by Ivan Faiella, Enrico Bernardini, Johnny Di Giampaolo, Marco Fruzzetti, Simone Letta, Raffaele Loffredo and Davide Nasti* (Research Papers)

n. 16 Cross-Currency Settlement of Instant Payments in a Multi-Currency Clearing and Settlement Mechanism, *by Massimiliano Renzetti, Fabrizio Dinacci and Ann Börestam* (Research Papers)

n. 17 What's ahead for euro money market benchmarks?, *by Daniela Della Gatta* (Institutional Issues) (in Italian)

n. 18 Cyber resilience per la continuità di servizio del sistema finanziario, *by Boris Giannetto and Antonino Fazio* (Institutional Issues) (in Italian)

n. 19 Cross-Currency Settlement of Instant Payments in a Cross-Platform Context: a Proof of Concept, *by Massimiliano Renzetti, Andrea Dimartina, Riccardo Mancini, Giovanni Sabelli, Francesco Di Stasio, Carlo Palmers, Faisal Alhijawi, Erol Kaya, Christophe Piccarelle, Stuart Butler, Jwallant Vasani, Giancarlo Esposito, Alberto Tiberino and Manfredi Caracausi* (Research Papers)

n. 20 Flash crashes on sovereign bond markets – EU evidence, *by Antoine Bouveret, Martin Haferkorn, Gaetano Marseglia and Onofrio Panzarino* (Research Papers)

n. 21 Report on the payment attitudes of consumers in Italy: results from ECB surveys, *by Gabriele Coletti, Alberto Di Iorio, Emanuele Pimpini and Giorgia Rocco* (Institutional Issues)

n. 22 When financial innovation and sustainable finance meet: Sustainability-Linked Bonds, *by Paola Antilici, Gianluca Mosconi and Luigi Russo* (Institutional Issues) (in Italian)

n. 23 Business models and pricing strategies in the market for ATM withdrawals, *by Guerino Ardizzi and Massimiliano Cologgi* (Research Papers)