Expert Systems with Applications: X 5 (2020) 100026

Contents lists available at ScienceDirect

Expert Systems with Applications: X

journal homepage: www.elsevier.com/locate/eswax

Using the European Commission country recommendations to predict sovereign ratings: A topic modeling approach

Ivan Pastor Sanz

University of Valladolid, School of Business and Economics, Avda. Valle del Esgueva 6, 47011 Valladolid, Spain

ARTICLE INFO

Article history:
Received 6 September 2019
Revised 27 February 2020
Accepted 27 February 2020
Available online 28 February 2020

JEL classification:
C35
G24
G15
H63

Keywords:
Credit ratings
European countries
Doc2vec
Country risk

ABSTRACT

This paper examines the role of textual and unstructured data in the credit risk assessment of sovereigns. Specifically, in this paper, a novel approach to understand and predict sovereign ratings is proposed. For that purpose, information embedded in the annual country reports issued by the European Commission is used. The model employs a neural-network-based document embedding known as document to vector (Doc2Vec) to convert each country report into a numerical vector, which is then used as features into a logistic regression. The model is trained using information from 2011 to 2019 and it correctly predicts the 70.27% of country ratings in the test sample, improving slightly the results obtained using only macroeconomic variables.

1. Introduction

In a more globalized and complex world, investors need standard metrics to compare the credit quality between issuers even if they belong to very different countries. Ratings are one of the most used metrics. Ratings summarize into a single label the ability and willingness to service the obligor's debts in full and on time where an implicitly probability of default can be associated (Chen, Chen, Chang & Yang, 2016). Although the number of credit rating agencies (CRAs) is high, only three collects more than the 90% of the market share: Moody's, Standard and Poor's (S&P) and Fitch. CRAs assign ratings for companies and even countries. In recent years, the number of rated obligors has increased considerably. The reliance on credit ratings has granted until now a high influence and political power to the credit agencies. Such influence has been enforcement by law, for example in the framework of Basel Accords, where credit agencies' ratings influence in the calculations of capital requirements for banks.

Credit ratings determine the cost of funding of obligors in capital markets and the access to derivative and loan contracts. However, despite its importance, credit rating agencies have been criticized as well. For example because their performance in important crises such as the Asian and Russian financial crisis in the late 1990s, or even in the recent global financial crisis that started in 2008. During these events, CRAs exacerbated unbalances, taking too much time to react or to predict the default events and then they reacted excessively, with severe downgrades (Pagano & Volpin, 2010; Reinhart, Levich & Majoni, 2002; Reinhart, 2002b). For example, in the most recent financial crisis, the European economies were downgraded by three notches in average. Greece and Italy were among the most affected countries, which were downgraded from A to CCC and from AA– to BBB respectively. Additionally, CRAs are questioned because the called "issuer-pays" model where agencies are paid by the issuers to publish a rating, which is a potential source of conflict of interest (Bernal, Girard & Gnabo, 2016; Haan & Amtenbrink, 2011).

The G-20 and the Financial Stability Board principles have recently encouraged the use of internal ratings rather than external credit ratings in the case of financial institutions, even for calculating own fund requirements to reduce overreliance on external credit ratings. The Basel Committee on Banking Supervision on December 2014 explicitly includes reducing mechanistic reliance on external ratings as one of its objectives.

In a rating process, many factors are considered, including political, growth, external debt, financial sector, structure of the public and private sector, social development and trade among others. CRAs provides a general information about the process fol-

E-mail address: ivanpastorsanz@gmail.com

https://doi.org/10.1016/j.eswax.2020.100026

lowed to rate a company or a country. For example, Standard and Poor's scores five key factors of a sovereign's degree of creditworthiness on a six-point scale (S&P, 2014b). Nevertheless, the different weights and the way factors are combined is unknown and it is not transparent, and also existing qualitative judgments which usually play a very important role in the rating assignation process. Moreover, the importance of these factors changes over time, in response to changes in macroeconomic situation being more point in time than pro-cyclical (Kiff, Nowak & Schumacher, 2012; Mora, 2006).

Concerns about the quality of ratings and their understanding have supposed an incentive to researches around the world since the last two decades. For example, it is considered as important to understand the determinants of ratings to assign a score to non-rated countries. The replication of credit ratings is mainly based on different econometric models trying to find a set of macroeconomic variables which better explain them. The most common approaches are usually ordinary least squares regression or ordinal response models.

Nevertheless, credit rating studies have widely ignored the qualitative and also more subjective information. This additional information could complement the current quantitative macroeconomic information and improve the performance of credit rating models. In this paper, the importance of qualitative information to explain the Standard and Poor's credit ratings is assessed. For this purpose, the yearly country reports issued by the European Commission for the European Member States are used. In Country Reports, the European Commission monitors progress made on social and economic issues in each Member State. There is a country Report per Member State, providing an overview of the general state of the EU's economy in a quantitative and qualitative way. In particular, the recently continuous space model named doc2vec is used (Le & Mikolov, 2014) which is an extension of the word2vec embedding model (Mikolov et al., 2013), a learning technique for natural language processing.

Proposed model firstly uses a shallow neural network to convert words contained in these reports in an n-dimensional space vector. Thus, each word in a document is converted in a vector such that words that share common contexts in the corpus are located in close proximity to one another in the space. In this case, not only words but the whole country reports are vectorized. It means that all the quantitative and qualitative information contained in each report will be summarized in a vector. As a final step, each country vector is used as the input of a logistic regression trying to predict country ratings.

According to our model, which is based on qualitative information provided in the country reports overcomes the accuracy of a model based only on a set of macroeconomic variables. The proposed model demonstrates the importance of qualitative information as a determinant of credit ratings and the potential of natural language processing to mitigate credit risk.

This paper is organized as follows. Section 2 reviews previous research on credit rating models. Section 3 describes most common ways to deal with textual data: the embedding approach. Section 4 details the data and main results obtained from the proposed model based on the European Country Reports. Section 5 compares the proposed model with two different models, first with an ordered logistic regression using only macroeconomic information and then with other combing macroeconomic information and information extracted from country reports. Finally, conclusions can be found in Section 6.

2. Literature review and CRA's criteria

External ratings have become increasingly important, where the number of rated countries have increased significantly since the 1990s, being relevant, for example, to determine the cost of funding for a country. At December 2018, there are more than 150 rated countries by at least one of the three main credit rating agencies.

Concerns about the quality of ratings and their understanding have supposed an incentive to researches around the world since the last two decades. Given its importance, quality of ratings should be assessed and, in some way, they should be transparently traced and replicated, which is not in practice an easy task.

Although credit rating agencies provides methodological guidelines about how they establish ratings, this information is not considered transparent enough or easily replicable. In a first step, rating agencies relies on macroeconomic, public and external finance as well as institutional factors to establish an initial score or rating which is then discussed by a committee of analysts. This committee decided the final rating mainly based on a simple majority vote. Their decision also includes a forward-looking opinion, considering if an issuer, for example a country, could experience unusual adverse conditions in case of a moderate stress. Thus, rating process can be divided in two main parts: based on fundamental analysis and subjective. Regarding the first part of the rating process, the fundamental factors, the three main agencies (S&P, Fitch and Moody's) measure similar factors. However, the way factors are combined is not exactly the same.

Standard and Poor's (2014) uses five fundamental determinants namely political, economic, monetary, external, and fiscal score key indicators as the basis of assigning sovereign ratings. Factors are assessed on a six-point numerical scale from one (the strongest) to six (the weakest) and grouped into two general categories: sovereign's institutional and economic profile. Then, the resulting two categories are combined using a matrix to determine the "indicative or basis rating level" which is usually modified with supplemental adjustment factors. S&P does not publish the weights of variables and even the complete list of variables they use, being difficult the replication of the rating process. On the other hand, Fitch (2014) bases their ratings on macroeconomic, public finances (general government), external finances and structural dimensions, including 18 different variables such as consumer price inflation, real GDP growth or GDP per capita among others. Fitch approach is, in this part, more transparent than S&P, publishing the weights of the four key factors of their methodology, although the importance of each variable is not provided. Finally, Moody's (2013) uses four fundamental determinants: economic strength, institutional strength, fiscal strength and susceptibility to event risk. Each determinant is based on a subset of well-defined list of variables. The rating agency publishes the weights and thresholds of variables being the most transparent agency in this fundamental part of the rating assignment.

Although the rating criteria are, in some cases, more detailed, such as Fitch and Moody's, some authors (Gärtner et al., 2011; D'Agostino & Lennkh, 2016; Vernazza & Nielsen, 2015) have demonstrated that the subjective importance of the rating process is high and even that it changes over the time.

Despite the high influence of the subjective component, some attempts to predict ratings using only a reduced set of variables have demonstrated a high ability to understand credit ratings. There are two main approaches in the literature related to the economic approach used. The first approach uses linear regression methods on a numerical representation of the ratings on a panel data by doing fixed or random effects. Linear regressions have obtained a good predictive power and they have helped to a better comprehension of the main determinants of sovereign credit ratings. Thus, in one of the first important papers on this topic, Lee (1993) estimated a linear regression model for 40 developing countries for 1979–87 using 9 variables obtaining a good adjustment. Cantor and Packer (1996) explained ratings of S&P and Moody's to 49 countries throughout of some macroeconomic variables. Six variables reached a high predictive power (per capita income, GDP growth, inflation, external debt, level of economic development, and default history). Later, other authors extended the number of countries, variables and period in their approaches. Thus, Monfort and Mulder (2000) extended the period of sample to 1994–1999 to include the Asian economic crisis. Eliasson (2002) used macroeconomic indicators to explain S&P ratings using a random effect panel data model, finding that variables originally used in Cantor and Packer (1996) still explains a significant part of the country ratings. Afonso (2003) wide these previous studies including 81 countries in June 2001. Alexe, Hammer, Kogan and Lejeune (2003) obtained a linear model which highly correlates with the S&P ratings using a sample of 69 countries at the end of December 1998 and Rowland and Torres (2004) used a panel data from 16 emerging market issuers to identify the determinants of the spread and the creditworthiness.

The second strand of the literature uses ordered response models trying to solve some limitations of the previous approach. Thus, in linear regression it is assumed that credit ratings categories are equally separated. It means that the difference in risk among all the rating categories is the same. This statement is not necessarily true. In fact, credit rating agencies specify that there is an important difference in the perceived risk of a country if it is categorized as investment grade or not. In the case of ordered response models, differences between the rating categories are estimated in the model itself. To treat ordered variables as continuous could cause errors in the inference (Afonso, Gomes & Rother, 2011; Bessis, 2002; Bissoondoyal-Bheenick, 2005; Mora, 2006; Trevino & Thomas, 2001). Using a simultaneous ordered probit model, Hu, Kiesel and Perraudin (2002) estimates rating transition matrices for the S&P's rated sovereigns and default experience. Bissoondoyal-Bheenick (2005) analyses the quantitative determinants of sovereign ratings provided by S&P and Moody's with an ordered response model. In this study a sample of 95 countries covering the period 1995–1999. According to the author, quantitative measures are only a part of the input into sovereign rating decisions. GNP per capita and inflation seem to be the most relevant economic variables. Moreover, economic variables do not maintain the same importance over the time and between the agencies. Bissoondoyal-Bheenick, Brooks and Yip (2006), compares ordered probit and case-based reasoning techniques for the modeling of the determinants of sovereign ratings. They also find that a proxy for technological development, specifically, mobile phone use, is an important determinant of credit ratings, traditionally not used in previous studies. Hill, Brooks and Faff (2010) employed data for 129 countries in the period 1990–2006 in order to explain differences in the credit quality assessment between S&P, Moody's and Fitch. They find that several variables have varying importance in explaining ratings across agencies using a cumulative probit model. More recently, Andreasen and Valenzuela (2016) investigated the effect of financial openness on corporate and debt ratings, finding that financial openness has a significant impact on credit ratings. Lastly, Teixeira et al. (2017) examines the determinants of sovereign for 1993–2013 using an ordered model. They demonstrate that credit rating differs for crisis and non-crisis periods, and that these differences also vary depending on the regions where the countries belong.

Other approaches have been used as well in this field different from the linear and ordered response models. Thus, Yim and Mitchell (2005) used different statistical techniques as neural networks or self-organizing maps among others to predict country ratings using a sample of 52 countries in 2002. They concluded that hybrid neural networks out-perform all other models including logit and probit models. Bennell, Crabbe, Thomas and Gwilym (2006) also compare the performance of ordered probit models and artificial neural networks (ANN). In this paper, ANN display a higher accuracy predicting sovereign ratings. Van Gestel et al. (2006) describes a process model to develop rating systems taking advantage the strengths of support vector machines in classification tasks. Polito and Wickens (2015) calculates the credit ratings for 14 European countries over the period 1995–2012 based on a model-based measure derived from the fiscal position of a country, calculating the forecast of future debt liabilities and the potential of each country to repay them.

From the previous studies, some conclusions can be extracted: first, many studies focused in the previous generally small set of variables used in the literature. This fact contributes to a reduced exploration of new variables, which it is likely to improve the accuracy of the models. It could be a limitation because credit rating agencies update its methodologies and new some variables can be added. Secondly, models are only based on quantitative variables due the difficulty to know and understand the criteria used by the external rating agencies, although the qualitative expertise of credit rating agencies represents an important part of sovereign ratings. Regarding the use of textual and non-structured data to predict sovereign ratings is scarce. For example, Apergis (2015) examined how news affects credit ratings of three European countries with sovereign debt problems (i.e. Greece, Ireland, and Portugal). It is concluded that news comes from market sources are a good determinant of credit ratings. As of the date of this paper, no articles have been found that attempt to extract text information to predict sovereign credit ratings.

3. Dealing with textual data

In the following Section, the proposed framework is explained. The model will predict sovereign credit ratings using textual information from European country reports. The proposed model consists of two steps. The first step comprises an unsupervised algorithm (doc2vec model) for learning vector representations of sentences contained in each report. The second step performs the classification task using a logistic regression, using the previous extracted vectors as input variables.

European country reports will be used to extract useful information to predict sovereign ratings. In this sense, this is a common prediction problem. However, the challenge is the way the textual and unstructured information is treated. Textual information should be properly transformed to be included in a classification algorithm. The transformation approach I follow is called "embedding", which assigns a unique vector to each word appearing in the vocabulary included in all the reports.

One of the simplest word embedding approach is named "bag of words", which consists of mapping each word from a vocabulary in a vector containing all zeros, except a 1 at the position corresponding to the index representing the word. Nevertheless, this approach has several drawbacks. Thus, the size and the sparsity of the resulting embedding is large, because it is fixed by the size of the vocabulary. Additionally, this approach does not learn similarities among different words. For example, words like "crisis" and "downturn" would be considered as totally different.

On 2013, Mikolov et al. proposed an efficient neural approach to learning high-quality embedding that overcomes previous model: the word2vec model. Word2vec is based on a three-layer neural network with one input, one hidden and an output layer. There two main network architectures used to train the model as shown in Fig. 1: the Skip-gram and Continuous Bag-of-Words (CBOW) architectures. The CBOW architecture attempts to predict a target word by combining the distributed representation of its surrounding words. On the other hand, Skip-gram attempts to predict the context of a word by using the distributed representation of an input word.

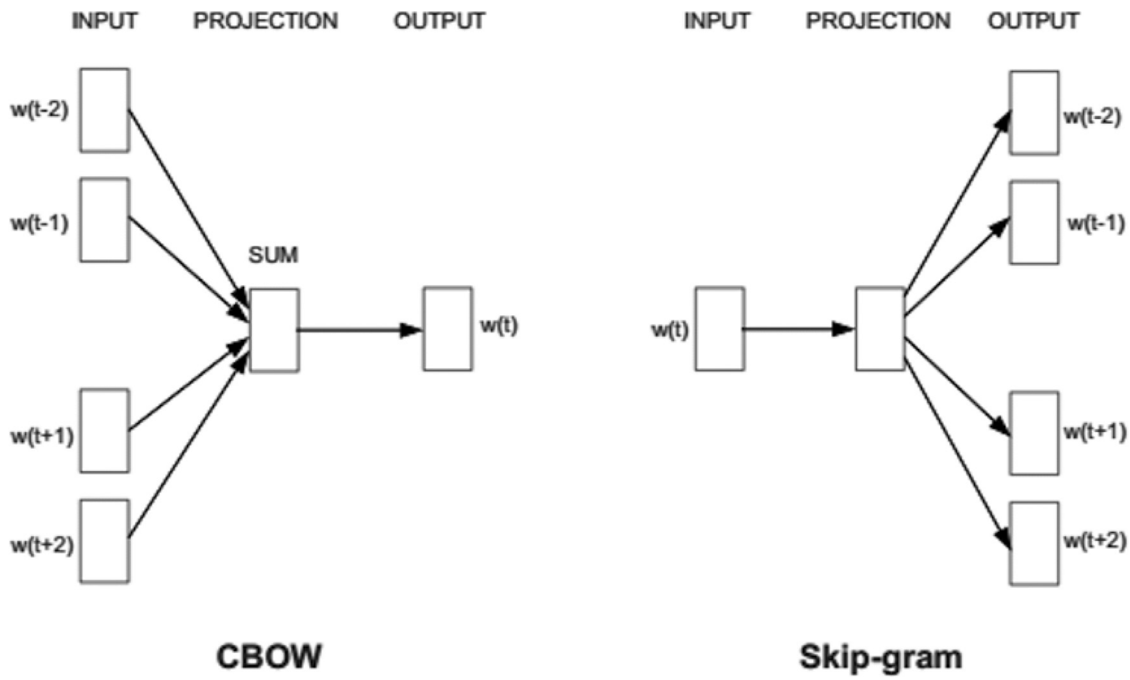

Fig. 1. Continuous Bag-of-Word (CBOW) and Skip-gram models.

Thus, being k the window size or the training context of the current word, the Skip-gram aims to maximize the average of log probability to predict the context words w_{t+j} based on the current word w_t as follows:

$$Skip-gram = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=-k}^{k}\log p\left(w_{t+j}|w_t\right)\right] \tag{1}$$

Whereas using the CBOW approach, the aim is to maximize the log probability of the target word w_j based on the surrounding words in the context:

$$CBOW = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{j=k}^{t-k}\log p\left(w_t|w_{t+j}\right)\right] \tag{2}$$

The output of the Word2vec neural net is a vocabulary in which each word has a numerical vector associated with. A well-trained set of word vectors will place similar words close to each other in that space. For example, words like “crisis” and “downturn” would have very close vectors. Inspired by the success of word2vec, Le and Mikolov (2014) extended word2vec into doc2vec, which produces a vector representation for each document, known as Paragraph vector or document embedding. Mikolov proposed two types of neural network architecture called Distributed Memory Model Paragraph Vectors (PV-DM) and Distributed Bag of Words models (PV-DBOW). PV-DM and PV-DBOW in the doc2vec model are analogous to Continuous bag of words and Skip-gram models in the word2vec model.

In PV-DM the model is trained to predict the center word using context words in a small window and also the paragraph vector as shown in the left side of Fig. 2. On the other hand, in PV-DBOW architecture, the paragraph vector is trained to predict context words directly (right side Fig. 2). The PV-DM model mostly performs better than the PV-BOW models which usually create a very high-dimensional representation leading to poorer generalization.

4. Data and proposed model

In this paper, a two-step model to predict country ratings is proposed. This model is based on textual information provided in the country reports issued by the European Commission. In country reports, the European Commission monitors progress made on social and economic issues in each Member State. There is a country Report per Member State, providing an overview of the general state of the EU’s economy in a quantitative and qualitative way.

These reports are usually published at the end of February of each year. For the model development, all the available reports over the time are collected, specifically, covering the period from 2011 to 2019. There are not existing reports from some European countries for every year. For example, there are not reports for Croatia before year 2013 and, in the case of Greece, for some specific years. Table 1 provides the list of countries with an available report by year. A total of 246 reports are used.

The preprocessing of the data is needed to extract useful information or features to train the model (Feinerer, Hornik & Meyer, 2008). Data cleaning or preprocessing data involves some steps as: removing formating, converting the data into plain text, removing whitespace and numbers, uppercase characters, removing stopwords and, finally, stemming words. Stopwords are defined as words in a language that are so common that their information value is practically null. As all the country reports are available in English, some examples of these stopwords are prepositions, determinants, conjunctions…

On the other hand, the stemming process refers to the process of erasing word suffixes to retrieve the root (or stem) of the words, which reduces the complexity of the data without significant loss of information. Thus, a word as the verb "argue" will be reduce to the stem “argu” regardless the form or complexity of the word in the text. Thus, other forms as "argued", "argues", "arguing", and "argus" are also reduced to the same stem. The stemming procedure reduces the number of words to consider and provides a better frequency representation. After the preprocessing steps, the dataset consists on of 60.516 unique words or tokens.

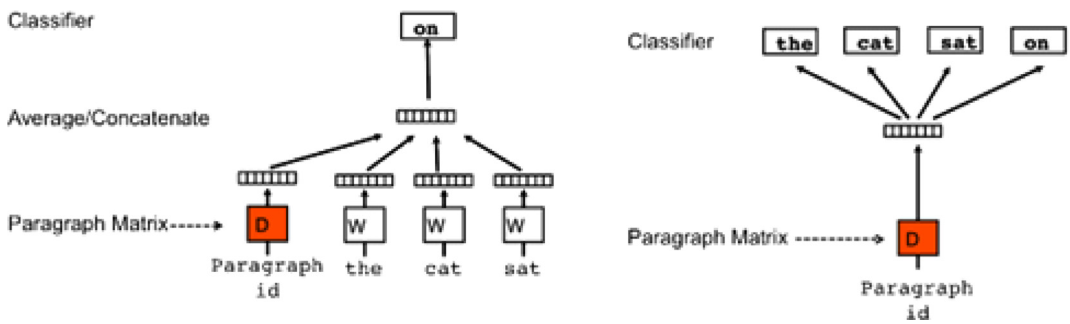

Fig. 2. Distributed Memory Model Paragraph Vectors (left) and Distributed Bag of Words models (right).

Table 1
List of available reports over the time.

Country\Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
Austria	x	X	x	x	x	x	x	x	x
Belgium	x	X	x	x	x	x	x	x	x
Bulgaria	x	X	x	x	x	x	x	x	x
Croatia			x	x	x	x	x	x	x
Cyprus	x	X	x	x	x	x	x	x	x
Czech Republic	x	X	x	x	x	x	x	x	x
Denmark	x	X	x	x	x	x	x	x	x
Estonia	x	X	x	x	x	x	x	x	x
Finland	x	X	x	x	x	x	x	x	x
France	x	X	x	x	x	x	x	x	x
Germany	x	X	x	x	x	x	x	x	x
Greece	x	X	x	x					x
Hungary	x	X	x	x	x	x	x	x	x
Ireland	x	X	x	x	x	x	x	x	x
Italy	x	X	x	x	x	x	x	x	x
Latvia	x	X	x	x	x	x	x	x	x
Lithuania	x	X	x	x	x	x	x	x	x
Luxembourg	x	X	x	x	x	x	x	x	x
Malta	x	X	x	x	x	x	x	x	x
Netherlands	x	X	x	x	x	x	x	x	x
Poland	x	X	x	x	x	x	x	x	x
Portugal	x	X	x	x	x	x	x	x	x
Romania	X	X	x	x	x	x	x	x	x
Slovakia	X	X	x	x	x	x	x	x	x
Slovenia	X	X	x	x	x	x	x	x	x
Spain	X	X	x	x	x	x	x	x	x
Sweden	X	X	x	x	x	x	x	x	x
United Kingdom	X	X	x	x	x	x	x	x	x
Number of countries	27	27	28	28	27	27	27	27	28

The rating to be predicted is the rating on June of the year when each European report is issued. As reports are issued during the first semester of each year, it is assumed that credit ratings will collect the updated opinion of the European Commission and the most recent macroeconomic information at the end of previous year. All the European countries are currently rated by the three CRA's. Moody's is not used because they estimate an expected loss more than a probability of default, as S&P and Fitch do. Credit ratings are summarized into a single label ranging from a AAA (the highest quality) to a D (worst quality). Ratings assigned to the European countries do not cover the whole range of possible ratings in Fitch and S&P rating levels. Therefore, credit ratings have been grouped in only 6 categories to have an enough number of observations for each of the rating levels in the sample and, in this way, to train a robust model. This approach has been commonly used in previous studies on country ratings (Bissoondoyal-Bheenick, 2005; Bennell et al., 2006; Boumparis, Milas & Panagiotidis, 2015; Sehgal, Mathur, Arora & Gupta, 2018). Table 2 displays a higher detail of the mapping between all the rating categories and the six categories which will be finally used.

Differences between ratings assigned by Fitch and S&P are low for European countries. In fact, on December 2019, ratings for 22 of the 28 countries is exactly the same in both rating agencies, and where differences exist, difference does not exceed of two notches. Once the rating mapping displayed in Table 2 is applied, differences are almost removed. As S&P is the credit rating agency with the highest number of rated countries over the world, S&P ratings are used as the dependent variable to be predicted in this paper.[1] Fig. 3 displays the distribution of credit ratings for the European countries from 2011 to 2019 according to S&P. S&P considers ratings as a forward-looking opinion about an obligor's overall creditworthiness.

To train the model, the dataset is divided into 2 partitions: training and test samples. The training sample consist on the 172 country reports, the 70% of the total dataset. The test sample consist on 74 reports. The proportion of credit ratings categories is kept in both samples. The proposed model is a two-step model: first a doc2vec model is trained using the training set. With this first step, each report is transformed in a numeric vector.

[1] The model was also tested using Fitch ratings and results were similar as the obtained using S&P. Thus, there was not observed any bias for the selection of the credit rating agency.

Table 2
Credit ratings mapping used to train the model.

Rating level	Final label
AAA	6
AA+	5
AA	5
AA-	5
A+	4
A	4
A-	4
BBB+	3
BBB	3
BBB−	3
BB+	2
BB	2
BB-	2
B+	1
B	1
B−	1
CCC	1
CC	1
SD/D	1

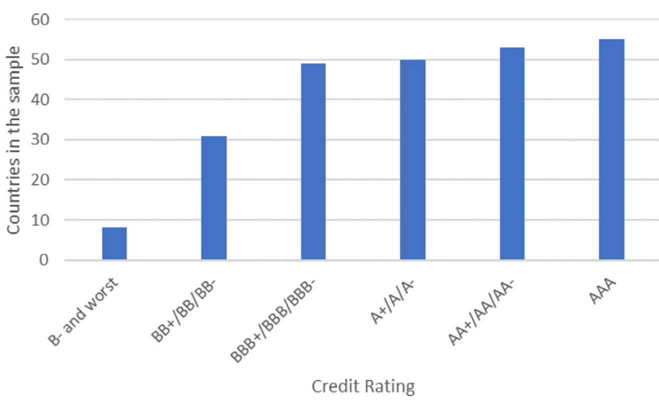

Fig. 3. Distribution of credit ratings for European countries from 2011 to 2019.

The doc2vec model requires of a set of hyper parameters which need to be tuned: the architecture used to train the model (Figs. 1 and 2 architectures), the vector and the window size, the minimum frequency of words to be considered in the training phase, the threshold to down-sample high frequency words, the number of negative word samples and the number of training epochs. The results obtained using a cross validation approach are used to optimize the hyper-parameters of the doc2vec model. The optimal configuration of these hyper parameters results in a doc2vec model where each country report will be transformed to a numerical vector of size 100.

Once each report is transformed, vectors are used as features to train a logistic regression, which dependent variable is the credit rating assigned by S&P. The accuracy of the proposed model is measured comparing predicted ratings with the real ones. Tables 3 and 4 display the results for the training and test samples respectively.

The model predicts exactly the 98.25% of ratings included in the training sample, predicting correctly the 70.27% of country ratings in the test sample. All the misclassifications are found within one rating notch in the training sample and until two notches in the test sample for two countries. Table 4 also exhibits how the model has some difficulties to classify countries placed in the rating category “3”, which contains BBB+, BBB and BBB- rating levels. These ratings establish the boundaries between investment and speculative grades. According to S&P,[2] the rating category which contains BBB+, BBB and BBB- display, historically, a yearly variability greater than observed for the rest of rating grades and where upgrades and downgrades are broadly balanced in the shorter time horizons, maybe making more difficult the classification for the model. Thus, a country can be upgraded one notch and to be downgraded again in the next year, returning to the initial rating. It is also observed some misclassifications for the worst and best rating grades. For example, for 5 countries with a rating AAA, the model assigns rating level one notch below. An additional weakness is that vectors are not interpretable, it means that it is not possible to identify what subjective factors, sentences or words in the reports explaining the resulting predictions, once reports have been transformed into vectors.

5. Rating models including macroeconomic information

In this Section, first, a new model based only in a set of macroeconomic variables is developed. This model, an ordered logit model, following the most common approach in the literature. Then, a new model will be trained, but this time, combining the macroeconomic information and previous vectors obtained from the doc2vec model. The former model serves as a benchmark of the model developed in previous section, while the later model tries to assess if the combination of quantitative variables and the subjective information extracted from reports result in a more predictive model.

For both models, a set of 14 macroeconomic variables are selected to be included in the models for the same period, covering from year 2011 to 2019. These variables are used by the credit rating agencies, especially by S&P, which is the rating agency of reference in this paper, and they have been relevant explanatory variables of sovereign ratings in previous analogous research (Armstrong, De Kervenoael, Li & Read, 1998; Dreisbach, 2007; Manasse & Roubini, 2009; Yim & Mitchell, 2005). The following variables are used: consumer prices (% annual change), GDP (% annual change), international reserves, inward foreign direct investment/GDP, terms of trade, budget balance (% of GDP), Public debt (% of GDP), GDP per head and a set of six governance indicators (control of corruption, government effectiveness, political stability and absence of violence/terrorism, regulatory quality, rule of law and voice and accountability indicators). Variables are extracted from the Worldbank databases.

A descriptive analysis of the selected variables is shown in Table 5, which also includes the expected sign (E.S.) of each variable from a univariate perspective on the creditworthiness of a country, where a positive sign means that the creditworthiness increases when the ratio increases. The sample of macroeconomic data is divided in a train and test samples in the same way as done in the model using country reports. Missing values present in some of the macroeconomic information are replaced by the mean of each variable calculated on the train sample. All the variables are then linearly scaled to have a zero mean and unit variance.

Regarding the first model, based only in macroeconomic information, Section 2 states that two main approaches are used to predict sovereign ratings: linear regression methods on a numerical representation of the ratings and ordered models. Although predictive enough in previous studies, linear regression models display more conceptual and methodological weaknesses than ordered response models and it is not recently very used. That’s because in this section and for the sake of simplicity, only an ordered logit model is estimated.[3]

[2] Default, Transition, and Recovery: 2018 Annual Sovereign Default And Rating Transition Study.

[3] Results of the linear regression model are available on request.

Table 3
Confusion matrix in the training sample.

Obs./Pred.	Below B+ (1)	BB+/BB/BB- (2)	BBB+/BBB/BBB (3)	A+/A/A (4)	AA+/AA/AA+ (5)	AAA (6)
Below B+ (1)	**6**	0	0	0	0	0
BB+/BB/BB− (2)	0	**21**	0	0	0	0
BBB+/BBB/BBB (3)	0	0	**34**	0	0	0
A+/A/A (4)	0	0	2	**33**	0	0
AA+/AA/AA+ (5)	0	0	0	1	**40**	0
AAA (6)	0	0	0	0	0	**35**

Table 4
Confusion matrix in the test sample.

Obs./Pred.	Below B+ (1)	BB+/BB/BB- (2)	BBB+/BBB/BBB (3)	A+/A/A (4)	AA+/AA/AA+ (5)	AAA (6)
Below B+ (1)	**1**	1	0	0	0	0
BB+/BB/BB- (2)	0	**8**	1	1	0	0
BBB+/BBB/BBB (3)	0	3	**6**	6	0	0
A+/A/A (4)	0	0	3	**12**	0	0
AA+/AA/AA+ (5)	0	0	1	1	**10**	0
AAA (6)	0	0	0	0	5	**15**

Table 5
Descriptive analysis of selected variables.

Variable number	Variable Name	Nobs	NAs	Minimum	Maximum	1. Quartile	3. Quartile	Mean	Median	Stdev	Skewness	Kurtosis	E.S
1	Consumer prices (% annual change)	246	0	−2.05	31.58	0.44	2.52	1.81	1.50	2.71	6.18	59.94	–
2	GDP (% annual change)	246	0	−9.20	25.02	1.20	3.50	2.17	2.13	2.87	1.55	16.84	+
3	International reserves	246	1	207.47	248,856	3445	63,144	47,164	24,943	56,763	1.43	1.15	+
4	Inward foreign direct investment/GDP	246	0	−672.90	1355.50	1.15	5.15	24.69	2.33	144.16	5.80	48.02	+
5	Terms of trade	246	18	40.08	117.60	89.52	101.77	93.65	97.94	14.22	−1.79	3.61	+
6	Budget balance (% of GDP)	246	0	−32.06	3.42	−4.20	−0.43	−2.74	−2.39	3.56	−2.84	17.92	+
7	Public debt (% of GDP)	246	0	6.07	181.77	40.15	85.93	66.82	63.39	33.93	0.62	0.27	–
8	GDP per head	246	0	0.01	0.12	0.02	0.05	0.03	0.03	0.02	1.52	3.10	+
9	Control of Corruption Index	246	0	−0.27	2.40	0.33	1.67	1.03	0.95	0.78	0.09	−1.26	+
10	Government Effectiveness Index	246	0	−0.33	2.24	0.81	1.58	1.14	1.12	0.55	−0.34	−0.37	+
11	Political Stability and Absence of Violence/Terrorism Index	246	0	−0.32	1.46	0.49	1.02	0.74	0.78	0.37	−0.43	−0.32	+
12	Regulatory quality index	246	0	0.24	2.05	0.83	1.64	1.20	1.16	0.45	0.02	−1.19	+
13	Rule of Law Index	246	0	−0.11	2.10	0.77	1.74	1.16	1.13	0.60	−0.26	−0.96	+
14	Voice and Accountability Index	246	0	0.31	1.69	0.96	1.37	1.11	1.11	0.34	−0.45	−0.44	+

Table 6 shows the variables, expected sign, coefficients, standard errors and significance of the ordered logit model based only on macroeconomic variables. The standard errors are based on the Eicker-White method, which produces consistent values whether or not the assumption of the homoscedasticity of the residuals holds (Bennell et al., 2006; Trevino & Thomas, 2001). According with Table 6 a total of 8 economic variables are significant across the six rating categories, specifically, international reserves, inward foreign direct investment/GDP, budget balance (% of GDP), public debt (% of GDP), GDP per head, government effectiveness, regulatory quality and voice and accountability index. From this list, only one variable shows a wrong sign compared with the expected one: regulatory quality index explained by the high correlation with the rest of governance indicators. The annual change of European growth is not significant due to the special situation of the euro area during the considered period and the volatility of the variable over these years, where countries with a weaker economy or with deeper unbalances are growing higher than the average of the euro area. Countries with lower ratings such as Cyprus, Hungary or Portugal display a GDP growth higher than a 2% in 2018, while other countries with high credit ratings such as France or Germany do not exceed a growth of 1.5% in the same year. Other variables such as terms of trade or inward foreign direct investment ratio do not display high differences among the European countries. The guidelines provided by credit rating agencies are commonly used to establish a rank ordering of countries around the world. As this model is developed on a specific economic area, it is not surprisingly that some variables are not very significant.

The −2 log likelihood for the model is 231.37 with a significant chi-square statistic, which indicates that the model gives a significant improvement over the baseline intercept only model. In the case of logistic and ordinal regression models, it not possible to compute the same R squared statistic as in linear regression so approximation should be computed instead. In this model, the McFadden pseudo R squared reaches the 60.06%. One of the assumptions underlying ordered models is that the relationship between each pair of outcome groups is the same, which is tested using the test of parallel lines. Parallel line test was carried out finding no enough evidence to reject the main assumption of model specification, therefore proportional odds assumption appears to have held for the model.

Table 6
Results based on an ordered logit model using macroeconomic variables.

Variable number	Variable name	E.S	Coefficients	t-statistics	p-value
1	Consumer prices (% annual change)	–	−0.051	−0.292	0.771
2	GDP (% annual change)	+	−0.291	−1.428	0.153
3	International reserves	+	1.985	6.097	0.000
4	Inward foreign direct investment/GDP	–	−1.032	−2.522	0.012
5	Terms of trade	+	0.227	0.964	0.335
6	Budget balance (% of GDP)	+	0.450	2.094	0.036
7	Public debt (% of GDP)	–	−3.865	−7.806	0.000
8	GDP per head	+	2.595	3.647	0.000
9	Control of Corruption Index	+	−1.000	−0.985	0.325
10	Government Effectiveness Index	+	2.028	3.074	0.002
11	Political Stability and Absence of Violence/Terrorism Index	+	−0.140	−0.447	0.655
12	Regulatory quality index	+	−1.533	−2.950	0.003
13	Rule of Law Index	+	0.946	1.071	0.284
14	Voice and Accountability Index	+	2.488	3.167	0.002
	Cut 1|2		−11.729	−7.045	0.000
	Cut 2|3		−7.492	−5.559	0.000
	Cut 3|4		−3.835	−3.571	0.000
	Cut 4|5		−0.634	−2.215	0.028
	Cut 5|6		4.356	3.643	0.000
	Accuracy (train sample)		71.51%		
	Accuracy (test sample)		67.57%		

In terms of explanatory power, the model exactly predicts 71.51% of country ratings in the training set and a 67.57% of ratings in the test sample. There is not a high difference comparing the accuracy of this model with the proposed model, based on country reports. The proposed model improves this benchmark model in only two countries measured in the test sample. Results are quite interesting because in the country reports there are unstructured textual data and opinions provided by the European Commission.

Finally, a final model is trained combining both macroeconomic variables and the resulting vectors of the doc2vec model in order to check if a model with different information improves the prediction accuracy of individual models. Although it was initially expected a better model, where macroeconomic variables are complemented with a more subjective information, results do not indicate a significant improvement. The accuracy reached in the training sample reaches the 100%., but this value decreases in the test sample, reaching an accuracy of 71.62%, which only improves in only one country the accuracy obtained using only the country reports. Results are mainly explained because information obtained in the country reports already collects all the information contained in the macroeconomic variables, being not so complement as initially expected. Country reports in fact contains all the macroeconomic information with additional explanations and forward looking opinion about the current situation. Therefore, macroeconomic information does not add much more value in this case.

6. Concluding remarks

In recent years, the number of credit ratings and its importance has increased considerably. Credit ratings summarize into a single label the government's ability and willingness to service its debts in full and on time where an implicitly probability of default can be associated. In a rating process, many factors are considered, including political, growth, external debt, financial sector, structure of the public and private sector, social development and trade among others. Macroeconomic variables have been a good proxy to understand and predict credit ratings in the past. Nevertheless, this information can only explain a part of the rating process.

In this paper, the importance of this unstructured and qualitative information to explain the Standard and Poor's credit ratings is assessed. For this purpose, the yearly country reports issued by the European Commission for the European Member States are used. In Country Reports, the European Commission monitors progress made on social and economic issues in each Member State, providing an overview of the general state of the EU's economy in a quantitative and qualitative way.

The proposed model firstly uses a shallow neural network to convert each report in an n-dimensional space vector. It means that all the quantitative and qualitative information contained in the reports will be summarized in a numerical vector. Once reports are transformed, they are used as an input of a logistic regression trying to predict country ratings.

This model slightly overcomes the accuracy of a model based only on a set of macroeconomic variables, the most common approach in the literature. The accuracy of the proposed model increase when macroeconomic data and the information contained in the country reports is combined. This paper demonstrates the importance of qualitative information as a determinant of credit ratings and the potential of natural language processing to mitigate credit risk. Finally, this paper opens new research lines to test the effect of using different embedding models in the accuracy of the model, or even the use of pre-trained models for predicting ratings, not exclusively for countries.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

CRediT authorship contribution statement

Ivan Pastor Sanz: Conceptualization, Methodology, Data curation, Writing - original draft, Visualization, Investigation, Validation, Writing - review & editing.

References

Afonso, A. (2003). Understanding the determinants of government debt ratings: Evidence for the two leading agencies. *Journal of Economics and Finance, 27*(1), 56–74.
Afonso, A., Gomes, P. M., & Rother, P. (2011). Short and long-run determinants of sovereign debt credit ratings. *International Journal of Finance and Economics, 16*(1), 1–15.

Alexe, S., Hammer, P. L., Kogan, A., & Lejeune, M. A. (2003). A non-recursive regression model for country risk rating. *Rutgers University Research Report, 9*, 1–40.
Andreasen, E., & Valenzuela, P. (2016). Financial openness, domestic financial development and credit ratings. *Finance Research Letters, 16*, 11e18.
Apergis, N. (2015). Newswire messages and sovereign credit ratings: Evidence from European countries under austerity reform programmes. *International Review of Financial Analysis, 39*, 54–62.
Armstrong, H., De Kervenoael, R. J., Li, X., & Read, R. (1998). A comparison of the economic performance of different micro-states, and between micro-states and larger countries. *World Development, 26*(4), 639–656.
Bennell, J. A., Crabbe, D., Thomas, S., & Gwilym, O. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications, 30*(3), 415–425.
Bernal, O., Girard, A., & Gnabo, J. Y. (2016). The importance of conflicts of interest in attributing sovereign credit ratings. *International Review of Law and Economics, 47*, 48e66.
Bessis, J. (2002). *Risk management in banking* (Second ed.). West Sussex: Wiley.
Bissoondoyal-Bheenick, E. (2005). An analysis of the determinants of sovereign ratings. *Global Finance Journal, 15*(3), 251–280.
Bissoondoyal-Bheenick, E., Brooks, R., & Yip, A. Y. N. (2006). Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches. *Global Finance Journal, 17*(1), 136–154.
Boumparis, P., Milas, C., & Panagiotidis, T. (2015). Has the crisis affected the behavior of the rating agencies? *Panel evidence from the Eurozone. Economics Letters, 136*, 118–124.
Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic Policy Review, 2*(2), 5–96.
Chen, S. S., Chen, H. Y., Chang, C. C., & Yang, S. L. (2016). The relation between sovereign credit rating revisions and economic growth. *Journal of Banking & Finance, 64*, 90e100.
D'Agostino, A., & Lennkh, R. A. (2016). Euro area sovereign ratings: An analysis of fundamental criteria and subjective judgement. *ESM Working Paper Series No.* 14/2016.
Dreisbach, D. (2007). A note on Chui, Gai and Haldane's sovereign liquidity crisis: Analytics and implications for public policy. *Journal of Banking and Finance, 32*(4), 624–629.
Eliasson, A.C. (.2002). Sovereign credit ratings. Deutsche Bank Research No. 02-1.
Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1–54.
Fitch (2014). Sovereign rating criteria: Master criteria.
Gärtner, M., Griesbach, B., & Jung, F. (2011). PIGS or lambs? the European sovereign debt crisis and the role of rating agencies. *International Advances in Economic Research, 17*(3), 288–299.
Haan, J.D., & Amtenbrink, F. (2011). Credit rating agencies. SSRN eLibrary DNB Working Paper No. 278.
Hill, P., Brooks, R., & Faff, R. (2010). Variations in sovereign credit quality assessments across rating agencies. *Journal of Banking and Finance, 34*, 1327–1343.
Hu, Y. T., Kiesel, R., & Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance, 26*(7), 1383–1406.
Kiff, J., Nowak, S.B., .& Schumacher, L.B. (.2012). Are rating agencies powerful? An investigation into the impact and accuracy of sovereign ratings. IMF Working Paper No. 12/23.
Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning* (pp. 1188–1196). Beijing, China.
Lee, S. H. (1993). Are the credit ratings assigned by bankers based on the willingness of ldc borrowers to repay? *Journal of Development Economics, 40*, 349–359.
Manasse, P., & Roubini, N. (2009). Rules of thumb for sovereign debt crises. *Journal of International Economics, 78*(2), 192–205.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *ICLR Workshop Papers.*.
Monfort, B., & Mulder, C. (2000). *Using credit ratings for capital requirements on lending to emerging market economies - Possible impact of a new Basel accord* Technical Report 00/69. IMF Working Papers.
Moody's (2013). Rating methodology: Sovereign bond ratings.
Mora, N. (2006). Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking and Finance, 30*(7), 2041–2062.
Pagano, M., & Volpin, P. F. (2010). Credit ratings failures and policy options. *Economic Policy. Wiley Online Library, 25*, 401–431.
Polito, V., & Wickens, M. (2015). Sovereign credit ratings in the European Union: A model based fiscal analysis. *European Economic Review, 78*, 220–247.
Reinhart, C. (2002b). Credit ratings, default, and financial crises: Evidence from emerging markets. *World Bank Economic Review, 16*(2), 151–170.
Reinhart, C., Levich, R., & Majoni, G. (2002). *Ratings, rating agencies and the global financial system: Summary and policy implications*. Massachusetts: Kluwer Academic.
Rowland, P., & Torres, J. L. (2004). Determinants of spread and creditworthiness for emerging market sovereign debt: A panel data study. *Banco de la República de Colombia*, 296.
Sehgal, S., Mathur, S., Arora, M., & Gupta, L. (2018). Sovereign ratings: Determinants and policy implications for India. *IIMB Management Review*, 140–159.
Standard & Poor's (2014). Sovereign rating methodology.
Standard & Poor's. (2014b). *Sovereign rating methodology. Technical Report. Standard and Poor's*.
Trevino, L., & Thomas, S. (2001). Local versus foreign currency ratings: What determines sovereign transfer risk? *The Journal of Fixed Income*, 65–75 June.
Van Gestel, T., Baesens, B., Van Dijcke, P., Garcia, J., Suykens, J., & Vanthienen, J. (2006). A process model to develop an internal rating system: Sovereign credit ratings. *Decision Support Systems, 42*(2), 1131–1151.
Vernazza, D. R., & Nielsen, E. F. (2015). The damaging bias of sovereign ratings. *Economic Notes, 44*(2), 362–408 Volume.
Yim, J., & Mitchell, H. (2005). Comparison of country risk models: Hybrid neural networks, logit models, discriminant analysis and cluster techniques. *Expert Systems with Applications, 28*(1), 137–148.