Article

Forecasting Credit Ratings of EU Banks

Vasilios Plakandaras [1,*], **Periklis Gogas** [1], **Theophilos Papadimitriou** [1], **Efterpi Doumpa** [2] **and Maria Stefanidou** [2]

[1] Department of Economics, Democritus University of Thrace, 69100 Komotini, Greece;
pgkogkas@econ.duth.gr (P.G.); papadimi@econ.duth.gr (T.P.)

[2] School of Economics, Business Administration and Legal Studies, International Hellenic University,
57001 Thessaloniki, Greece; e.doumpa@outlook.com (E.D.); stefanidoum94@gmail.com (M.S.)

* Correspondence: vplakand@econ.duth.gr

check for updates

Abstract: The aim of this study is to forecast credit ratings of E.U. banking institutions, as dictated by Credit Rating Agencies (CRAs). To do so, we developed alternative forecasting models that determine the non-disclosed criteria used in rating. We compiled a sample of 112 E.U. banking institutions, including their Fitch assigned ratings for 2017 and the publicly available information from their corresponding financial statements spanning the period 2013 to 2016, that lead to the corresponding ratings. Our assessment is based on identifying the financial variables that are relevant to forecasting the ratings and the rating methodology used. In the empirical section, we employed a vigorous variable selection scheme prior to training both Probit and Support Vector Machines (SVM) models, given that the latter originates from the area of machine learning and is gaining popularity among economists and CRAs. Our results show that the most accurate, in terms of in-sample forecasting, is an SVM model coupled with the nonlinear RBF kernel that identifies correctly 91.07% of the banks' ratings, using only 8 explanatory variables. Our findings suggest that a forecasting model based solely on publicly available financial information can adhere closely to the official ratings produced by Fitch. This provides evidence that the actual assessment procedures of the Credit Rating Agencies can be fairly accurately proxied by forecasting models based on freely available data and information on undisclosed information is of lower importance.

Keywords: credit ratings; machine learning; Support Vector Machines; banks

1. Introduction

Credit Rating Agencies (CRAs) have been around for more than 150 years. Their role progressed from simple information collectors to quasi-official evaluators of credit risk throughout the modern global financial system. CRAs were originally paid by potential investors to compile financial information and data at a time when such a service was too difficult and costly. Nonetheless, after the 1929 big market crash CRAs started to play a more formal role in the financial system. The stricter rules that were imposed by regulators with the Glass–Steagal Act in the mid-1930s limited banking, insurance and other financial institutions to only invest in "investment grade" securities, assessed by the CRAs. Since then, we have seen a growing reliance on CRAs ratings as they are increasingly incorporated in private contracts, investment guidelines for pension funds, endowment funds and other private entities that all came to rely on these CRAs ratings.

In the aftermath of the 2008 global financial crisis, the role of CRAs evolved to an increasingly important albeit a questionable one; they provide important financial information to market participants, mainly by issuing ratings on the probability of default for specific debt issuers. In recent years, there is an increased interest in the credit ratings process and specifically on the actual criteria used by the

CRAs to evaluate debt issuers. Focusing on banking institutions, rating agencies provide publicly available ratings associated with the ability of a banking institution to meet debt obligations on time. The supervisory framework of the Basel II and the Basel III accords expanded the role of credit rating agencies. As a result, now banks are required to calculate their risk weighted assets (RWA). This is done by either using the ratings provided by the CRAs or creating their own internal ratings approach. In either case, ratings are now in the epicenter of risk assessment and the resulting capital requirements for the banking institutions that are used by the supervising authorities.

Nevertheless, the experience from the 2008 financial crisis dictates that CRAs may underreact to financial information or significantly delay in downgrading debt issuers. In many occasions, they only downgrade a debt issuer long after the markets do. This happens especially in the case of "too-big-to-fail" financial and banking institutions. Long before the 2008 crisis, their integrity was under scrutiny in other major corporate collapses as well, such as the Enron failure in the U.S., the Asian financial crisis and Parmalat scandal and the subsequent failure in Europe. All these corporations were assessed and assigned high ratings just a few days before their collapse. The same was true even in major sovereign debt crises like the debt crises of Greece, Spain, Portugal and Italy in the early 2010's. The Securities and Exchange Commission (SEC) in a 2011 annual report on credit raters found "apparent failures" at each of the 10 credit rating agencies they examined. These included Standard & Poor's (S&P), Moody's, and Fitch, the "big three" credit rating agencies that hold 95% of the corresponding global ratings market. On top of these concerns, the fact that the criteria of the assessments are not fully disclosed and transparent, augments the mistrust on the actual quality of the ratings.

Based on the above criticism, the regulatory authorities in the U.S. decided to take some regulatory action in 2008 and since require the public disclosure of the information a CRA uses to determine a rating on a structured product. In response to the inability of CRAs to properly appreciate the risks in complex financial instruments before 2008, the European Commission strengthened the regulatory and supervisory framework for CRAs in the E.U. The new E.U. rules were introduced in three consecutive rounds. The first round of rules, which came into force at the end of 2009, established a regulatory framework for CRAs and introduced a regulatory oversight regime, whereby CRAs had to be registered and supervised by national competent authorities. In addition, CRAs were required to avoid conflicts of interest, and to have sound rating methodologies and transparent rating activities. In 2011, these rules were amended to take into account the creation of the European Securities and Markets Authority (ESMA), which supervised CRAs registered in the E.U. A further amendment was made in 2013 to reinforce the rules and address weaknesses related to sovereign debt credit ratings. In the case of European banks, total bank debt issuance amounted to approximately €881 billion in 2020. That included €220 billion of corporate bonds, €129 billion of medium-term notes, €160 billion of short-term debt and €510 billion of covered bonds (European Central Bank (European Central Bank ECB).

This study seeks to find the most important factors contributing to the ratings of European banking institutions. This was done using only publicly available information from the published financial statements of all banks. In doing so, we compiled a dataset of 112 E.U. banking institutions and attempted to fit a Probit and a Support Vector Machines (SVM) model in order to pinpoint the variables used in the true rating procedure. Thus, our focus was to use the publicly available data in order to accurately forecast the assigned credit ratings from the CRAs. To the best of our knowledge, our paper is the first that attempts to pinpoint the criteria used by rating agencies to assess the resilience of E.U. banks.

The remainder of the paper is organized as follows. Section 2 reviews the literature. In Section 3 we describe the data and the methodology, while the empirical findings are presented in Section 4. Section 5 concludes the paper.

2. Literature Review

While bank ratings are used extensively as explanatory variables in the economic literature, the nature of ratings per se remains largely ill-examined. A paper closely related to this study is Gogas et al. (2014) who examined the ratings of the Fitch Rating Agency for the case of 92 U.S. banks. The authors used ordered logit models to forecast bank credit ratings based on publicly available financial statements of the banks. Their empirical findings suggested that almost 84% of the actual ratings can be matched based on publicly available information. Bissoondoyal-Bheenick and Treepongkaruna (2011) analyzed the quantitative determinants of bank ratings, provided by Standard & Poor's, Moody's, and Fitch for U.K. and Australian banks. They instead based their analysis on an ordered probit model and found that accounting variables from the financial statements of banking institutions had more explaining power in identifying banks' ratings than macroeconomic ones. Pagratis and Stringa (2007) conducted an ordered probit analysis in order to evaluate the potential linkage between Moody's bank ratings and bank characteristics such as provisions, profitability, cost efficiency, liquidity, short-term interest rates and bank-size.

From a different perspective, Papadimitriou (2012) explored the clustering properties of 90 financial institutions using a correspondence analysis map. The goal was to correspond clustering groups with ratings from Fitch. The empirical findings support a correspondence between clusters and ratings, though the regions corresponding to the ratings are highly overlapped. Credit ratings have also been explored with the use of machine learning methods. Ravi et al. (2008) argued that almost 83.5% of bank failures can be foreseen based on a Support Vector Machine (SVM) model that utilized the information of 54 financial variables on a sample of 1000 banks, over the period 2005–2008.

Although the issue of identifying the exact structure of credit ratings for banks has not been studied to a large effect by the relevant literature, significant relevant papers can be found in the area of predicting bond ratings. Ederington (1985); Pinches and Mingo (1973); Belkaoui (1980) used statistical methods such us logistic regression and multivariable discriminant analysis (MDA) to predict bond ratings. Based on alternative sets of variables the prediction results vary in accuracy between 50% and 70%. Many studies on bond credit rating prediction build forecasting neural networks models (Dutta and Shekhar 1988; Surkan and Singleton 1990; Kim et al. 1993) that are more accurate than typical statistical methods. Moody and Utans (1995) used neural networks to forecast corporate bond ratings based on the ratings of S&P. Using 10 input variables they correctly forecasted 85.2% of the actual ratings. Maher and Sen (1997) compared neural networks to logistic regression models in forecasting bond ratings for the period 1990–1992. The most accurate model achieved 70% on a holdout sample. Kwon et al. (1997) compared ordinal pairwise partitioning (OPP) with back propagation and conventional neural networks for bond ratings of Korean firms. Using 126 financial variables for the period 1991–1993 they achieved 71–73% via neural networks with OPP and 66–67% via conventional neural networks. Huang et al. (2004) compared back propagation neural networks (BPNN) to SVMs in forecasting corporate credit ratings for the U.S. and Taiwan. The most accurate model was a linear SVM model achieving an 80% of correct bond classification.

He et al. (2012) examined the relationship between ratings and the business cycle on mortgage-backed securities (MBS) spanning the period from 2000 to 2006 and their respective ratings from Moody's, S&P and Fitch. The idea was that large financial institutions will persuade CRAs to issue a higher rating than the one dictated by the rating methodology. This discrepancy should be visible when the price of securities sold by big issuers drops more than the price of securities sold by small issuers (keeping everything else fixed). The empirical findings provided evidence in favor of a favorable rating for larger issuers in comparison to small ones, especially during the market boom period of 2004–2006. Hau et al. (2013) extended the previous setting to the banking sector, using a cross-sectional sample of 39,000 banking institution ratings for the period 1990–2011 from Moody's, S&P, and Fitch. The authors concluded that large banks systematically received higher ratings than they should have actually received. An important factor in this favorable rating scheme is the provision of large securitization to CRAs that affects the final outcome of the rating. This phenomenon is more

prevalent during economic booms, when the risk of reputational loss is lower. The erosion of the rating system due to the aforementioned practice leads to the adverse phenomenon where the upper investment grade range does not reflect expected default probabilities, i.e., a higher rating does not necessarily correspond to a lower risk of default.

From a different perspective, Kraft (2015) compared ratings to issuers with rating-based performance-priced loan contracts to issuers with contracts based on accounting ratios and other loan agreements. The study examined adjustments to ratings, i.e., the difference between actual rating and the hypothetical rating implied by reported financials. The study found that, after an adverse economic shock, the adjustments made for firms with rating-based contracts are more favorable than for firms with other types of contracts. This finding is consistent with the hypothesis of rating catering and suggests that reputational concerns are not sufficient to fully eliminate this phenomenon.

3. Data and Methodology

3.1. The Data

For our analysis we used a cross-section of 112 European banking institutions over the period 2013–2017. In order to train our forecasting models, we compiled observations for 34 variables from the Bank-Focus/Orbis[14] database that originate from the banks' financial statements up to 4 years prior to the 2017 actual rating grade. Thus, counting the lags of the 34 independent variables, we compiled a total of 136 explanatory variables considered as possible forecasters of bank ratings, where each lag was treated as an independent variable. The motivation for selecting up to 4 years of data prior to the 2017 Fitch rating stemmed from the fact that, as discussed in the introduction section, CRAs often react to the information reflected in financial statements with a delay. We obtained and used the ratings from Fitch for the year 2017 and the financial statements for the period 2013–2017, due to data availability issues. The independent variables can be classified into four general categories: Assets, Liabilities, Income statement and Financial Ratios. In Table 1 we report the compiled financial variables used as independent variables (or features in the machine learning terminology) to our models.

Table 1. Financial Variables.

No	Abbreviation	Description
Panel A: Assets		
1	TASSET	Total Assets
2	LO	Loans
3	GRLO	Gross loans
4	CBCB	Cash& Balances at Central Bank
5	LASSET	Liquid assets
Panel B: Liabilities		
6	DSF	Deposits and Short-term funding
7	EQ	Equity
8	TCDE	Total customer deposits
9	OIBL	Other interest-bearing liabilities
10	BDE	Bank deposits

Table 1. *Cont.*

No	Abbreviation	Description
Panel C: Income and Expenses		
11	NI	Net Income
12	NIM	Net interest margin
13	NIR	Net interest revenue
14	PBT	Profit before tax
15	OPIN	Operating income
16	ITEX	Income tax expense
17	OPPR	Operating profit
18	TOE	Total operating expenses
19	NOR	Net operating revenues
20	TIP	Total interest paid
21	TIR	Total interest received
Panel D: Financial Ratios		
22	NLTA	Net loans/Total assets
23	NLDSF	Net loans/Deposits and Short-Term funding
24	NLTDB	Liquid assets/Total deposits and borrowed
25	LADSF	Liquid assets/Deposits and Short-Term funding
26	LATDB	Liquid assets/Total deposits and borrowed
27	NIRAA	Net interest revenues/Average assets
28	OOPIAA	Other operating income/Average assets
29	NOEAA	Non-interest expenses/Average assets
30	ROAE	Return On Average Equity (ROAE)
31	ROAA	Return On Average Assets (ROAA)
32	ETA	Equity/Total assets
33	ENL	Equity/Net loans
34	EL	Equity/Liabilities

The dependent variable is ordinal and it is grouped in our case in four classes. These are assigned integer values from 0 to 3, such that lower values indicate a lower rating. The groupings of the four classes are depicted in Table 2.

Table 2. Grouping of Bank ratings in classes.

Class Identification	Rating Category					Number of Banks
3	AAA	AA–	AA	A+		24
2	A–	A	BBB+			34
1	BBB–	BBB				32
0	BB+	BB–	BB	B+	B–	22
	Total					112

The grouping is performed is such a way so that the four identified classes contain a quasi- balanced number of banking institutions, forming a balanced dataset that avoids micronumerosity issues.

3.2. Support Vector Machines

Support Vector Machines is a supervised machine learning method used in data classification. The basic concept of an SVM is to select a small set of data points from the initial dataset, called Support Vectors (SV), that defines a linear boundary separating the data points in two classes. In what follows we describe briefly the mathematical derivations of the SVM theory.

We consider a dataset of vectors $x_i \in R^2$ ($i = 1, 2, \ldots, n$) belonging to 2 classes (targets[1]) $y_i \in \{-1, +1\}$. If the two classes are linearly separable, we define a boundary as:

$$f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i - b = 0, \ y_i f(\mathbf{x}_i) > 0 \forall i \tag{1}$$

where $\mathbf{w}$ is the weight vector and b is the bias.

This optimal hyperplane is defined as the decision boundary that classifies each data vector to the correct class and has the maximum distance from each class. This distance is often called a "margin". In Figure 1, the SVs are represented with a contour circle, the margin lines (defining the distance of the hyperplane from each class) are represented by solid lines and the hyperplane is represented in the center.
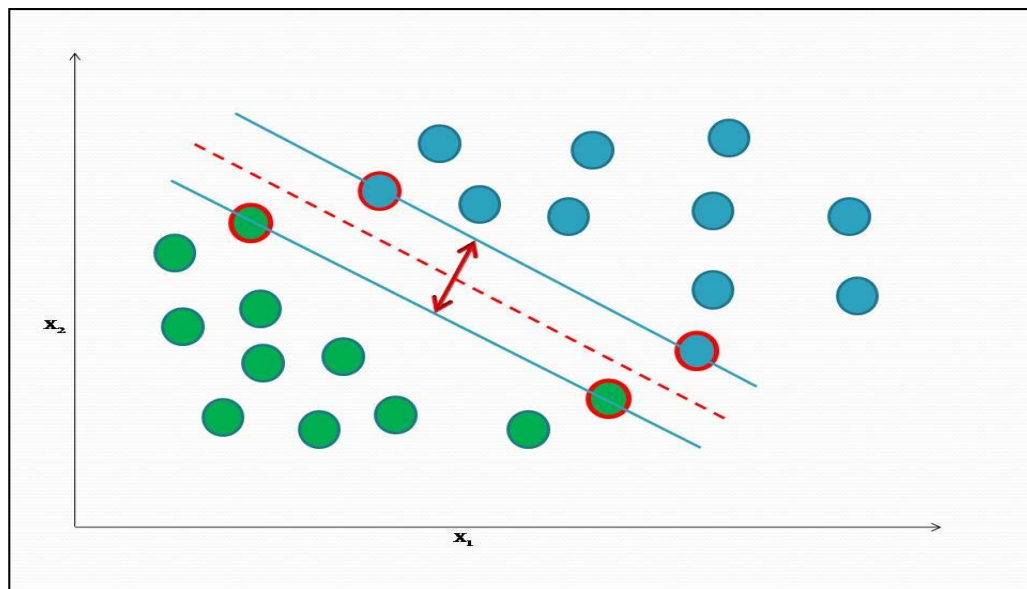

Figure 1. Hyperplane selection and support vectors. The SVs are indicated by the pronounced red circles, the margin lines are represented with the continuous lines, and the hyperplane is represented with the dotted line.

In order to allow for a predefined level of error tolerance in the training procedure, Cortes and Vapnik (1995) introduced non-negative slack variables, $\xi_i \geq 0$, $\forall i$, and a parameter, C, describing the desired tolerance to classification errors. The solution to the problem of identifying the optimal hyperplane can be dealt through the Lagrange relaxation procedure of the following equation:

$$\min_{\mathbf{w},b,\xi} \max_{a\mu} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{j=1}^{N} a_j\left[y_j\left(\mathbf{w}^T\mathbf{x}_j - b\right) - 1 + \xi_j\right] - \sum_{k=1}^{N}\mu_k\xi_k \right\} \tag{2}$$

where ξ_i measures the distance of vector x_i from the hyperplane when classified erroneously, and $a_1, \ldots, an$ are the non-negative Lagrange multipliers.

---

[1] In the SVM jargon.

The hyperplane is then defined as:

$$\hat{\mathbf{w}} = \sum_{i=1}^{N} a_i y_i \mathbf{x}_i \tag{3}$$

$$\hat{b} = \hat{\mathbf{w}}^{\mathrm{T}} \mathbf{x}_i - y_i, \ i \in V \tag{4}$$

where $V = \{i : 0 < y_i < C\}$ is the set of support vector indices.

When the two-class dataset cannot be separated by a linear separator, the SVM is paired with kernel methods. The concept is quite simple: the dataset is projected through a kernel function into a richer space of higher dimensionality (called a feature space), where the dataset is linearly separable. The solution to the dual problem with the projection of Equation (2) now transforms to:

$$\max_{\mathbf{a}} = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k y_j y_k \mathrm{K}(\mathbf{x}_j, \mathbf{x}_k) \tag{5}$$

under the constraints $\sum_{i=1}^{N} a_i y_i = 0$ and $0 \le a_i \le C$, $\forall i$, where $\mathrm{K}(\mathbf{x}_j, \mathbf{x}_k)$ is the kernel function. The SVM model can be extended to a multiclass classification method, using the one-against-the-rest approach; one class is kept aside and all others are grouped to form a new "grouped" class. After measuring the accuracy in forecasting the independent class kept aside, the second one is considered as independent and the others are grouped and so on until all classes are rotated. The overall accuracy is measured as the mean accuracy over all independent classes.

In our models we examined two kernels: the linear kernel and the radial basis function (RBf)[2]. The linear kernel detects the separating hyperplane in the original dimensional space of the dataset, while the RBF projects the initial dataset onto a higher dimensional space. The mathematical representation of each kernel is:

$$\text{Linear} \qquad K_1(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2 \tag{6}$$

$$\text{RBF} \qquad K_2(\mathbf{x}_1, \mathbf{x}_2) = e^{-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2} \tag{7}$$

4. Empirical Findings

4.1. Feature Selection

We identified the variables that contribute the most to the assigned bank ratings following a thorough regression-based variable selection procedure. The selected variables were then fed to both a Probit and an SVM model. As a first step, we measured the correlation coefficient, $r_{i,R}$, between each independent variable i and the assigned rating R. Based on the correlation values, we created six groups of regressors as follows: In group 1, we included all variables with $|r_{i,R}| \ge 0.4$. This resulted in 18 variables in group 1; TASSET, NIM, TIR, NOEAA for the period 2013–2016 and NIRAA for years 2014 and 2016. In a similar manner, in group 2 we did the same for $|r_{i,R}| \ge 0.4$ along with all the lags of variable NIRA. This group included 20 variables. In groups 3 and 4, we included the 30 variables with the highest positive correlation and the 30 variables with the highest negative correlation, respectively. In group 5, the variables included were the five with the highest correlation with the dependent variable and the five with the lowest one, a total of 10 variables. The last group, group 6, contained the entire sample of explanatory variables, a total of 136 features. Table 3 summarizes the variables' groups.

---

2   Our implementation of SVR models is based on LIBSVM (Chang and Lin 2011). The software is available at http://www.csie.ntu.edu.tw/~{}cjlin/libsvm/.

Table 3. Number of variables in each regressor group.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
18 variables	20 variables	30 variables	30 variables	10 variables	136 variables

The next step was to use the selected groups in order to identify the most significant variables in terms of identifying bank ratings. This was done in each group either by: (a) a combinatorial exhaustive search methodology of all possible sets of four variables in each one of the above six groups, hand-picking the ones with the highest R-square and (b) the same process but with all possible sets of eight variables from within each one of the six groups, and (c) a stepwise forward least squares technique where we kept the set of variables with a *p*-value greater than 0.1.

This variable selection procedure produced a total of 18 groups of regressors. Table 4 summarizes the variables selected from each method.

Table 4. Selected variables in each one of the 18 group.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Combinatorial 4	TASSET14	TASSET14	TASSET14	TIR16	TASSET14	TIR16
	TIR16	TIR16	TASSET13	DSF14	TIR16	DSF14
	NOEAA13	NIRAA13	OPPR13	NIRAA13	OIBL13	TASSET15
	NIM13	NOEAA13	OIBL16	NOEAA13	TASSET13	NOR14
	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
Combinatorial 8	TASSET14	TASSET14	TASSET14	DSF14	TASSET14	TIR16
	TIR16	TIR16	TASSET13	NIRAA13	TIR16	TASSET13
	TASSET13	NIRAA13	OIBL16	EQ16	OIBL13	ETA16
	TIR14	NOEAA13	LADSF16	TIR15	TASSET13	LO15
	NOEAA13	TASSET13	LADSF14	GRLO14	TIR14	GRLO15
	NIM13	TIR14	LADSF13	LO14	NOEAA14	NOEAA16
	NOEAA15	NOEAA15	OPPR14	OOPIAA13	TIR15	OOPIAA13
	NOEAA16	NOEAA16	NI13	OOPIAA16	TASSET16	NLTA14
	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
Stepwise-forward	TASSET14	TASSET14	TASSET14	TIR15	TASSET14	TASSET14
	TIR16	TIR16	TASSET13	DSF14	TIR16	TIR16
	NIRAA14	NIRAA14	OPPR13	NIRAA13	OIBL13	DSF14
	NOEAA14	NOEAA14	OIBL16	NOEAA13	TASSET13	PBT13
	(4)	(4)	(4)	EQ16	(4)	NIR13
				(5)		CBCB16
						CBCB13
						OOPIAA14
						TIR14
						NIR14
						(10)

4.2. Ordered Probit Model Results

The above selection procedure resulted in 18 different sets of regressors. These sets were fed to an ordered probit model that forecasts the credit bank rating assigned by Fitch for each institution for the year 2017. The evaluation of the forecasting accuracy of each forecasting model is depicted in Table 5. Each column corresponds to each one of the six groups of the prefiltered regressors while each row presents the forecasting results for the corresponding selection criterion. According to these results, the best accuracy using the probit model for all regressor selection criteria was achieved from the combinatorial search of eight variables from group 6.

Table 5. Bank Rating forecasting accuracy by the probit model (%).

Regressor Selection	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Combinatorial 4	57.14	54.46	50.89	49.11	51.79	54.46
Combinatorial 8	57.14	56.25	58.04	53.57	55.36	**66.07**
Stepwise-forward	57.14	57.14	50.89	46.43	51.79	57.14

Note: The highest accuracy is reported in bold. All values are percentages.

The best accuracy was 66.07% and the variables used were: (a) Total Interest Received 2016—a measure of bank profitability, (b) Total Assets 2013—a measure of capital invested in the bank, (c) Equity/Total Assets 2016—a measure of capital adequacy, (d) Loans 2015 and (e) Gross Loans 2015 both measures of market exposure, (f) Non-interest expenses/Average assets 2016—a feature representing operating efficiency, (g) Other operating income/Average assets 2013—a measure of bank profitability and finally, (h) Net Loans/Total Assets 2014—a measure of market exposure. Most variables refer to market exposure or the profitability of the bank.

In Table 6 we report the contingency table regarding the model's forecasts. The model achieved the best accuracy in class 2 that includes the A-, A, BBB+ ratings, with 70.59%. One might have expected this to be true for class 3, which includes the most creditworthy banking institutions. Thus, the model adheres closer to "mainstream" cases and identifies less accurately the classes 0 and 3. While accurately identifying class 3 can be of less importance, the accurate identification of soon-to-fail banks is of the utmost importance when it comes to financial risk management.

Table 6. Comparison of predicted to real rating categories.

	Correct	Incorect	% Correct	% Incorect
predicted 0	15	7	68.18%	31.82%
predicted 1	20	12	62.50%	37.50%
predicted 2	24	10	**70.59%**	29.41%
predicted 3	15	9	62.50%	37.50%
Total	74	38	66.07%	33.93%

Note: The highest accuracy is reported in bold.

4.3. Support Vector Machines Model Results

We used the same 18 groups of variables to train our SVM models. In this methodological setting, we employed both the linear and the nonlinear RBF (Radial Basis Function) kernel. We followed two popular training schemes. In the first approach, we used a 3-fold cross validation scheme, while in the second one we followed a bootstrapping scheme of 8000 replications.

Cross validation is a common training scheme in machine learning applications. The basic idea is to split the dataset used in training the model into *k* folds of similar length and train iteratively, keeping at each step one fold aside for validation. For instance, in the 3-fold cross validation scheme we started by keeping the first fold aside and tuned the model's parameters based on the second and the third folds. The first fold was used to measure the forecasting accuracy of the trained model. We repeated the procedure by keeping the second and the third fold aside, respectively. The training accuracy of the model is the average over all three folds that were kept sequentially aside. The main advantage of cross validation training is that it avoids overfitting the data; thus, the model adheres more to the data generating mechanism that produces the phenomenon under investigation and less to the specific sample at hand.

In bootstrap training we created a large number of surrogate random samples of the same length as the original samples which they replaced. In this paper we created and trained 8000 samples and

corresponding models. Then, from the distribution of the created forecasts we accessed the median and the 32nd and 68th percentiles in order to estimate the confidence intervals of the forecasts.

The forecasting accuracy of each model is depicted in Table 7 and for the cross validation and the bootstrapping training, respectively. In Panels A and B we present the results the linear and RBF kernels.

Table 7. Bank Rating forecasting accuracy by the Support Vector Machines (SVM) model (%).

Regressor Selection	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Panel A: Linear kernel (k-fold cross validation)						
Combinatorial 4	62.50	63.39	60.71	63.39	61.61	60.71
Combinatorial 8	58.04	64.29	58.04	50.00	63.39	53.57
Stepwise-forward	59.82	62.50	59.82	63.39	**67.86**	66.96
Panel B: nonlinear RBF kernel (k-fold cross validation)						
Combinatorial 4	64.29	63.39	61.61	62.50	63.39	61.61
Combinatorial 8	72.32	76.79	72.32	**91.07**	50.89	66.96
Stepwise-forward	68.75	60.71	68.75	61.61	62.50	79.46

Note: The highest accuracy is reported in bold. All values are percentages.

The most accurate model based on a cross-validation training was an SVM model coupled with the RBF kernel that achieved an accuracy of 91.07% based on Group 4 with a combinatorial search of eight variables. The independent variables of this model were: (a) Deposits and Short term funding 2014—a measure of liquidity, (b) Net interest revenues/Average assets 2013—a measure of operating efficiency, (c) Equity 2016—measuring the capital invested in the bank, (d) Total interest received 2015—a measure of income and profitability, (e) Gross Loans 2014 and (f) Loans 2014—both measures of market exposure and size of operations, and (g) Other operating income/Average assets 2013 and 2014—a ratio indicating the percentage of fees and other income other than interest from loans. Overall, most variables either measure market exposure or the bank's profitability.

The best forecasting model used eight variables. It is interesting that only one of them dated to the most recent ratings year. This was Equity 2016, a measure of the capital invested in the bank or bank size in terms of stockholders' stake. The next most recent forecaster was Total interest received two years prior to the rating. This is a measure of income quality and profitability. It is very interesting that from the other six identified best forecasting variables, four dated three years prior to the ratings and two four years back. The four that dated three years back are Loans 2014, Gross Loans 2014, both measures of core business exposure and size of operations, Deposits and short-term funding 2014, a measure of liquidity and Other operating income/Average assets 2014, relating to income other than the core business of the bank. Finally, two variables dated a full four years prior to the target rating: Net interest revenues/Average assets 2013—a measure of operating efficiency and again Other operating income/Average assets 2013, the non-core income of the bank.

According to these results, it seems that operating efficiency, as it is measured in the financial statements as net interest and other operating income over assets, has a long and lasting effect on the financial health of a banking institution, as this is reflected in its corresponding credit rating. Short term funding (deposits etc.) and the size of the bank's core business (loans, total interest received) affect the ratings over a period of two to three years and the only close rating determinant is total equity. Given that the actual data were classes ranging from zero to three [0, 3], the best forecasting SVM model classified only two banks on the brink of default (class 0) while they actually belong to the highest class 3 and in five cases misclassified five banks of class 0 as class 3 banks. The latter can be considered as a more severe misclassification issue and we will focus in a future version of this manuscript, on utilizing a different kernel that is not based on normal distribution like the RBF kernel.

In Table 8, we report the contingency table of the most accurate SVM model's forecasts.

Table 8. Comparison of predicted to real rating categories.

	Correct	Incorrect	%Correct	%Incorrect
predicted 0	22	0	**100%**	0%
predicted 1	26	6	81.25%	18.75%
predicted 2	31	3	91.18%	8.82%
predicted 3	23	1	95.83%	4.17%
Total	102	10	91.07%	8.93%

Note: The highest accuracy is reported in bold.

The highest forecasting accuracy was achieved in class 0 that included banks on the brink of default, with a 100% percentage in correct classification.

In Table 9 we depict the confusion matrix (actual and forecasted classes) of the most accurate SVM model. As we observed from Table 9, 26 of the 32 instances of class 1 were forecasted correctly, while four are classified as belonging to class 2, one to class 1 and one to class 3. For class 2, 31 are classified correctly while three are classified into class 3. At class 3 in only one instance was a bank classified as belonging to class 2 instead of the actual class 3. Thus, in most cases the model classified instances into neighboring classes and the tendency was to misclassify by assigning a higher (economically healthier) class.

Table 9. Confusion Matrix of the Best Model.

	Actual 0	Actual 1	Actual 2	Actual 3
predicted 0	22	1	0	0
predicted 1	0	26	0	0
predicted 2	0	4	31	1
predicted 3	0	1	3	23

Alternatively, instead of using solely the accuracy ratio, we can estimate the Area Under the Receiver Operating Curve (AUC-ROC). The higher the AUC, the more accurate the classification of the model. In Figure 2 we depict the AUC for the most accurate SVM and Probit model per class, respectively.

The SVM model achieved higher AUC for all four classes in comparison to the Probit model, reaching 0.95 for class 0, 0.79 for class 1, 0.77 for class 2 and 0.85 for class 3. The respective values of the Probit model are 0.86 for class 0, 0.64 for class 1, 0.66 for class 2 and 0.77 for class 3. Thus, either the accuracy of the AUC reached similar conclusions. Table 10 depicts the results of our bootstrapping training scheme.
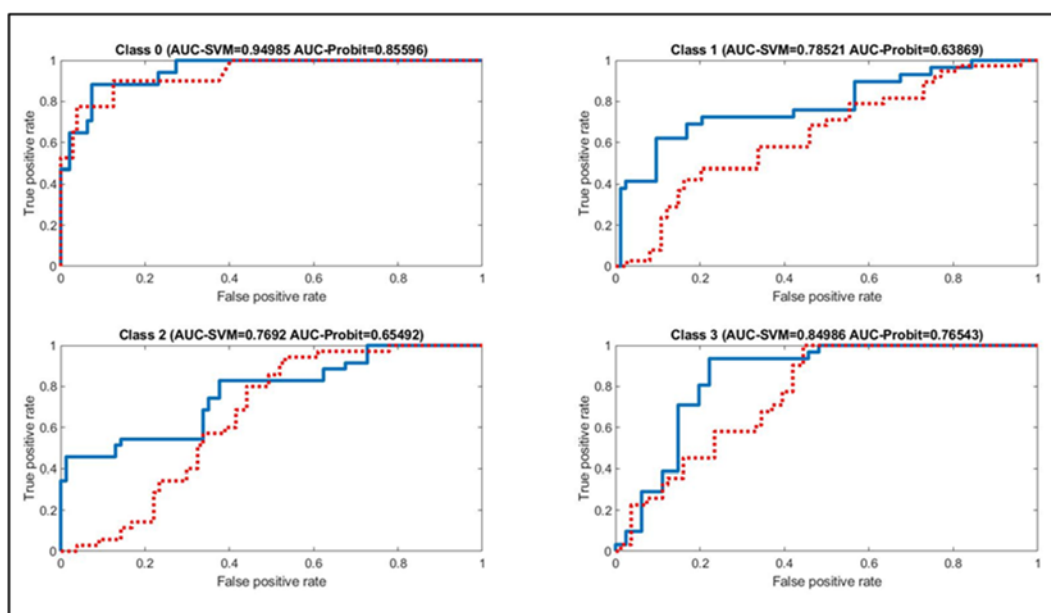
Figure 2. Area Under the Receiver Operating Curve. The line of the SVM model (blue continuous) outperforms the respective line of the Probit model (red dashed). The area under the curve (AUC) for each model is reported for each class and model, respectively.

Table 10. Bank Rating forecasting accuracy by the SVM model (%).

Regressor Selection	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Panel A: Linear kernel (bootstrap)						
Combinatorial 4	64.29 [61.61, 66.96]	70.54 [67.86, 73.21]	63.39 [60.71, 66.52]	64.29 [61.61, 66.96]	70.54 [66.96, 73.21]	63.40 [61.61, 66.96]
Combinatorial 8	61.61 [58.04, 65.18]	71.43 [68.75, 73.21]	59.82 [57.15, 63.39]	58.04 [55.36, 60.71]	70.54 [67.86, 70.66]	62.95 [60.27, 66.07]
Stepwise-forward	62.50 [59.82, 65.18]	66.96 [64.29, 69.64]	62.50 [59.82, 64.89]	65.18 [62.95, 68.75]	74.11 [71.43, 76.79]	**80.36** **[77.68, 85.04]**
Panel B: nonlinear RBF kernel (bootstrap)						
Combinatorial 4	81.25 [79.46, 83.93]	85.71 [84.82, 87.50]	78.13 [75.90, 80.36]	78.57 [75.89, 81.25]	86.60 [83.93, 87.50]	78.57 [75.89, 81.25]
Combinatorial 8	82.15 [80.36, 83.93]	91.96 [89.29, 93.75]	82.14 [80.36, 83.93]	91.52 89.29, 93.75]	97.32 [96.43, 98.21]	95.54 [94.64, 97.32]
Stepwise-forward	81.25 [79.46, 83.04]	88.39 [84.82, 90.18]	81.25 [79.46, 83.04]	97.32 [95.54, 97.32]	**98.21** **[97.32, 100]**	95.42 [92.11, 98.57]

Note: The highest accuracy is reported in bold. All values are percentages. One standard deviation confidence intervals in brackets.

As we observe from Table 10, the most accurate model based on the bootstrapping training method was an SVM model coupled with the nonlinear RBF kernel, that included the four variables selected from group 5, based on a stepwise-forward selection scheme. The forecasting accuracy was 98.21% with a 95% confidence interval of [97.32, 100], and the independent variables are (a) Total Assets 2013 and 2014—a measure of capital invested in the bank, (b) Total Interest received 2016—a measure of bank profitability, and (c) Other interest bearing liabilities 2013—a measure of market exposure. As with the most accurate cross validation training scheme, the most accurate bootstrapping model was based on financial variables that measure the market exposure or the bank's profitability. Interestingly, Fitch reacts to the information included in financial statements with a delay, as variables of 2013 and 2014 adhered closer to the unobserved, underlying rating mechanism than information included in 2016. Thus, while we would expect that a rating agency would update its rating model

yearly, we observe that this is not the case, and variables with a lag of 3 or 4 years are used. Given that the ratings can be proxied very closely by the publicly available data, an improvement in the disclosure of the data by the banking institutions could reduce the dependence on rating agencies.

Naturally, forecasting credit ratings (as accurate as it can be) does not bypass the problem that credit ratings themselves can be inaccurate in representing the true creditworthiness of borrowers (Parnes and Akron 2016; Parnes 2018). Nevertheless, the "true" creditworthiness of an E.U. banking institution will always be unknown, since it is dependent on private information that cannot be unveiled through public information. A natural extension of our work would be to compare the rating of Fitch to an alternative CRA in the framework of Parnes and Akron (2016), but we leave this avenue for future research.

5. Conclusions

In this study, we attempted to forecast the credit ratings of European financial institutions. To the best of our knowledge, this is the first time this has been done for European banks. In doing so we used a sample of 112 EU banks and tried to identify the most important factors contributing to their ratings. The target rating was the one provided by Fitch for the year 2017. In our approach, unlike what was done by CRAs, we only used publicly available data from the published financial statements of the banks. For each banking institution, we gathered data for 34 variables for 4 years prior to the 2017 rating, i.e., from 2013 to 2016. This resulted in 136 variables that were used as potential forecasters of Fitch ratings.

We followed a detailed variable selection procedure and created 18 alternative groups of regressors. First, we extracted six groups of variables based on different correlation criteria. Next, from each one of these six groups, we identified the most informative regressors using a combinatorial eight, combinatorial four and a stepwise forward procedure. This two-level variable selection scheme produced 18 alternative sets of explanatory variables. These regressors were then used in both a Probit model from classical econometrics and a Support Vector Machines (SVM) algorithm from the area of Machine Learning. In the case of the SVM model we used both the linear and the non-linear RBF kernel. Moreover, in this case and for the purpose of robustness of the results, we employed two training techniques: the standard cross-validation procedure with three folds to avoid overfitting and also bootstrapping with 8000 replications. The bootstrapping procedure enabled us to also produce confidence intervals for the forecasts.

Our empirical findings revealed that the SVM models vastly outperformed the Probit ones. The best Probit model reached an accuracy of 70.59%, while the best SVM with cross-validation reached 91.07%, and the best SVM with bootstrapping 98.21% with a 95% confidence interval of [97.32, 100]. The model based on the bootstrapping technique used only four independent variables as forecasters: (a) Total Assets of 2013 and 2014, implying that the size of a bank matters for the rating, (b) Total Interest received in 2016, which is a measure of bank income and (c) Other interest bearing liabilities from 2013, which is a measure of market exposure and capital expenses. Thus, the main drivers of bank ratings are the size, measured in total capital and profitability, measured by interest income and interest expense. It is interesting to see that the credit rating of 2017 was mostly determined by the size of the bank three and four years before, its capital expenses four years before and the interest income in the previous year.

Thus, capital and interest expense have a longer-term effect on the rating, explaining the apparent—and criticized—sluggishness of the CRAs in changing an assigned rating especially downwards. On the other hand, interest income has a more direct effect on the published rating. Moreover, we may infer from the apparent high accuracy of the best model that internal undisclosed information or other qualitative information used in the rating process by the CRAs, plays a very small role in the rating model.

References

Belkaoui, Ahmed. 1980. Industrial bond ratings: A new look. *Financial Management* 9: 44–51. [CrossRef]

Bissoondoyal-Bheenick, Emawtee, and Sirimon Treepongkaruna. 2011. Analysis of the determinants of bank ratings: Comparison across agencies. *Australian Journal of Management* 36: 405–24. [CrossRef]

Chang, Chih-Chung, and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2: 1–27. [CrossRef]

Cortes, Corina, and Vladimir Vapnik. 1995. Support vector networks. *Machine Learning* 20: 273–97. [CrossRef]

Dutta, Soumitra, and Shashi Shekhar. 1988. Bond rating: A non-conservative application of neural networks. Paper presented at the IEEE International Conference on Neural Networks, San Diego, CA, USA, July 21–24; pp. 1443–450.

Ederington, Louis H. 1985. Classification models and bond ratings. *Financial Review* 20: 237–62. [CrossRef]

European Central Bank (ECB). 2006. Dent Securities. Available online: https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/securities_issues/debt_securities/html/index.en.html (accessed on 20 July 2020).

Gogas, Periklis, Theophilos Papadimitriou, and Anna Agrapetidou. 2014. Forecasting bank credit ratings. *The Journal of Risk Finance* 15: 195–209. [CrossRef]

Hau, Harald, Sam Langfield, and David Marques-Ibanez. 2013. Bank ratings: What determines their quality? *Economic Policy* 28: 289–333. [CrossRef]

He, Jie, Jun Qian, and Philip E. Strahan. 2012. Are all ratings created equal? The impact of issuer size on the pricing of mortgage-backed securities. *The Journal of Finance* 67: 2097–137. [CrossRef]

Huang, Zan, Hsinchun Chen, Chia-Jung Hsu, Wun-Hwa Chen, and Soushan Wu. 2004. Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems* 37: 543–58. [CrossRef]

Kim, Jun Woo, H. Roland Weistroffer, and Richard T. Redmond. 1993. Expert systems for bond rating: A comparative analysis of statistical, rule-based and neural network systems. *Expert Systems* 10: 167–72. [CrossRef]

Kraft, Pepa. 2015. Do rating agencies cater? Evidence from rating-based contracts. *Journal of Accounting and Economics* 59: 264–83. [CrossRef]

Kwon, Young S., Ingoo Han, and Kun Chang Lee. 1997. Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating. *International Journal of Intelligent Systems in Accounting Finance and Management* 6: 23–40. [CrossRef]

Maher, John J., and Tarun K. Sen. 1997. Predicting bond ratings using neural networks: A comparison with logistic regression. *Intelligent Systems in Accounting, Finance and Management* 6: 59–72. [CrossRef]

Moody, John, and Joachim Utans. 1995. Architecture selection strategies for neural networks application to corporate bond rating. In *Neural Networks in the Capital Markets*. Edited by Apostolos Refenes. New York: John Wiley & Sons, Inc., pp. 277–300.

Pagratis, Spyros, and M. Stringa. 2007. *Modelling Bank Credit Ratings: A Structural Approach to Moody's Credit Risk Assessment*. London: Bank of England.

Papadimitriou, Theophilos. 2012. Financial institutions clustering based on their financial statements using multiple correspondence analysis. *Economics and Financial Notes* 1: 119–33.

Parnes, Dror. 2018. Observed Leniency among the Credit Rating Agencies. *The Journal of Fixed Income* 28: 48–60. [CrossRef]

Parnes, Dror, and Sagi Akron. 2016. Rating the Credit Rating Agencies. *Applied Economics* 48: 4799–812. [CrossRef]

Pinches, George E., and Kent A. Mingo. 1973. A Multivariate Analysis of Industrial Bond Ratings. *Journal of Finance* 28: 1–18. [CrossRef]

Ravi, Vadlamani, H. Kurniawan, Peter Nwee Kok Thai, and P. Ravi Kumar. 2008. Soft computing system for bank performance prediction. *Applied Soft Computing* 8: 305–15. [CrossRef]

Surkan, Alvin J., and J. Clay Singleton. 1990. Neural networks for bond rating improved by multiple hidden layers. Paper presented at IEEE International Conference on Neural Networks, San Diego, CA, USA, June 17–21; pp. 157–162.