

# Forecasting Bank Credit Ratings

---

**Periklis Gogas, Theophilos Papadimitriou and Anna Agrapetidou**

Department of Economics, Democritus University of Thrace,  
Komotini, Greece

## **Abstract**

**Purpose** - This study presents an empirical model designed to forecast bank credit ratings using only quantitative and publicly available information from their financial statements. For this reason we use the long term ratings provided by Fitch in 2012. Our sample consists of 92 U.S. banks and publicly available information in annual frequency from their financial statements from 2008 to 2011. **Methodology** - First, in the effort to select the most informative regressors from a long list of financial variables and ratios we use stepwise least squares and select several alternative sets of variables. Then these sets of variables are used in an ordered probit regression setting to forecast the long term credit ratings. **Findings** - Under this scheme, the forecasting accuracy of our best model reaches 83.70% when 9 explanatory variables are used. **Originality/value** - The results indicate that bank credit ratings largely rely on historical data making them respond sluggishly and after any financial problems are already known to the public.

## **1. Introduction**

Credit Rating Agencies (CRAs) provide a rating scale of risks associated with the ability of banks to meet debt obligations on time. These ratings are used by investors, borrowers, issuers and governments in making investment and financial decisions. Consequently, changes in ratings lead to important changes in capital allocation. The three major rating agencies are Moody's, Standard and Poor's and Fitch. The implementation of Basel II strengthened the demand for ratings and expanded the role of the CRAs. They play a key role in the pricing of credit risk. However their integrity has been under consideration due to cases such as Enron and Lehman Brothers which had been assessed, by the rating agencies, with high ratings just a few days before their collapse. For this reason, they are often blamed for not being able to provide the market with the appropriate ratings required for important investment decisions. Such cases

also occurred during the global financial crisis of 2008 and turn the interest in credit rating methodologies. It is generally accepted that when the credit rating of a bank is downgraded from the CRAs then things get worse for the specific banking institution and vice versa.

The aim of this paper is to find the most important variables that contribute to the long term ratings of U.S. banks as they are assigned by Fitch. The ability to forecast these ratings within reasonable margins of accuracy is of great importance first of all to investors. Correct identification of a future rating can help an investor to optimize her portfolio and anticipate any possible market reaction after the official rating is assigned by the rating agency. This may lead to a) lower hedging costs, b) improved profitability and c) a significant risk reduction in terms of portfolio volatility. Moreover, the forecasting model we aim to develop can provide significant information to the investors and any interested party in the case of banks that may not yet be rated by a major rating agency. Furthermore, even for the banks that were rated in the past, not all ratings are regularly updated and reported freely. By using only the publicly available information from the banking institutions' financial statements we can estimate a rating that can be used for investing decisions. The importance of a simple model that can forecast official ratings extends to the individual banking institutions as well: our analysis identifies the financial variables and ratios that are most important in the credit rating process. By using these to feed our proposed model, bank officials can replicate and anticipate their future ratings providing them with the ability to steer their institution towards the desired rating. Finally, the ability to forecast credit ratings based largely on publicly available information credit ratings enhances the transparency of the rating process and minimizes the possibility to exploit instances of conflict of interest between the credit rating agencies and the rated bank as was implied in the recent financial crisis.

The remainder of the paper is organized as follows. Section 2 reviews the literature and section 3 describes the data. Section 4 presents the methodology and the empirical results and section 5 concludes the paper.

## **2. Literature Review**

### *a) Credit rating methodologies*

Researchers have applied various empirical methodologies to explain and forecast credit ratings. Substantial papers can be found in predicting bond ratings. Ederington

(1985), Pinches and Mingo (1973), Belkaoui (1980) used statistical methods such as logistic regression and multivariable discriminant analysis (MDA) to predict bond ratings. In these studies different sets of variables were used and the prediction results were between 50% and 70%. Many studies on bond credit rating prediction with neural networks (Dutta et al. (1988); Surkan et al. (1990); Kim (1993)) show more promising results than statistical methods. Moody and Utans (1994) used neural networks to predict bond ratings of firms that had a rating from S&P. Using 10 input variables to predict 16 S&P subratings they managed to predict correctly just 36.2% of the ratings. However using 5-class and 3-class clustered partition their model predicted correctly 63.8% and 85.2% of the ratings respectively. Maher and Sen (1997) compared neural networks and logistic regression on predicting bond ratings for the period 1990-92. The best model was the neural network model yielding 70% accuracy on a holdout sample. Kwon et al. (1997) compared ordinal pairwise partitioning (OPP) based back propagation neural networks to conventional neural networks and MDA, using 126 financial variables for Korean companies for the period 1991-1993. They showed that the OPP neural network scheme provided an accuracy of 71-73% outperforming the other alternatives. Huang et al. (2004) applied back propagation neural networks (BPNN) and support vector machine (SVM) to corporate credit rating prediction for the United States and Taiwan markets and achieved a prediction accuracy approximately 80% via SVM. In a similar study, Chen et al. (2006) compared SVM to BPNN on Taiwan's issuer credit ratings. Once more the SVM outperformed the BPNN model with an accuracy rate of 84.62%.

The superiority of logit and probit models has been shown by Ederington (1985), McKelvey et al. (1975) and Trevino et al. (2000), who argue that logit and probit models are the most appropriate methods for the classification of credit ratings due to the nature of the dependent variable which is discrete and ordinal.

#### *b) Bank credit ratings*

Only recently, the issue of banks' credit ratings has attracted the attention of researchers. Statistical and machine learning techniques have been applied for the examination of bank credit ratings that are assigned by CRAs. Statistical techniques include linear regression (Poon et al. 1999) and ordered probit (Bissoondoyal-Bheenick and Treepongkaruna (2011); Pagratis and Stinga (2007)).

Multi-criteria decision techniques, namely UTilite's Additives DIScriminantes (UTADIS) and Multi-group Hierarchical DIScrimination (MHDIS) have been applied by Pasiouras et al. (2006, 2007). In the first paper they used UTADIS for the classification of banks into three coarse categories according to their financial soundness. The proposed methodology achieved a classification matching accuracy of 68.91%. The second study applies the MHDIS method for the classification of Asian banks into five categories. The accuracy of classification reached 66.03%. In both papers, the proposed methods outperformed MDA. Chen et al. (2012) applied a hybrid procedure based on Rough Set Theory to classify credit ratings in the Asian banking industry into five rating categories. They used an integrated feature selection approach to identify the actual determinants when calculating credit ratings. The proposed methodology yielded a classification accuracy of 83.84%.

Ioannidis et al. (2010) tried to classify banks into three categories using financial and country-specific variables based on Fitch individual ratings. Their data consisted of 944 banks from 78 countries. They compared various machine learning techniques such as multi-criteria decision aid, neural networks, classification trees and k-nearest NN. The highest accuracies were achieved using multi-criteria decision aid and neural networks. Several studies compared machine learning based methods with classic econometric methodologies. Hammer et al. (2012) compared multiple linear regression, ordered logistic regression, SVM and logical analysis on the classification of 800 banks from 70 different countries that had a rating from Fitch in 2001. The creditworthiness of banks was evaluated in terms of 23 financial variables and ratios and the S&P's country risk rating. They concluded that logical analysis outperformed the other approaches. In the same path, Bellotti et al. (2011) compared SVM with ordered choice models in forecasting individual bank ratings. They used financial and country specific variables for 629 international banks for the period 2000-2007. The SVM based model yielded more accurate forecasting than the ordered logit and probit models.

Papadimitriou (2013) explored the clustering properties of the correspondence analysis map of 90 financial institutions using data from their public financial statements. The goal was to compare the clustering groups with the rating assigned by the Fitch Rating Agency. Results showed that there is a visible relation between the clustering and the ratings, though it is not useful as it is, since the regions corresponding to the ratings are highly overlapping.

### *c) Determinants of bank credit ratings*

Financial ratios, macroeconomic factors and country specific variables have been used for the examination of bank credit ratings. Poon et al. (1999) used factor analysis to identify loan provisions, risk factors and profitability ratios as the most important variables to explain bank financial strength ratings. The study of Pasiouras et al. (2006) concludes that loan loss provisions, capitalization and region of operations are the most important variables for bank classification. In a similar research Pasiouras et al. (2007) deduce that the most important variables for credit rating forecasting are net interest margin, short-term funding, return on average equity, the number of shareholders and subsidiaries and the region of operations.

Bissoondoyal-Bheenick and Treepongkaruna (2011) analyzed the quantitative determinants of bank ratings, provided by Standard & Poor's, Moody's, and Fitch for the U.K. and Australian banks and they found that accounting variables have more explaining power in banks' ratings rather than macroeconomic variables. In the same path Chen et al. (2012) found that the most important feature of banks' credit ratings for the investment grade is asset quality while for the speculative grade the most important feature is capital adequacy.

Ioannidis et al. (2010) showed that country-specific variables have an important impact on classification accuracy. The same conclusion is supported in Bellotti et al. (2011). The authors demonstrated the role of country specific variables and identified four variables as the most important determinants of bank ratings: the lagged values of a) equity to total assets, b) liquid assets to total assets c) natural logarithm of total assets and d) net interest margin.

These studies demonstrate the importance of financial variables in the assessment of bank credit ratings. Although CRAs in their rating process take into account quantitative as well as qualitative information from various sources, empirical results show that there is an increased reliance on publicly available information and less reliance on confidential information.

### **3. The Data**

For our analysis we use a cross-section of 92 U.S. banks for which we could find freely available long-term ratings from Fitch. The ratings we attempt to forecast pertain to the

year 2012. In doing so, we collect for each one of the 92 banks in our sample 46 individual variables and ratios for four years prior to our target rating (2008 to 2011) that come from their publicly reported financial statements. Thus, in order to forecast the credit rating of a bank in 2012 we include in our regressor set for each variable its previous four years values: e.g., the Net Operating Income of 2011 (NOI11) and three years prior to that NOI10, NOI09 and NOI08. In this way, we effectively use 184 variables for each bank to forecast bank ratings. All financial data come from the database of the FDIC (Federal Deposit Insurance Corporation). The variables are reported in Table 1. The dependent variable is ordinal and initially has six categories that are grouped in our case in three coarse categories. They are assigned integer values from 0 to 2, such that lower values indicate a lower rating. The three rating categories (with assigned values in brackets) are: AA, A (2), BBB (1), BB, B, CCC (0).

**Table 1:** List of independent variables

<b>Assets and Liabilities</b>	<b>Income and Expenses</b>	<b>Performane Ratios</b>	<b>Condition Ratios</b>
Total employees (TOEM)	Total interest expense (TIE)	Yield on earning assets (YOEA)	Loss allowance to loans (LATL)
Total assets(TASSET)	Provision for loan and lease losses (PLLL)	Net interest margin (NIM)	Net loans and leases to deposits (NLLTD)
Cash and due from depository institutions (CASH)	Total noninteres tincome (TNI)	Net operating income to assets (NOIA)	Net loans and leases to core deposits (NLLCD)
Net loans & leases (NLL)	Service charges on deposit accounts (CHARGE)	Return on assets (ROA)	Equity capital to assets (EQCTA)
Loan loss allowance (LLA)	Trading account gains & fees (TRACC)	Return on equity (ROE)	Core capital (leverage) ratio (LEV)
Goodwill and other intangibles (GOI)	Additional noninterest income (ANI)	Assets per employee (ASSPE)	Tier 1 risk-based capital ratio (T1RBC)
Total deposits(TD)	Total noninterest expense (TNE)		Total risk-based capital ratio (TRBCR)
Interest-bearing deposits (IBD)	Salaries and employee benefits (SAL)		
Trading liabilities (TL)	Pre-tax net operating income (PTNOI)		
Subordinated debt (SD)	Securities gains (SEC)		
Total bank equity capital (TBEC)	Net income attributable to bank (NIA)		
Long-termassets(LTA)	Cash dividends (DIVDS)		
Average Assets (AA)	Net operating income (NOI)		
Volatile liabilities(VL)			

Loans and leases held for sale (LLHFS)			
Unused loan commitments (ULC)			
Tier 1 (core) risk-based capital (T1CRC)			
Tier 2 risk-based capital (T2RBC)			
Total unused commitments (TUC)			
Derivatives (DER)			

Note: All variables from the balance sheets (assets and liabilities) are expressed in terms of total assets except from total employees (TOEM) and total assets (TASSET). The log of total assets is used as a measure of bank size. All variables from the income statements (income and expenses) are expressed in terms of total interest income.

## 4. Methodology and empirical results

### 4.1 Feature Selection

In order to identify the variables that contribute the most to the assigned bank ratings of our sample we follow a thorough variable selection procedure. For the total of 184 regressors we first calculate the correlations,  $r_{i,R}$ , where  $i$  refers to each individual variable and R is the ratings dependent variable. Next, as a prefiltering step, we create six alternative groups of regressors as follows: In group 1 we include all variables with  $|r_{i,R}| \geq 0.5$  along with all their lags. E.g., if NOI10 has a correlation greater than 0.5 then in group 1 we include all time instances of this variable: NOI11, NOI10, NOI09 and NOI08. This group includes 20 variables. In a similar manner, in group 2 we do the same with for  $|r_{i,R}| \geq 0.4$ . This group includes 44 variables. For group 3 all variables with  $|r_{i,R}| \geq 0.4$  are included without their corresponding time instances: if only NOI10 has a correlation above 0.4 then only this variable is selected in group 3. In this group there are 15 variables. In group 4 we include the 5 variables with the highest positive correlation and the 5 variables with the highest negative correlation for a total of 10 variables. In group 5 we select the 30 variables with the highest correlation. Finally, in group 6 we have all 184 explanatory variables of our sample. Table 2 summarizes the number of variables in each group.

**Table 2:** Number of variables in each regressor group

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
20	44	15	10	30	184
variables	variables	variables	variables	variables	variables

The next step is to identify from each one of the above groups the most significant variables in terms of the ratings. This is done in each group by:

- a) A combinatorial exhaustive search methodology of all possible sets of 4 variables. We then select the one set that produces the highest  $R^2$  in a regression on the ratings dependent variable.
- b) Selecting in the same manner as above an augmented regressor set with 8 variables.
- c) We select using a stepwise forward method of least squares the set of variables with  $p$ -value greater than 0.1.

In Tables 3 and 4 we summarize the above results. Table 3 reports for each one of the six groups of the prefiltered regressors the  $R^2$  resulting from the regression of the best selected variables on the bank ratings ordered depended variable. The first line reports the  $R^2$  from selecting the best 4 regressors from each group using an exhaustive search of all possible combinations within the group. In line 2 this is done for an exhaustive search for 8 regressors. Finally, in the last line we report the corresponding  $R^2$  for the regression of the variables selected using the stepwise-forward variable selection criterion and the numbers in the parentheses indicate the number of variables that the stepwise-forward method selects for each one of the six groups of regressors.

**Table 3:**  $R^2$  values for the optimum set of regressors for the six groups of variables.

Regressor Selection	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Combinatorial 4	0.49	0.54	0.54	0.53	0.53	0.59
Combinatorial 8	0.53	0.61	0.57	0.49	0.60	0.70
Stepwise-forward	0.40	0.55	0.55	0.46	0.53	0.71
	(4)	(5)	(5)	(3)	(4)	(11)

Note: The numbers in the parentheses indicate the variables selected by the stepwise-forward criterion.

In Table 4 we identify all the variables selected as best regressors according to each one of the three variable selection methodologies described above and for each one of the six prefiltered groups of regressors. In Panel A we report the selected regressors from the combinatorial methodology that selects with an exhaustive search the set of the best 4 variables for each initial group of prefiltered regressors. Panel B reports the same

information when the combinatorial methodology selects the set of the best 8 variables for each group. Finally, in Panel C, we report the variables that are selected using the stepwise – forward method. This selects only three variables from group 4, four variables from groups 1 and 5, five variables from groups 2 and 3 and eleven variables from group 6 (the group that includes all 184 variables without any prefiltering). In Appendix A we report the detailed description of the variables presented in Table 4.

**Table 4:** Variables selected from each method.

<b>Panel A.</b> Methodology:Combinatorial 4 (set of 4 best regressors)					
Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
NOIA10	NOI10	NOI10	L_TASSET11	L_TASSET11	L_TASSET11
PTNOI10	L_TASSET11	L_TASSET11	L_TASSET10	L_TASSET10	TIE11
NOIA8	IBD11	IBD11	NOIA10	NOIA10	SEC8
ROA08	AA10	AA10	TNE11	TNE11	PTNOI10

  

<b>Panel B.</b> Methodology: Combinatorial 8 (set of 8 best regressors)					
Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
NIA10	L_TASSET11	L_TASSET11	NOI10	L_TASSET11	L_TASSET11
ROA10	IBD11	IBD11	AA10	L_TASSET10	L_TASSET10
PTNOI10	AA10	AA10	IBD11	TNE11	SEC8
ROA09	PTNOI10	L_TASSET10	NIA10	TNI10	LTA8
NOIA8	TNI11	NOIA10	PTNOI10	DIVDS11	YOEA11
ROA08	TNI9	L_TASSET9	ROA10	ROE09	NIM10
NIA11	ROE08	L_TASSET8	NOIA10	PTNOI9	NOIA10
ROA11	NOIA8	TNI11	PLLL10	PTNOI10	TBEC9

  

<b>Panel C.</b> Methodology: Stepwise Forward (set of regressors with $p > 0.1$ )					
Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
NOI10	NOI10	NOI10	NOI10	TNE11	NOI10
NIA10	L_TASSET11	L_TASSET11	IBD11	L_TASSET11	L_TASSET11
ROA10	IBD11	IBD11	AA10	L_TASSET10	TIE11
NOIA10	AA10	AA10		NOIA10	L_TASSET10
	L_TASSET10	L_TASSET10			SEC8
					GOI8
					ASSPE10
					LTA8
					TRL11
					SD11
					T2RBC8

## 4.2 Ordered Probit Models

The above selection procedure results in 18 different sets of regressors. These sets are next used in an ordered probit model and they are compared in terms of bank rating forecasting accuracy. The selection of ordered probit model is based on the nature of the dependent variable which is discrete and ordinal such as credit ratings. The ordered probit model takes into account the differences between the categories of the dependent variable. For example the difference between the rating categories 0 and 1 is not the same as 1 and 2. Thus, it is the most appropriate econometric methodology. (McKelvey et al. (1975); Trevino et al. (2000)). In an ordered probit model the dependent variable  $y$  represents ordered observations or in other words a ranking variable. In our case this is the credit rating assigned by Fitch for each individual banking institution for the year 2012. This dependent variable is modeled by a latent variable  $y_i^*$  that has a linear relation with the vector of explanatory variable  $\mathbf{x}_i$  as follows:

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon_i$$

Where  $\varepsilon_i$  is independently and identically distributed. The actual  $y_i$  is fitted from  $y_i^*$  where:

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq \gamma_1 \\ 1 & \text{if } \gamma_1 < y_i^* \leq \gamma_2 \\ 2 & \text{if } \gamma_2 < y_i^* \end{cases}$$

And the probabilities of having each value of  $y$  are given by:

$$Prob(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}, \gamma) = F(\gamma_1 - \mathbf{x}_i' \boldsymbol{\beta})$$

$$Prob(y_i = 1 | \mathbf{x}_i, \boldsymbol{\beta}, \gamma) = F(\gamma_2 - \mathbf{x}_i' \boldsymbol{\beta}) - F(\gamma_1 - \mathbf{x}_i' \boldsymbol{\beta})$$

$$Prob(y_i = 2 | \mathbf{x}_i, \boldsymbol{\beta}, \gamma) = F(\gamma_3 - \mathbf{x}_i' \boldsymbol{\beta}) - F(\gamma_2 - \mathbf{x}_i' \boldsymbol{\beta})$$

$F$  is the cumulative distribution function of  $\varepsilon$ . The marginal values  $\gamma$  are estimated with the  $\boldsymbol{\beta}$  coefficients by maximizing the log-likelihood function:

$$Logl(\boldsymbol{\beta}, \gamma) = \sum_{i=1}^N \sum_{j=0}^N \log(Prob(y_i = j | \mathbf{x}_i, \boldsymbol{\beta}, \gamma)) \cdot 1(y_i = j).$$

The last term corresponds to an indicator function which takes the value 1 if the condition is true and 0 if the condition is false.

### 4.3 Empirical Results

After an intricate feature selection process as described in section 4.1, we created three sets of regressors of each one of the prefiltered groups of variables resulting in a total of 18 forecasting models to be estimated and evaluated in terms of bank credit rating forecasting accuracy. We use the ordered probit methodology as described above to forecast the bank credit ratings for the year 2012. In Table 5 we report the forecasting accuracy of each one of the 18 models tested. Each column corresponds to each one of the six groups of the prefiltered regressors and each row presents the results for the corresponding variable selection criterion. It appears that all three regressor sets that are identified from group 6, the group that includes all initial 184 variables, outperform all regressor sets provided by the same criterion from the other five groups. For the combinatorial 4 criterion, group 6 has a forecasting accuracy of 71.74% while groups 2 and 3 both follow with an accuracy of 70.65%. In the combinatorial 8 selection criterion the regressors from group 6 provide a forecasting accuracy equal to 81.52% while the set from group 3 follows with an accuracy of 76.09%. Finally, in the stepwise-forward selection criterion, the best forecasting accuracy is achieved again with the regressors from group 6 at 81.52%. The regressor sets from groups 2, 3 and 4 follow with an accuracy of 71.74%. To conclude, according to these results, the best accuracy for all regressor selection criteria is achieved for the regressor sets from group 6 for the combinatorial selection of eight regressors and the stepwise forward method both with 81.52% correct bank rating forecast.

**Table 5:** Rating forecasting accuracy

Regressor Selection	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Combinatorial 4	69.57%	70.65%	70.65%	68.48%	68.48%	<b>71.74%</b>
Combinatorial 8	68.48%	75.00%	76.09%	71.74%	71.74%	<b>81.52%</b>
Stepwise-forward	65.22%	71.74%	71.74%	71.74%	68.48%	<b>81.52%</b>

For each one of the two forecasting models with the highest forecasting accuracy in the first column of Table 6 we list the corresponding variables used, in column two we report their source and in the third column we provide their description. From this we can observe that the log of the total assets in the last two years before the assigned rating

are significant in forecasting the bank rating in both models. Another common regressor in the two forecasting models are the securities gains (losses) over total interest income four years prior to the assigned rating. The log of total assets regressor is measured in absolute volume of total assets employed in the banking operations while the securities regressor is relative to the bank's interest income. Some additional interesting findings are that the variables that come from the balance sheet (A/L) are dominating both forecasting models (4 out of 8 in the first model and 7 out of 11 in the second) while the condition ratios (C/R) do not appear in both optimum models.

**Table 6:** List of regressors in the optimal forecasting models

Variable	Source	Description
<b>Panel A.</b> Group 6 using the combinatorial 8 criterion		
L_TASSET11	A/L	Log of total assets in 2011
L_TASSET10	A/L	Log of total assets in 2010
SEC8	I/E	Securities gains (losses) over total interest income in 2008
LTA8	A/L	Long-term assets (5+ years) over total assets in 2008
YOE11	P/R	Yield on earning assets in 2011
NIM10	P/R	Net interest margin in 2010
NOIA10	P/R	Net operating income to assets in 2010
TBEC9	A/L	Total bank equity capital over total assets in 2009
<b>Panel B.</b> Group 6 using the stepwise-forward criterion		
NOI10	I/E	Net operating income over total interest income in 2010
L_TASSET11	A/L	Log of total assets in 2011
TIE11	I/E	Total interest expense over total interest income in 2011
L_TASSET10	A/L	Log of total assets in 2010
SEC8	I/E	Securities gains (losses) over total interest income in 2008
GOI8	A/L	Goodwill and other intangibles over total assets in 2008
ASSPE10	P/R	Assets per employee (\$ millions) in 2010
LTA8	A/L	Long-term assets (5+ years) over total assets in 2008
TRL11	A/L	Trading liabilities over total assets in 2011
SD11	A/L	Subordinated debt over total assets in 2011
T2RBC8	A/L	Tier 2 risk-based capital over total assets in 2008

Note: the regressor source is classified as A/L when they come from Assets and Liabilities, I/E when they come from the Income and Expense accounts, P/R for a performance ratio and C/R for a condition ratio.

Finally, when we augment these models with additional regressors from the list of 184 total variables of our sample we manage to increase the forecasting accuracy and produce the best bank ratings forecasting model when in the list of the eight explanatory variables, selected from group 6 and the combinatorial 8 criterion we also include the long-term assets (5+ years) over total assets for the year 2009 (LTA9). The forecasting accuracy for this model reaches 83.70%. Specifically, the results show that the assigned ratings largely depend on historical data that span four years prior to the credit rating event: the size of the bank as it is measured by the natural log of total assets in the two years prior to the rating (L\_TASSET11 and L\_TASSET10), the ratio of long-term assets over total assets three and four years before the rating (LTA9 and LTA8), the structure of the bank's capital three years prior (TBEC9), gains or losses from securities as a percentage of total interest income four years before the rating (SEC8), the net interest margin and net operating income lagged two years (NIM10 and NOIA10 respectively) and the yield on earning assets from 2010 (YOEA10). In Table 7, we present an analysis of the forecasted classes against the actual bank ranking in 2012.

**Table7:**Comparison of predicted to real rating categories

	Rating category 0	Rating category 1	Rating category 2
predicted 0	14	1	0
predicted 1	3	21	5
predicted 2	0	6	42

In the rating category AA, A (2), 42 of the 47 observations were correctly classified. In the category BBB (1), BB,B and CCC (0), 21 of the 28 and 14 of the 17 observations were correctly classified.

## 5. Conclusion

Credit Rating Agencies play a significant role in the financial system. Their reports are an essential source of information to both sides of the transaction table. Investors' and capital or debt issuers' decisions and the overall success of such issues heavily rely on the assigned credit ratings. As a result, both parties may benefit from correctly forecasting future credit ratings using prior information. This was the main goal of our contribution: we tried to develop a relatively simple model to forecast Fitch's ratings.

We used 184 publicly available financial variables of U.S. banks from 2008 to 2011 in order to fit a model that will forecast the next year's long term rating of Fitch. In doing so, we prefilter this extensive list of possible regressors and produce six groups of variables. From each one of these groups we select three sets of explanatory variables by applying the combinatorial method for four and eight selected variables and the stepwise forward method. This results in eighteen sets of explanatory variables that are used in an ordered probit framework to forecast Fitch's long term ratings for 2012. According to the results the optimum model that reaches an 83.70% forecasting accuracy is the one that includes nine financial variables. Five of these nine forecasting regressors come from the banks' balance sheets, three are performance ratios and one comes from the income statement. No condition ratios appear significant in forecasting bank ratings. Moreover, the results show that the assigned ratings largely depend on historical data that span four years prior to the credit rating event. The variables that can be used to forecast the actual bank ratings are: the size of the bank as it is measured by the natural logarithm of total assets in the two years prior to the rating, the ratio of long-term assets over total assets three and four years before the rating, the structure of the bank's capital three years prior, the financial gains or losses from securities as a percentage of total interest income four years before the rating, the net interest margin and net operating income lagged two years and finally, the yield on earning assets lagged two years. Thus, the most important variables contributing to long term ratings of banks are –not surprisingly- size, performance ratios and asset quality. These results indicate that the assessment of credit ratings is largely relying on historical data that are widely available to investors, bank officials and policy makers. Thus, it is not surprising that during the recent financial crisis bank downgrading was announced rather late and long after their financial problems were already visible to all interested parties. Moreover, the results show that it is most easy to forecast banks with high ratings rather than banks with a low rating.

There are several possible extensions to this line of research. First, methodologically, the adoption and comparison in forecasting accuracy with the probit models of a multiclass Support Vector Machines (SVM) classifier from the area of Machine Learning (ML). Second, the extension of this forecasting modelling to bond ratings both corporate and sovereign.

## Acknowledgements

This research has been co-financed by the European Union (European Social Fund (ESF)) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

## References

- Belkaoui, A. (1980). Industrial bond ratings: a new look. *Financial Management*, 44-51.
- Bellotti, T., Matousek, R., & Stewart, C. (2011). Are rating agencies' assignments opaque? Evidence from international banks. *Expert Systems with Applications*, 38(4), 4206-4214.
- Bissoondoyal-Bheenick, E., & Treepongkaruna, S. (2011). An analysis of the determinants of bank ratings: comparison across ratings agencies. *Australian Journal of Management*, 36(3), 405-424.
- Chen, W. H., & Shih, J. Y. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, 30(3), 427-435.
- Chen, Y. S., & Cheng, C. H. (2012). Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. *Knowledge-Based Systems*, 39, 224-239.
- Chen, Y. S., (2012). Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach. *Knowledge-Based Systems*, 26, 259-270.
- Dutta, S., & Shekhar, S. (1988, July). Bond rating: a nonconservative application of neural networks. In *Neural Networks, 1988., IEEE International Conference on* (pp. 443-450). IEEE.
- Ederington, L. H. (1985). Classification models and bond ratings. *Financial Review*, 20(4), 237-262.
- Hammer, P. L., Kogan, A., & Lejeune, M. A. (2012). A logical analysis of banks' financial strength ratings. *Expert Systems with Applications*, 39(9), 7808-7821.

- Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision support systems*, 37(4), 543-558.
- Ioannidis, C., Pasiouras, F., & Zopounidis, C. (2010). Assessing bank soundness with classification techniques. *Omega*, 38(5), 345-357.
- Kim, J. W., Weistroffer, H. R., & Redmond, R. T. (1993). Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems. *Expert systems*, 10(3), 167-172.
- Kwon, Y. S., Han, I., & Lee, K. C. (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.
- Maher, J. J., & Sen, T. K. (1997). Predicting bond ratings using neural networks: a comparison with logistic regression. *Intelligent Systems in Accounting, Finance and Management*, 6(1), 59-72.
- McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology*, 4(1), 103-120.
- Moody, J., & Utans, J. (1994). Architecture selection strategies for neural networks: Application to corporate bond rating prediction. In *Neural networks in the capital markets* (pp. 277-300). John Wiley & Sons.
- Pagratis, S., & Stringa, M. (2007). Modelling bank credit ratings: A structural approach to Moody's credit risk assessment. Mimeo, Bank of England.
- Papadimitriou, Th., (2012). Financial Institutions Clustering based on their Financial Statements using Multiple Correspondence Analysis. *Economics and Financial Notes*, Vol. 1, no. 2, 119-133.
- Pasiouras, F., Gaganis, C., & Doumpos, M. (2007). A multicriteria discrimination approach for the credit rating of Asian banks. *Annals of Finance*, 3(3), 351-367.
- Pasiouras, F., Gaganis, C., & Zopounidis, C. (2006). The impact of bank regulations, supervision, market structure, and bank characteristics on individual bank ratings: A cross-country analysis. *Review of Quantitative Finance and Accounting*, 27(4), 403-438.
- Pinches, G. E., & Mingo, K. A. (1973). A multivariate analysis of industrial bond ratings. *The Journal of Finance*, 28(1), 1-18.

- Poon, W. P., Firth, M., & Fung, H. G. (1999). A multivariate analysis of the determinants of Moody's bank financial strength ratings. *Journal of International Financial Markets, Institutions and Money*, 9(3), 267-283.
- Surkan, A. J., & Singleton, J. C. (1990, June). Neural networks for bond rating improved by multiple hidden layers. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on* (pp. 157-162). IEEE.
- Trevino, L., & Thomas, S. (2000). Systematic differences in the determinants of foreign currency sovereign ratings by rating agency (No. 00-153).

<b>Appendix A: Description of variables in the 18 sets of regressors</b>					
<b>Panel A: Combinatorial 4</b>					
<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>	<b>Group 4</b>	<b>Group 5</b>	<b>Group 6</b>
<b>NOIA10</b> - Net operating income to assets in 2010	<b>NOI10</b> - Net operating income over total interest income in 2010	<b>NOI10</b> - Net operating income over total interest income in 2010	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET11</b> - Log of total assets in 2011
<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET10</b> - Log of total assets in 2010	<b>L_TASSET10</b> - Log of total assets in 2010	<b>TIE11</b> - Total interest expense over total interest income in 2011
<b>NOIA8</b> - Net operating income to assets in 2008	<b>IBD11</b> - Interest-bearing deposits over total assets in 2011	<b>IBD11</b> - Interest-bearing deposits over total assets in 2011	<b>NOIA10</b> - Net operating income to assets in 2010	<b>NOIA10</b> - Net operating income to assets in 2010	<b>SEC8</b> - Securities gains (losses) over total interest income in 2008
<b>ROA8</b> - Return on assets in 2008	<b>AA10</b> - Average assets (year to date) over total assets in 2010	<b>AA10</b> - Average assets (year to date) over total assets in 2010	<b>TNE11</b> - Total noninterest expense over total interest income in 2011	<b>TNE11</b> - Total noninterest expense over total interest income in 2011	<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010
<b>Panel B: Combinatorial 8</b>					
<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>	<b>Group 4</b>	<b>Group 5</b>	<b>Group 6</b>
<b>NIA10</b> - Net income attributable to bank (owners) over total interest income in 2010	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET11</b> - Log of total assets in 2011	<b>NOI10</b> - Net operating income over total interest income in 2010	<b>L_TASSET11</b> - Log of total assets in 2011	<b>L_TASSET11</b> - Log of total assets in 2011
<b>ROA10</b> - Return on assets in 2010	<b>IBD11</b> - Interest-bearing deposits over total assets in 2011	<b>IBD11</b> - Interest-bearing deposits over total assets in 2011	<b>AA10</b> - Average assets (year to date) over total assets in 2010	<b>L_TASSET10</b> - Log of total assets in 2010	<b>L_TASSET10</b> - Log of total assets in 2010

<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010	<b>AA10</b> - Average assets (year to date) over total assets in 2010	<b>AA10</b> - Average assets (year to date) over total assets in 2010	<b>IBD11</b> - Interest-bearing deposits over total assets in 2011	<b>TNE11</b> - Total noninterest expense over total interest income in 2011	<b>SEC8</b> - Securities gains (losses) over total interest income in 2008
<b>ROA9</b> - Return on assets in 2009	<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010	<b>L_TASSET10</b> - Log of total assets in 2010	<b>NIA10</b> - Net income attributable to bank (owners) over total interest income in 2010	<b>TNI10</b> - Total noninterest income over total interest income in 2010	<b>LTA8</b> - Long-term assets (5+ years) over total assets in 2008
<b>NOIA8</b> - Net operating income to assets in 2008	<b>TNI11</b> - Total noninterest income over total interest income in 2011	<b>NOIA10</b> - Net operating income to assets in 2010	<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010	<b>DIVDS11</b> - Cash dividends over total interest income in 2011	<b>YOE11</b> - Yield on earning assets in 2011
<b>ROA8</b> - Return on assets in 2008	<b>TNI9</b> - Total noninterest income over total interest income in 2009	<b>L_TASSET9</b> - Log of total assets in 2009	<b>ROA10</b> - Return on assets in 2010	<b>ROE9</b> - Return on equity in 2009	<b>NIM10</b> - Net interest margin in 2010
<b>NIA11</b> - Net income attributable to bank (owners) over total interest income in 2011	<b>ROE8</b> - Return on equity in 2008	<b>L_TASSET8</b> - Log of total assets in 2008	<b>NOIA10</b> - Net operating income to assets in 2010	<b>PTNOI9</b> - Pre-tax net operating income over total interest income in 2009	<b>NOIA10</b> - Net operating income to assets in 2010
<b>ROA11</b> - Return on assets in 2008	<b>NOIA8</b> - Net operating income to assets in 2008	<b>TNI11</b> - Total noninterest income over total interest income in 2011	<b>PLLL10</b> - Provision for loan and lease losses over total interest income in 2010	<b>PTNOI10</b> - Pre-tax net operating income over total interest income in 2010	<b>TBEC9</b> -Total bank equity capital over total assets in 2009
<b>Panel C: Stepwise Forward</b>					
<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>	<b>Group 4</b>	<b>Group 5</b>	<b>Group 6</b>
<b>NOI10</b> - Net operating income over	<b>NOI10</b> - Net operating income over	<b>NOI10</b> - Net operating income over	<b>NOI10</b> - Net operating income over	<b>TNE11</b> - Total noninterest expense over	<b>NOI10</b> -- Net operating income over

<i>total interest income in 2010</i>	<i>total interest income in 2010</i>	<i>total interest income in 2010</i>	<i>total interest income in 2010</i>	<i>total interest income in 2011</i>	<i>total interest income in 2010</i>
<b>NIA10</b> - <i>Net income attributable to bank (owners) over total interest income in 2010</i>	<b>L_TASSET11</b> - <i>Log of total assets in 2011</i>	<b>L_TASSET11</b> - <i>Log of total assets in 2011</i>	<b>IBD11</b> - <i>Interest-bearing deposits over total assets in 2011</i>	<b>L_TASSET11</b> - <i>Log of total assets in 2011</i>	<b>L_TASSET11</b> - <i>Log of total assets in 2011</i>
<b>ROA10</b> - <i>Return on assets in 2010</i>	<b>IBD11</b> - <i>Interest-bearing deposits over total assets in 2011</i>	<b>IBD11</b> - <i>Interest-bearing deposits over total assets in 2011</i>	<b>AA10</b> - <i>Average assets (year to date) over total assets in 2010</i>	<b>L_TASSET10</b> - <i>Log of total assets in 2010</i>	<b>TIE11</b> - <i>Total interest expense over total interest income in 2011</i>
<b>NOIA10</b> - <i>Net operating income to assets in 2010</i>	<b>AA10</b> - <i>Average assets (year to date) over total assets in 2010</i>	<b>AA10</b> - <i>Average assets (year to date) over total assets in 2010</i>		<b>NOIA10</b> - <i>Net operating income to assets in 2010</i>	<b>L_TASSET10</b> - <i>Log of total assets in 2010</i>
	<b>L_TASSET10</b> - <i>Log of total assets in 2010</i>	<b>L_TASSET10</b> - <i>Log of total assets in 2010</i>			<b>SEC8</b> - <i>Securities gains (losses) over total interest income in 2008</i>
					<b>GOI8</b> - <i>Goodwill and other intangibles over total assets in 2008</i>
					<b>ASSPE10</b> - <i>Assets per employee (\$ millions) in 2010</i>
					<b>LTA8</b> - <i>Long-term assets (5+ years) over total assets in 2008</i>
					<b>TRL11</b> - <i>Trading liabilities over total assets in 2011</i>

					<b>SD11-</b> <i>Subordinated debt over total assets in 2011</i>
					<b>T2RBC8-</b> <i>Tier 2 risk-based capital over total assets in 2008</i>