

## The Effectiveness of Artificial Credit Scoring Models in Predicting NPLs using Micro Accounting Data

Vasileios Giannopoulos\*

Department of Business Administration, University of Patras, Rio-Patras, Greece

### Abstract

In this paper we study the effectiveness of artificial credit scoring models in predicting SMEs default. We use a unique accounting dataset of small business loans granted by one of four systemic Greek Banks during expansion period. Comparing a neural network model (multilayer perceptron) and a decision tree model with the credit scoring model applied by the bank, we find that the bank's model had the relative worse performance in predicting loans default. Moreover the effectiveness of all models decreased significantly during the recession, indicating that the loan performance is no longer depended only on the quality of the borrowers but also on the economic conditions of the country.

**Keywords:** Non-performing loans; Credit scoring; Banks; Micro and small enterprises; Greek crisis; Neural network; Decision trees

### Introduction

The global financial crisis, and most importantly, domestic conjunctural and structural factors, inevitably affected the Greek economy from the September 2008 [1]. Greece recorded its worst economic performance in 2009 since joining the euro area. The paper explicitly focuses on the repayment behavior of SMEs loans in the Greek economy when macroeconomic conditions deteriorate. Greece constitutes a particular interesting country to study the default behavior of SMEs due to the critical importance and enormous volume of NPLs in the Greek banking sector along with our access to primary micro-level information about SMEs characteristics and loans repayment history from the bank under study. A key aim of this paper is to study the effectiveness of deferent credit scoring models during the recent Greek reception period.

To respond to our research questions, we compare the credit scoring model internally applied by the bank under examination with two other prediction models proposed by the extant literature such as binomial logistic regression and decision tree, in order to reveal weaknesses in bank decision making during expansion. This comparative analysis proves that the bank's credit scoring model performed worse than the other credit scoring models, thus confirming the existing literature.

The paper exploits manually gathered individual-level loan application and loan performance data from 3,294 SMEs loans that were granted in the expansion period (from January 2005 to December of 2005) by one of the four Greek systemic banks, which as oligopolistic players exhibit a similar strategic behavior offering very comparable products. During the span of this study, the Greek banking sector has undergone several development phases with alternating time periods [2]: initially, an expansion period (2002-2007) with very high GDP growth levels, then an unstable period caused by the global financial crisis on September 2008, afterwards the Greek sovereign crisis on April 2010 [3] which led to an accumulation of NPLs in the subsequent years (2010-2012) and finally the deep recession (2013-2016). In this turbulent environment, the analysis concentrates on the period 2010-2012 since NPLs in subsequent years increased to not-manageable levels due the vulnerable political scene that prevailed that period (so the incurred loan delays that period can be considered as mechanical). To sum up, the study explores the repayment behavior of the 3,294 SMEs loans few years later after they were granted (2005) and throughout the early recession period from August 2010 to July 2012.

Our paper focuses explicitly on the default behavior of SMEs and in this direction considers a relatively large number of SMEs borrower's idiosyncratic features as loan-specific characteristics thus substantially differing from those literature [4] that explores at aggregated level both macroeconomic factors and bank-specific characteristics for NPLs accumulation.

Finally, the paper contributes to the relevant credit scoring literature [5-8], evaluating the relative predictive ability of diverse credit scoring models. Our findings support evidence for emerging economies [9-10] that prediction models used internally by banks showed the relatively worst performance over time. Consequently, the study has important implications for bank management and policy makers aiming to ensure financial stability in Greece and in other euro area periphery economies in the post crisis era.

The rest of the paper is structured as follows. In section 2 we provide a literature review. In section 3 we present the data set and the employed variables. In section 4 we demonstrate the research methodology. In section 5 we provide the empirical results. Section 6 concludes the paper.

### Literature Review

In recent years, credit scoring has been become one of the primary ways for financial institutions to assess credit risk, improve cash flow, reduce possible risks and make managerial decisions. The accuracy of credit scoring is critical to financial institutions' profitability as this methodology is to classify loans to either good credit or bad credit, predicting the bad payers [11]. In particular, credit scoring shows specific benefits for evaluating micro and small business loans. Many credit scoring techniques have been used to build credit scorecards. The

**\*Corresponding author:** Giannopoulos Vasileios, Department of Business Administration, University of Patras, Rio-Patras, Greece, Tel: +306942547065; E-mail: [vgiannopoulos@upatras.gr](mailto:vgiannopoulos@upatras.gr)

**Received** October 05, 2018; **Accepted** December 11, 2018; **Published** December 22, 2018

**Citation:** Giannopoulos V (2018) The Effectiveness of Artificial Credit Scoring Models in Predicting NPLs using Micro Accounting Data. J Account Mark 7: 303. doi: [10.4172/2168-9601.1000303](https://doi.org/10.4172/2168-9601.1000303)

**Copyright:** © 2018 Giannopoulos V. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

most commonly used quantitative methods are statistical automatic credit scoring techniques and artificial intelligence techniques [7]. In the broad category of statistical automatic credit scoring techniques, three are the main methods: Linear Discriminant Analysis-Multivariate Discriminant Analysis [12], Logistic Regression Analysis-Logit and Probit models [11,13], Multivariate Adaptive Regression Splines [14]. Among them, the logistic regression model is the most commonly used in the banking industry due to its desirable features such as robustness and transparency [6].

Within the category of artificial credit scoring techniques, there are four main methods with empirical evidence to document their superior predictive accuracy: Artificial Intelligence techniques, such as Artificial Neural Networks [15,16], Decision Trees [17], Case Based Reasoning [18] and Support Vector Machines [19]. The neural networks are *inter alia* the alternative to the linear discriminant and logistic regression analysis due to the possible complex non-linear relationship between variables. Empirical studies on the accuracy of different credit scoring models report that neural networks techniques are more accurate than linear discriminant analysis and logit [11,8,15]. Although the usage of these techniques has increased in recent years, their weaknesses lie in their long training process and after obtaining the optimal network's architecture, their models act as a "black box", therefore, it is not easy to identify the relative importance of potential input variables [20].

As regards the predictive quality of a credit scoring model, this can be evaluated based both on statistical measures, such as sensitivity, specificity, correlation coefficients and information measures, such as relative entropy and mutual information. Generally, the choice of a particular technique depends on the data structure, the features used, the extent to which it is possible to segregate the classes by using those features and the purpose of the classification [21].

Apart from the aforementioned scoring models widely used by relevant literature, banks possess their own models that they internally exploit to separate loan applicants that are expected to pay back their debts from those who are likely to fall into arrears. Emerging literature on this specific field provides evidence suggesting a weak evaluation performance of such models. In particular, relevant results are identified in the paper of Louzada [6], where the employed credit scoring models, showed better performance than the model applied by the bank of Brazil that they examined. Mileris and Boguslauskas [9] demonstrated that the credit scoring models they used were more effective than the model applied by the studied Bank of Lithuania. Similarly, Abdou [10] showed that both the discriminant analysis and logistic regression methods showed a better performance than the model used by the researched bank of Egypt.

These findings motivate us thoroughly to investigate the credit scorecard the bank under examination and compare that with a neural network technique (Multilayer Perceptron) and a decision tree credit scoring model (CART), well established in the relevant literature.

## The Data set and the Employed Variables

The data set is collected manually from the internal Management Information System (MIS) of the bank under study and contains a very wide loan portfolio consisting of micro businesses and small enterprises as defined by the EU. The initial data set contained 4,102 loan applications granted in the late expansion period (2005) and we applied some filters to the data. In particular, we removed from the sample the repaid loans and those that were denounced before 08/2010, thus ending up with a final sample consisting of 3,294 applications

of micro and small enterprises spread across Greece, the repayment scheme of which we could easily follow during the entire investigation period. The present study is based on a joint project between academic researchers with previous professional banking experience and the top level lending management of the bank under investigation. This was carried out due to the necessity of identifying important drivers of credit risk related to borrowers' characteristics and re-evaluating the existing internal credit scoring model of the bank under study during recession.

In our analysis, we set as a dependent variable the 'performance of the loan' during the studied period. For the definition of a loan as non-performing, we use the basic rules of Basel I & II, where NPLs are those loans that are up to ninety days past due. As a time frame for the identification of the behavior of a NPL, empirical studies [22] specify either the performance of loans in a specific month or the performance of loans during a specific period, usually 12 months. In our analysis we utilize both identification methods.

As independent variables we use quantitative and qualitative loan characteristics derived from the loan application at the time of evaluation. In particular, qualitative information (such as the age of the borrower, the type of the loan etc.) is significant in explaining a firm's credit risk [23,24] justified by the "Five Cs of Credit" and used by lenders for credit worthiness evaluation of potential borrowers. These five respective criteria represent five general features of the borrower, attempting to gauge the chance of default: the character of the consumer, the capital, the collateral, the capacity and the economic conditions. In our research analysis, we utilize the ten main characteristics of the credit scoring model used by the bank under study as independent variables (loan characteristics). Table 1 summarizes the definition of these independent variables.

Given the onset of the Greek Crisis in April 2010 [3], we observe the loan repayment evolution and by extension the escalation of NPLs for the next two years 08/2010-07/2012. During this period, two crucial re-capitalizations of the domestic systemic banks took place in Greece since the domestic banking system was on the verge of collapse. The political scene at that time was particularly vulnerable and fragile further accentuating the negative economic and financial consequences of the crisis.

## Research Design

We utilize as dependent variable the performance of loans successively at three specific time points: August 2010, August 2011 and finally July 2012. Thereafter, we compare the specific credit scoring model of the bank under study with two credit scoring models—a multilayer perceptron neural network and a decision tree model, using as basic criterion the effectiveness of these models in the prediction of NPL's as the recession of the Greek economy deepens. Furthermore, we observe the time effect of the loan characteristics to the creation of new NPL's, testing in this way the impact of the crisis on the predictive ability of the credit scoring models.

### Multilayer perceptron neural network

The design of a neural network, called multilayer perceptron (MLP), is particularly suitable for the classification of variables and is widely used in practice. The network consists of an input layer, one or more hidden layers and an output layer, each of which consists of multiple neurons. Each neuron processes the input data and produces an output value that is transmitted to the neurons in the next layer.

Number	NPL determinant	Definition	Type of characteristic
1	Adverse	Borrowers who experienced adverse at the time of assessing the application	Character
2	Age	The age of the borrower	Character
3	Bankrel	Four indicators for the type of banking relationship: 1. no customer, 2. only loans, 3. only deposits 4. deposits and loans	Capital
4	Collateral	Four indicators for the type of collateral: 1. no collateral, 2. securities (checks-exchange), 3. "Mortgage on the property" 4. Cash collateral cover (deposits, bancassurance and investment savings products)".	Collateral
5	LTT	Loan to turnover ratio.	Economic Conditions
6	Owfac	Dummy variable taking the value 0 if the borrower had not owned facilities and the value 1 if the borrower had owned facilities at the time of assessing the application.	Capacity
7	Property	Dummy variable taking the value 0 if the borrower and the guarantor meet the criterion of real estate, or the value 1 otherwise.	Capital
8	Residence	Three indicators for the type of residence: 1. "rented house", 2. "live with parent" and 3. Private residence".	Capacity
9	Loan type	Three indicators for the loan type: 1. equipment, 2. facilities, 3. working capital fixed term 4. Working capital limit-overdraft.	Capital
10	Years	The years of vocational experience.	Economic conditions

**Table 1:** The association of micro and small firm's idiosyncratic features with the formation of NPL's.

Each neuron, in the input layer ( $I = 1, \dots, n$ ), yields the value of an estimator of the vector  $x$ . When we are referring to the examination for distinctness default/non-default, an output neuron is satisfactory. In each layer, the signal transmission is accomplished as follows. First, a weighted sum of the inputs to each neuron is calculated: the output value of each neuron of the network multiplied by the respective weight of the connection with this neuron. Then a transfer function  $g(x)$  applies in this weighted sum to determine the neuron output value. Thus, each neuron in the hidden layer ( $j=1, \dots, q$ ) produces the so-called activation

$$a_j = g \left( \sum_i w_{ij} x_i \right) \quad (1)$$

Neurons in the output layer ( $k=1, \dots, m$ ) behave in a manner similar to the neurons in the hidden layer to produce the network output result:

$$y_k = f \left( \sum_j w'_{jk} a_j \right) = f \left( \sum_j w'_{jk} g \left( \sum_i w_{ij} x_i \right) \right) \quad (2)$$

where  $w_{ij}$  and  $w'_{jk}$  are weights.

The logarithmic function:

$$f(x) = 1 / (1 + \exp(x)) \quad (3)$$

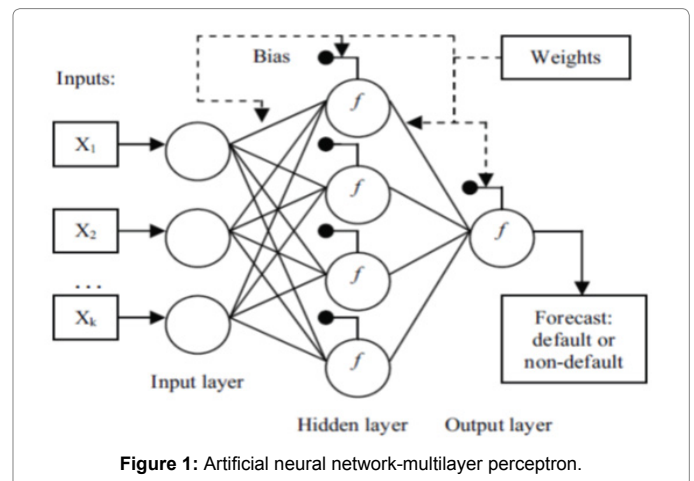
Or the alternative tangent hyperbolic function

$$f(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x)) \quad (4)$$

Are commonly used in the upper output of the network for the functions  $f$  and  $g$ . The logarithmic function is appropriate to the output layer, if we have a binary classification problem as in credit scoring, so that the output can be considered a default probability. The structure of a neural network with a single hidden layer is capable of approximating any continuous bounded integrable function (Figure 1).

## Decision trees

Decision tree is a popular technique for classification and has been widely used in the field of data mining. The trees try to maximize their average classification accuracy and consist of three main elements, the "decision nodes" corresponding to the characteristics; the "edges" or branches corresponding to different possible attributes and the "leaves" that include items usually belong in the same category.



**Figure 1:** Artificial neural network-multilayer perceptron.

Several algorithms have been developed for constructing a decision tree such as ID3, C5.0 and CART [25]. The CART models (classification and regression trees) are a classification method that has been used successfully in credit scoring [15]. In banking practice these models are mostly used as a supporting tool to accompany the parametric estimation methods and serve the selection of independent variables with the highest explanatory power. CART method typically uses binary trees and classifies a set of data into a finite number of categories. Originally it was developed as a tool that concerned itself with binary responses and therefore is suitable for use in the rating process of credit rating of a borrower [26,27].

In our paper, we use a CART model. The loan characteristics are the "decision nodes", the categories of each characteristic are the "edges" or branches and the performance of each loan (good or bad loan) represents the "leaves" [28-30]. From the set of the loan characteristics, we select the minimum number of these characteristics that minimize the estimated misclassification cost (Figure 2).

## Comparative Analysis of Credit Scoring Models

For the effective evaluation of the compared credit scoring models, we checked both the percentage of correct predictions (average accuracy) of each model, and the estimated misclassification cost.

Average accuracy measures the percentage of positives and negatives that are correctly identified as such.

$$\text{Average Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{\text{Total}} \quad (5)$$

Estimated misclassification cost (EMC) refers to the costs incurred by the incorrect categorization of a loan. Essentially, it is about the cost of not granting a good loan and the cost resulting from granting a bad loan. Particularly is calculated as:

$$\text{EMC} = \frac{\text{Performing Loans}}{\text{Total Loans}} \times \text{Type Error1} + \frac{\text{Non - Performing Loans}}{\text{Total Loans}} \times \text{Type Error2} \quad (6)$$

where type I error rate is the proportion of negatives that are incorrectly classified as positives.

$$\text{Type I error} = \frac{\text{False Negative}}{(\text{False Negative} + \text{True Positive})} \quad (7)$$

and type II error rate is the proportion of positives that are incorrectly identified as negatives.

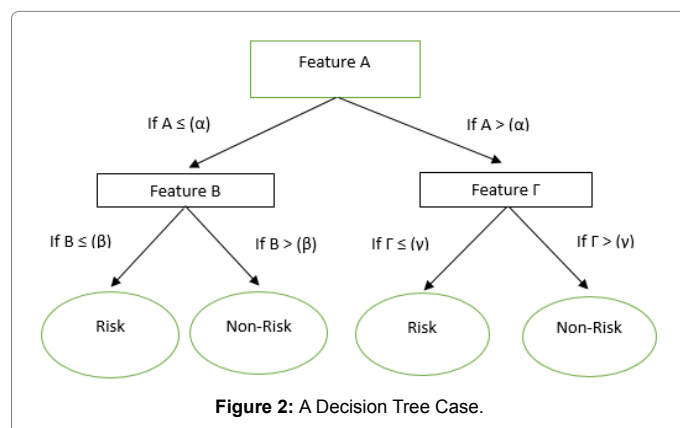
$$\text{Type II error} = \frac{\text{False Positive}}{(\text{False Positive} + \text{True Negative})} \quad (8)$$

## Results

Table 2 presents a comparative analysis of the predictive power of the three credit scoring models. The results show that both Multilayer Perceptron and the CART model perform better than the credit scoring model utilized by the case study bank in 2005. In general, taking into account both Average Accuracy (AA) and Estimated Misclassification Cost (EMC) we found that CART model pointed the best performance in August 2010 (AA: 92.26% / EMC: 0.0464), followed by the Multilayer Perceptron (AA: 92.23% / EMC: 0.0464). On the other hand, in August 2011, Multilayer Perceptron showed the best performance (AA: 87.40% / EMC: 0.0748), followed by the CART model (AA: 85.61% / EMC: 0.0847). In July 2012 decision tree exhibited the best performance (AA: 75.08% / EMC: 0.1355), followed by the Multilayer Perceptron (AA: 74.32% / EMC: 0.1572). The bank's credit scoring model performed the worst average accuracy and estimated misclassification cost in all cases. To sum up, the above analysis demonstrates that the credit scoring model applied by the bank under study recorded the worst performance overtime (Table 2).

Moreover, we observed that the predictive power of all credit scoring models was dramatically reduced since the trading behaviour of borrowers was largely influenced by external factors such as the downturn of the economy and the credit crunch of micro and small enterprises. More precisely, the average accuracy reduced from 92.23% (August 2010) to 74.32% (July 2012) for the Multilayer Perceptron, from 92.26% to 75.08% for the CART model, and from 88.95% to 72.01% for the bank's credit scoring model.

At the same time, the estimated misclassification cost increased as



Credit scoring model	August 2010	August 2011	July 2012
	Average Accuracy		
Multilayer Perceptron	92,23%	87,40%	74,32%
Decision tree	92,26%	85,61%	75,08%
Bank's Credit scoring model	88,95%	83,12%	72,01%
Estimated misclassification cost			
Multilayer Perceptron	0,0464	0,0748	0,1572
Decision tree	0,0464	0,0847	0,1355
Bank's Credit scoring model	0,0566	0,0927	0,1606

**Table 2:** Comparative analysis of Credit Scoring Models (average accuracy, misclassification cost).

the recession was evolved. Particularly this cost increased from 0.0464 (August 2010) to 0.1572 (July 2012) for the Multilayer Perceptron, from 0.0464 to 0.1355 for the CART model and from 0.0566 to 0.1606 for the bank's credit scoring model.

The significant reduction of the effectiveness of all studied models during the recession indicates that NPLs are influenced significantly by the deterioration of the external environment. This finding indicates that the good loan performance is no longer depended only on the loan-specific characteristics determined by bank management but also on the state of financial figures of the country.

## Conclusions and Discussion

The considerable rise of NPLs during the Greek recession motivated us to test the predictive ability of the bank's credit scoring model compared with a multilayer perceptron neural network and a decision tree model, widely used in relevant literature. To the best of our knowledge this paper is the first attempt to employ the appropriate methodological tools for the evaluation of bank internal and external forces that cause NPLs. In this context, we utilized a unique employed disaggregate data set capturing for the first time the small sized borrower's idiosyncratic features as compared to current literature that emphasizes macroeconomic elements (such as real GDP growth and unemployment) and bank specific characteristics at aggregated level.

A main conclusion of our study is that the bank's credit scoring model performed worse than the other two models. The considerable ineffectiveness of these models that our paper documented strongly supports new literature for emerging economies suggesting that prediction models owned and internally applied by banks performed relatively worst over time. In addition we found that as recession escalates, the predictive performance of all credit score models (included that of the case study bank) gradually weakens thus preventing even more the pursuit of a rational lending policy.

In future work, we could study the evolution of business loans granted during the recession, in order to find if the credit policy of banks has changed, compared to the expansion period. Similar studies could also be done in other EU countries, in order to identify similarities and differences between EU countries. Given that small and medium-sized enterprises are the engine of the European economies, we believe that their rational financing is an important factor in the development of the EU economy. Therefore, we consider very important to study the efficiency of the bank management regarding their credit policy. To this end, this paper shows that banks should invest in the development of artificial intelligence models, in order to create efficient credit scoring models.

Generally, the study provides valuable evidence and policy implications on the transformation of loans of micro and small businesses into NPLs taking into account management decisions and



environmental deterioration. These implications might be useful to practitioners when making difficult evaluation decisions of granting loans to diverse borrowers in changing environmental contexts.

## References

- Aggelopoulos E, Georgopoulos A (2015) The Determinants of Shareholder Value in Retail Banking During Crisis Years: The Case of Greece. *Multinational Finance J* 19: 109-147.
- Aggelopoulos E, Georgopoulos A (2017) Bank branch efficiency under environmental change: A bootstrap DEA on monthly profit and loss accounting statements of Greek retail branches. *Eur J Oper Res* 261: 1170-1188.
- Kosmidou KV, Kousenidis DV, Negakis CI (2015) The impact of the EU/ECB/IMF bailout programs on the financial and real sectors of the ASE during the Greek sovereign crisis. *J Banking & Finance* 50: 440-454.
- Louzis D, Vouldis A, Metaxas V (2012) Macroeconomic and bank specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *J Bank Financ* 36: 1012-1027.
- Crook J, Edelman D, Lyn T (2007) Recent developments in consumer credit risk assessment. *Eur J Oper Res* 183: 1447-1465.
- Louzada F, Ferreira-Silva P, Diniz C (2012) On the impact of disproportional samples in credit scoring models: An application to a Brazilian Bank data. *Expert systems with Applications* 39: 8071-8078.
- Roszbach K (2004) Bank lending policy, credit scoring, and the survival of loans. *Review of Economics and Statistics* 86: 946-958.
- Wu C, Wang X (2000) A neural network approach for analyzing small business lending decisions. *Review of Quantitative Finance and Accounting* 15: 259-276.
- Mileris, R, Boguslauskas, V (2011) Credit Risk Estimation Model Development Process: Main Steps and Model Improvement. *Engineering Economics* 22: 126-133.
- Abdou H, El-Masry A, Pointon J (2007) On the applicability of credit scoring models in Egyptian Banks. *Banks and Bank Systems* 2: 4-20.
- Lee T, Chiu C, Lu C, Chen I (2002) Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications* 23: 245-254.
- Altman E, Saunders A (1998) Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance* 21: 1721-1742.
- Gordy M (2000) A comparative anatomy of credit risk models. *J Banking and Finance* 24: 119-149.
- Friedman J (1991) Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19: 1-141.
- West D (2000) Neural Network Credit Scoring Models. *Comput Oper Res* 27: 131-1152.
- Zopounidis C, Doumpos M (2002) Multicriteria classification and sorting methods: A literature review. *Eur J Oper Res* 138: 229-246.
- Jiang L, Li C, Cai Z (2009) Learning decision tree for ranking. *Knowledge and Information Systems* 20: 123-135.
- Shin K (2001) A case-based approach using inductive indexing for corporate bond rating. *Decision Support Systems* 32: 41-52.
- Trustorff J, Konrad P, Leker J (2011) Credit risk prediction using support vector machines. *Review of Quantitative Finance and Accounting* 36: 565-581.
- Sustersic M, Mramor D, Zupan J (2007) Consumer credit scoring models with limited data. *Expert Systems with Applications* 36: 4736-4744.
- Twala, B (2010) Multiple classifier application to credit risk assessment. *Expert System with Applications* 37: 3326-3336.
- Makri V, Tsagkanos A, Bellas A (2014) Determinants of Non-Performing Loans: The case of Eurozone. *Panoeconomicus* 2: 193-206.
- Artavanis N, Morse A, Tsoutsoura M (2016) Measuring income tax evasion using bank credit: Evidence from Greece. *The Quarterly J Economics* 131: 739-798.
- Gupta J, Gregoriou A, Healy J (2015) Forecasting bankruptcy for SMEs using hazard function: To what extent does size matter? *Review of Quantitative Finance and Accounting* 45: 845-869.
- Tsaih R, Liu Y, Liu W, Lien Y (2004) Credit scoring system for small business loans. *Decision Support Systems* 38: 91-99.
- Konstantakis K, Michaelides P, Vouldis A (2016) Non-performing loans in a crisis economy: Long-run equilibrium analysis with a real time VEC model for Greece (2001-2015). *Physica A* 451: 149-161.
- Chen F, Li F (2010) Combination of feature selection approaches with SVM in credit scoring. *Expert systems with Applications* 37: 4902-4909.
- European Central Bank (2016) Draft guidance to banks on non-performing loans.
- European Banking Authority (2016) EBA report on SMES and SME supporting factor.
- European Banking Authority (2015) Risk assessment of the European banking system.