Ann Oper Res
https://doi.org/10.1007/s10479-018-2845-8

S.I.: RISK IN FINANCIAL ECONOMICS

Forecast bankruptcy using a blend of clustering and MARS model: case of US banks

Zeineb Affes[1] · **Rania Hentati-Kaffel**[1]

Abstract In this paper, we compare the performance of two non-parametric methods of classification and regression trees (CART) and the newly multivariate adaptive regression splines (MARS) models, in forecasting bankruptcy. Models are tested on a large universe of US banks over a complete market cycle and run under a K-fold cross validation. Then, a hybrid model which combines K-means clustering and MARS is tested as well. Our findings highlight that (i) Either in training or testing sample, MARS provides, in average, better correct classification rate than CART model (ii) Hybrid approach significantly increases the classification accuracy rate in the training sample (iii) MARS prediction underperforms when the misclassification of the bankrupt banks rate is adopted as a criteria (iv) Finally, results prove that non-parametric models are more suitable for bank failure prediction than the corresponding Logit model.

Keywords Bankruptcy prediction · MARS · CART · K-means · Early-warning system

JEL Classification C14 · C25 · C38 · C53 · G17 · G21 · G28 · G33

1 Introduction

The financial crisis, which started in 2007, has dramatically affected banks sector throughout the world. The shock wave epicenter was in the US and it took long time for regulators to

This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR-10-LABX-0095.

✉ Rania Hentati-Kaffel
rania.kaffel@univ-paris1.fr

Zeineb Affes
zeinebaffes@gmail.com

[1] Centre d'Economie de la Sorbonne, Université Paris1 Panthéon-Sorbonne, Maison des Sciences Economiques, 106-112 Boulevard de l'Hôpital, 75013 Paris, France

Published online: 10 April 2018

stop the default chain and save big banks. Therefore, the prevention against the systemic risk - failure of the banking system- becomes an ineluctable concern and the need of new forecasting tools is of major importance to not only regulators but also academics.

In this sense, federal banking supervisors [the Federal Reserve, the Federal Deposit Insurance Corporation (FDIC), and the Office of the Comptroller of the Currency (OCC)] and other financial supervisory agencies provide a supervisory rating (convenient summary of bank conditions at the time of an exam). This helps investors to detect banks that have a great default probability ratio. A key outcome of such an on-site exam is a CAMELS rating. The acronym "CAMEL" refers to the five components of a bank's condition that are evaluated: Capital adequacy, Asset quality, Management, Earnings, and Liquidity. A sixth component, a bank's Sensitivity to market risk, was added in 1997; hence the acronym was changed to CAMELS.

The FDIC developed also a Statistical CAMELS Off-site Rating system (SCOR) to perform the bank's stability evaluation. Collier et al. (2003) examine the performance of this model over the period 1986–2002 and point out the limitations of this model despite the usefulness of SCOR which is based only on financial ratios. Cole and Gunther (1995) prove the same results and report that the CAMELS ratings decay rapidly.

Predicting bank bankruptcy has reached a specific interest in financial literature. Thus, numerous models have been developed since the early 70s. All proposed models are based on classification methods in a multidimensional space defined by a set of specific variables.

The literature is rich of non-parametric and parametric models. With regard to the later, Beaver (1966) was one of the first researchers who focused on univariate analysis to study bankruptcy prediction. He tested the power of financial ratios to classify and predict bankrupt firms. Cash flow and debt ratios appeared to be the important predictors of bankruptcy. Altman (1968) used Multivariate Discriminant Analysis (MDA) to develop a five-factor model to calculate the well-known "Z-score" and predict bankruptcy of manufacturing firms.

As an examples of well-known statistical methods to predict failure, the logistic regression (logit) (Ohlson 1980; Demirgüç-Kunt and Detragiache 1997; Affes and Hentati-Kaffel 2017), Probit analysis (Zmijewski 1984; Hanweck et al. 1977) and factor analysis. West et al. (1985) demonstrated that that the combination of factor analysis and Logit estimation was promising in evaluating banks' condition. The factors identified by the Logit model as important descriptive variables for the banks' operations are similar to those used for CAMELS ratings.

Martin (1977) used both Logit and MDA statistical methods to predict bank failures. Results proved that the two models have similar classifications of defaulted and healthy banks. In the same sense, Jones and Hensher (2004) presented mixed Logit model for distress firm prediction and compared it with Multinomial Logit Models (MNL). They concluded that mixed Logit obtained substantially better predictive accuracy than Multinomial Logit models.

Affes and Hentati-Kaffel (2017) identified over the period 2008–2013 US banks leading to bankruptcy by conducting a comparative analysis based on both Canonical Discriminant Analysis and Logit models and highlighted also that suitability of models is improved by comparing different optimal cut-off score.

A second stream of literature tested non-statistical methods to set up a model for default prediction. For example, Kolari et al. (2002) use a Trait Recognition Model (TRA) (a kind of the image recognition algorithm). Empirical results recognized that these methods perform better than the statistical approach. Lanine and Vander Vennet (2006) showed that TRA approach outperforms Logit in predicting failures among Russian commercial banks.

Other non-statistical methods include Intelligence techniques such as induction of classification trees and Neural Networks methods (NM). NM procedures follow a process similar

to the human brain and contain mathematical and algorithmic elements that mimic the biological neural networks of the human nervous system (see Odom and Sharda 1990; Lenard et al. 1995; Zhang et al. 1999; McKee and Greenstein 2000; Anandarajan et al. 2001). Boyacioglu et al. (2009) tested various neural networks techniques and multivariate statistical methods to the problem of predicting bank failures in Turkey and presented a comparison of the classification performances of the techniques tested. They used similar financial ratios to those used in CAMELS ratings.

More recently, Decision Trees (DT) and regression tree (CART) methods were used to implement the financial distress prediction. Chen (2011) compared empirically DT with Logit for Taiwan firms, and found that DT got higher accuracy than Logit in short run (less than 1 year), while Logit performed better in long run (above one and half year).

Iturriaga and Sanz (2015) developed a hybrid neural network model to study the bankruptcy of U.S banks by combining a Multilayer Perceptrons (MLP) network and Self-Organizing Maps (SOM). They found that the MLP-SOM can detect 96.15% of the failures in the period between May 2012 and December 2013 and outperforms traditional models of bankruptcy forecast.

De Andrés et al. (2011) and Sánchez-Lasheras et al. (2012) combined Multivariate Adaptive Regression Splines (MARS) model with fuzzy clustering and Self Organized Map (SOM). They found that these hybrids models outperform a single classification models (i) LDA, NN (Feed-forward neural networks) and single MARS (ii) NN (back propagation neural networks) and MARS in terms of correct classification and of the identification of the companies that go bankrupt.

In this paper, we aim to model the relationship between ten financial variables and default probability of US banks by using the so-called non-parametric or flexible models. The main purpose is to test the accuracy of non-parametric methods of classification, to increase their prediction ability and to reduce misclassification problem. Here, we propose a blend of k-means and MARS model. We suggest, for the first time in the bank failure literature, the use of MARS and also a hybrid model which combines K-means and MARS model.

The method consists, in addition to Classification And Regression Trees (CART), to validate the Multivariate Adaptive Regression Splines (MARS) model which gained an increasing interest in financial literature. MARS was first proposed by Friedman (1991). The main advantage of this model is the capacity to explore the complex nonlinear relationships between response variable and various predictor variables.

Unlike the used methodology, the empirical part newly contributes to the existing literature by implementing all these models to a large universe of US banks, over the period spanning 6 years, from 2008 to 2013, under a K-fold Cross validation. According to the size of our data set we apply a 10-fold cross validation to separate it into training and testing sets. In order to evaluate models fitting, we consider the confusion matrix for both the training and the testing samples. Also, we use the Receiver Operating Characteristic Curves (ROC) for evaluating classification success. Finally, we evaluate the performance of each model according to the Area under the ROC Curve.

Our main findings highlights the promising functionality of MARS model and suggest that:

(i) Either in training or in the testing sample, MARS provide better correct classification than CART model in average (97.88–97.58% versus 95.02–93.4%)
(ii) Hybrid approach enhanced the classification accuracy for the training sample
(iii) Relying on misclassification rate of bankrupt banks, MARS underperformed, especially in 2008, 2009 and 2010

(iv) According to the AUC of the Receiver Operating Characteristic Curve we note the supremacy of the hybrid model
(v) Finally, CART provides a better interpretation of splitting values at the root node variables. Splitting values can be considered as early warning signals. CART method allows, in this sense, to carry out and to define the target values sheets that the regulators should take into account in order to be able to identify upstream banks in difficulty.

The paper is organized as follows. Section 2 presents the methodology and data used. Section 3 describes results of CART and MARS models. In Sect. 4, we analyze used models' predictability. Section 5 concludes. Models outcomes are in the Appendix.

2 The model

For empirical validation, we consider a large panel of US banks. We collect data for active banks (AB) from BankScope and non-active (NAB) ones from FDIC, over the period from 2008 to 2013.

We extract all ratios to build 10 financial variables, detailed as follow:

Categories CAMEL	Variables	Definition
Capital adequacy	EQTA	Total equity/total assets
	EQTL	Total equity/total loans
Assets quality	NPLTA	Non performing loans/total assets
	NPLGL	Non performing loans/gross loans
	LLRTA	Loan loss reserves/total assets
	LLRGL	Loan loss reserves/gross loans
Earnings ability	ROA	Net income/total assets
	ROE	Net income/total equity
Liquidity	TLTD	Total loans/total customer deposits
	TDTA	Total customer deposits/total assets

The choice of these ten ratios was conducted and justified by an abundant literature (Sinkey 1979; Martin 1977; Thomson 1991; Barr and Siems 1994; Pantalone et al. 1987; Godlewski et al. 2003).

We adopt the same rule of bank status selection as in Affes and Hentati-Kaffel (2017). Thus, the number of (NAB) banks was 410 over the entire period 2008–2013. The total number of active banks obtained in 2013 is 835, 850 in 2012, 883 in 2011, 955 in 2010, 1077 in 2009 and 1205 in 2008.

However, it was proved that classification tends to favor dominant category, here active banks (AB). This means that the original database has a highly skewed distribution. To create homogeneous groups, we apply K-fold Cross validation.

We apply a 10-fold cross validation to separate our data set into training and testing sets. 10-fold is the most widely used number of fold in cross validation.

The procedure for each model is the same and summarized as follow:

1. For each of 10 experiments, we use 9 folds for training and the remaining one for testing,
2. we reiterate 10 times for each 10-fold cross validation experiments,
3. we select parameters of the best model and then we minimize the cross validation error rate.

In order to evaluate the classification suitability of models, we established the confusion matrix for the training and the testing samples. Moreover, we use the Receiver Operating Characteristic Curves (ROC) to evaluate classification performance.

The ROC curve presents the possible distributions of scores for the banks. We determine the optimal cut-off value that maximize the correct classification rate (default and healthy banks correctly classified) and then classify the banks as a potential default bank when the score is higher than the cut-off or as healthy if the score is lower than the cut-off value.

3 Running classifications methods

3.1 Multivariate adaptive regression splines (MARS) implementation

Introduced by Jerome Friedman in 1991, Multivariate Adaptive Regression Splines is a form of stepwise linear regression which can model non-linearity between variables.

MARS is based on three parameters:

1. the maximum number of basis functions (term)
2. the smooth parameter (called also the penalty parameter), which is recommended to be equal to 3
3. and the maximum number of iteration between variables (equal to 2) (see Andalib and Atry 2009).

In MARS, the basis function (term) is fitted to segregate independent variable intervals by using recursive splits. In this model, all possible splitting points are estimated with a linear spline (also called piecewise polynomials). The best splitting point (knot positions) is the one for which the model extensions minimize a squared error criterion. Knot is the point at which two polynomial pieces connect. The best splitting point is the one for which the model extension minimize a squared error criterion. Knots uses two-sided truncated power functions as spline basis functions, described in Eqs. (1) and (2)

$$h^{+}(x;t) = [+(x-t)]_{+} \tag{1}$$

$$h^{-}(x;t) = [-(x-t)]_{+} \tag{2}$$

where $[q]_{+} = \max\{0, q\}$ and t is a univariate knot. MARS is represented as a combination of piecewise linear or hinge functions. The later have a knot or hinge at t, are zero on the one side of the knot and are linear on the other side.

The MARS fit a linear model in basis functions $\left\{h_m(x)_{m=1}^{M}\right\}$:

$$\widehat{f}_M(x;\beta) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x), \tag{3}$$

where $h_m(x)$ is a basis function of the form described below in Eqs. (1) and (2), M is the number of linearly independent basis functions, and β_m is the unknown coefficient for te mth basis function.

As mentioned above, a knot is the point in a range at which the slope of the curve changes. Both the number of the knots and their placement are unknown at the beginning of the process. A stepwise procedure is used to find the best points to place the spline knots. In its most general form, each value of the independent variable is tested as a possible point for the placement of a knot. The model initially developed is overfit (forward phase). A statistical criterion (generalized cross-validation) that tests for a significant impact on a goodness of fit

Table 1 MARS models: N basis functions

	Model	Optimal basis func.	Misclass	ROC
2008	7	4	0.0193	0.8685
2009	7	3	0.0571	0.9562
2010	3	9	0.0334	0.9682
2011	9	4	0.0189	0.9783
2012	9	7	0.0170	0.9947
2013	8	4	0.0129	0.9510

measure is used to remove knots. Only those that have a significant impact on the regression are retained (backward phase).

MARS can perform regressions on binary variables. In binary mode, the dependent variable is converted into a 0 (non-failed banks) or a 1 (failed bank). Ordinary Least Square (OLS) regression is then performed.

We conduct our methodology in three steps:

1. We apply a 10-fold cross validation to divide our data sets into training and testing, and we repeat this process 10 times to get different partition of the database.
2. We choose the best number of basis functions according to the model that minimizes the misclassification rate. MARS does not display relationships in terms of the original 10 financial ratios, but reclassifies the target-predictor variable dependencies into a set of basis functions (BFs to represent the calculated splines. To find the optimal number of BFs and target values, MARS employs a forward/backward stepwise approach to determine the knot points in the data set. At the beginning, the model is tested by allowing more basis functions than are needed (100 BFS). Afterward, this model is shrunken to describe an optimal model. During this stage, basis functions are removed one by one from the over-fit model based on a 'residual sums of squares' criterion. The best model will have a GCV R^2 score converging to 1 (see Table 1).
3. We conduct the analysis of variance (ANOVA) decomposition procedure to assess the parameter relative importance based on the contributions from the input variables and the BFs (see Tables 2 and 3). In fact, interpretation of the MARS model is greatly facilitated through its ANOVA decomposition.

For purpose of simplification, only 2013 basis functions and corresponding equations are detailed in this section. Results for other years are detailed in the Appendix.

We start our analysis by detecting interaction between variables.

Table 1 delivers the best MARS model selected based on on the 10 times repeated K-fold Cross validation by minimizing the misclassification rate and maximizing ROC. The optimal model has the lowest value of GCV, an index for measuring generalized mean square errors. We use a backward method by minimizing the value of GCV .

Tables 2 and 3 displays the ANOVA decomposition of the built MARS models and exhibits the importance of each ratio in the model. Generalized Cross Validation (GCV) gives the amount of degradation in the model when a ratio is deleted. A model with minimum GCV should be chosen. In fact, GCV reaches its lowest value where the quantity of error is most minimized in the model.

For 2013, the liquidity ratio (TDTA) is the most important variables. Both of capital adequacy variables and the assets quality proxy (NPLGL) have 51.22, 30.61 and 40.53 percent of importance, whereas percent of the rest of variables are null.

Table 2 MARS versus CART: variables importance

		2008		2009		2010		2011		2012		2013	
		MARS	CART	MARS	CART	MARS	CART	MARS	CART	MARS	CART	MARS	CART
Capital Adeq.	EQTA	21.82	11.53	100	100	77.37		13.62	100	100	100	51.22	100
	EQTL	78.63				62.80	100	100	8.02			30.61	
Asset Qual.	LLRGL												
	LLRTA									43.24			
	NPLGL	100	100					25.53		65.71		40.53	
	NPLTA				7.03		6.03		4.46	29.51	1.66		
Earnings	ROA		10.16	32.75		35.46		24.17					2.18
	ROE						5.15			22.51			
Liquidity	TDTA					15.62						100	
	TLTD	21.82				100							

Table 3 MARS: ANOVA decomposition

	fun	SD	gcv	#bsfns	Variable
2008	1	0.0221	0.0156	1	NPLGL
	2	0.1219	0.0228	2	EQTL
					NPLGL
	3	0.0329	0.0159	1	EQTA
					TLTD
2009	1	0.2175	0.0910	2	EQTA
	2	0.0772	0.0495	1	EQTA
					ROA
2010	1	0.0450	0.0278	1	ROA
	2	0.1779	0.0420	1	EQTL
					TLTD
	3	0.0633	0.0300	2	EQTA
					ROA
	4	0.0948	0.0317	3	EQTA
					EQTL
	5	0.1872	0.0443	1	EQTA
					TLTD
	6	0.0345	0.0280	1	EQTA
					TDTA
2011	1	0.2556	0.0651	1	EQTL
	2	0.0576	0.0216	1	EQTL
					NPLGL
	3	0.2993	0.0198	1	EQTL
					ROA
	4	0.2502	0.0194	1	EQTA
					ROA
2012	1	0.0400	0.0091	1	ROE
	2	0.1724	0.0173	2	EQTA
					NPLGL
	3	0.0517	0.0099	1	EQTA
					NPLTA
	4	0.1033	0.0121	2	EQTA
					LLRTA
	5	0.0021	0.0079	1	LLRTA
					ROE
2013	1	0.1296	0.0145	1	TDTA
	2	0.0936	0.0093	1	EQTA
					TDTA
	3	0.0598	0.0086	1	EQTA
					NPLGL
	4	0.0278	0.0081	1	EQTL
					TDTA

'fun' lists the ANOVA function number. 'SD' gives the standard deviation of the function indicating its importance to the overall model. 'gcv' indicates the GCV score of the ANOVA function. 'bsfns' gives the number of basis functions in the ANOVA function. 'variable' gives the predictor variables associated with the ANOVA function

According to the ANOVA decomposition outputs, we note that function 1 has the greatest effect on the model with a GCV score of 0.01452 which means that the most important variable (TDTA) impacts significantly the dependent variable.

Basis functions (BF) in 2013 are as follow:

$$BF1 = max(0, TDTA - 0.9180);$$
$$BF6 = max(0, 0.0132 - EQTA) * BF1;$$
$$BF28 = max(0, TDTA - 0.9310);$$
$$BF32 = max(0, NPLGL - 0.1040);$$
$$BF37 = max(0, 0.0618 - EQTA) * BF32;$$
$$BF48 = max(0, EQTL - 0.0297) * BF28;$$

The above basis functions prove the non-linear relationships between the dependent and independent variables.

The final model is expressed as follow:

$$Y = 0.0055 - 462.114 * BF6 + 19.3076 * BF28 + 59.9474 * BF37 - 357.153 * BF48; \quad (4)$$

It appears that, for example, in BF1, on variable TDTA (liquidity proxy), data is grouped into two sets: the first one is assigned 0 for all TDTA values that are below a threshold (e.g., c: 0.9180) and the second set contains values exceeding 0.9180. The BF1 does not appear in the final model but it contributes in the construction of BF6. Indeed, BF6 is defined as a combination between two variables EQTA and TDTA. This basis function has a negative effect on the target variable only when the value of EQTA is less than 0.0132 and the value of TDTA is greater than 0.9180.

In the final model MARS produces a single regression equation, taking into account only BF6, BF28, BF37 and BF48 (see Table 4). Thus, BF1 and BF32 were removed from the regression model because they have an indirect effect and they especially create BFs assigned in the model.

The viability of the bank depends positively on the (TDTA) variable with a positive beta coefficient of 19.3976. Thus, the greater the liquidity caused by the total customer deposit, the better is the financial health of the bank. The MARS model also suggests a negative correlation with BF6. Thus, if a bank has a TDTA > 0.9180 and a level of (Total equity/Total asset) < 0.0132 then this has a negative influence on the score attributed to each bank studied.

3.2 CART algorithm to build tree classifiers

The CART (Classification and Regression Trees) algorithm proposed by Breiman et al. (1984) is a widely used statistical procedure. It produces classification and regression models via tree-based structure. It is based on a hierarchy of univariate binary decisions and operates by selecting the best variable for splitting the data into two groups at the root node. CART is a form of binary recursive partitioning in which partitions can be split into sub-partitions.

This classifier assigns a predicted class membership obtained under a specific measurement $(x_1, x_2, \ldots, x_k)$. Let X the measurement space of all possible values of X. Tree's classifiers are constructed by making repetitive splits of X and the subsequently obtained subsets of X. As consequence, a hierarchical structure is formed.

In Finance feature, Frydman et al. (1985) were the first who employed decision trees to forecast default. After that, many research highlighted the accuracy of this method to predict bankruptcy (Carter and Catlett 1987; Gepp et al. 2010).

Table 4 MARS: final model (after backward stepwise elimination)

	Basis fun	Coefficient	Variable	Knot	Parent
2008	0	0.1248			
	3	− 525.0998	NPLGL	0.1100	EQTL
	7	534.7530	NPLGL	0.0823	EQTL
	12	− 0.4066	NPLGL	0.3166	
	36	− 209.9519	TLTD	0.7878	EQTA
2009	0	0.8895			
	15	− 48.1527	EQTA	0.0130	
	17	47.3615	EQTA	0.0297	
	45	2537.8658	ROA	− 0.0222	EQTA
2010	0	0.0083			
	10	31.5524	TLTD	0.9418	EQTL
	14	− 2.2644	ROA	− 0.0083	
	18	− 11481.2155	EQTA	0.0334	ROA
	34	− 130.0023	EQTA	0.0143	EQTL
	51	6537.1156	EQTL	0.0315	EQTA
	53	− 6486.4966	EQTL	0.0302	EQTA
	55	2271.6071	ROA	− 0.0135	EQTA
	59	38.4564	TLTD	0.5167	EQTA
	63	192.9708	TDTA	0.9019	EQTA
2011	0	0.0027			
	2	14.7799	EQTL	0.0707	
	4	− 301.0973	NPLGL	0.0773	EQTL
	12	− 413.2053	ROA	− 0.0249	EQTL
	27	577.8011	ROA	− 0.0217	EQTA
2012	0	0.1712			
	15	− 0.0873	ROE	− 1.9251	
	19	411.0032	NPLGL	0.0767	EQTA
	30	55.1609	NPLTA	0.1372	EQTA
	43	587.5851	LLRTA	0.0197	EQTA
	45	− 1132.5747	LLRTA	0.0000	EQTA
	49	− 449.1551	NPLGL	0.1598	EQTA
	74	− 0.0138	LLRTA	0.0000	ROE
2013	0	0.0055			
	6	− 462.1144	EQTA	0.0132	TDTA
	28	19.3076	TDTA	0.9308	
	37	59.9474	EQTA	0.0618	NPLGL
	48	− 357.1526	EQTL	0.0298	TDTA

To build a tree by CART, the procedure should specify a number of parameters:

1. the splitter that will allow to visualize the left branch if the splitter's variable $<$ value split.
2. the competitor identifier variable. In our bank failure problem, the dependent variable is either bankrupt or non-bankrupt, so classification tree is suitable for our case. Under this assumption and rules, we implement CART on our data.

CART methodology consists of three steps: (i) Construction of maximum tree (ii) Choice of the right tree size (iii) and classification of new data using constructed tree.

Table 5 identifies the node competitors by order of improvement. Thus, at the upper levels of the tree there are more significant variables, and less significant at the bottom of the tree. Splitting in regression trees is made in accordance with squared residuals minimization algorithm which implies that expected sum variances for two resulting nodes should be minimized.

In fact, the process of finding the smallest tree that fits the data could reduce the number of important variables. The tree obtained, is the one that yields the lowest cross-validated error.

Table 2 exhibits the importance of each ratio in the building of CART tree. For example, from 2009 to 2013, capital adequacy variables are the most important ones. However, in 2008, the (NPLGL) is the most important variable.

Figures 1, 2, 3, 4, 5 and 6 illustrate the final classification Tree.

In 2008, as showed by Fig. 1, for the first node, splits are formed on (NPLGL) ratio and in a particular location (0.08). NPLGL produces the greatest "separation".

This tree is split at each branch by a decision boundary [left is Yes (class "0"), right is No (class "1")]. Thus, if a bank has a ratio $NPLGL \leq 0.08$, then the bank is considered as healthy (AB) and non-active (NAB) otherwise.

Classification must normally converge to the following clusters: among the 1242 banks analyzed in 2008, 97% are in (AB) group (1205 banks) and only 37 banks (3%) are in the (NAB) group. According to the right branch of the tree (NPLGL $>$ 0.08), among the 1242 banks, 160 are classified in (NAB) (class 1). However, only 33 banks were actually NAB, yielding to a misclassification of 127 (AB). Looking at the left branch of the tree ($NPLGL \leq 0.08$), 1078 banks are classified in the (AB) group. Among these 1082 banks, 1078 banks are actually active banks meaning that 4 (NAB) are misclassified. In the left branch of the second level, a second distinction is based on the target value of "-0.05" of the ratio ROA. For banks having a ROA less than -0.05, among the 1082 (AB), only 25 are classified as (NAB) and for banks having a ROA ratio > -0.05, 1057 are classified as (AB). Among the 25 banks classified as (NAB), only 2 are actually (NAB) meaning that 23 (AB) are misclassified. Among the 1057 banks classified as (AB), 1055 are actually (AB) meaning that 2 (NAB) are misclassified. In the right branch of the second level, a second distinction is based on the target value of "0.07" of the ratio EQTA. For banks having an EQTA less than 0.07, among the 160 (NAB), 82 are classified as (NAB) and for banks having an EQTA ratio $>$ 0.07, 78 are classified as (AB). Among the 82 banks classified as (NAB), 32 are actually (NAB) meaning that 50 (AB) are misclassified. Among the 78 banks classified as (AB), 77 are actually (AB) meaning that only 1 (NAB) is misclassified.

In addition to this hierarchical clustering, CART classification allows regulator to provide a lot of information about banks by pointing to banks that show signs of financial fragility.

Table 5 CART: root node competitors in order of improvement

	Competitor	Split	Improvement	Left node	Right node
2008	NPLGL	0.0838	0.3093	1082	160
	EQTA	0.0640	0.3067	164	1078
	NPLTA	0.0638	0.2874	1081	161
	EQTL	0.0947	0.2661	230	1012
	ROA	− 0.0116	0.2480	301	941
	ROE	− 0.1807	0.2383	243	999
	LLRTA	0.0171	0.2372	997	245
	LLRGL	0.0188	0.2219	890	352
	TDTA	0.8376	0.0893	772	470
	TLTD	1.0055	0.0472	897	345
2009	EQTA	0.0483	0.6272	226	982
	EQTL	0.0788	0.6040	261	947
	NPLGL	0.0590	0.4501	790	418
	NPLTA	0.0362	0.4205	750	458
	ROA	− 0.0117	0.4086	433	775
	LLRGL	0.0282	0.3272	865	343
	LLRTA	0.0212	0.3144	907	301
	TDTA	0.9073	0.2856	1028	180
	ROE	− 1.5635	0.2091	105	1103
	TLTD	0.8590	0.0587	705	503
2010	EQTL	0.0491	0.3710	175	902
	EQTA	0.0421	0.3708	193	884
	NPLGL	0.0695	0.2729	780	297
	NPLTA	0.0420	0.2710	745	332
	ROA	− 0.0005	0.2483	403	674
	LLRTA	0.0187	0.2030	766	311
	LLRGL	0.0288	0.1909	775	302
	TDTA	0.9078	0.1653	903	174
	ROE	− 0.2230	0.1386	244	833
	TLTD	0.9005	0.0274	821	256
2011	EQTA	0.0641	0.5471	145	810
	EQTL	0.0794	0.5294	132	823
	NPLGL	0.0736	0.4205	761	194
	ROA	− 0.0010	0.4144	257	698
	NPLTA	0.0366	0.3925	711	244
	LLRGL	0.0282	0.2993	736	219
	LLRTA	0.0183	0.2797	731	224
	ROE	− 0.3415	0.2789	117	838
	TDTA	0.9137	0.2248	855	100
	TLTD	0.8364	0.0678	631	324

Table 5 continued

	Competitor	Split	Improvement	Left node	Right node
2012	EQTA	0.0421	0.9451	57	826
	EQTL	0.0682	0.9451	57	826
	NPLTA	0.0474	0.7875	780	103
	ROA	− 0.0007	0.7782	139	744
	NPLGL	0.0750	0.7388	784	99
	ROE	− 0.2107	0.6264	57	826
	LLRGL	0.0228	0.5691	655	228
	LLRTA	0.0168	0.5257	723	160
	TDTA	0.9259	0.5239	837	46
	TLTD	0.7686	0.2056	450	433
2013	EQTA	0.0562	0.4755	36	814
	EQTL	0.0958	0.4664	44	806
	ROA	− 0.0002	0.4176	90	760
	TDTA	0.9321	0.3185	825	25
	LLRGL	0.0283	0.3091	769	81
	NPLTA	0.0237	0.2936	693	157
	NPLGL	0.0370	0.2936	693	157
	ROE	− 0.0584	0.2906	51	799
	LLRTA	0.0160	0.2808	739	111
	TLTD	0.7685	0.1811	406	444

This tab presents only the root node. Competitor: variables according to their improvement strength. Split: the best split value of this variable. Improvement: the split improvement. N Left: the number of observations with values of the splitter going to the left. N Right: the number of observations with values of the splitter going to the right

Indeed, in our case we have checked that banks that were misclassified by CART actually default in the following years (from 2008 to 2013).

A second interpretation linked to the early warning system, values of splitters allows the regulator to set target values that should be used to detect these suspected banks.

For 2008, we can consider that banks with simultaneously an $NPLGL \leq 0.08$ and $ROA \leq -0.05$ or NPLGL > 0.08 and $EQTA \leq 0.07$ should be classified in the group of fragile banks (bordering on default zone) which must be controlled.

4 Models accuracy and prediction results

In this section, models' accuracy analysis relies on the confusion matrix estimation. Type I error happens when the model incorrectly predicts a (NAB) to survive, whereas a type II error occurs when the model predicts an (AB) to go bankrupt. Thus, the predictive capacity of model's classification is based on sensitivity and specificity rates. Accuracy rates are the proportion of the total number of predictions that were correct and specificity is calculated as the number of correct (NAB) predictions divided by the total number of (NAB). The best specificity is 1, whereas the worst is 0. We notice that the choice of the cut-off is

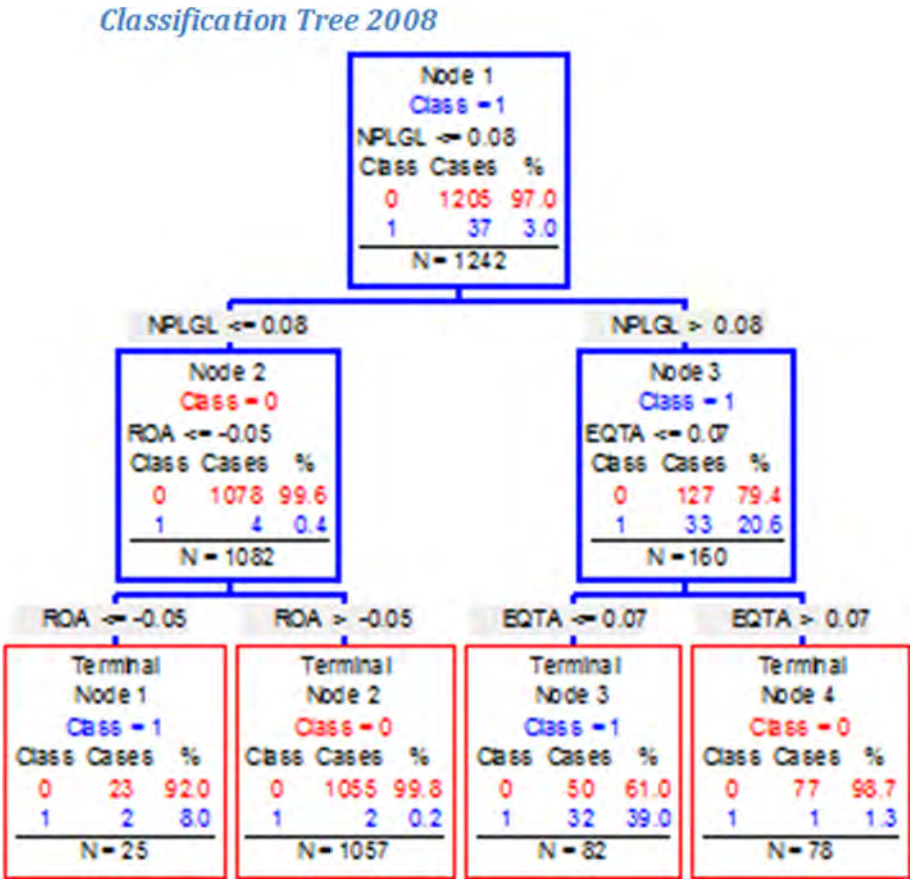

Fig. 1 Classification tree diagram 2008

crucial. Indeed, in a crisis period type I error decreases but at the same time, the number of (AB) classified as (NAB) increases, bringing on an suitableness cost in terms of economic policy.

Additionally, in this paper the quality of models has been measured with the use of the Receiver-Operating characteristic (ROC) curves and also the Area under (ROC) Curve. (ROC) curve shows the relation between specificity[1] and sensitivity[2] of the given test or detector for all allowable values of threshold (cut-off). In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (1-Specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a (sensitivity/specificity) pair corresponding to a particular decision threshold.

In this section, we test the performance of MARS, CART and K-MARS models by comparing its classification and predictions with the actual bankruptcies between 2008 and 2016.

[1] The specificity represents the number of the (AB) classified in the group of the (AB).

[2] The sensitivity represents the percentage of the (NAB) correctly classified.

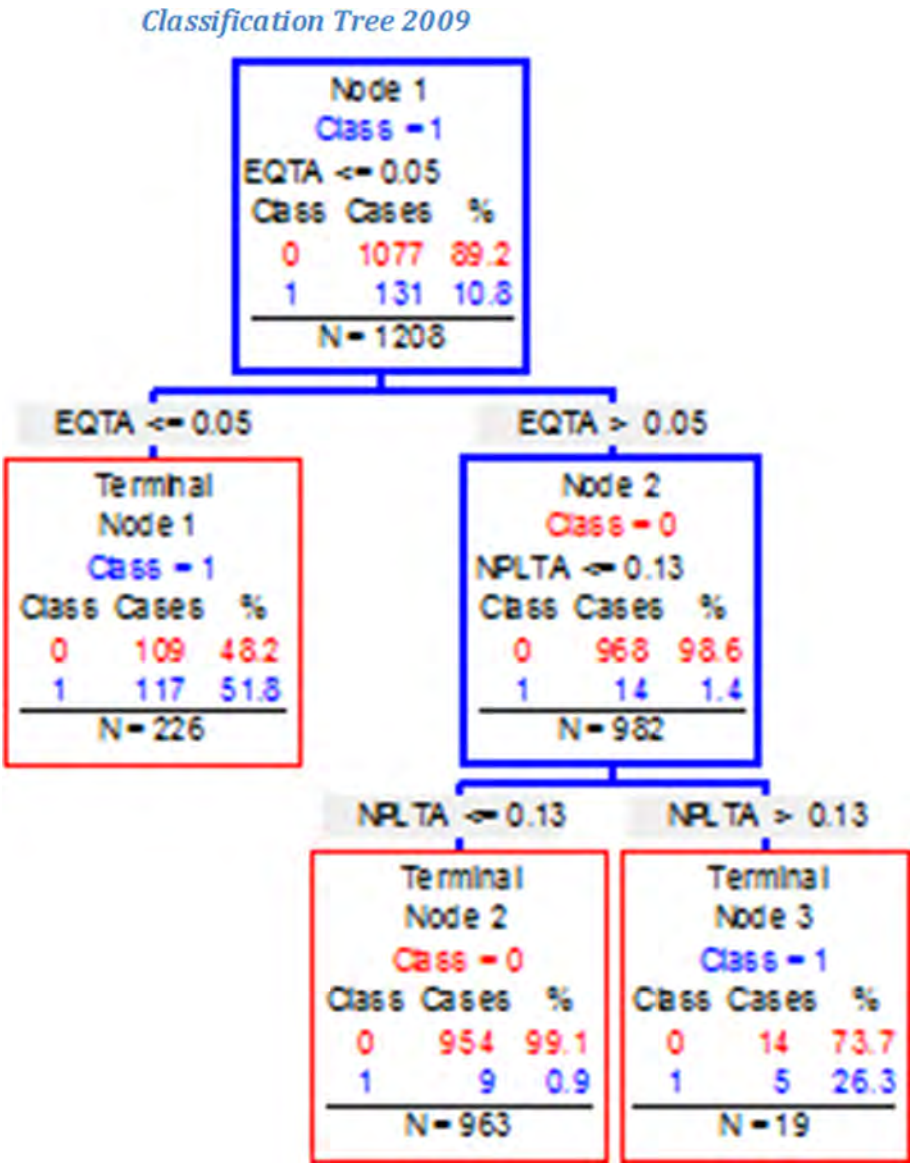

Fig. 2 Classification tree diagram 2009

4.1 MARS versus CART

Tables 6, 13 and 12 displays sensitivity, accuracy rates and misclassifications rates (Type I and II errors) both in Testing Sample and Training Sample.

For 2013 and in Testing Sample (TS), MARS model was able to correctly classify 99.29%. Only three (NAB) was misclassified (type I error: 20%) and three (AB) banks were considered by the model as (NAB) (type II error 0.36%). The same results were observed in

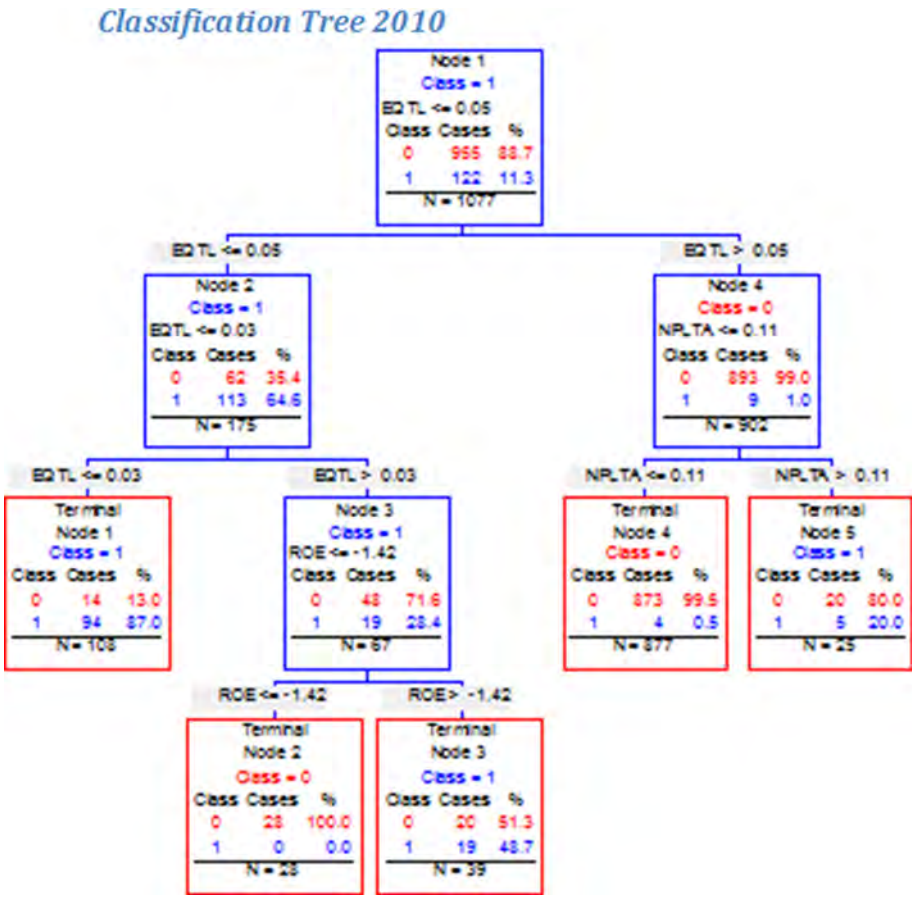

Fig. 3 Classification tree diagram 2010

the "Training Sample, (TRS)". The results for 2012 and 2011 were much better according to type I error, both in the "Testing Sample" and "Training Sample, (TRS)".

For 2008, 2009 and 2010, MARS underperformed CART by sensitivity scale. For example, Type I error rate was relatively high in 2008 [59.46% in the (TS)]. However, percentage of (AB) correctly predicted was 100, 98.05 and 98.32% respectively for the years 2008, 2009 and 2010 in the (TS). On the other hand, in terms of correct classification rate, MARS model performed better for 2008 (98.23%) than 2009 (94.7%) and 2010 (96.94%).

To summarize, we conclude that MARS model has a good predictive performance, measured by its ability to reduce the type I error and also by generating the best signal to to monitor fragile banks among the (AB).In this sense, MARS would be a powerful tool to enhance identification of the most outstanding patterns to forecast banking distress.

Tables 7, 13 and 12 highlighted also that (CART) produces a high level of correct classification for 2013.

We observe that 98.59% of the banks in the (TNS) are correctly classified and 97.53% of the banks in (TS) are classified in their adequate groups. CART for the (TNS) classify correctly all the failed banks (sensitivity: 100%) and for the testing only one (NAB) was

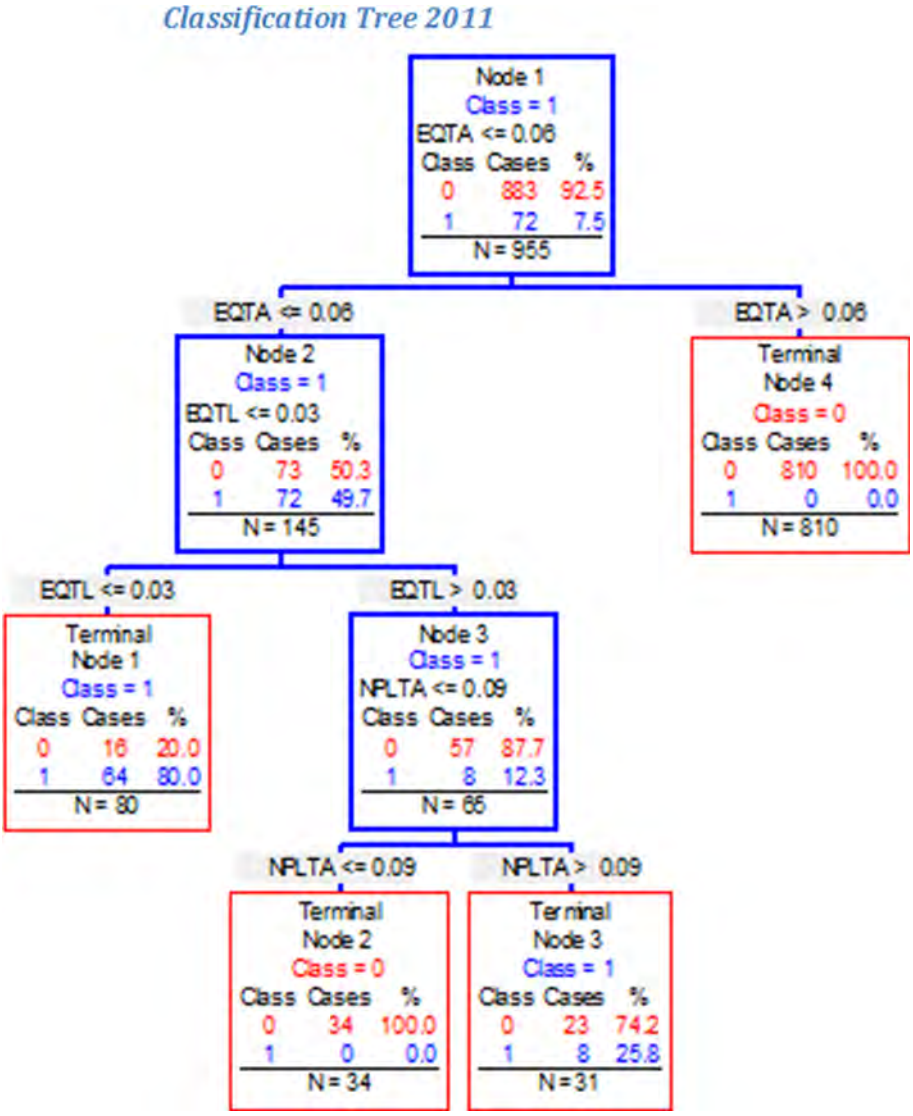

Fig. 4 Classification tree diagram 2011

predicted as (AB) (error type I: 6.67%). We also notice that, in the (TNS), Type I error is null for 2012.

However, results obtained for 2008, 2009 and 2010 showed that CART didn't procure a high correct classification rate. In fact, for 2009 we note a correct classification rate of about 89% in both (TNS) and (TS). Moreover, for this period, CART model provided a high misclassification rate. We noticed an average of 8.99% for Type II error and 8.59% for Type I error in the (TS).

Finally, in order to validate the failure prediction efficiency of the two non-parametric models, we propose to analyze the type II error as follows: From 2008 to 2013, we retrieve the

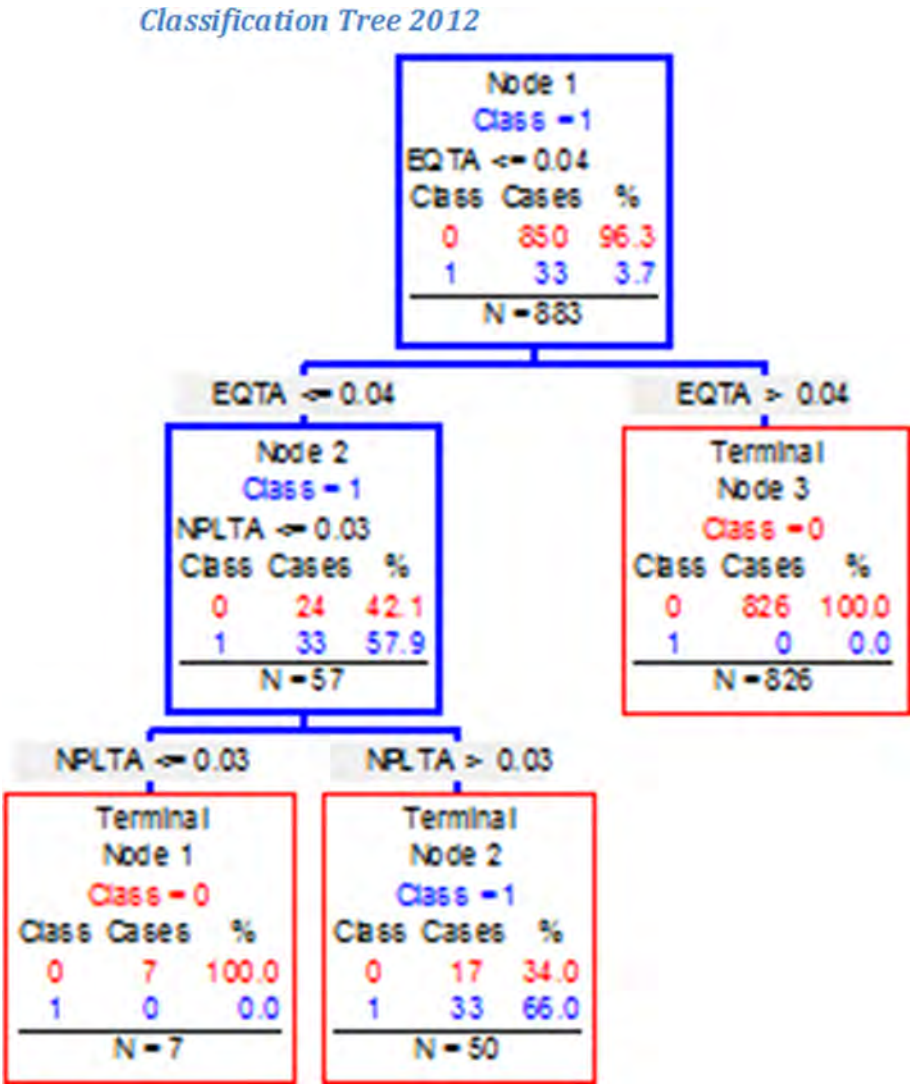

Fig. 5 Classification tree diagram 2012

(NAB) banks classified by MARS and CART models. We identify banks that are misclassified ((AB) that the model has considered as failing).

We check the survival of each bank for the next (N + 5) years. The results are summarized in Tables 8 and 9.

For example, for the CART model, 95 banks were incorrectly classified in 2009. 69.30% will default in 2010, 21.05% in 2011, 8.77% in 2012, and 0.88% in 2013. For MARS, the number of misclassified banks is less important. Among the 19 misclassified, which will be failing in the next 5 years, 94.74% will default in 2010 and the rest in 2012.

We can thus pick up the following conclusions: studied models provide a powerful EWS (Early Warning System). On average, CART put a red flag on 66.11% of unstable banks in

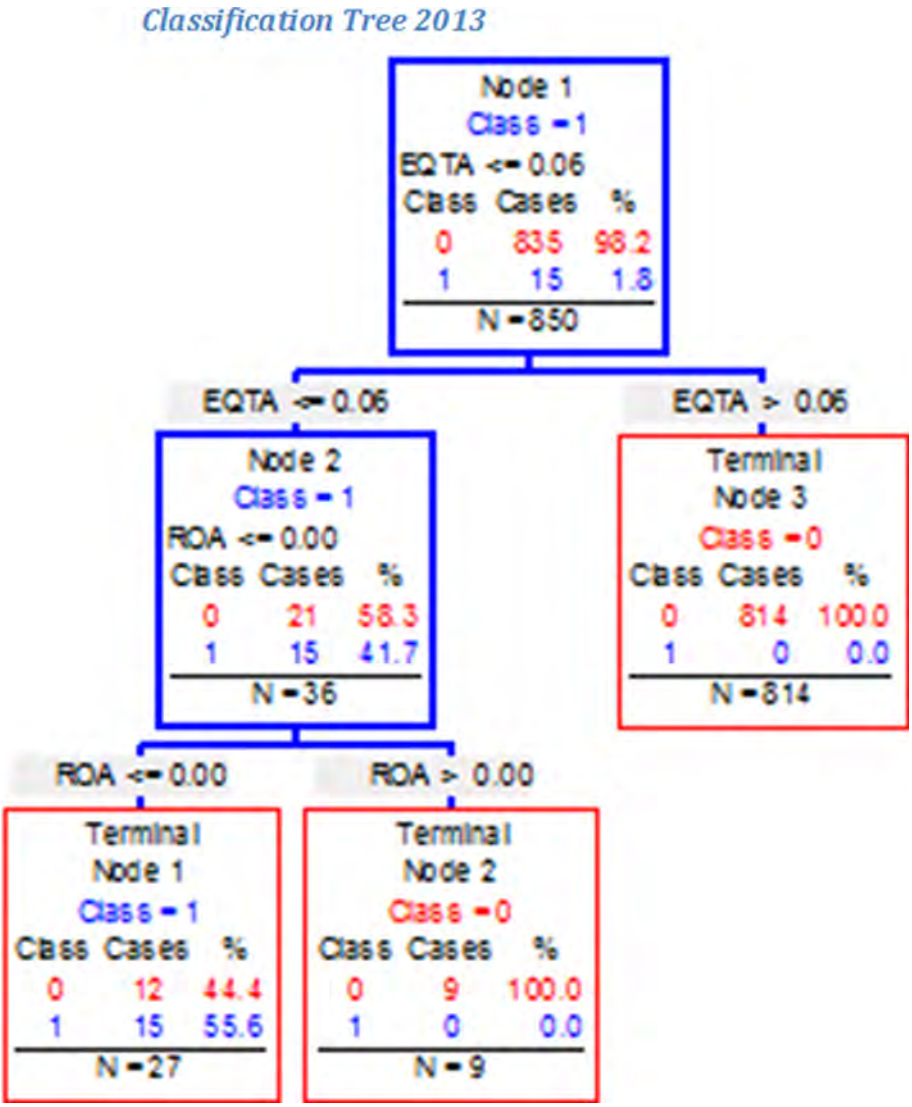

Fig. 6 Classification tree diagram 2013

year $N+1$. The MARS model is much more efficient and prevents the failure, in year $N+1$, of 80.81% of analyzed cases.

The CART model is much more penalizing in terms of classification. This explains the greater number of misclassified (AB).

4.2 Hybrid model accuracy

In order to improve the results of both models CART and MARS, we propose to build a hybrid model based on the classification model K-means and MARS.

Table 6 MARS accuracy and prediction results

MARS	2008		2009		2010	
	Learn	Test	Learn	Test	Learn	Test
C*(cut-off obtained from the ROC curve)	0.46139	0.58942	0.28475	0.44049	0.39741	0.46056
Failed banks correctly predicted	19	15	99	88	107	105
Non-failed banks correctly predicted	1204	1205	1048	1056	939	939
Failed banks incorrectly predicted	18	22	32	43	15	17
Non-failed banks incorrectly predicted	1	0	29	21	16	16
Incorrectly predicted total	19	22	61	64	31	33
Correctly predicted total	1223	1220	1147	1144	1046	1044
% of failed banks correctly predicted (Sensitivity)	51.35%	40.54%	75.57%	67.18%	87.70%	86.07%
% of type I error	48.65%	59.46%	24.43%	32.82%	12.30%	13.93%
% of non-failed banks correctly predicted (Specificity)	99.92%	100.00%	97.31%	98.05%	98.32%	98.32%
% of type II error	0.08%	0.00%	2.69%	1.95%	1.68%	1.68%
% of total incorrectly predicted (error rate)	1.53%	1.77%	5.05%	5.30%	2.88%	3.06%
% of total correctly predicted (correct classification rate)	98.47%	98.23%	94.95%	94.70%	97.12%	96.94%

Table 6 continued

MARS	2011		2012		2013	
	Learn	Test	Learn	Test	Learn	Test
C*(cut-off obtained from the ROC curve)	0.44601	0.48358	0.52336	0.24793	0.31903	0.24793
Failed banks correctly predicted	66	64	29	32	12	12
Non-failed banks correctly predicted	870	872	848	836	833	832
Failed banks incorrectly predicted	6	8	4	1	3	3
Non-failed banks incorrectly predicted	13	11	2	14	2	3
Incorrectly predicted total	19	19	6	15	5	6
Correctly predicted total	936	936	877	868	845	844
% of failed banks correctly predicted (Sensitivity)	91.67%	88.89%	87.88%	96.97%	80.00%	80.00%
% of type I error	8.33%	11.11%	12.12%	3.03%	20.00%	20.00%
% of non-failed banks correctly predicted (Specificity)	98.53%	98.75%	99.76%	98.35%	99.76%	99.64%
% of type II error	1.47%	1.25%	0.24%	1.65%	0.24%	0.36%
% of total incorrectly predicted (error rate)	1.99%	1.99%	0.68%	1.70%	0.59%	0.71%
% of total correctly predicted (correct classification rate)	98.01%	98.01%	99.32%	98.30%	99.41%	99.29%

Table 7 CART accuracy and prediction results

CART	2008		2009		2010	
	Learn	Test	Learn	Test	Learn	Test
Failed banks correctly predicted	34	33	122	120	118	114
Non-failed banks correctly predicted	1132	1096	954	954	901	893
Failed banks incorrectly predicted	3	4	9	11	4	8
Non-failed banks incorrectly predicted	73	109	123	123	54	62
Incorrectly predicted total	76	113	132	134	58	70
Correctly predicted total	1166	1129	1076	1074	1019	1007
% of failed banks correctly predicted (Sensitivity)	91.89%	89.19%	93.13%	91.60%	96.72%	93.44%
% of type I error	8.11%	10.81%	6.87%	8.40%	3.28%	6.56%
% of non-failed banks correctly predicted (Specificity)	93.94%	90.95%	88.58%	88.58%	94.35%	93.51%
% of type II error	6.06%	9.05%	11.42%	11.42%	5.65%	6.49%
% of total incorrectly predicted (error rate)	6.12%	9.10%	10.93%	11.09%	5.39%	6.50%
% of total correctly predicted (correct classification rate)	93.88%	90.90%	89.07%	88.91%	94.61%	93.50%

Table 7 continued

CART	2011		2012		2013	
	Learn	Test	Learn	Test	Learn	Test
Failed banks correctly predicted	72	70	33	32	15	14
Non-failed banks correctly predicted	844	811	833	827	823	815
Failed banks incorrectly predicted	0	2	0	1	0	1
Non-failed banks incorrectly predicted	39	72	17	23	12	20
Incorrectly predicted total	39	74	17	24	12	21
Correctly predicted total	916	881	866	859	838	829
% of failed banks correctly predicted (sensitivity)	100.00%	97.22%	100.00%	96.97%	100.00%	93.33%
% of type I error	0.00%	2.78%	0.00%	3.03%	0.00%	6.67%
% of non-failed banks correctly predicted (Specificity)	95.58%	91.85%	98.00%	97.29%	98.56%	97.60%
% of type II error	4.42%	8.15%	2.00%	2.71%	1.44%	2.40%
% of total incorrectly predicted (error rate)	4.08%	7.75%	1.93%	2.72%	1.41%	2.47%
% of total correctly predicted (correct classification rate)	95.92%	92.25%	98.07%	97.28%	98.59%	97.53%

Table 8 Analysis of type II error achieved with the CART and MARS

CART/testing sample	2008	2009	2010	2011	2012	2013
AB classified as DB	109	123	62	72	23	21
AB really bankrupt	95	114	49	47	14	5
AB really bankrupt (%)	87.16%	92.68%	79.03%	65.28%	60.87%	23.81%
MARS/testing sample	2009	2010	2011	2012	2013	
AB classified as DB	21	16	11	14	3	
AB really bankrupt	19	15	10	11	1	
AB really bankrupt (%)	90.48%	93.75%	90.91%	78.57%	33.33%	

Table 9 Analysis of predictability power of CART and MARS

N	N + 1 (%)	N + 2 (%)	N + 3 (%)	N + 4 (%)	N + 5 (%)
CART/testing sample					
2008	54.74	35.79	7.37	1.05	1.05
2009	69.30	21.05	8.77	0.88	
2010	67.35	18.37	14.29		
2011	59.57	34.04	2.13		4.26
2012	85.71	7.14		7.14	
2013	60.00		40.00		
Average	66.11	23.28	14.51	3.02	2.65
MARS/testing sample					
2009	94.74		5.26		
2010	86.67	6.67	6.67		
2011	60.00	30.00			10.00
2012	81.82	9.09		9.09	
2013			100.00		
Average	80.81	15.25	37.31	9.09	10.00

Clustering is a method of grouping (Anderberg 2014; Hartigan 1975; Jain and Dubes 1988) a set of objects into groups according to criteria predefined similarities between objects. Most clustering methods are based on a distance measure between two objects. Technically, clustering can be regarded as a minimization problem.

Let X the matrix of dimension data (N, n):

$$X = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ x_{N1} & \dots & x_{Nn} \end{pmatrix},$$

N corresponds to the number of banks, n the number of years and x_{ij} ratios variables.

Table 10 K-means accuracy and prediction results

K-means clustering	2008	2009	2010	2011	2012	2013
Failed banks correctly predicted	33	109	116	72	33	12
Non-failed banks correctly predicted	1075	894	821	794	800	806
Failed banks incorrectly predicted	4	22	6	0	0	3
Non-failed banks incorrectly predicted	130	183	134	89	50	29
Incorrectly predicted total	134	205	140	89	50	32
Correctly predicted total	1108	1003	937	866	833	818
% of failed banks correctly predicted (sensitivity)	89.19%	83.21%	95.08%	100.00%	100.00%	80.00%
% of type I error	10.81%	16.79%	4.92%	0.00%	0.00%	20.00%
% of non-failed banks correctly predicted (specificity)	89.21%	83.01%	85.97%	89.92%	94.12%	96.53%
% of type II error	10.79%	16.99%	14.03%	10.08%	5.88%	3.47%
% of total incorrectly predicted (error rate)	10.79%	16.97%	13.00%	9.32%	5.66%	3.76%
% of total correctly predicted (correct classification rate)	89.21%	83.03%	87.00%	90.68%	94.34%	96.24%

From N × n dimensional data set K-means algorithms allocates each data point to one of c clusters to minimize the within-cluster sum of squares:

$$\sum_{i=1}^{c} \sum_{k \in A_i} \| x_k - v_i \|^2 \tag{5}$$

where A_i is banks in the cluster i and v_i is the mean for these banks group over cluster i. This equation denotes actually a distance norm. In K-means clustering v_i is called the cluster prototypes, i.e. the cluster centers:

$$v_i = \frac{\sum_{x_k \in A_i} x_k}{N_i} \tag{6}$$

where N_i is the number of banks in A_i.

In our paper Z-score standardization is applied to find clusters and the number of cluster solutions is Two.

To implement the hybrid model, we proceed as follows:

Step 1 We apply the K-means clustering method on all banks. The cluster number is two (AB and NAB). In Table 10 we provide results of the classification based only on K-means.

Step 2 We do not run the MARS model on the data with their actual identifications. We use the classification generated by k-means. Hybrid's results are summarized in Table 11.

For 2013, the hybrid model provides a satisfactory rate of correct classification but we notice a slight gap between the training and the testing sample (98% against 96%). The model classifies correctly all the bankrupt banks both in (TNS) and (TS) testing sample (sensitivity:

Table 11 Hybrid model accuracy and prediction results

K-means MARS	2008		2009		2010	
	Learn	Test	Learn	Test	Learn	Test
C*(cut-off obtained from the ROC curve)	0.37009	0.20755	0.38761	0.35448	0.47587	0.40008
Failed banks correctly predicted	163	161	291	290	248	245
Non-failed banks correctly predicted	1071	1033	898	888	823	808
Failed banks incorrectly predicted	0	2	1	2	2	5
Non-failed banks incorrectly predicted	8	46	18	28	4	19
Incorrectly predicted total	8	48	19	30	6	24
Correctly predicted total	1234	1194	1189	1178	1071	1053
% of failed banks correctly predicted (Sensitivity)	100.00%	98.77%	99.66%	99.32%	99.20%	98.00%
% of type I error	0.00%	1.23%	0.34%	0.68%	0.80%	2.00%
% of non-failed banks correctly predicted (specificity)	99.26%	95.74%	98.03%	96.94%	99.52%	97.70%
% of type II error	0.74%	4.26%	1.97%	3.06%	0.48%	2.30%
% of total incorrectly predicted (error rate)	0.64%	3.86%	1.57%	2.48%	0.56%	2.23%
% of total correctly predicted (correct classification rate)	99.36%	96.14%	98.43%	97.52%	99.44%	97.77%

Table 11 continued

K-means MARS	2011		2012		2013	
	Learn	Test	Learn	Test	Learn	Test
C*(cut-off obtained from the ROC curve)	0.33772	0.36456	0.32136	0.23836	0.17253	0.08847
Failed banks correctly predicted	161	158	83	82	41	41
Non-failed banks correctly predicted	786	781	792	770	792	775
Failed banks incorrectly predicted	0	3	0	1	0	0
Non-failed banks incorrectly predicted	8	13	8	30	17	34
Incorrectly predicted total	8	16	8	31	17	34
Correctly predicted total	947	939	875	852	833	816
% of failed banks correctly predicted (Sensitivity)	100.00%	98.14%	100.00%	98.80%	100.00%	100.00%
% of type I error	0.00%	1.86%	0.00%	1.20%	0.00%	0.00%
% of non-failed banks correctly predicted (specificity)	98.99%	98.36%	99.00%	96.25%	97.90%	95.80%
% of type II error	1.01%	1.64%	1.00%	3.75%	2.10%	4.20%
% of total incorrectly predicted (error rate)	0.84%	1.68%	0.91%	3.51%	2.00%	4.00%
% of total correctly predicted (correct classification rate)	99.16%	98.32%	99.09%	96.49%	98.00%	96.00%

Table 12 Area under curve (AUC) results

ROC (area under curve)	MARS		CART		K-MARS	
	Training	Testing	Training	Testing	Training	Testing
2008	0.9453	0.8685	0.9464	0.8676	0.9999	0.9926
2009	0.9562	0.9297	0.9124	0.8865	0.9994	0.9947
2010	0.9757	0.9682	0.9717	0.9388	0.9998	0.9941
2011	0.9811	0.9783	0.9885	0.9693	0.9996	0.9926
2012	0.9983	0.9947	0.9900	0.9769	0.9997	0.9946
2013	0.9315	0.9510	0.9928	0.9589	0.9972	0.9971
Average	96.47%	94.84%	96.70%	93.30%	99.93%	99.43%

100%). Misclassification rate of (AB) in (TS) is higher than those of TNS (4.20% against 2.10%).

For the others years, we mainly observe the same results in term of correct classification rate. Moreover, the hybrid model provides a lower Type I error. However, we notice that the misclassification of (AB) (Type II error) is more important in the testing than in the learning sample.

Finally, according to the Area under Curve (ROC) (see Table 12), we conclude that on average MARS model provides a better accuracy results than CART model in the (TS) (94.84% against 93.3%). During all the period MARS outperforms CART, except for 2013, where the AUC of CART is slightly greater than MARS in the testing sample (95.89% against 95.10%).

The hybrid model K-MARS outperforms all the other models CART and MARS in both training and testing samples by using AUC.

To sum up, if we think in terms of average rates over the entire period we can confirm that MARS model provides better results than CART both in terms of correct classification rate and AUC for the testing sample. But it failed to classify correctly the bankrupt banks.

The combined model K-means and MARS is the best model in terms of accuracy in both training and testing sample in terms of average sensitivity (see Table 13; Figs. 7, 8, 9). Results of the training sample show the supremacy of the K-MARS model. It correctly classifies 98.91% of banks. It also provides a low type I error, meaning that only 0.19% of the bankrupt banks are misclassified. In (TS), results proved the supremacy of MARS in terms of correct classification rates (97.58% against 93.4% for CART and 97.04% for K-MARS) and in terms of type II error (1.15% against 6.7% for CART and 3.2% for K-MARS).

In (TS), the lowest level of type I error of K-MARS (1.16%) highlights the ability of the model to better classify (NAB). We note also that CART model delivers the highest level of type II error. This means that CART was able to predict the failure of banks in advance (The predictive power of the model). Also, in (TS), we observe the same trend during the period P1 (2008–2010) and P2 (2011–2013). Indeed, MARS model outperforms the other models in terms of specificity but it provides a less performance in terms of type I error. K-MARS was more accurate to classify banks during P1 that P2. In fact, in P2, we observe an upgrade of the correct classification rate only for MARS and CART and a decrease of the level of type I error for all the models.

Table 13 Comparative results in average

	2008–2013			2008–2010			2011–2013		
	MARS	CART	K-MARS	MARS	CART	K-MARS	MARS	CART	K-MARS
Training sample									
Sensitivity (%)	79.03	96.96	99.81	71.54	93.91	99.62	86.52	100	100
Type I error (%)	20.97	3.04	0.19	28.46	6.09	0.38	13.48	0	0
Specificity (%)	98.93	94.84	98.78	98.52	92.29	98.94	99.35	97.38	98.63
Type II error (%)	1.07	5.16	1.22	1.48	7.71	1.06	0.65	2.62	1.37
Error rate (%)	2.12	4.98	1.09	3.15	7.48	0.92	1.09	2.47	1.25
Correct classification (%)	97.88	95.02	98.91	96.85	92.52	99.08	98.91	97.53	98.75
Testing sample									
Sensitivity (%)	76.61	93.63	98.84	64.59	91.41	98.7	88.62	95.84	98.98
Type I error (%)	23.39	6.37	1.16	35.41	8.59	1.3	11.38	4.16	1.02
Specificity (%)	98.85	93.3	96.8	98.79	91.01	96.79	98.92	95.58	96.8
Type II error (%)	1.15	6.7	3.2	1.21	8.99	3.21	1.08	4.42	3.2
Error rate (%)	2.42	6.6	2.96	3.38	8.9	2.86	1.46	4.31	3.06
Correct classification (%)	97.58	93.4	97.04	96.62	91.1	97.14	98.54	95.69	96.94

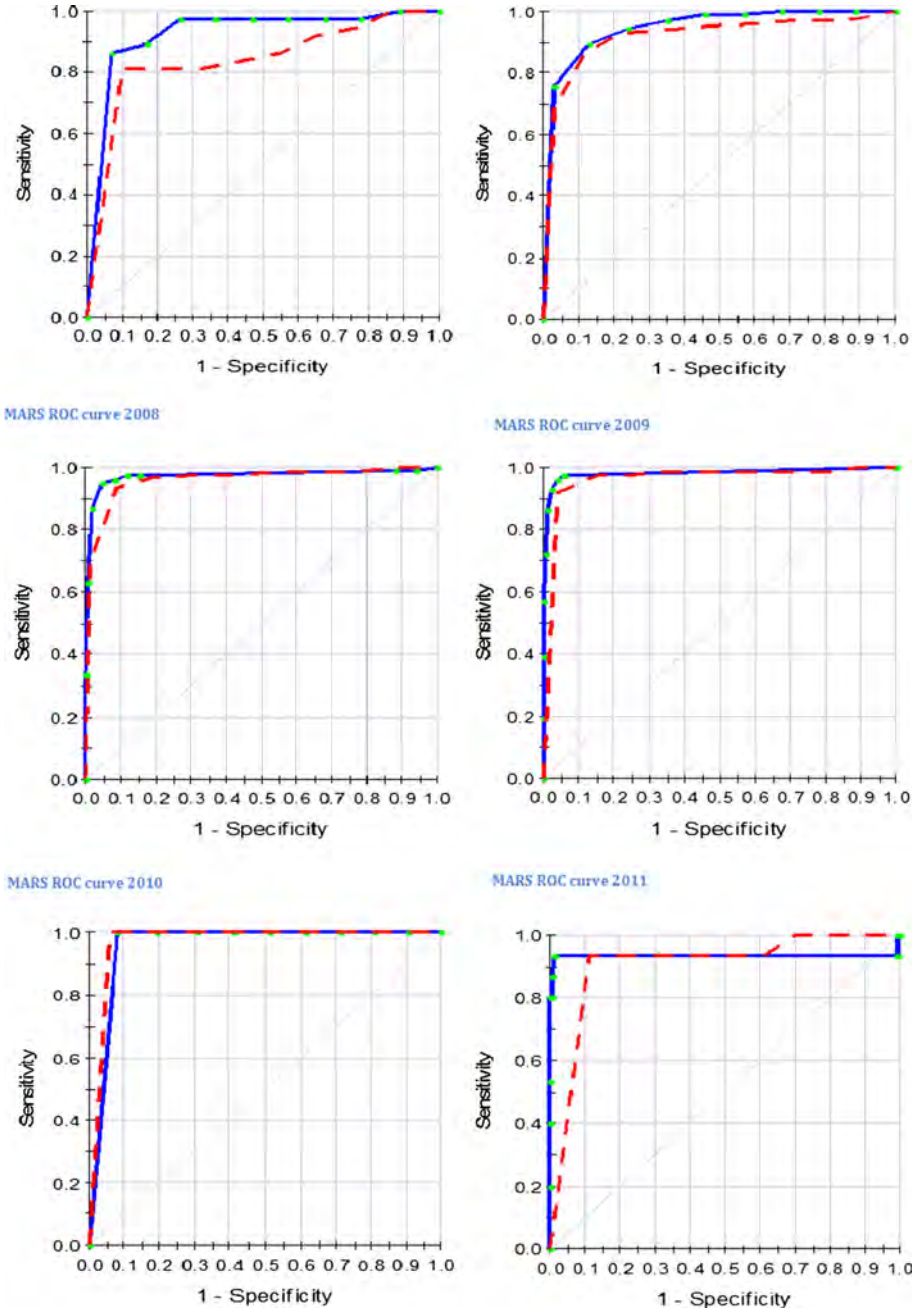

Fig. 7 ROC curve MARS

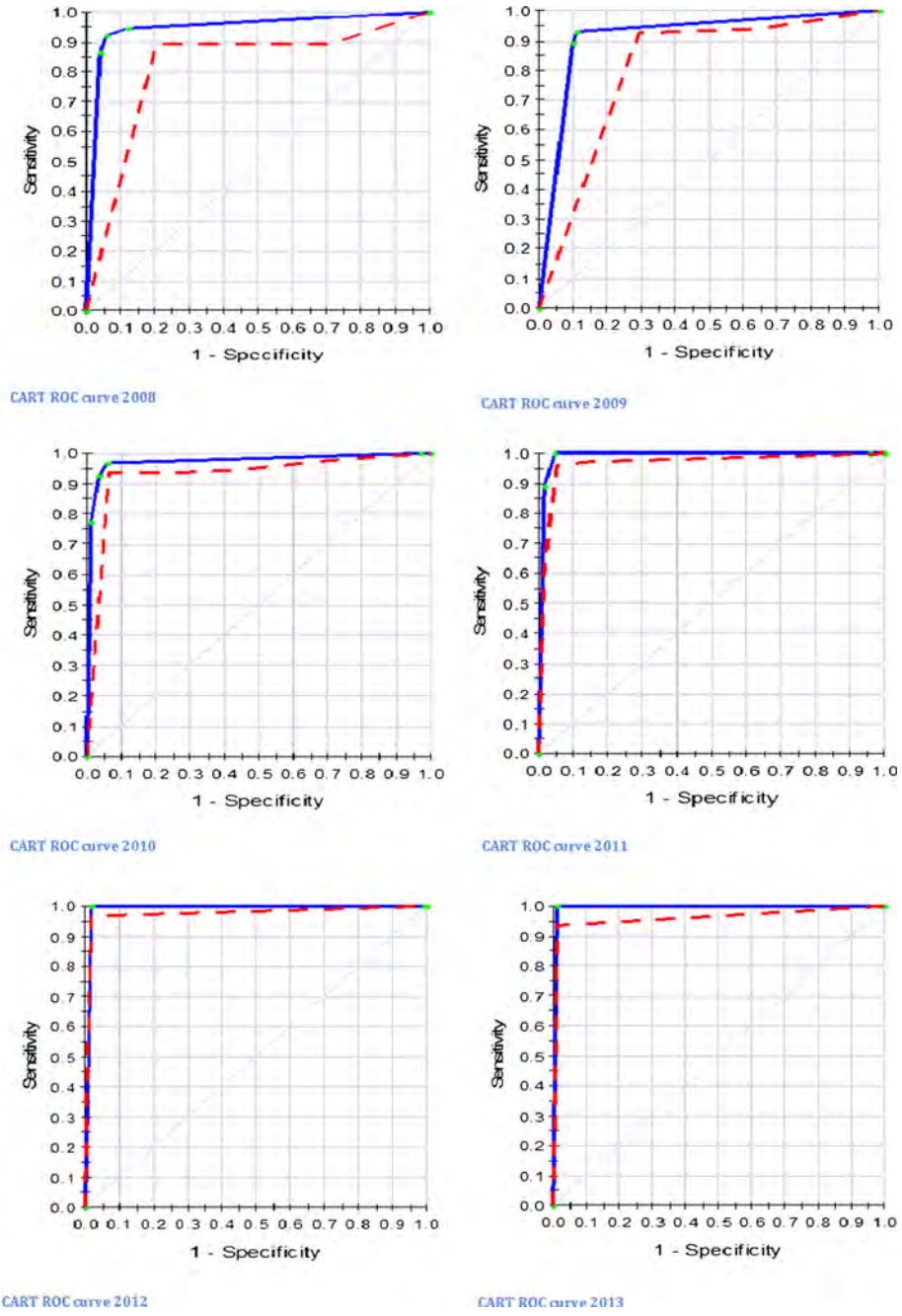

Fig. 8 ROC curve CART

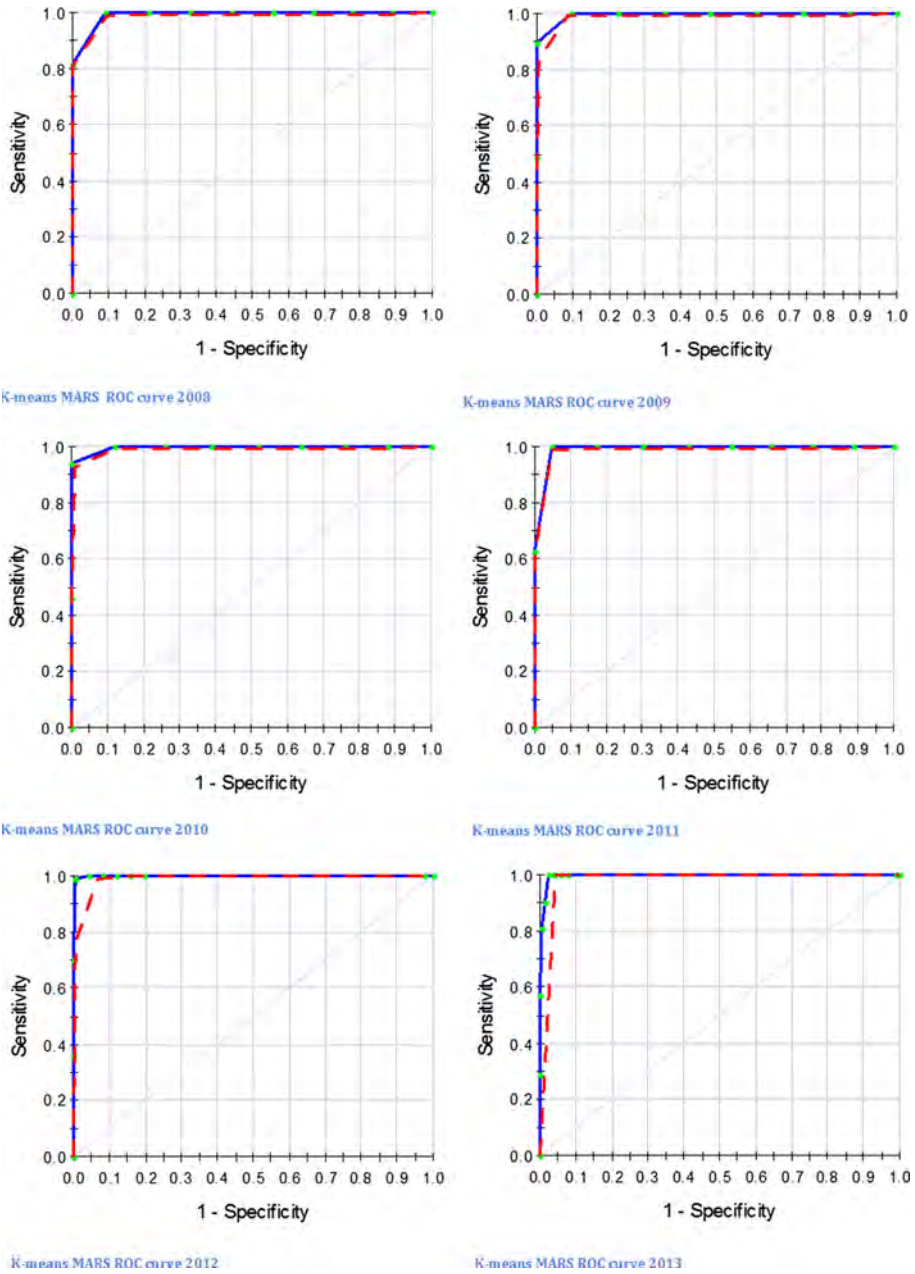

Fig. 9 ROC curve K-means MARS

5 Conclusion

In this paper, we developed a blend model based on two non-parametric classification models to study the bankruptcy of US banks. We provide a comparative approach between CART, MARS and K-means-MARS. Our main objective is to predict bank defaults some time before the bankruptcy occurs, and to build an early warning system based on CAMEL's ratios.

We based our empirical validation on a large panel of US banks gathered from both Bankscope and from the Federal Deposit Insurance Corporation.

The main contribution of our paper with regard to the existing literature is twofold:

- Methodological and conceptual: First, we propose, for the first time, a hybrid model that combines K-means and MARS models. We provide a comparative framework not only to non-parametric models but also to parametric models Logit and CDA (Affes and Hentati-Kaffel 2017).
- Empirical validation:
  - (i) Our study focuses on a large sample of US banks with different size. The paper analyses the behavior of banks over a 6-year period rich in events (it encompasses tow sub-periods, stress period 2008–2009 and a recovery one 2010–2013).
  - (ii) The comparative approach highlighted the supremacy of the proposed hybrid model in terms of accuracy classification for both training and validation samples.
  - (iii) The model enhanced the classification accuracy by 1% for the training sample
  - (iv) We established that MARS underperforms, by the misclassification rate of the bankrupt banks, notably for 2008 and 2009. Also, according to the Area under Curve (ROC), MARS model showed better accuracy results than CART model
  - (v) The results differ from 1 year to another, but a general behavior for all distressed banks could be conducted. CART classification shows that among the 10 tested ratios, the most important predictors are capital adequacy variables. Also, we note that the asset quality ratios (NPLTA) and (NPLGL) are much more important than the other two components (LLRTA) and (LLRGL). According to MARS the most important variables was also the components of the capital adequacy. The Liquidity variables (TLTD and TDTA) have an importance in detecting bank failure only in 2010 and 2013. We note that, with respect to parametric models (see Affes and Hentati-Kaffel 2017) the asset quality was also an important component to explain the financial conditions of banks (except for 2009 and 2010).

Finally, as mentioned in the introduction, the ultimate goal of this paper is to provide regulators and investors an early warning model. The study we carried out meets this objective in two ways:

First, our detailed study shows how for a CART model it is possible to identify and detect target variables that enable banks in fragile financial situations to be detected in advance. For example, in 2008 banks which are in fragile and alarming situations are those who presents the following characteristics simultaneously: (i) $NPLGL \leq 0.08$ and $ROA \leq -0.05$ (ii) NPLGL > 0.08 and $EQTA \leq 0.07$.

Second, MARS and CART models are a useful tool to identify in advance financial institutions in stress and so will be deserved with a special attention by supervisors. On average, CART put a red flag on 66.11% of unstable banks in year N + 1. The MARS model is much more efficient and prevents the failure, in year N + 1, of 80.81% of analyzed cases.

Finally, we believe that further extensions can be developed by including more financial variables and macroeconomic variables.

Appendix A: Basis function

Basis functions For 2008

$$BF2 = max(0, 0.0801 - EQTL);$$
$$BF3 = max(0, NPLGL - 0.1099) * BF2;$$
$$BF7 = max(0, NPLGL - 0.0823) * BF2;$$
$$BF10 = max(0, 0.0249 - EQTA);$$
$$BF12 = max(0, 0.3165 - NPLGL);$$
$$BF36 = max(0, TLTD - 0.7878) * BF10;$$

The final model is expressed as follows:

$$Y = 0.1248 - 525.1 * BF3 + 534.753 * BF7 - 0.4066 * BF12 - 209.952 * BF36; \quad (7)$$

The basis function BF2 does not appear in the model but it contributes in the construction of others basis function (BF3 and BF7). From BF2, on variable EQTL (capital adequacy proxy), data is grouped into two sets: the first one is assigned 0 for all EQTL values that are more than 0.08 and the second set contains the elevation values that are below a threshold (e.g., c = 0.08) Indeed, the BF3 is defined as a combination between NPLGL and EQTL. This basis function has a negative effect on the target variable only when the value of NPLGL exceeds 0.109 and the value of EQTL is lower than 0.08. The basis function (BF7) which is a combination of (BF1) with NPLGL, have a positive impact on the output. In other word, a value of NPLGL greater than 0.08 multiplied by a value of EQTL lower than 0.08 affects negatively the target 'Y'. Moreover, a decrease in the value of NPLGL above 0.316 (in BF12) will decrease the variable 'Y'. The negative effect of (BF36) appears only when the value of TLTD at least 0.787 and BF10 is positive (for EQTA lower than 0.0248).

Basis functions For 2009

$$BF15 = max(0, EQTA - 0.0129);$$
$$BF17 = max(0, EQTA - 0.0296);$$
$$BF44 = max(0, 0.0540 - EQTA);$$
$$BF45 = max(0, ROA + 0.0222) * BF44;$$

The final Model is expressed as below:

$$Y = 0.8895 - 48.1527 * BF15 + 47.3615 * BF17 + 2537.87 * BF45; \quad (8)$$

We note that BF15 is zero for value of EQTA lower than 0.0129. A negative sign for the estimated beta factor of BF15 indicates a decrease of the output variable. On the other hand a value of EQTA greater than 0.0296 (BF17) effects positively the target variable. We also note the presence of interaction between predictor variables, which means that the effect of a predictor on the target variable may depend on the value of another predictor. We see in the definition of the BF45 that the effect of the variable ROA on the output variable depends on the value of the ratio EQTA. The effect of this interaction can be explained as follow. If the value of EQTA is lower than 0.054 and the value of ROA is above -0.022, it has a positive impact on the target variable.

Basis functions For 2010

$$BF2 = max(0, 0.0799 - EQTL);$$
$$BF10 = max(0, 0.9418 - TLTD) * BF2;$$
$$BF13 = max(0, ROA + 0.0083);$$
$$BF14 = max(0, -0.0083 - ROA);$$
$$BF18 = max(0, 0.0333 - EQTA) * BF13;$$
$$BF32 = max(0, 0.0283 - EQTL);$$
$$BF34 = max(0, 0.0143 - EQTA) * BF32;$$
$$BF49 = max(0, EQTA - 0.0635);$$
$$BF50 = max(0, 0.0635 - EQTA);$$
$$BF51 = max(0, EQTL - 0.0315) * BF50;$$
$$BF53 = max(0, EQTL - 0.0301) * BF50;$$
$$BF55 = max(0, ROA + 0.0135) * BF50;$$
$$BF59 = max(0, TLTD - 0.5167) * BF50;$$
$$BF63 = max(0, TDTA - 0.9018) * BF49;$$

The final model is expressed as follows:

$$\begin{aligned} Y = {} & 0.0083 + 31.5524 * BF10 - 2.2643 * BF14 - 11481.2 * BF18 \\ & - 130.002 * BF34 + 6537.12 * BF51 \\ & - 6486.5 * BF53 + 2271.61 * BF55 + 38.4564 * BF59 + 192.971 * BF63; \end{aligned} \quad (9)$$

The BF2 does not appear in the model but it contributes to compute the BF10. Indeed, for values of TLTD and EQTL respectively lower than 0.94 and 0.0798, the BF10 has a positive effect on the output variable. For a value of ROA inferior than -0.00827 (BF14) impacts negatively the target value. We also note that the interaction between EQTA and ROA (BF18) has a negative effect on the target variable for a value of EQTA lower than 0.033 and ROA upper than -0.008. Moreover the negative impact of the BF34 appears only when EQTA is below 0.0143 and EQTA lower than 0.028. The BF49 and BF50 together define a piecewise linear function of EQTA with a knot of 0.0635. The values of these basis functions are positives when the value of EQTA is respectively superior to 0.0635 and inferior than 0.0635. The basis function BF50 is not involved in the model but it's used to compute the BF51, BF53, BF55 and BF59. Indeed, when the value of EQTL is above 0.0314 and the EQTA is lower than 0.0635, then the output variable will increase. On the other hand, a value of EQTL greater than 0.03 and EQTA below 0.0635, have a negative effect on the target variable. A positive impact of the BF55 and BF59 on the target variable appear respectively for ROA bigger than -0.0135 and TLTD higher than 0.516 multiplied by the BF50. The BF63 has a positive effect on the output variable for a value of TDTA superior to 0.9 and an EQTA higher than 0.0635.

Basis functions For 2011

$$BF2 = max(0, 0.0707 - EQTL);$$
$$BF4 = max(0, 0.0773 - NPLGL) * BF2;$$
$$BF12 = max(0, -0.02489 - ROA) * BF2;$$
$$BF25 = max(0, 0.0367 - EQTA);$$
$$BF27 = max(0, -0.0216 - ROA) * BF25;$$

The final model is expressed as follow:

$$Y = 0.0026 + 14.7799 * BF2 - 301.097 * BF4 - 413.205 * BF12 + 577.801 * BF27; \quad (10)$$

As it can be seen, the positive effect of the BF2 appears only when the value of EQTL is lower than 0.07. The BF2 was used in the construction of other basis functions (BF4 and BF12). In fact, we note that the BF4 and BF12 have a negative impact on the target variable for values of NPLGL inferior than 0.077 and ROA lower than -0.0248, multiplied by BF2. The BF27 is a function of the variable of ROA with a knot of -0.0216 multiplied by the BF25. It means that when the value of ROA is lower than -0.0216 and the value of EQTA is inferior than 0.0367, we note a positive impact on the target variable.

Basis functions For 2012

$$
\begin{aligned}
&BF2 = max(0, 0.0667 - EQTA);\\
&BF12 = max(0, 0.0049 - EQTA);\\
&BF14 = max(0, 0.0325 - EQTA);\\
&BF15 = max(0, ROE + 1.9250);\\
&BF16 = max(0, -1.9250 - ROE);\\
&BF19 = max(0, NPLGL - 0.0767) * BF14;\\
&BF30 = max(0, 0.1372 - NPLTA) * BF2;\\
&BF43 = max(0, LLRTA - 0.0197) * BF14;\\
&BF45 = max(0, LLRTA - 0) * BF12;\\
&BF49 = max(0, NPLGL - 0.1597) * BF14;\\
&BF74 = max(0, LLRTA - 0) * BF16;
\end{aligned}
$$

The final model is expressed as follow:

$$
\begin{aligned}
Y = {} & 0.1712 - 0.0873 * BF15 + 411.003 * BF19 + 55.1609 * BF30 + 587.585 * BF43\\
& - 1132.57 * BF45 - 449.155 * BF49 - 0.0138 * BF74; \qquad (11)
\end{aligned}
$$

The basis functions BF2, BF12 and BF14 appear in the construction of others basis functions (BF19, BF30, BF43, BF45 and BF49) only when the values of EQTA are respectively inferior to 0.0667, 0.00495 and 0.0325. The BF30 has a positive impact on the target when the value of NPLTA is lower than 0.137 multiplied by BF2. A positive impact on the output appears for values of NPLGL superior to 0.076 and a value of LLRTA above 0.197, multiplied by BF14. But for value of NPLGL upper than 0.159 and an EQTA lower than 0.0325, we note a negative effect on the target variable. Moreover we note the presence of interaction between the LLRTA ratio and two other variables (EQTA and ROE). The positive effect of BF45 appears for a value of LLRTA higher than 0 and EQTA lower than 0.00495. The BF74 have a positive effect on the target value when LLRTA is superior to 0 and ROE is less than -1.925. We also note that a value of ROA higher than -1.925 (BF15) impacts negatively the target.

References

Affes, Z., & Hentati-Kaffel, R. (2017). Predicting us banks bankruptcy: Logit versus canonical discriminant analysis. *Computational Economics*, 1–46. https://doi.org/10.1007/s10614-017-9698-0.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

Anandarajan, M., Lee, P., & Anandarajan, A. (2001). Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. *Intelligent Systems in Accounting, Finance and Management*, *10*(2), 69–81.

Andalib, A., & Atry, F. (2009). Multi-step ahead forecasts for electricity prices using narx: A new approach, a critical analysis of one-step ahead forecasts. *Energy Conversion and Management*, *50*(3), 739–747.

Anderberg, M. R. (2014). *Cluster analysis for applications: Probability and mathematical statistics: A series of monographs and textbooks* (Vol. 19). New York: Academic press.

Barr, R. S., Siems, T. F., et al. (1994). *Predicting bank failure using DEA to quantify management quality*. Federal Reserve Bank of Dallas: Technical report.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, *4*, 71–111.

Boyacioglu, M. A., Kara, Y., & Baykan, Ö. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications*, *36*(2), 3355–3366.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton: CRC Press.

Carter, C., & Catlett, J. (1987). Assessing credit card applications using machine learning. *IEEE Expert*, *2*(3), 71–79.

Chen, M.-Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, *38*(9), 11261–11272.

Cole, R. A. & Gunther, J. (1995). A camel rating's shelf life. Available at SSRN 1293504.

Collier, C., Forbush, S., Nuxoll, D., & O'Keefe, J. P. (2003). The SCOR system of off-site monitoring: Its objectives, functioning, and performance. *FDIC Banking Review Series*, *15*(3), 17.

De Andrés, J., Lorca, P., de Cos Juez, F. J., & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS). *Expert Systems with Applications*, *38*(3), 1866–1875.

Demirgüç-Kunt, A., & Detragiache, E. (1997). *The determinants of banking crises-evidence from developing and developed countries* (Vol. 106). Washington, D.C.: World Bank Publications.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1–67.

Frydman, H., Altman, E. I., & KAO, D.-L. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*, *40*(1), 269–291.

Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of Forecasting*, *29*(6), 536–555.

Godlewski, C. J., et al. (2003). *Modélisation de la prévision de la défaillance bancaire une application aux banques des pays emergents*. Document de travail LARGE: Université Robert Schuman Strasbourg.

Hanweck, G. A., et al. (1977). *Predicting bank failure*. Board of Governors of the Federal Reserve System (US): Technical report.

Hartigan, J. A. (1975). *Clustering algorithms*. New york: Wiley.

Iturriaga, F. J. L., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications*, *42*(6), 2857–2869.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Englewood Cliffs: Prentice-Hall Inc.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review*, *79*(4), 1011–1038.

Kolari, J., Glennon, D., Shin, H., & Caputo, M. (2002). Predicting large US commercial bank failures. *Journal of Economics and Business*, *54*(4), 361–387.

Lanine, G., & Vander Vennet, R. (2006). Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, *30*(3), 463–478.

Lenard, M. J., Alam, P., & Madey, G. R. (1995). The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences*, *26*(2), 209–227.

Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance*, *1*(3), 249–276.

McKee, T. E., & Greenstein, M. (2000). Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, *19*(3), 219–230.

Odom, M. D. & Sharda, R. (1990). A neural network model for bankruptcy prediction. In *IJCNN International Joint Conference on neural networks, 1990* (pp. 163–168).

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, *18*(1), 109–131.

Pantalone, C. C., & Platt, M. B., et al. (1987). Predicting commercial bank failure since deregulation. *New England Economic Review*, 37–47.

Sánchez-Lasheras, F., de Andrés, J., Lorca, P., & de Cos Juez, F. J. (2012). A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications*, *39*(8), 7512–7523.

Sinkey, J. F. (1979). *Problem and failed institutions in the commercial banking industry*. London: Wiley.

Thomson, J. B. (1991). Predicting bank failures in the 1980s. *Economic Review-Federal Reserve Bank of Cleveland*, *27*(1), 9.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, *80*(389), 73–83.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, *116*(1), 16–32.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, *22*, 59–82.