Research in International Business and Finance xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research in International Business and Finance

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

Full length Article

# Predicting bank failure: An improvement by implementing machine learning approach on classical financial ratios

Hong Hanh Le, Jean-Laurent Viviani*

*IGR-IAE-Rennes, CREM, 11 rue J. Macé, 35700 Rennes, France*

ARTICLE INFO

*JEL classification:*
G33

*Keywords:*
Failure prediction
Intelligent techniques
Artificial neural network
Support vector machines
K-nearest neighbors
US banks

ABSTRACT

This research compares the accuracy of two approaches: traditional statistical techniques and machine learning techniques, which attempt to predict the failure of banks. A sample of 3000 US banks (1438 failures and 1562 active banks) is investigated by two traditional statistical approaches (Discriminant analysis and Logistic regression) and three machine learning approaches (Artificial neural network, Support Vector Machines and k-nearest neighbors). For each bank, data were collected for a 5-year period before they become inactive. 31 financial ratios extracted from bank financial reports covered 5 main aspects: Loan quality, Capital quality, Operations efficiency, Profitability and Liquidity. The empirical result reveals that the artificial neural network and k-nearest neighbor methods are the most accurate.

## 1. Introduction

According to the Federal Deposit Insurance Corporation (FDIC), during 2008–2014 more than 500 banks declared as failures in the United States of America. The cost of failure per dollar of failed-bank assets is already high and may continue to rise. Consequently, the more banks go bankrupt, the higher the cost of resolving after-failure events. Year-end 2013, FDIC estimated that the total cost to the deposit insurance funds of resolving these failed banks is as high as 30 billion US dollars.

Banks are considered as failures if the state or bank regulator forces them to close because of insolvency problems. Because of the strong interconnection between banks and their essential role in financing the economy, the failure of banks is more threatening for the economy than the failure of other business firms. In some cases, the bankruptcy of one bank can cause a knock-on effect, which can spread quickly and have a negative impact on other banks (systemic risk). Hence, detection of bank failure before it occurs and try to avoid them is mandatory. In this research, we execute machine learning techniques which have been claimed to improve the prediction of sbank failure.

Starting with seminal research studies, Beaver (1966) and Altman (1968) built statistical models to predict firm failure based on accounting ratios. Since then, numerous studies have been advanced using different financial ratios, samples and periods. In parallel with the development of computational sciences, many different interesting approaches were explored to promote the power of technology. In order to help researchers better understand this complex field, Ravi Kumar and Ravi (2007) present a comprehensive review of the applications of prediction techniques to solve bankruptcy prediction problems of banks and firms. One of the intelligent technique families known as 'Machine learning' becomes more popular among researchers and practitioners. A commonly cited formal definition of machine learning, proposed by a computer scientist (Lantz, 2013, p. 10) explained that a machine is said to learn if it is able to take experience and utilize it such that its performance improves up on a similar experience in the future. More formally, according to Mitchell (1997), a computer program is said to learn from experience E with respect to some class of tasks T

* Corresponding author.
*E-mail addresses:* hanh_vn90@yahoo.com (H.H. Le), jean-laurent.viviani@univ-rennes1.fr (J.-L. Viviani).

http://dx.doi.org/10.1016/j.ribaf.2017.07.104
Received 17 February 2017; Received in revised form 3 June 2017; Accepted 4 July 2017

and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. In this study, we want to examine the effectiveness of these methods compared with more traditional statistical techniques.

The main contributions of the paper are the following. First it proposes a comparison of traditional statistical techniques: Linear Discriminant analysis (LDA) and Logistic regressions (Logistic) to machine learning techniques on predicting the failure of US banks. Machine learning techniques (k- Nearest neighbors (k-NNs), artificial neural networks (ANN) and Support Vector Machines (SVMs)) had not been systematically compared to predict bank failure. Moreover, the empirical results of previous studies are unclear (see the recent paper of López and Pastor Sanz, (2015)). Secondly, we use a large number of financial ratios (31) for 5 years before banks become inactive (inactivity can be due to bankruptcy, illiquidity, merging, or insolvency). The large number of ratios provides a means of covering all bank financial characteristics: loan quality, capital quality, operations, profitability and liquidity and of determining the ratios with the best failure prediction power. This diversity is justified by our analysis of the "Material loss review" of 102 banks from FDIC reports since 2009–2015, which shows that banks fail for various reasons (loan problems, profit reduction, credit risk, ineffective board of directors and management). Thirdly, we test these various techniques on a large sample of 3000 US banks (1438 failed, 1532 active) during the crisis and post-crisis period (2008–2014). This period deserves an in-depth study due to the change in financial environment and banking techniques: fall of real estate prices, biased pricing methods, new financial products and risks (Demyanyk and Hasan, 2010; López and Pastor Sanz, 2015), the reasons for bank failures could be different (or not) from those previously observed. Better knowledge of bank failure determinants is also very important for regulators (in the Basel 3, 4 reforms perspective).

This paper is constructed as follows: part 2 introduces a literature review. The methodologies are presented more detail in part 3. Part 4 is about the data and variables. The final result is mentioned in part 5. Finally, conclusion and discussion are included in part 6.

## 2. Literature review on bank failure prediction

Taking into account the fact that bankruptcy prediction is an important and widely studied topic, we will concentrate our literature review on the prediction of bank failure using financial ratios.[1] Moreover, we will focus on the studies that implement at least one of the five techniques compared in this paper. The history of bankruptcy prediction originated from predicting the failure of businesses. The important contribution of Altman (1968) motivated researchers to use multivariate analysis to predict the bankruptcy of firms. He provided an original Z-score formula (1968) and showed its advantage by analyzing five main financial and economic aspects of a firm: the liquidity, size dimensions; operating efficiency and profitability of the assets, financial leverage as well as considering the capability of management in dealing with competitive conditions (total asset turnover). Sinkey (1975) employed discriminant analysis to predict bank failures.

In comparison, Martin (1977) and Ohlson (1980) employed logistic regression to predict failures of firms and bank. Martin (1977) attempted to predict the US commercial bank failure within 2 years during 1970 and 1976 by using 25 financial ratios of asset risk, liquidity, capital adequacy and earning. He suggested that logistic regression has a higher percentage of correctly classified than linear discriminant. Since these initial studies, empirical studies have been conducted to compare the prediction accuracy of these two approaches (Boyacioglu et al., 2009).

Nevertheless empirical studies do not demonstrate a clear advantage for one of the two main traditional techniques: discriminant analysis versus Logit and Probit models. But Canbas et al. (2005) on a sample of 40 privately owned Turkish commercial banks showed, using 49 ratios, that discriminant analysis obtains slightly better results than Probit and Tobit models. On the same vein, a recent study Chiaramonte et al. (2015) revealed, on a big sample of 3242 banks across 12 European countries, that Z-score is a good predictive model to identify banks in distress (better than the Probit model) and also has the great advantage of simple calculation. According to the empirical study by Lo (1986), the equivalence between LDA and LR may not be rejected.

However, in some standpoints, statistical techniques are no longer preferred in view of their relatively low accuracy (Ravi Kumar and Ravi, 2007). The attention to and confidence in machine learning has increased enormously during the past 5–10 years. Numerous studies suggest that intelligent techniques perform more effectively than traditional statistical techniques. The main difference between intelligent and statistical techniques is that statistical techniques usually require researchers to define the structures of the model a priori, and then to estimate parameters of the model to fit the data with observations, while with intelligent techniques the particular structure of the model is learned directly from the data (Wang et al., 2015). Moreover, the statistical analysis depends on strict assumptions (normal distribution, no correlations between independent variables), that can result in poor prediction accuracy.

Among several machine-learning methods, the artificial neural network seems to be the most favored tool in prediction issues. Ky (1991) was among the first to implement a neural network on 118 banks (59 failed and 59 non-failed banks) in Texas during 1985–1987, and indicated that the neural network performed more effectively than other methods (Discriminant Analysis, factor-logistic, k-NNs and Decision tree). Several studies (Miguel et al., 1993; Bell, 1997; Olmeda and Fernandez, 1997; Swicegood and Clark, 2001; Aktas et al., 2003; Wu and Wang, 2000) compare ANN and the classical statistical techniques (Discriminant Analysis and Logistic Model) to predict bank failure. They generally conclude in the superiority of the neural network approach. In their survey Vellido et al. (1999), also suggest that ANN is better than the logit model for predicting commercial bank failures. More recently Lee and Choi (2013) compared the prediction accuracy of neural networks and linear discriminant analysis on a sample of Korean companies. Their results indicated that the bankruptcy prediction accuracy using neural networks is greater than that of LDA. Finally,

[1] Readers can refer to Ravi Kumar and Ravi (2007) and Fethi and Pasiouras (2009) for a broader and more in depth literature review.

a meta-analysis performed by Adya and Collopy (1998) reveals that neural networks outperformed alternative approaches in 19 out of the 22 analyzed studies.

Unlike Neural networks and Support Vector Machines, the k-nearest neighbor algorithm is not implemented widely in finance. This technique is implemented widely in biological and transportation fields. This method, however, can function appreciatively and obtain high prediction accuracy. Min and Lee (2005) proposed support Vector Machines for bankruptcy prediction. Boyacioglu et al. (2009) examined ANNs, SVMs and multivariate statistical methods to predict the failure of 65 Turkish financial banks. 20 financial ratios belonging to 6 main groups were chosen: Capital adequacy, Asset quality, Management, Earning, Liquidity and the sensitivity to the market risk. Overall, the result proved that SVMs achieved the highest accuracy. They concluded that this method outperforms neural network, discriminant analysis and logit methods. SVM was also proved to work better than neural networks through the research of Chiaramonte et al. (2015) for a sample of 3242 EU banks. Park and Han (2002) used k-nearest neighbor for company bankruptcy prediction but we do not find empirical studies specifically dedicated to the use of k-nearest neighbor to predict bank failure.

Finally some empirical studies compare the various predictions methods. Tam and Kiang (1992) compare discriminant analysis, Logit, k-nearest neighbor and artificial neural networks on bank failure prediction and find that the latter outperforms the other techniques. Martínez (1996) compares neural network back propagation methods with discriminant analysis, logit analysis and the k-nearest neighbor for a sample of Texan banks and concludes that the first set of methods outperforms. Zhao et al. (2009) compare Logit, ANN and k-NN. They find that ANN > Logit > k-NN when financial ratios rather than row data are used. These studies support neural networks as being the best methods of predicting bank failure. Serrano-Cinca and Gutierrez-Nieto (2011) compared 9 different methods to predict the bankruptcy of USA banks during the financial crisis, including Logistic Regression, Linear Discriminant Analysis, Support vectors Machines, k-nearest neighbor and Neural Networks. It can be concluded that no technique is clearly better than the others. Performance depends on the performance measure chosen; some techniques have more accuracy but less recall (1 minus Type II error rate).

Among numerous studies on predicting the bankruptcy of banks, history has shown that intelligent techniques (and specifically artificial neural networks) seem to work more effectively than statistical techniques. This study will execute both families of techniques in different methods and in a new attempt to make a comparison on two aspects: the accuracy and the importance of each ratio.

## 3. Methodology

### 3.1. Statistical techniques

Linear discriminant analysis and Logistic regression are popular methods for classifying objects based on their characteristics. These methods have been applied widely to predict the failure of firms and banks.

#### 3.1.1. Linear discriminant analysis (LDA) and logistic regression (LR)

Logistic regression (LR) is a regression model where the outcome is categorical (in our case the bank is active or inactive). More technically, the model assumes a linear relationship between the logarithm of the odds ratio (ratio of probabilities, see equation below) and one or more independent variables (bank characteristics, $x_j$).

$$g(x) = ln\frac{P(y=1)}{P(y=0)} = \sum_{j=1}^{m} \beta_j x_j + \beta_0$$

Linear Discriminant Analysis (LDA) derives a linear combination of ratios which best discriminate between failed and non-failed firms. Observations are assigned to one of the two groups in some 'optimal' way, for example, so as to minimize the probability or cost of misclassification. Logistic is often preferred to LDA as it is more flexible in assumptions and types of data that can be analyzed.

Canbas et al. (2005) propose an integrated model that combines LDA and LR in order to help predict bank failure. They demonstrate that this combination improves the prediction accuracy. Serrano-Cinca and Gutierrez-Nieto (2011) combine LDA with Partial Least Square analysis in order to predict the failure of US banks during the 2008 financial crisis.

Although the LDA and LR have become the most commonly used in bankruptcy prediction, their inherent drawbacks of statistical assumptions such as linearity, normality and independence among variables have constrained their practical applications (Lee and Choi, 2013). To solve the limitation of a linear approach, intelligent techniques (in this paper considered as machine learning approaches) achieve a forward movement by introducing nonlinear separation between groups.

Several methods have been implemented to classify companies or financial institutions and predict bankruptcy or failure. In this paper, three machine learning algorithms are applied: k-Nearest Neighbors, Artificial Neural Network and Support Vector Machines. Neural network is a well-known model and is considered to be one of the most powerful tools in prediction even when their conceptions are not easy to be translated. These models are referred to as black box processed because the mechanism that transforms the input into the output is obfuscated by a figurative box. On the contrary, k-nearest neighbors is regarded as a lazy learning technique (meaning that generalization beyond the training data is delayed until a query is made to the system). The idea is to classify unlabeled examples by assigning to them the class of the majority of its neighbors. Support Vector Machines are in between since, being not overly-complex, it is possible to enter into the black box.

#### 3.1.2. K-Nearest Neighbor (k-NN)

The k-Nearest Neighbor (k-NN) is an instance-based method, meaning that it assigns a new case to the majority class among the k-closest cases in the training set (Hand et al., 2001). In a brief description, nearest neighbor classifies by mapping the different characteristics of the dataset closely to different label groups, the given data with common features will then be placed in the same group. Each new case is classified based on the outcome of the majority of its neighbors.

Each bank is represented by a vector of its characteristics. Banks with similar characteristics tend to be placed closely together. The distance between each point to each group must be calculated in order to find out which group (active or inactive) the banks belong to. If the majority of neighbors of a given bank are classified as failed (active), this bank will be classified as failed (active).

There are three major decisions in the k-NN method: the set of stored cases, the distance metric used to compute the distance between cases, and the value of k (Weiss and Indurkhya, 1998).

There are several ways of calculating this distance. Traditionally, the k-NN algorithm deploys Euclidean distance. If $p$ and $q$ are two vectors of characteristics (two banks), each of them has $n$ features. The Euclidian distance between p and q is calculated as:

$$\mathrm{Dist}(\mathrm{p},\ \mathrm{q})=\sqrt{(p_1-q_1)^2+\ldots+(p_n-q_n)^2}$$

To classify the bank as active or inactive, we should begin by assigning the number of neighbors, k. We can select any value of $k$ to find the best grouping method. There are divergent hypotheses on selecting the ‘best’ $k$. Some researchers suppose that $k$ should be the square root of number of features. However, others assume that $k$ performs the best if it is between (2, 10). In this research, we experiment with various value of *‘k’* to find the optimal value.

#### 3.1.3. Artificial neural networks

Taking advantages of computer potential, Artificial Neural Networks (ANNs) are inspired by biological neural networks. ANN is applied widely on a variety of tasks such as: computer vision, speech recognition, etc. ANN is a machine learning technique, which can simulate any relationship. Although ANN is not the only technique that can do this it is often preferred for the ability to obtain a solution in a reasonable time.

The idea is to learn from examples using several algorithms just as a human being learns new things. The advantages of ANNs are their flexible nonlinear modeling capability, strong adaptability, as well as their learning and massive parallel computing abilities (Ticknor, 2013). However, they cannot explain the causal relationship among variables, which restricts its application to managerial problems (Lee and Choi, 2013).

A fully connected network includes series of neuron layers. While each unit in the same layer cannot interconnect, each layer can. The connection between one unit in a given layer and another in the following layer is represented by a number call a weight, which can be positive or negative. There are two ways to transfer information: feed-forwarding and back-propagation. Feed-forwarding will forward information from input layer to output layer and this processing can lead to a wrong result. However, back-propagation can fix these errors by sending back the information to optimize the outcome.

When designing a multilayer network, the decision on choosing the number of hidden layers is very important. Lee et al. (2005) and Zhang et al. (1999) show that one hidden layer is sufficient for most classification problems. Meanwhile, Vasu and Ravi (2011) suggested choosing 2 hidden layers in order to be sure that the network architecture will be sufficiently complex to cope with the complexity of bank failure prediction. In our study, we applied for both 1 hidden layer and 2 hidden layers to examine which one performs better.

To eliminate the possibility of being linear, we use an activation function which creates a non-linear decision boundary. Various types of activation function exist, for example: sigmoid, tanh, rectified linear unit, leaky rectified linear unit or max out. We decided to use a sigmoid function since its characteristics are suitable for our output. It is the most widely used function.

#### 3.1.4. Support Vector Machines (SVM)

The great advantage of Support Vector Machines (SVM) is that they combine the strengths of theory-driven conventional statistical methods and data-driven machine learning methods (Min and Lee, 2005). The method is based on the Vapnik’s (1995) structural risk minimization principle. SVM is highly appreciated for successful applications in many fields such as bioinformatics, text, image recognition, etc. SVMs are supervised learning models that analyze data used for classification and regression analysis. This method is developed from Statistical Learning Theory (Boser et al., 1992). The basic idea is that input vectors (a vector represents the financial characteristics of a given bank) are non-linearly mapped to a very high dimension feature space. A linear decision surface is constructed in this feature space thus SVMs transform complex problems (with complex decision surfaces) into simpler problems that can use linear discriminant functions.

Unlike numerous other methods which focus on whole training data, SVM pays attention to the most difficult to recognize data point based on the idea that if SVMs can figure out the toughest points, the others will be seen easily. The vectors most difficult to recognize are located close to the hyperplane separating active and failed banks, they are called support vectors. These points can be easily misclassified. The distance from the closest data points in each respective class to the hyperplane is called the margin. SVM will attempt to maximize these margins, so that the hyperplane is at the same distance from the 2 groups (failed and active banks). Intuitively, the more distant vectors are from the hyperplane, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it.

### 3.2. Implementation of statistical techniques

We use WEKA software to apply the listed statistical methods. WEKA is a collection of algorithms for data mining tasks. It can be run quickly for a big database. For more detailed information about the Weka package, the reader is referred to Witten and Frank (2005).

Firstly, 70% of the data from the 6-year period (5 year before failure plus the failure year) will be used on WEKA for training. The remaining 30% (5400 observations) will be used in order to test the prediction accuracy of each model. For each bank, we then checked whether it is correctly classified in the right group (active or inactive) at the right time (how many years before being inactive).

For the k-NN method, we tested several values of k to determine the appropriate value. As mentioned in part 3, there are several hypotheses on selecting k, for example: k should be the square root of total number of observation (in this paper are 43) or k should be between 1 and 10. We then tested both values of k from 1 to 10 and 43 For ANNs method, we set the hidden layer is equal to 1 or 2. The default training time is 500 on WEKA.

### 3.3. Data and variables

Initially, we collected over 5000 banks from Bankscope database. However, we set the condition that the number of inactive and active banks should be equivalent to test the performance of the machine learning approach. Finally, therefore, we select randomly a sample of 3000 banks including 1438 inactive and 1562 active banks. 6 year-periods include: year when banks go bankrupt and 5 years before being inactive was selected. Active banks were selected randomly with the criterion of being a US bank and still active until the first quarter of 2016.

After collecting and importing data in panel, we shuffled the order. Shuffling data is important to prevent bias learning process and predict more intelligently and in an integrated way. We then, divide the dataset into 2 subsets: a Training set (70% of data) and a Test set (30% of data). Theoretically, each method will learn 70% first training to create significant models. The remaining 30% will examine accuracy.

From bank financial statements we extract or construct 31 ratios. Zhao et al. (2009) demonstrate that the use of financial ratios, instead of raw accounting variables, significantly improves the performance of prediction techniques. Detailed accounting information is taken to forecast the status of banks and provide more adequate points of view. Ratios were selected by comparison with the lists of ratios used in previous empirical studies. Before 2007, these lists were presented in detail in the review by Ravi Kumar and Ravi (2007), after 2007 we take into account the lists presented in the papers referred in the literature review section.

Finally, the selected ratios cover: (i) loan quality, (ii) capital quality, (iii) operation efficiency, (iv) profitability and (v) liquidity. To shorten the name, we label each ratio by Z from Z1 to Z31. Each of these financial ratios is expected to have a strong influence on bank performances as well as possibly helping to predict the failure. For 31 ratios, we have dissimilar expectation signs on the bank's survival. Positive signs suggest that the higher the ratio the better the influence on the bank’s survival. Negative signs indicate the contrary (Table 1).

## 4. Empirical results

### 4.1. Descriptive statistics

Table 2 presents the means and standard deviations of the 31 selected financial ratios for active and failed banks groups for one year before becoming inactive. As in Canbas et al. (2005) the last two columns present the F-test for the equality of means among the two groups and the significance levels. We find that **25 of the 31 ratios** have a significant different mean (for failed and non-failed banks) at a level than 5%. Hence, the null hypothesis that the two group means are equal is rejected at the 5% significance level for these ratios. We find that one year before failure, the loan quality is significantly lower for inactive banks (especially Z2 and Z4). Equity can be seen as a general buffer against risk and we observe that these banks have less equity whatever the measure of equity and the comparison point (assets, loans, liability) (see Z6, Z7, Z8, Z12, Z15). Operational efficiency is also lower for inactive banks compared with active banks (Z19, Z20). However, contrary to expectations, liquidity is higher for inactive banks (Z29, Z30, Z31), possibly because the banks in the sample became inactive for solvability problems rather than for liquidity problems. Note that our results are quite similar to those of Canbas et al. (2005). On a sample of 40 Turkish banks during the period 1997–2003, they find that the ratios that are the most different between failed and active banks are interest expenses on assets and interest income on interest expenses, equity/TA, liquid assets total assets, standard capital ratio. López and Pastor Sanz (2015) employed data from the FDIC between 2002 and 2012, their results state that failed banks are more concentrated in real estate loans and have more provisions.

### 4.2. Comparison of accuracy

To analyze in detail the predictive performance of each method, we use several indicators (see Powers (2011) for more details on these measures). Precision is the fraction of those predicted positive by the model that are actually positive. Recall, also referred to as sensitivity, is the fraction of those that are actually positive which were predicted positive. F-measure is the harmonic mean of precision and sensitivity. The value of F-measure ranges from 0 to 1. A value of 1 indicates perfect prediction. MCC (Matthews Correlation Coefficient) is the measurement of the quality of binary classification. This indicator was first introduced by the

**Table 1**
Expected sign of ratios on bank's survival.

| Variables | Variables description | Expected sign |
|---|---|---|
| | Loan quality | |
| Z1 | Loan Loss reserve/Gross Loans | Negative |
| Z2 | Loan Loss provision/Net interest revenue | Negative |
| Z3 | Impaired Loans/Gross Loans | Negative |
| Z4 | Net charge off/Average Gross Loans | Negative |
| Z5 | Impaired Loans/Equity | Negative |
| | Capital quality | |
| Z6 | Tier 1 capital ratio | Positive |
| Z7 | Total capital ratio | Positive |
| Z8 | Equity/Total assets | Positive |
| Z9 | Equity/Net Loans | Positive |
| Z10 | Equity/Customer & short term funding | Positive |
| Z11 | Equity/Liabilities | Positive |
| Z12 | Capital funds/Total assets | Positive |
| Z13 | Capital funds/Net loans | Positive |
| Z14 | Capital funds/Deposit & Short term funding | Positive |
| Z15 | Capital funds/Liabilities | Positive |
| | Operations | |
| Z16 | Net interest margin | Positive |
| Z17 | Net interest revenue/Average Assets | Positive |
| Z18 | Other Operation income/Average Assets | Positive |
| Z19 | Non-Interest expense/Average Assets | Negative |
| Z20 | Pre-tax Operating Income/Average Assets | Positive |
| Z21 | Non-Operating Items & taxes/Average Assets | Negative |
| | Profitability | |
| Z22 | Return on Average Assets | Positive |
| Z23 | Return on Average Equity | Positive |
| Z24 | Inc. Net of Dist/Average Equity | Positive |
| Z25 | Cost to Income Ratio | Negative |
| Z26 | Recurring Earning Power | Positive |
| | Liquidity | |
| Z27 | Net Loans/Total Asset | Negative |
| Z28 | Net loans/Deposit & Short term funding | Negative |
| Z29 | Net Loans/Total Deposit and Borrowing | Negative |
| Z30 | Liquid Assets/Deposit & Short term Funding | Positive |
| Z31 | Liquid Assets/Total Deposit & Borrowing | Positive |

A positive sign indicates that when the ratio increases, the probability to fail decreases.

biochemist Brian Matthews in 1975. The value of MCC is between $[-1, 1]$. MMC = 1 indicates a prefect prediction; MMC = 0 indicates that the prediction is not better than random prediction; MMC = −1 indicates disagreement between prediction and observation. ROC Area (Receiver Operation Characteristic curve) is usually used for a binary classifier with the value between [0, 1]. This curve is created with y-axis is true positive rate, and x-axis is false positive rate. The closer to 1 the values of ROC are, the better the prediction. PRC Area (Precision/Recall plots): this indicator is used less frequently than others. However, Saito and Rehmsmeier (2015) suggested that PRC is more informative than a ROC plot when evaluating binary classifiers. PRC plots evaluate the fraction of true positives among positive predictions and hence can provide an accurate prediction of future classification performance.

#### 4.2.1. *Choice of parameters*

Firstly, we made a decision on choosing the number of hidden layer for ANNs and the number of neighbors for the k-NN method. Regarding ANN methods, we test whether the number of hidden layers is 1 or 2. The result in Table 3 shows that for 1 or 2 hidden layers, the difference is small. Overall, with 1 and 2 hidden layers, ANNs can recall 74.4% and 75% respectively and the precision ratio is 75.7% and 75.7% respectively. Consequently, we may conclude as in Lee et al. (2005) and Zhang et al. (1999) that using 1 hidden layer is sufficient. Finally, we will use the result from ANNs with 2 hidden layers to compare with other methods.

For the k-NN method, we also implemented various values of *k* in order to try to find the 'best k'. The first assumption is that k should equal the square root of the total number of observations, which is 43. The second assumption is that k should be between 1 and 10. The first one brings only 72.9% precision, while the others obtained around 74% (see Table 4). We therefore state that the number of k-nearest neighbors should be between 1 and 10. In this case, we choose the k with the greatest precision which is k = 8 and denote it 8_NN.

#### 4.2.2. *Comparison of the five bank failure prediction methods*

Following Vasu and Ravi (2011) we decompose the accuracy ratio into two dimensions: false positive (Type I error, the classifier

**Table 2**
Descriptive statics for the 31 financial ratios for active and failed banks – one year before being inactive.

| Ratio | Inactive banks | | Active banks | | F | Sig. |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| Z1 | 1.60 | 0.944 | 1.45 | 0.913 | 20.058 | 0.000 |
| Z2 | 13.45 | 24.780 | 4.09 | 8.065 | 200.135 | 0.000 |
| Z3 | 2.03 | 3.015 | 2.00 | 2.115 | 0.115 | 0.735 |
| Z4 | 0.53 | 1.077 | 0.19 | 0.654 | 110.766 | 0.000 |
| Z5 | 14.88 | 24.632 | 12.55 | 14.892 | 10.005 | 0.002 |
| Z6 | 13.17 | 4.952 | 15.14 | 6.743 | 81.813 | 0.000 |
| Z7 | 14.53 | 4.827 | 16.40 | 6.713 | 75.745 | 0.000 |
| Z8 | 10.09 | 3.213 | 11.10 | 3.398 | 70.150 | 0.000 |
| Z9 | 16.61 | 8.880 | 18.29 | 13.153 | 16.502 | 0.000 |
| Z10 | 12.26 | 4.964 | 13.41 | 7.119 | 25.630 | 0.000 |
| Z11 | 11.40 | 4.295 | 12.71 | 5.188 | 56.090 | 0.000 |
| Z12 | 10.37 | 3.221 | 11.48 | 3.313 | 85.988 | 0.000 |
| Z13 | 17.08 | 9.010 | 18.90 | 13.129 | 19.327 | 0.000 |
| Z14 | 12.63 | 5.148 | 13.87 | 7.186 | 29.358 | 0.000 |
| Z15 | 11.73 | 4.326 | 13.14 | 5.132 | 66.062 | 0.000 |
| Z16 | 3.95 | 1.088 | 3.76 | 1.436 | 16.040 | 0.000 |
| Z17 | 3.52 | 0.967 | 3.35 | 1.213 | 17.886 | 0.000 |
| Z18 | 1.00 | 1.190 | 1.12 | 1.474 | 6.184 | 0.013 |
| Z19 | 3.52 | 1.810 | 3.18 | 1.605 | 30.231 | 0.000 |
| Z20 | 1.00 | 1.663 | 1.30 | 0.998 | 38.089 | 0.000 |
| Z21 | −0.29 | 0.622 | −0.24 | 0.572 | 4.716 | 0.030 |
| Z22 | 0.67 | 1.238 | 0.98 | 0.780 | 67.328 | 0.000 |
| Z23 | 7.28 | 11.764 | 8.83 | 6.525 | 20.391 | 0.000 |
| Z24 | 1.88 | 11.626 | 5.27 | 5.961 | 103.154 | 0.000 |
| Z25 | 69.85 | 32.613 | 67.83 | 14.711 | 4.909 | 0.027 |
| Z26 | 1.46 | 1.416 | 1.48 | 1.260 | 0.081 | 0.777 |
| Z27 | 65.16 | 13.438 | 65.97 | 13.338 | 2.763 | 0.097 |
| Z28 | 78.41 | 20.623 | 78.06 | 17.342 | 0.263 | 0.608 |
| Z29 | 73.60 | 15.244 | 75.59 | 15.583 | 12.394 | 0.000 |
| Z30 | 9.72 | 10.426 | 7.71 | 8.077 | 35.414 | 0.000 |
| Z31 | 9.24 | 9.080 | 7.50 | 7.843 | 31.809 | 0.000 |

Table presents the means and standard deviations (SD) of the 31 ratios (Z1 to Z31) used to compare active and inactive banks. The F-test (F) is used for comparison of means. The p-value for the F-test (Sig.) is given in the last column.

**Table 3**
The comparison of ANNs 1 hidden layer and 2 hidden layers.

| Method | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area |
|---|---|---|---|---|---|---|
| ANNs_1 | 0.758 | 0.741 | 0.734 | 0.493 | 0.771 | 0.739 |
| ANNs_2 | 0.757 | 0.753 | 0.75 | 0.506 | 0.819 | 0.803 |

Table gives the accuracy measures for ANN with 1 hidden layer (ANNs_1) and two hidden layers (ANN_2). **Precision** is the fraction of those predicted positive that are actually positive. **Recall** is the fraction of those that are actually positive which were predicted positive. **F-measure** is the harmonic mean of precision and sensitivity. **MCC**: Matthews correlation coefficient. **ROC area**: Receiver Operation Characteristic curve. **PRC Area**: Precision/Recall plot.

misclassifies an actual active bank as a failed bank, FP in Table 5) and false negative (Type II error, the classifier misclassifies a failed bank as an active bank, 1-TP rate in Table 5). Note that for banks false negative is considered by banking regulators to be far more costly than false positive. As foreseeable from the literature review and from the previous results, Logistic and ANN obtain the best performance with the lowest values of type 1 and type 2 errors for both ratios. Table 5 highlighted that ANNs performed better than all the other methods whatever the performance measure (75.3% of TP rate and 25.9% of FP rate, which lead to 75.7% precision and 75.3% recall). As can be seen, k-NN and LR achieved similar results: around 74% precision. SVM and LDA obtain the lowest performance with only 71.6% and 72% precision respectively (TP). The distance among these results is not too significant, however we notice that the traditional logistic approach can predict more accurately than some of the machine learning approaches (as already observed by Zhao et al., 2009).

We then extract the result into years as in Table 6 to summarize the total errors of each method by year. The error here is defined when the active is classified as inactive and vice versa. ANNs make fewer errors than the others in the year that banks go inactive, and make the most errors 3 years before. Meanwhile, other methods can make errors evenly over the years. As noticed from the previous comment, SVMs make the most mistakes for most of every year. Surprisingly, the maximum number of incorrect classification occurred at the year or one year before failure.

We investigate to find out which method can recognize the failure when the other methods predict wrongly (Table 7). This

**Table 4**
Comparison of k-NNs with different number of neighbors.

| Method | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area |
|---|---|---|---|---|---|---|
| 1_NN | 0.731 | 0.731 | 0.73 | 0.459 | 0.728 | 0.668 |
| 2_NN | 0.74 | 0.712 | 0.698 | 0.442 | 0.774 | 0.719 |
| 3_NN | 0.741 | 0.74 | 0.739 | 0.477 | 0.791 | 0.743 |
| 4_NN | 0.736 | 0.722 | 0.714 | 0.451 | 0.8 | 0.759 |
| 5_NN | 0.736 | 0.736 | 0.734 | 0.469 | 0.804 | 0.768 |
| 6_NN | 0.738 | 0.727 | 0.721 | 0.459 | 0.808 | 0.775 |
| 7_NN | 0.741 | 0.739 | 0.737 | 0.476 | 0.81 | 0.78 |
| 8_NN | 0.741 | 0.731 | 0.725 | 0.467 | 0.811 | 0.783 |
| 9_NN | 0.741 | 0.739 | 0.737 | 0.476 | 0.81 | 0.783 |
| 10_NN | 0.739 | 0.73 | 0.724 | 0.464 | 0.812 | 0.787 |
| 43_NN | 0.729 | 0.724 | 0.72 | 0.448 | 0.806 | 0.798 |

Table gives the accuracy measures of k-NNs with a number of neighbors from 1 (1_NN) to 43 (43_NN). **Precision** is the fraction of those predicted positive that are actually positive. **Recall** is the fraction of those that are actually positive which were predicted positive. **F-measure** is the harmonic mean of precision and sensitivity. **MCC**: Matthews correlation coefficient. **ROC area**: Receiver Operation Characteristic curve. **PRC Area**: Precision/Recall plots.

**Table 5**
Performance of bank failure prediction methods.

| Method | Confusion matrix | | Precision | Recall | ROC Area | PRC Area |
|---|---|---|---|---|---|---|
| ANNs_2 | 2415 | 444 | 75.7% | 75.3% | 81.9% | 80.3% |
| | 892 | 1649 | | | | |
| 8_NN | 2455 | 404 | 74.1% | 73.1% | 81.1% | 78.3% |
| | 1048 | 1493 | | | | |
| LDA | 2185 | 674 | 72.0% | 72.0% | 77.6% | 75.8% |
| | 836 | 1705 | | | | |
| LR | 2235 | 624 | 73.9% | 73.9% | 79.6% | 77.3% |
| | 785 | 1756 | | | | |
| SVM | 2121 | 738 | 71.6% | 71.6% | 71.5% | 65.5% |
| | 794 | 1747 | | | | |

Table gives the accuracy measures for the five bank failure prediction techniques: ANN with two hidden layers (ANNs_2), k-NN with 8 neighbors (8_NN), Linear discriminant analysis (LDA), Logistic Regression (LR), Support Vector Machine, (SVM). **Precision** is the fraction of those predicted positive that are actually positive. **Recall** is the fraction of those that are actually positive which were predicted positive. **ROC area**: Receiver Operation Characteristic curve. **PRC Area**: Precision/Recall plots.

**Table 6**
Number of banks misclassified by year for the 5 predictions techniques.

| Year | ANNs_2 | KNN_8 | LDA | Logistic | SVM | Total |
|---|---|---|---|---|---|---|
| 0 | 199 | 252 | 285 | 262 | 295 | 1293 |
| 1 | 231 | 252 | 272 | 255 | 279 | 1290 |
| 2 | 220 | 258 | 248 | 212 | 225 | 1165 |
| 3 | 240 | 261 | 258 | 238 | 279 | 1279 |
| 4 | 234 | 213 | 218 | 221 | 219 | 1109 |
| 5 | 212 | 216 | 229 | 221 | 235 | 1118 |
| Total | 1336 | 1452 | 1510 | 1409 | 1532 | |

**Table 7**
Right when other methods are wrong.

| Methods | ANNs_2 | 8_NN | LDA | Logistic | SVM |
|---|---|---|---|---|---|
| Total | 469 | 92 | 3 | 14 | 12 |
| Year 0 | 73 | 24 | 1 | 3 | 2 |
| Year 1 | 85 | 24 | 0 | 2 | 1 |
| Year 2 | 94 | 10 | 0 | 3 | 1 |
| Year 3 | 82 | 15 | 0 | 0 | 1 |
| Year 4 | 67 | 9 | 1 | 4 | 5 |
| Year 5 | 68 | 10 | 1 | 2 | 2 |

Number of failed banks that one method can detect when all the others cannot.

criterion is important and not used widely to date. Our purpose is to observe how dominant the method is. Surprisingly, ANNs can recognize 469 instances while the other methods cannot. After ANNs, k-NNs can also predict 92 observations while the other methods cannot. However, LDA can recognize only 3 banks, which is very poor.

## 5. Conclusion

This paper proposes an empirical study on the prediction of bank failure through 2 approaches: machine learning and two traditional statistical approaches. We observed firstly that machine learning, ANNs and k-NN methods perform more effectively than traditional methods. However, the difference in prediction accuracy between ANNs and k-NN methods and the traditional logistic regression method is not very big. In addition, we observed that SVM does not perform better than traditional methods. Nevertheless, ANN and k-nearest neighbor demonstrate their remarkable ability when they can detect the failure correctly but the other methods cannot.

All 31 ratios are important to predict bank failures. Each group has at least one significant ratio that affects the survival of the banks. Among them, three groups play a more important role, namely operation efficiency, profitability and liquidity. Notably, the ratios Z3 (Impaired Loans/Gross Loans), Z6 (Tier 1 capital ratio), Z12 (Capital funds/Total assets), Z18 (Other Operation Income/Average Assets), Z17 (Net interest revenue/Average Assets), Z21 (Non Operation Items & taxes/Average Assets), Z22 (Return on Average Assets), Z25 (Cost to income ratio) Z27 (Net Loans/Total Asset), Z28 (Net loans/Deposit & Short Term funding) and Z29 (Net Loans/Total Deposit & Borrowing) are more relevant than the others.

Our results have important institutional and policy implications. In effect, banks and bank supervisors developed early warning systems to prevent individual bank failure and banking crisis. Sahajwala and Van den Bergh (2000) provide an overview of the different approaches that are being used or developed in this field. Our study can help banks and bank supervisors to design such early warning systems because it shows that the traditional logistic regression models perform quite well and they can be complemented by machine learning techniques (ANNs and k-NN) to detect the most difficult cases. Moreover, these methods are based on ratios analysis and our study provides some information on the financial ratios that could help to better predict bank failures.

The limitation of this study is that we emphasize accounting information and ignore bank market data. Moreover, we could not determine the role of each ratio in machine learning techniques.

## References

Adya, M., Collopy, F., 1998. How effective are neural networks at forecasting and prediction?: A review and evaluation. Int. J. Forecast. 17 (5–6), 488–495.
Aktas, R., Doganay, M., Yildiz, B., 2003. Predicting the financial failure: a comparison of statistical methods and neural networks. Ankara Univ. J. SBF 58, 1–24.
Altman, E.I., 1968. Financial ratios: discriminant analysis and the prediction of corporate bankruptcy. J. Finance 23, 589–609.
Beaver, W.H., 1966. Financial ratios as predictors of failures. Emp. Res. Account. 4, 71–111.
Bell, T.B., 1997. Neural nets or logit model?: A comparison of each model's ability to predict commercial bank failures. Int. J. Intell. Syst. Account. Finance Manag. 6, 249–264.
Boser, B.E., Guyon, I., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Pittsburgh, ACM. Proceedings of the Fifth Annual Workshop of Computational Learning Theory 5. pp. 144–152.
Boyacioglu, M.A., Kara, Y., Baykan, O.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 Syst. Appl. 36, 3355–3366.
Canbas, S., Cabuk, A., Kilic, S.B., 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: the Turkish case. Eur. J. Oper. Res. 166, 528–546.
Chiaramonte, L., Croci, E., Poli, F., 2015. Should we trust the Z-score? Evidence from the european banking industry. Global Finance J. 28, 111–131.
Demyanyk, Y., Hasan, I., 2010. Financial crises and bank failures: a review of prediction methods. Omega 38, 315–324.
Fethi, M.D., Pasiouras, F., 2009. Assessing Bank Performance with Operational Research and Artificial Intelligence Techniques: A Survey. University of Bath School of Management (Working Paper Series).
Hand, D., Mannila, H., Smyth, P., 2001. Principles of Data Mining. MIT Press, Cambridge, MA.
Ky, T., 1991. Neural network models and the prediction of bank bankruptcy. Omega 19 (5), 429–445.
López, F.J., Pastor Sanz, I.I., 2015. Bankruptcy visualization and prediction using neural networks: a study of U.S. commercial banks. Expert Syst. Appl. 42, 2857–2869.
Lantz, B., 2013. Machine Learning with R. Packt Publishing Ltd.
Lee, S., Choi, W.S., 2013. A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. Expert Syst. Appl. 40, 2941–2946.
Lee, K., Booth, D., Alam, P., 2005. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. Expert Syst. Appl. 29, 1–16.
Lo, A.W., 1986. Logit versus discriminant analysis. a specification test and application to corporate bankruptcies. J. Econometr. 31 (2), 151–178.
Martínez, I., 1996. Forecasting company failure: neural approach versus discriminant analysis: an application to Spanish insurance companies. In: Sierra Molina, G., Bonsón Ponte, E. (Eds.), Intelligent Systems in Accounting and Finance, pp. 169–185 Huelva.
Martin, D., 1977. Early warning of bank failure: a logit regression approach. J. Bank. Finance 1 (3), 249–276.
Min, J.H., Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Syst. Appl. 28, 603–614.
Mitchell, T.M., 1997. Machine Learning. McGraw-Hill.
Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res. 18, 109–131.
Olmeda, I., Fernandez, E., 1997. Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction. Computational Econ. 10, 317–335.
Park, C.-S., Han, I., 2002. A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. Expert Syst. Appl. 23 (3), 255–264.
Powers, D., 2011. Evaluation: from precision, recall and F-Measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. 2 (1), 37–63.
Ravi Kumar, P., Ravi, V., 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. Eur. J. Oper. Res. 180, 1–28.
Sahajwala, R., Van den Bergh, P., 2000. Supervisory Risk Assessment and Early Warning Systems. Basel Committee on Banking Supervision Working Papers. 53p.
Saito, T., Rehmsmeier, M., 2015. The precision – recall plots is more informative than the ROC Plot when evaluating binary classifiers on Imbalanced datasets. Plos One J. 11/024, 1–22. http://dx.doi.org/10.1371/journal.pone.0118432.
Serrano-Cinca, C., Gutierrez-Nieto, B., 2011. Partial Least Square Discriminant Analysis (PLS-DA) for Bankruptcy Prediction. CEB Working Paper. (n° 11/024).
Sinkey, J.F., 1975. A multivariate analysis of the characteristics of problem banks. J. Finance 30, 21–36.

Swicegood, P., Clark, J.A., 2001. Off-site monitoring for predicting bank under performance: a comparison of neural networks, discriminant analysis and professional human judgement. Int. J. Intell. Syst. Account. Finance Manag. 10, 169–186.

Tam, K.Y., Kiang, M., 1992. Predicting bank failures: a neural network approach. Decis. Sci. 23, 926–947.

Ticknor, J., 2013. A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. 40 (14), 5501–5506.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vasu, M., Ravi, V., 2011. Bankruptcy prediction in banks by principal component analysis threshold accepting trained wavelet neural network hybrid. In: Proceedings of the 7th International Conference on Data Mining. Las Vegas, July, 18–21.

Vellido, A., Lisboa, P., Vaughan, J., 1999. Neural networks in business: a survey of applications (1992–1998). Expert Syst. Appl. 17, 51–70.

Wang, G., Ma, J., Yang, S., 2015. Improved boosting based on feature selection for corporate bankruptcy prediction. Expert Syst. Appl. 41, 2353–2361.

Weiss, S.M., Indurkhya, N., 1998. Predictive Data Mining: A Pratical Guide. Morgan Kaufmann, San Francisco, CA.

Witten, I.H., Frank, E., 2005. Data Mining: Pratical Machine Learning Tools and Techniques, 2nd ed. Morgan Kaufamn, San Fransisco, CA.

Wu, C., Wang, X.M., 2000. A neural network approach for analyzing small business lending decisions. Rev. Quant. Finance Account. 15, 259–276.

Zhang, H., Hu, M.Y., Patuwo, B.E., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. Eur. J. Operational Res. 116, 16–32.

Zhao, H., Sinha, A.P., Ge, W., 2009. Effects of feature construction on classification performance: an empirical study in bank failure prediction. Expert Syst. Appl. 36 (2), 2633–2644.