Neural networks versus logistic regression: the best accuracy in predicting credit rationing decision

Wafa Sayeh\textsuperscript{1*}, Annie Bellier\textsuperscript{2}

\textsuperscript{1}THEMA, University of Cergy-Pontoise, France  
\textsuperscript{2}THEMA, University of Cergy-Pontoise, France  
* Tel. 00331786567; e-mail: wafa.sayeh@u-cregy.fr

Abstract

In this paper, we establish a general framework for explaining the functioning of Artificial Neural Networks (ANNs) in binomial classification. We compare this approach with one of the conventional techniques, namely logistic regression (LR) to predict credit rationing (CR) decisions. We use leave-one-out cross validation to ensure the robustness of the two classifiers. Based on data from a survey of 246 French Small and Medium Enterprises (SMEs), we highlight the imbalanced learning issue as the fact that the class distributions of rationed and non-rationed firms are skewed. We apply a comprehensive approach using a combination of oversampling and undersampling methods called ROS/RUS. We balance the data in the learning phase and compare the results with those obtained on the original data. We estimate the classifiers’ performance based on the confusion matrix, the accuracy rate with paired t-tests, the ROC curve and the Area under the ROC curve (AROC). Our study can contribute to the choice of the best approach in CR prediction for small businesses.

Keywords: Credit rationing, prediction, artificial neural network, logistic regression, classification, Small Business Enterprises

JEL Classification: B23, C45, D82, G30
Introduction

Financial decision making and the inner functioning of the banking system remain opaque for most customers. Many companies are unaware of, for example, the existence of internal rating systems for identifying bad risks and thus are unable to anticipate the precise selection criteria which are behind the decision to issue or to refuse the loan. Our goal presupposes the idea that if companies can predict what their bank’s finance decision will be, they may take the necessary measures to put in place a strategy which maximizes the likelihood of obtaining bank funding while negotiating more favorable credit terms. Still on the basis of the company's financial position, questions that need to be answered are: can we perfectly predict the future CR decision? If so, what is the best algorithm that discriminates between and predicts classifications?

CR decisions are subject to considerable uncertainty for borrowers. Most of the studies which analyze the CR decision have focused on its measures and determinants by using either logistic regression (LR) (Cole, 1998; Cavalluzzo and Cavalluzzo, 1998; Cavalluzzo et al., 2002; Chakravarty and Scott, 1999; Blanchflower et al., 2003; Berkowitz and White, 2004; De Bodt, Lobez and Statnik, 2005; Becchetti et al., 2010; Han, Fraser and Storey, 2009) or probit regression (Angelini et al., 1998; Levenson and Willard, 2000; Lehmann and Neuberger, 2001; Blumberg and Letterie, 2007; Freel et al., 2012), without taking into account the predictive and the classification power of these models. However, there is a sizeable empirical literature that highlights the scale of the challenge of finding a dominant pattern in bankruptcy prediction (see; Ravi Kumar and Ravi, 2007 for a detailed review of the literature).

From traditional statistical methods (linear discriminant analysis, LR, factor analysis) to more sophisticated algorithms (the architecture of different artificial neural networks -ANNs-, decision trees, support vector machines, k-Nearest Neighbors algorithms, etc.), the comparison between such models has been central to many studies and is a particular feature in the selection of the most efficient model. We will confine our analysis to two algorithms, namely LR and ANN. Our choice is not based on random selection but on the fact that among traditional statistical methods, LR is used extensively in modeling CR. On the other hand, ANNs have become increasingly popular since their introduction in finance in the early 1990s and have been proven successful in predicting bank and corporate failure. The comparison between LR and ANN will be carried out using more than one measure covering the confusion matrix and its ancillary measures, the ROC curve and the Area under the ROC curve (AROC). If there is no evidence of the dominance of ANN on LR, it would be beneficial to use the less complicated model, LR, that produces easily understandable parameters. Although this stage seems to be straightforward in LR, it is much less
easy for ANN because of its black-box nature. Illuminating the black-box is identified as a matter of focus and various studies propose different techniques to provide user feedback (Garson, 1991; Goh, 1995; Olden and Jackson, 2002; Olden, Joy and Death, 2004; Paliwal and Kumar, 2011).

This study aims to provide a detailed comparison between LR and ANN prediction models on CR and to identify the theoretical key criteria needed to make credit decisions. First and foremost, we need to highlight some points that have raised a number of concerns related to the binomial classification approach. The first is related to the sample itself: when the classes are not approximately equally represented, which is our view, we have to deal with the problem of imbalanced learning. As described in Menardi and Torelli (2014), several techniques have been proposed in the literature to correct the error rate and therefore to deal with the skewed class distribution. Each approach carries with it benefits and shortcomings, and with the availability of many techniques which deal with the problem of imbalanced data, it is not easy to choose among them. In this study, we used a combination of the oversampling and undersampling method called the ROS/RUS approach, which consists of random oversampling with replacement of the minority class and random undersampling (without replacement) of the majority class. Leave-one-out cross validation (LOOCV) was used to test the robustness of our results during the learning phase. This technique avoids the overfitting problem and provides an accuracy that objectively reflects the model’s performance.

The remainder of the paper is organized as follows. First, we present some studies from a wide review of the literature that compares the accuracy of these techniques in predicting bankruptcy. Second, a brief discussion of the selected classifiers (ANN and LR) is given. Third, we provide details on the data and the methodology applied and briefly discusses some of the concepts used. Fourth, we present our results. Sixth, we summarize our findings and conclude.

**Literature review**

There is a sizeable empirical literature which uses and compares different techniques to identify the most suitable technique, based on a kind of evaluation of the criteria in financial decision making (Odom and Sharda, 1990; Tam and Kiang, 1992; Fletcher and Goss, 1993; Altman et al., 1994; Fernandez and Olmeda, 1995; Bell, 1997; Maher and Sen, 1997; Youn and Gu, 2010; Abdou and Pointon, 2011). The better forecasting performance of ANNs is supported by some studies such as Tam and Kiang (1992), Salchenberger et al. (1992), Fernandez and Olmeda (1995), Barniv et al., (1997), Maher and Sen (1997), Zhang et al.,(1999) and Youn and Gu (2010). Tam and Kiang (1992) tested the accuracy of various algorithms in predicting bank failures. These include ANN, LR, discriminant analysis, k Nearest Neighbor and decision trees. The starting point of a classification model is the choice of inputs. Tam and Kiang rely on the variables used in CAMEL rating as an early warning system to detect bank failures and rank algorithms selected in order of performance. ANN with 10 hidden units and 3 layers is the best classifier in the one- and two-year period, with the exception of discriminant analysis which gives the lowest type II error rate, that is, the percentage for misclassifying a non-failed bank as a failed one in the one-year period. Maher and Sen (1997) tested the accuracy of these two models on bond ratings. The observation is classified in the category with the
highest predicted probability, given that there are six bond rating categories. Using the holdout validation, LR is able to correctly classify 61.66% of the holdout sample compared with 70% for ANN. Along the lines of the previous studies, Youn and Gu (2010) compared these two algorithms by applying them to predict business failures for Korean lodging firms and achieved a similar result. The overall accuracy for the LR model is equal to 77.27%, with type I and type II errors equal to 9.09% and 36.36% respectively. ANN gives better overall accuracy and correctly classified 54 of the 66 of the holdout sample. Three of the 33 failed firms are classified as non-failed which gives a percentage of 9.09, and nine of the 33 non-failed firms are classified as failed. Both models give the same type I error which is the most costly. But ANN gives higher predictive ability in classifying non-failed firms. Zhang et al. (1999) also affirmed that ANN is a superior technique to deal with the bankruptcy classification problem, compared with LR. The average overall classification rate of ANN is equal to 81.82% compared to 78.18% for logistic regression.

But such results have not been confirmed by additional research (Desai et al., 1996; Bell, 1997; Laitinen and Kankaanpaa, 1999; Salehi, 2011). Desai et al. (1996) compare the performance of Multilayer perceptron, LR and linear discriminant analysis. Using a paired t-test for the comparison of means, they conclude that for good and bad loans, the differences in means between ANN and LR are not statistically significant. Bell (1997) also used LR and ANN to predict commercial bank failure and concludes that both models are equally efficient and neither can be considered as a dominant predictive model. Salehi (2011) compared the accuracy of these techniques to identify bad and good debtors after the loan was obtained, using 11 input variables relating to the company’s financial position and the industry it belonged to. ANN and LR have similar prediction efficiency. Laitinen and Kankaanpaa (1999) compared the accuracy of six classifiers including ANN and LR to predict one year, two years and three years prior to financial failure. In predicting one year prior to failure, LR gives the lowest error rate, but ANN outperforms LR in the second and third year prior to default. However, discrepancy of means appears to be not significant. On the other hand, Desai et al. (1997) lean towards traditional techniques in classifying loan quality (bad, poor, and good). By comparing LDA, LR, ANN and genetic algorithms, they found that LR dominates the other algorithms in predicting bad and good loans while ANN is better for poor loans. For the total sample, the accuracy rate of LR is greater than that of the other methods.

One conclusion that may be drawn from the preceding research is that despite its greater flexibility compared to LR, ANN produces contrasted results. This shows, consequently, that ANN does not completely dominate LR, depending on the purpose of the study, the choice of predictor variables and the complex architecture of ANNs (Smith, 1993).

**Design of logistic regression and Artificial Neural Networks**

**Logistic regression**

Binary logistic regression is very widely known and used for modeling specific binary decisions (Bell, 1997). It builds the relation between a binary or ordinal variable (outcome) and a set of discrete and/or continuous attributes.
If \( X = (x_0, x_1, x_2, \ldots, x_n) \) denotes \( n \) dependent variables, \( Y \) denotes the absence \((Y=0)\) or the presence \((Y=1)\) of CR with a probability, \( \Pr \). The main objective is to determine the probability of CR for a firm under its features introduced in the model.

The relation between the predictor variables and the probability, \( \Pr \) is described as:

\[
\text{Logit}\{\Pr(Y = 1|X)\} = \ln \frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)} = wX
\]  
(1)

\[
\frac{\Pr(Y = 1|X)}{1 - \Pr(Y = 1|X)} = e^{wX}
\]  
(2)

\[
\Pr(Y = 1|X) = \frac{e^{wX}}{1 + e^{wX}} = \frac{\exp\left(\sum_{i=0}^{n} w_i x_i\right)}{1 + \exp\left(\sum_{i=0}^{n} w_i x_i\right)}
\]  
(3)

The probability estimates will always be between 0 and 1 due to the logistic transformation. The coefficient of the predictor variables is calculated by using maximum likelihood estimates. Such a technique consists of choosing the values of \( \omega \) that maximize the probability function given the observed data \( Y \) and \( X \). Unlike linear regression, there is no closed-form solution for the estimates, and it seems necessary to use an iterative algorithm which draws near to the maximum likelihood solution (Landwehr, Hall and Franck, 2005).

The decision boundary between the two classes is linear for LR and is represented by a hyper plane in the \( X \) vector space. \( X \) is fitted to class 1 if:

\[
\frac{P(Y = 1|X)}{1 - P(Y = 1|X)} > 1
\]

\[
e^{wX} > 1
\]

\[
wX > 0
\]

\[
\sum_{i=0}^{n} w_i x_i > 0
\]  
(4)

The decision boundary between the two classes of rationed and non-rationed firms is represented by the hyper-plane defined by the equation \( wx=0 \) where the points with points \( x \) above belong to the class of rationed firms.
Artificial Neural Network

ANN is an intelligent non-linear technique which simulates the functions of the human brain. There are many possible ANN architectures; the most well-known and often-used for classification problems is the multilayer perceptron network (MLP) and the most common neural network is the single hidden layer back-propagation network. Its structure would be defined by an input layer, an output layer and one hidden layer. Each neuron is only connected to all subsequent neurons. The neurons are organized in layers in such a way that the outputs in one layer serve as the inputs in the next layer. The baseline of an ANN is a model neuron which is mathematical adjustment built on a package of input data and a single output. For each neuron, every input has an associated weight that defines its relative importance. The neuron will sum the input value multiplied by its weight, then applies to this value a nonlinear transformation function to determine the output according to a threshold value. The default value of the threshold is equal to 0.5. Each model allows the decision-maker to calculate the probability that the target variable (Y=1) and the probability that the variable (Y=0). The decision-maker can then adjust the value of the threshold to be tolerated.

The MLP network is based on a non-linear activation function for the hidden layers which transforms the weighted sum of inputs to an output value. It is an essential part of the network because it introduces non-linearity into the network. Without hidden layer(s), the network with logistic activation functions will act exactly like LR.

The learning algorithm of ANN can be either supervised or unsupervised. In supervised learning algorithms, we once had the target output. The neural "learns" the relationship between the inputs and the output and determines the causal link.

Assuming that our network is defined as one single hidden layer back-propagation network, X is the vector of input variables with the bias term, \( w_{hi} \) corresponds to the weight matrix connecting the ith input node to the hth hidden node, \( w_{oh} \) corresponds to the weight matrix between the hth hidden unit to the other output unit, Y is the vector of output variables, D is the number of input variables and M is the number of hidden units.

The input of the hidden layer is given by:

\[
Z_h = \sum_{i=1}^{D} w_{hi}X_i
\]

(5)

We apply the activation (or transfer) function to \( Z \) to obtain the output vector of the hidden layer:

\[
Z_h = f^{(1)}(Z) = f^{(1)}(\sum_{i=1}^{D} w_{hi}X_i)
\]

(6)

The output Y is equal to:
\[ Y = f^{(2)} \left( \sum_{j=1}^{M} w_{o,j} Z_h \right) = f^{(2)} \left( \sum_{j=1}^{M} w_{o,j} f^{(1)} \left( \sum_{i=1}^{P} w_{h,i} x_i \right) \right) \] (7)

Where \( f^{(2)} \) and \( f^{(1)} \) are the transfer functions for the output node and hidden node, respectively. A wide range of transfer functions is proposed in the literature. The logistic sigmoid function is probably the best known:

\[ f^{(2)}(x) = f^{(1)}(x) = \frac{1}{1 + \exp(-x)} \] (8)

Learning in Multilayer Perceptron (MLP) is made using the back propagation algorithm which is a supervised learning technique that uses a gradient descent method to adjust the connection weights between neurons in order to reduce the value of the error function. The algorithm compares the predicted value with the actual value in order to compute the value of some predefined error function. The information is then passed to the network and the connection weights are adjusted. The process is repeated according to the number of the training cycles until the error rate is small enough. The default performance function for MLP is mean squared errors (MSE), defined as:

\[ MSE = \frac{1}{N} \sum_{j=1}^{N} (e_j)^2 = \frac{1}{N} \sum_{j=1}^{N} (t_j - Y_j)^2 \] (9)

Where \( t_j \) and \( Y_j \) are the target value and the network output for the jet training case, respectively, and \( N \) is the number of training cases.

ANN versus LR hypothesis

ANN and LR have been largely used in a number of fields other than finance. LR is the standard comparison for machine learning techniques. From a technical point of view, ANN and LR have several key opposite and common characteristics. An ANN model without hidden layer(s) and with a sigmoidal activation function is the same as an LR model (Bishop, 1995). Both models enable decision-makers to compute the conditional probability of class membership. ANNs have significant competitive advantages over LR which are, firstly, their ability to capture complex nonlinear relationships between the dependent and independent variables through hidden nodes, whereas in LR the relationship between the predictor variables and the target variables is already specified. In a classification approach, this means that LR can only be adequate in so far as the classes are linearly separable because the decision boundary of the LR model is linear and is represented by a hyper-plane. In the case of a non-linear relationship, the use of an LR model becomes inappropriate. The second advantage of ANNs is that they can identify the correlation between independent variables through hidden nodes. The larger the number of predictor variables, the more likely an interaction exists which makes an LR model difficult to implement because all possible interactions between
variables should be tested. If this is the case, it can be expected that reliance on an ANN model is a good choice and will give better performance than an LR model.

The difference between LR and ANN also lies in the fact that the former is considered a parametric technique while the second is considered a non-parametric technique, making it difficult to interpret. The first disadvantage is that illuminating the black-box is a delicate question and several approaches have been proposed to determine the most important inputs in the ANN. Considerable efforts to illuminate the black-box and to find the effect of each independent variable on the target one have emerged since the end of the 1980s (see, Oldon and Jackson, 1992 for more detailed descriptions). However, LR has the advantage of being able to identify and test the parameter significance of predictor variables. The second disadvantage of ANN is that it is considered an empirical method (Tu, 1996) and this makes optimal parameter selection difficult. With various ANN algorithms, it is sometimes impractical to define the optimal algorithm and one can use an ANN that is not the best-performing network.

**Design of the study**

**Sample**

We analyzed the selected data from the replies by French firms to a survey about CR. They were asked two fundamental questions. The first question is (i) did you ever apply for a loan in 2008? If the answer was "yes", the respondent had to choose between three outcomes: (ii) you received the entire funding requested (iii) you received only a part of the financing asked for (iv) your application was refused by the bank. The resulting CR thus obtained corresponds to the size CR, if response (iii) and to the pure CR, if response (iv). Size CR occurs when some or all of the applicants receive a smaller loan than they desire, even if they are willing to pay the quoted price. Pure CR occurs when some applicants are denied a loan even though they are willing to pay the quoted price and cannot be distinguished by banks from other applicants who do receive loans. As there were only a small number of responses about CR, we build a discrete variable RAT, which is equal to zero if the firm obtained the loan requested and is equal to one if the loan application was denied in whole or in part by the bank.

The survey was conducted in 2010 and was addressed to firms from a diverse range of business sectors: industry, agricultural products, textile, food and beverages, etc. After eliminating firms with missing data, the final sample size was 246 enterprises with complete information. From the 246 companies studied, 59 firms were rationed by banks in 2008, in whole or in part, namely about 24 percent of the sample. The sample only included firms with fewer than 250 employees and annual balance sheets below 43 million EUR. In the survey, we also formulated questions related to the lending relationship with the main bank and manager characteristics. Information to supplement those data, relating to the default risk of a company, was obtained from Insead Oee Data Services (IODS). Financial variables were taken from financial statements: balance sheet, income sheet and attached tables for the year 2007.
Design of ANN

For the classification task, ANN and binary LR were trained using data mining techniques implemented in Rapidminer (for ANN) and Knime (for LR), two of the most popular data mining tools. The empirical framework starts with a feed-forward neural network trained by a back propagation algorithm (multilayer perceptron). The optimization of the weights is done by backward propagation of the error across a number of the training cycles. Undertraining or overtraining the network can lead to poor performance. In the case of a small number of training cycles, the network does not sufficiently adjust the synaptic weights from their initial randomized states to minimize the overall error rate. Furthermore, a large number of training cycles will over train the network which cannot think outside the box. In our model, after having tested several versions of MLP, the number of training cycles is set at 45. In the architecture of our MLP, we selected one hidden layer to control nonlinear input/output relationships. On the number of hidden layers to be introduced in the ANN, empirical literature argues that one hidden layer is sufficient to address a number of issues, and therefore, increasing the number of hidden layers by more than one or two might yield worse results (Hornick, Stinchcombe and White, 1989; Bishop, 1995). The sigmoid function is used as the activation function and the ANN is composed of three layers: input layer, one hidden layer and output layer. The number of nodes in the input layer refers to the number of attributes and the number of nodes in the output layer represents the number of classes of the target variable. The number of hidden nodes depends on the number of input and output nodes, the number of training cases, the training algorithm and the architecture of the ANN. Classification results can differ according to the number of hidden units used. A large number of hidden units can produce overfitting due to the inability of ANN to correctly classify cases that are not in the training set. A small number of hidden units also gives the same poor results and produces under-fitting (Geman, Bienenstock and Doursat, 1992). There is no general rule about the optimal number of hidden units and we will therefore have to train some networks and to select those which minimized the generalization error obtained from cross-validation. In Rapidminer, the basic setting is equal to the (number of attributes+ number of classes)/2+ 1. A rule of thumb provided by Berry and Linoff (1997) is that the number of hidden units shall not exceed twice the number of input units. In our study, we start with no hidden units and increase the number until the back propagation error is minimized. In the end, we retain 11 hidden nodes.

Before building such a system, we normalize our data to make the values of our attributes scaled to -1 and +1 before training as we used here a sigmoid as an activation function and training algorithms give better results when input variables are scaled (Sola and Sevilla, 1997). Because MLPs do not automatically handle qualitative variables, we translate k categories to k-1 binary variables.

Imbalanced learning problem

In the literature, there is no clear definition about the cutoff value below which the data are considered to be imbalanced. Ding (2011) argues that if the minority class represents only 5% of the entire sample, we are confronted with an imbalanced learning problem. Even having, a priori, the same distribution of rationed and non-rationed firms...
among classes does not correspond to the real world in CR. But a learning method can in our case correctly classify 76% of observations by simply sorting all cases as being non-rationed and therefore it is completely useless. The accuracy appears here as an inappropriate measure of performance. Menardi and Torelli (2014) give a complete review of the literature about the imbalanced learning problem and argue that the tricky bit here is that during the learning phase, standard classification algorithms are based on balanced data and assume equal misclassification costs. Lawrence et al. (1998) explain that the MLPs perform poorly when the training sample is imbalanced because MLP, like many other classification techniques, assumes symmetrical distribution between the two classes. If this assumption is not retained, the classifier tends to ignore or treat the minority class as noise. As described in Menardi and Torelli (2014), several techniques have been proposed in the literature to correct the error rate and therefore to deal with the skewed class distribution (resampling techniques, cost sensitive learning, adjusting thresholds, SMOTE algorithms, etc.). Each approach carries with it benefits and shortcomings, and with the availability of many techniques which deal with the problem of imbalanced data, it is not easy to choose among them. In this study, we used a combination of the oversampling and undersampling method called the ROS/RUS approach, which consists of random oversampling with replacement of the minority class and random undersampling (without replacement) of the majority class such as the learning sample, which consists of 123 rationed and 123 non-rationed firms, and we come back to our original sample in the testing phase. We finally chose this method after having tested several approaches (cost-modifying approach, oversampling and undersampling, threshold-modifying approach) that did not yield good results. It can be expected that the resampling technique would significantly improve the classification performance of the minority class (Weiss and Provost, 2001; Laurikkala, 2001; Estabrooks and Japkowicz, 2004).

Cross-validation

A cross-validation technique is useful for estimating the generalization error. It is based on resampling and allows us to examine the classification performance of a classifier or to decide, from a range of classifiers, which one is best. The best classifier is the one that gives the least generalization error. The method involves randomly dividing the sample into one training subsample and one testing subsample and avoiding over fitting. During the training phase, the model is fitted and connection weights are computed. The obtained model is then evaluated using the testing subsample. With small size data, we cannot afford to split the data into training and testing sets and we preferred to use all the data during the learning process. The leave-one-out cross validation technique (LOOCV) was developed as a good alternative to avoid the over fitting problem which can be found in the naive validation and the holdout method and to ensure the stability of a trained model. The naive approach consists of using the same sample for training and testing the model. The problem with this technique is that it gives an optimistic prediction error. The holdout method can also be problematic, especially in the case where the size sample is small. Thus, cross validation allows for an honest error rate rather than optimistic error rate estimates. LOOCV takes place in several stages:

- The partitioning consists of the $k=1$ observation for the validation sample and the remaining observations $(n-k)$ as the training data.
- Run the model with the training data and calculate the true error rate of the $k$ observation ($E_k$)
- Repeat these two stages such that each observation is used once as the validation data.

- The true error rate is estimated as the average error rate on validation data: \( E = \frac{1}{N} \sum E_i \)

Leave-one-out cross validation (LOOCV) technique is a particular case of K-fold cross-validation.

Selection of variables

For each business we selected 12 variables from a set of financial ratios and qualitative variables largely discussed in the literature as acting on the credit granting decision. If those variables capture the borrower's credit quality, then it is on the basis of this information that the CR decision will be made. However, we will be faced here with a shortage of information, especially for the qualitative aspects. The explanatory variables comprise but are not limited to the following: we used one proxy for the banking relationship, which is the variable number of banks with which the firm holds an account (MULTI) and is a dummy variable equal to zero if the firm has a relationship with one bank and one if the firm has a relationship with more than one bank. The size of the firm is measured by its total assets (TA), in log form. As in Agostino et al. (2008), we expect that the greater the firm’s total assets, the lower the probability of its being rationed. We also include the variable of the firm's age (AGEFIRM).

As an indicator of the firm's liquidity, we take the ratio of liquid assets to total assets (LIQUI). It is expected that the more liquid the firm, the lower the probability of being rationed because liquid assets help to cover a part of the loss in the case of default. But a high level of liquidity assets can also be perceived by the bank as a signal of transferring the firm’s risk to the lender. The debt level is measured by the ratio of bank debt to total assets (BDEBT). Its effect on CR is, however, ambiguous: it may increase the bank's reluctance to grant a loan, as it increases the probability of default; it may also demonstrate the company's capacity to obtain bank loans. A firm’s financial characteristics also include a measure of profitability proxied by the return on assets (ROA), computed as the net income from total assets. As in Agostino et al. (2008), we also used the ratio of tangible assets to total assets as a measure of the capacity of the firm to collateralize (TANG). A borrower's repayment ability is measured by the variable of financial debt and internal financing capacity (CARE). We also used the ratio of equity to total assets (EQUITY) and we assume that the higher the ratio, the stronger the firm will be financially.

As another alternative source of financing, we take the use of trade credit (TRADE) which is proxied by the number of credit days available to the firm. The effect of trade credit on CR is controversial. Some studies argue that trade credit acts as a substitute for bank credit, and the volume of trade credit will be positively correlated with CR. Because trade credit is more expensive than bank loans, the use of this alternative source of financing can reflect the inability of the company to obtain bank loans (spill-over assumption). But an extensive use of trade credit does not mean that the firm is credit rationed and may be viewed as an attempt to reduce transaction costs. The expected effect of trade credit on CR can also be positive as described in Cook (1999), showing that firms using trade credit have a lower probability of CR. We also consider indicators related to human capital such as borrower's gender.
(GENDER) and borrower's age (AGE). We also take into account the industry type as a control variable for a firm's sector of activity by introducing industry fixed effects.

Results and discussion

As a first step, a series of tests were run. Table I presents a summary of statistics on these variables. To better understand our predictor variables, univariate analysis is carried out in order to compare means between rationed and non-rationed firms using the parametric student's t-test. The results are presented in Table II. Six of the 12 attributes show a significant difference in means between rationed and non-rationed firms. We can notice a significant disparity between rationed and non-rationed groups according to the variable trade credit. The difference between the averages of the two groups is equal to -14.81 and is significant. Rationed firms seem to pay their suppliers later, compared to the average for trade credit. The proxy of banking relationships (multi-bancarity) shows that 54% of non-rationed firms deal with more than one bank and 40.6% of rationed firms have multiple banking relationships. The average difference is significant. At first sight, it may be said that multi-bancarity reduces CR decisions. Liquid assets are also higher for non-rationed firms than for rationed ones. The owner's characteristics appear here to be non-significant and do not act on the decision to grant credit or not. The firm's size expressed by the natural logarithm of the total assets is significant and demonstrates that on average non-rationed groups contain large-scale firms. The variables EQUITY, ROA and TA have the expected sign and show good consistency with the empirical literature on credit scoring. To better understand the CR decision, we make the presumption that this will depend specifically on the signal transmitted by a company about its quality. Thus, the bank classification system is based on a set of relevant variables that will determine "lemon" projects.

Leave-one-out cross validation results on the predictive performance for both ANN and LR are given in Table III. Based on a comparison between the two classifiers, this table illustrates that the overall classification rate of LR is 74.80% while ANN allowed proper classification of 71.14%. For non-rationed firms, LR outperforms ANN with a correct classification percentage of 89.30% compared with 87.17% for ANN. For rationed firms, LR also gives better prediction and correctly classifies 28.81% of cases, compared with 20.34% for ANN. To test the level of significance of our results, we carry out a paired t-test (given in Table IV) to determine if there is a statistically significant difference between means. Overall, LR dominates ANN and the difference in means of 3.66% is not statistically significant at the 5% level (p-value is 0.139). Regarding rationed and non-rationed firms, LR gives better prediction than ANN and the mean difference is significant for rationed classes. This result is also evidenced in Desai et al. (1997). Our findings also suggest that the two classifiers perform poorly when predicting the rare class (rationed firms). The prediction of the two models is more likely to be non-rationing leaning than rationing leaning. This could be attributed to the fact that rationed firms are under-represented in comparison to the non-rationed firms and the results could be biased toward the second class of firms (the majority class). We used a combination of the oversampling and undersampling method called the ROS/RUS approach to balance the data such that the learning sample consists of 123 rationed and 123 non-rationed firms and we come back to our original sample in the testing phase. We finally choose this method after having tested several approaches (cost-modifying approach, oversampling
and undersampling, threshold-modifying approach) that did not yield good results. The results of cross-validation with the ROS/RUS technique demonstrate that in the two classifiers, the correct classification of the minority class increases, in particular for the LR model. The mean difference of 28.81% is statistically significant, demonstrating that LR outperforms ANN in predicting rationed firms. For the non-rationed class, the difference in means is also significant at the 5% level in favor of the ANN model. But when looking at the overall accuracy, the mean difference seems to be non-significant and does not allow a firm conclusion regarding the dominance of one of the two classifiers. The criterion of accuracy is based here on a specific threshold value of 0.5. We will look at other criteria, which are the ROC curve and the Area under the ROC Curve (AROC), which are independent of the decision threshold. These criteria are attractive for several reasons: they are considered as a more relevant performance measure when the class distribution is skewed and for unequal misclassification error costs. The ROC curve offers a more comprehensive view of the classifier's performance and the AROC summarizes the ROC curve and makes the comparison between classifiers easy. All the indicators set out above are placed only for this cutoff point and do not explain how the probabilities of True Positive (TP) and False Positive (FP) differ when the decision threshold varies. The ROC curve plots the different values of probabilities of True Positive (TP) and False Positive (FP) according to different threshold decisions and the AUC gives a single measure of classifier performance which makes comparison clearer. Some studies estimate that AROC provides a better measure than accuracy and should substitute for accuracy as an instrument for evaluating and comparing classifiers (Bradley, 1997; Ling et al., 2003). Figure 1 presents the ROC curves for ANN and LR without balanced data in the first instance and with balanced data in the second instance. ROC analysis shows that the ROC curve of LR lies above the ROC curve of ANN for nearly all threshold points, demonstrating that LR dominates ANN for balanced and imbalanced data. However, for imbalanced data, the curves draw near the diagonal which represents random guessing, with the evidence that both LR and ANN perform poorly in predicting CR decisions when the data are imbalanced. Learning from balanced data improves the performance of both classifiers. The second criterion that links directly to the ROC curve is the AROC. With random resampling in the learning phase, the best AROC was obtained using the LR model, 75.1% compared with 64.4% for ANN. Without the ROS/RUS technique, the results dropped to only 63.5% for LR compared with 53.5% for ANN. That shows that a resampling technique used to balance data improves the classification performance. Table V compares the AROC for ANN and LR using the Delong, Delong and Clarke-Pearson (1988) non-parametric test computing on Sigma Plot 12 software. The level of significance is set to 0.05. The test shows that the difference in AROC is statistically significant for the dependent samples. The p-value is equal to 0.004 for imbalanced data and 0.011 for balanced data.

Conclusion

The objective of this study is primarily to assess the predictive ability of LR and ANN to discriminate between rationed and non-rationed firms. As with many real-world classification problems, the distributions of the two classes in CR are skewed (Cole, 1998; Angelini et al. 1998; Cavalluzzo et al.; 2002; Blanchflower et al. 2003; Rand, 2007; Agostino et al. 2008, Muravyev et al. 2009; Hashi and Toci, 2010; Freel et al. 2012; Krasniqi, 2010, Becchetti et al. 2010). The Random Oversampling of the minority class and the Random Undersampling of the majority class

13
technique are carried out before the training phase in order to obtain the same number of observations in each class and, therefore, allow algorithms to recognize the minority class. We compared the classification performance of ANN and LR by using accuracy, the ROC curve and AROC. Our main findings are that when learning from imbalanced data, both LR and ANN give low accuracy of classification for the minority class (rationed firms) and are unable to discriminate between rationed and non-rationed firms by leaning on the majority class side. In this context, it is crucial to deal with the class imbalance problem by rebalancing the sample artificially before the training phase or by using other approaches that will reinforce the minority class. Our results show that when balancing distribution of the two classes, the performance of classifiers in recognizing the minority class increases compared to imbalanced data sets, particularly for the LR model which, ultimately, outweighs ANNs according to the ROC curve and AROC criteria. In the literature, it cannot be concluded that one technique outperforms the other without an empirical comparison (Tu, 1996). Unfortunately, to the best of our knowledge, there have been no studies to compare the effect of the classification performance on the CR decision and we cannot match up our findings with the empirical literature. The preference for LR may be explained by the fact that the CR task seems to have low complexity and there are no important interactions or complex nonlinearities supporting the use of ANN. We suggest using the methodology developed in our paper on a more complete sample with all the different types of CR, particularly self-rationing which is recognized by the recent literature to be more widespread than expected.
References


Bell, T. B. (1997). Neural nets or the logit model? A comparison of each model’s ability to predict commercial bank failures. Intelligent Systems in Accounting, Finance and Management, 6(3), 249-264.


### Table I Variable descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAT</td>
<td>0.24</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MULTI</td>
<td>0.51</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FIRMAGE</td>
<td>21.67</td>
<td>15.33</td>
<td>2</td>
<td>208</td>
</tr>
<tr>
<td>TA</td>
<td>14.09</td>
<td>1.33</td>
<td>10.16</td>
<td>17.54</td>
</tr>
<tr>
<td>LIQUI</td>
<td>0.16</td>
<td>0.18</td>
<td>0</td>
<td>0.99</td>
</tr>
<tr>
<td>BDEBT</td>
<td>0.12</td>
<td>0.12</td>
<td>0</td>
<td>0.70</td>
</tr>
<tr>
<td>ROA</td>
<td>0.05</td>
<td>0.11</td>
<td>-0.54</td>
<td>0.67</td>
</tr>
<tr>
<td>TANG</td>
<td>0.14</td>
<td>0.14</td>
<td>0</td>
<td>0.81</td>
</tr>
<tr>
<td>TRADE</td>
<td>69.91</td>
<td>43.72</td>
<td>0</td>
<td>365</td>
</tr>
<tr>
<td>AGE</td>
<td>48.26</td>
<td>9.43</td>
<td>24</td>
<td>77</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.09</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CARE</td>
<td>1.76</td>
<td>13.68</td>
<td>-154</td>
<td>128</td>
</tr>
<tr>
<td>EQUITY</td>
<td>0.34</td>
<td>0.22</td>
<td>-0.92</td>
<td>0.97</td>
</tr>
</tbody>
</table>

### Table II Univariate comparisons

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean: RAT=0</th>
<th>Mean: RAT=1</th>
<th>Difference</th>
<th>T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRADE</td>
<td>66.51</td>
<td>80.69</td>
<td>-14.18**</td>
<td>-2.18</td>
</tr>
<tr>
<td>CARE</td>
<td>1.94</td>
<td>1.20</td>
<td>0.74</td>
<td>0.36</td>
</tr>
<tr>
<td>MULTI</td>
<td>0.54</td>
<td>0.40</td>
<td>0.133*</td>
<td>1.79</td>
</tr>
<tr>
<td>GENDER</td>
<td>0.09</td>
<td>0.084</td>
<td>0.006</td>
<td>0.144</td>
</tr>
<tr>
<td>AGE</td>
<td>48.47</td>
<td>47.59</td>
<td>0.87</td>
<td>0.62</td>
</tr>
<tr>
<td>FIRMAGE</td>
<td>21.59</td>
<td>21.93</td>
<td>-0.33</td>
<td>-0.14</td>
</tr>
<tr>
<td>BDEBT</td>
<td>0.12</td>
<td>0.11</td>
<td>0.01</td>
<td>0.50</td>
</tr>
<tr>
<td>EQUITY</td>
<td>0.38</td>
<td>0.24</td>
<td>0.14***</td>
<td>4.22</td>
</tr>
<tr>
<td>TA</td>
<td>14.28</td>
<td>13.51</td>
<td>0.77***</td>
<td>3.95</td>
</tr>
<tr>
<td>TANG</td>
<td>0.138</td>
<td>0.141</td>
<td>-0.003</td>
<td>-0.17</td>
</tr>
<tr>
<td>ROA</td>
<td>0.065</td>
<td>0.031</td>
<td>0.033**</td>
<td>2.125</td>
</tr>
<tr>
<td>LIQUI</td>
<td>0.184</td>
<td>0.114</td>
<td>0.069**</td>
<td>2.498</td>
</tr>
</tbody>
</table>

This table reports the univariate tests on the difference between rationed and non-rationed firms. We report the mean, the mean difference, the p-value of the t-test with the assumption of normally distributed data. ***, **, and *, indicate significance at the 1%, 5% and 10% levels, respectively.
### Table III Leave-one-out cross validation of classifier’s performance

**Artificial Neural Network**

<table>
<thead>
<tr>
<th></th>
<th>With resampling</th>
<th>Without resampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Y=0)</td>
<td>True (Y=1)</td>
</tr>
<tr>
<td>Pred (Y=0)</td>
<td>149</td>
<td>37</td>
</tr>
<tr>
<td>Pred (Y=1)</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>Class Recall</td>
<td>79.68%</td>
<td>37.29%</td>
</tr>
<tr>
<td>Class Precision</td>
<td>80.11%</td>
<td>36.67%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>69.51%</td>
<td></td>
</tr>
<tr>
<td>Error rate</td>
<td>30.49%</td>
<td></td>
</tr>
<tr>
<td>AROC</td>
<td>0.644</td>
<td></td>
</tr>
</tbody>
</table>

**Logistic regression**

<table>
<thead>
<tr>
<th></th>
<th>With resampling</th>
<th>Without resampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True (Y=0)</td>
<td>True (Y=1)</td>
</tr>
<tr>
<td>Pred (Y=0)</td>
<td>130</td>
<td>20</td>
</tr>
<tr>
<td>Pred (Y=1)</td>
<td>57</td>
<td>39</td>
</tr>
<tr>
<td>Class Recall</td>
<td>69.51%</td>
<td>66.10%</td>
</tr>
<tr>
<td>Class Precision</td>
<td>86.66%</td>
<td>40.62%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>68.70%</td>
<td></td>
</tr>
<tr>
<td>Error rate</td>
<td>31.30%</td>
<td></td>
</tr>
<tr>
<td>AROC</td>
<td>0.751</td>
<td></td>
</tr>
</tbody>
</table>

### Table IV Pairwise comparisons between ANN and LR

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Overall</th>
<th>Rationed</th>
<th>Non-rationed</th>
<th>Without resampling</th>
<th>With resampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
</tr>
<tr>
<td>t-statistic</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
</tr>
<tr>
<td>sig (2-tailed)</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
<td>ANN</td>
<td>LR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>LR</th>
<th>ANN</th>
<th>LR</th>
<th>ANN</th>
<th>LR</th>
<th>AN</th>
<th>LR</th>
<th>AN</th>
<th>LR</th>
<th>AN</th>
</tr>
</thead>
<tbody>
<tr>
<td>74.80%</td>
<td>71.14%</td>
<td>28.81%</td>
<td>20.34%</td>
<td>89.30%</td>
<td>87.17%</td>
<td>-0.242</td>
<td>4.489</td>
<td>-2.762</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.139</td>
<td>0.096</td>
<td>0.451</td>
<td>0.809</td>
<td>0.000</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1-ROC curves for LR and ANN

ROC (Receiver Operating Characteristic) curve is plotted using the values of False positive rate (1-specificity) on the X-axis and the True positive rate (sensitivity) on the Y-axis by moving the score threshold. A (0, 1) corresponds to the ideal model with 1 true positive rate and 0 false positive rate. The line which connects the endpoints (0,0) and (1,1) represents a classifier making randomly guesses. If the ROC Curve is below the diagonal, then its performance is worse than random for all the different score of threshold.

Table V ROC Curve Area Comparison

<table>
<thead>
<tr>
<th>Pair (LR, ANN)</th>
<th>without resampling</th>
<th>with resampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Difference</td>
<td>0.1</td>
<td>0.107</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.034</td>
<td>0.042</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>[0.031, 0.168]</td>
<td>[0.023, 0.189]</td>
</tr>
<tr>
<td>Chi-square, DF = 1</td>
<td>8.182</td>
<td>6.332</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.004</td>
<td>0.011</td>
</tr>
</tbody>
</table>