Expert Systems with Applications 40 (2013) 356–364

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Credit scoring models for the microfinance industry using neural networks: Evidence from Peru

Antonio Blanco [a,*], Rafael Pino-Mejías [b], Juan Lara [c], Salvador Rayo [c]

[a] Department of Financial Economics and Operations Management, Faculty of Economics and Business Studies, University of Seville, Avda. Ramon y Cajal, 1, 41018 Seville, Spain
[b] Department of Statistics and Operational Research, Faculty of Mathematics, University of Seville, Avda. Reina Mercedes, s/n 41012 Seville, Spain
[c] Department of Financial Economics and Accounting, Faculty of Economics and Business Studies, University of Granada, Campus Cartuja, s/n 18071 Granada, Spain

ARTICLE INFO

Keywords:
Microfinance institutions
Classification rules
Multilayer perceptron
Linear discriminant analysis
Quadratic discriminant analysis
Logistic regression

ABSTRACT

Credit scoring systems are currently in common use by numerous financial institutions worldwide. However, credit scoring with the microfinance industry is a relatively recent application, and no model which employs a non-parametric statistical technique has yet, to the best of our knowledge, been published. This lack is surprising since the implementation of credit scoring should contribute towards the efficiency of microfinance institutions, thereby improving their competitiveness in an increasingly constrained environment. This paper builds several non-parametric credit scoring models based on the multilayer perceptron approach (MLP) and benchmarks their performance against other models which employ the traditional linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR) techniques. Based on a sample of almost 5500 borrowers from a Peruvian microfinance institution, the results reveal that neural network models outperform the other three classic techniques both in terms of area under the receiver-operating characteristic curve (AUC) and as misclassification costs.

1. Introduction

Over the last decade, the microfinance sector[1] has grown dramatically, and is currently considered as a booming industry. In the period 1998–2008, the number of microfinancial institutions (hereinafter, MFIs) grew by 474%, and the number of customers increased by 1048%. Attracted by this rapid growth, a large number of international commercial banks have started operating in the microfinance sector, viewing it as a potential for profitable investment. This injection of interest has increased the competition between the players in this industry, and has negatively affected the MFIs. The MFIs therefore need to increase their efficiency in all their processes, minimize their costs, and control their credit risk if they want to survive in the long-term. One way for the MFIs to become more efficient in order to compete with the commercial banks is the implementation of automatic credit scoring[2] systems to evaluate their credit applicants since credit scoring reduces the cost of credit analysis, improves cash flow, enables faster credit decisions, reduces the losses, and also results in the closer monitoring of existing accounts and the prioritization of repayment collection. To this end, Rhyne and Christen (1999) suggest that credit scoring is one of the most important uses of technology that may affect microfinance, and Schreiner (2004) affirms that experiments carried out in Bolivia and Colombia show that the implementation of credit scoring improves the judgment of credit risk and thus cuts, by more than $75,000 per year, the costs of MFIs. Nevertheless, and in contrast to the concentration of research on financial institutions, the development of credit scoring models in the microfinance sector has only undergone minor advances. Furthermore, those models in existence are based on traditional parametric statistical techniques, mainly linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR), despite the overwhelming evidence found in numerous studies which indicates that the non-parametric methodologies usually outperform these classic statistical models (for example, see Lee & Chen, 2005; West, 2000). That is, to the best of the authors' knowledge, in the existing literature

* Corresponding author. Tel.: +34 954 559 875; fax: +34 954 557 570.
E-mail address: aj_blanco@us.es (A. Blanco).

[1] In the microfinance sector, operate the Microfinance institutions (hereinafter MFIs) which offer savings services and small loans (namely microcredits) to those sectors of the population with the greatest problems of access to financial resources. Therefore, the MFIs exercise relevant social work since they financially support the poorest people, who, by creating a microenterprise, can escape the socioeconomic situation of exclusion in which they find themselves. For this reason, the goals and management criteria of the many MFIs lay less emphasis on business components and greater emphasis on social components than those used by their new competitors (international commercial banks).

[2] The objective of credit scoring models is to assign credit applicants to one of two groups: either to a 'good credit' group that is likely to repay the financial obligation or a 'bad credit' group that should be denied credit because of a high likelihood of defaulting on the financial obligation (Hand & Henley, 1997).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.07.051

no credit scoring model designed for the microfinance industry applies a non-parametric methodology, and therefore, the microfinance industry has not yet benefited from the advantages of non-parametric techniques to improve the performance of credit scoring models, and hence are failing to compete on equal terms with their new competitors, the international commercial banks. Of the few credit scoring models developed for MFIs, all have used parametric methodologies, particularly LDA and LR (Kleimeier & Dinh, 2007; Rayo, Lara, & Camino, 2010; Reinke, 1998; Sharma & Zeller, 1997; Viganò, 1993; Vogelgesang, 2003; Zeller, 1998). However, the strict assumptions (linearity, normality and independence among predictor variables) of these traditional statistical models, together with the pre-existing functional form relating response variables to predictor variables, limit their application in the real world. Several authors (for example, Karels & Prakash, 1987; Reichert, Cho, & Wagner, 1983) point out that two basic assumptions of LDA are often violated when applied to credit scoring problems: (a) the independent variables included in the model are multivariate and normally distributed, (b) the group dispersion matrices (or variance–covariance matrices) are equal across the failing and the non-failing groups (for a detailed analysis of the problems in applying discriminant analysis in credit scoring models, see Eisenbeis, 1978). In the cases where the covariance matrices of the two populations are unequal, theoretically, QDA should be adopted, although LDA is reported to be a more robust and precise technique (Dillon & Goldstein, 1984). In the same way as LDA, LR is also optimal under the assumption of multivariate normal distributions with equal covariance matrices, and LR also remains optimal in a wider variety of situations. However, logistic regression requires larger data sets to obtain stable results, interactions between predictor variables must be formulated, and complex non-linear relations between the dependent and independent variables could be incorporated through appropriate but not evident transformations. For these reasons, in recent years, non-parametric statistical models, such as the *k*-nearest neighbor algorithm (Henley & Hand, 1996), support vector machines (Vapnik, 1998), decision tree models (Davis, Edelman, & Gammerman, 1992), and neural network models (Patuwo, Michael, & Ming, 1993), have been successfully applied to credit scoring problems. Of these, artificial neural networks (ANNs) constitute one of the most powerful tools for pattern classification due to their non-linear and non-parametric adaptive-learning properties. Many studies have been conducted that have compared ANNs with other traditional classification techniques in the field of credit scoring models, since the default prediction accuracies of ANNs are better than those using classic LDA and LR (Arminger, Enache, & Bonne, 1997; Desai, Conway, Crook, & Overstreet, 1997; Desai, Crook, & Overstreet, 1996; Hand & Henley, 1997; Lee & Chen, 2005; Lee, Chiu, Lu, & Chen, 2002; Malhotra & Malhotra, 2002; Markham & Ragsdale, 1995; Patuwo et al., 1993; Piramuthu, 1999; Srinivasan & Ruparel, 1990; West, 2000). However, despite yielding satisfactory results, ANNs also feature certain disadvantages, such as its black box nature and the long training process involved in the design of the optimal network topology (Chung & Gray, 1999).

The main goal of this paper is therefore to develop a credit scoring model specially designed for the microfinance industry by using multilayer perceptron neural networks (hereinafter, MLP). Moreover, we also compare the performance of MLP models against the three parametric techniques most widely used: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR). Based on a large sample which contains financial and non-financial variables of almost 5500 borrowers from a Peruvian MFI, seventeen credit scoring models are created, of which fourteen are MLP-based models.

The remainder of our paper proceeds as follows. In Section 2, details of our data set are provided, and a detailed examination of the variables available is undertaken in order to predict the default. In Section 3, several credit scoring models specifically designed for MFIs are developed. To his end, various methodologies are employed: Fisher discriminant analysis, logistic regression, and multilayer perceptron. In Section 4, the results of different models are shown and their comparison is made. An extensive discussion on the results is also carried out. Finally, Section 5 provides the main conclusion of this study and future research lines are analyzed.

2. Data and variables

2.1. The data set

We use a data set of microcredits from a Peruvian Microfinance Institution (*Edpyme Proempresa*). Our dataset contains customer information during the period 2003–2008 related to: (a) personal characteristics (marital status, gender, etc.); (b) economic and financial ratios of their microenterprise; (c) characteristics of the current financial operation (type interest, amount, etc.); (d) variables related to the macroeconomic context; and (e) any delays in the payment of a microcredit fee. After eliminating missing and abnormal cases, 5451 cases remain. From among these, 2673 (49.03%) are default cases, and 2778 (50.97%) are not. In line with other studies (for example, Schreiner, 2004), a microcredit presenting a delay in repayment of at least fifteen days is defined as default microcredit. To perform an appropriate comparison of the classification models, (LDA, QDA, LP, and MLP), our final data set is randomly split into two disjoint sub-sets; a training set of 75% and a test set of 25%. The test sample contains a total of 1363 cases (51.80% failed and 48.20% non-failed). The configuration of parameters of each model is selected through a 10-fold cross-validation procedure, as described in Sections 3.1–3.3. One advantage of cross-validation is that the credit scoring model is developed with a large proportion of the available data (75% in this case).

2.2. Description of input variables

Table 1 shows the input variables used in this study.[3] They provide the various characteristics of borrowers, lenders, and loans. Numerous qualitative variables are considered in our study, since: (a) Schreiner (2004) suggests that the input variables of the credit scoring forces the microfinance sector to be more qualitative and informal than those considered by traditional banks; and (b) recent literature concludes that the inclusion of qualitative variables improves the prediction power of models. Moreover, since the default of borrowers has a close relationship with the general economic situation, variables linked to the macroeconomic context are also considered as input variables. With respect to the dependent variable, default of the microcredit, this takes a value of 1 if the microcredit fails, and 0 otherwise.

The first ratio indicates the number of times the income exceeds total assets. Therefore, we estimate that the ratio ($R1$) is inversely related with respect to the probability of default. The ratio $R2$ measures the relationship between the gross and operating costs of the microenterprise. As with the previous ratio, we expect that the sign of its coefficient is negative since the higher the value of this ratio, the more solvent the income/loss of the firm, and the lower the financial difficulties. The third financial ratio ($R3$) measures the liquidity of the microenterprise. Due to the design of this ratio, the higher its value, the lower the probability of default. Therefore, the sign of the estimator is expected to be negative. The fourth

[3] This table also shows the expected sign of the relationship between each input variable and the probability of default. The statistical descriptions of all the input variables are shown in Table 1 and Table 2 of Appendix 1. These statistics are presented for each group (failed and non-failed).

Table 1
Description of financial, non-financial and economic variables.

Variable	Description	Expected estimator sign (β)
Financial ratios		
*R*1	Asset rotation: income sales/total assets	–
*R*2	Productivity: gross utility/operating costs	–
*R*3	Liquidity: cash/total asset liquidity	–
*R*4	Liquidity rotations: cash/income Sales $\times$ 360	+
*R*5	Leverage1: total liabilities/(total liabilities + shareholders' total equity)	+
*R*6	Leverage2: total liabilities/shareholders' equity	+
*R*7	ROA: net income/total assets	–
*R*8	ROE: net income/shareholders' equity	–
Non-financial information		
Zone	Geographical location of the agency or branch. Dummy variable: (0) central zone, (1) Outskirts	+
Old	Duration as a borrower of the MFI. Numeric variable	–
Previous_Loan_Granted	Previously granted credits. Numeric variable	–
Loan_Granted	Loans granted in the last year. Numeric variable	–
Loan_Denied	Previously denied loans. Numeric variable	+
Sector	Activity sector of the micro-business. Categorical variable: (0) commerce, (1) agriculture, (2) production, (3) service	±
Purpose	Destination of microcredit. Dummy variable: (0) work capital, (1) fixed asset	+
Mfi_Class	MFI customer classification Dummy variable: (0) normal customer, (1) customer with repayment problems of any sort	+
Total_Fees	Total number of fees paid in credit history. Numeric variable	–
Arrears	Number of arrears. Numeric variable	+
Ave_Arrear	Average (days) of customer default. Numeric variable	+
Max_Arrear	Number of days of major default. Numeric variable	+
Gender	Borrower gender. Dummy variable: (0) male, (1) female	–
Age	Age at time of application. Numeric variable	±
Marital_St	Marital Status. Dummy variable: (0) single, (1) family unit	–
Employm_St	Employment Status of borrower. Dummy variable: (0) owner, (1) dependent	±
Guarantee	Guarantee presented. Dummy variable: (0) sworn declaration, (1) real guarantee	+
Currency	Type of currency for loan granted. Dummy variable: (0) Peruvian Nuevos Soles (PEN) (1) US Dollar ($)	+
Amount	Amount of microcredit. Numeric variable	–
Duration	Number of monthly fees for applied loan. Numeric variable.	+
Interest_R	Monthly interest rate for microcredit. Numeric variable	+
Forecast	Loan officer forecast: credit situation at expiration. Dummy variable: (0) without problems, (1) with problems	+
Macroeconomic indicators		
GDP	Rate of annual change of Gross Domestic Product (GDP) during loan term	–
CPI	Rate of annual change of Consumer Price Index (CPI) during loan term	+
Empl_R	Rate of annual change of variation of employment rate (ER) during loan term	–
ER	Rate of annual change of variation of exchange rate (ER) PEN[a]-$ during loan term	+
IR	Rate of annual change of interest rate (IR) during loan term	+
SEI	Rate of annual change of stock exchange index (SEI) during loan term	–
Water	Rate of annual change in cost of municipal water during loan term	+
Electricity	Rate of annual change in cost of electricity during loan term	+
Phone	Rate of annual change in cost of telephone consumption during loan term	+

[a] Peru's currency is the Nuevo Sol, denoted by the ISO code PEN.

financial ratio ($R4$) indicates the number of days the microenterprise takes to recover its treasury. In this case, the larger the value of this variable, the greater the likelihood of default. Therefore, the expected sign of the estimator is positive. The fifth financial ratio ($R5$) represents the percentage of liabilities that have microenterprises in their financial structure. We understand that a high level of liabilities inversely affects the ability of micro-entrepreneurs to pay. Consequently, a positive sign of the estimator of this variable is expected. The sixth financial ratio ($R6$) measures the ratio between the amount of debt and equity, and thus complements the information provided by the previous variable. We estimate that a high debt ratio results in an increase in the likelihood of default, which would be a negative estimate. The seventh financial ratio ($R7$) measures the return on assets (ROA). A higher return on assets should help reduce the likelihood of default. A negative sign of this variable estimator is therefore expected. The final financial variable ($R8$) measures the return on equity (ROE), that is, the return accrued by property of the company. The greater the financial return of a firm, the smaller its probability of default. We therefore consider that the sign of the estimator of this variable should be negative. Customers who both live in a central area and locate their microenterprise in a central area usually run less risk of financial distress than those in rural areas. Therefore, the variable *Zone* is expected to have a positive sign in the estimator. The age of the relation MFI-customer implies that the bank knows the payment history of a customer in detail, and this is why the variable *Old* is inversely related to the probability of default. The variables *Previous_Loan_Granted* and *Loan_Granted* are expected to have a negative sign in the estimator for the same reasons as for the variable *Old*; bearing in mind that a lasting relationship with the financial institution involves the lender knowing all the risk inherent to the customer and also believing that this customer is reliable, Crook, Hamilton, and Thomas (1992). For customers with loans denied in the past, the risk of financial problems is more present. Thus, it is considered that the sign of the variable *Loan_Denied* is positive. Since there is no previous reference that suggests a criterion for the consideration of a sector with more financial problems than others, the sign of the variable *Sector* remains undetermined. For the variable *Purpose*, we propose a positive sign as we understand that the microcredit destined to the acquisition of an asset implies a greater risk than a credit destined for working capital because the process of asset recovery through depreciation takes longer. Borrowers with any problem of payment in the past (greater risk) take the value of 1 in the variable *Mfi_Class*, thus, we consider that the sign of the estimator is positive. The higher the amount in fees the customer has paid, the greater the experience as a customer, and the less the probability of default. Therefore, the sign of the estimator of the variable *Total_Fees* is negative. However, the variables *Arrears*, *Ave_Arrears* and *Max_Arrears* are closely related to the probability of non-payment, and hence their estimator has positive signs. According to Schreiner (2004), women are better payers than men. Consequently a negative sign is considered for the variable *Gender*. There is no empirical evidence about the relationship between the variable *Age* and the probability of default; therefore the sign of the estimator for this variable cannot be determined. Customers responsible for a family unit usually have better payment behavior of their debts than those who are single, that is, those without family obligations (Kleimeier & Dinh, 2007). For this reason, the variable *Marital_St* must have a negative estimator. A positive estimator is expected in the variable *Employ_St*, since customers who have some experience in the running of a microenterprise, have a lower probability of default than those who have only worked as an employee (that is, without any experience as micro-entrepreneurs). In microfinance, the reputation of the borrower is the main guarantee. Hence, *E. Proempresa* asks for only a sworn statement of their property from those customers who rarely have problems in the fulfillment of their payment obligations. On the other hand, real guarantees are demanded from both new customers and those who in the past have had problems with payments. Therefore, the sign of the estimator of the variable *Guarantee* must be positive. A microcredit granted in foreign currency (not in local currency) is affected by a risk in the rate of exchange and, for that reason a positive sign is expected in the estimator of the variable *Currency*. On the other hand, *E. Proempresa* only accepts a microcredit request for high amounts if customers have paid their previous microcredits without any problem. Therefore, microcredits of high amounts correspond to old customers and good payers, since these customers have a lower probability of default than those with microcredits whose amount is lower. Therefore, a negative sign is expected in the estimator of the variable *Amount*. It is widely supported by the literature on credit risk that the bank which lends money over the long-term runs a greater risk of default than those banks that give short-term loans. Therefore, the coefficient of the variable *Duration* must be positive. The higher the interest rate of a financing source is, the more difficulties the borrower has repaying it. Consequently, the variable *Interest_R* must have a positive estimator. Finally, we believe that an important variable, albeit totally subjective, is the risk analyst's opinion on the probability that a customer may have financial problems. Just as defined for the variable *Forecast*, we expect a positive sign in its estimator.

On the other hand, we also introduce variables with information about the economic cycle since the absence of this kind of variable has historically implied a major limitation of financial distress models. Furthermore, as stated by Kim and Sohn (2010) the macroeconomic environment is a key factor that directly affects the payment behavior of any borrower. The macroeconomic variables under consideration are calculated through the following expression:

$$\Delta VM_{i,j} = \frac{VM_{i+j} - VM_i}{VM_i} \tag{1}$$

where $\Delta VM_{i,j}$ is the variation rate of the considered macroeconomic variable and VM is the considered macroeconomic variable and i is the moment of the granting of the loan and j is the microcredit duration.

3. Research methodology and experimental design

3.1. Discriminant analysis credit scoring model

Given two multivariate independent samples where p quantitative predictor variables have been observed for n_i cases, $i = 1, 2$, $n = n_1 + n_2$, the LDA model supposes that both populations are multivariate normal with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and common covariance matrix $\boldsymbol{\Sigma}$. The LDA rule classifies a p-dimensional vector $\mathbf{x}$ to class 2 if

$$\mathbf{x}^t \widehat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2}\hat{\mu}_2^t \widehat{\Sigma}^{-1}\hat{\mu}_2 - \frac{1}{2}\hat{\mu}_1^t \widehat{\Sigma}^{-1}\hat{\mu}_1 + \log \hat{\pi}_1 - \log \hat{\pi}_2 \tag{2}$$

where the prior probabilities of class memberships π_1 and π_2 are usually estimated by the class proportions in the training set. Linear Discriminant Analysis provides the minimum misclassification rate, and therefore it is optimal under the hypothesis previously described. This rule can be expressed as class 2 if $D > 0$, where D is the linear discriminant function, computed through a linear combination of the inputs. The classification rule can also be formulated by predicting class 1 if the estimated probability for the class 1 is greater than a threshold probability p_c. This last value can be selected by the empirical optimization of the classification error. As suggested in Hastie, Tibshirani, and Friedman (2001), a K-fold cross-validation may be followed. The training data set is randomly split into K roughly equal-sized parts. For the kth part, the LDA model is fitted to the other $K-1$ parts, and the classification error for each possible p_c is computed on the kth part. The mean classification error of the K parts is obtained for each p_c. Ninety-nine possible values for p_c (0.01, 0.02, ..., 0.99) are considered in our study, and the value minimizing the 10-fold cross-validation classification error set is selected, namely 0.35.

Linear Discriminant Analysis is fitted with the R function *lda* (Venables & Ripley, 2002), available in the MASS library. A variable selection process with the function 'greedy.wilks' of the package 'klaR' of R (Weihs, Ligges, Luebke, & Raabe, 2005) is first performed. In this case the initial model is defined by starting with the variable which separates the groups most. The model is then extended by including further variables depending on the Wilk's lambda criterion: select the one which minimizes the Wilk's lambda of the model and the variable is included if its p-value still shows statistical significance.

The LDA model can be explained through the coefficients of the linear discriminant function. However, the simplicity of the model can be insufficient to capture complex structures in the dataset. Moreover, the optimality of the LDA classification rule requires the data to be independent and normally distributed while the covariance matrices are also required to comply with the homoscedastic assumption (Johnson & Wichern, 2002). When the covariance matrices are not assumed to be equal, quadratic discrimination functions are computed, and hence the QDA rule yields

$$\arg\max_i \delta_i(\mathbf{x}), \quad \delta_i(\mathbf{x}) = -\frac{1}{2}\log|\widehat{\Sigma}_i| - \frac{1}{2}(\mathbf{x}-\hat{\mu}_i)^t\widehat{\Sigma}_i^{-1}(\mathbf{x}-\hat{\mu}_i) + \log\hat{\pi}_i \tag{3}$$

The R function *qda* (Venables & Ripley, 2002) in the MASS library is used in our case study. A similar search for the cut point is also carried out for the QDA model through the same set of 99 threshold probabilities as in LDA, thereby obtaining 0.99.

3.2. Logistic regression credit scoring model

For a binary response and p quantitative predictors $x_1, \ldots, x_p$, (some of which may be dummy variables for coding qualitative variables, as in LDA and QDA), the LR model assumes that the probability of the target response is

$$\pi(x_1,\ldots,x_p) = \frac{e^{\beta_0+\beta_1x_1+\cdots+\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\cdots+\beta_px_p}} \tag{4}$$

There are several inferential procedures to test the statistical significance of the whole model and of the individual significance of each variable. The model may also be interpreted for which a great family of diagnostics and criteria are available to identify influential and outlying observations. Logistic regression can be fully embedded in a formal decision framework, but in order to perform a comparison with the other models, a threshold probability needs to be specified, which corresponds to varying the prior class probabilities. Thus 99 possible values for this threshold probability (0.01, 0.02, ..., 0.99) are also considered, and that value which minimizes the 10-fold validation error is selected, thereby obtaining 0.58.

We have fitted the LR model with the *glm* function in R (Venables & Ripley, 2002), in an attempt to compute the maximum likelihood estimators of the $p+1$ parameters by an iterative weighted least squares (IWLS) algorithm. In the same way as in LDA, a previous stepwise procedure is run in order to select the most significant variables. The function 'step.glm' of R is employed, which applies a forward sequential procedure based on the Akaike Information Criterion.

Again, in the same way as LDA, LR is also optimal under the assumption of multivariate normal distributions with equal covariance matrices; although LR remains optimal in a wider variety of situations. However, LR requires larger data sets in order to obtain stable results, and complex nonlinear relations between the dependent and independent variables could be incorporated through appropriate but not evident transformations.

3.3. Artificial neural networks credit scoring models

Artificial Neural Networks (ANNs) constitute a computational paradigm which provides a great variety of mathematical nonlinear models, useful for tackling a wide range of statistical problems. Several theoretical results support a particular architecture, namely the multilayer perceptron (MLP), an example being the universal approximate property, as in Bishop (1995). Moreover, MLP is the most commonly used type of neural network in business studies (Vellido, Lisboa, & Vaughan, 1999; Zhang, Patuwo, & Hu, 1998). Following these results, we have considered a three-layered perceptron where the output layer is formed of one node which provides the estimation of the probability of default. This value is computed with the logistic activation function $g(u) = e^u/(e^u+1)$, also used in the hidden layer. By denoting H as the size of the hidden layer, $\{v_{ih}, i = 0,1,2,\ldots,p, h = 1,2,\ldots,H\}$ as the synaptic weights for the connections between the p-sized input an1d the hidden layer, and $\{w_h, h = 0,1,2,\ldots,H\}$ as the synaptic weights for the connections between the hidden nodes and the output node, then the output of the neural network from a vector of inputs $(x_1,\ldots,x_p)$ is

$$\hat{y} = g\left(w_0 + \sum_{h=1}^{H} w_h g\left(v_{0h} + \mathop{m}_{j=1}^{p} v_{ih}x_j\right)\right) \tag{5}$$

The output of this model provides an estimation of the probability of default for the corresponding input vector. A final decision can be obtained by comparing this output with a threshold, usually set at 0.5, thereby reaching a decision of default if $\hat{y} > 0.5$.

One major disadvantage of MLP is the fact that there is no known procedure which guarantees that a global solution can be attained for the problem of finding a configuration of synaptic weights that minimizes the usual error criteria, and hence one of the many possible local minima is often obtained through one of the many learning rules proposed in the literature. A further drawback is its black-box nature, which makes it very difficult to interpret the resulting model, although certain relevant proposals exist, from among which stand out Bayesian neural networks (Neal, 1996).

As input nodes, our MLP models use the set of variables selected for the sequential parametric model that has the highest area under the receiver operating characteristic curve[4] (LR model). Nevertheless, since performance of the MLP can be improved with normalization of the quantitative input variables, the range of each predictor variable is mapped into the $[-1, 1]$ interval. No general rule exists for the determination of the optimal number of hidden nodes: a crucial parameter for the optimal network performance (Kim, 2003). The most common way to determine the size of the hidden layer is via experiments or trial and error (Tang & Fishwick, 1993; Wong, 1991). The number of hidden nodes determines the complexity of the final model, and networks of a more complex nature fail to ensure better generalization capability. One well-known strategy is based on some type of validation study (Hastie et al., 2001) and therefore we selected the size of the hidden layer (H) through a 10-fold cross-validation search in $\{1, 2, \ldots, 20\}$.

Two different programs are used in the construction of the MLP credit scoring models. The first choice is the freely available R system. The *nnet* R function (Venables & Ripley, 2002) fits single-hid-

[4] Hereafter, AUC.

den-layer neural networks by means of the BFGS procedure, a quasi-Newton method also known as a variable metric algorithm, in an effort to minimize an error criterion which allows a decay term λ in order to prevent overfitting problems.[5] For classification problems, one appropriate error function is the conditional maximum likelihood (or entropy) criterion (Hastie et al., 2001). Defining $W = (W_1, \ldots, W_M)$ as the vector of all M coefficients of the net, and given n targets $y_1, \ldots, y_n$, where $y_i = 1$ for microcredit default, and $y_i = 0$ otherwise, the BFGS method is applied to the following problem:

$$\underset{\mathbf{W}}{Min}\sum_{i=1}^{n}(y_i \ln \hat{y}_i + (1-y_i)\ln(1-\hat{y}_i)) + \lambda\left(\sum_{i=1}^{M} W_i^2\right) \quad (6)$$

The R implementation of an MLP model requires the specification of two parameters: the size of the hidden layer (H) and the decay parameter (λ), and therefore a 10-fold cross-validated search of the size of the hidden layer (H) and the decay parameter (λ) is carried out over a grid defined as $\{1,2,\ldots,20\} \times \{0,0.01,0.05,0.1,0.2,\ldots,1.5\}$. In this case, we have also considered training without regularization, where $\lambda = 0$.

The Neural Network Toolbox (Demuth & Beale, 1997) with MATLAB R2010b constitutes the other tool employed to fit MLP. This commercial system offers a great variety of learning rules, and we have considered the following six main learning algorithms to train the MLP: gradient descent, gradient descent with momentum, BFGS quasi-Newton (similar to R), Levenberg–Marquardt, scaled conjugate gradient, and resilient back-propagation. The first algorithm is the traditional back-propagation method originally proposed with MLP, and hence it is included in our study, accompanied by the variant based on a momentum term. These two learning rules require a key parameter, the learning rate. Rumelhart, Hinton, and Williams (1986) concluded that lower learning rates tend to give the best network results and the networks are unable to converge when the learning rate is greater than 0.012. For this reason, learning rate 0.010 is tested during the training process of MLPs that use the gradient descent and its variant based on a momentum term as training algorithms. In our case, as recommended by MATLAB, the momentum takes the value 0.90. The other four methods are recommended in the MATLAB documentation for classification problems, and are widely known as second-order training algorithms. These six learning rules try to minimize a sum of squared errors (SSE):

$$\underset{\mathbf{W}}{Min}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

As in R, there remains the problem of selecting H, and therefore the size of the hidden layer (H) is chosen through a 10-fold cross-validation search in $\{1,2,\ldots,20\}$ for each learning method.

MATLAB allows the use of early stopping in MLP training. This well-known strategy splits the training data set into effective training and validation sets, and the error on the validation set is monitored during training. When the validation error begins an increasing trend, the training process is stopped because an overfitting phenomenon may have been initiated. We have trained the MATLAB neural nets both with early stopping (25% of size) and without early stopping.

The basic parameters of all the fitted MLP models can be seen in Table 2. Firstly, several MLP models are fitted using the traditional gradient descendent back-propagation training algorithm. Secondly, second-order training algorithms (quasi-Newton back-propagation, Levenberg–Marquardt back-propagation, resilient back-propagation, and scaled conjugate gradient back-propagation) are implemented in order to develop further MLPs. These six learning rules are made through MATLAB. And finally, another two MLPs are then fitted using the R system, one of which applies the regularization procedure.

[5] The BFGS algorithm can be found in Bishop (1995).

3.4. Model evaluation measures

The area under the ROC curve (AUC) is often employed in classification problems. In this paper, the AUC is computed with the aid of the ROCR library available in R (Sing, Sander, Beerenwinkel, & Lengauer, 2005). However, it is well known that, in order to evaluate the overall default prediction capability of the designed models, the prior probabilities and the misclassification costs should also be considered (West, 2000). It is apparent that the cost associated with a Type I error (a customer with good credit is misclassified as a customer with bad credit) and a Type II error (a customer with bad credit is misclassified as a customer with good credit) are frequently very different. Generally, the misclassification costs associated with Type II errors are much higher than those associated with Type I errors. According to West (2000), the relative ratio of misclassification costs associated with Type I and Type II errors must be 1:5,[6] and hence special attention should be paid to Type II errors of all models constructed. In accordance with West (2000), we express the function on computing the expected misclassification cost when only two populations are considered as:

$$Cost = C_{21}P_{21}\pi_1 + C_{12}P_{12}\pi_2 \quad (8)$$

where π_1 and π_2 are prior probabilities of good and bad credit populations, P_{21} and P_{12} measures the probability of making Type I errors (a customer with good credit is misclassified as a customer with bad credit) and Type II errors (a customer with bad credit is misclassified as a customer with good credit), respectively, and C_{21} as well as C_{12} are the corresponding misclassification costs of Type I and Type II errors. In order to compute the expected misclassification costs of the various default prediction models, the estimates of misclassification probability and misclassification costs have first to be calculated. The most commonly adopted estimates for P_{21} and P_{12} are the fraction of good-credit customers misclassified as bad-credit customers and the fraction of bad-credit customers misclassified as good-credit customers, where the two coefficients differ and are independent from each model.

4. Results and discussion

In this section, the performance of the three parametric models (LDA, QDA and LR) are first discussed and compared, and, secondly, the various MLPs developed are benchmarked with respect to the classic techniques. Finally, the statistical characteristics of the best credit scoring models are described.

The input variables selected in the sequential selection process and the values of their coefficients for LDA and LR models are shown in Tables 3 and A1 of Appendix 1. Table 3 contains the AUC, Types I–II errors and misclassification costs of all the models built. Focusing on the parametric models, we observe that the AUC of LDA and QDA models are 93.03% and 91.98%, both of which are lower than the AUC of the LR model (93.22%). Therefore, in line with other authors (Lee et al., 2002; Ohlson, 1980), we find that the LR model outperforms LDA and QDA.[7] However, when the misclassification cost criteria are employed, QDA has the lowest misclassification costs (50.77%) of the all parametric models. Thus, in contrast with the results obtained with the AUC criteria, QDA shows better performance, according to misclassification costs, than the

[6] Many other authors use this ratio (1:5); however, the real costs associated to each type of error depend on each individual lender.

[7] Since the LR approach has the highest AUC; all the MLPs use only the significant variables of the LR model as input nodes (for more details see Section 3.2. above).

Table 2
Basic parameters of multilayer perceptron models.

Models	Training algorithm	Software	Hidden nodes	Early stopping	Regularization	% Training	% Validation
MLP 1	Gradient descent	Matlab	14	No	No	100	0
MLP 2	Gradient descent	Matlab	14	Yes	No	75	25
MLP 3	Gradient descent with momentum	Matlab	10	No	No	100	0
MLP 4	Gradient descent with momentum	Matlab	10	Yes	No	75	25
MLP 5	BFGS quasi-Newton	Matlab	9	No	No	100	0
MLP 6	BFGS quasi-Newton	Matlab	9	Yes	No	75	25
MLP 7	Levenberg–Marquardt	Matlab	2	No	No	100	0
MLP 8	Levenberg–Marquardt	Matlab	2	Yes	No	75	25
MLP 9	Scaled conjugate gradient	Matlab	14	No	No	100	0
MLP 10	Scaled conjugate gradient	Matlab	14	Yes	No	75	25
MLP 11	Resilient	Matlab	9	No	No	100	0
MLP 12	Resilient	Matlab	9	Yes	No	75	25
MLP 13	BFGS quasi-Newton	R	10, $\lambda = 0$	No	No	100	0
MLP 14	BFGS quasi-Newton	R	3, $\lambda = 0.2$	No	Yes	100	0

Table 3
AUC, Type I–II errors, and misclassification costs in the test sample.

MODELS	AUC	Type I errors (%)	Type II errors (%)	Misclassification costs
LDA (greedy.wilks)	0.9303	8.52	18.27	0.5143
QDA (qda)	0.9198	11.72	17.42	0.5077
LR (glm)	0.9322	5.94	20.96	0.5715
MLP 1	0.9023	9.40	24.40	0.6772
MLP 2	0.9124	8.20	22.90	0.6326
MLP 3	0.9015	15.30	21.50	0.6305
MLP 4	0.9458	7.60	16.70	0.4691
MLP 5	0.9079	11	15.70	0.4597
MLP 6	0.9427	7.60	17.10	0.4795
MLP 7	0.9389	4.40	22.40	0.6014
MLP 8	0.9413	3.70	22.40	0.5980
MLP 9	0.9148	12.60	18.30	0.5347
MLP 10	0.9459	7.60	16.70	0.4692
MLP 11	0.9395	10.70	15.30	0.4478
MLP 12	0.9357	8.50	17.60	0.4968
MLP 13	0.9236	6.68	22.81	0.6230
MLP 14	0.9543	7.76	15.30	0.4337

LDA and LR models.

With respect to the non-parametric methodology, the results show that, in at least several cases, the accuracy performance of the MLP models is better than that of the LDA, QDA and LR models. However, in term of AUC, the results obtained for all methodologies are similar. Relevant differences are obtained in terms of the misclassification costs.[8] For the MLP models, the highest AUC and lowest misclassification cost are obtained when the second-order algorithms re implemented. That is, our results suggest that the gradient descendent algorithm is less efficient than the second-order algorithms considered in this study. However, when the gradient descendent algorithm is implemented with momentum, then the performance, both in terms of AUC and misclassification costs, improves considerably (see model MLP 4 in Table 3). Therefore, the traditional gradient descent is clearly superseded in our data set. According to Table 3, the model with the highest performance is the MLP 14. It is a three-layer perceptron, with 20 input nodes, 3 hidden nodes and one output node. The training has been performed with R, using a BFGS quasi-Newton learning rule, and both the size of the hidden layer and the regularization parameter are selected by 10-fold cross-validation, the value of this latter parameter being 0.2. Table 3 shows that early stopping in MATLAB models improve the AUC, but misclassification costs are not lower in all learning rules. In the R model regularization improves both AUC and misclassification costs, so it is worthwhile this added parameter selection process.

In brief, we conclude, in line with other authors (for example, see Lee & Chen, 2005; West, 2000), that, in general, not only do MLP models have a greater AUC but also lower misclassification costs than the traditional LDA, QDA and LR approaches. These empirical results confirm the theoretical superiority (principally, non-linear and non-parametric adaptive-learning properties) of the MLP models over the parametric and widely used LDA, QDA and LR models when applied to pattern classification problems. Moreover, there is no requirement for MLP models to assume the strict assumptions of traditional statistical models, nor to assume pre-existing functional forms by relating response variables to predictor variables which result in their limited application in the real world. However, the major disadvantages of an MLP model include: (a) its black-box nature, which renders the resulting model very difficult to interpret; and, (b) its long training process in designing the topology of the optimal network. However, despite these disadvantages of MLP models, we consider MFIs should use these models instead of the traditional parametric models since even a minor improvement in predictive accuracy of the MLP default-prediction model is of critical value. Just a mere 1% improvement in accuracy would reduce losses in a large loan portfolio and save millions of dollars (West, 2000). The differences, in terms of the misclassification costs, between the best MLP (model MLP 14) with respect to the LDA, QDA and LR models, are 8.06%, 7.04%, and 13.78%, respectively. That is, the implementation of neural network approaches help to reduce the MFI losses significantly, and therefore, provides a way

[8] In this study, the values selected for the calculation of the misclassification costs are: $C_{21} = 1$ and $C_{12} = 5$ (as recommended by West (2000)), P_{21} and P_{12} are dependent of each model; and $\hat{\pi}_1 = 0.482$ and $\hat{\pi}_2 = 0.518$. For further details on these coefficients, see Section 3.4 above.

to obtain a competitive advantage over other MFIs which fail to implement this methodology.

5. Conclusion and futures research lines

Credit scoring systems are currently in common use by the majority of financial institutions worldwide. However, the application of credit scoring within the microfinance industry is a relatively recent issue. In recent years, the use of non-parametric methodologies and the introduction of non-financial variables into credit scoring models have boomed in the specialized literature. However, very little research deals with both issues, and, to the best of the authors' knowledge, this is the first study which applies a non-parametric methodology (MLP) to create a credit scoring systems for the microfinance industry. For this reason, in this paper, 14 multilayer perceptron (MLP) credit-scoring models are fitted and compared by using a Peruvian microfinance institution sample which contains financial and non-financial variables. In addition, these non-parametric models are benchmarked with the results of the traditional LDA, QDA and LR methodologies.

Our findings show that multilayer-perceptron credit scoring can work for microfinance institutions, and obtain higher accuracy in performance and lower misclassification costs than the classic LDA, QDA and LR models. These results imply major consequences for the efficiency of MFIs due to the cost savings. Thus, the best MLP involved provides a misclassification cost with a reduction of 8.06%, 7.04%, and 13.78% in comparison with the LDA, QDA, and LR models, respectively. That is, the implementation of a neural network approach supposes that the MFIs reduce their losses in terms of millions of dollars, and therefore provides a way for the MFIs to achieve a competitive advantage over their competitors (mainly commercial banks), since it constitutes a key to an increas-

Table A1
Statistical description of quantitative independent variables.

Variable	Failed		Non-Failed	
	Mean	Standard deviation	Mean	Standard deviation
R1	0.7637	0.8055	0.8436	0.8528
R2	3.9421	4.8284	3.8881	6.8548
R3	0.0683	0.0689	0.1448	3.2438
R4	0.1301	0.1368	0.1654	1.9812
R5	0.1421	0.1617	0.1196	0.1474
R6	0.2242	0.3227	0.1810	0.2789
R7	0.1531	0.1764	0.1771	0.2756
R8	0.1799	0.2015	0.2012	0.2911
Old	2.3468	1.5110	2.2397	1.5099
Previous_Loan_Granted	5.3900	5.0040	5.0600	4.6940
Loan_Granted	3.4600	2.3040	4.3400	2.3400
Loan_Denied	0.3200	0.5380	0.3300	0.5360
Mfi_Class	0.3500	0.4770	0.1100	0.3110
Total_Fees	36.1800	25.8510	31.7100	22.8390
Arrears	13.0400	10.7870	13.3400	11.1700
Ave_Arrear	8.0000	8.1510	6.8600	6.4340
Max_Arrears	20.2000	27.7650	16.5500	21.5030
Age	43.0175	10.6148	42.5628	10.4770
Amount	0.7338	0.6548	0.6458	0.5998
Duration	8.1100	4.7950	7.0300	3.5520
Interest_R	4.9242	0.9183	5.1255	0.8801
GDP	8.8985	29.7134	4.8139	26.3989
CPI	2.6377	2.2101	3.1247	2.1318
Empl_R	3.5702	10.6827	2.8671	9.6861
ER	−2.4123	4.4517	−5.5607	3.8899
IR	5.9631	13.9525	12.1717	11.7493
SEI	44.5991	32.4527	49.5322	33.3754
Water	2.4576	3.7483	3.1681	4.2243
Electricity	3.6054	12.2598	8.5162	10.4552
Phone	-7.1809	8.0019	-1.7179	3.8308

Table A2
Statistical description of qualitative independent variables.

Variable	Categories	Failed (%)	Non-Failed (%)
Zone	Center	46.94	53.06
	Outskirts	55.84	44.16
Sector	Commerce	48.53	51.47
	Agriculture	60.68	39.32
	Production	53.22	46.78
	Service	54.31	45.69
Purpose	Work capital	47.07	52.93
	Fixed asset	77.51	22.49
Gender	Male	51.32	48.68
	Female	50.71	49.29
Marital_St	Single	50.73	49.27
	Family unit	51.06	48.94
Employm_St	Owner	50.81	49.19
	Dependent	70.73	29.27
Guarantee	Sworn declaration	58.50	41.50
	Real guarantee	43.47	56.53
Currency	PEN	89.30	92.10
	$	10.70	7.90
Forecast	Without problems	42.94	57.06
	With Problems	97.27	2.73

Table A3
Significant variables using linear discriminant analysis.

Linear discriminant analysis model	
Variable[a]	Coefficient
Forecast	2.2062*
ER	0.1684*
CPI	−0.0956*
Total_Fees	0.0125*
Arrears	−0.0232*
Mfi_Class	0.7577*
Guarantee	−0.2508*
Duration	−0.0684*
IR	−0.0461*
Empl_R	−0.0290*
Electricity	−0.0125*
Purpose	0.3559*
SEI	0.0040*
GDP	−0.0052*
Zone	0.1412*
R8	−0.3811*
Max_Arrears	−0.0022*
R2	0.0077*

[a] ***p-Value < 0.001; **p-value < 0.01. *p-value < 0.05.

Table A4
Significant variables using logistic regression.

Logistic regression model	
Variable[a]	Coefficient
Forecast	4.2624***
ER	0.3477***
Total_Fees	0.0221***
Arrears	−0.0449***
Mfi_Class	1.2592***
Guarantee	−0.6117***
IR	−0.1011***
Empl_R	−0.0247**
Purpose	0.6048**
GDP	−0.0235***
Zone	0.4209***
Water	0.0346*
Duration	−0.1275***
Intercept	0.2685

[a] ***p-Value < 0.001. **p-value < 0.01. *p-value < 0.05.

ingly constrained environment. Moreover, empirical evidence has also been attained which supports the fact that MLP models trained with second-order algorithms obtain a significantly better performance (both in terms of AUC and misclassification costs) than those that use the traditional gradient descent. Therefore, we suggest that microfinance institutions apply neural network approaches, especially those using second-order training rules, when setting up their credit scoring models, instead of employing the parametric LDA, QDA and LR models.

This paper offers an appropriate solution so that the MFIs can benefit from all the positive aspects that the implementation of the credit scoring systems involves, such as the increase in efficiency, profitability and market share, reduction of costs and losses, and professional-image management. Hence MFIs will be able to create competitive advantages and compete with commercial banks by using advanced risk-management tools.

This study can be further improved in future research in several ways. Firstly, more relevant variables may be collected in an effort to increase the prediction accuracies of the models. And secondly, other newly developed classification methodologies, such as other kinds of artificial neural networks (e.g. radial basis function, learning vector quantization, fuzzy adaptive resonance, and Bayesian learning neural networks), classification and regression trees (CART), and support vector machines (SVM), can be employed and their results can then be compared with those of the MLP, LDA, QDA and LR models established in this paper.

Appendix A

Tables A1–A4.

References

Arminger, G., Enache, D., & Bonne, T. (1997). Analyzing credit risk data: A comparison of logistic discriminant classification tree analysis and feedforward networks. *Computational Statistics, 12*, 293–310.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.

Chung, H. M., & Gray, P. (1999). Special section: Data mining. *Journal of Management Information Systems, 16*, 11–16.

Crook, J. N., Hamilton, R., & Thomas, L. C. (1992). A comparison of discriminations under alternative definitions of credit default. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), *Credit scoring and credit control* (pp. 217–245). Oxford: Oxford University Press.

Davis, R. H., Edelman, D. B., & Gammerman, A. J. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Mathematics Applied in Business & Industry, 4*(1), 43–51.

Desai, V. S., Conway, J. N., Crook, J. N., & Overstreet, G. A. (1997). Credit scoring models in the credit-union environment using neural networks and genetic algorithms. *IMA Journal of Mathematics Applied in Business & Industry, 8*(4), 323–346.

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*(1), 24–37.

Demuth, H., & Beale, M. (1997). *Neural network toolbox for use with Matlab.user's guide*. The Math Works Inc..

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis methods and applications*. New York: Wiley.

Eisenbeis, R. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal of Banking and Finance, 2*(3), 205–219.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society, Series A, 160*(3), 523–541.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Henley, W. E., & Hand, D. J. (1996). A *k*-nearest neighbor classifier for assessing consumer credit risk. *Journal of the Royal Statistical Society, Series D (The Statistician), 44*(1), 77–95.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River: Prentice-Hall.

Karels, G., & Prakash, A. (1987). Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance Accounting, 14*(4), 573–593.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*(1–2), 307–319.

Kim, H. S., & Sohn, S. Y. (2010). Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research, 201*(3), 838–846.

Kleimeier, S., & Dinh, T. A. (2007). Credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis, 16*(5), 471–495.

Lee, T. S., & Chen, I. F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications, 28*(4), 743–752.

Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications, 23*(3), 245–254.

Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *European Journal of Operational Research, 136*(1), 190–211.

Markham, I. S., & Ragsdale, C. T. (1995). Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis. *Decision Sciences, 26*(2), 229–242.

Neal, R. M. (1996). *Bayesian learning for neural networks*. New York: Springer.

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

Patuwo, E., Michael, Y. H., & Ming, S. H. (1993). Two-group classification using neural networks. *Decision Sciences, 24*(4), 825–845.

Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research, 112*(2), 310–321.

Rayo, S., Lara, J., & Camino, D. (2010). A credit scoring model for institutions of microfinance under the Basel II normative. *Journal of Economics, Finance & Administrative Science, 15*(28), 89–124.

Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business and Economic Statistics, 1*(2), 101–114.

Reinke, J. (1998). How to lend like mad and make a profit: A micro-credit paradigm versus the start-up fund in South Africa. *Journal of Development Studies, 34*(3), 44–61.

Rhyne, E., & Christen, R. P. (1999). *Microfinance enters the marketplace*. USAID Microenterprise Publications.

Rumelhart, D. E., Hinton, D. E., & Williams, R. J. (1986). *Learning internal representations by error propagation in parallel distributed processing*. Cambridge, MA: MIT Press.

Schreiner, M. (2004). Scoring arrears at a microlender in Bolivia. *Journal of Microfinance, 6*(2), 65–88.

Sharma, M., & Zeller, M. (1997). Repayment performance in group-based credit programs in Bangladesh: An empirical analysis. *World Development, 25*(10), 1731–1742.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing the performance of scoring classifiers, R package version 1.0-1. <http://rocr.bioinf.mpi-sb.mpg.de/>.

Srinivasan, V., & Ruparel, B. (1990). CGX: An expert support system for credit granting. *European Journal of Operational Research, 45*(2–3), 293–308.

Tang, Z., & Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing, 5*(4), 374–385.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Springer.

Vellido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural network in business: A survey of applications (1992–1998). *Expert Systems with Applications, 17*(1), 51–70.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S-PLUS*. New York: Springer.

Viganò, L. A. (1993). Credit scoring model for development banks: An African case study. *Savings and Development, 17*(4), 441–482.

Vogelgesang, U. (2003). Microfinance in times of crisis: The effects of competition, rising indebtness, and economic crisis on repayment behaviour. *World Development, 31*(12), 2085–2114.

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR. Analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support* (pp. 335–343). Berlin: Springer-Verlag.

West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27*(11–12), 1131–1152.

Wong, F. S. (1991). Time series forecasting using backpropagation neural networks. *Neurocomputing, 2*, 147–159.

Zeller, M. (1998). Determinants of repayment performance in credit groups: The role of program design, intra-group risk pooling, and social cohesion. *Economic Development and Cultural Change, 46*(3), 599–620.

Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*, 35–62.