Neural Networks, Ordered Probit Models and Multiple Discriminants

Evaluating Risk Rating Forecasts of Local Governments in Mexico.

Alfonso Mendoza-Velázquez*
Economics Department
Universidad Popular Autónoma del Estado de Puebla (UPAEP)

Pilar Gómez-Gil**
Computational Science Department
National Institute of Astrophysics, Optics and Electronics (INAOE)

Abstract

Credit risk ratings have become an important input in the process of improving transparency of public finances in local governments and also in the evaluation of credit quality of state and municipal governments in Mexico. Although rating agencies have recently been subjected to heavy criticism, credit ratings are indicators still widely used as a benchmark by analysts, regulators and banks monitoring financial performance of local governments in stable and volatile periods. In this work we compare and evaluate the performance of three forecasting methods frequently used in the literature estimating credit ratings: Artificial Neural Networks (ANN), Ordered Probit models (OP) and Multiple Discriminant Analysis (MDA). We have also compared the performance of the three methods with two models, the first one being an extended model of 34 financial predictors and a second model restricted to only six factors, accounting for more than 80% of the data variability. Although ANN provides better performance within the training sample, OP and MDA are better choices for classifications in the testing sample respectively.

Keywords: Credit Risk Ratings, Ordered Probit Models, Artificial Neural Networks, Discriminant Analysis, Principal Components, Local Governments, Public Finance, Emerging Markets.

JEL classification: H79, C25, C63.

---

* *Contact author:* Center for Research and Economic Intelligence (CIIE). Universidad Popular Autónoma del Estado de Puebla (UPAEP), 21 Sur 1103 Col. Santiago, C.P. 72160 Puebla, Pue. México. Phone: +52 (222) 229 94 00, ext. 7780. Toll free number (in Mexico): 01800 224 2200. Email: amv101@yahoo.com, alfonso.mendoza@upaep.mx.
** Computational Science Department. National Institute of Astrophysics, Optics and Electronics (INAOE), Luis Enrique Erro # 1, Tonantzintla, Puebla, México, Tel.: +52 (222) 266 31 00 ext. 8228. E-mail: pgomez@ccc.inaoep.mx, pgomez@acm.org

I. **Introduction**

Credit risk assessment is one important element in the financial and fiscal arrangement of local governments in Mexico. Since 2001 credit risk ratings have become a rather compulsory prerequisite for every State or Municipal government looking for cheaper bank or capital markets financing. Any local government issuing debt needs to be rated at least by two separate credit risk rating agencies.

A public finance credit risk rating is an opinion about a local government's ability and willingness to pay. With no credit rating these two attributes would be difficult to evaluate in Mexico due to the lack of timely and reliable information about public finances. As the events in the ongoing worldwide crisis have revealed, failing to assess credit quality based on quality information can lead to financial bankruptcy, default, crisis and contagion.

Despite the heavy criticism to rating agencies, credit risk ratings—with all their imperfections—are tools still widely considered by analysts and are among the very few parameters available to monitor the health and soundness of local govenrments' public finances.[1] Commercial banks and financial creditors for instance use risk ratings as a benchmark to calculate capital reserves and to manage default risk. The bigger the gap between the State credit risk rate and the sovereign risk rate, the bigger will be the required reserve capital and, therefore, a higher interest rate the local government will be charged for such credit.

Regulatory bodies are also interested in monitoring the financial performance of Mexican State governments in order to detect liquidity problems and potential defaults, especially during economic crisis as the one Mexico is facing since mid – 2008. All states in Mexico count now with at least two credit ratings provided by international agencies. Standard & Poor's, Moodys and FitchRatings—the most widely used international rating agencies—have been operating in Mexico since 2001. These credit ratings are reassessed with a time frequency that becomes too slow in times of distress. Given the importance and need of monitoring the health of local finances in the aftermath of the crisis, it would be very useful to count with reliable methods to approximate the credit quality of local governments at any moment, without waiting for the next 'official' credit opinion.

[1] Official information for instance is not readily available and is not fully trustworthy due to a lack of consistent accountancy principles among other factors.

Empirical finance researchers have used different classification methods to estimate credit ratings of corporate firms and banks such as Multiple Discriminant Analysis (MDA)—a widely popular and accepted method among practitioners—and Limited Dependent-Variable Models, such as Ordered Probit (OP) and Logit models. Although research has been extensive examining assets, firms and even sovereign governments, very little has been investigated with respect to the risk rating of local governments. Formal literature on this is practically inexistent and, to our knowledge, there is no research using Artificial Neural Networks (ANN) methods in the study of local public finances.

In this article we apply three methods to classify and predict credit quality on using risk ratings of local public finance in Mexico. We employ credit risk ratings and other financial data freely available from the rating agency Fitch Ratings related to public finances in the States from 2001. This is the first formal research examining the forecasting properties of three models of credit ratings applied to local governments: Multiple Discriminant Analysis (MDA), Ordered Probit (OP) and Artificial Neural Networks (ANN). Several financial factors are used as explanatory variables and, in order to control for multicollinearity, we also use the method of Principal Components. This last approach also allows us to investigate the performance of these methods using a large vs. a small set of explanatory variables.

The paper is divided as follows: The next section briefly reviews the literature on forecasting methods and their application to credit risk ratings, as well as their applications in economics and finance. Section three describes the methods used throughout this paper to evaluate the credit quality of local public finances based on risk ratings. Section four analyzes the evaluation results while section five presents some conclusions.

2. Brief Literature Review

Artificial Neural Networks (ANN) algorithms have gained some popularity in applications to social science, economics and business. Various surveys have reported the use of ANN for modeling foreign exchange, capital markets, investments, macroeconomics, bankruptcy forecasts and credit risk assessment.[2] It has been found in

[2] See Wood and Bhaskar (2006) and Wong et al. (1997) for applications in finance, business and operations. The first author provides around 100 references with applications to ANN in studies published since 1995. Since then, the number of papers related to this subject has grown significantly.

general that these algorithms tend to provide better outcomes than other statistical or mathematical methods. Among the competing methodologies used successfully in the analysis of credit quality of issuers and issues, we find Multiple Discriminant Analysis (MDA), categorical models or limited dependent variable models such as Probit or Logit models, genetic algorithms, linear programming (LP), among others.

Empirical applications of these methods range from sovereign credit ratings, to standard issuers such as banks, insurance companies and several financial and non-financial firms. In sovereign credit ratings for instance, Benell, et al. (2006) developed two ANN algorithms and compared them with an Ordered Probit model. The authors found that ANN algorithms provide much better forecasts than Ordered Probit models. These methods have also been applied in financial firms to the analysis of bankruptcy in savings associations. One instance is the contribution by Salchenberger et al. (1992), who report that ANN over perform Ordered Logit models.

Lee (2007) compares Multiple Discriminant Analysis with an ANN algorithm based on backpropagation to examine risk ratings of corporate credits. He reports that ANN algorithms provide better results than MDA. Also, in the evaluation of corporate credit, Kumar and Bhattacharya (2006) compared MDA with ANN and found that ANN performs better than MDA. In terms of the method, Patuwo (1993) found that the size of the training data (in contrast with the neural network architecture), help to maximize the correct classification rate and concluded that ANN provide better forecasts than MDA.

On the performance of the methods used in the analysis of credit risk there is some growing literature. Comparing MDA, genetic algorithms and logistic regression, Desai (1997) found that ANN have a better classification performance than other classification models. Markham (1995), among other authors, has combined ANN and MDA to get joint models that are able to provide better results than both techniques individually. Some other authors, such as Ting Peng, et. al. (1992), have proposed to integrate ANN, computer simulations and optimization simultaneously to obtain better results. Support vector machines (SVM)[3] have also been employed trying to get better credit risk classifications. For example, in the study of Taiwanese high technology companies using SVM, Huang (2009) reported that integration of nonlinear graphs for reduction of dimensionality provides better results than other classical methods. Some other techniques based on genetic programming have also been explored, for instance

[3] Support Vector Machines are a subclass of artificial neural networks that map data into a high dimensional space, where linear classification is performed (Huang, 2009).

Tokinaga (2005) combined these methods with ANN for binary classifications. In such an application, the author analyzes the probability of bankruptcy in Japanese industries from 1970 to 1986. As in other applications, this study found that ANN provides a better fit and forecasts than Multiple Discriminant Analysis.

Kotsiantis et al. (2006) use ANN to examine fraudulent finance declarations and corporate bankruptcy forecasts, they also find a better fit with respect to other classification methods. Bharat and Barin (1995) used ANN to model the market price of Initial Public Offers (IPOs) associated to a vector of economic variables. They found that sub-valuations, commonly observed in IPO’s, are reduced by around 8% using ANN. Gutierrez and Serrano (2007) used several financial and social indicators to forecast credit risk of micro-credit institutions using ANN. These authors also found that, in contrast with social variables, financial factors significantly explain credit ratings.

International research to evaluate credit ratings of sub-sovereign governments is very limited, and research examining the case of Mexico is practically non-existent. It is evident that most of the studies related to credit risk ratings focus on corporate or debt markets in developed countries, while very little has been found or written about developing markets. In Mexico there are some pioneering studies analyzing local public finances using Ordered Probit models—see García-Romo, et al. (2010) and Yorio (2006)—but, to our knowledge, there is not a single formal academic reference comparing the forecasting ability of the three most popular methods in emerging markets and none on local government public finances in particular.

We regard this gap in the literature as an opportunity to examine the case of an emerging market and a small sample size. We compare three methods: Artificial Neural Networks, Ordered Probit Models and Multiple Discriminant Analysis. In addition, by using the method of principal components, we investigate whether a small set of explanatory variables is more effective than a set of several predictors. On this we report in striking results.

3. Artificial Neural Network Algorithm and Ordered Probit Model

We briefly describe in this section, an algorithm known as backpropagation used in this research to train a feedforward artificial neural network. Also, the main features of Ordered Probit Models are presented.

3.1 Artificial Neural Network

Artificial Neural Networks (ANN) are mathematical models inspired in the behavior of biological neurons, that have shown success in the modeling of systems where the governing rules are unknown, but at the same time there are many reliable empirical examples describing them (Gómez-Gil 2009). The modeling is achieved through the ANN ability to learn, that is, ANN adjust their behavior and produce an output according to situations shown to them by examples. ANN can summarize and numerically represent essential information obtained from examples. Such information is obtained by a process known as “learning” or “training.” In addition to this, ANN, once trained, are able to generalize their outcomes, so that they are able to deal with variations or noise in the incoming information, producing ‘correct’ answers in spite of these variations.

ANN are composed of basic processing elements known as neurons, which are in turn inter-connected via numerical values called weights. ANN receive incoming data, which is processed following specific evaluation rules and strategies of connections in the neurons (commonly known as network topology or topography). One popular style of neuron is the perceptron (Haykin 1999). Neural nets made with perceptrons are able to produce outputs representing classifications, forecasts, evaluations, etc. Topological connections in ANN may take different forms; in one of them, neurons are organized in groups called layers, where members of one layer connect to members of the next layer. This model is known as multi-layer perceptrons (MLP). One of the most commonly used algorithms to train MLP is called *backpropagation* (Haykin 1999). There are however other models of ANN used for classification and forecasting, with different connection strategies and training algorithms, for example: Radial-basis Function Networks (Light 1992), Self Organizing Maps (Kohonen 1988) or Recurrent Neural Networks (Mandic and Chambers 2001). In this study we prefer MLPs using backpropagation algorithms, due to its proven ability to accurately train the network, to learn input-output mappings from training samples (Chen and Jain 1994) and its strength over noisy input data (Werbos 1994).

More formally, a perceptron is a neuron y_j with an output defined as:

$$y_j = \tau\left(\sum_{i=1}^{m} w_{ji}x_i + b_j\right) \tag{1}$$

where:

w_{ji} are weights connecting to neuron *j* from *m* other neurons located at a previous layer or from *m* external inputs to the network.

b_j is an additional weight, commonly known as bias, associated to the *j-th* neuron. The value of *j* goes from 1 to the number of neurons in the layer where the node is located.

x_i is the output of *i-th* neuron found in a layer previous to the one where the perceptron *j* is found. Both neurons are connected by w_{ji}. Alternatively, x_i may be the *i-th* external input to the network.

$\tau(.)$ is a continuous and differentiable function defining the activation rule of neuron y_j.

m is the number of connections to neuron *j*.

A multi-layer perceptron with *m* inputs, one hidden level with *h* neurons and one neuron in the output layer, as the one shown in figure 1, defines a system able to approximate the value of any arbitrary function $f(x_1, x_2, ..., x_m)$ (Haykin 1999). Inputs to this ANN correspond to values of the independent variables $x_1, x_2, ..., x_m$; its output (the function approximation) is defined as:

$$F(x_1, x_2, ..., x_m) = \sum_{j=1}^{h} \alpha_j \tau(\sum_{i=1}^{m} w_{ji} x_i + b_j) \tag{2}$$

where:

w_{ji}, b_j, for $j = 1,...,h;\ i = 1,...,m$ are weights connecting neurons in the hidden layer to external inputs,

α_j, for $j = 1,...,h$ are weights connecting the single neuron in the output layer with neurons in the hidden layer,

$\tau(u) = \frac{1}{1 + e^{-\lambda u}}$ is the activation function used for neurons in the hidden layer; λ is a scaling coefficient controlling the behavior of the activation function in a range where $\tau'(u) \neq 0$, an important condition to facilitate training.

It should be noted that the activation function of the single neuron located in the output layer of the MLP defined in (2) is a linear function of the type $\mu(x) = x$.

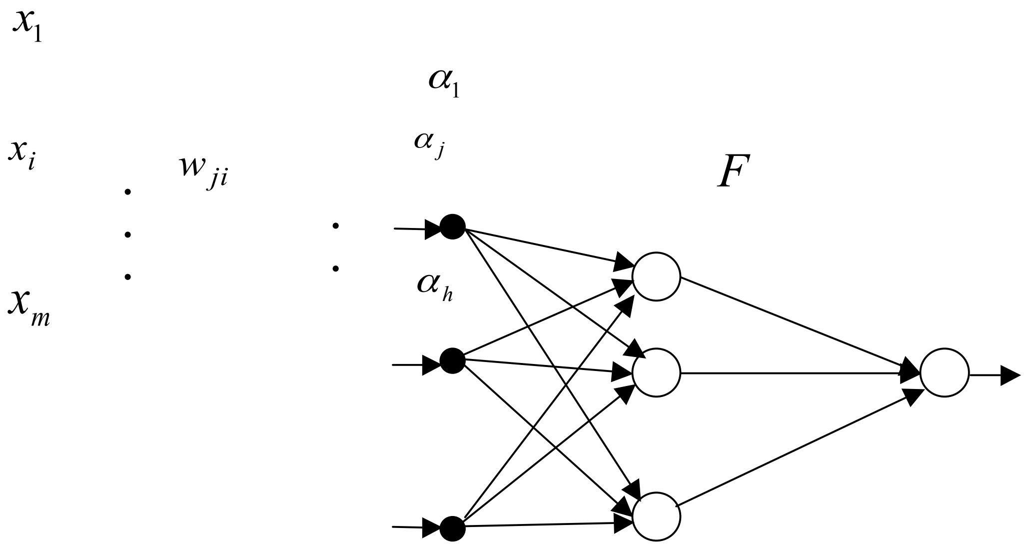

Figure 1. A multi-layer perceptron (MLP) with m inputs and one hidden layer with h neurons, able to approximate a function $f(x_1,...,x_m)$.

The ANN, defined by equation (2) and shown in figure 1, is used in this work to approximate the unobserved variable credit rating of local governments in Mexico. In such model the observed financial variables are inputs to the ANN, and the output is the calculated credit rating, that is, the approximated function. As defined by equation (2), the output of the ANN is a non-linear function of these input variables. The approximation capability of this ANN is supported by the universal approximation theorem (Cybenko 1989), which ensures that there exist values $w_{ji}, \alpha_j,$ and b_j such that:

$$\left| F(x_1,...,x_m) - f(x_1,...,x_m) \right| < \varepsilon \text{ for } \varepsilon > 0 \quad , \tag{3}$$

where F is defined according to equation (2), f is the approximating unknown function and ε is a small number.

3.1.1 Training Strategy

In this research we use the backpropagation algorithm developed by Werbos (1990) to adjust weights in the MLP, according to the derivation described by Rumelhart *et al.* (1986). Backpropagation is a supervised learning method based on repetitive presentation of examples, derived from a gradient descendent minimization of a cost function. This algorithm aims at progressively reducing the output error (E_{total}) generated by the network when a set of P training samples are presented to the network where the correct outputs (in this case, the credit ratings) D_p are known; that is, the algorithm aims to minimize:

$$E_{total} = \sum_{p=1}^{P} E_p \tag{4}$$

where:

$$E_p = \left| F_p - D_p \right| \tag{5}$$

F_p is the output of the ANN when the *p-th* training sample is evaluated; D_p is the correct or desired output. Modifications to network weights are done iteratively, using the training samples several times until a desired minimum error E_{total} is achieved or until a maximum of sweeps over the training set is executed.

Next we describe an algorithm to train, that is, to find the right weights w_{ji}, b_j and α_j of the ANN defined by equation (2) using *P* training samples. The MLP, once trained, will be able to find the unobserved credit rating of a given State. This algorithm is based in the backpropagation derivation defined by Rumelhart *et al*. (1986).

The training algorithm takes the following steps:

Step 1. Initialize weights.

1.1 Assign small random values to weights:

$w_{ji} = random(-0.01, 0.01)$

$b_j = random(-0.01, 0.01) \quad i = 1,\ldots,m, j = 1,\ldots,h$

m is the number of input variables,

h is the number of neurons in the hidden layer. This value is experimentally defined, as described at section 4.2.

$random(a,b)$ is a function generating random numbers, uniformly distributed in the interval [a, b].

1.2 Initialize a counter for iterations:

$sweeps = 0$

Step 2. Repeat:

2.1 Set the value of accumulated error among expected and desires values of the network for this iteration to zero:

$E_{total} = 0$

2.2 For $\forall p \in P$, (the training set) do:

$$O_j = \sum_{i=1}^{m} w_{ji} x_{ip} + b_j \text{ , for each } j = 1..h$$

x_{ip} is the *i-th* explanatory variable of the *p-th* sample in the training set P

$$F_p = \sum_{j=1}^{h} \alpha_j \tau(O_j)$$

$\tau(u) = \dfrac{1}{1+e^{-\lambda u}}$, λ is the scaling coefficient (see description following equation 2)

$$E_p = \left|F_p - D_p\right|$$

$$E_{total} = E_{total} + E_p$$

$$\delta_1 = (D_p - F_p)F_p(1 - F_p)$$

$$\alpha_j = \alpha_j + \eta\delta_1 O_j \text{ , } j = 1..h$$

$$\delta_{2j} = \tau'(O_j)\delta_1\alpha_j \text{ , } j = 1..h$$

$$b_j = b_j - \eta\delta_{2j} \; j = 1..h$$

$$w_{ji} = w_{ji} + \eta\delta_j x_i \text{ , } j = 1..h, i = 1..m$$

2.3 $sweeps = sweeps + 1$

Until $E_{total} \le 0.05$ or *sweeps* reaches a maximum desired number of iterations over the training set[4].

Step 3. End.

3.1.2 Algorithm to Assign a Credit Rating

Once trained, the network is ready to assign a credit rating, according to the scale defined by FitchRatings[5]. Since the ANN is being used as an approximation realization of a function $f(\mathbf{x}) : \Re^m \to \Re$, it is required to transform the output of the ANN, which is a real value, to the best integer value corresponding to a value in the scale at table A.1. The algorithm to do so is next described:

[4] E_{total} is used as a convergence criteria for this learning algorithm. The value 0.05 is experimentally chosen.

[5] See table A.1 in appendix.

Step 1. Read observed or input variables x_k with $k = 1,...,m$; where m is the number of observed variables.

Step 2. Calculate $F = \sum_{j=1}^{h} \alpha_j \tau(\sum_{i=1}^{m} w_{ji} x_i + b_j)$

Step 3. Calculates credit rating:

$$PossibleCategory = floor(F + 0.5)$$

where $floor(x)$ is a function returning the maximum integer, less than or equal to its argument x.

Step 4. Ensure a valid credit rating:

If $PossibleCategory < 1$ or $PossibleCategory > MAX\_CATEGORY$

$AssignedRating = null$

else

$AssignedRating = PossibleCategory$

where $MAX\_CATEGORY$ corresponds to the maximum numerical value allowed according to table A.1. If the ANN generates an output value greater that the highest possible rating allowed (that is 9 corresponding to rating BBB), or smaller than 1 (corresponding to rating AAA), the network declares itself unable to assign a rating (null value).

3.2 Ordered Probit Model

Ordered Probit models are built using a *latent or unobserved variable model* satisfying the assumptions of the classical lineal regression model (Wooldridge 2001):

$$y_t^* = \sum_{k=1}^{K} \beta_k x_{kt} + \varepsilon_t \tag{6}$$

It is assumed that the unobserved credit rating for year t, y_t *, is a linear function of a series of explanatory variables x_k, observed in year t, and an error term ε_t, which, in the case of Probit models, is assumed to follow a normal distribution.

During the period of analysis there is a rating range from AA+ to BBB. Therefore, a discrete number is assigned to each rating: 1 for AAA, 2 for AA+ and so on until 9 is assigned to BBB.[6] The relationship between the unobserved rating y^* and the observed rating y_t is as follows:

$$y_t = \begin{cases} \text{AAA} & =1 & \text{if } \ y^* \le \theta_1 \\ \text{AA}+ & =2 & \text{if } \ \theta_1 < y^* \le \theta_2 \\ \text{AA} & =3 & \text{if } \ \theta_2 < y^* \le \theta_3 \\ \dots & \dots & \dots \\ \text{BB} & =12 & \text{if } \ \theta_{11} < y^* \le \theta_{12} \end{cases} \tag{7}$$

Parameters θ_i represent the index cut points and mark the threshold for every rating. Such parameters, as well as the coefficients associated to the explanatory variables β_k, are estimated via Maximum Likelihood.[7]

The probability of obtaining each rating is given by:

$$\begin{aligned} P(y_t = 1) &= F(-\sum_{k=1}^{K} \beta_k x_{kt}) \\ P(y_t = 2) &= F(\theta_1 - \sum_{k=1}^{K} \beta_k x_{kt}) - F(-\sum_{k=1}^{K} \beta_k x_{kt}) \\ P(y_t = 3) &= F(\theta_2 - \sum_{k=1}^{K} \beta_k x_{kt}) - F(\theta_1 - \sum_{k=1}^{K} \beta_k x_{kt}) \end{aligned} \tag{8}$$

…

$$P(y_t = 9) = 1 - F(\theta_8 - \sum_{k=1}^{K} \beta_k x_{k,t})$$

where $F(\dots)$ is a cumulative distribution function. That is, the probability to obtain a rating AAA is equal to the probability that y_t is less or equal to 1; the probability of obtaining a AA+ rating is equal to the probability of y_t being less or equal to 2, minus

[6] See table A.1 in appendix.

[7] Estimation of Ordered Probit models using Maximum Likelihood provides better estimators and forecasts than estimations using Ordinary Least Squares (OLS). This allows us to overcome the two main drawbacks of Linear Probability Models: adjusted probabilities from Probit models are strictly at [0,1]—due to the use of the normal distribution function—and this model allows the partial effects of independent variables to vary (see Wooldridge, 2001).

the probability of being less or equal to 1 and so on, up to the probability of obtaining a rating BBB equal to 1 minus probability that y_t is less or equal to 8.

Due to the non-linear nature of *F(z)*, the cumulative distribution function, the coefficients associated to all dependent variables in the *latent variable model* do not represent the corresponding marginal effects. Therefore, the partial effect becomes the partial derivative of the probability of obtaining a rating *j* with respect to variable x_k:

$$\begin{aligned}
\frac{\partial P(y=1)}{\partial x_k} &= -[f(-\sum_{k=1}^{K}\beta_k x_{kt})]\beta_k \\
\frac{\partial P(y=2)}{\partial x_k} &= [f(-\sum_{k=1}^{K}\beta_k x_{kt}) - f(\theta_1 - \sum_{k=1}^{K}\beta_k x_{kt})]\beta_k \\
\frac{\partial P(y=3)}{\partial x_k} &= [f(\theta_1 - \sum_{k=1}^{K}\beta_k x_{kt}) - f(\theta_2 - \sum_{k=1}^{K}\beta_k x_{kt})]\beta_k \\
\frac{\partial P(y=12)}{\partial x_k} &= [f(\theta_{11} - \sum_{k=1}^{K}\beta_k x_{kt})]\beta_k
\end{aligned} \tag{9}$$

where *f*(…) is the normal probability density function. In this manner, the change in probability of a local government given a unit variation in variable x_k, holding constant (at their mean values) the rest of the variables, can be estimated. It should be noted that the sign of the marginal effect of a given variable will remain unaffected over the probability that $y = 9$, and an opposite sign over the probability that $y = 1$. For the rest of the variables, sign concordance is ambiguous and can be determined only after estimation, because it depends on the values taken by the rest of variables.

Similarly, if the sign of coefficient β_k is negative, the marginal effect of variable x_k will increase the probability of obtaining a rating y=1 and decrease the probability of obtaining y=9; the contrary occurs if β_k is positive. For the rest of the ratings, the sign of the marginal effect may be determined only after estimation using the exact values taken by the variables. Therefore, the sign in the marginal effect determines whether variations of explanatory variables are related to increases or decreases in the probability of obtaining a given rating.

4. Evaluation of ANN, Ordered Probit and Multiple Discriminat Analysis

This section presents the evaluation of forecasts made via Artificial Neural Network (ANN), Ordered Probit and Multiple Discriminant Analysis (MDA) to local public finance variables in Mexico. In order to test the performance of each method in and out of the sample, the data is divided in two sub-samples, referred to hereafter as the "training set" and the "testing set" respectively.

4.1 Data Analysis and Variable Definitions.

As a first step, information related to State public finances in Mexico was collected from the database provided by FitchRatings, freely available through its web page on the public finances' section (www.fitchmexico.com). This information is preferred over other data sources such as INEGI (National Institute of Statistics and Geography, by its initials in Spanish) or SHCP (the Treasury, by its initials in Spanish) in order to keep consistency in data and because it is primary information, obtained directly by the agency from States Treasuries during the rating process.

Several financial ratios were calculated following definitions provided by García-Romo, et al. (2010) and some other variables proposed by FitchRatings in their public credit risk analysis—see Table 1. The final database contains 22 States, rated in Mexico by FitchRating from 2001 to 2007, for each of the 35 financial variables in every State, including credit ratings.

Table 2 shows basic descriptive statistics of all financial variables considered. Skeweness and kurtosis indicate that most variables do not seem to individually follow a normal distribution, which is confirmed by the values of normality statistics (Jarque-Bera and Shapiro Wilk). Also, there is an evident and strong dispersion among variables; the statistical ranges and variance are high. Hence in order to prevent unwanted effects we standardize all the explanatory variables used in this paper. It was noticed that this variability is heavily influenced by the inclusion of the two biggest states in Mexico: Mexico City Government and the State of Mexico Government. In addition, it is widely recognized the fact that Mexico City is the highest rating government (AAA), but mainly due to the full financial support of all credit issues by the federal government. We believe this last feature affects the estimation of credit risk, so Mexico City is dropped from our sample in the rest of the analysis.

In addition to the estimation process using all 34 financial variables, in order to avoid multicollinearity and also to provide a more parsimonious analysis, *ad hoc* with the small sample size, we employ Principal Components Analysis (PCA) as described in Mendoza (2010). This method reduces the 34 explanatory public finance variables in México to only six financial factors, accounting for more than 80% of the variation in the sample. We present both set of results in the following sections.

Table 1. Definition of State financial variables.

Name	Definition	Detailed Information
State Dimension		
IT	Total Revenue	Own income + Federal Income
IFOS	Ordinary tax income	Own incomes (taxes, rights of use, products, etc.) + federal and State federal shares, (non including municipal transfers) + other federal incomes as (Federal contributions, Branch 33 y others)
GPRI	Primary expenditure	Current expenditure, transfers, investment expenditure and ADEFAS.
GCR	Current Expenditure	Millions of Pesos of 2006.
AHOIN	Internal Savings	Total income minus primary expenditure
TRIB	Federal Participations / Total Transfers	Share of federal taxes collected in the state (%)
Income, Saving and Investment Generation		
IEIT	Own incomes/Total incomes	Own incomes (taxes, right of use, products, etc.) + Federal Shares to States + Fund for State strengthening (F-IV branch 33)
IEGO	Own incomes/Current Expenditure	Own Incomes by each Peso spent
INVI	Investment expenses/Own income	
INVB	Investement expenses/ Gross Domestic Product	
INVP	Investment expenses/Primary expenses	State investment (without Transfers from branch 33)/Own investment
AHOINIFO	Internal savings/IFOS	Primary balance minus interest payments
Ordinary Expenses		
GOIFO	Current Expenditure/IFOS	
GOTNEIFO	(Current Expenditure + Non labeled Transfers)/IFOS	
CORP	Current Expenditure/Primary expenses	Current Expenditure, Transfers, Total investments and debts from past Fiscal Years (ADEFAS)
Leverage		
DEU	Total Debt	
DAH	DEU/Internal Savings	
DPIB	DEU/PIB	Total Debt / GDP
DPAR	DEU/Federal income	
DIFOS	DEU/IFOS	Direct public debt from State Bodies + county public debt
DD	Direct debt	Millions of pesos of 2006
DIOD	Indirect debts of no centralized organizations	Millions of pesos of 2006
DIM	Indirect debts of counties	Millions of pesos of 2006
DDIFO	DD / IFO	Direct Debt / IFO
DDAI	DD /AI	Direct Debt/ Internal Savings
Sustainability of the Debt		
SDEU	Debt service	IFOS — operative expenses — non labeled transfers and others
SDEUAI	SDEU/Internal Savings	
SAHO	SDEU/IT-GPRI + INV	INV represents investment expenses
SPAR	SDEU/Federal income	Interest payment + debt amortizations
SIFOS	SDEU/IFOS	
Results		
BPRI	Primary Balance	Levels in millions of pesos of 2006
BFIN	Financial Balance	Levels in millions of pesos of 2006
PIB	Gross Domestic Product	Levels in millions of pesos of 2006
PIBPER	GDP Per capita	Levels in pesos of 2006

Source: García-Romo, et al. (2005) and Credit Analysis on Public Finances by FitchRatings.

Table 2. Descriptive Statistics of Mexican States Public Finances 2001-2007.

	Mean	Standard Deviation	Bias	Kurtosis	Max value	Min value	Normality
IT[a]	18,650	10,773	0.7982	0.1199	48,695	4,180	0.9421
IFOS	7,542	4,861	0.9003	0.0724	21,501	393	0.9156
GPRI	18,640	11,026	0.7913	0.1106	50,051	1,145	0.9406
GCR	3,690	2,694	0.5567	-0.8519	10,378	126	0.9031
AHOIN	1,777	1,225	1.0189	0.6485	5,571	0.0000	0.9102
TRIB	1.9519	0.4360	0.7343	1.6336	3.5658	1.0400	0.9472
IEIT	0.0673	0.0372	1.2265	1.3481	0.1909	0.0050	0.8729
INVI	1.7181	0.9882	1.8865	4.6434	6.0981	0.2831	0.8322
INVB	0.0131	0.0160	6.7553	56.5614	0.1599	0.0000	0.8084
INVP	0.1091	0.0634	2.6699	10.3917	0.4725	0.0157	0.8045
IEGO	0.3888	0.1883	0.9052	0.2587	0.9616	0.0906	0.9217
AHOINIFO	0.2458	0.0906	-0.4274	-0.0302	0.4440	0.0000	0.9736
GOIFO	0.4570	0.1302	-0.4374	-0.3444	0.7261	0.1200	0.9802**
GOTNEIFO	0.9709	0.5145	1.8915	2.2664	2.6904	0.5930	0.6416
CORP	0.1834	0.0627	0.0692	-0.0289	0.3332	0.0400	0.9864*
DEU	2,100	2,074	2.0000	4.0000	10,658	0.0000	0.7902
DAH	1.441	1.460	2.958	13.049	10.344	0.0000	0.7144
DPIB	0.022	0.075	8.118	67.849	0.708	0.0000	0.8064
DPAR	0.396	0.307	0.267	-0.193	1.181	-0.539	0.9438
DIFOS	6.575	72.466	11.53	132.998	836.010	0.0000	0.9594
DD	1,513.20	1,473.66	1.930	5.630	8,181.20	0.0000	0.8117
DIOD	401.91	738.44	2.500	6.770	4,310.50	0.0000	0.5942
DIM	145.91	232.83	2.010	3.810	1,121.88	0.0000	0.7173
DDIFO	0.2172	0.1484	0.3189	-0.4432	0.6063	0.0000	0.9699
DDAI	1.0317	1.0554	3.3378	17.3855	8.1009	0.0000	0.6998
SDEU	359.60	488.08	3.2200	14.3600	3528.67	0.0000	0.6524
SDEUAI	0.2290	0.2372	1.7458	3.1734	1.2037	0.0000	0.8194
SAHO	0.2586	0.3470	3.4530	15.9867	2.4861	0.0000	0.6417
SPAR	0.0595	0.0896	0.9336	6.7695	0.4921	-0.2641	0.8741
SIFOS	0.0515	0.0524	2.0684	5.4257	0.2918	0.0000	0.8160
BPRI	-68.7684	786.1661	-0.9761	2.6892	1,921.20	-3,003.10	0.9476
BFIN	-194.2367	788.6629	-1.1801	3.0117	1,604.90	-3,227.40	0.9552
PIB	2,140,141	11,309,570	7	59	107,092	1,188	0.1970
PIBPERC	6.0762	2.4007	0.5830	-0.4994	12.7854	2.6000	0.9385
Rating	6	1	-0.3810	0.3911	9	2	0.9893*

[a]See Table 1 for variable definitions.
*,** and *** Significant to 1%, 5% y 10% respectively.

4.2 Neural Network Implementation

The topology of the ANN described in section 3.1 was first trained using a sample consisting of 112 observations, those cases from 2001 to 2006. Each case has 34 input values corresponding to the explanatory financial variables presented in previous sections and one output value corresponding to the desired output (credit rating). This first sample data will be referred as the *'training set'* for this and the other two methods considered in this paper (i.e., ordered probit and discriminant analysis). In order to find the optimal number of neurons in a *hidden layer*, all possible networks containing 3 to 40 hidden nodes were trained and evaluated over a *"testing set",* with 21 allocations consisting of the cases not considered in the training set. This sample data is sometimes referred to as the 'out–of–sample' data.

In order to avoid overtraining the neural network, the training process was stopped when $E_{total} \leq 0.05$ (see equation 4 and algorithm in section 3.1.1). In the beginning of the network training, weights (w_i and b_j) were set to random values in the range [-0.01, 0.01]; also a coefficient $\eta = 0.01$ (see algorithm in section 3.1.1) and a scaling factor for the activation function $\lambda = 0.01$ (see equation 2) were used. These initial values for weights, learning coefficient and scaling factor respectively, were chosen using an experimental process, analyzing training results for different initial values, until a satisfactory error value was found. Figure 2 shows the percentage of correct classifications obtained by networks with different number of hidden nodes using the *Testing Sample* data, with the best performance obtained by a network with 10 hidden nodes. It must be noted that this "best" network contains just one out of an infinite number of possible sets of weights able to accurately approximate the training set with a precision of $E_{total} \leq 0.05$.[8]

[8] Several ANN applications in finance and economics do not report details of this exercise to determine the number of neurons in a hidden layer. Some others use heuristic procedures to choose the number of neurons *a priori*. It is important to insist that the aim of the ANN is to find the weights w_i and b_j that minimize E_{total}. This approach is somewhat different to the optimization and classification criteria used in the two competing methods.

4.2.1 Discussion on Neural Network Classification

This section shows the classification results obtained by the ANN. Table 3 details a classification matrix with a 100% of correct classification in the training set. As it was explained in section 3.1, a feedforward ANN with one hidden layer is able to fully approximate any arbitrary function, described by a finite set of samples, as in the case of this experiment. This ANN model may then accurately adjust its parameters by an iterative process, called training, to a function that adjusts perfectly to training data (Chen and Jain 1994, Haykin 1999).

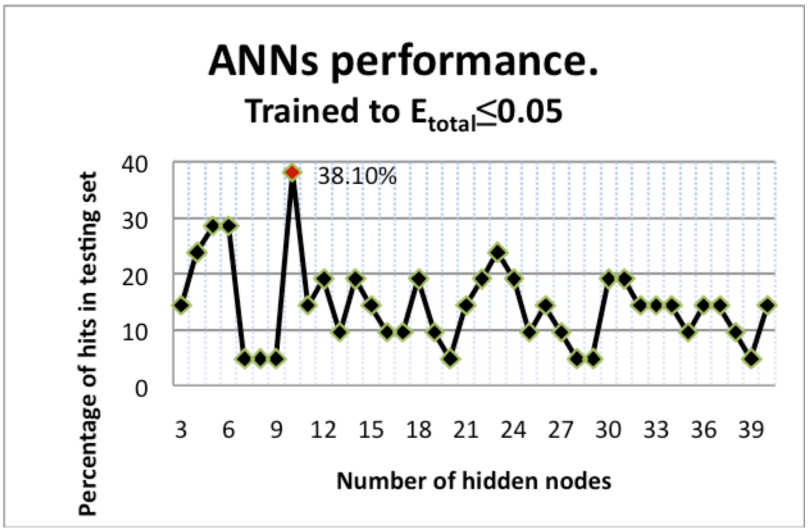

Figura 2. ANN performance in the *Testing Set* for different numbers of hidden nodes

It must be pointed out that this does not mean that the model accurately defines the system generating the data, but it just accurately adjusts to data representing the system introduced presented to the network during training. If some important information about the system was not presented in the training data, the ANN would not capture such information. This situation conveys the neural net designer to make an accurate analysis of the training set (Pullum et al. 2007) and to follow an empirical process to find the "best" network (Mayosky 2006). Section 4.5 shows the forecasting ability of this network to classify observations that are *not* part of the training sample.

Table 3. Classification Matrix: Artificial Neural Network (data in the Training Set).*

		True credit rating							
		AA+	AA	AA-	A+	A	A-	BBB+	BBB
Forecast	AA+	3	0	0	0	0	0	0	0
Credit	AA	0	7	0	0	0	0	0	0
Ratings	AA-	0	0	3	0	0	0	0	0
	A+	0	0	0	35	0	0	0	0
	A	0	0	0	0	26	0	0	0
	A-	0	0	0	0	0	28	0	0
	BBB+	0	0	0	0	0	0	6	0
	BBB	0	0	0	0	0	0	0	4
Total obs.		3	7	3	35	26	28	6	4
Correct Class.		3	7	3	35	26	28	6	4
Percentage		100%	100%	100%	100%	100%	100%	100%	100%

* ANN contains 10 hidden nodes, training to adjust at $E_{total} \leq 0.05$

4.3 Ordered Probit Model Implementation

Table 4 shows the classification matrix of the Ordered Probit using the training set with all 34 financial variables. Overall, out of the 112 observations in the training set, the model is able to correctly classify 58 observations, which is just slightly above a coin toss. The ratings reflect that the probability of classification is greater for ratings in A +/- 1 notch. It is evident that if we want to fully classify ratings in the training set, the neural network is a much better option than categorical models. It would be natural to expect that the performance of logit models, or improvements of these, would underperform the neural network in the training set.

The distribution of failures conveys interesting information about the accuracy of the method. In this case, from table 4 we observe that the Ordered Probit Model tends in general to underestimate credit ratings, that is, a greater proportion of forecasts lies below the actual ratings. In particular, for ratings greater than A+ the underestimation is more pronounced, while for ratings of A and lower, ordered probit tends to assign greater ratings than the actual ratings received. It is obvious that the classification ability of this method is very much less satisfactory than the one obtained by ANN.

Table 4. Classification Matrix: Ordered Probit Model (data in Training Set).

		True credit rating							
		AA+	AA	AA-	A+	A	A-	BBB+	BBB
Forecast	AA+	1	0	0	0	0	0	0	0
Credit Rating	AA	0	1	1	1	0	0	0	0
	AA-	0	0	0	0	0	0	0	0
	A+	2	6	3	23	7	1	0	0
	A	0	0	3	11	19	5	1	0
	A-	0	0	0	0	4	11	4	2
	BBB+	0	0	0	0	0	0	1	0
	BBB	0	0	0	0	0	3	0	2
Total of observations		3	7	7	35	30	20	6	4
Correct forecasts		1	1	0	23	19	11	1	2
Percentage		33.3%	14.3%	0.0%	65.7%	63.3%	55.0%	16.7%	50.0%

4.4 Multiple Discriminant Analysis

Table 5 below shows the results obtained for classifications of credit ratings using discriminant functions, also known as Multiple Discriminant Analysis (MDA). This method is widely used by practitioners to classify observations. As in the classification matrices presented before, successful classifications of credit ratings in the training set are shown in the main diagonal of the matrix. It can be noticed that the highest percentage of success in the training sample (correct classifications) is again obtained for credit ratings A+, A and A- (numbers 5, 6 and 7 respectively). From the distribution of failures (misclassifications), it is observed again the asymmetric behavior of classifications in the MDA algorithm as it tends to overestimate/underestimate credit ratings. This time however, MDA forecasts greater ratings than the true ratings originally assigned by FitchRatings on classifications lower than rating 6 (A). On the contrary, for ratings greater than 6 (A), MDA seems to assign ratings lower than the true assigned ratings. The classification ability of this method is less satisfactory than the one obtained by ANN, but much better than the one obtained using Ordered Probit models in the section before. Overall, MDA correctly forecasts 71.4% of the ratings in state governments, while Ordered Probit models forecast 51.8% of ratings.

Table 5. Classification Matrix: Multiple Discriminant Analysis (data in Training Set).

		True credit rating							
		AA+	AA	AA-	A+	A	A-	BBB+	BBB
Forecast	AA+	3	2	1	0	0	0	0	0
Credit	AA	0	5	1	4	5	1	0	0
Rating	AA-	0	0	1	4	0	0	0	0
	A+	0	0	0	23	1	0	0	0
	A	0	0	0	4	20	1	0	0
	A-	0	0	0	0	0	21	1	0
	BBB+	0	0	0	0	0	4	4	1
	BBB	0	0	0	0	0	1	1	3
Total of observations		3	7	3	35	26	28	6	4
Correct forecasts		3	5	1	23	20	21	4	3
Correct Percentage		100%	71.4%	33.0%	65.7%	76.9%	75.0%	66.7%	75.0%

4.5 Goodness of Fit and Forecasting Ability

As it was explained before, the forecasting ability and goodness of fit of the three methods presented here (Artificial Neural Networks, Ordered Probit and Discriminant Analysis) is tested dividing the overall sample in two parts. The first susbample, named "Training Set", is composed of 112 observations covering up to year 2006 and it is used to produce classifications within the sample as we have just examined. The second subsample is known as the "Testing Set" and contains data corresponding only to observations of the year 2007.

Results in the Training Set

Table 6 shows the percentage of successful classifications for each method and some additional statistics to evaluate the forecasting error. Focusing first on the *training set* (first panel of the table), it can be noticed that MDA correctly classifies 71.4% of data (with an Absolute Mean Error—AME—of 0.4554 notches). When we relax the forecasting criteria and allow classification forecasts to be valid within an interval of +/-1 notch, the percentage of correct hits using MDA increases importantly to 89.3% in the training set. Similarly, if the interval of analysis is +/-2 notches, correct classification increases to 94.6% and, for intervals of +/- three notches correct classification increases to 99.1%, an almost perfect classification.

The forecasting performance of Ordered Probit models for data in the *training set* shows a similar behavior. However, only 51.8% of correct point classification was obtained for data in the training set, which makes the forecasting ability of this method no better than a flip of a coin. The forecasting ability of Ordered Probit models however improves after we allow for interval classifications. In the training set, when there is a +/- one notch interval allowed, correct classification increases to 83.0% of the cases, 6.3 percentage points lower than MDA. If intervals increase to +/- two notches, the forecasting ability of Ordered Probit models continues to improve and we now observe correct classification of 98.2% of the cases. With three notches Ordered Probit models are able to fully classify credit ratings. All in all, the forecasting ability of ANN in the training set is best, both in terms of accuracy and variability, while MDA provides more successful classifications, both point and interval, than ordered probit models. This last method however reports a slightly lower variability than MDA.

Results in the Testing Set

The *Testing Set* shown in table 6 is the set where one is really able to evaluate the forecasting ability of the three methods. In this case Ordered Probit models are able to provide the best point classifications of the three methods considered, reporting a percentage of successful hits of 50.0%, compared to 38.1% of ANN and 28.6% of MDA respectively. Examining classifications by intervals, Ordered Probit models turn out to be the best forecasting alternative: they give greater accuracy with lower absolute errors. ANN provide a better classification rate than MDA (28.10% vs. 28.57%), but if we measure correct classifications by intervals, a lower rate of success of ANN is observed once it is compared with MDA. In terms of variability, Ordered Probit models are a better choice than MDA or ANN. In fact, ANN turns out to be the last choice given the greatest variability among the three alternatives considered.[9]

All in all, ANN is a better choice to classify ratings within the training set, while ordered probit models provide a better classification than the other two alternatives in the testing set.

[9] It must be pointed out that the results of the ANN reported here are the best results obtained by our network topology in the validation set. It should be noticed that the process of finding such a network is heuristic and hand-crafted, in the sense that there is no a theoretical way to ensure that the best possible initial values of parameters or number of nodes in a hidden layer, have been chosen (Mayosky 2006). Also, it must be considered that the process of finding a network with acceptable performance is computationally costly and time-expensive.

Table 6. Performance of Classification Methods for Training and Testing Sets using 34 financial variables.

Criteria	Discriminant Analysis	Ordered Probit	Neural Network*
% of hits in **TRAINING SET**	71.4%	51.8%	100%
% of hits in +/- 1 notch	89.3%	83.0%	100%
% of hits in +/- 2 notches	94.6%	98.2%	100%
% of hits in +/- 3 notches	99.1%	100.0%	100%
Mean Absolute Error	0.4554	0.2143	0.0
Relative Mean Absolute Error	0.0826	0.2231	0.0
Maximum Absolute Error	4	3	0
% of hits in **TESTING SET**	28.57%	50.0%	38.10%
% of hits in +/- 1 notch	61.90%	80.0%	61.90%
% of hits in +/- 2 notches	76.19%	88.5%	71.43%
% of hits in +/- 3 notches	95.24%	88.5%	76.19%
Mean Absolute Error	1.4300	1.0000	1.6191
Relative Mean Absolute Error	0.2549	0.1568	0.2892
Maximum Absolute Error	4	8	5

* Ten hidden nodes.

4.6 Principal Component Analysis

One important criticism to the results above is that the number of financial variables used as predictors is relatively high compared to the small sample size. Also, it would be logic to expect that many of the predictors are strongly associated with one and other, affecting the final classification performance of each method.

Table A.2 in the appendix presents the correlation matrix of the variables employed in this study. Due to the evident linear association among several variables it would be of interest to reduce the number of predictors using multivariate methods. In fact, the linear association of explanatory variables (multicollinearity) justifies the use of principal components and factorial analysis to reduce the number of variables and avoid multicollinearity.

Employing factorial analysis we found six factors explaining the behavior of State public finances in Mexico.[10] Table 7 shows the six linearly independent factors together with their economic interpretation, the expected impact on the probability of

[10] For a detailed description of principal component analysis and factor analysis with an application to public estate finances in Mexico see Mendoza (2010). This subsection is based on that paper.

obtaining the highest credit rating (relevant for probit models and MDA) and the individual explained variation of the data.[11]

Table 7. Factors Describing State Public Finances in Mexico.

Factor [a]	Economic Interpretation	Impact[b]	Variance
1. Dimension	Measures the dimension of the State. It is composed of variables in levels such as total income, expenses, debt, etc.	(?)[c]	28.4%
2. Debt Sustainability (debt service)	Measures the ability of a given State to service its debt. It is composed of financial ratios associating debt service with fiscal income, share transfers, internal savings among others.	(+)[d]	14.9%
3. Leverage	Measures the leverage of a given State. It is composed of debt variables in levels and financial ratios of debt.	(+)	15.5%
4. Current Expense	Captures the propensity of a given government to spending or saving. The greater the score of this factor the greater the propensity to save as opposed to saving and viceversa.	(+)	8.8%
5. Results	Captures the tradeoff between maintaining a balanced budget and investment. The greater the score of this factor the better in terms of balanced budget (or surplus) but less the investment realized, and viceversa.	(-)	6.0%
6. Investment	This factor captures the propensity to investment by the State. It is composed of investment financial ratios.	(-)	7.2%

[a] The name of the factor is assigned according to the variables contained in a given factor. For more details and economic intuition of these factors the reader is kindly referred to Mendoza (2010). [b] It refers to the impact on the probability to obtain the highest credit rate. [c] Undefined impact. [d] The higher the leverage, the lower the probability to receive the highest credit rating—see scale defined at table A.1.

Table 8 below shows the results of the three methods using six factors as predictors.[12] For the training set the ANN provides once more the best results in terms of prediction and variability. Ordered Probit is the second best and the last choice is MDA. This time however, in contrast with the ANN experiments using all financial variables, several networks did not classify all the observations correctly. The last column in Table 8 for instance shows the results of a network that was not able to fully classify ratings in the training set, but on the other side reported the greatest accuracy in the testing set.

In the testing set obtained from six financial factors—and in clear contrast with the case of all financial predictors included—, MDA is the method with the best point classification rate (41.1%), better than Ordered Probit or ANN with 28.57%, 23.81% and

[11] We do not show here the size and direction of the estimated parameters, factor loadings or marginal effects to save space. For details on these factors refer to Mendoza (2010). Detailed estimation results are readily available from authors.

[12] We do not show classification matrices of these results but they are available from authors upon request.

28.57% respectively. However, allowing for interval classification shows a better performance of Ordered Probit models, becoming the best choice among the three alternatives considered. Ordered Probit models also provide in general the lowest variability.

Table 8. Performance of Classification Methods for Training and Testing Sets using six financial factors.

Criteria	Discriminant Analysis	Ordered Probit	Neural Network 37 nodes*	Neural Network 5 nodes**
% of hits in **TRAINING SET**	41.10%	50.04%	100.00%	57.14%
% of hits in +/- 1 notch	61.60%	80.18%	100.00%	93.75%
% of hits in +/- 2 notches	80.40%	97.29%	100.00%	100%
% of hits in +/- 3 notches	98.00%	100.00%	100.00%	100%
Mean Absolute Error	1.1964	0.7411	0.0	0.4900
Relative Mean Absolute Error	0.2053	0.1660	0.0	0.1080
Maximum Absolute Error	5	3	0	2
% of hits in **TESTING SET**	41.10%	28.57%	23.81%	28.57%
% of hits in +/- 1 notch	47.60%	80.09%	28.57%	47.62%
% of hits in +/- 2 notches	61.90%	95.23%	33.33%	85.71%
% of hits in +/- 3 notches	76.20%	100.0%	38.09%	90.47
Mean Absolute Error	2.1429	1.1429	6.6670	1.619
Relative Mean Absolute Error	0.4286	0.2656	1.4370	0.3960
Maximum Absolute Error	7	4	20	7

* Thirty seven hidden nodes. **Five hidden nodes.

5. Conclusions

Credit risk ratings have become an important input in the process of improving transparency of public finances in local governments and also in the evaluation of credit quality of state and municipal governments in Mexico. Although they have recently been subjected to heavy criticism, credit ratings are indicators still widely used by regulators and banks to monitor financial performance in stable and volatile periods.

In this work we have compared and evaluated the performance of three forecasting methods frequently used by practitioners to estimate credit ratings of local governments in Mexico. Financial data and credit ratings provided by FitchRatings were used to define explanatory variables in the estimation of Ordered Probit models (OP) and Multiple Discriminant Analysis (MDA) and also, as input variables for Artificial Neural Networks (ANN). We have also compared the performance of the three methods using 34 variables as predictors and also, in order to account for potential multicollinearity, using only six factorial components accounting for more than 80% of the data variation.

It was found in general that ANN provides better point forecasts than OP and MDA in the training sets, that is, in classifications within the estimation sample. In contrast, MDA is the best choice when 34 financial variables are used as predictors, while OP is the best alternative when only six factors are considered (also when examining training sets).

In the testing set however we observe that OP is a better choice than ANN or MDA when the whole financial variables are considered and MDA is the best alternative when only six factors are considered. In other words, it seems that OP's point classification performance is better with extended models, while MDA's performance is best with parsimonious models. In general, OP improves substantially when interval classification is allowed and, in fact, for all interval cases considered in the testing sets, OP is the best choice for practitioners wanting to classify state credit ratings.

All in all, if a finance practitioner aims at forecasting credit ratings with a small sample size (as the one for local governments in Mexico), her best choice in terms of computational cost, variability and forecasting ability, are Ordered Probit Models.

References

Bharat and Barin (1997). "Performance Evaluation of Neural Network Decision Models", Journal of Management Information Systems", Vol. 14, Issue 2, pp. 201-216.

Benell, Julia A., David Crabe, Stephen Thomas and Owain ap Gwilym (2006). "Modelling sovereign credit ratings: Neural networks versus Ordered Probit," Expert Systems with Applications, 30, pp. 415-425.

Chen, David S.,Ramesh C. Jain (1994) "A Robust Back Propagation Learning Algorithm for Function Approximation," IEEE Transactions on Neural Networks, Vol. 5, No. 3, pp.467-479.

Cybenko, G. 1989 "Approximation by superposition of a sigmoidal function," Mathematics of Control, Signals and Systems, vol. 2 pp. 303-314

Desai et al., 1996 V.S. Desai, J.N. Crook and G.A. Overstreet Jr., "A comparison of neural networks and linear scoring models in the credit union environment", European Journal of Operational Research **95** (1) (1996), pp. 24–37 22.

García-Romo, G., Ibarra-Salazar, J. and Sotres-Cervantes, L. (2010). "Determinants of Mexican States Government Credit Ratings", in Mendoza (2010), *Las Finanzas Públicas Estatales en México, Herramientas de Diagnóstico y Respuesta en un Entorno de Crisis*, Editorial Porrúa, por publicarse, Septiembre.

Gómez-Gil, P. (2009) Tutorial: An introduction to the use of Artificial Neural Networks. Draft available at
[url:http://ccc.inaoep.mx/~pgomez/cursos/redes%20neuronales%20 artificiales/index_archivos/Page530.html]

Gutiérrez Nieto, B.; Serrano Cinca, C. (2007). “Factors explaining the rating of Microfinance Institutions”, Nonprofit & Voluntary Sector Quarterly, *forthcoming*.

Haykin, S. 1999. Neural Networks, Second Edition. Upper Saddle River, NJ: Prentice-Hall.

Huang, Shian-Chang (2009) “Integrating nonlinear graph based dimensionality reduction schemes with SVMs for credit rating forecasting.” Expert Systems with Applications, 36, 7515-7518.

Kotsiantis, S., E. Koumanakos, D. Tzepelis and V. Tampakas (2006). “Financial Application of Neural Networks: Two Case Studies in Greece”. Lecture Notes in Computer Science, Artificial Neural Networks ICCAN 2006, part II, pp.672-681.

Kumar, Kuldeep and Sukato Bhattacharya (2006). “Artificial Neural Network vs. Linear Discriminant Analysis in Credit Rating Forecasts”. Review of Accounting and Finance, Vol. 5, No. 3, pp. 216-227.

Lee, Young-Chan (2007). “Application of support vector machines to corporate credit rating prediction”. Expert Systems with Applications, 33, 67-74.

Light, W. (1992) “Ridge functions, sigmoidal functions and neural networks” in E.W. Cheney, C.K. Chui and L.L. Schumaker, eds. Approximation Theory VII, Boston:Academic Press. pp 163-206.

Mandic, Danilo P. and Jonathon A. Chambers (2001). Recurrent Neural Networks for Prediction. Chinchester: Wiley and Sons.

Markham, I.S. and C.T. Ragsdale (1995). “Combining Neural Networks and Statistical Predictions to Solve the classification problem in Discriminant Analysis” Decision Sciences, 26, p. 229.

Mayosky, MA (2006). “Introducción a las Redes Neuronales” pp. 101-131 in: Aprendizaje Automático: Conceptos Básicos y Avanzados. Edited by B.Sierra Araujo Madrid:Pearson Prentice Hall.

Mendoza-Velazquez, Alfonso (2009). “The Information Content and Redistribution Effects of State and Municipal Rating Changes in Mexico”. Economics Discussion Papers, No 2009-17. Available at http://www.economics-ejournal.org/economics/discussionpapers/2009-17.

Mendoza-Velazquez, Alfonso (2010). “Indicators of Performance, Pressure and Vulnerability of Public Finance in the States of México”, El Trimestre Económico, forthcoming, Vol. 307, Jul-Sep.

Patuwo, M.Y. Hu and M.S. Hung (1993). “Two-group classification using neural networks”, Decision Sciences, 24 p. 825.

Pullum, L.L. Taylor, B. J and Darrah, M.A. (2007) Guidance for the Verification and Validation of Neural Networks. IEEE Computer Society, Hoboken:John Wiley.

Rosch, Daniel (2004). “An Empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies. International Journal of Forecasting, Vol. 21, No. 1, (January-March 2005), pp. 37-51

Rumelhart, D.E., G. E. Hinton, and R.J. Williams, 1986. “Learning internal representations by error propagation,” in D. E. Rumelhart and J.L. McCleland, eds. Vol 1, Chapter 8, Cambridge, MA: MIT Press.

Salchenberger, L.M., E.M. Cinar and N.A. Lash (1992), “Neural networks: A new tool for predicting thrift failures”, Decision Sciences 23, pp. 899–916

Stephen Figlewski, Halina Frydman and Weijian Liang (2008). “Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions”, Default Risk.com, available at: URL: http://www.defaultrisk.com/pp_model146.htm [Last accessed at May 2009]

Ting Peng Liang, Herbert Moskowitz and Yuehwern Yih (1992). "Integrating Neural Networks and Semi-Markov Processes for Automated Knowledge Acquisition: An Application to Real-time Scheduling", Decision Sciences, Vol. 23, issue 6, pp. 1297-1314.

Tokinaga S., J. Lu and Y. Ikeda (2005) "Neural network rule extraction by Using Genetic Programming and its applications to exploratory classifications", IEICE, Trans. Fundamentals, Vol. E88-A, No 20, pp. 2627-2635.

Werbos, P.J. 1990. "Backpropagation through time: What it does and how to do it." Proceedings of the IEEE, Vol. 78, pp. 1550-1560.

Werbos, P.J. (1994). The Roots of Backpropagation. From ordered derivatives to Neural Netwroks and Political Forecasting. New York:Wiley Interscience.

Wong B.K. and Y. Selvi, (1998) Neural network applications in finance: A review and analysis of literature (1990–1996), Information and Management **34** (3), pp. 129–139.

Wong B.K., T.A. Bodnovich and Y. Selvi, (1997). "Neural network applications in business: A review and analysis of the literature (1988–1995), Decision Support Systems **19**, pp. 301–320

Wood, Douglas and Dasgupta, Bhaskar (2006). "An innovative tool for financial decision making: The case of Artificial Neural Networks", Creativity and Information Management 4, No 3, pp.172-183.

Wooldridge, Jeffrey (2001). "Introductory Econometrics: A modern Approach", Southwestern College Pub., U.S.

Yorio, Gabriel (2006). "Calificaciones de riesgo crediticio de los gobiernos estatales en México". Tesis Maestría en Economía. El Colegio de México.

Appendix

Table A.1. Ratings Scale used by FitchRatings

Scale	Category	Definition
AAA	1	Highest credit rating quality
AA+[a]	2	Very high credit rating quality
AA	3	
AA-	4	
A+	5	High credit rating quality
A	6	
A-	7	
BBB+	8	Fair credit rating quality
BBB	9	
BBB-	10	
BB+	11	Speculative
BB	12	
BB-	13	
B+	14	Highly speculative
B	15	
B-	16	
CCC	17	High risk for non compliance
CC	18	Very high risk for non compliance
C	19	Highest risk for non compliance
D	20	Non compliance
E	21	Credit Rating Suspended

Source: Created using information provided by Fitch Ratings. [a] Signs + and – indicates strength or relative position into ratings. For federal entities ratings oscillate from B a AA.