REVISTA DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA (13). Páginas 3–23.
Junio de 2012. ISSN: 1886-516X. D.L: SE-2927-06.
URL: http://www.upo.es/RevMetCuant/art.php?id=56

Extreme Learning Machine to Analyze the Level of Default in Spanish Deposit Institutions

Montero-Romero, Teresa
Department of Management and Quantitative Methods, ETEA, Córdoba (Spain)
Correo electrónico: tmontero@etea.com

López-Martín, María del Carmen
Department of Economics, Legal Sciences and Sociology, ETEA, Córdoba (Spain)
Correo electrónico: mclopez@etea.com

Becerra-Alonso, David
Department of Management and Quantitative Methods, ETEA, Córdoba (Spain)
Correo electrónico: dbecerra@etea.com

Martínez-Estudillo, Francisco José
Department of Management and Quantitative Methods, ETEA, Córdoba (Spain)
Correo electrónico: fjmestud@etea.com

ABSTRACT

The level of default in financial institutions is a key piece of information in the activity of these organizations and reveals their level of risk. This in turn explains the growing attention given to variables of this kind, during the crisis of these last years. This paper presents a method to estimate the default rate using the non-linear model defined by standard Multilayer Perceptron (MLP) neural networks trained with a novel methodology called Extreme Learning Machine (ELM). The experimental results are promising, and show a good performance when comparing the MLP model trained with the Leverberg-Marquard algorithm.

Keywords: level of default; financial institutions; neural networks; Extreme Learning Machine.
JEL classification: G21; G01; C45.
MSC2010: 62P05; 68T01; 82C32.

Artículo recibido el 26 de julio de 2011 y aceptado el 23 de diciembre de 2011.

Análisis de la morosidad de las entidades financieras españolas mediante Extreme Learning Machine

RESUMEN

La morosidad en las entidades financieras es un dato muy importante de la actividad de estas instituciones pues permite conocer el nivel de riesgo asumido por éstas. Esto a su vez explica la creciente atención otorgada a esta variable, especialmente en los últimos años de crisis.

Este artículo presenta un método para estimar el nivel de la tasa de morosidad por medio de un modelo no lineal definido por la red neuronal *Multilayer Perceptron* (MLP) entrenada con una nueva metodología llamada *Extreme Learning Machine* (ELM). Los resultados experimentales son prometedores, mostrando un buen resultado si se compara con el modelo MLP entrenado con el algoritmo de Leverberg-Marquard.

Palabras clave: nivel de morosidad; instituciones financieras; redes neuronales; Extreme Learning Machine.
Clasificación JEL: G21; G01; C45.
MSC2010: 62P05; 68T01; 82C32.

1. INTRODUCTION

The aim of this paper is to study the default rate in financial institutions through variables that define the entities' financial state. More specifically, we try to analyze how certain financial data (in the form of independent variables) is related to the level of default in each institution. The relevance of this study is justified by the importance that this last variable has in the context of today's economic crisis.

Linear economic models are not able to capture non-linear patterns. Moreover, linear techniques cannot distinguish effectively between random noise and non-linear relationships, and usually make assumptions about the independence and normality of data. In the last few years, artificial neural networks (ANNs) have emerged as a powerful statistical modeling technique (Bishop, 1995). ANNs are useful to detect the underlying functional relationships within a dataset and to perform several tasks like pattern recognition, classification, modeling, and prediction. The most popular neural network model may be the Multilayer Perceptron (MLP) (Bishop, 1995), trained with the well-known back-propagation algorithm, due to its simple architecture. Besides, a number of recent papers use ANNs to analyze traditional problems in Accounting and Finance (Coakley, 2000; Herbrick, 2000; McNelis, 2005; Parisi, 2006; Martínez-Estudillo *et al*., 2007; Gutiérrez *et al*., 2009; López-Martín *et al*., 2011).

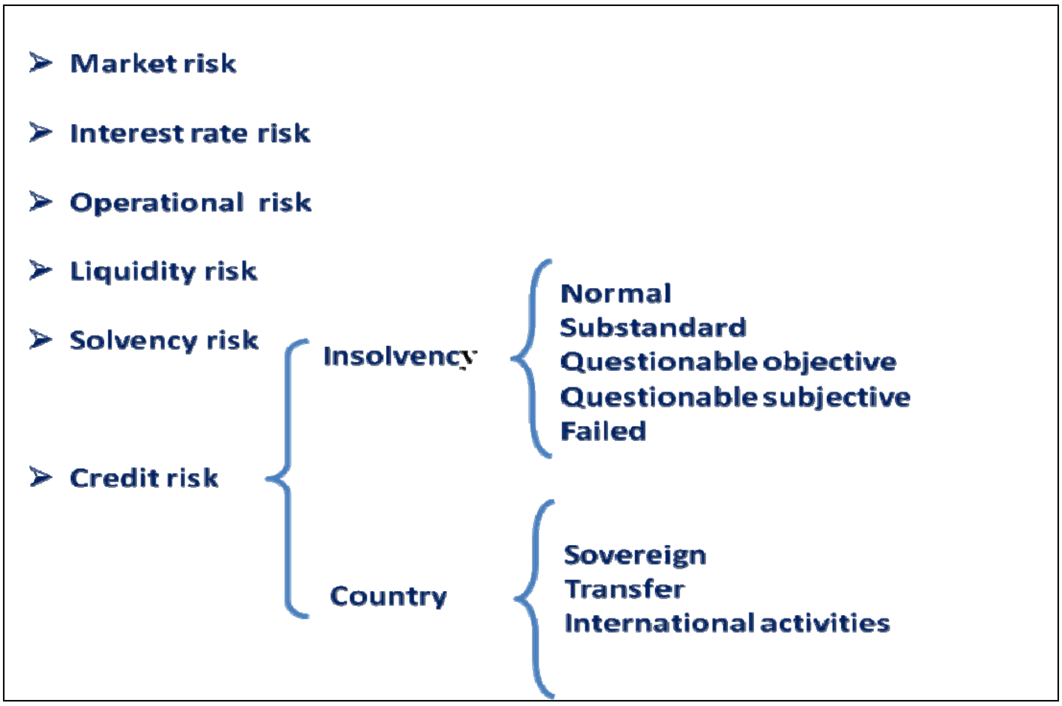
Taking all the above into account and as part of a greater research topic, let us introduce the main elements. The financial institutions studied in this study are Spanish deposit institutions with data since 2008. The methodology used is the non-linear model defined by MLP neural network trained with the Extreme Learning Machine (ELM) algorithm (Huang, 2004). ELM for MLP neural networks is a new algorithm that randomly chooses hidden nodes and analytically determines the output weights of the network. Theoretically, the ELM algorithm tends to provide a good generalization performance at an extremely fast learning speed. The experimental results based on artificial and real benchmarking problems show that ELM can result in a better generalization performance in many cases and can learn thousands of times faster than traditional learning algorithms for feed-forward neural networks (Huang *et al*., 2011).

The structure of the rest of the paper is the following: Sections 2 and 3 analyze the level of default in Spanish deposit institutions and the importance of this variable in the current economic context. Section 4 shows the state of the art with respect to studies using the ELM algorithm in the field of Business and Finance. Section 5 describes the mathematical model defined by the MLP neural network used in the prediction task and the ELM algorithm. Then Section 6 presents the experimental application where the performance of the proposed approach is tested. Finally, the paper closes with some remarks and conclusions.

2. CREDIT RISK AND DEFAULT PAYMENT IN THE SPANISH FINANCIAL SYSTEM

The uncertainty characterising the present economic context has turned risk into a critical variable that must be kept under control. It is known that financial institutions face different kinds of risk, including market risk, credit risk, liquidity risk, and operational risk. Figure 1 shows various types of risk undergone by financial institutions. The main categories are defined in the following figure.

Figure 1. Financial institutions exposure risk



Source: elaborated by authors.

Market risk is the potential loss on an investment due to adverse changes in the factors affecting price and/or value. Interest rate risk is the potential loss from adverse changes affecting the interest rate. Operational risk shows potential loss as a result of failures in internal processes or potential loss when these internal processes are inadequate. The liquidity risk is the risk of not being able to turn an asset into cash. Solvency risk is the risk of debts not met at maturity. It refers to a position in the long term. And finally, credit risk is the potential loss assumed by the financial institution as a consequence of breach of contract obligations to counterparts.

One of the most important kinds of risk and the most studied in the field of financial institutions is credit risk; see some examples: Rodríguez Fernández, 1987; Cruz González, 1998; López and Saidenberg, 2000; Boal Velasco and González Sánchez, 2001; Soler and Miró, 2001; García Céspedes, 2005, among other studies.

This risk is the object of our study, i.e. the risk of not recovering money that has been lent, along with its corresponding interest. This work is focused on credit risk, not only because it can be the cause of great losses in financial institutions, but also because it has already caused interventions by *Banco de España* in the last two crisis-ridden years. It also indirectly influences the capital requirements demanded by Basel III regulations, as well as the recent Law 2/2011 for the reinforcement of financial institutions.

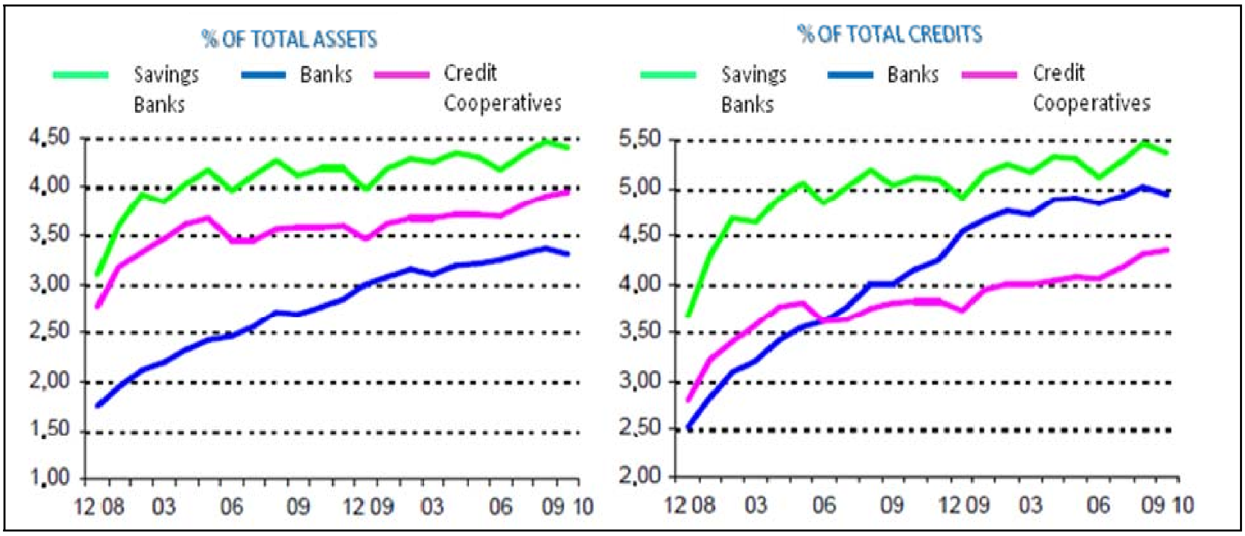
The tensions tackled by the national and international financial systems since the beginning of the crisis have obstructed its function as an essential driver of credit in the economy. In the Spanish scenario, credit institutions have undergone great difficulties in obtaining financial support. In turn, whatever financial support was granted came accompanied by deteriorated assets (mostly associated to the housing sector). Many loans were labeled as doubtful, and default rates reached surprising levels before the beginning of 2007. These problems have resulted in increasing difficulties for families and small companies to access credit, which is fundamental in order to generate jobs in the Spanish business network.

In this context, the crisis has emphasized the need for a clean, competitive and solid financial system that allows for mediation between those who own the financial resources and those who need them. A well-structured financial system is the main warranty for a country’s productive economy to be able to access the financial support needed for investment projects that generate wealth and jobs. However, in order for this to take place, it is necessary to trust in the integrity of institutions and proper performance of the markets.

The data gathered over the last few years confirm these problems. Figures 2 and 3 show how default rates in deposit institutions have risen over the last three years (2008-2010). These are seen to be higher in savings banks, than in banks and credit cooperatives. These latter two differ depending on whether rates are considered in terms of total assets or in terms of approved loans. Although the level of default has increased since 2007, it is also true that in the last two years there has been a dramatic decrease in the number of these three institutions’ doubtful loans since their peak in late 2008 and early 2009. This phenomenon is more related to the greater number of requisites prior to credit approvals, than to the solvency of clients in general. This explains the constant and even the reduced outstanding balance of credits granted by institutions.

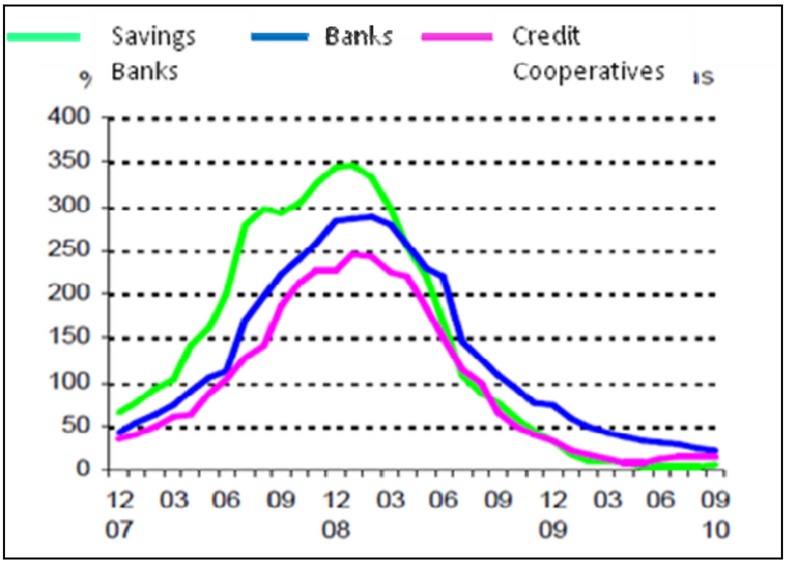
The data shown and the context in which we place this study prove how important it is for deposit institutions to have plenty of capital in order to be prepared for unexpected losses, such as those in times of decreased or moderately increased economic activity. This is one of the reasons that explain the capital regulation reforms that took place in the Basel Committee (Basel III), to increase demands to maintain resources, especially for those institutions that have a greater capacity to absorb losses. Minimum solvency requisites in Spain were complemented at the beginning of this year with a new capital coefficient (that is regulated by 2/2011 Law, of Financial Institution Reinforcement), that in practice is even more demanding than the one imposed by Basel III, in order to dissipate any doubts that financial markets may have about the health of our financial system.

Figure 2. Level of default in institutions in terms of total assets and total credit



Source: CECA (2010).

Figure 3. Level of default variations from year to year in deposit entities



Source: CECA (2010).

3. INSTITUTIONS STUDIED, SOURCES, AND EXPERIMENTAL DESIGN

3.1. Institutions Analysed and Study Stages

As indicated in the introduction to this paper, the aim of this work is to estimate the level of default in Spanish deposit institutions using variables that define each entity's financial information. In order to do this, we start by studying data from 2008. This year is chosen since it is the first term in which, despite the already growing level of default, financial reports of deposit institutions still did not reflect the real magnitude of the level of default caused by the economic crisis. It is thus a term not too far from the transition to a recognizable state of crisis, and one where the level of default in financial institutions was still representative.

As we know, institutions that can approve credits or similar measures constitute what is called the group of credit institutions. Even so, this study has not taken this group into consideration, but instead that made up of deposit institutions, which account for more than 90% of the balance of all credit institutions. Thus, there is a deliberate elimination of some credit institutions that do not have the same kind of activity (such as specialized credit institutions) or those that have different features (such as the government's credit institution). With this criterion, the total number of deposit institutions available with annual accountings was 152. Out of these the level of default has been obtained for 111 of them (73.03% of the total). Table 1 shows distributions by types of institution, and the average level of default of each type.

Table 1. Sample distribution classified by the kind of institution, percentage of each type, and default payment average per institution in 2008

Institutions	# institutions	Percentage of each type	Default payment average
Banks	31	27.93%	2.86%
Savings banks	46	41.44%	3.06%
Credit cooperatives	34	30.30%	1.81%
TOTAL	**111**	**100.00%**	**2.29%**

Source: elaborated by authors.

3.2. The Dependent Variable: Financial Institutions' Level of Default

We can define credit risk as the risk of not recovering the money lent with its corresponding interest. Within this kind of risk, special attention is paid to insolvency risk, which refers to refund difficulties, either due to the level of default, or for subjective reasons (such as reasonable doubts). Then, within insolvency risk, there is normal risk, substandard, questionable objective, questionable subjective, and failed risk. The questionable objective risk has an objective component: nonpayment. This includes all of the debt instruments, regardless of the client and the warranty, which are overdue either due to the main loan, interests or other expenses agreed upon in the contract, for more than 3 months and unless classified as failed. It also includes the risks (normally endorsements or third party involvements) in which the endorsed is in default payment. In questionable risk, operations are included for reasons other than default payment (whether overdue or not) whenever reasonable doubts over refund exist. This can be due to negative wealth, continued losses, delayed payments, or inadequate economic or financial structure.

Clearly, in strict terms every client in default is questionable, although not every questionable client is in default payment. This is because there is the possibility of classifying a risk as questionable for subjective reasons other than nonpayment. Still, in general, this distinction is not made when referring to “default payment”, and thus, in order to determine the default payment rate, the total of doubtful loans and default payers have been considered in all the loans. Thus, the level of default obtained in this way has been chosen as the identifying variable for credit risk for each institution. The information on each variable in 2008 has been obtained from reports on each of the institutions analysed.

3.3. Independent Variables in the Classification Model

The selection of independent variables has been carried out by carefully defining the elements in the financial statements, with prior identification of the operations generating default payment in deposit institutions (such as credits given by the institution), or those that indicate the existence of default payment (such as loans available for sale).

The information referring to these variables has been extracted from the balance sheets and the profit and losses accounts published by the institutions. Such financial statements have been obtained from *Confederación de Cajas de Ahorros* (Federation of Savings Banks), the *Asociación Española de Banca* (Spanish Banking Association), and the *Unión Nacional de Cooperativas de Crédito* (National Union of Credit Cooperatives).

The independent variables selected will be grouped into four categories: the kind of institution, assets, liabilities, and profit and loss, all expressed in terms of the percentage of the total balance.

The process of including variables in this model entails a previous selection based on discriminant analysis. From the preliminary reports of the financial states that could be related to the level of default, as mentioned above, the co-linearity between them was analysed using discriminant analysis. The independent variables chosen after that are included in Table 2.

Table 2. Independent variables

Aggrupation	Denomination	Variable type
Kind of institution	Institution credit cooperative	Binary
	Institution saving bank	Binary
Assets variables	Total holdings of securities (to deal)	Quantitative
	Holdings of securities (to deal): customers credits	Quantitative
	Total other financial assets at fair value (with changes in profits and loss)	Quantitative
	Other financial assets at fair value (with changes in profits and loss): credits to customers	Quantitative
	Financial assets on sale	Quantitative
	Loans (total)	Quantitative
	Loans: customers credits	Quantitative
	Loans: securities other than shares	Quantitative
	Loans: securities held in pawn	Quantitative
	Non-current assets available for sale	Quantitative
	Shares	Quantitative
	Real estate investment	Quantitative
Liability variables	Others liabilities at fair value (with changes in profits and loss)	Quantitative
	Derivatives	Quantitative
	Allowances	Quantitative
	Capital and reserves	Quantitative
	Contingent risks	Quantitative
	Contingent commitment	Quantitative
Profit and loss variables	Net interest income	Quantitative
	Return on equity instruments and non-interest income	Quantitative
	Provisioning expense (net)	Quantitative
	Operating profit	Quantitative
	Financial assets impairment losses (net)	Quantitative
	Profit (losses) in disposal of others assets not classified in non-current assets available for sale	Quantitative
	Losses in business combination	Quantitative
	Profit (losses) in non-current assets available for sale not classified as discontinued operations	Quantitative

Source: elaborated by authors.

3.4. Experimental Design

A database including 28 independent variables and 111 financial institutions is built, along with their corresponding level of default. Nominal variables have been turned into binary, one for each category, as shown in Table 2. In order to study the generalization of this model the database was split into two subsets. One will be used to train our neural network, and the other will be used to test it. There are 73 institutions assigned to the training set, and 38 to test. In order to make the partition, subsets were chosen that included extreme cases in both training and test sets, in order to improve generalisation.

4. EXTREME LEARNING MACHINES IN FINANCIAL PROBLEMS

Over the last decade, computational intelligence techniques have been used in a wide range of applications (marketing and sales, risk assessment and accounting, manufacturing, finance, etc.). Of numerous computational intelligence techniques, artificial neural networks (ANNs) have been playing a dominant role. However, it is known that ANNs face some challenging issues such as slow learning speed and/or poor computational scalability. ELM, as an emergent technology that overcomes some challenges faced by other techniques, has recently attracted the attention of computational intelligence and machine learning communities, in both theory and applications.

This section shows "the state of the art" of studies that use the ELM algorithm to solve problems in the field of business and finance. It must be highlighted that the number of studies in this area today is limited. The following are the most relevant from our point of view:

a) In Li *et al*., 2009, the authors apply the ELM classifier to the field of credit scoring. They compare two ELM algorithms with two well-known classifiers, namely, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), obtaining better results over two credit data sets from the University of California, Irvine (UCI).

b) ELM has also been used in forecasting to overcome the drawbacks of existing neural network forecasting models. In Wong and Guo, 2010, a hybrid intelligent model combines a novel meta-heuristic optimisation technique, the harmony search (HS) algorithm (Mahdavi *et al.*, 2007), with ELM to construct a learning algorithm to obtain optimal neural network weights and achieve a better generalisation performance. The model is developed to tackle sales forecasting problems in the fashion retail supply chain.

c) In Van Heeswijk *et al.*, 2009, the authors investigate the application of adaptive ensemble models of ELMs to the problem of one-step-ahead prediction in non-stationary time series. The proposed model achieves a test error comparable to the best methods with a low computational cost.

d) Another work (Sorjamaa *et al*., 2008) proposes a hybrid model that combines ELM with various intelligent heuristics such as the variable selection using the Partial Least Squares (PLS) and a projection based on Nonparametric Noise Estimation (NNE), to ensure proper results by the ELM method. Then, after the network is first created using the original ELM, the selection of the most relevant nodes is carried out using a Least Angle Regression (LARS) ranking of the nodes and a Leave-One-Out estimation of the performances, leading to an Optimally-Pruned ELM (OP-ELM). Finally, the prediction accuracy of the global methodology is demonstrated using the ESTSP 2008 Competition and Poland Electricity Load datasets.

e) Finally, in a previous work (López-Martín *et al*., 2011), we addressed the problem from the perspective of classification. This paper uses the ELM model for the classification of the default status of financial institutions. The technique proposed is shown to perform better than logistic regression for this problem.

This is, to the authors' knowledge, the first paper that applies the ELM algorithm to make a prediction about the level of default in deposit institutions from available financial information. We believe that the good performance obtained so far by the ELM algorithm in benchmark datasets will increase its use in applications on real problems in the field of finance and economics.

5. MATHEMATICAL MODELS

This section starts with a brief introduction to the MLP neural network model and regression with neural networks. Then the ELM algorithm used to estimate the neural network weights is presented in detail.

5.1. MLP Neural Network Model

Let us consider the training set given by N patterns $D = \left\{ \left(\mathbf{x}_i, t_i \right) : \mathbf{x}_i \in \mathbb{R}^n, t_i \in \mathbb{R}, i = 1, 2, \ldots, N \right\}$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in}) \in \mathbb{R}^n$, represents the input and t_i is the output for each one of these input vectors. The model used in order to do regression here is given by the function:

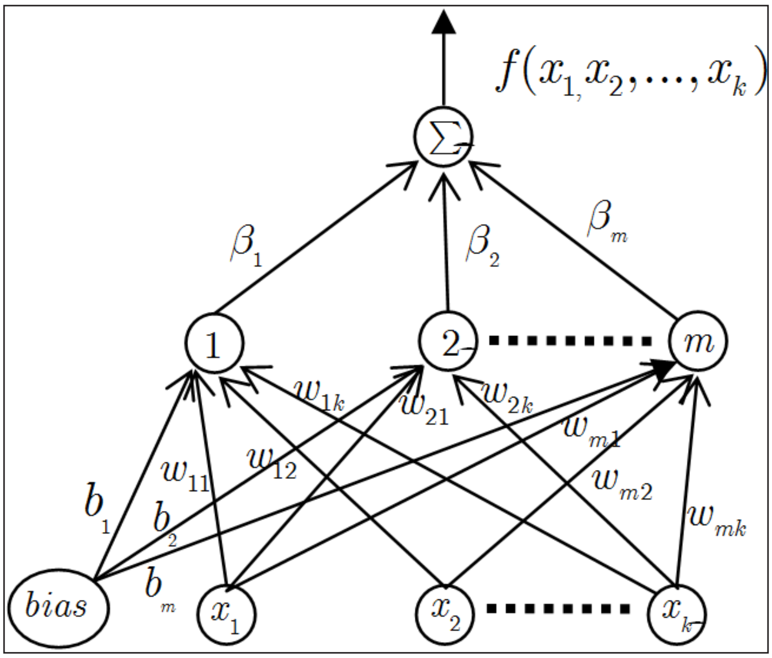
$$f(\mathbf{x}) = \sum_{i=1}^{M} \beta_i \, g(\langle \mathbf{w}_i, \mathbf{x} \rangle + b_i),$$

where the function g is the sigmoidal function given by:

$$g(t) = \frac{1}{1 + e^{-t}}.$$

This model is called the multilayer perceptron (MLP). In fact, this can be seen as a feed-forward neural network with n-nodes in the input layer (one for each independent variable), m nodes in the hidden layer, and one node in the output layer. $\mathbf{w}_i = (w_{i1}, ..., w_{in})$ are the weight vectors that connect each i-th node in the hidden layer with the nodes in the input layer corresponding to independent variables. β_i are the weights connecting each i-th node in the hidden layer with the output node. We define $\langle \mathbf{w}_i, \mathbf{x} \rangle$ as the Euclidean scalar product of input and weight vectors, and b_i as the independent bias associated to node i of the hidden layer. Figure 4 shows the structure of the feed-forward neural network at hand.

Figure 4. Structure of the MLP neural network proposed



Source: elaborated by authors.

The MLP model can be interpreted as a generalised linear model obtained as a combination of basis functions. Each one of these is a nonlinear function of a linear combination of the inputs, where the coefficients in the linear combination are adaptive parameters. For more details about the MLP model, the readers can see Bishop, 1995.

5.2. Regression and Neural Network Models

The fact that a neural network with sigmoidal transfer functions can approximate any given continuous function with the desired accuracy (Hornik, 1989) is a powerful basis for the application of neural networks to regression. Neural network regression is an instance of model-free or non-parametric regression. A model-free regression problem can be stated as follows. Given n pairs of vectors:

$$(\mathbf{x}_l, y_l) = (x_{l1}, x_{l2}, \ldots, x_{lk}, y_l), \quad l = 1, 2, \ldots, n,$$

that have been generated from unknown models

$$y_l = f(\mathbf{x}_l) + \varepsilon_l, \quad l = 1, 2, \ldots, n \quad ,$$

where y is the response variable, $\mathbf{x}_l$ is the independent variables vector, f is an unknown smooth non-parametric function that transforms a k-dimensional Euclidean space into $\mathbb{R}$:

$$f : \mathbb{R}^k \to \mathbb{R}$$

and ε_l are random variables with null mean, $E\left[\varepsilon_l\right] = 0$, and independent of $\mathbf{x}_l$. The aim of the regression is to construct an estimator $\hat{f}$, which is a function of the data $(\mathbf{x}_l, y_l)$, to approximate the unknown function f, and use this estimation to predict a new y given a new $\mathbf{x}$:

$$\hat{y} = \hat{f}(\mathbf{x}).$$

Let us consider a feed-forward neural network with k inputs and a hidden layer with m nodes. The hidden layer carries out a non-linear projection of the input vector $\mathbf{x}$ to a vector $\mathbf{h}$ where:

$$h_j = g\left(\sum_{i=1}^{k} w_{ji} x_i\right), \quad j = 1, \ldots, m.$$

As previously stated, each node performs a non-linear projection of the input vector. $\mathbf{h} = g(\mathbf{x})$ tells how the output layer obtains its values from vector $\mathbf{h}$. Moreover, the projection performed by the hidden layer of a multilayer perceptron distorts data structure and inter-pattern distances in order to achieve a better approximation.

5.3. Extreme Learning Machine

Huang *et al.*, 2004, is the reference for the description of ELM. The regression problem can be formulated as an attempt to find solutions for $\mathbf{w}_i = \left(w_{i1}, ..., w_{in}\right)$ and β_i using the following system of equations:

$$f(\mathbf{x}_j) = \mathbf{t}_j, \quad j = 1, 2, ..., N$$

where

$$f(\mathbf{x}_j) = \sum_{i=1}^{m} \beta_i g\left(\left\langle \mathbf{w}_i, \mathbf{x}_j \right\rangle + b_i\right), \quad j = 1, 2, \ldots, N.$$

This system can also be expressed more concisely as $\mathbf{H\beta} = \mathbf{T}$, where $\mathbf{H}$ is the hidden layer's output matrix of the neural network given by:

$$\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m, b_1, ..., b_m, \mathbf{x}_1, ..., \mathbf{x}_N) = \begin{pmatrix} g(\left\langle \mathbf{w}_1, \mathbf{x}_1 \right\rangle + b_1) & \ldots & g(\left\langle \mathbf{w}_m, \mathbf{x}_1 \right\rangle + b_m) \\ \vdots & \ddots & \vdots \\ g(\left\langle \mathbf{w}_1, \mathbf{x}_N \right\rangle + b_1) & \ldots & g(\left\langle \mathbf{w}_m, \mathbf{x}_N \right\rangle + b_m) \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \mathbf{T} = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}.$$

Each column on matrix $\mathbf{H}$ is made of the values of the corresponding hidden layer node, evaluated for each one of the patterns $\mathbf{x}_i$ in the training set.

The ELM algorithm randomly selects the values for $\mathbf{w}_i = (w_{i1}, ..., w_{in})$ and b_i, and then obtains corresponding values for $\beta_0, \beta_1, ..., \beta_M$, from the generalized linear model. This is done by calculating the minimum quadratic solution of the linear system, given by:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$$

where $\mathbf{H}^{\dagger} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}$ is the generalized Moore-Penrose inverse matrix (Serre, 2002). The solution $\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$ obtained has the following properties:

a) It minimizes the training error:

$$\hat{\boldsymbol{\beta}} = \arg\min \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|.$$

b) It is the minimum Euclidean norm among all the possible solutions of the linear system:

$$\|\hat{\boldsymbol{\beta}}\| = \|\mathbf{H}^{\dagger}\mathbf{T}\| \leq \|\boldsymbol{\beta}\|.$$

In short, the corresponding algorithm for this method is as follows (Figure 5).

Figure 5. ELM algorithm

Given a training set $D = \{(\mathbf{x}_i, t_i) : \mathbf{x}_i \in \mathbb{R}^n, t_i \in \mathbb{R}, i = 1, 2, \ldots, N\}$, the activation function $g(t)$, and m neurons in the hidden layer:
Step 1: Assign arbitrary input weights for w and bias b.
Step 2: Calculate the hidden layer output matrix $\mathbf{H}$.
Step 3: Calculate the output weights $\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$.

Source: elaborated by authors.

Geometrically speaking, the solution obtained corresponds to the orthogonal projection of vector $\mathbf{T}$, which determines the class corresponding to each pattern in the *m*-dimensional vector subspace (given by the number of nodes in the hidden layer) made by the column vectors of matrix $\mathbf{H}$.

Observe that when $N = m$ (i.e. there are as many nodes in the hidden layer as patterns in the training set), matrix $\mathbf{H}$ is square and the corresponding system of equations has a unique solution, which is equivalent to saying that the training error is equal to zero. This happens because vector $\mathbf{T}$ is in the subspace made by the m column vectors of matrix $\mathbf{H}$**,** and thus it can be expressed as the only possible linear combination. In this case, however, the generalisation error over the test set will be greater since overfitting takes place. It is therefore essential to determine the optimal number of nodes to avoid overfitting.

The ELM algorithm has been shown to have a good generalisation capability while it significantly reduces the time needed to train the neural network. For more details on this method the reader can see: Huang *et al*., 2004; and Huang *et al.*, 2006.

5.4. The Levenberg-Marquard Algorithm

The problem of learning in neural networks has been formulated in terms of the minimisation of an error function depending on the adaptive parameters (weights and biases) in the network. The Levenberg-Marquardt (LM) algorithm is a modification of the gradient descendent method specifically designed to a sum-of-squares error function (Levenberg, 1944; Marquardt, 1963). The LM algorithm seeks to minimize the error function while at the same time trying to keep the step size small so as to ensure that the linear approximation of the error function remains valid. It has become a standard technique for nonlinear least-squares problems and can be thought of as a combination of steepest descent and the Gauss-Newton method. For this reason it has been chosen to optimize the MLP model in the comparison carried out in the next section. A description of the LM algorithm applied to neural network optimization can be seen in Bishop, 1995.

6. EXPERIMENTS

Several experiments are carried out in order to evaluate the performance of the regression method in predicting the default rate. The only preprocessing applied to the original database is the (-1,1) normalization of each one of the attributes, to facilitate the fitting processes of the neural networks.

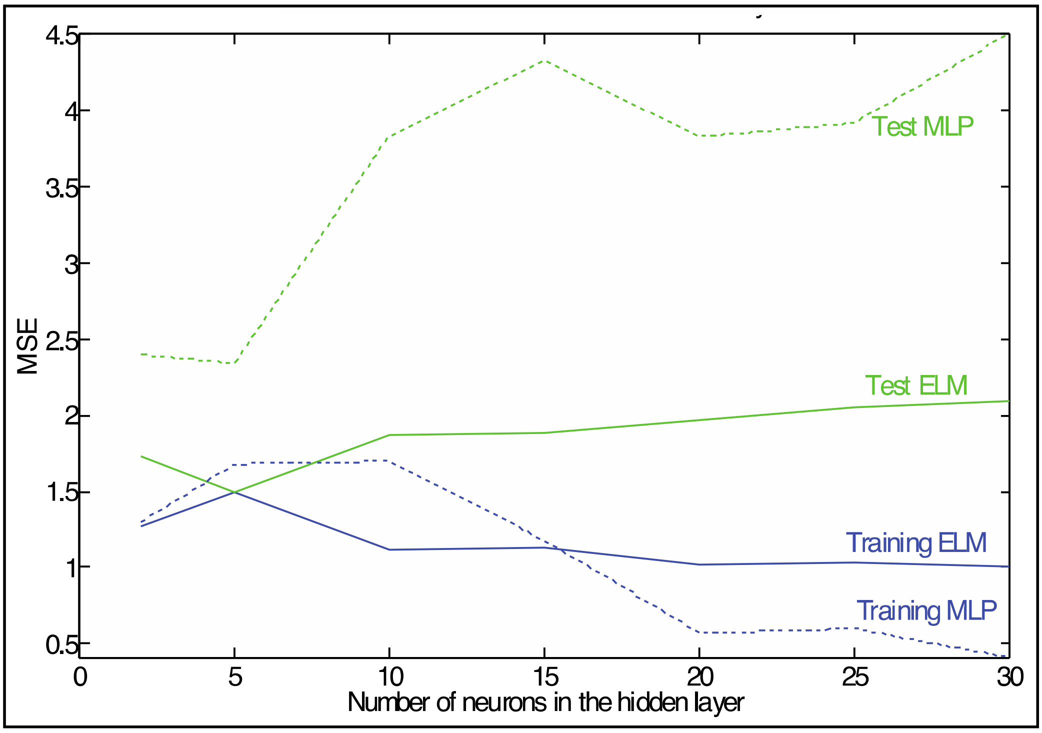
Firstly, the non-linear model defined by the ELM algorithm is compared to the standard MLP neural network training with the Levenberg-Marquard algorithm (MLP-LM). The MLP-LM used in this experiment can be found in the MATLAB Toolbox for Neural Networks included in the 2010 release of MATLAB®. The code for ELM, also in MATLAB, can be downloaded from G. B. Huang’s website (http://www.ntu.edu.sg/home/egbhuang/ELM_Codes.htm). For the comparison of the models the Mean Squared Error (MSE) has been used as given by:

$$MSE = \frac{1}{n}\sum_{l=1}^{n}(y_l - \hat{y}_l)^2 ,$$

where y_l are the observed values and $\hat{y}_l$ the predicted ones.

In order to give both ELM and MLP-LM a fair chance to perform accurately, we run both neural networks for a number of neurons in the hidden layer ranging from $m = 2$ to $m = 30$ neurons. The data set was divided into training (73 patterns) and testing (38 patterns) subsets to evaluate the generalization capability of the models. Both ELM and MLP-LM algorithms have been executed 30 times for the basic partition previously mentioned, and the average value for the regression results has been considered. Figure 6 shows the averaged MSE values for each method (after 30 runs), using the same training and test samples.

Figure 6. ELM *vs.* MLP-LM compared for different number of neurons



Source: elaborated by authors.

We can see how, as we increase the number of neurons in the hidden layer, the complexity of the model increases accordingly. This decreases the MSE in the training set, while it increases in the test set, giving the expected overfitting. The MLP-LM method has a more pronounced overfitting, while ELM works better in terms of generalisation.

Figure 6 shows how both methods perform best for $m = 5$ neurons in the hidden layer. A more detailed look at this particular instance can be seen in Table 3, which contains MSE average values in training and test subsets of the two models for 30 runs of each algorithm. Each model shows the mean, standard deviation, and best and worst results of MSE.

Table 3. MSE values in the training and test subsets of the two models for $m = 5$

Model	# runs	MSE training	Std. dev. MSE training	MSE testing	Std. dev. MSE testing	Best model (testing)	Worst model (testing)
ELM	30	1.501	0.053	1.493	0.235	1.211	2.494
MLP(L-M)	30	1.723	0.532	2.343	1.064	1.162	6.510

Source: elaborated by authors.

6.1. Statistical Comparison

In order to confirm that the differences observed both in the mean and the variance between ELM and MLP-LM (for $m = 5$) are significant, we have made the following verifications using the SPSS package. We have chosen the comparison for $m = 5$ since the best performance for both methods is obtained for this value. The comparison between the two procedures requires, as a first step, a test of normal fit for both distributions. The K-S test shows that both methods present normal distributions (p-values of 0.610 and 0.160 for MLP-LM and ELM, respectively).

Once the normality is verified, and since the partitions are the same for both methods in each repetition, it is now necessary to determine the possible relation between samples, which in turn depends on the method used for their comparison (t-test of either independent or related samples). We can assume the independence of the variables due to the low correlation coefficient (0.081 with p-value 0.670).

The comparison of the two measures, given the difference found in their variances, is significant (zero p-value in both contrasts). We can then conclude that the difference observed both in dispersion and in mean values for both methods is statistically significant, and thus the values of both parameters are greater for the MLP-LM experiment.

6.2. Prediction Capability

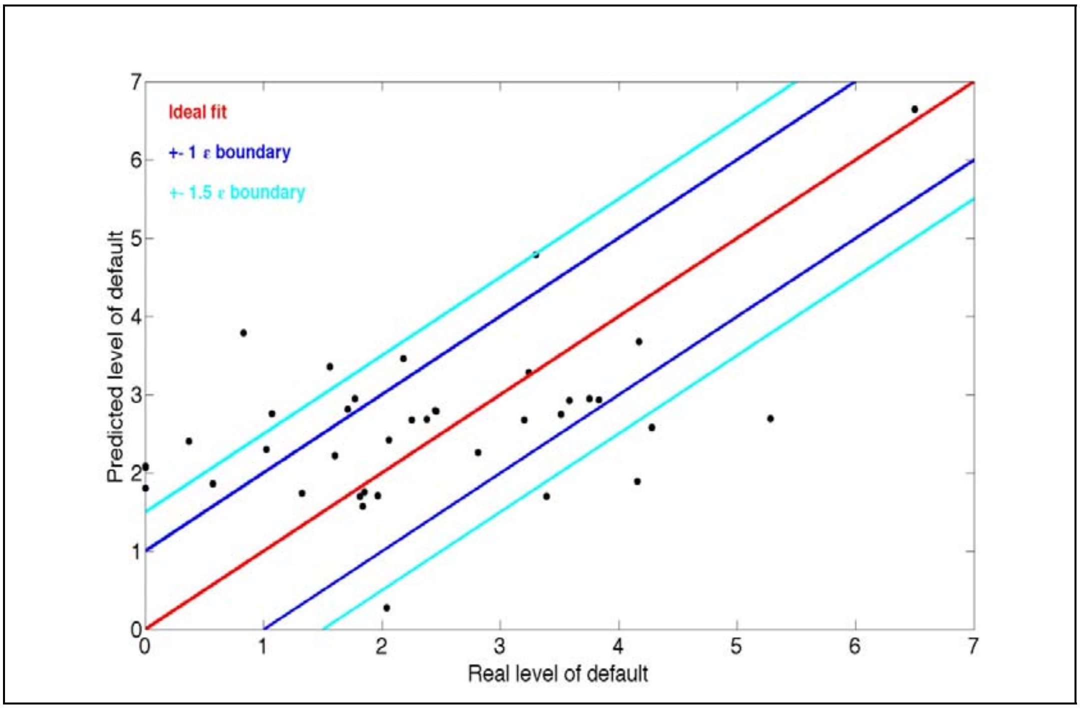
In order to learn more about the ability of this model to predict the default rate, the following measure has been considered. From the best model obtained with ELM (with $m = 5$ hidden nodes), the percentage of patterns is calculated for the training and testing sets whose predicted value belongs to an interval ε around the ideal fit. Table 4 shows the values of this measure for $\varepsilon=1$ and $\varepsilon=1.5$. Moreover, Figure 7 depicts the real and predicted values in our test set for the 30 runs of the ELM method.

Table 4. Values of measure for different ε

Sets	$\varepsilon=1$	$\varepsilon=1.5$
Training	60.3%	75.4%
Testing	55.2%	71.0%

Source: elaborated by authors.

Figure 7. Predicted *vs*. real levels of default in a sample test set



Source: elaborated by authors.

7. CONCLUSION

Today the Spanish financial system is in a process of rearrangement due to the problems caused by the current economic crisis. This is particularly relevant in terms of the relationship between financial institutions and the real estate and construction industries. A critical element for analysis is the level of default, which in turn indicates the quality of the assets held by the entities and the subsequent credit risk they are exposed to. We must bear in mind that, from the point of view of financial institutions, an adequate management of this risk is critical to guarantee the future of the entities themselves. It is also fundamental for the overall strength and reliability of the entire financial system is the fact that the institutions need to improve their performance within the economic system.

Financial authorities also need to be permanently aware of each institution's credit risk situation, and thus demanding greater controls and restrictions in the case where any of them drift towards more problematic scenarios. The demand for solvency needs to be founded on clear and reliable methods of evaluation. These are the motivations that explain the interest in methods that permit the level of default (especially for institutions that anticipate a greater credit risk) to be estimated, given certain magnitudes related to the accounting balance of each entity.

In order to carry out this estimate, a neural network has been used that is based on a non-linear and non-parametric model trained by the ELM algorithm. There are two benefits to be derived from the use of this approach:

a) From a financial point of view, this algorithm provides an acceptable estimate of the level of default. This can be useful for both the entities and the authorities involved.

b) The method allows for a good generalisation and a fast learning adaptation of the data in use. The results obtained by the ELM method provide better performance when compared to those of the MLP-LM method.

ACKNOWLEDGEMENTS

This work was supported in part by the Spanish Inter-Ministerial Commission of Science and Technology under Project TIN2011-22794, the European Regional Development fund, and the *Junta de Andalucía*, Spain, under Project P2011-TIC-7508. We would also like to thank the board of reviewers for their careful study and their ideas for improving both the writing and the experiments presented in this paper.

REFERENCES

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

Boal Velasco N. and González Sánchez, M. (2001) “Estimación del riesgo de crédito mediante modelos internos”, *Banca & Finanzas*, n. 66, pp. 40–45.

Coakley, J.R. and Brown, C.E. (2000) “Artificial Neural Networks in Accounting and Finance: Modelling Issues”, *International Journal of Intelligent Systems in Accounting, Finance & Management*, n. 9, pp. 119–144.

Confederación Española de Cajas de Ahorro (CECA) (2010) “Informe sobre la morosidad en las cajas de ahorros”, Mimeo.

Cruz González, F.J. de la (1998) “Enfoques cuantitativos para la predicción del riesgo de crédito”, en *Predicción de la insolvencia empresarial*, Madrid, Monografías AECA.

García Céspedes, J.C., (2005) “Nuevas técnicas de medición del riesgo de crédito”, *Revista Economía Financiera*, n. 5, abril, pp. 86–114.

Gutiérrez, P.A., Segovia-Vargas, M.J., Salcedo-Sanz, S., Hervás-Martínez, C., Sanchis, A., Portilla-Figueras, J.A., Fernández-Navarro, F. (2009) “Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises”, *Omega*, doi:10.1016/j.omega.2009.11.001.

Herbrich, D., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., and Obermayer, K. (2000) “Neural Networks in Economics: Background, applications and new developments”, in

Advances in Computational Economics: Computational techniques for Modelling Learning Economics, T. Brenner, Editor, Kluwer Academics. pp.169–196 .

Hornik, K. (1989). “Multilayer feedforward neural networks are universal approximators”, *Neural Networks*, 2 (5), pp. 359–366.

Huang, G.B., Wang, D.H., and Lan, Y. (2011) “Extreme learning machines: a survey”, *International Journal of Machine Learning and Cybernetics*, pp. 1–16.

Huang, G., Zhu, Q., and Siew, C. (2004). “Extreme learning machine: a new learning scheme of feedforward neural networks”, *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, *70*, pp. 985–990.

Huang, G., Zhu, Q., and Siew, C. (2006). “Extreme learning machine: Theory and applications”, *Neurocomputing*, *70* (1-3), pp. 489–501.

Levenberg, K. (1944) “A Method for the Solution of Certain Non-linear Problems in Least Squares”, *Quarterly of Applied Mathematics*, 2(2) , Jul, pp. 164–168.

Li, F.C. Wang, P.K., and Wang, G.E. (2009) “Comparison of the primitive classifiers with extreme learning machine in credit scoring”, en *Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference*, pp. 685–688.

López, J.A. and Saidenberg, M.R. (2000) “Evaluating credit risk models”, *Journal of Banking & Finance*, vol. 24, n. 1-2, pp. 151–165.

López-Martín, M.C., Montero-Romero, M.T., Becerra-Alonso, D., and Martínez Estudillo, F.J. (2011), “Clasificación por nivel de morosidad de las entidades de depósito españolas mediante redes neuronales”, *Anales de Economía aplicada 2011*, p. 373.

Mahdavi, M. Fesanghary, M., and Damangir, E. (2007) “An improved harmony search algorithm for solving optimization problems”, *Applied Mathematics and Computation* 188, nº. 2, pp. 1567–1579.

Marquardt, D.W. (1963). “An Algorithm for the Least-Squares Estimation of Nonlinear Parameters”, *SIAM Journal of Applied Mathematics*, 11(2), pp. 431–441, Jun.

Martínez Estudillo, F.J., Hervás Martínez, C., Torres Jiménez, M., and Martínez Estudillo, A.C. (2007) “Modelo no lineal basado en redes neuronales de unidades producto para la clasificación. Una aplicación a la determinación del riesgo en tarjetas de crédito”, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, n. 3, junio, pp. 40–62.

Mcnelis, P.D. (2005) *Neural Networks in Finance: Gaining Predictive Edge in the Market*, Advanced Finance Series, Elsevier Academic Press

Parisi, A., Parisi, F., and Díaz, D. (2006), "Modelos de Algoritmos Genéticos y Redes Neuronales en la Predicción de Índices Bursátiles Asiáticos", *Cuadernos de Economía*, n. 43, pp. 251–284.

Rodríguez Fernández, J.M. (1987) "Crisis en los bancos privados españoles: un modelo logit", *Investigaciones Económicas*, suplemento, pp. 59–64.

Serre, D. (2002) *Matrices: theory and applications*, New York, Springer.

Soler, M. and Miró. A. (2001) "Enfoques cuantitativos para riesgo de crédito de particulares y su aplicación a realidades nacionales diferentes", *Perspectivas del sistema financiero*, n. 72, pp. 43–56.

Sorjamaa, A., Miche, Y., Weiss, R., and Lendasse, A. (2008) "Long-term prediction of time series using NNE-based projection and OP-ELM", en *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference,* pp. 2674–2680.

Van Heeswijk, M., Miche, Y., Lindh-Knuutila, T., Hilbers, P., Honkela, T., Oja, E., and Lendasse, A. (2009) "Adaptive ensemble models of extreme learning machines for time series prediction", *Artificial Neural Networks–ICANN 2009*, pp. 305–314.

Wong, W.K. and Guo, Z.X. (2010) "A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm", *International Journal of Production Economics,* pp. 614–624.

APPENDIX: DATA SOURCES

Asociación Española de Banca: "Estados financieros públicos individuales 2008" (balance y cuenta de pérdidas y ganancias): http://www.aebanca.es/es/EstadosFinancieros/index.htm?pAnio=2008

Confederación Española de Cajas de Ahorro (CECA): "Estados financieros públicos individuales 2008" (balance y cuenta de pérdidas y ganancias): http://www.cajasdeahorros.es/balance.htm

Unión Nacional de Cooperativas de crédito (UNACC): "Estados financieros públicos individuales 2008" (balance y cuenta de pérdidas y ganancias): http://www.ruralvia.com/rsi_data/downloadPDF?p_report=4310_unacc_1.rdf&p_modelo=4310&p_periodo=20081231, http://www.ruralvia.com/rsi_data/downloadPDF?p_report=2300_unacc_1.rdf&p_modelo=2300&p_periodo=20081231