Omega 38 (2010) 333–344

Contents lists available at ScienceDirect

Omega

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

# Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises

P.A. Gutiérrez [a], M.J. Segovia-Vargas [b], S. Salcedo-Sanz [c,*], C. Hervás-Martínez [a], A. Sanchis [d], J.A. Portilla-Figueras [c], F. Fernández-Navarro [a]

[a] Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain
[b] Department of Financial Economics and Accounting I, Universidad Complutense de Madrid, Madrid, Spain
[c] Department of Signal Theory and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain
[d] Bank of Spain, Madrid, Spain

ARTICLE INFO

*Article history:*
Received 30 May 2009
Accepted 4 November 2009
Processed by B. Lev
Available online 20 November 2009

*Keywords:*
Banking crises prediction
Product unit neural networks
Radial basis function neural networks
Logistic regression
Hybrid methods

ABSTRACT

As the current crisis has painfully proved, the financial system plays a crucial role in economic development. Although the current crisis is being of an exceptional magnitude, financial crises are recurrent phenomena in modern financial systems. The literature offers several definitions of financial instability, but for our purposes we identity financial crisis with banking crisis as the most common example of financial instability. In this paper we introduce a novel model for detection and prediction of crises, based on the hybridization of a standard logistic regression with product unit (PU) neural networks and radial basis function (RBF) networks. These hybrid approaches are fully described in the paper, and applied to the detection and prediction of banking crises by using a large database of countries in the period 1981–1999. The proposed techniques are shown to perform better than other existing statistical and artificial intelligence methods in this problem.


## 1. Introduction

The recent financial collapse has stressed the crucial role of the financial system to guarantee the economic development. The financial system is responsible for the allocation of resources over time and among different alternatives of investment, by pricing the postposition of consumption (free risk rate) and pricing the risk (risk premium). A correct functioning of the financial system allows economies to reach higher levels of real growth, as well as more stable macroeconomic conditions. On the other hand, good macroeconomic policies are a prerequisite for financial stability. Therefore sound macroeconomic policies together with a developed financial system reinforce each other guaranteeing financial stability and sustainable growth.

Apart from the current financial crisis, in the last 20 years at least 10 countries have experienced the simultaneous onset of banking and currency crisis, with contractions in gross domestic product of between 5% and 12% in the first year of the crisis, and negative or only slightly positive growth for several years thereafter [1,2]. This emphasizes the fact that preserving financial stability is one of the main goals for policy-makers from the beginning of the monetary systems. It is the special role that banks play in the financial system and their specificities as money issuers that explain why a great number of financial crises have had the banking sector as protagonist. In the 1980s and 1990s, several countries including developed economies, developing countries, and economies in transition, have experienced severe banking crises. Such a proliferation of large scale banking sector problems has raised widespread concern, as banking crises disrupt the flow of credit to households and enterprisers, reducing investment and consumption, and possibly forcing viable firms into bankruptcy. Banking crises may also jeopardize the functioning of the payments system and, by undermining confidence in domestic financial institutions, they may cause a decline in domestic savings and/or a large scale capital outflow. Finally, a systemic crisis may force even solid banks to go bankrupt.

In most countries, policy-makers have attempted to diminish the consequences of banking crises through various types of interventions, ranging from the pursuit of a loose monetary policy to the bail out of insolvent financial institutions with public funds. However, even when they are carefully designed, rescue operations have several drawbacks: they are often very costly for the budget; they may allow inefficient banks to remain in business; they are likely to create the expectation of future bail outs, reducing incentives for adequate risk management by banks and other markets participants (moral hazard); managerial incentives are also weakened when, as it is often the case, rescue operations force healthy banks to bear the losses of ailing institutions.

* Corresponding author. Tel.: +34 91 885 6731; fax: +34 91 885 6699.
*E-mail address:* sancho.salcedo@uah.es (S. Salcedo-Sanz).

0305-0483/$ - see front matter 
doi:10.1016/j.omega.2009.11.001

Finally, loose monetary policy to shore up banking sector losses can be inflationary and, in countries with an exchange rate commitment, it may trigger a speculative attack against the currency. This way, preventing the occurrence of systemic banking problems is undoubtedly a major objective for policy-makers, and understanding the mechanisms that are behind banking crises in the last 15 years is a first step in this direction.

A number of previous studies have analysed different aspects of banking systems [3–5] and episodes of banking sector distress occurred. Most of these works consist of case studies, many of them applying econometric analysis of different situations. For example in [6] an econometric model is used to predict bank failures using Mexican data for the period 1991–1995. In a more recent work [7], the behaviour of a number of macroeconomic variables of the months before and after a banking crisis is analysed. Thus, the authors try to identify variables that act as "early warning signals" for crises. Other studies apply classical statistical techniques such as discriminant, logit or probit analysis [8–11], etc. However, although the obtained results have been satisfactory, all these techniques present the drawback that they make some assumptions about the model or the data distribution that are not usually satisfied. So in order to avoid these inconveniences of statistical methods, it has been recently suggested in the economic field the use of soft-computing techniques, mainly neural networks or evolutionary computation algorithms [12,13].

In recent years, artificial neural networks (ANNs) have been successfully used for modelling financial time series [14–16], for controlling complex manufacturing processes [17] and bankruptcy prediction [18,19]. The most popular neural network model is maybe the back propagation (BP) neural network [20] due to its simple architecture yet powerful problem-solving ability. In ANNs, the hidden neurons are the functional units and can be considered as generators of function spaces. Most existing neuron models are based on the summing operation of the inputs, and, more particularly, on sigmoidal unit functions, resulting in what is known as the multilayer perceptron (MLP). However, alternatives to MLP have arisen in the last few years: product unit neural network (PUNN) models are an alternative to MLPs and are based on multiplicative neurons instead of additive ones. They correspond to a special class of feed-forward neural network introduced by Durbin and Rumelhart [21]. While MLP network models have been very successful, networks that make use of product units (PUs) have the added advantage of increased information capacity [21]. That is, smaller PUNNs architectures can be used than those used with MLPs [22]. They aim to overcome the non-linear effects of variables by means of non-linear basis functions, constructed with the product of the inputs raised to arbitrary powers. These basis functions express the possible strong interactions between the variables, where the exponents may even take on real values and are suitable for automatic adjustment. Another interesting alternative to MLPs are radial basis function neural networks (RBFNNs). RBFNNs can be considered a local approximation procedure, and the improvement in both its approximation ability as well as in the construction of its architecture has been noteworthy [23]. RBFNNs have been used in the most varied domains, from function approximation to pattern classification, time series prediction, data mining, signals processing, and non-linear system modelling and control [24]. RBFNNs use, in general, hyper-ellipsoids to split the pattern space. In many cases, PU and RBF networks are trained by using evolutionary algorithms (EAs), obtaining with this method advantages respect to traditional training approaches [25–27].

In this paper we consider the hybridization of these novel networks (PUs and RBFs) with a standard logistic regression to improve the performance of the classifiers in the problem of bank crises prediction. Logistic regression (LR) has become a widely used and accepted method of analysis of binary or multi-class outcome variables as it is more flexible and it can predict the probability of the state of a dichotomous variable (in our case, the probability of crisis) based on the predictor variables (in our case, macroeconomic variables). The hybridization of LR and PUNNs or RBFNNs is done by considering a recent work in classifier construction [28], where the hybridization of the LR model and evolutionary PUNNs (EPUNNs) to obtain binary classifiers is proposed. In a first step, an evolutionary algorithm [25] is used to determine the basic structure of the product unit model. That step can be seen as a global search in the space of the model coefficients. Once the basis functions have been determined by the EA, a transformation of the input space is considered. This transformation is performed by adding the non-linear transformations of the input variables given by the PU functions obtained by the EA. The final model is linear in these new variables together with the initial covariates. On the other hand, the hybridization of the LR and evolutionary RBFNNs is also tested in this paper, in such a way that we combine a linear model with a radial basis function neural network (RBFNN) non-linear model and then we estimate the coefficients using logistic regression. In this paper we show that the hybrid models involving LR, PUNNs and RBFNNs outperforms several other existing classification techniques in the problem of banking crises prediction, and they are therefore a very interesting tool to take into account in this field.

The structure of the rest of the paper is as follows: next section briefly describes the main variables that are considered as key in banking crises detection. Section 3 describes in detail the hybrid models LR-PUNNs and LR-RBFNNs proposed in this paper. A brief description of the evolutionary algorithm used in the first training of the networks is also included. Section 4 presents the experimental section of the paper, in which we test the good performance of the proposed approaches in a Financial Crisis Database, formed by a sample of data of 79 countries in the period 1981–1999. Finally, the paper is closed with some remarks and conclusions in Section 5.

## 2. Data and variables involved in banking crises detection

This section defines the independent and the dependent variables involved in the present study. As for the dependent variables, the literature offers several definitions of financial instability. In [29] financial stability is defined in terms of its ability to help the economic system allocate resources, manage risks, and absorb shocks. Another strand of the literature focuses on extreme realizations of financial instability. According to [30] a financial crisis is a disruption to financial markets in which adverse selection and moral hazard become much worse, so that financial markets are unable to efficiently channel funds to those who have the most productive investment opportunities. However, in this paper we identity financial instability with banking crisis, because it is the most common example of financial instability given the especial role banks play in the financial system. Once more, academics, central banks and officials offer several definitions of banking crisis [11,31]. Many of these definitions completely solve the problem of how to summarize such description in one single quantitative indicator, or a set of them. On the other hand, these indicators are not readably available for a large number of countries, or there are not enough comparable cross-country data to construct some of the indicators. The empirical literature identifies banking crises as events, expressed through a binary variable, constructed with the help of cross-country surveys [32]. This is the approach that we follow in this work as well. The dependent variable is defined as: systemic and non-systemic banking crises dummy equals one during episodes identified as in [32].

The independent variables included are dictated by the theory on the determinants of banking crisis. All the independent variables are connected with the risks faced by banks. There are risks common to all, no matter the business. However, the typical risks faced by banks are the following ones:

- The credit risk is the risk that borrowers become unable or unwilling to service their debt. Thus, the theory predicts that a shock which adversely affects the economic performance of bank borrowers, and which cannot be diversified, should be positively correlated with systemic banking crises. The empirical literature has highlighted a number of economic shocks associated with episodes of banking sector problems: cyclical output downturns, terms of trade deteriorations, declines in asset prices such as equity and real estate [7,11].
- The interest rate risk: because the asset side of bank balance sheets usually consists of loans of longer maturity at fixed interest rates, the rate of return on assets cannot be adjusted quickly enough, and banks must bear losses. Thus, a large increase in short-term interest rates is likely to be a major source of systemic banking sector problems. The increase in short-term interest rates may be due to several factors, such as an increasing in inflation rate, a shift towards more restrictive monetary policy which raises real rates, an increase in international interest rates, the removal of interest rate controls due to financial liberalization, or the need to defend the exchange rate against a speculative attack[1] [7].
- The currency risk occurs when banks borrow in foreign currency and lend in domestic currency. In this case, an unexpected depreciation of the domestic currency threatens bank profitability. Foreign currency debt was a source of banking problems in Mexico in 1995, in the Nordic Countries in the early 1990s, and in Turkey in 1994 [30].
- Liquidity risk: when bank deposits are not insured, deterioration in the quality of a bank's asset portfolio may trigger a run, as depositors rush to withdraw their funds before the bank declares bankruptcy. Because bank assets are typically illiquid, runs on deposits accelerate the onset of insolvency. The possibility of self-fulfilling runs make banks especially vulnerable financial institutions. A run on an individual bank should not threaten the banking system as a whole unless partially informed depositors take it as a signal that other banks are also at risk (contagion). In these circumstances, bank runs turn into a banking panic.

For all these reasons, as independent variables we employ a set of macroeconomic and financial variables, both qualitative (exchange rate target, monetary policy target, central bank independence and previous crisis) and quantitative (the rest of the variables). Among the macroeconomic variables we include: the real growth of Gross Domestic Product (GDP), the level of real GDP per capita, the inflation rate and the real interest rate to capture the external conditions that countries face. Among the financial variables we include domestic credit growth, bank cash to total assets, bank foreign liabilities to foreign assets and the previous crises as a measure of the degree of confidence of the depositors in the financial stability.

We also explore the role of monetary policy. The existing literature on monetary policy has concentrated on issues more different than financial stability (mainly price stability but also output stabilization). The impact of the monetary policy design on financial stability is related to the very much debated question of the relation between price stability and financial stability. The economic literature is divided as to whether there are synergies or a trade-off between them [30,33]. If synergies existed between the two objectives it would seem safe to argue that the same monetary policy design which helps achieve price stability also fosters financial stability [34]. However, if there were a trade-off, it would be much harder to establish an a priori on the impact of price stability on financial stability. Among the variables related with monetary policy we include the monetary policy regime and the degree of independence of the central bank. This debate has become more important in light of the recent turbulences.

## 3. Description of the hybrid methodologies proposed

The hybrid models we analyse in this paper for the prediction of banking crises are LR models based on the hybridization of the standard linear model and non-linear terms. These non-linear models are constructed with basis functions obtained from evolutionary product unit neural networks (EPUNNs) and evolutionary radial basis functions neural networks (ERBFNNs). In this section we describe the main characteristics of the elements considered for the hybridization: binary LR approach, EPUNNs and ERBFNNs.

### 3.1. Binary logistic regression (LR)

Typically, in supervised classification, a set of $n_{\mathrm{T}}$ training samples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{n_{\mathrm{T}}}, y_{n_{\mathrm{T}}})$ is given. The inputs $\mathbf{x}_i$ (i.e. the set of macroeconomic and financial variables) form a feature space $\mathbf{X}$, and the output $y_i$ (i.e. the "systemic crisis" class) has a class label $c$, which belongs to a finite set $\mathbf{C}$. A classification rule is designed based on the training data, so that, given a new input $\mathbf{x}_i$ with the corresponding values for the macroeconomic variables, a class $c$ from $\mathbf{C}$ with the smallest probability of error is assigned to it.

In this paper the situation considered is the following: a binary outcome variable $y$ ("systemic crisis" (crisis) or "non-systemic crisis" (non-crisis)) is observed together with a vector $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$ of covariates for each of the $n_{\mathrm{T}}$ training samples (assuming that the vector of inputs includes the constant term 1 to accommodate the intercept). The two-class is coded via a 0/1 response $y_i$, where $y_i = 1$ for a crisis sample and $y_i = 0$ for non-crisis samples. Let $p$ be the conditional probability associated with the first class. Logistic regression (LR) is a widely used statistical modelling technique in which the probability $p$ of the dichotomous outcome event is related to a set of explanatory variables $\mathbf{x}$ in the form:

$$\operatorname{logit}(p) = \ln\left(\frac{p}{1-p}\right) = f_{\mathrm{LR}}(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ is the vector of the coefficients of the model, $\boldsymbol{\beta}^{\mathrm{T}}$ is the transpose vector and $f_{\mathrm{LR}}(\mathbf{x}, \boldsymbol{\beta})$ is the LR model. We refer to $p/(1-p)$ as odds-ratio and to expression (1) as the log-odds or logit transformation. A simple calculation in Eq. (1) shows that the probability of occurrence of an event as a function of the covariates is non-linear and is given by

$$p(\mathbf{x}; \boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}}}{1 + e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}}} \tag{2}$$

The complementary event probability can therefore be obtained as $(1-p(\mathbf{x}; \boldsymbol{\beta}))$. Once the conditional probability function defined in (2) is known, the Bayesian (optimal) decision rule can

[1] Higher real interest rates are likely to hurt bank balance sheets even if they can be passed on to borrowers, as higher lending rates result in a larger fraction of non-performing loans.

be constructed as

$$r(\mathbf{x}) = \operatorname{sign}\left\{\ln\left(\frac{p(\mathbf{x};\boldsymbol{\beta})}{1-p(\mathbf{x};\boldsymbol{\beta})}\right)\right\}$$

Given the set of macroeconomic variables $\mathbf{x}$ for a specific bank, the probability $p$ that the bank belongs to the first class can be determined from (2). Similar to the maximum-likelihood classification, these class probabilities for each new bank may be outputted to produce a soft classification. The results from this paper advocate the utility of the LR as a potential approach for the soft classification similar to other recent approaches such as the MLP neural networks or the decision tree regression. A hard classification can be produced by considering a cut-off or threshold probability to assign the corresponding class. Observe that LR not only constructs a decision rule but it also finds a function that for any input vector defines the probability $p$ that the vector $\mathbf{x}$ belongs to the first class.

Let $D=\{(\mathbf{x}_l, y_l); 1 \le l \le n_T\}$ be the training dataset, where the number of samples is $n_T$. Here it is assumed that the training sample is a realization of a set of independent and identically distributed random variables. The unknown regression coefficients $\beta_i$, which have to be estimated from the data, are directly interpretable as log-odds ratios or, in term of $\exp(\beta_i)$, as odds ratios. That log-likelihood used as the error function is

$$l(\boldsymbol{\beta}) = \sum_{l=1}^{n_T} y_l \log p(\mathbf{x}_l;\boldsymbol{\beta}) + (1-y_l)\log(1-p(\mathbf{x};\boldsymbol{\beta})) \quad (3)$$

The estimation of the coefficient vector $\boldsymbol{\beta}$ is usually carried out by means of an iterative procedure like the Newton–Raphson algorithm or the iteratively reweighted least squares (IRLS) [35]. In this paper, two different algorithms have been considered for obtaining the maximum likelihood solution for the logistic regression model, both available in the WEKA machine learning workbench [36]:

- *MultiLogistic*: It is an algorithm for building a multinomial logistic regression model with a ridge estimator to guard against overfitting by penalizing large coefficients, based on the work by le Cessie and van Houwelingen [37]. In order to find the coefficient vector $\boldsymbol{\beta}$ for which $l(\boldsymbol{\beta})$ in (3) is minimized, a quasi-Newton method is used.
- *SimpleLogistic*: It is based on applying LogitBoost algorithm with simple regression functions and determining the optimum number of iterations by a five fold cross-validation: the data are equally splitted five times into training and test, LogitBoost is run on every training set up to a maximum number of iterations (500) and the classification error on the respective test set is logged. Afterwards, LogitBoost is run again on all data using the number of iterations that gave the smallest error on the test set averaged over the five folds. Further details about the algorithm can be found in [38].

### *3.2. Neural network models to hybridize with LR*

This subsection presents the evolutionary product unit neural networks (EPUNNs) and evolutionary radial basis functions neural networks (RBFNNs) considered in the paper. First we briefly introduce the PUNNs and RBFNNs, then we describe the Evolutionary Algorithm used to train these models, and finally we present the hybrid models proposed and used in the paper.

#### *3.2.1. Product unit neural networks (PUNNs) and radial basis functions neural networks (RBFNNs)*

PUNNs are an alternative to MLPs, and are based on multiplicative neurons instead of additive ones. A multiplicative neuron is given by $\prod_{i=1}^{k} x_i^{w_{ji}}$, where $k$ is the number of the inputs. When the exponents are $\{0,1\}$, a higher-order unit is obtained, namely the sigma–pi unit [39]. In contrast to the sigma–pi unit, in the PU the exponents are not fixed and may even take real values. Product unit based neural networks have several advantages, including increased information capacity and the ability to express strong interactions between input variables. Furthermore, it is possible to obtain upper bounds of the Vapnik–Chervonenkis (VC) dimension of product unit neural networks similar to those obtained for MLP [40]. Despite these advantages, PUNNs have a major handicap: they have more local minima and more probability of becoming trapped in them [22]. The main reason for this difficulty is that small changes in the exponents can cause large changes in the total error surface and therefore their training is more difficult than the training of standard MLPs. Several efforts have been made to carry out learning methods for product units [22]. The back propagation algorithm, which is the most common algorithm for training multilayer neural networks, does not work very well with the product units because of its complex error surface.

RBFNNs can be considered a local approximation procedure, and the improvement in both its approximation ability as well as in the construction of its architecture has been noteworthy [23]. RBFNNs use, in general, hyper-ellipsoids to split the pattern space. This is different from MLPs which build their classifications on hyper-planes, defined by a weighted sum. In RBFNNs, the main problem is how to determine the hidden centres' number and their locations. If all the training samples are selected as hidden centres, the generalization capability of the network will become very poor so that many noised or deformed samples will be unable to be recognized, although the network is guaranteed to converge to some satisfying solution. Alternatively, there are many approaches to determine the hidden centres. For instance, the number and position of the RBFs may be fixed and defined a priori [41]; they may be determined by input clustering ($k$-means clustering, fuzzy $k$-means clustering, hierarchical clustering and self-organizing map neural networks) [42]. An interesting alternative is to evolve RBFNNs using evolutionary algorithms (EAs). In [43] a very complete state of the art of the different approaches and characteristics of a wide range of EA and RBFNN combinations is given.

In this work, we define a common framework for both PUNNs and RBFNNs. The structure of the neural network considered is described in Fig. 1: an input layer with $k$ neurons, a neuron for every input variable, a hidden layer with $m$ neurons and an output layer with one neuron.

There is no connection between the neurons of a layer and none between the input and output layers either. The activation function of the $j$-th neuron in the hidden layer is given by $B_j(\mathbf{x}, \mathbf{w}_j)$ and the activation function of the output neuron is given by

$$f(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \sum_{j=1}^{m} \beta_j B_j(\mathbf{x}, \mathbf{w}_j) \quad (4)$$

where $\beta_j$ is the weight of the connection between the hidden neuron $j$ and the output neuron and $\boldsymbol{\theta}$ is the vector of parameters of the neural net. The transfer function of all hidden and output neurons is the identity function. The difference between PUNNs and RBFNNs is related to the activation function considered in the hidden layer. In this way, Product Units (PUs) are considered for PUNNs:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \prod_{i=1}^{k} x_i^{w_{ji}} \quad (5)$$

where $w_{ji}$ is the weight of the connection between input neuron $i$ and hidden neuron $j$ and $\mathbf{w}_j = (w_{j1}, \ldots, w_{jk})$ is the weight vector.

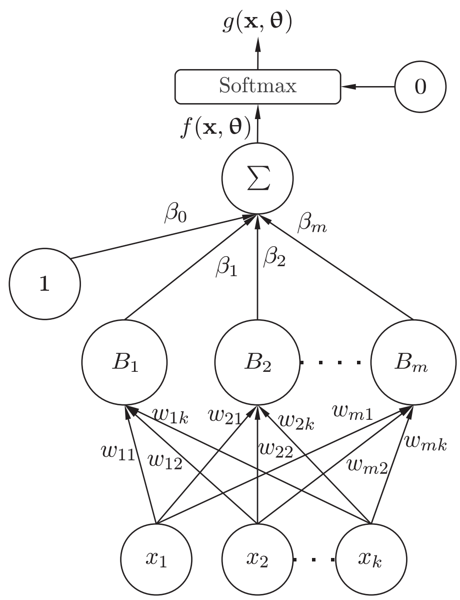


**Fig. 1.** Structure of the neural networks considered: an input layer with $k$ input variables, a hidden layer with $m$ basis function and an output layer with one neuron.

On the other hand, Gaussian RBFs are considered for RBFNNs:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_j\|^2}{r_j^2}\right) \tag{6}$$

where $\mathbf{w}_j = (\mathbf{c}_j, r_j)$, $\mathbf{c}_j = (c_{j1}, c_{j2}, \ldots, c_{jk})$ is the centre or average of the $j$-th Gaussian RBF transformation, $r_j$ is the corresponding radius or standard deviation and $c_{ji}, r_j \in \mathbb{R}$.

We consider the softmax activation function [20] given by

$$g(\mathbf{x}, \boldsymbol{\theta}) = \frac{\exp f(\mathbf{x}, \boldsymbol{\theta})}{1+\exp f(\mathbf{x}, \boldsymbol{\theta})} \tag{7}$$

where $f(\mathbf{x}, \boldsymbol{\theta})$ is the output of the output neuron for pattern $\mathbf{x}$ and $g(\mathbf{x}, \boldsymbol{\theta})$ is the probability that a pattern $\mathbf{x}$ belongs to the crisis class. With this model, the cross-entropy error function is defined by Eq. (3), but substituting $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ with $g(\mathbf{x}, \boldsymbol{\theta})$.

### 3.2.2. *Evolutionary programming algorithm*

In order to estimate the parameters and the structure of the PUNNs or RBFNNs that minimizes the classification error function, an evolutionary algorithm (EA) has been considered. The algorithm is similar to the one proposed by [44]. The population-based evolutionary algorithm for architectural design and the estimation of real-coefficients have points in common with other evolutionary algorithms in the bibliography [26,45,46]. The search begins with an initial population. This population is updated in each generation using a population-update algorithm, and is subject to the evolutionary operations of replication and mutation. Crossover is not used due to its potential disadvantages in evolving artificial networks [45]. For this reason, this EA belongs to the evolutionary programming (EP) paradigm. The general structure of the EA is detailed next:

*Evolutionary programming* (EP)

(1) Generate a random population of size $N_{\mathrm{P}}$.
(2) Repeat until the maximum number of generations.
  (a) Apply parametric mutation to the best 10% of individuals.
  (b) Apply structural mutation to the remaining 90% of individuals.
  (c) Calculate the fitness of every individual in the population.
  (d) Add best fitted individual of the last generation (elitist algorithm).
  (e) Rank the individuals with respect to their fitness.
  (f) Best 10% of population individuals are replicated and substitute the worst 10% of individuals.
(3) Select the best individual of the population in the last generation and return it as the final solution.

First, the initial population is generated: the algorithm begins with the random generation of a larger number of networks than the number of networks used during the evolutionary process. $10N_{\mathrm{P}}$ networks are generated, where $N_{\mathrm{P}}$ is the number of individuals of the population to be trained during the evolutionary process. We consider the cross-entropy $l(\boldsymbol{\theta})$ as the error function of an individual $g(\mathbf{x}, \boldsymbol{\theta})$ of the population, $g$ being a PUNN or a RBFNN; and then, the fitness measure is a decreasing strictly transformation of the error function $l(\boldsymbol{\theta})$ given by $A(g) = 1/(1+l(\boldsymbol{\theta}))$, where $0 < A(g) \leq 1$.

The adjustment of both weights and structure of the NNs is performed by the complementary action of two mutation operators: parametric and structural mutation. Parametric mutation implies a modification in the coefficients ($\beta_j$ and $w_{ji}$) of the model, using a self adaptive simulated annealing algorithm [47]. Structural mutation modifies the topology of the neural nets, helping the algorithm to avoid local minima and increasing the diversity of the trained individuals. Five structural mutations are applied sequentially to each network: neuron deletion, connection deletion, neuron addition, connection addition and neuron fusion. In order to define the topology of the neural networks generated in the evolution process, two parameters are considered: $m$ and $M_{\mathrm{E}}$. They correspond to the minimum and the maximum number of hidden neurons in the whole evolutionary process. In order to obtain an initial population formed by models simpler than the most complex models possible, parameters must fulfil the condition $m \leq M_E$.

More details about the EA can be found in [27,44]. Moreover, some specific characteristics of the algorithm need to be adapted for evolving RBFNNs. These characteristics can be found in [48]. For the experimental section, we will denote the RBFNNs obtained by the EP algorithm as evolutionary radial basis functions (ERBF method) and the PUNNs obtained by this EP algorithm as evolutionary product units (EPU method).

### 3.2.3. *Logistic regression using product units (LRPU) and radial basis functions (LRRBF)*

Logistic regression using product units (LRPUs) and using radial basis functions (LRRBFs) are two hybrid methods that considers the EP presented in the previous section in order to obtain a PUNN or a RBFNN structure and hidden neuron weights accurate enough. When these are obtained, it applies a multilogistic regression maximum likelihood (ML) optimization over the basis functions (PUs or RBFs) of the NN selected. So the generic model is given by

$$f(\mathbf{x}, \boldsymbol{\theta}) = \alpha_0 + \sum_{j=1}^{m} \beta_j B_j(\mathbf{x}, \mathbf{w}_j) \tag{8}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{W})$, $\boldsymbol{\alpha} = (\alpha_0, \beta_1, \ldots, \beta_m)$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$, with $\mathbf{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jk})$, $w_{ji} \in R$. The coefficients $\mathbf{W}$ are given by the EP algorithm, they not being adjusted by the ML method. The ML method only optimizes the linear part of the model, i.e. the $\boldsymbol{\alpha}$ coefficients.

Two different multilogistic regression algorithms are applied, both available in the WEKA machine learning workbench [36]: (1) MultiLogistic [37], which considers all initial and RBF covariates and (2) SimpleLogistic [38], which incrementally constructs the model and applies cross-validation, resulting in an automatic covariate selection. $B_j(\mathbf{x}, \mathbf{w}_j)$ is defined with (5) or (6), what, together with the two ML algorithms, results in four different methods: MultiLogistic regression using product units (LRPU), SimpleLogistic regression using product units (LRPU*), MultiLogistic regression using RBFs (LRRBF) and SimpleLogistic regression using RBFs (LRRBF*).

*3.2.4. Logistic regression using initial covariates and product units (LRIPU) and radial basis functions (LRIRBF)*

These models extend those presented in the previous subsection, considering the initial covariates $\mathbf{x}$ of the problem. Their generic expression is given by

$$f(\mathbf{x}, \boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^{k} \alpha_i x_i + \sum_{j=1}^{m} \beta_j B_j(\mathbf{x}, \mathbf{w}_j) \tag{9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{W})$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m)$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$, with $\mathbf{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jk})$, $w_{ji} \in R$. The values adjusted with the ML method correspond again to the $\boldsymbol{\alpha}$ vector, the coefficients $\mathbf{W}$ being given by the EP algorithm.

In the same way that with the previously presented models, two different logistic regression algorithms are applied (SimpleLogistic and MultiLogistic) over the two different models defined by considering (5) or (6) for $B_j(\mathbf{x}, \mathbf{w}_j)$, what finally results in four different methods: MultiLogistic regression using initial covariates and PUs (LRIPU), SimpleLogistic regression using initial covariates and PUs (LRIPU*), MultiLogistic regression using initial covariates and RBFs (LRRBF) and SimpleLogistic regression using initial covariates and RBFs (LRIRBF*).

## 4. Computational experiments and results

This section is structured in several subsections. First, we describe the Database used to evaluate the performance of our hybrid approaches. Several existing algorithms for comparison purposes are described in Section 4.2, and the complete experimental design organization is shown in Section 4.3. Finally, the main results obtained with the proposed techniques are presented in Section 4.4.

### *4.1. Database description*

The Financial Crisis Database used in this study is formed by a sample of 79 countries in the period 1981–1999 (annual data). The database has 521 samples, 164 are crisis samples, and the rest are non-crisis cases.

The dependent variable is:

- *Systemic and non-systemic banking crises dummy*: equals one during episodes identified as in [32]. They present information on 117 systemic banking crises (defined as much or all of bank capital being exhausted) that have occurred since the late 1970s in 93 countries and 51 smaller non-systemic banking crises in 45 countries during that period. The information on crises is cross-checked with that of [9] and with International Monetary Fund staff reports and financial news.

And the independent variables are the following:

- *Monetary policy strategies*: these variables (Exchange rate target, Monetary policy target) are dummies. The exchange rate target takes four values depending on the exchange rate regime: free floating (FF), managed floating (MF), pegged currencies (PC) and currency board (CB). The Monetary policy target equals one during periods in which targets were based on monetary aggregates, two when the objective was inflation, three when the two variables are into the objective function and zero in other cases, according to the chronology of the Bank of England survey of monetary frameworks, in [49]. Since it provides a chronology for the 1990s, we have complemented it with information from other sources for the previous years. Regarding exchange rate arrangements, we use classifications of exchange rate strategies in [50,51], and [52] for Latin America countries. Data for monetary and inflation targets were complemented with the information taken from [51,53]. It should be noted that some judgement has gone into the classification of regimes.
- *Central bank independence*: measures to what extent the central banks are legally independent according to their charters, following the approach in [33]. This variable goes from 0 (least independent) to 1 (most independent) and is taken from [33], for the 1970s and 1980s and for the 1990s from [49,54]. The index of independence is assumed to be constant through every year of each decade.
- *Inflation*: percentage change in the GDP deflator. Source: International Monetary Fund, International Financial Statistics, line 99bir.
- *Real interest rate*: nominal interest rate minus inflation in the same period, calculated as the percentage change in the GDP deflator. Source: International Monetary Fund, International Financial Statistics. Where available, money market rate (line 60B); otherwise, the commercial bank deposit interest rate (line 60l); otherwise, a rate charged by the central bank to domestic banks such as the discount rate (line 60).
- *Net capital flows to GDP*: capital account plus financial account plus net errors and omissions. Source: International Monetary Fund, International Financial Statistics, lines (78bcd + 78bjd + 78cad).
- *Real GDP per capita in 1995 US dollars*: this variable is expressed in US dollars instead of purchasing power parity (PPP) for reasons of data availability. GDP per capita in PPP was available only for two points in time. Source: The World Bank, World Tables; and The European Bank for Reconstruction and Development (EBRD), Transition Report, for some transition countries.
- *Real GDP growth*: percentage change in GDP Volume (1995 = 100). Source: International Monetary Fund, International Financial Statistics (line 99bvp) where available; otherwise, The World Bank, World Tables; and EBRD, Transition Report, for some transition countries.
- *World real GDP growth*: percentage change in GDP volume (1995 = 100). Source: International Monetary Fund, International Financial Statistics (line 99bvp) where available; otherwise, The World Bank, World Tables; and EBRD, Transition Report, for some transition countries.
- *Domestic Credit growth*: percentage change in domestic credit, claims on private sector. Source: International Monetary Fund, International Financial Statistics, line 32d.
- *Bank Cash to total assets*: Reserves of deposit money banks divided by total assets of deposit money banks. Source: International Monetary Fund, International Financial Statistics, line 20 divided by lines (22a + 22b + 22c + 22d + 22f).
- *Bank foreign liabilities to foreign assets*: deposit money banks foreign liabilities to foreign assets. Source: International Monetary Fund, International Financial Statistics, lines (26c + 26cl) divided by line 21.
- *Previous crises*: this variable equals zero if the country has not have a previous crisis; one, if the country has suffered one previous crisis; two, in case of two or three previous crisis, and, three, otherwise.

*4.2. Alternative statistical and artificial intelligence methods used for comparison purposes*

Different state-of-the-art statistical and artificial intelligence algorithms have been implemented for comparison purposes. Specifically, the results of the following algorithms have been compared with the soft-computing techniques presented in this paper:

(1) The logistic model tree (LMT) [38] classifier.
(2) The C4.5 classification tree inducer [36].
(3) The $k$ nearest neighbour ($k$-NN) classifier.
(4) The support vector machine (SVM) classifier [35] with RBF kernels and using the sequential minimal optimization (SMO) algorithm.
(5) A Gaussian radial basis function network (RBFNetwork) [36], deriving the centres and width of hidden units using $k$-means, and combining the outputs obtained from the hidden layer using logistic regression.
(6) The multilayer perceptron (MLP) neural network [20], trained with a simple Back Propagation (BP) algorithm.
(7) The Naive Bayes standard learning algorithm (NaiveBayes) [36].
(8) The RoughSet methodology for classification [55].

These algorithms have been selected for comparison since they are some of the best performing algorithms of recent literature on classification problems. Many of these approaches have also been tested before in bankruptcy detection problem. The detailed description and some previous results of these methods can be found in [35,36,38,56–58].

*4.3. Experimental design*

The evaluation of the different models has been performed using four different measures: correctly classified rate (*CCR*) or accuracy, area under the receiver operating characteristic curve (*AUC*), specificity (*Sp*) and sensitivity (*Se*). Part of these measures are obtained from the contingency or confusion matrix $M(g)$ of a classifier $g$:

| Target | Predicted | |
|---|---|---|
| | Non-crisis | Crisis |
| Non-crisis | $n_{\mathrm{TN}}$ | $n_{\mathrm{FP}}$ |
| Crisis | $n_{\mathrm{FN}}$ | $n_{\mathrm{TP}}$ |

where $n_{\mathrm{TN}}$ represents the number of true negative examples (i.e. patterns correctly predicted as non-crisis), $n_{\mathrm{FP}}$ represents the number of false positive examples (i.e. patterns predicted as crisis when they are really non-crisis patterns), $n_{\mathrm{FN}}$ represents the number of false negative examples (i.e. patterns predicted as non-crisis when they are really crisis patterns), and $n_{\mathrm{TP}}$ represents the number of true positive examples (i.e. patterns correctly predicted as crisis). The diagonal corresponds to correctly classified patterns and the off-diagonal to mistakes in the classification task.

Using this confusion matrix, the following measures can be derived:

- The correctly classified rate (*CCR*) or accuracy is defined as

$$CCR = \frac{n_{\mathrm{TN}} + n_{\mathrm{TP}}}{n_{\mathrm{TN}} + n_{\mathrm{FN}} + n_{\mathrm{FP}} + n_{\mathrm{TP}}} \tag{10}$$

that is, the rate of all the correct predictions.

- The specificity (*Sp*) is defined as

$$Sp = \frac{n_{\mathrm{TN}}}{n_{\mathrm{TN}} + n_{\mathrm{FP}}} \tag{11}$$

that is, the rate of non-crisis examples that are correctly classified.

- The sensitivity (*Se*) is defined as

$$Se = \frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}} + n_{\mathrm{FN}}} \tag{12}$$

that is, the rate of crisis examples that are correctly classified.

By using these measures, it is possible to analyse the two kind of errors that can be committed in a binary classification problem: type I error or false negative (FN) rate (negative instances that were erroneously reported as being positive, $\mathrm{FN} = 1 - Sp$) and type II error or false positive (FP) rate (positive instances that were erroneously reported as being negative, $\mathrm{FP} = 1 - Se$). This analysis is important because the cost of classifying a banking sector in crisis as non-crisis is more serious than classifying a healthy one as crisis [59]. Note that all these measures are obtained using a 0.5 probability threshold for the classification of the banking crises. However, the use of arbitrary cut-off probabilities makes the computed error rates difficult to interpret [60]. However, the use of a relevant pay-off function and prior probabilities to determine the optional cut-off probability could introduce some kind of bias on the results, which could favour the models proposed.

Receiver operating characteristics (ROC) graphs are useful for organizing classifiers and visualizing their performance. ROC graphs are two-dimensional graphs in which TP rate is plotted on the $Y$ axis and FP rate is plotted on the $X$ axis. An ROC graph depicts relative trade-offs between benefits (true positives) and costs (false positives). The diagonal line $y = x$ represents the strategy of randomly guessing a class. For example, if a classifier randomly guesses the positive class half the time, it can be expected to get half the positives and half the negatives correct; this yields the point $(0.5, 0.5)$ in ROC space.

Some classifiers naturally yield an instance probability or score, a numeric value that represents the degree to which an instance is a member of a class. Such a ranking or scoring classifier can be used with a threshold to produce a discrete (binary) classifier. Each threshold value produces a different point in ROC space, and if we join all these points we obtain an ROC curve. To compare classifiers we may want to reduce ROC performance to a single scalar value representing expected performance. A common method is to calculate the area under the ROC curve, abbreviated *AUC*. Since the *AUC* is a portion of the area of the unit square, its value will always be between 0 and 1. However, because random guessing produces the diagonal line between $(0, 0)$ and $(1, 1)$, which has an area of 0.5, no realistic classifier should have an *AUC* less than 0.5. More details about the ROC curves and the *AUC* measure can be found in [61]. Bradley [62] has compared popular machine learning algorithms using AUC, and found that AUC exhibits several desirable properties compared to accuracy. For example, AUC has increased sensitivity in analysis of variance (ANOVA) tests, is independent to the decision threshold, and is invariant to a priori class probability distributions [62]. Moreover, the *AUC* measure is more sensitive to the errors on the positive class, since it has an important statistical meaning: it is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance, considering also all possible thresholds.

Because of these reasons, the *AUC* has been selected as the main comparison measure in this work, since it avoids the use of arbitrary cut-off probabilities in prediction tests, what makes the computed error rates difficult to interpret.

The different soft computing experiments were conducted using a software package developed in JAVA by the authors, as an extension of the JCLEC framework (http://jclec.sourceforge.net/) [63]. This software package is available in the non-commercial

JAVA tool named KEEL (http://www.keel.es) [64]. We used WEKA [36] (http://www.cs.waikato.ac.nz/ml/weka) for obtaining the results of the rest of methods, considering each method default parameter values.

The parameter values used in the hybrid techniques proposed are the following: to start processing data, each of the input variables was scaled in the rank $[1, 2]$ for PUNNs and $[-2, 2]$ for RBFNNs. Weights are assigned using a uniform distribution defined throughout two intervals, $[-2, 2]$ for connections of RBFNNs between the input layer and hidden layer, $[-5, 5]$ for connections of PUNNs between the input layer and hidden layer, and, for all kinds of neurons, $[-10, 10]$ for connections between the hidden layer and the output layer. The initial value of the radii $r_j$ for RBFNNs is obtained as a random value in the interval $(0, d_{\max}]$, where $d_{\max}$ is the maximum distance between two training input examples. The minimum and maximum number of hidden neurons for RBFNN models are $m = 7$ and $M_E = 9$ respectively, and $m = 1$ and $M_E = 3$ for PUNNs.

The size of the population is $N_P = 500$. We have considered $\alpha(0) = 0.5$, $\lambda = 0.1$ and $s = 5$ for the parametric mutations. For the structural mutations, the number of hidden neurons that can be added or removed in a structural mutation is within the $[\Delta_{\min}, \Delta_{\max}] = [1, 2]$ interval. The ratio of the number of connections to add or delete in the hidden and the output layer during structural mutations is $\Delta_o = 0.05$ and $\Delta_h = 0.3$.

The stop criterion is reached whenever one of the following two conditions is fulfilled: the variance of the fitness of the best 10% of the population is less than $10^{-4}$ or 500 generations are completed.

In order to assess the ability of the models in an out-of-time dataset, the dataset has been split in two subsequent time periods, holding the later for evaluation of the model only. Consequently, the training information table consisted of 442 samples (305 non-crisis and 137 crisis samples) from 79 countries in the period 1981–1997 (annual data) described by the variables explained in Section 2. The test or generalization information table consisted of 79 data described by the same variables in the period 1998–1999 (52 non-crisis and 27 crisis samples). Since the LRPU, LRPU*, LRRBF, LRRBF*, LRIPU, LRIPU*, LRIRBF and LRIRBF* methods are stochastic, the training process was repeated 30 times in order to evaluate the randomness of the evolutionary method.

### 4.4. Results

The experimental results presented in this paper are structured in two subsections. The first subsection studies which of the hybrid techniques proposed yields a better trade-off between performance and complexity of the model obtained. The second subsection compares the best performing techniques to the other statistical and artificial intelligence methods.

#### 4.4.1. Comparison of the different hybrid techniques proposed

In this subsection, the different hybrid techniques presented in Section 3.2 are compared to each other, with regards to the basis function used for the hidden layer of the models. In this way, the following methods are compared:

- Product unit methods: EPU, LRPU, LRPU*, LRIPU and LRIPU*.
- Radial basis function methods: ERBF, LRRBF, LRRBF*, LRIRBF and LRIRBF*.

The results obtained with all these methods are included in Table 1, where the four measures previously presented are obtained over the generalization set ($CCR_G$, $AUC_G$, $Sp_G$ and $Se_G$). The training results has not been considered, since it is well-known that the results in the training sample are upwards biased [65]. The mean and the standard deviation values of these measures for the best models obtained in the 30 executions of the different algorithms were calculated and included in the table. Similar results have been included for the number of links or connections (#conn.) of these models. From a purely quantitative point of view, the best mean $CCR_G$ is obtained by ERBF, the best mean $AUC_G$ by LRIPU*, the best mean $Sp_G$ by LRPU*, the best mean $Se_G$ by LRIRBF* and the lowest mean number of connections is associated to the LRPU* method. Moreover, RBF methods result, in general, in a higher number of connections and the use of the SimpleLogistic algorithm (included in those methods with a final *) always reduces the number of connections and avoids over-fitting.

We have used a statistical independent *t*-test for analysing whether there is a significant difference between the means of the hybrid RBF and PU algorithms. As we previously observed, the $AUC_G$ measure is independent of the threshold selected, so the $AUC_G$ will be the test variable to analyse the differences of performance between the different algorithms. The results of these tests are included in Table 2. From the analysis of these results, it can be concluded that LRIRBF* method obtains significant better results than LRIRBF and LRRBF methods, and that LRIPU* obtains significant better results than all the other methods. Taking into account that the highest mean $AUC_G$ is obtained by the LRIPU* method and that it also results in almost the lowest number of connections (what means a higher interpretability, as we will check in Section 4.5), the method selected for banking crises prediction is LRIPU*.

#### 4.4.2. Comparison with other statistical and artificial intelligence methods

In this section we focus on the LRIPU* method, as it is the best in terms of $AUC_G$ among the different hybrid approaches proposed in this paper. In order to complete the experimental section, a comparison of this method with well-known techniques for classification given in Section 4.2 has been carried out. These other techniques are deterministic, so we have selected the best $AUC_G$ model of the 30 runs to perform the comparison. Table 3 presents the results obtained with the different techniques, and the result obtained by the LRIPU* network. The results of the second best hybrid method (LRIRBF*) are also shown for reference. The $AUC_G$ results for the RoughtSets method could not be obtained, as this method does not include a threshold that allows to represent the ROC curve. However, there are standard techniques for obtaining probability estimates (and consequently ROC curves) for those methods which do not directly output a probability value (LMT, C4.5, *k*-NN and LibSVM, see [36] for more details).

Note that the LRIPU* network obtains the best result in terms of $CCR_G$ and $Sp_G$ over all techniques compared. The best results in terms of $Se_G$ are obtained by the LMT and C4.5 methods, but with a high decrease in $CCR_G$ and $Sp_G$. RoughSets are an interesting alternative, obtaining a slightly lower $Se_G$ than LMT and C4.5, but without losing so much accuracy. However, the LRIPU* and LRIBF* networks results in the highest $AUC_G$, what make them the most competitive classifiers, given that the $AUC_G$ measure comprises the behaviour of the classifier for all possible thresholds. The differences in $AUC_G$ are really significant with respect to techniques such as LMT, C4.5, *k*-NN, RBFNetwork or MLP. In general, these results show that the proposed hybrid approaches based on LR and PU or RBF networks are robust approaches to tackle the prediction of banking crises, and obtain better results than the majority of the existing alternative methods.

**Table 1**
Statistical results (mean and standard deviation, SD) of the $CCR_G$, $AUC_G$, specificity ($Sp_G$), sensitivity ($Se_G$) and number of connections (#conn.) obtained using the different soft computing methods proposed.

| Method | $CCR_G$ Mean ± SD | $AUC_G$ Mean ± SD | $Sp_G$ Mean ± SD | $Se_G$ Mean ± SD | #conn. Mean ± SD |
|---|---|---|---|---|---|
| ERBF | **84.94 ± 3.05** | 0.8950 ± 0.0270 | 91.86 ± 3.60 | 71.60 ± 4.60 | 72.93 ± 15.30 |
| LRRBF | 83.76 ± 3.16 | 0.8842 ± 0.0323 | 90.45 ± 3.44 | 70.86 ± 5.65 | 71.67 ± 16.56 |
| LRRBF* | 84.39 ± 2.87 | 0.8804 ± 0.0681 | 91.73 ± 3.58 | 70.25 ± 6.56 | 66.37 ± 16.03 |
| LRIRBF | 77.93 ± 3.80 | 0.8434 ± 0.0316 | 81.35 ± 5.32 | 71.36 ± 5.32 | 93.67 ± 16.56 |
| LRIRBF* | 83.88 ± 3.05 | 0.9027 ± 0.0272 | 89.68 ± 4.59 | **72.72 ± 3.44** | 64.97 ± 12.75 |
| EPU | 82.19 ± 3.97 | 0.8768 ± 0.0428 | 92.37 ± 5.43 | 62.59 ± 10.03 | 22.03 ± 8.88 |
| LRPU | 82.78 ± 4.12 | 0.8766 ± 0.0461 | 92.95 ± 5.61 | 63.21 ± 10.15 | 22.03 ± 8.88 |
| LRPU* | 83.08 ± 4.30 | 0.8819 ± 0.0349 | **93.72 ± 5.20** | 62.59 ± 9.19 | **21.03 ± 8.90** |
| LRIPU | 80.76 ± 2.22 | 0.8612 ± 0.0432 | 85.64 ± 2.89 | 71.36 ± 5.50 | 44.03 ± 8.88 |
| LRIPU* | 84.81 ± 3.03 | **0.9043 ± 0.0239** | 91.47 ± 4.91 | 71.98 ± 4.09 | 25.10 ± 11.77 |

The best quantitative result is represented in bold face.

**Table 2**
$p$-Value results of a pair-wise relationated $t$-test comparing the $AUC_G$ values of the different variants of the methods associated for each basis function considered (PU and RBF).

| PU methods | | | RBF methods | | |
|---|---|---|---|---|---|
| Methods compared | | $p$-Value | Methods compared | | $p$-Value |
| ERBF vs. | LRRBF | 0.010[a] | EPU vs. | LRPU | 0.904 |
| | LRRBF* | 0.235 | | LRPU* | 0.270 |
| | LRIRBF | 0.000[a] | | LRIPU | 0.011[a] |
| | LRIRBF* | 0.076 | | LRIPU* | 0.000[b] |
| LRRBF vs. | LRRBF* | 0.663 | LRPU vs. | LRPU* | 0.342 |
| | LRIRBF | 0.000[a] | | LRIPU | 0.010[a] |
| | LRIRBF* | 0.004[b] | | LRIPU* | 0.001[b] |
| LRRBF* vs. | LRIRBF | 0.001[a] | LRPU* vs. | LRIPU | 0.016[a] |
| | LRIRBF* | 0.097 | | LRIPU* | 0.000[b] |
| LRIRBF vs. | LRIRBF* | 0.000[b] | LRIPU vs. | LRIPU* | 0.000[b] |

[a] the first method is significantly better than the second one with $\alpha = 0.05$.
[b] the second method is significantly better than the first one with $\alpha = 0.05$.

**Table 3**
Results of the $CCR_G$, $AUC_G$, specificity ($Sp_G$) and sensitivity ($Se_G$) of the best hybrid method proposed compared to those obtained using different statistical and artificial intelligence methods.

| Method | $CCR_G$ | $AUC_G$ | $Sp_G$ | $Se_G$ |
|---|---|---|---|---|
| LMT | 74.68 | 0.7347 | 67.31 | **88.89** |
| C4.5 | 78.48 | 0.7244 | 73.08 | **88.89** |
| KNN | 63.29 | 0.6232 | 65.38 | 59.26 |
| LibSVM | 83.54 | 0.7949 | 92.31 | 66.67 |
| SLogistic | 83.54 | 0.8875 | 88.46 | 74.07 |
| RBFNetwork | 67.09 | 0.6182 | 98.08 | 7.41 |
| MLP | 69.62 | 0.7265 | 71.15 | 66.67 |
| NaiveBayes | 67.09 | 0.8775 | 98.08 | 7.41 |
| MultiLogistic | 82.28 | 0.8782 | 86.54 | 74.07 |
| RoughSets | 86.08 | – | 88.46 | 81.48 |
| LRIRBF* | 84.81 | 0.9209 | 90.38 | 74.07 |
| LRIPU* | **91.14** | **0.9295** | **100.00** | 74.07 |

The best quantitative result is represented in bold face.

**Table 4**
Probability expression of the best LRIPU* model, $CCR_G$, $AUC_G$ values and testing confusion matrix associated with this model.

Best LRIPU* crisis probability model

$$p_{\text{Crisis}} = 1 - \frac{e^{f(\mathbf{x},\boldsymbol{\theta})}}{1+e^{f(\mathbf{x},\boldsymbol{\theta})}}$$
$$f(\mathbf{x},\boldsymbol{\theta}) = 6.15 - 2.14x_{15} - 4.72PU_1 + 2.29PU_2$$
$$PU_1 = x_{11}^{-4.72}x_{19}^{-0.41}$$
$$PU_2 = x_4^{0.65}x_{10}^{-0.25}x_{14}^{-1.29}x_{22}^{-3.86}$$

$x_4 \leftarrow (\text{exchange}=\text{MF}); x_{10} \leftarrow (\text{cbanIndep}); x_{11} \leftarrow (\text{realIntRat});$
$x_{14} \leftarrow (\text{forLiabRev}); x_{15} \leftarrow (\text{gdpGrowth}); x_{19} \leftarrow (\text{previousCrisis}=0);$
$x_{22} \leftarrow (\text{previousCrisis}=3);$

$x_i \in [1.0, 2.0]$

$CCR_G = 91.14\%, AUC_G = 0.9295$
Generalization confusion matrix

| | Predicted | |
|---|---|---|
| Target | 0 | 1 |
| 0 | 52 | 0 |
| 1 | 7 | 20 |

## 4.5. Discussion

The economical interpretation of the model is based on the output of the LRIPU* network, presented in Table 4. This is the model that shows the best performance in terms of $AUC_G$. It is important to note that this model is highly non-linear, and so a direct interpretation in terms of individual variables is not adequate, since the model must be considered as a whole. With this in mind, an analysis of the variables included in the model could be useful to extract some conclusions.

In the model considered it is possible to see that the most important set of variables is represented by the first PU ($PU_1$) since its coefficient has the highest absolute value. A three-dimensional graphical representation of the non-linear relationship between the value associated to this PU and the two corresponding input variables has been included in Fig. 2, where it is important to take into account that the input variables are scaled ($X_i \in [1,2]$). Starting with the interpretation of this $PU_1$ node, it shows that low real interest rates positively contribute to an increase of the probability to enter in a crisis from a situation of non-crisis. This result contradicts findings of some previous studies where high interest rates appear to be one of the usual triggers of a financial crisis [66,67]. However, the fact that real interest rates enter in this model with opposite sign could be pointing at an interesting result more in line with what we have seen in the current crisis: too low interest rates could be an incentive for investor's search for yield, therefore reducing the risk aversion and constituting the seed for a financial bubble that when burst could trigger a financial crisis.

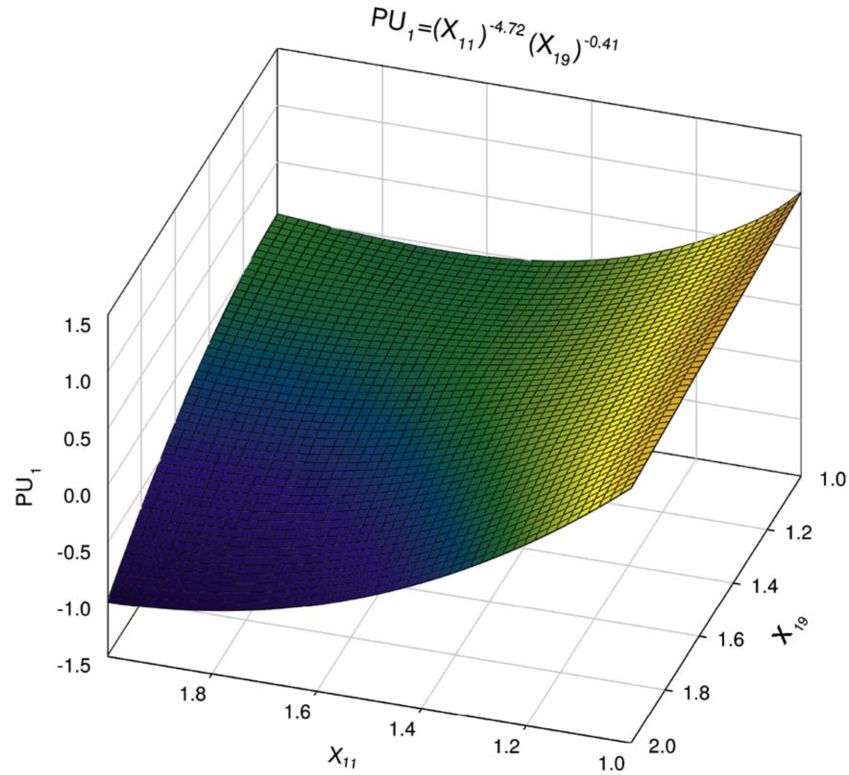


**Fig. 2.** Three-dimensional graphical representation of the non-linear relationship between the value associated with the $PU_1$ and the two corresponding input variables, in the crisis probability model given by the LRIPU* network.

The second node ($PU_2$) predicts that, when a country is in a lasting crisis (more than three years), a managed floating rate regime together with a less independent central bank and low foreign liabilities ratio reduces the probability of continuing in crisis next year. Again, this result is not fully in line with previous analysis, where it is proposed that a stronger currency policy commitment (pegged currency regime, currency board) and a more independent central bank positively contribute to overcome a crisis. It is commonly accepted that a stronger policy commitment allows a country to gain credibility (initialization of economic stabilization policies that could be unpopular), whereas an independent central bank contributes again to restore confidence in the these policies [68]. Regarding the foreign liabilities conclusions, they are more mixed since some of them justify a positive contribution to the crisis. This is due to the currency risk associated with foreign liabilities could materialize if a depreciation occurs, this being the origin of a crisis. However, others justify that foreign liabilities could be a stabilizing factor when local deposits fell. The interpretation here could be related with the interaction of the two policy variables and the situation of low foreign liabilities. When the currency risk is moderate because the volume of foreign liabilities is not too high, monetary policies that allow for certain degree of flexibility allow for a degree of discretion to adapt the currency value that could be more credible than tougher commitments.

The ability of this kind of models to capture non-linear relationships could well explain the apparent contradictory results we find. Therefore, we can extract some useful conclusions from the above results:

- Currency and monetary policies (currency regime, monetary policy design and central bank independence) are very important determinants in explaining a crisis.
- The effect of these policies is not linear regarding the probability of crisis. Depending on the circumstances (how willing the social agents are to accept short term sacrifices, climate of confidence, development of the country, etc.) and the exact policy mix, the effect of a specific tool is different.
- Among the relevant circumstances that determine the effect of the different policy tools is the length of the crisis, which in is turn related with the willingness of the people to accept policies that imply short term sacrifices in order to overcome the crisis and therefore with the credibility of certain policies.

These results should be taken with caution as, similar to other empirical studies focused in the identification of potential crisis, our model is also subjected to some weaknesses and limitations, especially as potential tools for policy-makers. First, the concentration on banking crisis in the emerging economies has influenced the type of indicators selected since the available data for these economies are typically much less complete in terms of quantity and quality. Moreover, the availability of data influences the sampling process, introducing a selection bias. A second weakness of these models is the lack of variables which capture the contagion effect, given again the lack of available information on direct linkages between banks (via interbank exposures) or indirect links (via payment systems). A third drawback is that models do not usually distinguish between periods of fragility and periods of crisis. This would be a distinction very useful for policy-makers, because it could help to identify what policies could avoid that a situation of fragility turns into a crisis.

## 5. Conclusions

In this paper we have presented several new hybrid algorithms mixing logistic regression with product unit neural networks

(PUNNs) or radial basis function neural networks (RBFNNs) for detecting and predicting banking crises. Both hybrid approaches consist in the evolutionary training of a PUNN or a RBFNN using an evolutionary algorithm (EA). Once the basis functions have been determined by the EA, a transformation of the input space is considered. This transformation is performed by adding the non-linear transformations of the input variables given by the PU or RBF functions obtained by the EA. The final model obtained is linear in these new variables together with the initial covariates. In this paper we have tested these hybrid approaches in a Financial Crisis Database, formed by macroeconomic variables of 79 countries in the periods 1981–1997 (training data) and 1998–1999 (test data), and the corresponding crisis/no crisis decision variable. The hybrid models have been shown to be very strong in the problem of bank crisis prediction. The *AUC* measure has been presented and applied to the evaluation of the different models, since it avoids the use of arbitrary cut-off probabilities, evaluating the models for all possible thresholds. The results obtained have proven the good performance of the proposed approaches, that improve the results of other existing Statistical and Artificial Intelligence techniques in the problem of detecting and predicting banking crises.

## Acknowledgements

This work has been partially supported by Universidad de Alcalá and Comunidad de Madrid under Grant no. CCG07-UAH/AMB-3993, by the TIN 2008-06681-C06-03 project of the Spanish Inter-Ministerial Commission of Science and Technology (MICYT), FEDER funds and by the P08-TIC-3745 project of the "Junta de Andalucía" (Spain). The authors also wish to thank the Editor-in-Chief, the Associate Editor and anonymous reviewers for the excellent review process carried out to our paper, with very helpful and valuable comments and suggestions.

## References

[1] Stiglitz J, Furman J. Evidence and insights from East Asia. Brookings papers on economic activity, number 2; 1998.
[2] Hanson J. Postcrisis challenges and risks in East Asia and Latin America. In: Caprio et al., editors. Financial crisis; 2006.
[3] Tsionas EG, Papadakis EN. A bayesian approach to statistical inference in stochastic DEA. Omega 2010;38(4), doi:10.1016/j.omega.2009.02.003.
[4] Delis MD. Competitive conditions in the Central and Eastern European banking systems. Omega 2010;38(4), doi:10.1016/j.omega.2008.09.002.
[5] Fiordelisi F, Molyneux P. Total factor productivity and shareholder returns in banking. Omega 2010;38(4), doi:10.1016/j.omega.2008.07.009.
[6] González-Hermosillo B. Banking sector fragility and systemic sources of fragility. International Monetary Fund, Working paper number 12; 1996.
[7] Kaminsky G, Reinhart C. The twin crises: the causes of banking and balance-of-payments problems. American Economic Review 1999;89(3):473–500.
[8] Pasiouras F, Tanna S. The prediction of bank acquisition targets with discriminant and logit analyses: methodological issues and empirical evidence. Research in International Business and Finance 2009;24(1):39–61.
[9] Domaç I, Martínez-Peria MS. Banking crises and exchange rate regimes: is there a link? The World Bank, Working paper number 2489; 2000.
[10] Eichengreen B, Arteta C. Banking crisis in emerging markets: presumptions and evidence. Center for International and Development Economics Research, University of California, Berkeley; 2000.
[11] Caprio G. Klingebiel D. Bank insolvencies: cross-country experience and policy research. The World Bank, Working paper number 1620; 1997.
[12] Liang W-Y, Huang C-C. A hybrid approach to constrained evolutionary computing: case of product synthesis. Omega 2008;36(6):1072–85.
[13] Balana S, Vratb P, Kumarc P. Information distortion in a supply chain and its mitigation using soft computing approach. Omega 2009;37(2):282–99.
[14] Trippi R, Turban E. Neural networks in finance and investing. Probus Publishing Co.; 1994.
[15] Tay F, Cao L. Application of support vector machines in financial time series forecasting. Omega 2001;29(4).
[16] Van Gestel T, Suykens J, Baestaens D, Lambrechts A, Lanckriet B, Vandaele B, et al. Financial time series prediction using least squares support vector machines within the evidence framework. IEEE Transactions on Neural Networks 12(4);2001:809–21.
[17] West D, Mangiameli P, Chen S. Control of complex manufacturing processes: a comparison of SPC methods with a radial basis function neural network. Omega 1999;27(3):349–62.
[18] Falavigna G. An analysis of the key-variables of default risk using complex systems. International Journal of Business Performance Management 2008;10(2–3):202–30.
[19] Charalambous C, Charitou A, Kaourou F. Comparative analysis of artificial neural network models: application in bankruptcy prediction. Annals of Operations Research 2000;99(1–4):403–25.
[20] Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.
[21] Durbin R, Rumelhart D. Products units: a computationally powerful and biologically plausible extension to backpropagation networks. Neural Computation 1989;1(1):133–42.
[22] Ismail A, Engelbrecht AP. Pruning product unit neural networks. In: Proceedings of the 2002 international joint conference on neural networks; 2002. p. 257–62.
[23] Bishop CM. Improving the generalization properties of radial basis function neural networks. Neural Computation 1991;3(4):579–81.
[24] Howlett R, Jain LC. Radial basis function networks 1: recent developments in theory and applications. Heidelberg: Physica-Verlag; 2001.
[25] Yao X. Evolving artificial neural networks. Proceedings of the IEEE 1999;87(9):1423–47.
[26] Yao X, Liu Y. A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 1997;8(3):694–713.
[27] Martínez-Estudillo AC, Martínez-Estudillo FJ, Hervás-Martínez C, García N. Evolutionary product unit based neural networks for regression. Neural Networks 2006;19(4):477–86.
[28] Hervás-Martínez C, Martínez-Estudillo FJ, Carbonero-Ruz M. Multilogistic regression by means of evolutionary product-unit neural networks. Neural Networks 2008;21(7):951–61.
[29] Houben A, Kakes J, Schinasi G. Toward a framework for safeguarding financial stability. International Monetary Fund, Working paper number 101; 2004.
[30] Mishkin FS. Understanding financial crises: a developing country's perspective, National Bureau of Economic Research, Working paper number 5600; 1996.
[31] Gupta P. Currency crises, banking crises and twin crises: a comprehensive review of the literature. International Monetary Fund; 1996.
[32] Caprio G, Klingebiel D. Episodes of systemic and borderline financial crises. Dataset mimeo, The World Bank; 2003.
[33] Cukierman A, Webb SB, Neyapti B. Measuring the independence of central banks and its effect on policy outcomes. The World Bank Economic Review 1992;6:353–98.
[34] Issing, O. Monetary and financial stability: is there a trade-off? In: Proceedings of the conference on monetary stability, financial stability and the business cycle, March 28–29. Basle: Bank for International Settlements, 2003.
[35] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Berlin: Springer; 2001.
[36] Witten I, Eibe F. Data mining: practical machine learning tools and techniques, 2nd ed. Los Altos, CA, Amsterdam: Morgan Kaufmann, Elsevier; 2005.
[37] le Cessie S, van Houwelingen J. Ridge estimators in logistic regression. Applied Statistics 1992;41(1):191–201.
[38] Landwehr N, Hall M, Frank E. Logistic model trees. Machine Learning 2005;59(1–2):161–205.
[39] Lenze B. On local and global sigma–pi neural networks a common link. Advances in Computational Mathematics 1994;2(4):479–91.
[40] Schmitt M. On the complexity of computing and learning with multiplicative neural networks. Neural Computation 2002;14:241–301.
[41] Hunt K, Sbarbaro D, Zbikowki R, Gawthrop P. Neural networks for control system—a survey. Automatica 1992;28:1083–112.
[42] Gomm JB, Yu DL. Selecting radial basis function network centers with recursive orthogonal least squares training. IEEE Transactions on Neural Networks 2000;11(2):306–14.
[43] Buchtala O, Klimek M, Sick B. Evolutionary optimization of radial basis function classifiers for data mining applications. IEEE Transactions on Systems Man, and Cybernetics, Part B 2005;35(5):928–47.
[44] Martínez-Estudillo FJ, Hervás-Martínez C, Gutiérrez PA, Martinez-Estudillo AC. Evolutionary product-unit neural networks classifiers. Neurocomputing 2008;72(1–2).
[45] Angeline P, Sauders G, Pollack J. An evolutionary algorithm that constructs recurrent neural networks. IEEE Transactions on Neural Networks 1994;5:54–65.
[46] García-Pedrajas N, Hervás-Martínez C, Muñoz-Perez J. Multi-objective cooperative coevolution of artificial neural networks. Neural Networks 2002;15(10):1259–78.
[47] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science 1983;220(4598):671–80.
[48] Gutiérrez PA, Hervás-Martínez C, Carbonero C, Fernández JC. Combined projection and Kernel basis functions for classification in evolutionary neural networks. Neurocomputing 2009;72(13–15):2731–42.
[49] Mahadeva L, Sterne G, editors. Monetary policy frameworks in a global context. London: Routledge; 2000.

[50] Reinhart CM, Rogoff KS. The modern history of exchange rate arrangements: a reinterpretation. The National Bureau of Economic Research, Working paper number 8963; 2002.
[51] Kuttnner KN, Posen AS., Beyond bipolar: a three-dimensional assessment of monetary frameworks. National Bank of Austria, Working paper number 52; 2001.
[52] Berg A, Borensztein E, Mauro P. An evaluation of monetary regime options for Latin America. International Monetary Fund, Working paper number 211; 2002.
[53] Carare A, Stone M. Inflation targeting regimes, International Monetary Fund. Working paper number 9; 2003.
[54] Cukierman A, Miller GP, Neyapti B. Central bank reform, liberalization and inflation in transition economies: an international perspective. Journal of Monetary Economics 2002;49:237–64.
[55] Slowinski R, Zopounidis C. Application of the rough set approach to evaluation of bankruptcy risk. International Journal of Intelligent Systems in Accounting, Finance and Management 1995;4(1):27–41.
[56] Tam KY. Neural network models and the prediction of bank bankruptcy. Omega 1991;19(5):429–45.
[57] Dimitras A, Slowinski R, Susmaga R, Zopounidis C. Business failure prediction using rough sets. European Journal of Operational Research 1999;114: 263–80.
[58] Sanchis A, Segovia-Vargas MJ, Gil JA, Heras A, Villar JL. Rough Sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem). European Journal of Operational Research 2007;181(3): 1554–73.
[59] Nanda S, Pendharkar P. Linear models for minimizing misclassification costs in bankruptcy prediction. International Journal of Intelligent Systems in Accounting, Finance & Management 2001;10:155–68.
[60] Palepu KG. Predicting takeover targets: a methodological and empirical analysis. Journal of Accounting and Economics 1986;8:3–35.
[61] Fawcett T. An introduction to ROC analysis. Pattern Recognition 2006;27:861–74.
[62] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 1997;30:1145–59.
[63] Ventura S, Romero C, Zafra A, Delgado J, Hervás-Martínez C. JCLEC: a Java framework for evolutionary computation. Soft Computing 2008;12:381–92.
[64] Alcala-Fdez J, Sanchez L, Garcia S, del Jesus JM, Ventura S, Garrell J, et al. KEEL: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing 2008;13(3):307–18.
[65] Altman EI. Corporate financial distress: a complete guide to predicting, avoiding, and dealing with bankruptcy. New York: Wiley; 1993.
[66] Hardy D, Pazarbasioglu C. Leading indicators of banking crises: was Asia different? IMF Working paper, 98/91; 1998.
[67] Bell J, Pain D. Leading indicator models of banking crises—a critical review. Financial Stability Review 2000:113–29.
[68] Hutchinson M, McDill K. Are all banking crises alike? The Japanese experience in international comparison, National Bureau of Economic Research, Working paper 7253; 1999.