

# **Study of Discrete Choice Models and Adaptive Neuro-Fuzzy Inference System in the Prediction of Economic Crisis Periods in USA**

*Eleftherios Giovanis*

*Royal Holloway University of London, Department of Economics,  
TW20 0EX, Egham, Surrey,  
United Kingdom  
(Email: [eleftherios.giovanis.2010@live.rhul.ac.uk](mailto:eleftherios.giovanis.2010@live.rhul.ac.uk))*

*Abstract:* In this study two approaches are applied for the prediction of the economic recession or expansion periods in USA. The first approach includes Logit and Probit models and the second is an Adaptive Neuro-Fuzzy Inference System (ANFIS) with Gaussian and Generalized Bell membership functions. The in-sample period 1950-2006 is examined and the forecasting performance of the two approaches is evaluated during the out-of sample period 2007-2010. The estimation results show that the ANFIS model outperforms the Logit and Probit model. This indicates that neuro-fuzzy model provides a better and more reliable signal on whether or not a financial crisis will take place.

## **I. INTRODUCTION**

The economic facts and the financial crisis during the last three years have led to doubts among the economic society and governments of which methodology must be followed and what regulations should be set up. The sub-prime mortgage crisis which took place in USA and became apparent in 2007 has led to great weakness in the financial system and the financial industry regulations. The purpose of this paper is the examination of the methodology and the models which can be applied in the prediction of crisis periods. Early-warning models exploit systematic relationships apparent in historical data between variables associated with the build-up to crises and the actual incidence of crisis. Their aim is to forewarn the policy makers about the future financial crisis and help them to take pre-emptive measures. This study proposes an alternative framework, an Adaptive Network-Based Fuzzy Inference System or Neuro-Fuzzy System (ANFIS) that can be used as an early-warning system for financial and economic crises.

The paper is organized as follows: Section II provides a briefly synopsis of previous researches of financial and economic crisis predictions. Section III reviews the methodology of binary Logit and Probit approaches and their respective advantages and disadvantages. The rest part of section III, describes the procedure and methodology of ANFIS model. Section IV

presents the data, and the research sample used in the estimations and prediction procedures, while in section V the empirical estimated and forecasting results are reported. In the final section, the general conclusions of the empirical findings are discussed.

## II. LITERATURE REVIEW

Various approaches have been developed and applied in financial, banking and currency crisis. One of these approaches is the application of Logit and Probit models (Eichengreen and Rose 1998, Demirguc-Kunt and Detragiache 1998, Frankel and Rose 1996, Glick and Rose 1998, Glick and Moreno 1999). In this approach dependent variable is a binary “crisis” variable equal to 1 if a crisis ensues and 0 otherwise. More specifically Demirguc-Kunt and Detragiache (1998) found that low GDP growth, high inflation and high interest rates might lead to economic downturns, which are the most significant variables among other macroeconomic factors. Eichengreen and Rose (1998) found that the Northern interest rates are strongly connected with the onset of banking crises in the developing countries. The methodology of Probit and Logit models is that allows for statistical testing, identifying the sign, the magnitude and the marginal distributions of the explanatory variables to the onset of crisis. On the other hand, this approach confronts the problem of misspecification errors and serial correlation.

Another approach which has been used in the crisis prediction is the “noise-to-ratio” model (Kaminsky and Reinhart 1996, Kaminsky *et al.* 1998). The “noise-to-ratio” approach, has been recently applied to currency and banking crises to examine the impact of deteriorating fundamentals and identify variables as strong potential indicators of crisis. The method works effectively when there are sharp changes between crises episodes and periods of tranquillity. Kaminsky and Reinhart (1996) examined monthly data for 23 crises, and they found that the economic downturn precedes the banking crisis, where specifically, the output tends to peak roughly eight months before a crisis take place. Kaminsky *et al.*, (1998) found that exports, the deviations of the real exchange rates from the trend, the output and equity prices are strong potential indicators providing a warning signal that a crisis will take place. Also they found that when an indicator exceeds a specific threshold, then a banking crisis might take place in the following 24 months. The advantage of this approach is that we can directly rank the possible candidate variables as potential crisis periods, but it does not allow for statistical testing, and it is impossible to examine the magnitude of each explanatory variable to the crisis phenomena.

The history of Artificial Intelligence began when McCulloch and Walter Pitts proposed a model of artificial neurons in 1943 (McCulloch and Pitts 1943). Significance of this work is that each neuron is characterised as being “on” or “off”. Switching to “on” occurred when significant number of neighbouring neurons stimulated. McCulloch and Pitts showed that any computable function could be computed by network of connected neurons (McCulloch and Pitts, 1943). Newell, Shaw and Simon developed a reasoning program called Logic Theorist. It was meant for automatic theorem proving which led the development of Information Processing Language, the first list-processing language. Next Newell *et al.* (1959) wrote the General Problem Solver (GPS). It imitated the way humans solve the problems.

During the 1960s and 1970s, success was had in building natural language understanding systems in limited domains. For example, the STUDENT program of Bobrow (1967) could

solve high school algebra problems expressed in natural language. Winograd's SHRDLU system could, using restricted natural language, discuss and carry out tasks in a simulated blocks world (Winograd 1972). CHAT-80 could answer geographical questions placed to it in natural language (Warren and Pereira 1982). Other expert systems were the 1965 Dendral programs (Feigenbaum and Buchanan 1993), which determined molecular structure from mass spectrometer data and R1 (McDermott 1980) used to configure computer systems among others. During the 1990s and the 2000s there was great growth in the sub-disciplines of Artificial Intelligence such as perception, probabilistic and decision-theoretic reasoning, planning, embodied systems, machine learning, and many other fields.

Human knowledge nowadays becomes increasingly important—we gain it from experiencing the world within which we live and use our ability to reason to create order in the mass of information, i.e., to formulate human knowledge in a systematic manner. Since we are all limited in our ability to perceive the world and to profound reasoning, we find ourselves everywhere confronted by *uncertainty* which is a result of lack of information, as lexical impression, incompleteness, in particular, inaccuracy of measurements. The other limitation factor in our desire for precision is a natural language used for describing and sharing knowledge, communication, etc. We understand core meanings of word and are able to communicate accurately to an acceptable degree, but generally we cannot precisely agree among ourselves on the single word or terms of common sense meaning. In short, natural languages are vague.

Since 1990, new approaches have been introduced in economics and finance, like computational intelligence including neural networks, fuzzy logic and genetic algorithms. We propose the scientific findings and methods of artificial intelligence because most studies have found superior results, especially in stock returns and economic data prediction than the common Logit models and Multiple Discriminant Analysis among others (Salchenberger *et al.* 1992, Coats and Fant 1993, Zhang *et al.* 1999, Fan and Palaniswami 2000, Brockett *et al.* 2006, Ni and Yin 2009, Giovanis 2010). Thus, economists and financial managers should adopt in their portfolio of research tools the artificial intelligence methods and approaches.

### III. METHODOLOGY

#### 3.1 Binary Logit and Probit Regressions

In this section, a brief description of the binary Logit and Probit models estimation is provided. The logistic distribution is defined as (Greene 2008):

$$\text{Prob}(Y = 1 | x) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \Phi(x'\beta) \quad (1)$$

Logit model can be written in a general form regression as:

$$y = \alpha + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (2)$$

where variable  $y$  is a binary dummy variable taking value 1 if the economy is on the crisis or economic recession period and value zero otherwise (no crisis period),  $x_i$  indicates the explanatory variables,  $\alpha$  is the constant,  $\beta_i$  are the regression estimators. Next the Probit model is presented (Greene, 2008), which is defined as:

$$\Phi^{-1}(p_i) = Z = \alpha + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (3)$$

where  $\Phi^{-1}(p_i)$  is the inverse cumulative distribution function (CDF) of the standard normal,  $\alpha$ ,  $\beta_i$  and  $x_i$  are defined as in (2).

One of the advantages and features of Logit and Probit models is the non-linearity of the effect of the right-hand side variables on the left-hand side variable. This is an attractive property of the model because the literature on currency crises has demonstrated that such effects are at work (Eichengreen and Rose 1998; Demircuc-Kunt and Detragrache 1998). Furthermore, there is complete freedom to choose which variables enter into the vector  $X_t$ , with the selection being based on various theories in the financial and economic crisis literature. Another positive point about the Probit and Logit approach is that it summarizes all the information in one easily interpretable number – the probability of a crisis occurring. In addition, Probit and Logit models consider all variables together and look only at the marginal contributions of each indicator; the discrete choice models disregard variables that do not contribute information that is not already captured in the other variables. Lastly, Probit and Logit lend themselves readily to statistical testing.

On the other hand there are disadvantages about this approach. Kaminsky *et al.* (1998) note that unlike the “noise-to-ratio” approach, one is unable to rank indicators on the basis of forecasting accuracy – they either enter the regression significantly, or do not. In addition, measures of statistical significance cannot distinguish whether an indicator is good at correctly calling crises, or merely sends few false signals. Lastly, Kaminsky *et al.* (1998) note that the nonlinear nature of the model makes it difficult to assess the marginal contribution of an indicator to the probability of a crisis. More specifically, measuring the marginal effect at the mean may not be appropriate for this application, as we are most concerned with what happens when the indicator variables are far from their mean.

The classification of dummy dependent variable, denoting the economic recession and expansion periods, is based on the definition by National Bureau of Economic Research (NBER), where a recession begins when the economy reaches a peak of activity. Based on the definitions of NBER the most important and conceptual measures of the economic activity is the employment rate and the domestic production. Furthermore, according to NBER, the last economic recession began in December of 2007, so this sample is included in the fourth quarter of 2007 as quarterly data are used in the analysis. The recession officially ended in June 2009, according to the Business Cycle Dating Committee of the National Bureau of Economic Research, the official arbiter of such dates. The prediction or the classification percentage is done based on the estimated coefficients from the in-sample period using as the cutoff point the value of 0.5. For the forecasting and the classification performance of the binary Logit and Probit models is:

If  $y^* > 0.5$ , then the economy is on crisis or economic recession period

If  $y^* \leq 0.5$ , then the economy is on no crisis or economic expansion period.

Variable  $y^*$  denotes the predicted values.

The main problem with this measure is that the choice of the cutoff point is arbitrary. Traditionally, a cut-off point used has been 0.5. However, there is no reason why this cutoff is the appropriate one. Cramer (2003) suggests that a more appropriate cutoff point is the sample

frequency, which is defined as the sum of the percentage of correct prediction over the sum of the sample. The bottom line is that the uncertainty about the proper choice of cut-off point is a major problem with using the percentage correctly predicted as a measure of goodness-of-fit. Furthermore, even if this cut-off point is chosen the empirical results are not changed in the current study, as the predicted values are either very close to one or zero.

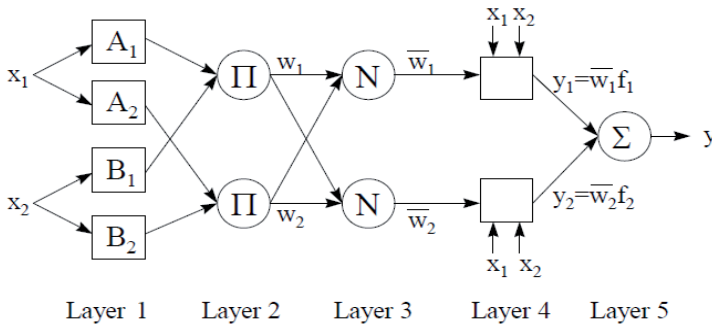
There is no basis for this procedure, finding the appropriate cut-off point; it is insoluble, and it is of no practical interest. In practice, the aim of empirical working forecasts and policy analyses is the conditional prediction of the aggregate incidence of success, and the assessment of the effect of changes in the covariates on the result.

### 3.2 Adaptive Network-Based Fuzzy Inference System (ANFIS)

The reason of proposing the neuro-fuzzy logic is because the traditional classification of one and zero can be misleading. More specifically in economic literature, economic depressions are discriminated from economic recessions. More precisely, the current crisis of 2007 and the crisis of 1929-1933 are considered as economic depressions, so the dummy binary crisis variable takes value 1. However, the crises of 1973-1975 and 1979-1981, which are owned to the petroleum crisis, are considered as economic recessions and not depressions, thus the dummy variable takes value 1. At this point, it should be noticed that the effects of economic recessions and depressions might have a significant different impact in gross domestic product and unemployment rate among other macroeconomic and microeconomic factors. Therefore, a better classification could be, for example, 0.8-1 for economic depressions and 0.6-0.8 for economic recessions. On the other hand the post-economic recovery period is very different from an economic expansion period, where the dummy variable takes the value 0. A post-economic recovery period can be characterized by medium gross domestic product growth rate and medium or even high unemployment rate, while an economic expansion period can be characterized by a high gross domestic product growth rate and low unemployment rate. For this reason, the classification of one and zero can be sometimes misleading.

Jang (1993) and Jang and Sun (1995) introduced the adaptive network-based fuzzy inference system (ANFIS). This system makes use of a hybrid learning rule to optimize the fuzzy system parameters of a first order Sugeno system. An example of a two input with two rules first order Sugeno system can be graphically represented by *Figure 1*.

*Figure 1: Example of ANFIS Architecture for a Two-Input, Two-Rule First-Order Sugeno Model*



where the consequence parameters  $p$ ,  $q$ , and  $r$  of the  $n$ th rule contribute through a first order polynomial of the form:

$$f_n = p_n x_1 + q_n x_2 + r_n \quad (4)$$

The ANFIS architecture is consisted of two trainable parameter sets, the antecedent membership function parameters and the polynomial consequent parameters  $p, q, r$ . The ANFIS training paradigm uses a gradient descent algorithm to optimize the antecedent parameters and a least squares algorithm to solve for the consequent parameters. Because it uses two very different algorithms to reduce the error, the training rule is called a hybrid. The consequent parameters are updated first using a least squares algorithm, and the antecedent parameters are then updated by backpropagating the errors that still exist. The ANFIS architecture consists of five layers with the output of the nodes in each respective layer represented by  $O_i^l$ , where  $i$  is the  $i^{th}$  node of layer  $l$ . Two inputs are selected, the gross domestic product growth rate and the unemployment rate and three linguistic terms {low, medium, high} are defined. Therefore, the rules are nine, and these are:

*IF GDP is low AND unemployment rate is low THEN  $f_1 = p_1 x_1 + q_1 x_2 + r_1$*   
*IF GDP is low AND unemployment rate is medium THEN  $f_2 = p_2 x_1 + q_2 x_2 + r_2$*   
*IF GDP is low AND unemployment rate is high THEN  $f_3 = p_3 x_1 + q_3 x_2 + r_3$*   
*IF GDP is medium AND unemployment rate is low THEN  $f_4 = p_4 x_1 + q_4 x_2 + r_4$*   
*IF GDP is medium AND unemployment rate is medium THEN  $f_5 = p_5 x_1 + q_5 x_2 + r_5$*   
*IF GDP is medium AND unemployment rate is high THEN  $f_6 = p_6 x_1 + q_6 x_2 + r_6$*   
*IF GDP is high AND unemployment rate is low THEN  $f_7 = p_7 x_1 + q_7 x_2 + r_7$*   
*IF GDP is high AND unemployment rate is medium THEN  $f_8 = p_8 x_1 + q_8 x_2 + r_8$*   
*IF GDP is high AND unemployment rate is high THEN  $f_9 = p_9 x_1 + q_9 x_2 + r_9$*

where  $GDP$  denotes the gross domestic product growth rate. Basically, there are two types of fuzzy set operation that are usually used in the antecedent rules. These are the *AND* and *OR*. Mathematically, the *AND* operator can be realized using *Min* or *Product* operation while *OR* can be realized using *Max* or *Algebraic sum* operator. The *AND* operator is selected, and the *Product* operator is chosen instead of *Min* in order to avoid monotonic results. It should be noted that more inputs with more linguistic terms can be obtained, but this is not necessary means that it will improve the forecasting performance, as also taking two inputs the computation time is significant reduced. On the other hand, more inputs can be very useful to examine various factors under specific rules and investigate the causes of financial crises and there is the possibility to avoid them. Because there are three linguistic terms and two inputs in the case examined, the steps for ANFIS system computation, which are captured in *Figure 1*, are:

In the first layer, the fuzzy layer, the membership grades are generated.

$$O_i^1 = \mu_{A_i}(x_1), \mu_{B_i}(x_2) \quad (5)$$

where  $x_1$  and  $x_2$  are the inputs. In layer 2, the production layer, the firing strengths or weights are generated

$$O_i^2 = w_i = \prod_{j=1}^m (\mu_{A_i}(x_1), \mu_{B_i}(x_2)) = \text{and method}(\mu_{A_i}(x_1), \mu_{B_i}(x_2)) = (\mu_{A_i}(x_1) \cdot \mu_{B_i}(x_2)) \quad (6)$$

In layer 2 the *AND* rule is used, therefore the *product* operator is obtained. In layer 3, the normalised layer, the firing strengths are normalized. Because there are nine rules it will be:

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2 + w_3 + \dots + w_8 + w_9} \quad (7)$$

In layer 4, the defuzzy layer, the rule outputs based on the consequent parameters are calculated.

$$O_i^4 = y_i = \overline{w_i} f = \overline{w_i} (p_i x_1 + q_i x_2 + r_i) \quad (8)$$

Layer 5 is the total output layer and is defined as the summation of all inputs. It can be expressed as:

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i \overline{w_i} f_i}{\sum_i \overline{w_i}} \quad (9)$$

In the last layer the consequent parameters can be solved using a least square algorithm as:

$$Y = X \cdot \theta \quad (10)$$

where X is the matrix

$$X = [w_1 x + w_1 + w_2 x + w_2 + \dots + w_9 x + w_9] \quad (11)$$

and  $\theta$  is a vector of unknown parameters as:

$$\theta = [p_1, q_1, r_1, p_2, q_2, r_2, \dots, p_9, q_9, r_9]^T \quad (12)$$

where  $T$  indicates the transpose. Because the normal least square method leads to a singular inverted matrix, the Singular Value Decomposition (SVD) method with Moore-Penrose pseudoinverse of a matrix is used (Moore 1920, Penrose 1955, Petrou and Bosdogianni 2000). Other methods that can be used to eliminate the particular problem are Lower Triangular, Upper Triangular and Orthogonal decomposition (QR) among others. More specifically the Singular Value Decomposition (SVD) procedure is:

$$X = USVT \quad (13)$$

The singular values in  $S$  are positive and arranged in decreasing order. Their magnitude is related to the information content of the columns of  $U$  -principle components- that span  $X$ . Therefore, to remove the noise effects on the solution of the weight matrix, we simply remove the columns of  $U$  that correspond to small diagonal values in  $S$ . The weight matrix is then solved for using the following:

$$\theta = VS^{-1}U^T Y \quad (14)$$

For the first layer and relation (5), the Gaussian and Generalized Bell membership functions are used. Additional membership functions can be used as the triangular, the trapezoidal and the log-sigmoid among others. In the previous part, the computation procedure for the consequent



parameters by using the least squares algorithm with Moore-Penrose pseudoinverse matrix has been described. The next step is the description of the training procedure for the antecedent parameters with the error back-propagation algorithm. The symmetrical Gaussian membership function is defined as:

$$\mu_{ij}(x_j; c_{ij}, \sigma_{ij}) = \exp\left(-\frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right) \quad (15)$$

where  $c_{ij}$  is the center parameter and  $\sigma_{ij}$  is the spread parameter. The next function is Generalized Bell (Gbell) such as:

$$\mu_{ij}(x_{ij}; a_{ij}, b_{ij}, c_{ij}) = \exp\left(-\left(\frac{x_{ij} - c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right) \quad (16)$$

where  $c$  locates the center of the curve and parameters  $a$  and  $b$  vary the width of the curve. The error back-propagation algorithm with the simple steepest descent method is used in order to find the optimized antecedent parameters (Tsoukalas and Uhrig 1996, Denai *et al.* 2004, Khan *et al.* 2010). The antecedent parameter  $c$  update is:

$$c_{ij}(n+1) = c_{ij}(n) - \frac{\eta_c}{p} \cdot \frac{\partial E}{\partial c_{ij}} \quad (17)$$

where  $\eta_c$  is the learning rate for the parameter  $c_{ij}$ ,  $p$  is the number of patterns and  $E$  is the error function which is:

$$E = \frac{1}{2}(y - y^t)^2 \quad (18)$$

where  $y^t$  is the target-actual and  $y$  is the ANFIS output variable. The chain rule in order to calculate the derivatives used to update the membership function parameters are:

$$\frac{\partial E}{\partial c_{ij}} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial y_i} \cdot \frac{\partial y_i}{\partial w_i} \cdot \frac{\partial w_i}{\partial \mu_{ij}} \cdot \frac{\partial \mu_{ij}}{\partial c_{ij}} \quad (19)$$

The partial derivatives for two inputs are derived below:

$$\frac{\partial E}{\partial y_{ij}} = y - y^t = e \quad (20)$$

For the output is:

$$y = \sum_{i=1}^n y_i, \text{ hence it will be } \frac{\partial y}{\partial y_i} = 1 \quad (21)$$

$$\sum_{i=1}^n w_i, \text{ hence it will be } \frac{\partial y_i}{\partial w_i} = \frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^n w_i} \quad (22)$$

$$w_i = \prod_{j=1}^m \mu_{A_{ji}}, \text{ then it will be } \frac{\partial w_i}{\partial \mu_{ji}} = \frac{w_i}{\mu_{ji}} \quad (23)$$



The last partial derivative,  $\frac{\partial \mu_{ij}}{\partial c_{ij}}$ , depends on the membership function we examine. The update equations for antecedent  $c_{ij}$ , and  $\sigma_{ij}$  parameters of Gaussian function are:

$$c_{ij}(n+1) = c_{ij}(n) - \eta_c \cdot e^{\frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^n w_i}} \cdot \left( \frac{(x_{ij} - c_{ij})}{\sigma_{ij}^2} \right) \mu_{ij}(x_j) \quad (24)$$

$$\sigma_{ij}(n+1) = \sigma_{ij}(n) - \eta_\sigma \cdot e^{\frac{(p_i x_1 + q_i x_2 + r_i) - y}{\sum_{i=1}^n w_i}} \cdot \left( \frac{(x_{ij} - c_{ij})^2}{\sigma_{ij}^3} \right) \mu_{ij}(x_j) \quad (25)$$

Similarly update equations can be derived for Generalized Bell function. For the parameter initialization of Gaussian membership function, the values of GDP and unemployment rate in specific intervals are taken. In *Table 1* the procedure is reported.

Based on the samples of *Table 1* the respective average and standard deviations for each membership function in each rule are obtained. GDP in *Table 1* denotes the gross domestic product. For the generalized bell function and the parameters  $c$  as initial values the average value is taken, as in the case of Gaussian function. For parameter  $b$ , 1.1 is set up as the initial value in all cases, while for parameter  $a$  the following formula (Denai *et al.* 2004) is obtained:

$$\alpha = \frac{range}{mfs * 2} \quad (26)$$

where the *range* is the well known statistical measure defined as the difference between the maximum and minimum value. Therefore, the range of the input data is obtained, while *mfs* is the number of membership functions. Because each rule has two parameters plus the constant term, there are  $3^3=27$  parameters. The learning rates for all parameters are set up at 0.1 and the number of the maximum epochs at 50.

*Table 1: Samples for Set Up the Initial Parameters of Gaussian Membership Function*

First Rule Low	Second Rule Medium	Third Rule High
If GDP <0	If GDP>0 and GDP<0.02	If GDP>0.02
If unemployment rate< 4.5	If unemployment rate >4.5 and unemployment rate <6	If unemployment rate >6

The ANFIS is a class of adaptive networks, which are functionally equivalent to the fuzzy inference system (FIS), which is a popular computing framework based on the concepts of fuzzy set theory, fuzzy if-then rules, and fuzzy reasoning. It is a very powerful approach for building complex and nonlinear relationship between a set of input and output data. Another advantage of ANFIS is the fact that it can be trained with no need for the expert knowledge usually required for the standard fuzzy logic design, and both numerical and linguistic

knowledge can be combined into a fuzzy rule base by employing fuzzy methods. The fuzzy membership functions can be tuned optimally by using optimization algorithms, as the error back-propagation algorithm used in this study. Other advantages of the ANFIS include its nonlinear ability, its capacity for fast learning, and its adaptation capability. Generally, the strength of ANFIS is that it uses the neural network's ability to classify data and finds patterns. Moreover, it develops a fuzzy expert system that is more transparent to the user and also less likely to produce memorization errors than a neural network. Furthermore, ANFIS keeps the advantages of a fuzzy expert system, while removing or at least reducing the need for an expert (Tsoukalas and Uhrig 1996).

The artificial intelligence has its advantages and disadvantages, including ANFIS as well. One of the problems with the ANFIS design is that a large amount of training data might be required to develop an accurate system, depending always on the research study. Another disadvantage, which has been discussed, is the fact that there is only one output from an ANFIS. Thus, ANFIS can only be applied to prediction tasks or the approximation of nonlinear function where there is only one output. It should be mentioned that in the current study this is not necessary a drawback, as one output is the main part of interest. Another disadvantage that it should be highlighted is the determination of membership functions shape and location for each fuzzy variable involved with fuzzy logic. Therefore, the efficiency of the ANFIS approach depends on the estimated parameters of premise and consequent parts. Moreover, the membership functions associated with each input and output node cannot be adjusted; only the values of the rules can be.

Furthermore, another drawback has to do with the optimization algorithm used in this study. More specifically, the error back-propagation algorithm is based on gradient descent method which according to the error surface tries to find the best weight and bias composition in order to minimize the network error, but there are some disadvantages like slow converges, lack of robustness and inefficiency (Rumelhart *et al.* 1986). Thus, alternative methods have been proposed, which might be able to improve the training process, like Levenberg-Marquardt (LM) that is a method which is based both Gauss-Newton nonlinear regression and gradient steepest descent method. At this point it should be noticed that the error main disadvantage of error back-propagation and LM algorithm, among others, is that are derivative based procedures, which means that sometimes are unable to find the optimum solution, minimum or maximum. For the solution of this possible problem, genetic algorithms have been proposed, which are free-derivative procedures (Mitchell, 1996).

#### IV. DATA AND RESEARCH DESIGN

The period 1950-2006 is set up as the estimation or the in-sample period. Then all the models are tested during the out-of-sample period 2007-2010 in order to compare their predicting performance. The data come directly from the *Federal Reserve Bank of St. Louis* and the *National Bureau of Economic Research*. The choice of variables is based on various research papers and studies (Demirguc-Kunt and Detragrache, 1998, Eichengreen and Rose 1998, Glick and Moreno 1999, Fioramanti 2006). Furthermore, the decision of the variable choice is based on National Bureau of Economic Research (NBER), which defines real GDP, real income,

unemployment rate, industrial production and retail sales as the most important factors defining the economic activity in US economy. Moreover, all the candidate variables are tested and the most significant among them are chosen. Specifically, the same variables are used in Logit and Probit regressions.

Table 2: Variables used in Estimated Regressio

industrial production	unemployment rate
inflation rate	total investments at all commercial banks
total borrowings of depository institutions from federal reserve system	oil prices
interest rates of 3-monthly treasury bills	bank prime loan rate
public debt	balance of accounts
total loans at all commercial banks	gdp growth

## V. EMPIRICAL RESULTS

In *Table 3* the estimation results of binary Logit and Probit regressions are reported. It can be noticed that an increase in the *bank prime loan rate* and the *balance of accounts* leads to an increase of the probability of crisis occurrence, while an increase on remaining variables leads to decrease of the probability. More specifically, the strongest positive effects are exhibited by the *bank prime loan rate* as shown in *Table 3*. Therefore, one concluding remark is that might have led to the current financial crisis is the *bank prime loan rate*. More precisely, this factor increased from 5.25 in 2005 to 8.25 in 2007, which was relatively high. Additionally, the *bank prime loan rate* has been decreased at 3.25 through 2009 and 2010, and this probably might be one of the factors among others, which can lead to the growth of the US economy.

On the other hand, if the *industrial production* and *GDP growth* increase, then the risk of financial crisis occurrence is decreased. It should be noted that the marginal effect of *GDP growth* in Logit regression might be wrong. The reason is that *GDP growth* is expected to lead in significant decreasing of the probability of a crisis occurrence based on the findings of other studies. However, the marginal effect, provided in *Table 3*, is almost zero. This problem is eliminated with Probit regression, where a strong negative effect is found, thus the last model is more appropriate.

In *Table 3* the likelihood-ratio chi-square is displayed. More specifically the likelihood-ratio (LR) chi-square statistic is defined as (Agresti 2002):

$$LR_k(x^2) = 2 \sum f \cdot \ln\left(\frac{f}{f_i}\right) \quad (27)$$

where  $f$  and  $f_i$  indicate respectively the observed and the expected frequency of crisis and no-crisis periods and  $k$  denotes the number of independent variables in the logistic regression. Using the likelihood-ratio statistic of (27) the following hypotheses are tested:

$H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$ , indicating that Logit or Probit regression is statistically insignificant, against the alternative hypothesis  $H_1 : \beta_0 \neq \beta_1 \neq \dots \neq \beta_k \neq 0$ , indicating that Logit or Probit regression is statistically significant.

Likelihood-ratio chi-square statistic follows chi-square distribution with  $k$  degrees of freedom. In the case of *Table 3* there are four independent variables, so it is  $k=4$ . The value of  $LR_{(4)}$  chi-square is 118.51 and 126.62 respectively and it is higher than the critical values of  $\chi^2_{(4)}$  at  $\alpha=0.10$ ,  $\alpha=0.05$  and  $\alpha=0.01$ , which are respectively 7.779, 9.488 and 13.277. Moreover, the  $p$ -values of  $LR_{(4)}$  chi-square are zero and the null hypothesis is rejected, indicating that the Logit and Probit regressions are statistically significant.

*Table 3: Results of Binary Logit and Probit Regressions for In-Sample Period*

Model	Variable	Estimated Coefficients	Standard Errors	z-statistics	P-Value	Marginal Effects	Diagnostic Tests
Logit	Constant	-0.464	0.545	-0.85	0.395		Pseudo $R^2$ 0.5206
	Industrial Production	-7.653*	1.696	-4.51	0.000	0.00047	$LR(4) \chi^2$ 118.51 [0.000]
	Bank Prime Loan Rate	0.189**	0.073	2.56	0.010	0.010	Log-Likelihood -35.141
	Balance of Accounts	0.008*	0.002	3.14	0.002	0.002	
	GDP Growth	-129.425*	37.072	-3.49	0.000	0.000	
Probit	Constant	-0.341	0.299	-1.14	0.253		Pseudo $R^2$ 0.5506
	Industrial Production	-4.031*	0.843	-4.78	0.000	-0.3776	$LR(4) \chi^2$ 126.62 [0.000]
	Bank Prime Loan Rate	0.110*	0.040	3.20	0.001	0.0108	Log-Likelihood -51.669
	Balance of Accounts	0.004*	0.001	3.20	0.002	0.0004	
	GDP Growth	-71.397*	19.888	-3.59	0.000	-6.6882	

Note: \* denotes statistically significant in 0.01 level, \*\* denotes statistically significant in 0.05 level

In *Table 4* the estimated fuzzy antecedent and consequent parameters with Gaussian membership function are reported. Based on *Table 4*, various estimations for different rules can be captured. For example, if the gross domestic growth is low and unemployment rate is medium, then it will be:

$$f_2 = p_2x_1 + q_2x_2 + r_2 = 2.1776x_1 + 0.1158x_2 - 2.0985$$

In a similar way the other rules can be examined. The estimations and procedures are similar with Generalized Bell function, thus the estimations are not provided.

It can be concluded that ANFIS approach with more inputs can be more useful in the financial crisis predictions, under specific conditions and rules when many factors are available, as the bank prime loan rate, exchange rates, public debt and balance of accounts among many others. The purpose of this paper is to present a model which can be very efficient in forecasting the crisis periods, while further research can be made in order to build a system with many inputs, avoiding in a successful manner possible crises in the future, as financial and economics crises, food crises even and war or political crises and instabilities. Additionally, in *Table 4* the estimations of antecedent parameters are provided. More specifically, parameters  $c_1$ ,  $c_2$  and  $c_3$  are the center parameters for gross domestic product corresponding in linguistic terms {low, medium, high} respectively, while parameters  $c_4$ ,  $c_5$  and  $c_6$  are the respective center parameters for unemployment rate. The same definition is followed for base parameters  $\sigma$ .

*Table 4: Estimated Fuzzy Antecedent and Consequent Parameters with Gaussian Function*

<i>Consequent parameters</i>								
$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$	$p_9$
-1.8755	2.1776	-5.7874	-7.6787	6.6937	-1.9796	-3.9859	-5.5927	1.0928
$q_1$	$q_2$	$q_3$	$q_4$	$q_5$	$q_6$	$q_7$	$q_8$	$q_9$
0.2220	0.1158	-0.2909	-0.0940	-0.3252	1.4562	0.2525	-0.4999	0.4138
$r_1$	$r_2$	$r_3$	$r_4$	$r_5$	$r_6$	$r_7$	$r_8$	$r_9$
0.6209	-2.0985	4.7834	-0.9467	3.4275	-9.7193	-2.8258	4.9072	-2.0257
<i>Antecedent parameters</i>								
$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$			
0.0279	0.0132	-0.0062	3.7604	5.6557	7.4601			
$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\sigma_6$			
0.0062	0.0047	0.0081	1.2943	3.6030	1.2995			

In *Table 5* the correct prediction percentage or the forecasts of the three models are reported during the in-sample period 1950-2006, while in *Table 6* the forecasts of the same models for the out-of sample period 2007-2010 are provided. It is observed that the forecasting performance of Logit model during the in-sample period is relatively significant, while in the out-of sample is poor. More specifically, with Logit model the correct percentage prediction, during the in-sample period, is 70.00 per cent for the crisis periods and 95.74 for the no-crisis periods.

Concerning Probit model it is observed that during the in-sample period 1950-2006 slightly outperforms Logit model. In the out-of sample period 2007-2010 Probit predicts 71.42 and 88.88 per cent correct the crisis and no crisis periods respectively. The overall percentage, which is 81.25, is significant higher to the predicted overall percentage of Logit model, which is 68.75. Additionally, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values in the case of Probit model, are well below the respective values of Logit, especially in the out-of-sample period.

*Table 5: Prediction Results for In-Sample Period*

Model				
Logit	Crisis	28	12	70.00
	No Crisis	8	180	95.74
	Overall percentage			91.22
	MAE	0.1233	RMSE	0.3511
Probit	Crisis	30	10	75.00
	No Crisis	6	182	96.81
	Overall percentage			92.98
	MAE	0.1124	RMSE	0.3352
ANFIS Gaussian	Crisis	36	4	90.00
	No Crisis	23	165	87.76
	Overall percentage			88.15
	MAE	0.1370	RMSE	0.3701
ANFIS Gbell	Crisis	34	6	85.00
	No Crisis	18	170	90.42
	Overall percentage			89.47
	MAE	0.1358	RMSE	0.3685

*Table 6: Prediction Results for Out-of-Sample Period*

Model				
Logit	Crisis	4	3	57.14
	No Crisis	2	7	77.77
	Overall percentage			68.75
	MAE	0.2500	RMSE	0.500
Probit	Crisis	5	2	71.42
	No Crisis	1	8	88.88
	Overall percentage			81.25
	MAE	0.1556	RMSE	0.3887
ANFIS Gaussian	Crisis	7	0	100.00
	No Crisis	3	6	66.67
	Overall percentage			81.25
	MAE	0.0975	RMSE	0.3122
ANFIS Gbell	Crisis	7	0	100.00
	No Crisis	2	7	77.77
	Overall percentage			87.50
	MAE	0.0850	RMSE	0.2915

In *Tables 5* and *6* the forecasting results of ANFIS with Gaussian and Gbell membership functions for in-sample and out-of-sample periods are reported. It is observed that ANFIS with Gaussian membership function correctly classifies the crisis periods at 90.00 and 100.00 per cent during the in-sample and out-of-sample period respectively. This indicates that the forecasting performance of ANFIS for crisis periods is significant superior to Logit and Probit models. On the contrary, ANFIS with Gaussian membership function predicts correctly the no-crisis periods at 87.76 and 66.67 per cent during the in-sample and out-of-sample period respectively, significant lower to Logit and Probit, indicating that ANFIS is not very efficient for the prediction of no-crisis periods. On the other hand, the ANFIS with Generalized Bell function classifies correctly the crisis periods at 85.00 and 100.00 per cent during the in-sample and out-of-sample period respectively, while the correct predicted percentages for the no-crisis periods are 90.42 and 77.77 per cent.

It should be mentioned that the correct percentage of ANFIS is low for the no-crisis periods, as it gives a warning signal through 2010, while based on NBER the crisis has officially ended in June of 2009. However, even if the crisis has ended in 2009, US economy faced a very slow pace of economic recovery after this period, according to various macroeconomic factors and especially unemployment rate, which was seemed to be more like an economic stagnation.

## VI. CONCLUSIONS

This study has analyzed and examined two approaches in financial crisis prediction modeling. The first approach was Logit and Probit models. The forecasting performance of Logit regression is poor in the out-of sample period, while it is more significant during the in-sample period. Probit regression exhibits higher forecasting performance than Logit in both in-sample and out-of sample periods.

On the other hand, ANFIS models with Gaussian and Generalized bell membership functions have been estimated. The conclusion is that ANFIS with both fuzzy membership functions examined, and especially with Gbell membership function, outperforms significant the discrete choice models, Logit and Probit, in the out-of-sample period, which period is of much more significant interest, as it can be observed by the lower RMSE and MAE values, as well as, by the percentage of correct prediction.

This paper suggests the practice of fuzzy logic and neural networks plus the examination in other fields of artificial intelligence for further study and research. Furthermore, this paper suggests the adaptation of this approach by governments, central banks and the responsible personnel for the economic policy plans. Furthermore, ANFIS can be examined with additional fuzzy membership functions, as the trapezoidal, the triangular and the S-shaped function among others. Moreover, genetics can be used, as an alternative optimization algorithm, instead of error back-propagation used in this study.

## REFERENCES

- Agresti, A. (2002). *Categorical Data Analysis*, Second Edition. NJ: John Wiley & Sons.  
 Bobrow, D.G. (1967). Natural language input for a computer problem solving system, in: M. Minsky (ed.), *Semantic Information Processing*. Cambridge MA: MIT Press: 133-215.



- Brockett, P.L., L.L. Golden, J. Jang, and C. Yang (2006). A comparison of neural network, statistical methods and variable choice for life insurers' financial distress prediction, *Journal of Risk and Insurance*. 73: 397-419.
- Coats, P.K. and F.L. Fant (1993). Recognizing financial distress patterns using a neural network tool, *Financial Management*. 22: 142-155.
- Cramer, J.S. (2003). *Logit Models from Economics and Other Fields*. Cambridge: Cambridge University Press.
- Demirguc-kunt, A. and E. Detragiache (1998). The Determinants of Banking Crises in Developing and Developed Countries, IMF Staff Papers. 45: 81-109.
- Denai, M.A., F. Palis and A. Zeghib (2004). ANFIS Based Modelling and Control of Non-Linear systems: A Tutorial, *IEEE Conference on Systems, Man, and Cybernetics*. 4: 3433-3438.
- Eichengreen, B. and A.K. Rose (1998). Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking Crises. NBER Working Papers 6370. National Bureau of Economic Research, Cambridge.
- Fan, A. and M. Palaniswami (2000). A new approach to corporate loan default prediction from financial statement. in *Proceedings Computational Finance/Forecasting Financial Markets Conference CF/FFM-2000*. London, UK, May.
- Feigenbaum, E.A. and B.G. Buchanan (1993). DENDRAL and Meta-DENDRAL: roots of knowledge systems and expert systems applications, *Artificial Intelligence*. 9:233-240
- Fioramanti, M. (2006). Predicting Sovereign Debt Crises Using Artificial Neural Networks: A Comparative Approach, Working paper 72. Istituto Di Studi e Analisi Economica. Roma.
- Frankel, J. and A.K. Rose (1996). Currency Crashes in emerging Markets: An Empirical Treatment, International Finance Discussion Papers 534. Board of Governors of the Federal Reserve System. Washington. D.C.
- Glick, R. and A.K. Rose (1998). Contagion and trade: Why currency crises are regional?, NBER, Working Papers 6806. National Bureau of Economic Research. Cambridge. MA.
- Glick, R. and R. Moreno (1999). Money and Credit, Competitiveness and Currency Crises in Asia and Latin America, Federal Reserve Bank of San Francisco papers BP99-01. Center for Pacific Basin Money and economic Studies.
- Giovanis, E. (2010). Application of Logit Model and Self-Organizing Maps (SOM) for the prediction of Financial Crisis Periods in US Economy, *Journal of Financial Economic Policy*. 2: 98-125.
- Greene, W.H. (2008). *Econometric Analysis*, Sixth Edition, NJ: Prentice Hall
- Jang, J.-S.R. (1993). ANFIS: Adaptive-Network-based Fuzzy Inference Systems, *IEEE Transactions on Systems, Man, and Cybernetics*. 23: 665-685.
- Jang, J.-S. R. and C. T. Sun (1995). Neuro-fuzzy Modeling and Control, *Proceedings of the IEEE*. 83: 378-406.
- Kaminsky, L.G. and C.M. Reinhart (1996). The Twin Crises: The Causes of Banking and Balance of Payments Problems, Federal Reserve Board Discussion Papers 544, Board of Governors of the Federal Reserve System. Washington. D.C.
- Kaminsky, L.G., S. Lizondo, and C.M. Reinhart (1998). Leading Indicators of Currency Crises, IMF Staff Papers. 45: 1-48.
- Khan, L., S. Anjum, and R. Bada (2010). Standard Fuzzy Model Identification using Gradient Methods, *World Applied Sciences Journal*. 8: 01-09.
- McCulloch, W. and W. Pitts (1943). A Logical Calculus of Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*. 5:115-133.
- McDermott, J. (1980). R1: an expert in the computing system domain, In *Proceedings of the National Conference on Artificial Intelligence*. 269-271.
- Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society*. 26: 394-39.

- Newell, A., J.C. Shaw, and H.A. Simon (1959). Report on a general problem-solving program, *Proceedings of the International Conference on Information Processing*. 256-264.
- Ni, H. and H. Yin (2009). Exchange rate prediction using hybrid neural networks and trading indicators, *Neurocomputing*. 72: 2815-2823.
- Penrose, R. (1955). A generalized inverse for matrices, *Proceedings of the Cambridge Philosophical Society*. 51: 406-413.
- Petrou, M. and P. Bosdogianni (2000). *Image Processing: The Fundamentals*, UK: John Wiley & Sons.
- Rumelhart, D.E., G.E. Hinton, and R.J. Williams (1986). *Learning internal representations by error propagation, Explorations in the microstructure of cognition*, Volume 1. Foundations. Cambridge, MA: MIT Press.
- Salchenberger, L., E. Cinar, and N. Lash (1992). Neural networks: A new tool for predicting thrift failures, *Decision Sciences*. 23: 899-916.
- Tsoukalas, L.H. and R.E. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*, First Edition. NY: John Wiley & Sons.
- Warren, D.H.D. and F.C.N. Pereira (1982). An efficient easily adaptable system for interpreting natural language queries. *Computational Linguistics*. 8: 110-122.
- Winograd, T. (1972). *Understanding Natural Language*. NY: Academic Press
- Zhang, G., Y. Hu, and E. B. Patuwo (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, *European Journal of Operation Research*. 116: 16-32.

