Application of Adaptive Network-Based Fuzzy Inference System in Macroeconomic Variables Forecasting

E. Giovanis

Abstract— In this paper we apply an Adaptive Network-Based Fuzzy Inference System (ANFIS) with one input, the dependent variable with one lag, for the forecasting of four macroeconomic variables of US economy, the Gross Domestic Product, the inflation rate, six monthly treasury bills interest rates and unemployment rate. We compare the forecasting performance of ANFIS with those of the widely used linear autoregressive and nonlinear smoothing transition autoregressive (STAR) models. The results are greatly in favour of ANFIS indicating that is an effective tool for macroeconomic forecasting used in academic research and in research and application by the governmental and other institutions

Keywords—Linear models, Macroeconomics, Neuro-Fuzzy, Non-Linear models

I. INTRODUCTION

FUZZY logic is an effective rule-based modelling in soft computing, that not only tolerates imprecise information, but also makes a framework of approximate reasoning. The disadvantage of fuzzy logic is the lack of self learning capability. The combination of fuzzy logic and neural network can overcome the disadvantages of the above approaches. In ANFIS, is combined both the learning capabilities of a neural network and reasoning capabilities of fuzzy logic in order to give enhanced prediction capabilities. ANFIS has been used by many researchers to forecast various time-series comparing with Autoregressive (AR) and Autoregressive Moving Average (ARMA) models finding superior results in favour of ANFIS [1]-[3]. On the other hand empirical analysis in macroeconomics as well as in financial economics is largely based on times series. This approach allows the model builder to use statistical inference in constructing and testing equations that characterize relationships between economic variables. There is a few number of researches made in the field of macroeconomics prediction with ANFIS, where artificial intelligence procedures are still not used by the national statistical services and central banks in many countries, while conventional econometric modelling is still in use, especially the Autoregressive models. Their argument is that economics and finance are described better by statistical modelling and properties. This is not absolute necessary,

Eleftherios Giovanis is with the Royal Holloway University of London, department of Economics, Egham, Surrey TW20 0EX, UK, e-mail: giovanis@freemail.gr, Eleftherios.Giovanis.2010@live.rhul.ac.uk
.

because macroeconomic and financial times-series can be characterized by non-linearities, which might be more efficient described by neural networks. Additionally, economics and finance are social sciences, so imprecision of the human behaviour is widely present, which can be described by fuzzy logic. For this reason we propose the neuro-fuzzy approach.

In section II we present the methodology of stationarity and unit root tests, as well as the estimating and forecasting procedure of the models examined. In section III the frequency and the type of data are described. In section IV the estimated and forecasting results are reported, while in the last section the concluding remarks of this study and some proposals are presented.

II. METHODOLOGY

A. Unit Root and Stationary Tests

It is possible that the variables are not stationary in the levels, but probably are in the first or second differences. To be specific we confirm this assumption by applying Augmented Dickey-Fuller-ADF [4] and KPSS stationary test [5]. The ADF test is defined from the following relation:

$$\Delta y_t = \mu + \gamma y_{t-1} + \phi_1 \Delta y_{t-1} + ... + \phi_p \Delta y_{t-p} + \beta t + \varepsilon_t \quad (1)$$

, where y_t is the variable we examine each time. In the right hand of (1) the lags of the dependent variable are added with order of lags equal with p. Additionally, (1) includes the constant or drift μ and trend parameter β. The disturbance term is defined as ε_t. In the next step we test the hypotheses:

H_0: φ=1, β=0 => $y_t \sim I(0)$ with drift

against the alternative

H_1: |φ|<1   => $y_t \sim I(1)$ with deterministic time trend

The KPSS statistic is then defined as:

$$KPSS = T^{-2} \sum_{t=1}^{T} s^2{}_t / \hat{\sigma}^2 (p) \cdot \quad (2)$$

, where T is the number of sample and $\hat{\sigma}^2(p)$ is the long-run

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

variance of ε_t and can be constructed from the residuals ε_t as:

$$\hat{\sigma}^2(p) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + \frac{2}{T}\sum_{t=1}^{p}w_j(p) + \sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j} , \qquad (3)$$

, where p is the truncation lag, $w_j(p)$ is an optional weighting function that corresponds to the choice of a special window [6]. Under the null hypothesis of level stationary,

$$KPSS \to \int_0^1 V_1(r)^2\,dx \qquad (4)$$

, where $V_1(x)$ is a standard Brownian bridge: $V_1(r) = B(r) - rB(1)$ and $B(r)$ is a Brownian motion (Wiener process) on $r \in [0, 1]$. Because relation (4) is refereed in testing only on the intercept and not in the trend and as we are testing with both intercept and trend we have the second-level Brownian bridge $V_2(x)$ and it is:

$$KPSS \to \int_0^1 V_2(r)^2\,dx \qquad (5)$$

, where $V_2(x)$ is given by:

$$V_2(r) = W(r) + (2r - 3r^2)W_1 + (-6r + 6r^2)\int_0^1 W_s(s)ds \quad (6)$$

B. Autoregressive (AR) Models

We consider a series $y_1, y_2, \ldots, y_n$. An autoregressive model of order p denoted AR(p), states that y_t is the linear function of the previous p values of the series plus an error term:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \qquad (7)$$

, where $\varphi_1, \varphi_2 \ldots, \varphi_p$ are weights that we have to define or determine, and ε_t denotes the residuals which are normally distributed with zero mean and variance σ^2 [7]. Conditioned on the full set of information available up to time i and on forecasts of the exogenous variables, the one-period-ahead forecast of y_t would be

$$\hat{y}_{t+1|t} = \hat{\phi}_{0\,t+1|t} + \hat{\phi}_1 y_t + \hat{\phi}_2 y_{t-1} + \ldots + \hat{\phi}_p y_{t-p+1} + \hat{\varepsilon}_{t+1|t} \quad (8)$$

C. Moving Average (MA) Models

We consider the q order moving average MA(q) specification [7]:

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \qquad (9)$$

, where the $\theta_1, \ldots, \theta_q$ are the parameters of the model, μ is the constant and $\varepsilon_1, \ldots, \varepsilon_q$ are again the white noise error terms. The forecasts are given by (10)

$$\hat{y}_{t+j} = \hat{\mu} - \hat{\theta}_1\varepsilon_t - \hat{\theta}_2\varepsilon_{t-1} - \ldots - \hat{\theta}_q\varepsilon_{t-q+1} \qquad (10)$$

D. Autoregressive Moving Average (ARMA) Models

From the previous two sections we combine Autoregressive (AR) Moving Average (MA) Models and the Autoregressive Moving Average (ARMA) which encompasses (7) and (9) is defined as:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-p} \qquad (11)$$

ARMA(p, q) process has p autoregressive, lagged dependent-variable, terms and q lagged moving-average terms. The series R_t is said to be integrated of order one, denoted $I(1)$, because taking a first difference produces a stationary process. A nonstationary series is integrated of order d, denoted $I(d)$, if it becomes stationary after being first differenced d times autoregressive integrated moving-average model, or ARIMA (p, d, q) and will be [7]:

$$\Delta^d y_t = \mu + \phi_1\Delta^d y_{t-1} + \phi_2\Delta^d y_{t-2} + \ldots + \phi_p\Delta^d y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \qquad (12)$$

The forecasts for ARMA (p, q) model is given by (13)

$$\hat{y}_{t+j} = \hat{\mu} + \hat{\phi}_1 y_t + \hat{\phi}_2 y_{t-1} + \ldots + \hat{\phi}_p y_{t-p+1} + \hat{\theta}_1\varepsilon_t + \hat{\theta}_2\varepsilon_{t-1} + \ldots + \hat{\theta}_q\varepsilon_{t-q+1} \qquad (13)$$

Various procedures have been suggested for determining the appropriate lag length in ARMA models. The first is the addition of Akaike Information Criterion (AIC) and Schwarz Criterion (SC), which are often used in ARIMA model selection and identification. That is, AIC and SC are used to determine if a particular model with a specific set of p, d, and q parameters is a good statistical fit. SC imposes a greater penalty for additional coefficients than the AIC but generally, the model with the lowest AIC and SC values should be chosen. Specifically for Autoregressive Moving Average, Autoregressive and Moving Average models we choose Akaike criterion which is defined as:

$$AIC(p) = \ln\frac{e'e}{T} + \frac{2p}{T} \qquad (14)$$

, where e denotes the residuals, T is the sample and p indicates the lag number. We use for the optimum setting up to 5 lags.

E. Smoothing Transition Autoregressive (STAR) Models

The smoothing transition auto-regressive (STAR) model was introduced and developed by Chan and Tong [8] and is defined as:

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

$$y_t = \pi_{10} + \pi_1' w_t + (\pi_{20} + \pi_2' w_t) F(y_{t-d}; \gamma, c) + u_t \qquad (15)$$

, where $u_t \sim (0, \sigma^2)$, π_{10} and π_{20} are the intercepts in the middle (linear) and outer (nonlinear) regime respectively, $w_t = (y_{t-1....} \; y_{t-j})$ is the vector of the explanatory variables consisting of the dependent variable with $j=1...p$ lags, R_{t-d} is the transition variable, parameter c is the threshold giving the location of the transition function and parameter γ is the slope of the transition function. We shall consider two transition functions, the logistic and the exponential [9], which are defined by (16) and (17) respectively. Parameter d indicates the delay and we divide parameter γ with $\sigma(r)$, which is the standard deviation of y_t. We follow this procedure, recommended by Teräsvirta [11] because the estimation of parameter γ may cause problems like overestimations. First we apply a test to examine if y_t is linear or not proposed by Teräsvirta et al. [10]. The STAR model estimation is consisted by three steps according to Teräsvirta [11].

$$F(y_{t-d}) = (1 + \exp[-\gamma(y_{t-d} - c)])^{-1}, \gamma > 0 \qquad (16)$$

$$F(y_{t-d}) = 1 - \exp(-\gamma(y_{t-d} - c)^2), \gamma > 0 \qquad (17)$$

a) The specification of the autoregressive (AR) process of $j=1,...\;p$. One approach is to estimate AR models of different lag orders and the maximum value of j can be chosen based on the AIC information criterion Besides this approach, j value can be selected by estimating the auxiliary regression (18) for various values of $j=1,...p$, and choose that value for which the P-value is the minimum, which is the process we follow.

b) The second step is testing linearity for different values of delay parameter d. We estimate the following auxiliary regression:

$$y_t = \beta_o + \beta_1 w_t + ... + \sum_{j=1}^{p} \beta_{2j} y_{t-j} y_{t-d} +$$

$$\sum_{j=1}^{p} \beta_{3j} y_{t-j} y_{t-d}^2 + \sum_{j=1}^{p} \beta_{4j} y_{t-j} y_{t-d}^3 + \varepsilon_t \qquad (18)$$

The null hypothesis of linearity is H_0: $\beta_{2j} = \beta_{3j} = \beta_{4j} = 0$. In order to specify the parameter d the estimation of (18) is carried out for a wide range of values $1 \le d \le D$ and we choose $d=1,...,6$ In the cases where linearity is rejected for more than one values of d, then d is chosen by the minimum value of $p(d)$, where $p(d)$ is the P-value of the linearity test. Additionally if there are more than one zero P-values we choose this one with the highest F-statistic. We examine for $j=1,2...5$ and we choose those values of j similarly with parameter d. In order to compute parameters c and γ, we apply a grid search procedure for equation (18) with non linear squares and Levenberg-Marquardt algorithm. The grid search for parameter c takes place in the interval between the minimum and maximum value of each variable with

increment 0.01 and for parameter γ in the interval [1 10] with increment 0.05. The initial value for parameter c is the mean value of our data and for parameter γ is 1.

c) The third and last step is the specification of STAR model. We test the following hypotheses by [9,11]

$$H_{04}: \beta_{4j} = 0, j = 1,..., p \qquad (19)$$

$$H_{03}: \beta_{3j} = 0 \mid \beta_{4j} = 0, j = 1,..., p \qquad (20)$$

$$H_{02}: \beta_{2j} = 0 \mid \beta_{3j} = \beta_{4j} = 0, j = 1,..., p \qquad (21)$$

If we reject the (19) hypothesis then we choose LSTAR model. If (19) is accepted and (20) is rejected then ESTAR model is selected. Finally accepting (19) and (20) and rejecting (21) we choose LSTAR model. We examine both LSTAR and ESTAR models, to examine the forecasting performance and to show that the difference between their predicting performances might be almost zero.

F. Adaptive network-based fuzzy inference system (ANFIS)

Jang [12] and Jang and Sun [13] 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 Fig. 1.
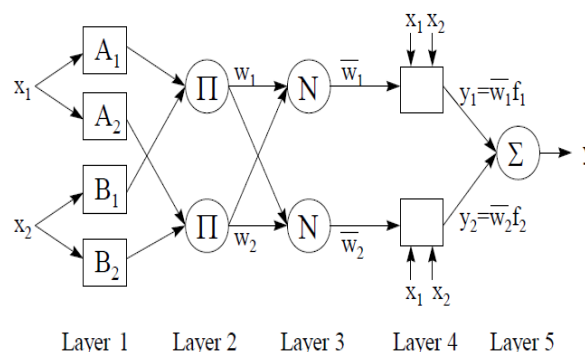

Fig. 1. Example of ANFIS architecture for a two-input, two-rule first-order Sugeno model

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

$$f_n = p_n x_1 + q_n x_2 + r_n \qquad (22)$$

The ANFIS architecture is consisted of two trainable parameter sets, the antecedent membership function parameters and the polynomial parameters p,q,r, also called the consequent parameters. The ANFIS training paradigm

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

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. We define five linguistic terms {very low, low, medium, high, very high}. Because we examine ANFIS with only on input, the dependent variable with one lag, and we a\have defined five linguistic terms the rules will be:

IF y_{t-1} is very low THEN $f_1 = p_1x + r_1$

IF y_{t-1} is low   THEN $f_2 = p_2x + r_2$

IF y_{t-1} is medium THEN $f_3 = p_3x + r_3$

IF y_{t-1} is high   THEN $f_4 = p_4x + r_4$

IF y_{t-1} is very high THEN $f_5 = p_5x + r_5$

, where y_{t-1} denotes the dependent or target variable with one lag. Because we have only one input we do on take the *AND-OR* operators. We could take more inputs, but one input is enough to get very satisfying forecasts, as we show in the next part of the study. 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 ith node of layer l. Because we have five linguistic terms in the case we examine the steps for ANFIS system computation are:

$$O_i^1 = \mu_{A_i}(x) \tag{23}$$

The adjustable parameters that determine the positions and shapes of these node functions are referred to as the premise parameters. In layer 2 we have:

$$O_i^2 = w_i = \prod_{j=1}^{m} \mu_{A_i}(x) \tag{24}$$

Each node output represents the firing strength of the reasoning rule. In layer 3, each of these firing strengths of the rules is compared with the sum of all the firing strengths. Therefore, the normalized firing strengths are computed in this layer as:

$$O_i^3 = \overline{w_i} = \frac{O_i^2}{\sum_i O_i^2}$$

(25)

Layer 4 implements the Sugeno-type inference system, i.e., a linear combination of the input variables of ANFIS, $x_1, x_2, \ldots x_p$ plus a constant term, $r_1, r_2, \ldots r_p$, form the output.

$$O_i^4 = y_i = \overline{w_i} f = \overline{w_i}(p_i x + r_i) \tag{26}$$

, where parameters $p_1, p_2, \ldots, p_i$ and $r_1, r_2, \ldots, r_i$, in this layer are referred to as the consequent parameters. In layer 5 we take:

$$O_i^5 = \sum_i \overline{w_i} f_i = \frac{\sum_i \overline{w_i} f}{\sum_i \overline{w_i}} \tag{27}$$

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

$$Y = X \cdot \theta \tag{28}$$

, where X is the matrix

$$X = [w_1 x + w_1 + w_2 x + w_2 + \ldots + w_5 x + w_5] \tag{29}$$

, where x is the matrix of inputs and θ is a vector of unknown parameters as:

$$\theta = [p_1, q_1, r_1, p_2, q_2, r_2, \ldots, p_9, q_9, r_9]^T \tag{30}$$

, where T indicates the transpose. Because the normal least square method leads to singular inverted matrix we use the singular value decomposition (SVD) with Moore-Penrose pseudoinverse of matrix [14]-[16]

For the first layer and relation (23) we use the triangular membership function. The triangular function is defined as:

$$\mu_{ij}(x_j; a_{ij}, b_{ij}) = \begin{cases} 1 - \dfrac{|x_j - a_{ij}|}{b_{ij}/2}, if |x_j - a_{ij}| \le \dfrac{b_{ij}}{2} \\ 0, otherwise \end{cases} \tag{31}$$

, where α_{ij} is the peak or center parameter and b_{ij} is the spread or support parameter. We use error back propagation algorithm with steepest descent method in order to find the optimum parameters *a* and *b*. The peak parameter update for the triangle membership function is:

$$a_{ij}(n+1) = a_{ij}(n) - \frac{\eta_\alpha}{p} \cdot \frac{\partial e}{\partial a_{ij}} \tag{32}$$

, where ηa is the learning rate for the parameter α_{ij} and E is the error functions which is:

$$e = \frac{1}{2}(y - y^t)^2 \tag{33}$$

, where y^t is the target-actual and y is ANFIS output variable. The chain rule used in order to calculate the derivatives and update the membership function parameters are [17]-[19]:

$$\frac{\partial E}{\partial a_{ij}} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial y_i} \cdot \frac{\partial y_i}{\partial w_i} \cdot \frac{\partial w_i}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial a_{ij}} \tag{34}$$

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

After some partial derivatives computations, the update equations for a_{ij} are b_{ij} are respectively

$$\alpha_{ij}(n+1) = \alpha_{ij}(n) - \eta_a \cdot e \frac{(p_i x + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \frac{w_i}{\mu_{ij}(x_j)} \cdot \frac{2|x_j - a_{ij}|}{b_{ij}} \quad (35)$$

$$b_{ij}(n+1) = b_{ij}(n) - \eta_b \cdot e \frac{(p_i x + r_i) - y}{\sum_{i=1}^{n} w_i} \cdot \frac{w_i}{\mu_{ij}(x_j)} \cdot \frac{1 - \mu_{ij}(x_j)}{b_{ij}} \quad (36)$$

In Table I we present the initial values for parameters α_{ij} and b_{ij} in each variable we examine. The learning rates for parameters a, b and consequent parameters (RHS) are set up at 0.1, 0.5 and 0.5 respectively. The number of maximum epochs is 50

TABLE II
THE INITIAL VALUES FOR PARAMETERS α_{ij} AND b_{ij} IN EACH VARIABLE

	Very Low	Low	Medium	High	Very High
GDP	a=-2.5 b=1.5	a=-0.5 b=1.5	a=1.5 b=1.5	a=3.5 b=1.5	a=5.5 b=1.5
Inflation	a=0 b=0.15	a=0.2 b=0.15	a=0.4 b=0.15	a=0.6 b=0.15	a=0.8 b=0.15
TB interest rates	a=2 b=1.5	a=4 b=1.5	a=6 b=1.5	a=8 b=1.5	a=10 b=1.5
Unemployment rate	a=5 b=1.5	a=6 b=1.5	a=7 b=1.5	a=8 b=1.5	a=9 b=1.5

,where GDP in Table I denotes the gross domestic product growth rate.

The forecasting performance of Autoregressive (AR) models and Feed-forward neural networks Autoregressive (FFNN-AR) models in both in-sample and out-of-sample periods is counted based on the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) described respectively by (37) and (38).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (37)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (38)$$

III. DATA

We examine four macroeconomic variables of US economy, the Gross Domestic Product growth rate, the inflation rate, the six monthly treasury bills interest rates and the unemployment rate. The data we use in our analysis are monthly, except from Gross Domestic Product which is in quarterly frequency. We examine the period 1991-2009 for Gross Domestic Product and period 1950-2009 for all the other variables. The periods 1991 to 2005 and 1950 to 2005 are used for the in-sample period and the period 2006-2009, which is a period of 16 observations for Gross Domestic Product and 48 observations for the remaining variables, is left for the out-of-sample forecasting period. It should be noticed that the forecasting step is two-step ahead. More specifically all models present a very good performance in one ahead period. In this paper we try to examine the forecasting performance for two-step ahead, One period ahead, monthly or quarterly, is not very useful as two periods can be. Additionally, testing more periods as three or four periods ahead can be even more useful.

IV. EMPIRICAL RESULTS

In Table II we present the results of ADF and KPSS tests, while in Table III their critical values are reported. The results are mixed. For gross domestic product we reject unit root in α=0.05 and 0.10 based on ADF test, while we accept stationarity only in α=0.01 based on KPSS test. For unemployment and inflation rates we reject unit root based on ADF statistic in all statistical significance levels, but we reject stationary hypothesis based on KPSS test. We accept that treasury bills interest rates are stationary in first differences, $I(1)$, based on both ADF and KPSS tests. Beside these results we examine both ARMA and ARIMA processes to compare the forecasts. In Table IV we present the Autoregressive and Moving average processes for the variables we examine.

TABLE II
ADF AND KPSS TESTS

Indices	ADF-statistic	KPSS-statistic
GDP Growth Rate Levels	-4.008	0.1767
Inflation Rate Levels	-5.040	0.3759
Inflation Rate First differences		0.1102
Treasury Bills Levels	-2.089	0.5547
Treasury Bills First differences	-8.461	0.0335
Unemployment Rate Levels	-4.245	0.2776
Unemployment Rate First differences		0.0531

TABLE III
CRITICAL VALUES FOR ADF AND KPSS STATISTICS

Critical values for ADF[1]	Critical values for KPSS[2]
-4.086, α= 0.01	0.216, α= 0.01
-3.471, α= 0.05	0.146, α= 0.05
-3.162, α= 0.10	0.119, α= 0.10

1 MacKinnon [20], 2 Kwiatkowski *et al.,*[5]

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

TABLE IV
AR, MA AND ARMA PROCESSES FOR VARIABLES

Indices	AR(p)	MA(q)	ARMA(p,q)	ARIMA(p,d,q)
GDP Growth Rate	1	4	1,4	1,1,3
Inflation Rate	5	4	5,4	3,1,3
Treasury Bills	5	5	4,5	5,1,5
Unemployment Rate	5	5	5,5	2,1,0

In Tables V and VI we report the linearity tests for the four macroeconomic variables. We observe that we accept the linearity process for GDP and that it could not be found a non-linear process for the specific macroeconomic variable. Also the value of lag order p is chosen based on the minimum p-value and in the cases where there are more than one zero p-values lag order p is chosen based on the highest *F-statistic*.

In Table VII we present the results of STAR model specification tests and the hypotheses (19)-(21). In all cases we accept that LSTAR is the appropriate model. In the case of inflation rate we accept the LSTAR model in $\alpha=0.05$ and $\alpha=0.10$, while we could accept ESTAR marginally for $\alpha=0.01$. In Table VIII the estimated parameters γ and c, with non-linear squares and Levenberg-Marquardt algorithm, are reported. Generally, from the value of parameter γ we observe that there is a smoothly transition between the regimes.

TABLE V
LINEARITY TESTS FOR GDP AND INFLATION RATE

Indices	GDP Growth Rate	Inflation Rate
p	1	1
d=1	0.9247	24.820
	(0.4353)	(0.000)
d=2	0.8589	13.458
	(0.4681)	(0.000)
d=3	1.333	3.470
	(0.2749)	(0.0160)
d=4	1.331	6.278
	(0.2809)	(0.0003)
d=5	0.8226	16.085
	(0.4856)	(0.000)
d=6	1.204	18.618
	(0.3182)	(0.000)

*p-values in parentheses

TABLE VI
LINEARITY TESTS FOR TREASURY BILLS AND UNEMPLOYMENT RATE

Indices	Treasury Bills	Unemployment Rate
p	1	1
d=1	7.503 (0.0001)	0.0140

		(0.9977)
d=2	17.551 (0.000)	3.295
		(0.0292)
d=3	5.941 (0.0001)	15.669
		(0.000)
d=4	3.069 (0.0274)	21.580
		(0.000)
d=5	5.228 (0.0014)	25.197
		(0.000)
d=6	2.813 (0.0387)	27.923
		(0.000)

*p-values in parentheses

TABLE VII
STAR SPECIFICATION TESTS

Indices	Inflation Rate	Treasury Bills	Unemployment Rate
Delay (d)	1	2	6
F-statistic	6.630	18.558	7.944
$H_{04} : \beta_4 = 0$	(0.0103)	(0.000)	(0.000)
F-statistic	66.627	10.457	27.705
$H_{03} : \beta_3 = 0 / \beta_4 = 0$	(0.000)	(0.0013)	(0.000)
F-statistic	0.5120	22.404	45.196
$H_{02} : \beta_2 = 0 / \beta_3 = \beta_4 = 0$	(0.4745)	(0.000)	(0.0899)
Type of model	LSTAR	LSTAR	LSTAR

TABLE VIII
ESTIMATED VALUES OF PARAMETERS γ AND c

	Inflation Rate	Treasury Bills	Unemployment Rate
ESTAR			
γ	1.4196	0.5313	1.1972
c	0.2348	-0.0066	1.2109
LSTAR			
γ	1.4605	1.1000	1.5241
c	0.9292	0.0010	1.3868

In Tables IX and X the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are reported. The results are mixed, but the general conclusion is that ANFIS outperforms significant all the models in out-of-sample periods. The exception is the case of unemployment rate where Logistic STAR model outperforms ANFIS only in the in-sample period. In the case of the gross domestic product (GDP), among the linear models, ARMA and ARIMA present the best forecasting performance in the in-sample period, but in the out-of-sample period Moving Average presents the best performance followed by Autoregressive (AR) model, indicating that the last models might be more appropriate for out of sample forecasts in GDP. In the case of inflation rate and the in-sample period, the linear models present a similar performance, as the worst performance is reported by Logistic STAR model. ANFIS present the best performance, as it was

World Academy of Science, Engineering and Technology
International Journal of Economics and Management Engineering
Vol:4, No:4, 2010

mentioned, followed by exponential STAR model. In the out-of sample period and the inflation rate ANFIS presents the best forecasting performance followed by ARIMA and exponential STAR model.

TABLE IX
RMSE AMD MAE FOR GDP AND INFLATION RATE

	GDP		Inflation	
In sample period				
	RMSE	MAE	RMSE	MAE
AR	1.9216	1.4282	0.3506	0.2457
MA	1.9285	1.4291	0.3560	0.2498
ARMA	1.8344	1.3221	0.3491	0.2520
ARIMA	1.8528	1.3372	0.3681	0.2548
ESTAR			0.2907	0.2067
LSTAR			0.3999	0.2658
ANFIS	0.8003	0.6403	0.2245	0.1478
Out-of-sample period				
	GDP		Inflation	
	RMSE	MAE	RMSE	MAE
AR	3.8269	2.1184	0.4999	0.2276
MA	2.5212	1.9123	0.2162	0.2010
ARMA	4.6297	2.7162	0.2251	0.2231
ARIMA	3.9451	2.7993	0.1922	0.1773
ESTAR			0.2130	0.1774
LSTAR			0.2781	0.2471
ANFIS	1.6531	1.0651	0.1759	0.1551

TABLE X
RMSE AMD MAE FOR INTEREST RATES AND
UNEMPLOYMENT RATE

	Interest rates		Unemployment	
In sample period				
	RMSE	MAE	RMSE	MAE
AR	2.7655	1.9509	1.5450	1.2172
MA	2.7784	2.0630	1.5550	1.2015
ARMA	2.7604	1.9439	1.5491	1.2154
ARIMA	2.7441	1.9407	1.5420	1.2912
ESTAR	2.7960	2.7302	0.2073	0.1541
LSTAR	3.2374	2.8784	0.5313	0.4307
ANFIS	1.8595	1.0343	0.3063	0.2861
Out-of-sample period				
	Interest rates		Unemployment	
	RMSE	MAE	RMSE	MAE
AR	2.0401	1.6104	1.1816	0.8530
MA	1.9569	1.7818	0.9673	0.9033
ARMA	1.7050	1.2110	1.7261	1.1395
ARIMA	1.9824	1.4260	2.2424	1.6378
ESTAR	1.8967	1.7119	0.4142	0.3915
LSTAR	1.4558	1.3049	0.1841	0.1421
ANFIS	1.2600	1.1643	0.1165	0.0999

For the 6-month treasury bills interest rates and linear models, ARMA and ARIMA present the best performance in both in-sample and out-of-sample periods, while the best performance among all models, presents ANFIS followed by Logistic STAR model.

Finally, in the last macroeconomic variable we examine, the unemployment rate, Exponential STAR model followed by Logistic STAR model and ANFIS, presents the best forecasting performance in the in-sample period. In the out-of-sample period ANFIS outperforms the other models, followed by Logistic STAR model. RMSE and MAE in the case of ANFIS and the out-of-sample period, are significant lower relatively to the other models.

V. CONCLUSIONS

In this paper we proposed the Adaptive Network-Based Fuzzy Inference Autoregressive System (ANFIS) which is known also as Adaptive Neuro-Fuzzy Inference System. We examined the forecasting performance of linear and non-linear models, as the autoregressive, moving average, autoregressive moving average and smoothing transition autoregressive models and we have shown that ANFIS outperforms significant the other models in the most cases. Furthermore, we concluded that the non-linear smoothing transition autoregressive models present a superior forecasting performance to the linear models in some cases. We examined only one fuzzy membership function, the triangular, while other membership functions can be tested as well as the trapezoidal, the Gaussian or the Generalized Bell functions among others. Furthermore, genetic algorithms can be used instead to error backpropagation algorithm we used in the current study, for the training process. Generally, ANFIS outperforms significant the conventional linear and non-linear econometric modelling in all macroeconomic variables we examined in the out-of-sample period which is our main interest. For this reason ANFIS technology should be adopted and used from the national statistical services, the central banks and the financial industry in the countries, where the conventional econometric modelling is still mainly used. Further research applications should be made and the principles and techniques of artificial intelligence should be introduced in economic academic departments in the future.

REFERENCES

[1] Atsalakis, S.G. and G.I. Atsalakis, "Fruit production forecasting by neuro-fuzzy techniques," Paper prepared for presentation at the 113th EAAE Seminar "*A resilient European food industry and food chain in a challenging world*", Chania, Crete, Greece, September, pp. 3-6, 2009
[2] Ballini, R., I. Luna, L.M. Lima and R.L.F. da Silveira, "A comparative analysis of neurofuzzy, ANN and ARIMA models for Brazilian stock index forecasting," *SCE - Computing in Economics and Finance,* 1995
[3] D. Marček, "Stock Price Forecasting: Autoregressive Modelling and Fuzzy Neural Network," *Proceedings of the 1999 EUSFLAT-ESTYLF Joint Conference*, Palma de Mallorca, Spain, September, pp. 22-25, 1995
[4] D. A. Dickey, and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*. vol. 74, pp. 427–431, 1979
[5] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal of Econometrics,* vol. 54, pp. 159-178, 1992
[6] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika,* vol. 37, pp. 1-16, 1950
[7] W.H Greene, *Econometric Analysis*, Sixth Edition, Prentice Hall, New Jersey, 2008, pp. 560-579
[8] K.S. Chan and H. Tong, "On estimating thresholds in autoregressive models", *Journal of Time Series Analysis*, vol. 7, pp. 178-190, 1986
[9] T. Teräsvirta and H.M. Anderson, "Characterizing nonlinearities in business cycles using smooth transition autoregressive models", *Journal of Applied Econometrics*, vol. 7, pp. 119-136, 1992
[10] T. Teräsvirta, C.F. Lin and C.W.J. Granger "Power of the Neural Network Linearity Test", *Journal of Time Series Analysis*, vol. 14, pp. 209-220, 1993

[11] T. Teräsvirta, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association*, vol. 89, no. 425, 208–218, 1994

[12] J.-S.R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, 1993

[13] J.-S. R. Jang and C.-T. Sun, "Neuro-fuzzy Modeling and Control," *Proceedings of the IEEE*, vol. 83, no. 3, 378-406, March, 1995

[14] E.H. Moore, "On the reciprocal of the general algebraic matrix," *Bulletin of the American Mathematical Society,* vol. 26, pp. 394–395, 1920

[15] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 406–413, 1955

[16] M. Petrou, and P. Bosdogianni, *Image Processing: The Fundamentals*, John Wile & Sons, 2000

[17] L. Khan, S. Anjum and R. Bada, "Standard Fuzzy Model Identification using Gradient Methods," *World Applied Sciences Journal*, vol. 8, no. 1, pp. 01-09, 2010

[18] M.A. Denai, F. Palis and A. Zeghbib, ANFIS Based Modelling and Control of Non-Linear systems: A Tutorial, *IEEE Conf. on Systems, Man, and Cybernetics*, vol. 4, pp. 3433-3438, 2004

[19] L.H. Tsoukalas, and R.E. Uhrig, *Fuzzy and Neural Approaches in Engineering*, First Edition, John Wiley & Sons, 1997, pp. 445-470

[20] J. G. MacKinnon, "Numerical Distribution Functions for Unit Root and Cointegration Tests", *Journal of Applied Econometrics*, vol. 11, pp. 601-618, 1996