Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Exchange Rates Forecasting Using a Hybrid Fuzzy and Neural Network Model

An-Pin Chen[1] and Hsio-Yi Lin[2]

[1)] Institute of Information Management, National Chiao-Tung University, HsinChu, Taiwan 300
apc@iim.nctu.edu.tw

[2)] Department of Finance, Ching-Yun University, Jung-Li, Taiwan 320
hylin@cyu.edu.tw

***Abstract*- Artificial neural networks (ANNs) are promising approaches for financial time series prediction and have been widely applied to handle finance problems because of its nonlinear structures. However, ANNs have some limitations in evaluating the output nodes as a result of single-point values. This study proposed a hybrid model, called Fuzzy BPN, consisting of backpropagation neural network (BPN) and fuzzy membership function for taking advantage of nonlinear features and interval values instead of the shortcoming of single-point estimation. In addition, the experimental processing can demonstrate the feasibility of applying the hybrid model-Fuzzy BPN and the empirical results show that Fuzzy BPN provides a useful alternative to exchange rate forecasting.**

***Keywords*- backpropagation neural network, Fuzzy Membership Function, Exchange rate.**

I. INTRODUCTION

Recently international investment activities are more frequent and global trades become more liberal, floating exchange rate system cause uncertainty of exchange rate in the international trade and investment. Thus, exchange rates forecasting, using linear time series models, non-linear time series models, and artificial intelligence models, becomes an important financial problem and has been a recurrent subject of research during the last two decades.

Meese and Rogoff[15] demonstrated the forecasts of exchange rate predictability from structural model based on monetary and asset pricing theories of exchange rate determination perform no better than the ones generated by the simplest of all models in terms of out-of sample forecasting ability. Further, many literatures[1,3,17,21] also pointed out the standard econometric methods are unable to produce significantly better forecasts than the random walk model and supportive of the efficient-market hypothesis.

Although these findings are strength advocated that the exchange rates trend is random walk, many researchers have attempted to search various alternative methods for modeling of exchange rates forecasting. One of the first studies to overthrow the random walk model is the proposal made by MacDonald and Taylor[14]. Many literatures have proposed several proofs explaining that Exchange rates belong to nonlinear behavior. In addition, Kilian and Taylor[12] also signified that the forecast efficiency of econometric exchange models is not able to achieve its optimum because it is constrained by the linear quality of the traditional statistics models. Afterwards, the exchange rates time-series property has been proven to exist in the family of Autoregressive conditional heteroskedasticity(ARCH) effect.

In the past ten years, following the rapid advancement of technology and the vast application of artificial intelligence, researchers have become more tend to use artificial neural network(ANN) as an alternative method in exchange rates forecasting and Backpropagation neural networks(BPNs) is one of the most popular ANN used. Lisi and Schiavo[13] used BPNs , chaotic models were separately applied on the exchange rate prediction and the results from both were better than the random walk hypothesis. Funahashi[8] and Hornik et al.[10] believed that ANN is more suitable for time series prediction. In addition, most of the studies done recently hybridize several artificial intelligence techniques, for instance [9, 20], or integrate ANN statistics methods, for example, Chen and Leung[5] used the General Regression Neural Network(GRNN) to predict foreign exchange rates and through actual proofs discovered that GRNN approach not only results better exchange rate forecasts but also products in higher investment returns than the single-stage model.

However, the predictive outputs of ANN are generally single-point values. It seems unreasonable that “single-point values” outperform an interval for forecasting certain financial predicting problems, that is, stock prices indexes, returns, and exchange rates. A single-point value indeed has more difficulty than an interval value in for forecasting a target value. In order to take advantage of BPNs non-linear feature and improve the single-point values problems in BPNs, this paper attempts to propose a BPNs using a fuzzy set architecture, and modified neural network is designed to combine the non-linear learning characteristic of BPNs and the interval estimation of statistics, thus can be a dynamical model for recognize the financial time series patterns and for forecast the exchange rate trend.

The remainder of this paper consists of five sections. Section2 introduces the basic concept of BPN and GARCH models. Section3 then describes a fuzzy set interval approach based on the BPN model for forecasting exchange rates movement in this part, a case study of the US /New Taiwan Dollar exchange rates is also designed to examine the influence of the predictive performance of the modified BPNs(short-call Fuzzy BPNs below) suggested by this study, and a comparison is drawn between the traditional BPN model, Fuzzy BPNs and AR-GARCH model. Subsequently, the empirical results are presented and discussed in section4. Finally, the concluding remarks are presented in section 5.

II. ARTIFICIAL NEURAL NETWORK AND GARCH MODEL

A. *Artificial Neural Network Model*

The ANN used in this paper is BPN, which uses Backpropagation trained by gradient descent algorithm. This algorithm supposes that the j th neuron of the hidden layer receives that activation function:

$$H_j = \sum_i x_i w_{ij}^h \tag{1}$$

Where x_i is the signal to the input neuron i and w_{ij}^h is the weight of the connection between the i th input neuron and the he j th neuron of the hidden layer., then this activation function produces as output by a transfer function f of the hidden layer

$$h_j = f_j(H_j) = f_j^h \tag{2}$$

Then each output neuron k receives as input from the output of the previous layer (hidden layer) and produces the final result

$$O_k = \sum_j w_{jk}^o \times h_j \tag{3}$$

where w_{jk}^o is the weight of the connection between hidden neuron j and output neuron k, and it is transformed again to

$$o_k = f_k(O_k) = f_k^o \tag{4}$$

The goal of the learning process is to determine a set of weights when the actual output y_k by the network given x_i as input be as close as possible to the desired output o_k, the function of squared errors for each neuron, which is to be minimized,

$$E = \frac{1}{2}\sum_k (y_k - o_k)^2 , \tag{5}$$

The data fed to an input node are multiplied by a set of weights; all such weighted inputs are totaled using an activation function that depends on the learning algorithm at each node of the next layer. The output of the activation function then transforms the raw input for a node in the next layer, this process is called "feed-forward"

In addition, the weights are modified to reduce the squared error. The change in weights,

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} \tag{6}$$

Where η is the learning rate, $0 < \eta < 1$, Rumelhart et al.(1986)[16] introduced a momentum term α in (6), thus obtaining the following learning rule,

$$\Delta w_{kj}(t+1) = -\eta \frac{\partial E}{\partial w_{kj}} + \alpha \Delta w_{kj}(t) \tag{7}$$

The momentum α is usually set in the interval [0,1] and it can also be helpful to keep the learning process from fell into the local minima.

That is, in the final layer, the predictive values of the output nodes may differ from the target values owing to the weights being randomly initialed. The error between the predictive and the target values can be adjusted by adjusting the weights of learning epochs, using a delta rule derived from a cost function of the error. This process is termed "backward".

B. *GARCH Model*

The GARCH model of Engle[7] and Bollerslev[2] requires joint estimation of the current conditional mean model as formula (8) and the past conditional variance (9) in order to capture the non-linearity involved the distribution of financial data is leptokurtic. The GARCH(p,q) model can be represented by the following model:

$$E_t = a_0 + \sum_{i=1}^m a_i E_{t-i} + \sum_{j=1}^n b_j \varepsilon_{t-j} \tag{8}$$

where E_t is series of continuous exchange rate(normalized), the a_0, a_i and b_j are the constant parameters, $\varepsilon_t \sim N(0,h)$ and the conditional variance of errors, h_t is given by:

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \qquad (9)$$

Where $\alpha_0 > 0, \alpha_i, \beta_j \geq 0$ and $\sum_i \alpha_i + \sum_j \beta_j < 1$.

These restrictions on the parameter prevent negative variances and the GARCH(1,1) was found to be the most popular.

III. THE HYBRID METHODOLOGY AND RESEARCH DESIGN

A. Fuzzy BPNs

This paper proposes fuzzy-interval architecture using fuzzy set for improving the single-point shortcoming of BPNs, call Fuzzy BPN below. Further, a fuzzy set is completely characterized by it membership function (MF), the MF of fuzzy-interval approach is defined in this paper is the Gaussian MF and specified by two parameters {c, σ}:

$$f(x; c, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \qquad (10)$$

where c is the Gaussian MFs center and σ determines the MFs width. In this paper, the c indicates the mean of weekly exchange rates and the σ intents the standard deviation of weekly exchange rates, the MF of fuzzy-interval is also decided completely by c and σ. Note that Gaussian MF is a direct generalization of the normal distribution use in probability theory , when fuzzy-interval MF is centered on c and the extent to which it spreads out around c is added and subtracted 1.96σ ($\pm 1.96\sigma$) of 95% probability of confidence interval (see Fig. 1).

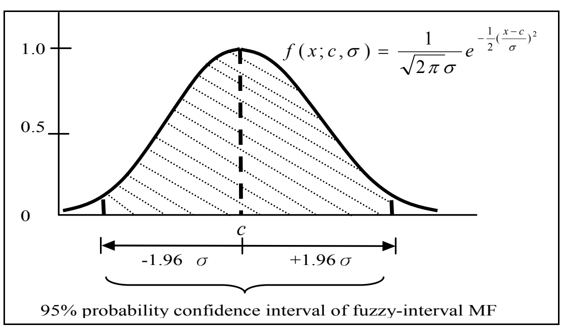

Fig. 1 Gaussian MF of fuzzy-interval approach in this paper

According to the assumption of the MF above mentioned, this research tries to learn the parameter c and σ using BPN. Fig. 2 is shown the BPNs frame for producing the fuzzy-interval MF, then used c and σ to find the fuzzy-interval MF, in this way, not only can it maintain the BPNs non-linear feature, at the same time, it can improve the single-point values problems in BPNs. Here, the above framework is called Fuzzy BPNs, as seen in Fig. 3.

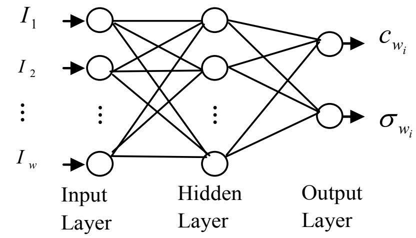

Fig. 2 BPNs frame for producing the fuzzy-interval MF

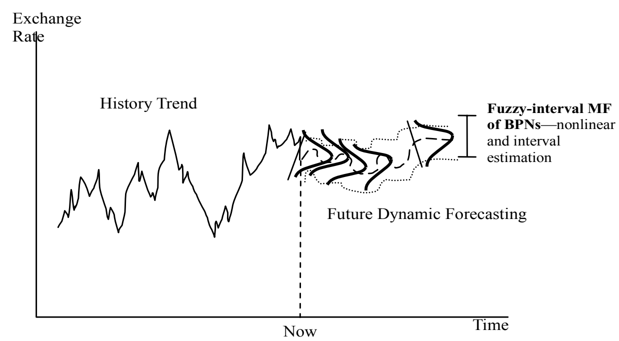

Fig. 3 the property of fuzzy BPNs in this paper

B. Data and Experimental Design

The data sets were bilateral exchange rates between New Taiwan Dollar and US dollar (NTD/USD), and composed of daily rates covering almost 14-year period from the beginning of Central Bank of China, Republic of China (Taiwan), on January 3, 1993 to October 14, 2006 and including 3425 observations.

This study attempts to take w days for predict the following weekly (5 trading days) exchange rate. To put it plainly, when we want to forecast the next unknown weekly exchange rate, we can use the past w days ahead the future next weekly days to training model for get predicted values. Consequently, a "sliding window" was proposed as shown in Fig. 4 with different window width $w+5$ moving from the first period to the last period of the entire data set labeled by S_i (i is from one to $N-w-4$) resulting in all N (N =3425) observations being divided again into $N-w-4$ samples. Consulting Chen and Tsao[4] and Tay and Cao[19], there are five different w, their being 5,10,15,20 and 25, considered in this paper. Many investigations have used a convenient ratio to separate in-samples form out-of samples ranging from

70%:30% to 90%:10%[22]. Hence, about approximately 25% of the samples are used for test, 75% for training in this paper and every sample comprises a time series data containing $w + 5$ exchange-rate observations.

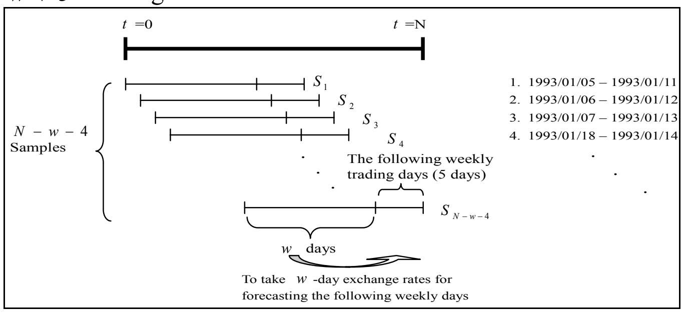

Fig. 4. Sliding window

For effective predictive performance of BPN and GARCH processing, this paper takes the natural logarithmic transformation to stabilize the time series of exchange rate via normalization. The normalizations of two output variables of the exchange rates in this paper separately are

$$mean_{s_i} = \frac{1}{5}\sum_{j}^{5} \ln\left(\frac{P_{w+j}}{P_w}\right) \tag{11}$$

$$SD_{s_i} = \sqrt{\frac{\sum_{j}^{5}\left(\ln\left(\frac{P_{w+j}}{P_w}\right) - mean_{s_i}\right)^2}{4}} \tag{12}$$

where P_w denotes the normalized basic day of the following weekly exchange rates for the previous w days, while $mean_{S_i}$ and SD_{S_i} represents the mean and standard deviation for the following week exchange rates during period S_i.

IV. EMPIRICAL RESULTS

This section interprets and presents the best specifications of Fuzzy BPNs, traditional BPNs and GARCH model for daily NTD/USD exchange-rate series.

A. *BPNs Model*

The BPNs model used in this study is a three-layer feed forward network, and is trained to map the next weekly-day mean and standard deviation for the coming w days using a backpropagation algorithm. This study varies the number of nodes in the hidden layer and stopping criteria for training, TABLE I is the parameters setting list and Matlab7.0 program language was run for the experiments of BPNs in this study.

TABLE I BPNs parameters setting in this study

parameters	setting value
Hidden layer	1 layer and 2 layers
Hidden nodes	1 layer : 5, 15, 30, 50, 100 2 layers : 5, 15, 30
Learning epochs	10000
Learning rate	0.1, 0.3, 0.5, 0.7, 1.0
Momentum term	Defaulted by Matlab program language
Total number of trial-and-error	350 { 1 layer : 5 × 5 × 5 (w) = 125 2 layers 3 × 3 × 5 × 5 (w) = 225

Several performance criteria are used to model BPNs, this study including the Mean squared error (MSE) suggested by Coakley et al.[6]to determine the point at which the training stops and assess the forecasting performance.

$$MSE = \frac{\sum_{i}^{n_w} (F_i - O_i)^2}{n_w - 1} \tag{13}$$

where n_w is the number of the example sequences, $n_w = N - w - 4$, O_i is the target value, F_i is the predicted value, the final determined parameters of each w -days BPNs are based on the smallest converged MSE their own respectively. Since the major purpose of this paper is to investigate the effects of BPNs parameters on the modeling and forecasting performance of BPNs, the values of MSE between training set (in-sample) and testing set (out-of-sample) will be compared, with the emphasis put on the out-of-sample analysis, because it is only using the testing data that the BPN parameter setting with the best forecasting capability can be proven and found.

All the set parameters after passing through Trail and Error, then based on the smallest MSE value of the 5 different w, the MSE value is chosen as its first measurement standard, if the training data MSE value is the same then the training data becomes the second screening standard, TABLE II is the best parameter setting model (the best performance) chosen and arranged as follows.

TABLE II BPNs best parameter setting model

w	training data MSE	testing data MSE	Hidden	learning rate
5	0.000010357	0.000010723	30	0.5
10	0.000009428	0.000009566	15×15	0.3
15	0.000005957	0.000008403	30×30	0.7
20	0.000007632	0.000012630	15	1.0
25	0.000008058	0.000010245	5×15	0.5

B. *GARCH Model*

Various goodness-of-fit statistics are used to compare the all estimated GARCH model in this paper, the diagnostics are the MSE, the likelihood-ratio tests, tests for the standard residuals, Schwarz's Bayesian information criterion (SBC) by Schwarz(1978)[18] and Akaike's information criterion (AIC)[11].The GARCH models were tried for $p = 1,2,\cdots,5$ and $q = 1,2,\cdots,5$ using SAS program software, TABLE III shows that the statistically significant parameters for every AR(w)-GARCH(p,q) model and the last results was listed , the estimated values of parameters α_0, α_1 and β all satisfy $\alpha_0 > 0,\ \alpha_1 \geq 0,\ \beta \geq 0$ and $\alpha + \beta < 1$. This indicates the weaknesses of imposing the parameter estimates of a GARCH model to certain constraints such as stationary.

TABLE III Estimation results of GARCH models for NTD/USD exchange rates

Model	$\alpha_0(\times 10^{-6})$	t-value	α_1	t-value	β_1	β_2	β_3	β_4	β_5	t-value				
										β_1	β_2	β_3	β_4	β_5
AR(5)GARCH(1,4)	4.5487	17.5	0.001401	0.03	0.0421	0.0276	0.0231	0.0281	-	27.95	7.07	4.24	6.72	-
AR(10)GARCH(1,4)	4.5564	17.21	0.001382	0.03	0.0421	0.0275	0.0229	0.0278	-	27.13	6.69	3.98	5.67	-
AR(15)GARCH(1,5)	4.5185	13.35	0.001567	0.02	0.0423	0.0256	0.0183	0.0171	0.0228	25.64	5.23	2.83	2.95	5.39
AR(20)GARCH(1,5)	5.9404	9.55	0.0172	0.17	0.0399	0.0256	0.0186	0.0182	0.0212	20.59	3.73	2.3	2.23	3.44
AR(25)GARCH(1,3)	0.67168	13.51	0.4395	14.28	0.1156	0.0804	0.0907	-	-	24.38	4.67	3.86	-	-

TABLE IV indicates all final AR(w)-GARCH(p,q) models that their own MSE values, Log L values, the lowest AIC and SBC, dividedly. In the next section, the Fuzzy BPNs and traditional BPNs models will be compared the forecasting performance with final AR-GARCH models.

TABLE IV The goodness-of-fit statistics values of all final AR(w)-GARCH(p,q) models

Model	Goodness-of-Fit Statistics			
	MSE	Log L	SBC	AIC
AR(5)-GARCH(1,4)	0.0000204	11061.0956	-22028.009	-22098.191
AR(10)-GARCH(1,4)	0.0000204	11045.7559	-22057.512	-21958.113
AR(15)-GARCH(1,5)	0.0000203	11044.3184	-22042.637	-21908.183
AR(20)-GARCH(1,5)	0.0000199	11031.2324	-22006.465	-21842.826
AR(25)-GARCH(1,3)	0.0000198	11739.1774	-23416.355	-23235.244

C. Forecasting Performance

Fuzzy BPNs, traditional BPNs, and AR-GARCH models all used similar measurement standard- MSE values as its measurement standard. It can be known from TABLE V that the MSE value of different w BPNs models are all lower than those with AR-GARCH models, which shows that the forecasting ability of the BPNs models are better than the AR-GARCH models; in addition, from the point of view of forecasting accuracy rate as the judgment standard, the exchange rates of training data of the Fuzzy BPNsare between the fuzzy-interval MF's forecasting areas, which are 83.3669%, 83.1798%, 82.4129%, 83.8702%, and 83.5577% respectively, while the accuracy rate of the exchange rates of training data to be guessed correctly are 70.6909%, 68.1455%, 70.2941%, 60.9482%, 63.7264%, while the accuracy rate of the traditional BPN models and the AR-GARCH models is 0%. It can be known than that aside from the Fuzzy BPNs having a better forecasting ability than the AR-GARCH models, the study made use of the sector characteristic of fuzzy MF to improve the single point forecasting shortcoming of the traditional BPN models.

TABLE V The performance comparison of Fuzzy BPNs, traditional BPNs and AR-GARCH models

W-day	MSE			Accuracy Rate of NTD/USD Exchange-rate Prediction				
	BPNs		AR-GARCH	Fuzzy BPNs*		BPNs		AR-GARCH
	training	testing		training	testing	training	testing	
5	0.000010357	0.000010723	0.000020400	83.3669%	70.6909%	0%	0%	0%
10	0.000009428	0.000009566	0.000020400	83.1798%	68.1455%	0%	0%	0%
15	0.000005957	0.000008403	0.000020300	82.4129%	70.2941%	0%	0%	0%
20	0.000007632	0.000012630	0.000019900	83.8702%	60.9482%	0%	0%	0%
25	0.000008058	0.000010245	0.000019800	83.5577%	63.7264%	0%	0%	0%

* Assumption of 95% probability in Gaussian distribution, the Fuzzy-interval MFs were extended based on $c \pm 1.96\ \sigma w$

V. CONCLUSIONS

The applications of ANNs in financial area have obtained increasing popularity in the past decades. Nevertheless, a strict methodology on how to properly design a system of ANNs for forecasting time series data is still a difficult problem; the disadvantages of ANNs also have be widely discussed and solved, such as "black box", single-point prediction, etc. In this study, a method called Fuzzy BPNs consisted of fuzzy-interval MF was suggested for the purpose of improving upon the shortcomings of single-point estimations in conventional artificial neural networks, and still has possession of ANNs nonlinear capabilities. This paper also provides evidence for the forecast performance of Fuzzy BPNs in terms of interval evaluation is not only much better than traditional BPNs in terms of single-point evaluation, but more well than AR-GARCH models. To conclude, this contribution presents that a combination of BPNs with Fuzzy membership function proposed by this research offers a useful approach for predicting time series patterns in exchange market data.

VI. REFERENCES

[1] D. Alexander, L.R. Thomas, "Monetary/asset Models of Exchange Rate Determination: How Well Have They Performed in the 1980's?" *International Journal of Forecasting*, vol. 3, pp53-64, 1987.

[2] T. Bollerslev, "Generalised Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, pp307-327, 1986.

[3] G. Canarella, S. K. Pollard, "Efficiency in Foreign Exchange Markets: a Vector Autoregression Approach," *Journal of International Money and Finance*, vol. 7, pp331-346, 1987.

[4] S.-H. Chen, C. Y. Tsao, "Information Content of the Trajectory-Domain Models," *9th international Conference on Computing in Economics and Finance, Seattle, USA*, July 2003.

[5] A.-S. Chen, M. T. Leung, "Regression Neural Network for Error Correction in Foreign Exchange Forecasting and Trading," *Computers and Operations Research*, vol. 31, pp1049-1068, 2004.

[6] J. R., Coakley, D. D. Mcfarlane, W. G. Perley, Alternative criteria for evaluating artificial neural network performance, presented at TIMS/ORSA Joint National Meeting , April 1992.

[7] R. F. Engle, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, vol. 50, pp267-287, 1992.

[8] K. Funahashi, "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, vol. 2, pp183-192, 1989.

[9] K. Huarng, T. H.-K. Yu, "The Application of Neural Networks to Forecast Fuzzy Time Series," *Physica A*, vol. 363, pp481-491, 2006.

[10] K. Hornik, M. Stinchcombe, H. White, "Multilayer Feedforward Networks Are Universal Approximateors," *Neural Networks*, vol. 2, pp359-366, 1989.

[11] G. G. Judge, W. E. Griffiths, R. C. Hill, H. Lutkepohl, T. C. Lee, :The Theory and Practice of Econometrics," 2nd ed., New York: Wiley, 1985.

[12] L. Kilian, M. P. Taylor, "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates?" in: *European Central Bank Working Paper Series*, WP No. 88, November 2001.

[13] F. Lisi, R. A. Schiavo, "A Comparison between Neural Networks and Chaotic Models for Exchange Rate Prediction," *Computational Statistics and Data Analysis*, vol. 30, pp87-102, 1999.

[14] R. MacDonald, M. P. Taylor, "The Monetary Model of the Exchange Rate: Long-run Relationships, Short-run Dynamics and How to Beat a Random Walk," *Journal of International Money and Finance*, vol. 13, pp276-290, 1994.

[15] R. A. Meese, K. Rogoff, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, vol. 14, pp3-24, 1983.

[16] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning Internal Representations by Error Propagation." in: *D. E. Rumelhart, J. L. McClelland, (Eds.), Parallel Distributed Processing*, vol. 1, MIT press, Cambridge, MA, pp 318-362, 1986.

[17] N. Sarantis, C. Stewart, "Structural, VAR and BVAR Models of Exchange Rate Determinate: A Comparison of Their Forecasting Performance," *Journal of Forecasting*, vol. 14, pp201-215, 1995.

[18] G. Schwarz, "Estimating the Dimension of a Model", Annals of Statistics, vol. 6, pp461-464, 1978.

[19] F. E. H. Tay, L. J. Cao, "Modified Support Vector machines in Financial Time Series Forecasting," *Neurocomputing*, vol. 48, pp.847-861, 2002.

[20] M. Versace, R. Bhatt, O. Hinds, M Shiffer, "Predicting the Exchange Traded Fund DIA with a Combination of Genetic Algorithms and Neural Networks," *Expert systems with Applications*, vol. 27, pp417-425, 2004.

[21] C. C. P. Wolff, "Time Varying Parameters and the Out-of-sample Forecasting Performance of Structural Exchange Rate Models," *Journal of Business Economics and Statistics*, vol. 5, pp87-103, 1987.

[22] G. P. Zhang, "Business Forecasting with Artificial Neural Networks: An overview", in: G. P Zhang(Ed.), Neural Networks in Business Forecasting, Idea Group Publishing , Hershey, PA, 2004.