

Forecasting Exchange Rates Using Artificial Neural Networks

N.V. CHANDRASEKARA AND C.D. TILAKARATNE

Department of Statistics, University of Colombo, Sri Lanka

ABSTRACT

In the dynamic global economy, the accuracy in forecasting the foreign currency exchange rates is of crucial importance for any future investment. The use of computational intelligence based techniques for forecasting has been proved extremely successful in recent times. The aim of this study is to identify a neural network model which has ability to predict the US Dollar against Sri Lankan Rupee (USD/LKR) with higher accuracy level. In this study both static and dynamic neural network architectures were considered. Three types of neural network models were employed: (I) Feedforward neural network (FFNN; static neural networks) with the Backpropagation (BPR) algorithm; (II) FFNN with the Scaled Conjugate Gradient (SCG) algorithm; (III) Time Delay neural network (TDNN; dynamic neural network). Best performed models were found from each approach and forecastability of these models were compared to come up with the best model to predict the USD/LKR. It was found that the FFNN trained with the SCG algorithm performs better than FFNN trained with the BPR algorithm. Therefore the best static neural network for predictions is the FFNN trained with the SCG algorithm. The best TDNN outperforms the best static neural network model and finally this model can be proposed as the best model to predict the USD/LKR. The best performed TDNN model contains two hidden layers, three neurons in each layer and six time delays. This model has ability to forecast unseen data with 76% prediction accuracy.

Keywords: Feedforward Neural Network, Modelling exchange rates, Non-linear models, Time Delay Neural Network

INTRODUCTION

The foreign exchange market has experienced unexpected development over the last few decades as the world moves into a concept of a "global village" and it has been one of the leading financial markets in the world. The exchange rate, which is the rate at which the currency of one country can be exchanged for the currency of another country, plays an important role in controlling dynamics of the exchange market. Therefore, appropriate prediction of exchange rates is a crucial factor to achieve the success in many business, investments, etc. all over the world. Nowadays there exists significant evidence that the exchange rates are predictable to

some extent. Most of the researches related to financial market predictions in this era suggest that foreign exchange rates are predictable with high accuracy (for example Macdonald, 1995).

The prediction of exchange rate is one of the most challenging applications of modern time series forecasting as the rates are inherently noise, non-stationary and deterministically chaotic. These characteristics suggest that there is no complete information that could be obtained from the past behavior of such markets to fully capture the dependency between the future rates and that of the past. In such situations, it is widely assume that the historical data captures the past behavior up to some extent. As a result, the historical data is the major player or the input in the prediction process. Time series modelling is the conventional technique used to model the exchange rates. However, these techniques have limitations as the exchange rate series are non-stationary and noisy (Pan *et al.*, 2005).

Many researches have been carried out in the literature to compare traditional time series models with computational models for forecasting. These models are unsatisfactory because they are parametric and some are based on the assumption that the time series been forecasted are linear and stationary. Many past studies (Hill *et al.*, 1996; Kohara *et al.*, 1997; Yao *et al.*, 1999; Giles *et al.*, 2001; Kaboudan, 2005; Binner *et al.*, 2005; Tilakaratne *et al.*, 2007) suggest that non-linear models such as neural network models perform better than traditional time series linear models. Therefore, neural networks have been advocated as an alternative to traditional statistical forecasting methods.

Artificial Neural Network (ANN) is more effective in describing the dynamics of non stationary time series due to its non parametric, noise tolerance, data driven and adaptive properties. Neural network models are different from traditional linear models and other parametric nonlinear approaches, which are often limited in scope when handling nonlinear or nonstandard problems. ANNs are universal function approximates that can map any nonlinear function without prior assumptions about the data. Past studies reveal that the foreign exchange rates can be forecast with high accuracy using artificial neural networks (Kuan, and Liu, 1995; Kodogiannis and Lolis, 2001; Kamruzzaman and Sarker, 2003; Kaboudan, 2005). Past studies employed feed forward neural network (FFNN) with back propagation and scaled conjugate gradient algorithms (for example Kamruzzaman and Sarker, 2003). However time delay neural network was not applied in the published work.

The main objective of this study is to propose a neural network model to predict exchange rate of the US Dollar against Sri Lankan Rupee. This objective was further elaborated into finding the best static neural network model, finding the best dynamic neural network and comparing those best models.

Many studies have been carried out to predict exchange rates using artificial neural networks in many countries and no published research which was done relating to Sri Lankan Rupee was found. This would be the first study which was carried out to predict exchange rate of the USD/LKR using feedforward neural networks.

When considering time delay neural networks, past studies relating to exchange rate predictions even in other countries are hardly found, but they have been commonly used for other time series predictions.

Architecture and Training of Neural Networks

ANNs can be broadly classified into two main categories using their architecture which are namely static ANNs and dynamic ANNs. Static ANNs such as feedforward neural networks (FFNN) have no feedback elements and contain no delays. Their output is calculated directly from its inputs. In dynamic ANNs such as time delay neural networks (TDNN) the output not only depend on the current inputs, it also depend on the previous inputs and outputs or state of the network. TDNN uses delay elements with several time steps and there are two types of time delay neural networks which are namely Focused time delay neural networks and Distributed time delay neural networks. Focused time delay neural networks are appropriate for time series forecasting and Distributed time delay neural networks are more suitable for pattern recognition. Both feedforward and time delay neural networks are widely used for time series modelling recently.

Neural network models are data driven models and usually these models trained using training data, validate using validation data and finally test on test data. Training set was to train the neural network while the validation set was employed to avoid overtraining of neural networks. The forecastability of fitted neural network models were tested using the test data set.

There are so many factors which should be discussed under the architecture of a neural network which are namely number of layers (input, output, hidden), number of nodes in each layer, etc.

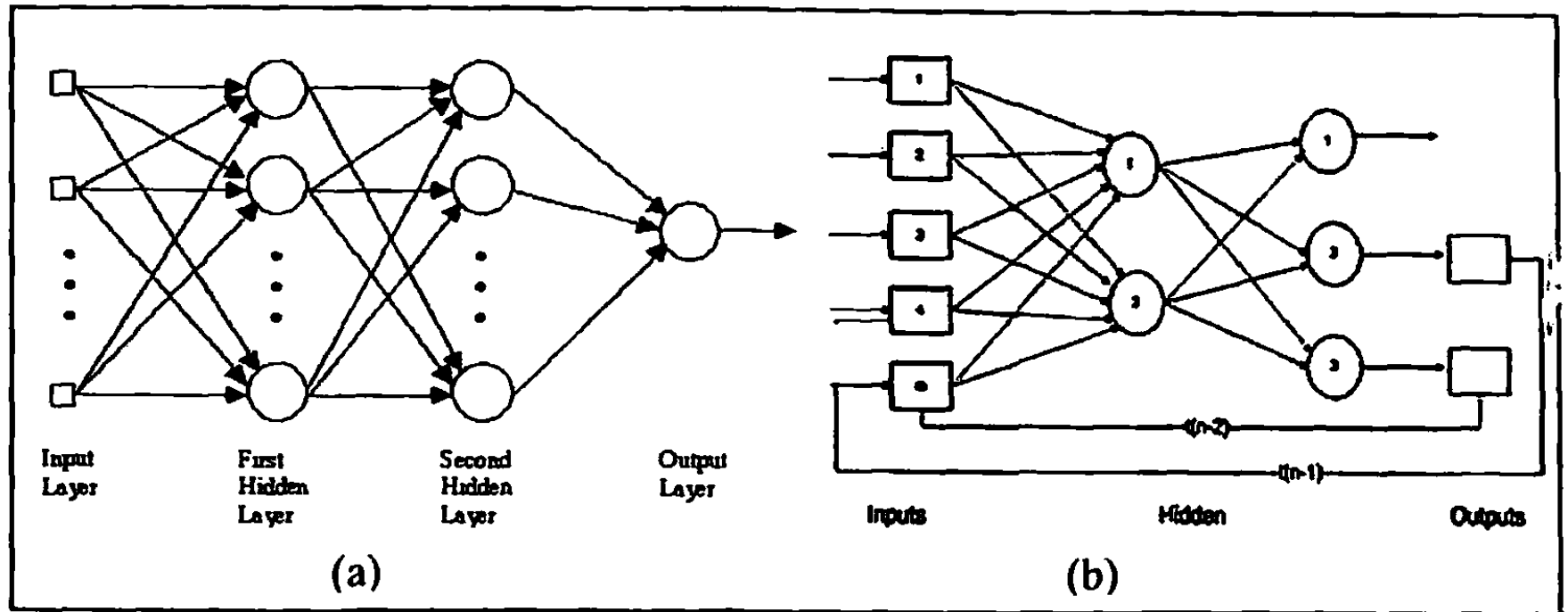


Figure 1: Architecture of (a) FFNN and (b) TDNN

There are large numbers of factors that could affect neural network training and model selections which include network architecture, activation function, training algorithm, initial weights, number of hidden layers, etc. Therefore all these factors should be utilized to come up with a best neural network model which can be used for prediction. Thus, a trial-and-error approach or cross-validation experiment is often help to find the best model. Typically a large number of neural network models are considered. The one with the best performance in the validation set is chosen as the winner, and the others are discarded.

Steepest gradient descends technique is used in learning algorithm when training neural networks which is based on minimizing the sum of squared error, E (see equation (1)) over all training data. During training, each desired output is compared with actual output and E is calculated as sum of squared error at the output layer:

$$E = \sum_{i=1}^N (a_i - p_i)^2 \quad (1)$$

where a_i is the actual value and p_i is the corresponding predicted value.

As ANN training and model selection affect by many factors which are mentioned above, we can use appropriate performance criteria to measure the performance in the validation process such as statistical metrics will be discussed.

The organization of this paper as follows: The next section describes the materials and methods. Third section illustrates the results and discussion. Last section contains conclusions and recommendations.

MATERIALS AND METHODS

Literature reveals that, the moving averages and historical data can be considered as inputs to neural networks in the process of predicting exchange rates (Kuan and Liu, 1995; Kamruzzaman and Sarker, 2003). Therefore, several potential input variables such as moving averages (5MA, 10MA, 20MA, 30MA, 40MA, 60MA, etc) and lag exchange rates were taken into account in the process of selecting suitable inputs to networks under this study. Descriptive statistics such as moving average graphs, auto correlation plots were used to identify the most suitable inputs among them in the preliminary analysis and those variables were used to further identify the combine effect of them towards the prediction of exchange rate of the USD/LKR.

Nearly ten years daily data of exchange rate of the USD/LKR (1st January 1998 to 28th November 2008) directly taken from Central Bank of Sri Lanka, were used in this study. This study period contains 2634 daily observations of the variable excluding holidays since exchange rates are not issued by Central Bank on holidays. If we use this whole data set to train, validate and test neural networks, it will train for much older period data and it suppose to predict for a very later period. It will lead to reduce accuracy of predictions. This problem can be overcome by using the concept of moving windows. Therefore, whole data set has been divided into three non-overlapping moving windows and each of these windows is used at a time to train, validate and test neural networks. Each window consists of training, validation and testing samples of appropriate proportions. 5% of the most recent data of a window was allocated for testing. The next most recent 15% of data was used for validation and remaining was used for training.

Three types of neural network models would be employed to predict the exchange rate of the USD/LKR which are namely feedforward neural network model trained with the backpropagation algorithm, feedforward neural network model trained with the scaled conjugate gradient algorithm and time delay neural network model.

To find the best performed static neural network model which uses the backpropagation algorithm (BPR), feedforward neural network was built with BPR as the learning algorithm. In the process of selecting the best model, setting

parameters of the training algorithm, setting number of hidden layers and neurons in each layer and selecting appropriate input combination are crucial factors which affect the accuracy of predictions. Firstly the neural networks were trained with different input combinations to identify the most suitable set of input combination. Then the best performed neural network architecture was selected from each approach separately. Most suitable number of hidden layers and number of neurons in each layer were identified in this process. Then parameters associated with learning algorithm were varied in different iterations and training and testing process were carried out a large number of times until come up with a satisfactory model with the use of performance metrics (see Performance Criteria) which were implemented as functions. Training, validation and testing of FFNN trained with the BPR algorithm was done iteratively a large number of times to find the above suitable factors. The model which showed highest performance was selected as the best performed model of FFNN with BPR to predict the USD against the LKR.

Same procedure was carried out using the Scaled Conjugate Gradient algorithm (SCG) and finally come up with the best performed model of FFNN with SCG algorithm to predict USD against LKR.

Similarly TDNN model was trained for different number of hidden layers, different number of neurons in hidden layers and different time delays and finally came up with the best performed dynamic neural network model to predict the USD/LKR.

The best performed FFNN models which use two learning algorithms were compared and found the best static neural network model for predictions. Finally a comparison was done using best performed static neural network model and best performed dynamic neural network model to find the most suitable model to predict exchange rate of the USD against the LKR.

Learning Algorithms

Backpropagation (BPR): Backpropagation is a commonly used systematic method of training multilayer artificial neural networks. It was built on high mathematical foundation and has very good application potential. The backpropagation learning algorithm is simple to implement and computationally efficient algorithm which uses steepest gradient descends technique. Even though it has its own limitations, it is applied to a wide range of practical problems and has successfully demonstrated its power.

$$\Delta w_j(n) = -\eta \frac{\partial E}{\partial w_j} + \alpha \Delta w_j(n-1) \quad (2)$$

The weight w_j is adjusted in the n^{th} training cycle according to the Equation (2) where the parameters η and α are learning rate parameter and momentum factor respectively. The learning rate is multiple times the negative of the gradient to determine the changes to the weights and biases and the parameter momentum is a constant that defines the amount of momentum which should be set between 0 and 1.

There are two different ways in which this gradient descent algorithm can be implemented which are namely incremental mode and batch mode. In incremental mode, the gradient is computed and the weights are updated after each input is applied to the network. In batch mode, all the inputs are applied to the network before the weights are updated. Batch mode is theoretically proven good method which is widely used when implementing backpropagation algorithm.

Scaled Conjugate Gradient (SCG): SCG algorithm was introduced as an improvement to the BPR algorithm which does not produce the fastest convergence (Møller, 1990). In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions.

There are four variations of conjugate gradient algorithms which are namely Fletcher-Reeves Update, Polak-Ribière Update, Powell-Beale Restarts and Scaled Conjugate Gradient. Each of the conjugate gradient algorithms except scaled conjugate gradient requires a line search at each iteration. This line search is computationally expensive, because it requires that the network response to all training inputs be computed several times for each search. The scaled conjugate gradient algorithm (SCG), developed by Moller was designed to avoid the time-consuming line search. SCG needs to calculate Hessian matrix which is approximated by the following equation.

$$E''(w_i)p_i = \frac{E'(w_i - \sigma_i p_i) - E'(i)}{\sigma_i} + \lambda_i p_i \quad (3)$$

σ and λ are the two parameters associated with this algorithm. The parameter σ determines the change in the weight for the second derivative approximation. The

parameter λ regulates the indefiniteness of the Hessian matrix. These two parameters should be as small as possible ($< 10^{-4}$) to come up with a better model for predictions.

Performance Criteria

Measuring performance is a crucial factor to come up with a best model and to draw conclusions. There are number of statistical metrics which are used to measure the performance of models. Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE), Directional symmetry (DS), Correct Up Trend (CU) and Correct Down Trend (CD) are some of the widely used performance metrics in time series predictions. Following equations used to calculate performance in each method.

$$NMSE = \frac{1}{\sigma^2 N} \sum_k (x_k - \hat{x}_k)^2 \quad (4)$$

where x_k is the actual value and \hat{x}_k is the corresponding predicted value of the k^{th} observation of the test set. N is the sample size and $\sigma^2 = \frac{\sum_k (x_k - \bar{x}_k)^2}{N}$

$$MAE = \frac{1}{N} \sum |x_k - \hat{x}_k| \quad (5)$$

$$DS = \frac{100}{N} \sum_k d_k^{(DS)} \quad (6)$$

where $d_k^{(DS)} = \begin{cases} 1 & \text{if } (x_k - x_{k-1})(\hat{x}_k - \hat{x}_{k-1}) \geq 0 \\ 0 & \text{Otherwise} \end{cases}$

$$CU = 100 \frac{\sum_k d_k^{(CU)}}{\sum_k l_k^{(CU)}} \quad (7)$$

$$\text{where } d_k^{(CU)} = \begin{cases} 1 & \text{if } (\hat{x}_k - \hat{x}_{k-1}) > 0, (x_k - x_{k-1}) (\hat{x}_k - \hat{x}_{k-1}) > 0 \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{and } t_k^{(CU)} = \begin{cases} 1 & \text{if } (x_k - x_{k-1}) > 0 \\ 0 & \text{Otherwise} \end{cases}$$

$$CD = 100 \frac{\sum_k d_k^{(CD)}}{\sum_k t_k^{(CD)}} \quad (8)$$

$$\text{where } d_k^{(CD)} = \begin{cases} 1 & \text{if } (\hat{x}_k - \hat{x}_{k-1}) < 0, (x_k - x_{k-1}) (\hat{x}_k - \hat{x}_{k-1}) > 0 \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{and } t_k^{(CD)} = \begin{cases} 1 & \text{if } (x_k - x_{k-1}) < 0 \\ 0 & \text{Otherwise} \end{cases}$$

NMSE and MAE measure the deviation between actual and forecasted value. Smaller values of these metrics indicate higher accuracy in forecasting. CU and CD measure the correctness of upwards and downwards trends respectively. DS indicates the overall prediction accuracy of the model. Larger values of DS, CU and CD show higher performance.

RESULTS AND DISCUSSION

Results of the preliminary analysis provided evidence about the significance of the potential influence variables to exchange rates of the USD/LKR. 5MA, 10MA, 20MA, 30MA, 40MA and previous day's exchange rate were identified as the most suitable inputs to the neural networks.

These variables were used as input features to the neural networks to identify the combined effect of them towards the prediction of exchange rate in each approach separately. The most suitable input combination for the feedforward neural network (FFNN) trained with the backpropagation (BPR) algorithm is 5MA, 10MA, 20MA, 30MA and 40MA. The same input combination shows highest performance in the FFNN trained with the scaled conjugate gradient (SCG) algorithm.

Whenever number of hidden layers in a neural network increases, it becomes more complex and slower to converge. Therefore many researchers (for example Kamruzzaman and Sarker, 2003) suggest designing neural networks with one or two hidden layers or using minimum number of hidden layers in a design. Therefore numbers of hidden layers were varied from one to four while keeping other factors constant and the most suitable number of hidden layers was selected in each approach using the performance measures. Static FFNN model which used the BPR algorithm shows highest performance when there are two hidden layers and FFNN trained with the SCG algorithm shows highest performance when there is only one hidden layer. When there are two hidden layers, TDNN model shows the highest performance.

After selecting number of hidden layers for networks, the number of neurons in each hidden layer needs to be fixed. For this purpose, number of neurons in each layer were varied and tested while keeping other factors constant for each approach. FFNN model trained with the BPR algorithm needed three neurons in first hidden layer and three neurons in second layer to achieve the highest performance. When there are four neurons in the hidden layer FFNN trained with the SCG algorithm shows highest performance. TDNN shows highest performance when there were three neurons in both hidden layers. These numbers are very subjective which means that they are totally depend on the data set used for training and can be take some other values when considering a different problem with a different data set.

When considering neural network models, selecting the most suitable parameters of the learning algorithm is an important factor which affects the accuracy of predictions of neural network models. Therefore, most suitable parameters of each learning algorithm were selected by varying and measuring performance of them in several iterations. Most appropriate parameters for the BPR algorithm used in FFNN model are learning rate (η) of 0.005 and the momentum (α) of 0.7. FFNN model trained with the SCG algorithm showed highest performance when σ and λ which are the parameters associated with the learning algorithm were set to 1×10^{-12} and 5×10^{-15} respectively.

Selecting appropriate time delay affects the prediction accuracy of TDNN. Therefore, number of time delays was varied while keeping other factors constant to come up with most suitable time delay to predict the USD/LKR. This process was done iteratively a large number of times and finally found that TDNN model shows highest performance with six time delays than the other number of time delays.

Table 1 shows the forecasting results measured in terms of the performance metrics of best performed models selected from each approach.

Table 1: Performance of best performed models selected from each approach

Model	Performance				
	NMSE	MAE	DS	CU	CD
FFNN(BPR)	1.7185	0.3769	62.8596	57.5476	59.1198
FFNN(SCG)	0.4849	0.1493	64.9259	56.0714	53.1515
TDNN	0.08646	0.06834	76.0917	72.8575	68.8132

The best performed FFNN trained with the BPR algorithm showed 62% accuracy of directional predictions (Table 1). Upward prediction accuracy was 57% and downward prediction accuracy was 59% of that model. Prediction errors (Normalized Mean Squared Error-NMSE and Mean Absolute Error- MAE) also showed smaller values around one.

According to Table 1, NMSE and MAE are 0.48499 and 0.14935, respectively for the best performed FFNN model trained with the SCG algorithm. It can predict upward trends with 56% accuracy and downwards trends with 53% accuracy. Furthermore, this model shows 65% of directional prediction accuracy.

The best performed time delay neural network model showed 76% prediction accuracy in the context of directional symmetry (Table 1). It is able to predict up trend with 73% accuracy and down trend with 69% accuracy. Furthermore, prediction errors are very small values (NMSE= 0.08646 and MAE=0.06834).

As per the results obtained in this study, FFNN trained with the SCG algorithm performs better than the FFNN trained with the BPR algorithm in terms of directional symmetry and errors. The best performed TDNN model performs better than both FFNN best performed models in terms of all performance metrics.

Time series plots of predicted/actual exchange rates against index were drawn in order to give clear idea of the accuracy of the predictions of best performed models selected from each approach.

Figures 2(a), 2(b) and 2(c) illustrates the time series plots correspond to best performed models selected from FFNN trained with the BPR algorithm, FFNN trained with the SCG algorithm and TDNN respectively. Vertical lines were used to separate three test sets correspond to three moving windows to avoid misinterpretations. By referring Figure 2(a) and 2(b), it can be identified that there is a visible difference in predicted and actual values of Window 2. High deviation of predicted values may be due to variation of exchange rate due to the political disturbances during the period January 2001 to January 2004. In Figure 2(a), predictions correspond to Window 3 are more accurate in first half and predicted values deviate from their actual values in the second half of data. Figure 2(c) displays that, the predicted values by the best performed TDNN model are very much closer to their actual values in all three windows.

Therefore, FFNN model trained with the SCG can be suggested as better static neural network to predict the USD/LKR than FFNN trained with BPR algorithm and dynamic time delay neural network model can be identified as the most appropriate neural network model among all three models considered to predict exchange rate.

The government of Sri Lanka can make use of this best neural network model for future planning and investments. Export/Import businessmen, bankers and investors can make use of this model to make their decisions.

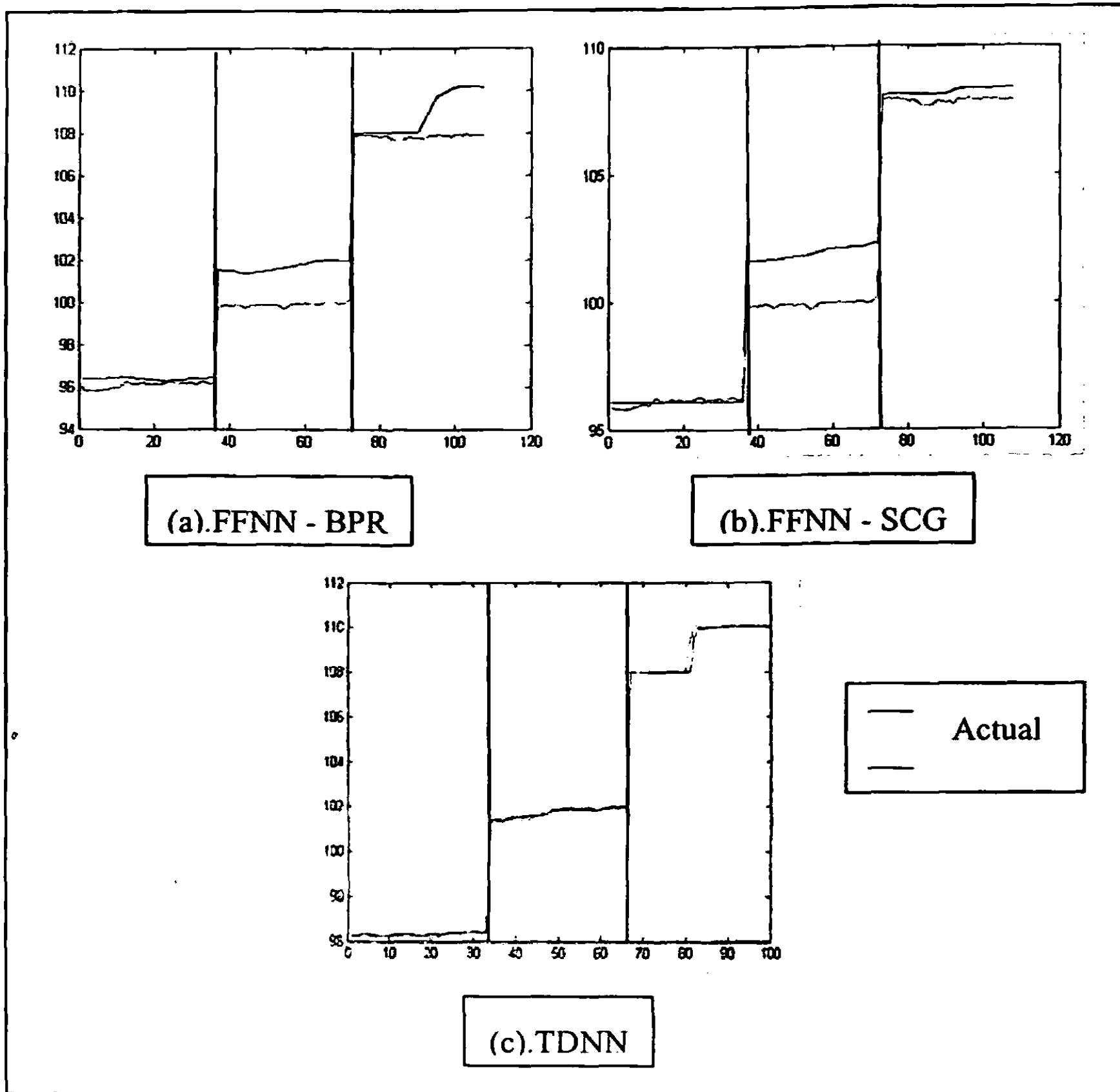


Figure 2: Time series plots of actual/predicted of test data in best performed models of (a) FFNN with the BPR algorithm, (b) FFNN with the SCG algorithm and (c) TDNN

CONCLUSIONS AND RECOMMENDATIONS

The best performed feedforward neural network model trained with the backpropagation algorithm shows 62% prediction accuracy. Prediction accuracy of the best performed feedforward model trained with the scaled conjugate gradient

algorithm is 65%. Experimental results show that the best static neural network model to predict the USD/LKR is the feedforward model trained with the SCG algorithm. The best performed time delay neural network model shows 76% prediction accuracy which contains two hidden layers, three neurons in each layer and six time delays. Dynamic neural network model is the most appropriate model to predict exchange rate of the US Dollar against the Sri Lankan Rupee.

These conclusions are based on the considered study period and the input data set. However these results may differ according to the market, the study period and the two currencies considered for the exchange rate. Therefore, conclusions of this study are subjective.

In this study exchange rate of the US Dollar to the Sri Lankan Rupee was considered as the US Dollar is the major player of the exchange market. This study can be further extended to predict exchange rates of other currencies as well.

Only the feedforward and the time delay neural network architectures were used in this study. There are many other neural network models which can be used for time series predictions such as recurrent neural network, FFNN trained with the Bayesian Regulation algorithm. These techniques can be used to find the best performed models from each approach and they can be compared to find which neural network architecture would predict exchange rate more accurately.

As literature reveals, the neural network can be used to build hybrid systems (Kodogiannis and Lolis, 2001). As a further improvement a hybrid neural network model can be applied to predict exchange rates.

REFERENCES

- Binner, J.M., R.K. Bissoondeal, T.Elger, A.M. Gazely, A.W. Mullineux (2005). A comparison of linear forecasting models and neural networks: An application to Euro inflation and Euro divisia. *Applied Economics*, 37: 665-680.
- Giles, C.L., S. Lawrence and A.C. Tsoi (2001). Noisy time series prediction using a recurrent neural network and gramatical inference. *Machine Learning*, 44: 161-183.

- Hill, T., M. O'Connor and W. Remus (1996). Neural network models for time series forecasts. *Management Science*, 42: 1082-1092.
- Kaboudan, M. (2005). Computational forecasting of two exchange rates. Retrieved on 25/09/2008 from School of Computer Science, University of Birmingham. < <http://www.cs.bham.ac.uk>>.
- Kamruzzaman, K. and R.A. Sarker (2003). Comparing ANN based models with ARIMA for prediction of Forex Rates. *Asor bulletin*, 22.
- Kodogiannis, V. and A. Lolis (2001). Prediction of exchange rates by neural network and fuzzy System based techniques. *European Symposium on Artificial Neural Networks*, 25: 245-250.
- Kohara, K., T. Ishikawa, Y. Fukuhara, Y. Nakamura (1997). Price prediction using prior knowledge and neural Networks. *Intellegent systems in Accounting, finance and Management*, 6: 11-22.
- Kuan, C.M. and T. Liu (1995). Forecasting exchange rates using feedforward and recurrent neural networks, *Journal of Applied Econometrics*, 10: 347-364.
- Macdonald, R. (1995). Long-run exchange rate modelling: A survey of the recent evidence. *Palgrave Macmillan Journals*, 42: 437-489.
- Møller, M.F., (1990). A scaled conjugate gradient algorithm for fast supervised learning. Department of Computer Science, University of Aarhus, Denmark.
- Pan, H., C. Tilakaratne and J. Yearwood (2005). Predicting Australian Stock Market Index using neural networks exploiting dynamical swings and intermarket influences. *Journal of Research and Practice in Information Technology*, 37: 43-54.
- Tilakaratne, C.D., M.A. Mammadov and S.A. Morris (2007). Effectiveness of using quantified intermarket influence for predicting trading signals of Stock Market. Conference in Research and Practice in Information technology (CR.PIT). 70: 167-175.
- Yao, J., C.H. Tan and H.L. Poh (1999). Neural networks for technical analysis: A study on KLCT. *International Journal of Theoretical and Applied Finance*, 2 (2): 221-241.