Predicting Exchange Rates with Modified Elman Network※

Beum-Jo Park*

수정된 엘만신경망을 이용한 외환 예측*

박 범 조**

ABSTRACT

This paper discusses a method of modified Elman network(1990) for nonlinear predictions and its application to forecasting daily exchange rate returns. The method consists of two stages that take advantages of both time domain filter and modified feedback networks. The first stage straightforwardly employs the filtering technique to remove extreme noise. In the second stage neural networks are designed to take the feedback from both hidden-layer units and the deviation of outputs from target values during learning. This combined feedback can be exploited to transfer unconsidered information on errors into the network system and, consequently, would improve predictions. The method appears to dominate linear ARMA models and standard dynamic neural networks in one-step-ahead forecasting exchange rate returns.

Keywords : Predictions of Daily Exchange Rate Returns, Nonlinearity, BDS statistic, Time Domain Filter, Artificial Neural Networks, Backpropagation.

※ I thank Chung-Min Kuan for his helpful comments and Halbert White and Francis Diebold for their concerns. The remaining errors, of course, are mine. An earlier version of this paper was presented at the Korean Econometric Association Meeting.

** Department of Economics, Dankook University

Ⅰ. INTRODUCTION

Exchange rate predictions have been the subject of one of the most active bodies of literature in economics since the inception of floating exchange rate system in the early 1970s. The prediction models have usually been constructed under assumption of linearity. In practice, however, their results of predicting the short-term movements of exchange rates have been unsatisfactory although they are straightforward to practice and they are supported by a theorem of Wold (1938) that any stationary process can be represented as a linear system generating uncorrelated impulses.

Krugman (1988) showed that the failure of forecasting short-term exchange rates with the standard models might be largely due to nonlinearity in time patterns of exchange rates. This indication has received extensive support from both theoretical studies such as a speculative bubble model (Blanchard and Watson, 1982), target zone theory (Krugman, 1988), etc. and empirical studies (Diebold, 1988; Baillie and Bollerslev, 1989; Hsieh, 1988, 1989; Diebold and Nason, 1990; Chinn, 1991; Meese and Rose, 1991; Park, 1997 etc.). Thus, nonlinear methods make it possible to predict short-term movements of exchange rates, giving significant improvements over the usual linear methods.

The one of the most appealing nonlinear methods is artificial neural network (ANN) presented by Rummelhart and McCelland (1986). Since artificial neural networks were originally designed for reproducing some flexibility and power of the human brain by artificial means, they have been widely used for a nonlinear method in many fields and their numerous empirical applications have been satisfactory. In recent days, there has also been rapidly increased concern about applying them to economics and finance after the first effort to use them in predicting IBM stock prices by White (1988).

Their application to exchange rates stems from Kuan and Liu (1992), who investigate the out-of-sample forecasting ability of feedforward and recurrent networks on five daily exchange rates against the U.S. dollar, including the British pound, the Canadian dollar, the Deutsche mark, the Japanese yen and the Swiss franc. The empirical results show that the neural networks perform reasonably well in terms of sign forecasts. However, as far as out-of-sample MSE being concerned, the results are consistent with the conclusion of Diebold and Nason (1990) that nonlinearities of exchange rates, if any, may not be exploited to improve point forecast. The results in Tsibouris (1993) also support the evidence that the neural networks are useful in forecasting the sign of the exchange rate changes, but not the magnitude. Weigend et al. (1992)

present results from forecasting sunspots and exchange rates using a form of neural net algorithms that avoids overfitting to data by the method of weight-elimination. They show that the result of sign prediction is better than chance (i.e., random walks). Zhang and Hutchinson (1993) apply multilayer perceptrons to forecast exchange rates with data, which are tick-by-tick Swiss franc to U.S. dollar exchange rates. On average for 1-minute, 15-minute, and 60-minute predictions, the predictions of the networks have about 5% improvement in Root-Mean-Squared errors over the random walk model, although the error distribution is quite different in two cases.

Other studies of forecasting exchange rates using ANN are Abu-Mostafa (1995), who reports a statistically significant improvement in performance in four major foreign exchange markets by an ANN with a simple symmetry hint, and Hsu et al. (1995) who use an ANN to select predictive indicators and show that the forecasting accuracy of direction is better than that from the unprocessed universe of indicators.

The reasons for wide application of the ANN to prediction of exchange rates are obvious. Artificial neural networks are data-driven modeling approaches and let the data decide the structure and parameters of a model without any restrictive parametric modeling assumptions. Therefore, they seem fully suited to fit a wide variety of nonlinear functions. More practical reason is that, in contrast to other nonlinear methods such as polynomial expansion, the speed complexity of artificial neural networks may be given as a linear form because the number of parameters goes up almost linearly with respect to the number of inputs (Barron, 1991).

Despite their adaptability to various nonlinearities, like other nonlinear methods they may have unavoidable limitation in predicting exchange rates. The idea of this paper is that models for only nonlinearity in exchange rates may not significantly improve predictions. Thus, this paper proposes a certain enhancement to go with a method of modified Elman network (1990) (MMEN) for overcoming the limitation of prediction models. That is, this method (MMEN) has a two-stage procedure that takes advantages of both time domain filter and modified recurrent networks. Since high noise in time series may have an undesirable effect on the learning process of neural networks, it may give rise to a poor capacity to generalize in respect of out-of-sample. The first stage straightforwardly employs the filtering technique to remove the high noise. The time domain filtering, then, can produce better inputs to neural networks. On the other hand, simple time delayed inputs in prediction models may not enough to explain a variety of fluctuating patterns of exchange rate returns, which introduce a complex and highly dynamic behavior, because the inputs cannot take into account all economic and non-economic factors greatly influencing exchange rate returns.

The key point of this paper is that it is at least that undetected regularities caused by the omitted factors exist in errors and the modification of feedback neural networks can extract the regularities from the errors in the models during learning. That is, in the second stage the recurrent networks are designed to take the feedback from not only hidden-layer units but also the deviation of outputs from target values during learning. This combined feedback can be exploited to transfer the information in errors into the network system and to considerably improve convergence in terms of both accuracy and speed.[1]

Empirical studies are conducted by using changes of exchange rates of the Korea won relative to the US dollar (UD), the Japanese yen (JY), the Deutsche mark (DM), and the British pound (BP). It is well known that different exchange rates follow different behaviors so that the statistical features of the exchange rate data used here should be analyzed carefully. In order to diagnose the possibility of prediction, the dependence tests of the exchange rate returns are particularly carried out by Ljung-Box Q, Tsay (1986), and Brock-Dechert-Scheinkman (BDS, 1987) tests. For the UD and the JY there is strong evidence of linear and nonlinear dependence but for the DM and the BP not any dependence. This implies that predictions of the UD and the JY can be improved by nonlinear time series models but the DM and the BP cannot be predicted well by any simple time series model because they seem follow random walk processes. As expected, in one-step-ahead predictions of all the exchange rate returns, the MMEN appears to dominate linear ARMA models and dynamic neural networks such as time delay neural networks (TDNN).

The structure of this paper is as follows. In the second section, standard dynamic neural networks are introduced. The third section is devoted to propose the approach for predictions of exchange rates. In fourth section empirical studies are carried out and the results are reported. Finally, the fifth section draws some conclusions.

II. ARCHITECTURE OF TIME DELAY NEURAL NETWORKS

This section introduces dynamic neural networks briefly because most of the economists are still not familiar with artificial neural networks.[2] Neural networks can approximate a wide variety of

1) It is widely recognized that for highly dynamic series such as exchange rate series neural networks generally suffer from extremely slow learning by using algorithms based on gradient descent networks.
2) For further details see Kuan and White (1994).

nonlinear functions arbitrarily well, proved that their architectures include sufficiently large hidden units (Hornik, Stinchcombe, and White, 1989; Funahashi, 1989). Thus, neural networks have been receiving growing attention for modeling or forecasting nonlinear system in many fields.

For time series predictions, the most popularly used neural networks are clearly time delay neural networks (TDNN; Weigend, Huberman, and Rumelhart, 1990) and recurrent neural networks (RNN; Elman, 1990). The time delay neural networks can be analyzed by using standard statistical methods and more the results of such analysis can be applied for time series predictions directly, but they may not be sufficient to characterize the patterns of highly dynamic time series. On the other hand, the recurrent neural networks are suited for applications that refer to the patterns of genuinely time dependent inputs such as time series predictions due to their dynamic feature. That is, the recurrent neural networks, in which the input patterns pass through the network more than once before generating a new output pattern, can learn extremely complex patterns. Several experts on neural networks have confirmed the superiority of the recurrent neural networks over feedforward networks when performing nonlinear time series forecast (for example, Connor and Atlas, 1991)[3].

This section needs to tensely describe the time delay neural networks because they have more basic architecture than the recurrent neural networks and Kuan and Liu (1992) illustrate that the time delay neural networks are performed well in predicting sign of exchange rate returns. Architecture of time delay neural networks with single hidden layer is shown in figure 1[4].

- The initial p inputs are given a time delay vector,

$$x = [x_t x_{t-1} \cdots x_{t-p+1}]$$

- The inputs are connected to L hidden units (N_l) via a nonlinear activation function that is generally a monotonic nondecreasing function such as hard limiter function, threshold logic function, or sigmoid function. The goal of nonlinearity is to enhance the features of inputs.
- The hidden units are also connected to a target via a linear function and an output is produced as the weighted sum of the activations of the hidden units.
- Adjustable biases can be used between the inputs and the output.

3) While in the dynamic context the recurrent neural networks can outperform the time delay neural networks, they occasionally are difficult to be trained optimally by a standard backpropagation algorithm due in part to the dependence of their network parameters (Kuan and Hornik, 1991).

4) It turned out that single hidden layer may be enough to analyze time series (Hornik, Stinchcombe, and White, 1989).

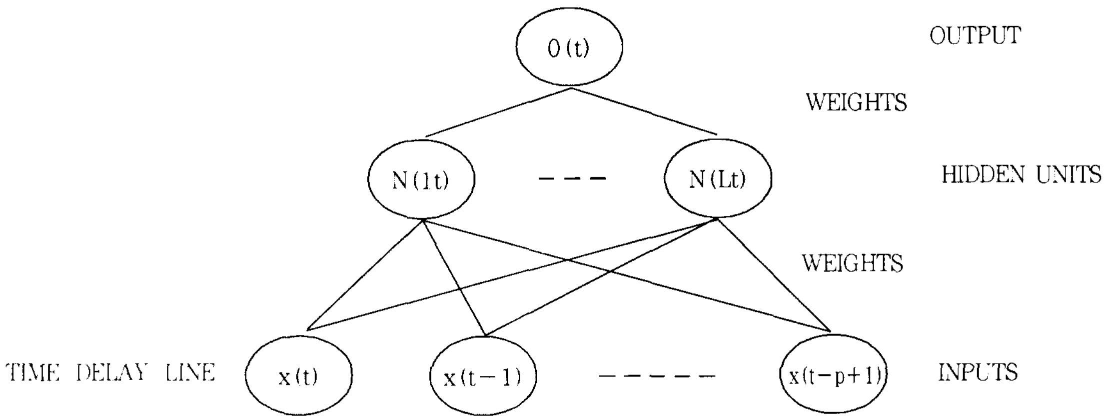

Figure 1. Architecture of TDNN

Architecture of the time delay neural networks can be mathematically rewritten as

$$N_{it} = h(\sum_{i=1}^{p} W_{ij} x_{jt}), \qquad t=1, \cdots T$$

$$O_t = g(\sum_{i=1}^{L} B_i N_{it}),$$

where N_i is the vector of hidden units, x_j is the vector of inputs, (Adjustable biases, x_0 and N_0 can be added as a vector of constant ones), W_{ij}, β_i are parameters called weights, and h, g are known activation functions.

As indicated earlier, the time delay neural networks may have limitations of accurately recognizing the patterns of highly dynamic time series. To improve the recognition of the patterns, we can consider recurrent neural networks that take results of processing at a particular time step and feed data back into the network inputs at the next time step. Since the output of the recurrent neural networks is a function of current input and its entire history, in the dynamic context the recurrent neural networks can outperform the time delay networks. Mathematically, the architecture of the recurrent neural networks appears as

$$N_{it} = h(\sum_{i=1}^{p} W_{ij} x_{jt} + \sum_{l=1}^{q} v_{il} N_{i(t-1)}), \qquad t=1, \cdots T$$

$$O_t = g(\sum_{i=1}^{L} \beta_i N_{it}),$$

The time delay neural networks are nonlinear expansions of least squares algorithm so that they can statistically approximate a desirable function through error-backpropagation learning algorithm that is used most commonly. This algorithm is a steepest descent algorithm, which minimizes the total quadratic error,

$$E = \frac{1}{2}\sum_{t=1}^{T}(y_t - O_t)^2$$
$$= \frac{1}{2}\sum_{t=1}^{T}(y_t - g(N_{ot}))^2$$

To get the steepest descent direction of the total quadratic error, we need to calculate the gradient of E with respect to β_i, which is realized using the chain rule.

$$\frac{\partial E}{\partial \beta_i} = -(y_t - O_t)\frac{\partial g}{\partial N_{ot}}\frac{\partial N_{ot}}{\partial \beta_i}$$

When $\partial g \partial N_{ot} = g', \partial N_{ot} \partial \beta_i = I_i$, and $(y_t - O_t)g' = \delta^o$, β_i is updated using the following equation,

$$\beta_i(k+1) = \beta_i(k) + \eta\,\delta^o I_i$$

where η is the learning rate.

By a similar way, we can calculate the gradient of E with respect to W_i.

$$\frac{\partial E}{\partial W_i} = -h' x_t (y_t - O_t) g' \beta_i,$$

where $h' = \partial h / \partial W_i$. When $h' x_t (y_t - O_t) g' \beta_i = \delta^h$ W_i is also updated using the following equation,

$$W_i(k+1) = W_i(k) + \eta\,\delta^h x_t$$

III. THE METHOD OF MODIFIED ELMAN NETWORK

In practice, even dynamic neural networks with reasonable complexity fail to learn exchange rate returns satisfactorily, and therefore their prediction ability is not sufficiently powerful. Possible reasons for this may be that exchange rate returns include extreme noise, and more seriously, most of them may not be strongly correlated over short time period. The main idea of this paper is that simple time delayed inputs are not enough to explain the complex dynamics of exchange rate returns, and intuitively we need additional inputs which can be unmeasurable but important factors for recognizing the complex dynamics.[5]

How can we consider the unmeasurable factors as inputs? Although this seems to be a dilemma, we can deal with it by modifying the standard recurrent neural networks introduced by Elman in 1988. At one particular epoch k, feedback from the hidden layer units becomes additional input units to the networks at the next epoch $k+1$ and so does the deviation of output from the target value. With the combined feedback, it is clearly possible to regard the deviation (i.e., error) from the previous epoch as an input and to allow the network system to keep unattainable information on the error. That is, at the outset of training, the error being fed back to an input may significantly influence the network output but other additional inputs from the hidden layer may not. As training with backpropagation algorithm is going on, the fit improves and the error sufficiently decreases. Then the error little affects the network output because information on the error is likely to be transferred into other additional inputs. This means that the networks can extract unmeasurable but relevant information from the error. Therefore, it is expected that the networks predict the future movements in time series well based upon accurate pattern recognition from the past.

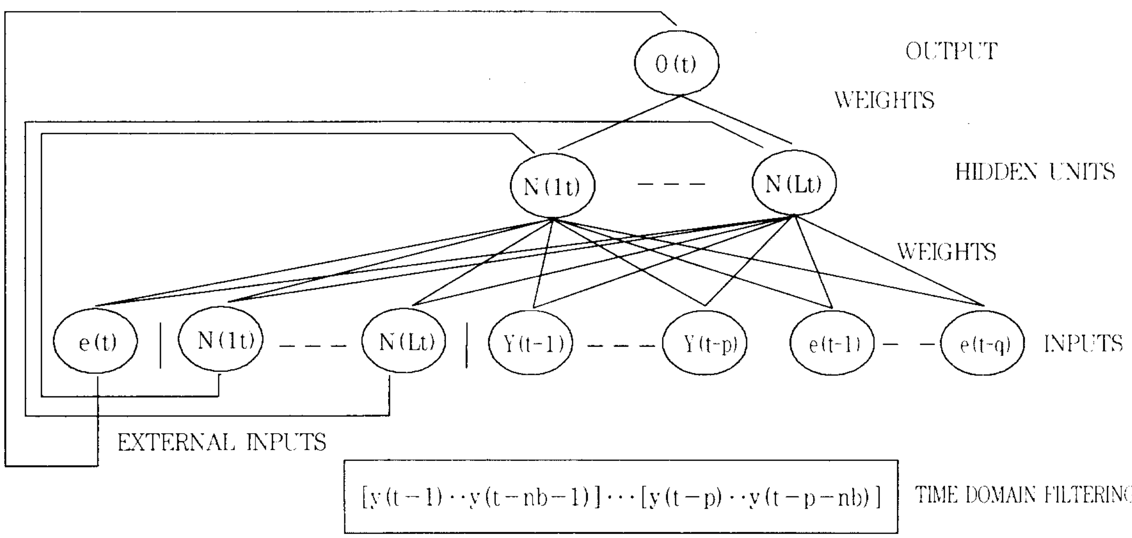

Figure 2. Architecture of MMEN

5) Economic variables can be considered as additional and measurable inputs but they alone may hardly play a role in predicting the complex dynamics because, since the beginning of 1970s, economic models have little explained the exchange rate returns but as a very long run theory.

For network computing I propose a two-stage procedure shown in figure 2. It takes advantages of both time domain filtering and the networks on the basic idea that networks do not have to be constructed by one complicated architecture. First, since the neural networks proposed here can lead to an arbitrarily accurate fit on data with undesirable noise and thus this overfitting may have an unwanted effect upon out-of-sample generalization, the time domain filter is carried out to enhance inputs by removing noise.

This filter structure is the general tapped delayline filter described by the equation,

$$Y_t = \sum_{i=1}^{nb} b_{i+1} y_{t-i} + \sum_{i=1}^{na} a_{i+1} Y_{t-i},$$

where y_t is the original series, Y_t is the filtered series, and b_i, a_i are the filter coefficients that can be obtained from a filter design such as a digital Butterworth filter (Parks and Burrus, 1987). Next, to understand underlying patterns from the dynamic series, we consider a nonlinear ARMA (p, q)model with the neural networks:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}, e_{t-1}, e_{t-2}, \cdots, e_{t-q}) + e_t,$$

where f is the neural networks. In cases where a high order is required in AR process, a ratio of low-order AR and MA processes frequently leads to parsimonious representations. Thus, the model compares favorably with the simple time delay neural networks: $y_t = f(x_t, x_{t-1}, \cdots, x_{t-p+1}) + e_t$.

The neural networks with single hidden layer in figure 2 transform their inputs into an output by the equation below,

$$N_{it}^{k+1} = h\left(\sum_{j=1}^{p+q} W_{ij}^{k+1} x_{jt} + \sum_{l=1}^{L} v_{il}^{k+1} N_{lt}^{k} + \delta^{k+1} e_t^k\right),$$

$$O_t^{k+1} = g\left(\sum_{l=1}^{L} \beta_i^{k+1} N^{k+1}\right),$$

$$i = 1, \cdots, L;\ t = 1, \cdots, T,$$

where x_t is a $p+q$ vector of input variables (Bias inputs are not considered here) and superscript k denotes the index of epoch. More compactly, the above equations can be rewritten as

$$O_t^{k+1} = g\left[\sum_{l=1}^{L} \beta_i^{k+1} h\left(\sum_{j=1}^{p+q} W_{ij}^{k+1} x_{jt} + \sum_{l=1}^{L} v_{il}^{k+1} N_{lt}^{k} + \delta^{k+1} e_t^k\right)\right]$$

$$= : \mathrm{f}(x, \xi)$$

where ξ is the vector of all parameters, $\xi = (\beta, W, v, \delta)$.

The inputs x_t at k epoch activate each hidden unit in the hidden layer through the tan-sigmoid function h and the hidden-unit activations N^{t} feed back to the input layer at $k+1$ epoch. Then, they serve to keep unmeasurable information from the error that is the deviation of network output O^t from the target value Y^t. As network experience accumulates, the error fades out and it can be treated as zero. Consequently, the network output is jointly determined by x^t and N^t. This means that we can directly apply the networks to predict future changes of time series because x^t are inputs and N^t are created from the networks.

Feedback variables are created from the networks and so they depend upon the network parameters β and W. Sometimes, the dependence of feedback variables can make the standard backpropagation algorithm infeasible because the derivatives of f with respect to the network parameters are able to be calculated incorrectly and the algorithm can take an appropriate gradient search direction (Rumelhart, Hinton, and Williams, 1986). Feeding error back into input, however, make it possible to overcome the problem by considerable improvement of convergence, and the recurrent neural networks, which feed error back into input, can be trained to respond to outputs by the standard backpropagation algorithm because if the standard backpropagation algorithm takes a wrong gradient search direction, feeding error back into input makes it to take a new search direction compensating for the wrong search direction at the next step.

Ⅳ. PREDICTING EXCHANGE RATES

The data used here are based upon daily prices of the Korea won relative to four major currencies: the US dollar (UD), the Japanese yen (JY), the Deutsche mark (DM), and the British pound (BP). The data contain a total of 1,138 observations from March 2 1990 to December 29 1993[6]. It has been well known that foreign exchange rates are usually nonstationary. For avoiding problems arising from the nonstationarity I took the natural logarithmic differences between two successive trading days. Let s_t be exchange rate at time t and the log differences of exchange rates are defined as $y_t = \log(s_t / s_{t-1}) \cdot 100$.

6) In Korea fixed rate system was changed into managed floating in 1980 and exchange rates have been able to be more freely determined in the foreign exchange markets since March 2 1990.

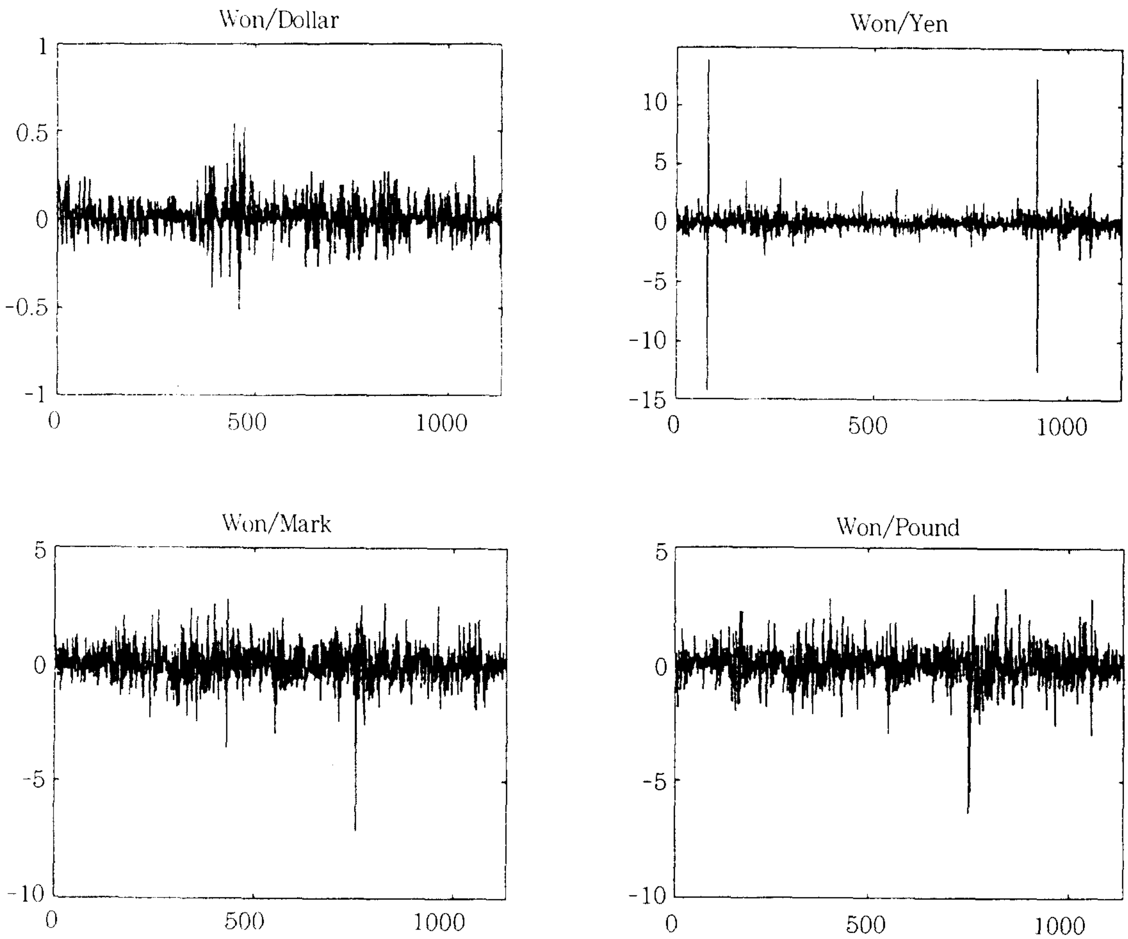

Figure 3. Data Plots : $y_t = log(s_t / s_{t-1}) \cdot 100$

Each time series is graphed in figure 3. These data plots show that the exchange rate returns are stationary but highly dynamic. Further, they seem to include extreme noise. The UD and the JY tend to be clustered together over time and this tendency may exhibit time series dependence such as ARCH process (Engle, 1982). Table 1 also reports some statistical properties of the data. The means and the medians are very small but the ranges of daily changes except the UD are relatively high. The range for the JY is, especially, between -14.2455 and 13.8666. It would appear that all the daily returns are not normally distributed according to the estimates of skewness and kurtosis. The high values of the kurtosis estimates also suggest a possibility that the data are not independent and identically distributed over time (Hsieh, 1988). The somewhat expected results imply that it may be quite difficult to analyze basic behaviors of the exchange rate returns and to predict their future changes.

Table 1. Summery Statistics of Changes of Log Exchange Rates

	UD	JY	DM	BP
Mean	0.0134	0.0398	0.0146	0.0048
Median	0.0000	0.0127	0.0144	0.0159
Variance	0.0090	1.0514	0.5826	0.6029
Skewness	0.2014	-0.2749	-0.6897	-0.7195
Kurtosis	3.6633	99.7181	7.7141	7.1897
Maximum	0.5421	13.8666	2.7746	3.2839
Minimum	-0.5140	-14.2455	-7.1246	-6.3775

In order to examine the presence of any linear or nonlinear dependence several test methods are carried out. At first, to detect any linear dependence standard errors of the autocorrelation coefficients, ρ_M, are computed and the Ljung-Box Q statistics for M lags are also computed as

$$Q(M) = T(T+2) \sum_{j=1}^{M} \frac{1}{T-j} \hat{\rho}_j^2.$$

Under the null hypothesis of non autocorrelation, Q is asymptotically a chi-square distribution with M degrees of freedom. According to the autocorrelation coefficients and their standard errors shown in table 1, the first coefficient for the UD and the JY is statistically different from zero at the 5% significance level but for the MD and the BP no coefficient may be statistically different from zero. The joint test that the first M autocorrelation coefficients are zero is carried out for M=50 and M=100. At the 5% significance level, the null hypothesis is rejected for the UD and the JY but not rejected for the MD and the BP. Consequently, it appears that the UD and the JY are linearly correlated but the MD and the BP are not.

According to a number of previous studies, it is possible for the exchange rate returns to have linear and nonlinear dependence simultaneously. Thus, even if evidence of linear dependence is not detected, the exchange rates can have nonlinear dependence alone, so we need to investigate carefully whether they exhibit nonlinear dependence. The fundamental techniques are adopted to compute standard errors of the autocorrelation coefficients and the Ljung-Box Q statistics of the squared data because the presence of nonlinear dependence can imply the correlation of the squared data in part. Table 2 indicates that for the all exchange rates except the DM the autocorrelation coefficients and the Ljung-Box Q statistics of the squared data are remarkably larger than those of the original data and thus, with the exception of the DM, all the returns may have nonlinear dependence.

Table 2. Autocorrelation Coefficients and the Ljung-Box Q

Original Data : y_t				
Lags	UD	JY	DM	BP
ρ_1	2169*(.0296)	-.3049*(.0296)	-.0212(.0296)	.0117(.0296)
ρ_2	-.0393(.0310)	-.0153(.0322)	-.0249(.0296)	.0006(.0296)
ρ_3	-.0610(.0310)	-.0054(.0323)	.0075(.0296)	.0491(.0296)
ρ_4	-.0654(.0311)	-.0181(.0323)	-.0432(.0296)	-.0286(.0297)
ρ_5	.0186(.0312)	.0029(.0323)	.0392(.0297)	.0550(.0297)
ρ_6	.0614(.0313)	-.0010(.0323)	-.0197(.0297)	-.0067(.0298)
ρ_7	-.0352(.0314)	-.0048(.0323)	-.0396(.0298)	-.0690(.0298)
ρ_8	-.0586(.0314)	.0098(.0323)	.0368(.0298)	.0196(.0299)
ρ_9	-.0692(.0315)	.0011(.0323)	-.0235(.0298)	-.0111(.0299)
ρ_{10}	.0289(.0316)	-.0039(.0323)	.0209(.0299)	.0590(.0300)
ρ_{15}	-.0245(.0320)	.0246(.0323)	-.0349(.0300)	.0010(.0302)
ρ_{20}	.0051(.0322)	.0127(.0324)	.0073(.0302)	-.0071(.0302)
Q(50)	207.15[.000]	130.91[.000]	51.01[.4334]	47.71[.5655]
Q(100)	248.48[.000]	171.16[.000]	95.94[.5961]	103.33[.3897]

Squared Data : y_t^2				
Lags	UD	JY	DM	BP
ρ_1	1663*(.0296)	.4921*(.0296)	.0287(.0296)	.0491(.0296)
ρ_2	.1749*(.0304)	-.0069(.0361)	.0339(.0296)	.1065*(.0297)
ρ_3	.1959*(.0313)	-.0057(.0361)	.0325(.0297)	.3012*(.0300)
ρ_4	.0895*(.0323)	-.0053(.0361)	.0496(.0297)	-.0047(.0326)
ρ_5	.0680(.0326)	-.0063(.0361)	.0422(.0298)	.0828(.0326)
ρ_6	.0820*(.0327)	-.0049(.0361)	.0341(.0298)	-.0025(.0327)
ρ_7	.1004*(.0329)	-.0057(.0361)	.0028(.0299)	.0221(.0327)
ρ_8	.0744 (.0331)	-.0025(.0361)	.0265(.0299)	.0076(.0328)
ρ_9	.1158*(.0333)	-.0007(.0361)	-.0047(.0299)	.0119(.0328)
ρ_{10}	.0711(.0336)	-.0050(.0361)	.0050(.0299)	-.0060(.0328)
ρ_{15}	.1431*(.0375)	-.0020(.0361)	.0133(.0303)	.0110(.0333)
ρ_{20}	-.0038(.0386)	-.0030(.0361)	.0453(.0304)	-.0013(.0334)
Q(50)	556.14[.0000]	277.24[.0000]	54.78[.2978	173.33[.0000]
Q(100)	655.87[.0000]	278.83[.0000]	75.99[.9647]	221.23[.0000]

Notes: SEs of ρ are in parentheses and significance levels are in brackets.
* Significantly different from 0 at the 5% level.

To find strong support for it, we can run Tsay (1986) test that includes more combinations of cross-product terms. The first step of the test is to fit an AR(2) to y_t and to save the estimated residuals, $\hat{e}_t$. The next step is linear regression of the vector (y^2_{t-1}, $y_{t-2}y_{t-1}$, y^2_{t-2}) on the vector(1, y_{t-1}, y_{t-2}), saving the estimated residuals as $\hat{z}_t$. *The final step is to regress* $\hat{e}_t$. on $\hat{z}_t$. and to run $F_{3,\ T-4}$ at conventional significance level. As expected, the results of Tsay test also provide strong evidence of nonlinear dependence for the UD and the JY but not for the DM and the BP.

Now, to merely pick up nonlinear dependence BDS test is applied to the residuals of AR(5) in which linear dependence may be fully eliminated. Brock, Dechert, and Scheinkman (1987) show that the BDS statistic has a limiting standard normal distribution under the null hypothesis of i.i.d.. In a finite sample the choice of imbedding dimension m and distance ε can have an effect on the BDS statistic. Choosing imbedding dimension too large compared with the sample size may yield the unreliable BDS statistic because there are too few nonoverlapping observations.

Table 3. Tsay and BDS Statistics

	m	UD	JY	DM	BP
BDS	2	0.8259	1.8863	-0.2127	0.2648
	3	1.3206	2.4594*	-0.1107	0.4046
	4	1.9707*	2.6676*	0.0808	0.6997
	5	2.4766*	3.0423*	0.2804	1.0151
	6	3.3584*	2.9718*	0.5603	1.5205
Tsay		6.3371**	16.6657**	1.3544	0.8004

* The null hypothesis of i.i.d. is rejected at 5% significance level.
* The null hypothesis of linear independence is rejected at 5% significance level.

Thus, according to our data the proper choice for m is from 2 to 6. On the other hand, if ε is too small or large, the BDS statistic can also be ill-behaved. Brock, Hsieh, and LeBaron (1991) suggest the best choice of ε is between 0.5 and 1.5 times the standard deviation. In this paper the BDS statistics are calculated for ε = 0.5 times the standard deviation. In the case of the JY only four extreme observations affect the standard deviation greatly so that they are excluded for calculation of the standard deviation. Interestingly but not surprisingly, for the DM and the BP there is little evidence of serial correlation of the AR residuals but the UD and the JY exhibit stronger serial correlation as m increases. This means the UD and the JY have nonlinear dependence significantly. Now to conclude, the strong evidence of serial dependence in the UD and the JY series allows time series models to predict the future changes of the exchange rates, but we can expect the apparent limitation of standard linear models due in part to the presence of considerable nonlinearities. Moreover, it would be hard to predict the future changes of the DM and the BP series, which have little serial dependence. As has been mentioned before, this unpredictability can partially arise from randomness of the exchange rates.

One of main interest in this section is to apply the MMEN proposed in this paper for predicting the future changes of the exchange rates and evaluate its prediction results in comparison with those of the ARMA and the TDNN models. For in-sample estimation 1037 observations are used and for one-step-ahead predictions the last 100 observations are reserved. The method consists of the two stage procedures. In the first stage, the original data are filtered by the general time domain filter described in section 3. The filter coefficients, b_i, a_i, are created from 5th Butterworth function. It is expected that this filtering operation will reduce extreme noise effectively and produce data with better quality, which can be used as inputs to the neural networks. In next step, the filtered data are standardized to lie between -1 and 1. The neural network has single hidden layer with 10 different hidden units ($L=1, 2, \cdots, 10$). A tansigmoid function and a linear function are defined as activation functions, h and g, respectively For the network inputs, 3 different AR processes ($p=1, 2, 3$) and 3 different MA processes ($q=1, 2, 3$) are chosen. Hence, 90 neural networks are applied to predictions. The initial values for parameters are random numbers following normal distributions.

The TDNN has same design except network inputs. For the TDNN inputs are time delayed vector ($x_t, x_{t-1}, \cdots, x_{t-9}$). In addition, we experience 8 different ARMA(p, q) models ($p=0, 1, 2$, $q=0, 1, 2$ except $p=0$ and $q=0$). To decide an appropriate ARMA model we consider Schwarz information criterion (SIC), $SIC(k)=\log(\hat{\sigma}^2_i)+K\log(T)/T$, where k is the number of parameters. For all series ARMA(0, 1) is the best model based on the SIC and it is applied to one-step-ahead predictions.[7]

To assess the quality of predictions we use a general measure, normalized mean squared error,

$$NMSE = \sum_{t=1}^{T} \frac{(observation_t - prediction_t)^2}{(observationt_t - mean)^2} = \frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^{T}(y_t-\hat{y}_t)^2$$

where *mean* and $\hat{\sigma}^2$ are average and variance of the target values. If the average of the data is simply used as predictor, a value of $NMSE=1$ is obtained.

7) For the UD, JY, DM, and BP, the values of SIC of ARMA(0,1) are -4.7434, -0.0521, -0.5274, and -0.4927, respectively.

Table 4. One-step-ahead Prediction Results : NMSE

UD										
	MMEN (p, q)									TDNN
L	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	
1	0.721	0.677	0.672	0.657	0.676	0.652	0.222	0.213	0.206	0.934
2	0.589	0.219	0.732	0.656	0.654	0.435	0.235	0.201	0.193	0.938
3	0.667	0.611	0.625	0.594	0.328	0.174	0.199	0.189	0.310	0.938
4	0.757	0.220	0.220	0.699	0.200	0.600	0.561	0.241	0.580	0.935
5	0.647	0.791	0.829	0.702	0.244	0.636	0.184	0.620	0.550	0.928
6	0.735	0.719	0.730	0.689	0.720	0.440	0.739	0.226	0.653	0.913
7	0.733	0.257	0.606	0.763	0.692	0.625	0.282	0.241	0.576	0.940
8	0.736	0.223	0.669	0.332	0.225	0.388	0.192	0.238	0.288	0.938
9	0.243	0.269	0.770	0.190	0.273	0.615	0.256	0.217	0.211	0.934
10	0.731	0.238	0.686	0.239	0.310	0.660	0.478	0.662	0.334	0.934
ARMA (0.1)										0.920

JY										
	MMEN (p, q)									TDNN
L	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	
1	0.800	1.017	1.019	0.918	1.052	1.035	1.134	1.112	1.082	0.997
2	0.857	0.985	0.993	0.885	1.013	0.549	1.159	1.067	1.224	0.991
3	0.787	0.987	0.919	0.886	1.029	1.064	1.024	1.032	1.298	0.968
4	0.819	1.035	1.020	0.909	0.982	1.166	0.973	1.097	0.888	0.952
5	0.818	0.902	0.919	0.916	1.046	1.019	1.264	0.964	1.312	0.995
6	0.869	1.024	0.832	0.764	1.154	1.423	0.845	1.026	1.225	1.065
7	0.791	0.978	0.955	0.838	1.221	0.903	1.456	1.022	0.944	0.970
8	0.844	1.142	1.164	0.868	1.345	1.064	1.094	0.748	1.145	1.072
9	0.812	1.010	1.021	0.853	0.908	0.904	1.105	1.119	0.969	1.041
10	0.883	0.984	0.930	0.845	1.046	1.100	0.695	1.087	0.701	0.995
ARMA (0.1)										0.984

DM										
	MMEN (p, q)									TDNN
L	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	
1	0.740	0.974	0.976	0.902	0.992	0.997	1.015	1.015	1.006	0.980
2	0.678	0.977	0.975	0.870	1.002	0.936	0.863	1.033	1.059	0.976
3	0.731	0.998	0.969	0.740	1.019	1.050	0.783	0.973	0.920	0.982
4	0.703	0.914	0.959	0.735	1.016	0.896	0.862	1.164	0.970	0.976
5	0.663	0.956	1.028	0.803	0.859	0.865	0.708	0.625	0.937	0.979
6	0.746	1.037	0.766	0.781	0.969	1.168	0.888	1.170	1.017	0.977
7	0.677	1.048	0.964	0.874	0.974	1.047	0.835	0.968	0.940	0.978
8	0.573	0.897	1.015	0.855	0.641	0.857	0.910	1.274	0.878	0.976
9	0.582	0.998	0.821	0.815	1.336	1.209	0.841	0.690	0.716	0.977
10	0.714	0.896	0.929	0.610	1.131	0.600	1.059	1.003	1.037	0.982
ARMA (0.1)										0.996

Table 5. Continued

BP										
	MMEN (p, q)									TDNN
L	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)	
1	0.743	0.961	0.970	0.940	0.963	0.987	0.859	1.141	1.055	1.046
2	0.724	0.987	0.970	0.756	0.983	0.879	0.941	0.688	1.097	1.590
3	0.701	0.912	0.972	0.888	0.988	0.942	0.976	0.820	0.841	1.061
4	0.694	0.917	0.925	0.908	0.999	1.025	1.084	1.284	1.393	1.060
5	0.669	0.958	0.916	0.691	1.110	0.844	1.056	1.135	1.055	1.015
6	0.826	0.913	0.925	0.961	0.963	0.831	0.874	1.060	1.075	1.055
7	0.783	0.950	0.903	0.840	1.111	0.854	0.870	0.939	0.959	1.053
8	0.740	1.016	0.827	0.786	0.645	0.825	0.797	0.625	1.122	1.062
9	0.751	1.040	0.896	1.055	1.232	1.198	1.061	1.024	1.219	1.056
10	0.761	1.009	1.062	0.931	1.137	1.250	1.211	0.490	0.839	1.062
ARMA (0.1)										1.003

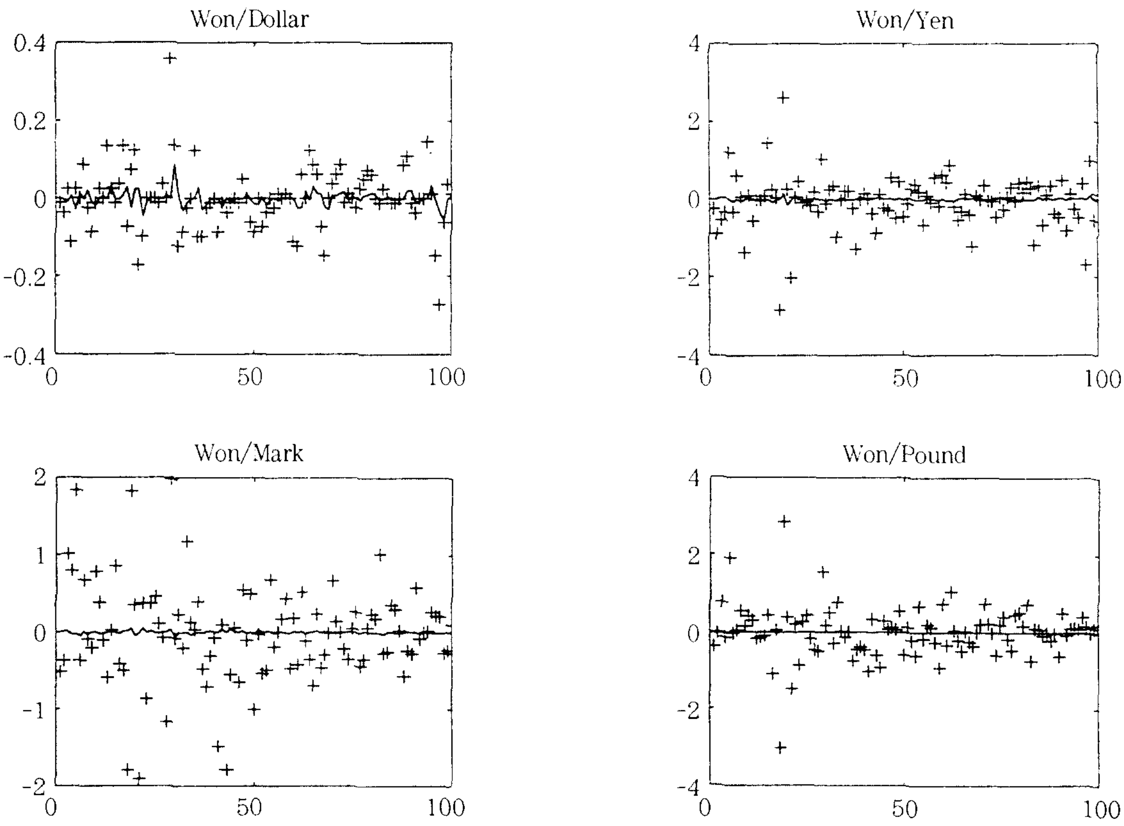

Figure. 4. One-step-ahead Predictions of linear ARMA

Table 4 sets out the results of the NMSE estimates. According to the results, for all exchange rates predictions of the ARMA model do not significantly dominate those of average of data. In the worst case, the NMSE of the ARMA model exceeds 1 for the BP. Figure 4 also supports the collapse of the ARMA model in predictions. Meanwhile, it appears that the TDNN slightly improve upon the standard ARMA results with respect to the NMSE and during learning process they suffer from very slow convergence on target values. By contrast, as one would expect, the MMEN reveals quite dramatic advance on the predictive power of the TDNN and the ARMA models. In view of figures 5 and 6, it can be shown that, confirming our expectations, the new NN approach has very satisfactory performance of one-step-ahead predictions. For the all series, it turns out that the MMEN has the lowest value of NMSE, 0.173(p=2, q=2, L=3), 0.549(p=2, q=3, L=2), 0.573(p=1, q=1, L=8), and 0.490 (p=3, q=2, L=10), respectively. The estimates, however, may still be unstable except the UD because they have huge variances.

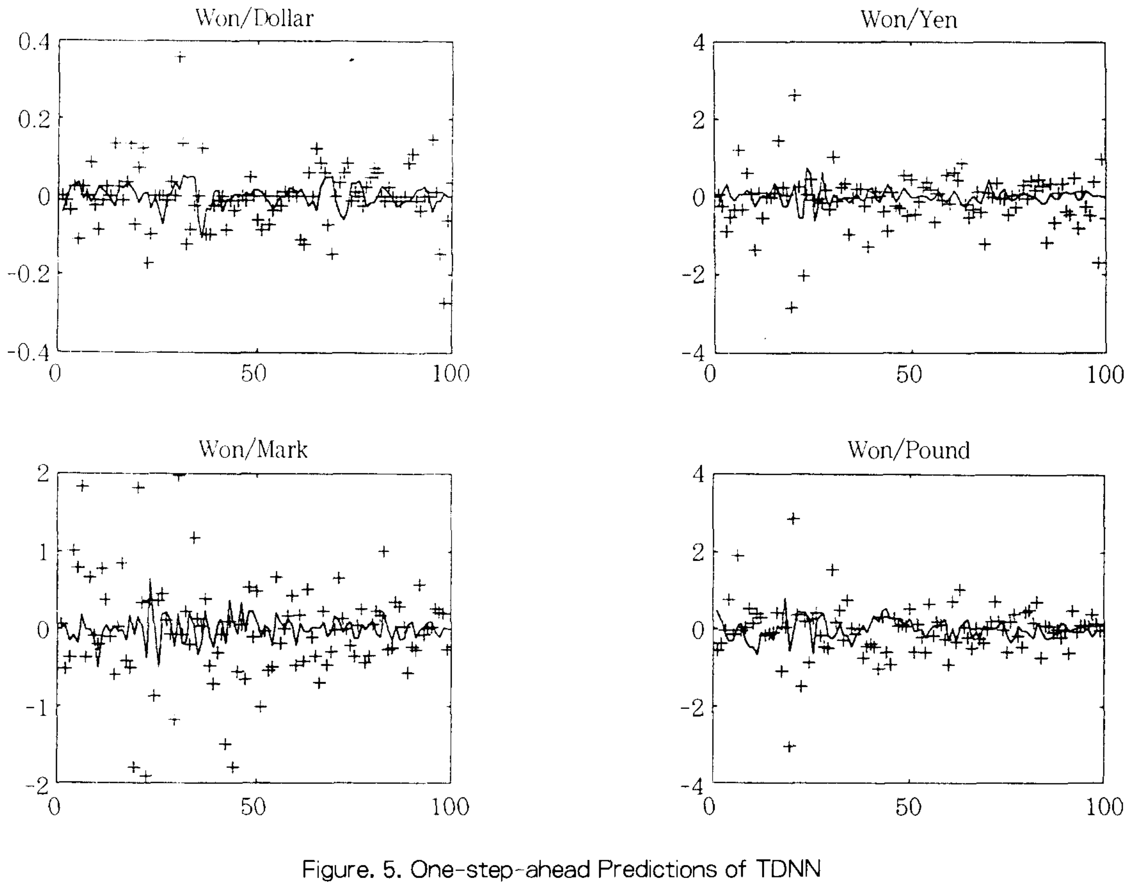

Figure. 5. One-step-ahead Predictions of TDNN

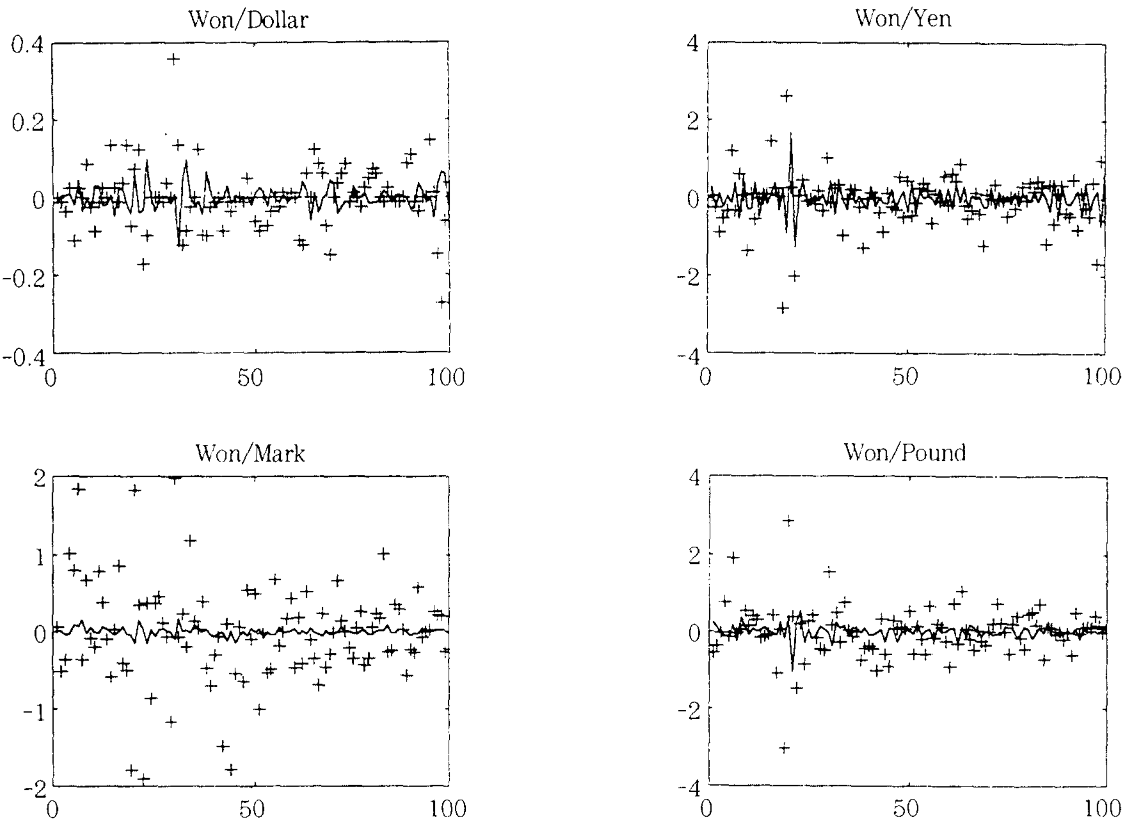

Figure. 6. One-step-ahead Predictionys of MMEN

Ⅵ. CONCLUDING REMARKS

Strong nonlinearity in changes of exchange rates is detected by several tests for the UD and the JY. Despite the presence of the nonlinearity standard TDNN cannot significantly improve predictions. The results are quite consistent with the idea of this paper that standard nonlinear models cannot lead to any dramatic advance on the predictive power of linear models. This paper therefore proposes a method of modified Elman network that can take advantages of both time domain filter and feedback neural networks. Especially, the recurrent networks are designed to take account of unconsidered information in the prediction model during its learning process. Our empirical works provide that the MMEN, as one would expect, has remarkably superior predictions compared to the TDNN and the ARMA models for all exchange rate series. Consequently, we can draw a conclusion that the MMEN is particularly useful for economic and financial fore casts.

Instead of completely forecasting future exchange rates from present ones, a more realistic goal is to build the best model for the exchange rate data. However, there is no formal theory for de termining optimal networks. Further, Elman-type networks are apt not to be trained optimally by a standard algorithm due in part to the dependence of the network parameters. These limitations call for further studies. First, to specify the optimal modified Elman network in application to ex- tremely dynamic time series data, one need to develop a method of choosing appropriate network structure including the number of hidden layer and units. Second, genetic algorithm may be use- ful for ensuring the convergence of the modified Elman network on the global minimum.

Finally, for future research state-space theory can be combined with neural network models in stead of the neural network ARMA models because state-space embedding can preserve an un- derlying geometrical structure and it may remarkably improve predictions for sufficiently low-di- mensional dynamics.

REFERENCES

Abu-Mostafa, Y., "Financial Market Applications of Learning from Hints," Neural Networks in the Capital Markets, edited by A. Refenes, John Wiley & Sons (1995), 221-232.

Barron, A., "Universal Approximation Bounds for Superpositions of a Sigmoidal function," Techni- cal Report No. 58, Department of Statistics, University of Illinois at U-C (1991).

Baillie, R. and T. Bollerslev, "The Message in Daily Exchange Rates : A Conditional Variance Tale," *Journal of Business and Economic Statistics* (1989), 7, 297-305.

Blanchard, O. and M. Watson, Bubbles, Rational Expectations and Financial Markets, ed. Wachtel, P. : Crises in the Economic and Financial Structures, Lexington Books, Lexington, MA (1982).

Brock, W., A. Hsieh, and B. LeBaron, Nonlinear Dynamics, Chaos, and Instability : Statistical Theory and Economic Evidence, The MIT Press (1991).

Brock, W., Dechert, W., and J. Scheinkman, "A Test for Independence Based on the Correlation Dimension," University of Wisconsin, Madison, University of Houston, and University of Chicago (1987).

Chinn, M. D., "Some Linear and Nonlinear Thoughts on Exchange Rates," *Journal of International Money and Finance* (1991), 10, 214-230.

Diebold, F. X., Empirical Modeling of Exchange Rate Dynamics, Berlin, Springer-Verlag (1988).

Diebold, F. and J. Nason, "Nonparametric Exchange Rate Prediction?" *Journal of International Economics* (1990), 28, 315-332.

Elman, J., "Finding Structure in Time", *Cognitive Science* (1990), 14, 179-211.

Engle, R., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* (1982), 50, 987-1007.

Funahashi, K.I., "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks* (1973), 2, 183.

Hornik, K., M. Stinchcombe and H. White, "Multi-layer Feedforward Networks Are Universal Approximators, *Neural Networks* (1989), 2, 359-366.

Hsieh, D. A., "The Statistical Properties of Daily Foreign Exchange Rates: 1974-1983," *Journal of International Economics* (1988), 129-145.

_______, "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *Journal of Business* (1989), 62, 339-368

Hsu, W., L. Hsu and F. Tenorio, "A Neural Network Procedure for Selecting Predictive Indicators in Currency Trading," Neural Networks in the Capital Markets, edited by A. Refenes, John Wiley & Sons (1995), 245-257.

Krugman, P., "Target Zone and Exchange Rate Dynamics," NBER Working Paper No 2481 (1988).

Kuan, C.-M. and K. Hornik, "Learning in a Partially Hard-wired Recurrent Network," BEBR Working Paper, 91-0114, University of Illinois, Urbana-Champaign (1991).

Kuan, C.-M. and T. Liu, "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks," BEBR Working Paper, 92-0128, University of Illinois, Urbana-Champaign (1992).

Kuan, C.-M. and H. White, "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews* (1994), 13(1), 1-91.

Meese, R and A. Rose, "Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation," *American Economic Review* (1991), v80, No2, 192-196.

Park, B.-J., "Robust Forecasts of Exchange Rates Using Neural Network Regression Quantiles" (1997).

Parks, T. and C. Burrus, Digital Filter Design, John Wiley and Sons, (1987).

Rumelhart, D., G. Hinton, and R. Williams, "Learning Internal Representations by Error Propagation," In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume I, edited by Rumelhart and J. McClelland (1986), 318-362.

Rumelhart, D. and J. McClelland, Explorations in the Microstructures of Cognition, Cambridge, MA, MIT Press (1986).

Tsay, R. S., "Nonlinearity Tests for Time Series," *Biometrika* (1986), 73(2), 461-466.

Tsibouris. G., "Essays on Nonlinear Models of Foreign Exchange," *Ph.D. Dissertation*, University of Wisconsin-Madison (1993).

Weigend, A., B. Huberman, and D. Rumelhart, "Predicting the Future: A Connectionist Approach," *Intl. J. Neur. Sys.* (1990), 1, 193-209.

_______, "Predicting Sunspots and Exchange Rates with Connectionist Networks," Nonlinear Modeling and Forecasting, edited by M. Casdagli and S. Eubank, Addison-Wesley (1992), 395-432.

White, H., "Economic Prediction Using Neural Networks: The Case of IBM Stock Prices," Proceedings of the IEEE Second International Conference on Neural Networks, II (1988), 451-458.

Wold, H., A Study in the Analysis of Stationary Time Series, Stockholm, Almqvist and Wiksell (1938).

Zhang, X. and J. Hutchinson "Simple Architectures on Fast Machines: Practical Issues in Nonlinear Time Series Prediction," Time Series Prediction: Forecasting the Future and Understanding the Past, edited by A. Weigend and N. Gershenfeld, Addison-Wesley (1993), 219-241.