Foreign Exchange Rates Forecasting with Neural Networks

Jingtao Yao, Hean-Lee Poh, Teo Jašić
Department of Information Systems and Computer Science
National University of Singapore
Singapore 119260
E-mail : yaojt@iscs.nus.sg

***Abstract*— In this paper, a neural network based foreign exchange rates forecasting method is discussed. Neural networks with time series and technical indicators as inputs are built to capture the underlying "rules" of the movement in currency exchange rates. Before using historical data to train the neural networks, the traditional R/S analysis is used to test the "efficiency" of each market. The study shows that without the use of extensive market data or knowledge, useful prediction can be made and significant paper profit can be achieved with simple technical indicators.**

1 Introduction

Since 1973, with the abandonment of the fixed foreign exchange rates and the implementation of the floating exchange rate system by industrialized countries, researchers have been striving for an explanation of the movement of exchange rates. Thus, many kinds of forecasting methods are developed by thousands of researches and experts. Technical and fundamental analysis are among the major forecasting methods which are popularly used in the financial area. Foreign exchange rates are affected by many highly correlated factors. These factors could be economic, political and even psychological factors. The interaction of these factors is in a very complex fashion. Therefore, to forecast the change of foreign exchange rates is generally very difficult. Neural networks are an emerging and challenging computational technology and they offer a new avenue to explore the dynamics of a variety of financial applications. Actually, they are simulated networks with interconnected 'neurons' which try to mimic the function of the brain's central nervous system. Neural networks have been shown to have great potential for financial forecasting. Examples using neural networks in currency applications include Refenes[3], Weigend[4], and Zhang[8]. Feed-forward backpropagation networks are the most commonly used networks and meant for the widest variety of applications.

In this paper, the research results on using neural networks to forecast the exchange rates between the US dollar and five other major currencies, Japanese Yen (JPY), Deutsch Mark (DEM), British Pound(GBP), Swiss Franc(CHF) and Australian Dollar(AUD) are presented. This study shows that without the use of extensive market data or knowledge, useful prediction can be made and significant paper profit can be achieved with simple technical indicators.

2 Foreign Exchange Rate Forecasting

From the very beginning, Forex was determined by the balance of payments. The balance of payments was merely a way of listing receipts and payments in international transactions for a country. The balance was determined mainly by the import and export of goods. Therefor it was not difficult to predict Forex at that time. Later on, interest rates and other demand-supply factors had become more relevant to each currency. Increased Forex trading, and hence speculation due to liquidity and bonds, had also contributed to the difficulty of forecasting Forex.

The application of forecasting method includes two basic steps: analyze data series and select the forecasting method that best fits the data series. To maximize profits from the liquidity market, more and more 'best' forecasting techniques are used by traders. Nowadays, traders no longer rely on a single technique to provide information about the future of the markets but rather use a variety of techniques to obtain multiple signals. Neural networks are often trained by using both technical and fundamental indicators to produce trading signals. To improve the profit gain and to decrease risk is the most important motivations for developing the neural networks. Neural networks can make contributions to the maximization of returns, while reducing costs, and limiting risks. In this paper, a mixed technical method which takes not only the delayed time series data as inputs but also the technical indicators is illustrated. The work discussed in this paper would represent a violation of the *efficient market hypothesis*. The inclusion of fundamental factors will be studied in a different paper.

Variable	Mean	Std. Dev.	Variance	Max(*)
AUD/USD	0.7424	0.0605	0.003642	2.6616
CHF/USD	1.6203	0.3828	0.1466	3.3503
DEM/USD	1.9450	0.4770	0.2263	3.081
GBP/USD	1.6025	0.1978	0.03938	2.7220
JPY/USD	152.240	42.2177	1782.4335	2.6187

Table 1: Data Statistics of Weekly Foreign Exchange Rates; * Maximum value for normalized data - zero mean and unit variance

3 Data Set Construction and Efficient Testing

Historical data are divided into three portions: training, validation and testing sets. The training set contains two thirds of the collected data, while the validation and the testing sets contain two fifteenths and three fifteenths respectively. The division is based on the experience of the authors which can be considered as a rule of thumb. A model is considered good if the error of out-of-sample testing is the lowest compared with the other models. If the trained model is the best one for validation and also the best one for testing, one can assume that it is a good model for future forecasting. The data sets of five currencies studied in this paper comprise 2910 daily rates for a sampling period of between 18 May 1984 and 7 July 1995. The data are chosen and segregated in time order. In other words, the data of the earlier period are used for training, the data of the later period are used for validation, and the data of the latest time period are used for testing. The statistics summary of the weekly data used in this study are shown in Table 1.

In this paper, the weekly closing prices are used as the prediction target of our experiment. They refer to each Friday's closing prices in the Singapore market. In the event of Friday being a holiday, the most recently available closing price for the currency was used. The data set for each currency in this study consist of 510 weekly data. It is segregated as training set: 18 May 1984 to 12 July 1991, validation set: 19 Nov 1991 - 29 Oct 1993 and testing set: 5 Nov 1993 - 7 July 1995.

In a real situation, there is no closing price for Forex. Forex trading takes place 24 hours a day over the world. The 24 hour data should be used in order to capture the underlying rules of the movement in Forex rates. The more data you use the more rules you can get. In this research, the weekly data are used assuming that they have enough information to capture the "rules". Due to the volatility of the currency movement, a different frequency of data maybe needed than the weekly data. The data could be sampled according to the market character, e.g. bullish, bearish, or trading, etc. In other words, when the market is volatile, we sample more data for training, and vice versa. Nonlinear or volatility time scale[7] will be taken into consideration in our further research.

3.1 Testing the Efficient Market Hypothesis

The most famous, widely tested and little believed hypothesis are *Random Walk Hypothesis* and *Efficient Market Hypothesis*[5]. The *Random Walk hypothesis* states that the market prices wander in a purely random and unpredictable way. The *efficient market hypothesis* states that the markets fully reflect all of the available information and prices are adjusted fully and immediately once new information become available. In the actual market, some people do react to information immediately after they have received the information while other people wait for the confirmation of information. The waiting people do not react until a trend is clearly established.

H. E. Hurst, who was a hydrologist, found that most natural phenomena, including river discharge, temperatures, rainfall, and sunspots, follow a *biased random walk* which is a trend with noise. The Hurst Exponent H [1] is a measure of the bias in fractional Brownian motion. The method could be used in economic and financial market time series to see whether these series are also biased random walks which indicates the possibility of forecasting.

The rescaled range analysis (R/S analysis) [1] is able to distinguish a random series from a fractal series, irrespective of the distribution of the underlying series (Gaussian or non-Gaussian). It can be used to detect the long-memory effect in the foreign exchange rate time series over a time period. R captures the maximum and minimum cumulative deviations of the observations x_t of the time series from its mean

Exchange	Hurst Exponent	Correlation
AUD/USD	0.532681	0.046347
CHF/USD	0.553941	0.077645
DEM/USD	0.554672	0.078737
GBP/USD	0.544408	0.063497
JPY/USD	0.540706	0.058053

Table 2: Hurst exponent and Correlation for the experimented five currencies

(μ), and it is a function of time (the number of observations N):

$$R_N = \max_{1 \leq t \leq N}[x_{t,N}] - \min_{1 \leq t \leq N}[x_{t,N}] \tag{1}$$

where $x_{t,N}$ is the cumulative deviation over N periods. The R/S ratio of R and the standard deviation S of the original time series can be estimated by the following empirical law: $R/S = N^H$ when observed for various N values. H describes the probability that two consecutive events are likely to occur. The type of series described by $H = 0$ is random, consisting of uncorrelated events. A value of H different from 0.50 denotes the observations that are not independent. When $0 < H < 0.5$, the system is an antipersistent or ergodic series with frequent reversals and high volatility. For the case $(0.5 < H < 1.0)$, H describes a persistent or trend-reinforcing series which is characterized by long memory effects. However, even in the case that the Hurst process describes a biased random walk, the bias can change abruptly either in direction or magnitude. Therefore, only the average cycle length of observed data can be estimated.

As shown in Table 2 , The value of Hurst Exponent for the logarithmic returns of daily exchange rates data is higher than 0.5 for all the observed time series. The highest value is 0.554 for the exchange rate of CHF/USD which denotes a long-memory effect in time series. Hence, there exist possibilities for conducting time series forecasting in the studied data sets.

4 Resulst of Weekly Exchange Rates Forecasting

Time series forecasting is perhaps the most exciting application of neural networks. The objective is to discover the underlying "structure" of the mechanism generating the data, i.e., to discover the relationship between present, past and future observations. In this paper, a purely time delayed time series and a simple technical indicators based time series method are experimented.

4.1 Measurement of Neural Networks

A usual measure to evaluate and compare the predictive power of the model is the Normalized Mean Squared Error (NMSE)[2][6]. Additional evaluation measures include the calculation of correct matching number of the actual and predicted values, x_t and $\hat{x}_t$ respectively, in the testing set with respect to the sign and directional change (expressed in percentages). Directional change statistic is the average of a_k where $a_k = 1$ if $(x_{t+1} - x_t)(\hat{x}_{t+1} - x_t) > 0$, and $a_k = 0$ otherwise. These statistics are desirable because the NMSE measure prediction only in terms of levels. Hence, the quality of the forecast can be measured by the correctness of gradient predictions (D_{stat}) or simply by the return one can expect if one starts with either the USD or the currency in consideration.

To simulate the real profit, a paper profit is used in this study. Assume that a certain amount of seed money is used in this program. The seed money is used to buy a certain amount of another currency when the prediction shows a rise in that currency. At the end the of testing period, the currency should be converted to the original currency of the seed money using the exact direct or cross rate of that day. The results obtained are shown in Table 3 and Table 4. The paper profit is calculated as follows:

$$Return = \left(\frac{money2}{money1}\right)^{\frac{52}{nw}} - 1 \tag{2}$$

where money1 = Seed money on first the testing day; money2 = Money after trading on last the testing day; nw= No. of weeks in testing period.

4.2 Forecasts Using Purely Time Delayed Time Series and Indicators

The purely time delayed forecast method is one of the simplest technical analysis methods. The real targets of the previous periods are used as inputs to the neural network to forecast the next period

Exchange	Model	Test. NMSE (R^2)	Gradient	$Ret1_{US}$
AUD/USD	5-3-1	0.0543 (0.9456)	55.00 %	1.09 %
CHF/USD	5-3-1	0.1100 (0.8900)	56.00 %	8.40%
DEM/USD	6-3-1	0.3153 (0.6847)	51.00 %	4.36 %
GBP/USD	6-3-1	0.1555 (0.8445)	54.74 %	2.30 %
JPY/USD	5-3-1	0.1146 (0.8853)	53.40 %	3.00 %

Table 3: The Testing Results for Neural Network Models (Delay Method) for Weekly Foreign Exchange Data

Exchange	Model	Test. NME	Gradient	$Ret1_{US}$	$Ret1$	$Ret2_{US}$	$Ret2$
AUD/USD	5-3-1	0.035105	73.86 %	8.82%	12.19%	12.43 %	15.90 %
AUD/USD	6-4-1	0.032362	76.14 %	8.97%	12.34%	12.67 %	16.16 %
CHF/USD	5-3-1	0.068819	65.91 %	28.49%	9.99%	22.49 %	4.85 %
CHF/USD	6-4-1	0.065962	64.77 %	32.36%	13.31%	21.64 %	4.15 %
DEM/USD	5-3-1	0.063462	61.36 %	22.86%	8.86%	15.20 %	2.07 %
DEM/USD	6-4-1	0.061730	64.77 %	27.84%	13.27%	18.00 %	4.55 %
GBP/USD	5-3-1	0.061370	73.86 %	7.22%	2.87%	14.78 %	10.13 %
GBP/USD	6-4-1	0.053650	72.73 %	10.62%	6.13%	16.48 %	11.76 %
JPY/USD	5-4-1	1.966195	46.59 %	19.71%	3.47%	0.00 %	-13.57 %
JPY/USD	6-4-1	1.242099	46.59 %	23.42%	6.67%	0.00 %	-13.57 %

Table 4: The Testing Results for Neural Network Models using indicators for Weekly Foreign Exchange Data. $Ret1$: Return using the *Stategy 1*; $Ret2$: Return using the *Stategy 1*; Ret_{US} denotes the seed money is in USD.

exchange rate. In our experiment, five to eight weeks of time delayed data are used. Some of the measurements of the forecasting results are shown in Table 3.

This method sometimes leads to prediction that seems to generate a time-delayed time series of the original time series. With the inclusion of some popular indicators used by traders, it might help to remove some of the time delay characteristics of the prediction.

The father of Dow-Jones, Charles Dow, divided the trend into three different levels, namely the primary trend, the secondary trend, and the minor trend. The advantage of moving average is that it tends to smooth out some of their irregularities that exist between market days. Moving averages are used as inputs to the neural network. MA5, MA10, MA20, MA60, and MA120 are used as the inputs to neural networks. They refer to moving averages for one week, two weeks, one month, one quarter and half a year respectively. One of the diagrams showing the predicted (out) and the actual (tar) time series for the period of Nov 1993 - July 1995 (out of sample) is shown in Figures 1 AUD/USD.

Further, the forecasts for each of the currencies were repeated, but with a hybrid of indicators and one time delay term. The configuration of the neural network is 6-4-1. For example, for Australian Dollar, the results showed that the hit rate of 6-4-1 was slightly higher than that of the pure indicator forecasting method with a configuration of 5-3-1. The hit rate of the former is 76.14% and that of the latter is

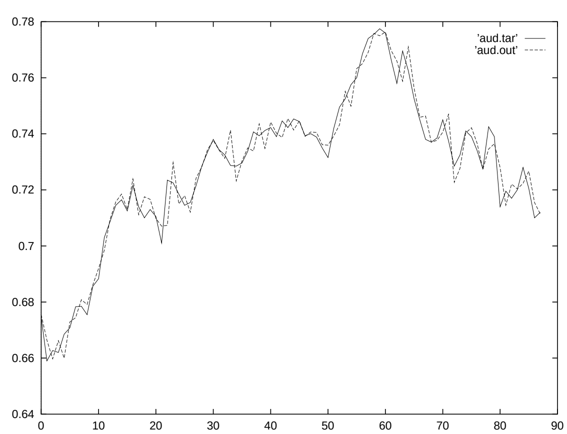

Figure 1: Prediction of the Weekly AUD/USD Nov 1993 - July 1995 (Indicators Method)

73.86%. In general, the addition of an additional term of time delay does not contribute much to the improvement of the hit rate. So, we can safely conclude that hit rates of approximately 70% can be achieved consistently for AUD, GBP, and somewhat lower for CHF and DEM.

Also, from the graphs, one can conclude that the forecasts for the first twenty weeks of the pure testing period look "impressive". This means that, the neural network needs to be retrained, probably every twenty weeks (or half a year) with the latest data to increase the chance of achieving a better forecast. Notice that, in actual application, only validation data sets would be required together with the training data sets.

4.3 Limitations

The data are chosen and segregated in time order in the experiment, This method may have some *recency problems*. Neural networks were only trained using data up till the end of October 1993. In forecasting the Forex after November 1993, the neural network is 'forced' to use knowledge up till 1993 only.

A very small NMSE does not necessarily imply good generalization. The sum of the NMSE of the three parts of data (training, validation and testing) must be kept small, not just the training NMSE alone. Sometimes having small NMSEs for testing and validation is more important than having small NMSE for training.

After experimenting with the choice of data, a very good testing result may not predict well. On the other hand, a model which is trained with randomly chosen data may predict well even with average testing results. Further, better testing results are demonstrated in the period near the end of the training sets. This is a result of the 'recency' problem.

4.4 Trading Strategies

Trading is an art. As there is no perfect forecasting technique, trading profit is ensured only by a good trading strategy taking "full" advantage of a good forecasting method. There are two kinds of trading strategies used in this study. One uses the difference between predictions, and another uses the difference between the predicted and the actual levels to trade.
Strategy 1:

$$\text{if}(\hat{x}_{t+1} - \hat{x}_t) > 0 \ \text{ then } \ \textit{buy} \ \text{ else } \ \textit{sell} \tag{3}$$

Strategy 2:

$$\text{if}(\hat{x}_{t+1} - x_t) > 0 \ \text{ then } \ \textit{buy} \ \text{ else } \ \textit{sell} \tag{4}$$

In actual trading, practitioners may choose one of the strategies. A conservative trading strategy would require a trader to act only when both strategies recommend the same actions.

In this paper, 1% of transaction cost was included in the calculation. The transaction cost of a big fund trading, and thus affecting the market rates was not taken into consideration To be more realistic, a specific amount of transaction cost has to be included in the calculation. If the output of neural network is given in percentage of changes, we can use positive or negative output to show that the currency is going up or down.

5 Comparison with ARIMA

The Box-Jenkins methodology, or *Autoregressive Integrated Moving Average*(ARIMA) Model, provides a systematic procedure for the analysis of time series that was sufficiently general to handle virtually all empirically observed time series data patterns. To compare the forecasting results of the neural networks, a number of ARIMA models were built. Table 5 is the results of ARIMA models with different trading strategies. The entire data set was used as fitting data for the ARIMA models. In other words, the data forecast by ARIMA were already used in the fitting stage of ARIMA model building. hence, the ARIMA models should deliver worse out-of-sample forecasting returns than the ARIMA results indicated in Table 5. Focusing on the gradients, the ARIMA methods can achive about 50% of correctness while up to 73% of correctness can be achieved using neural network models.

From practitioners' point of view, returns are more important than gradient. With reference to Table 5 and 4, the differences between ARIMA models and neural network models are significant. The best return using *Strategy 1* regardless the devaluation and strategies for ARIMA models is only 6.94%, while for neural network models is 28.49%.

Model	Gradient	$Ret1_{US}$	$Ret2_{US}$
AUD101	52.27 %	1.43 %	1.36 %
AUD202	54.32 %	1.53 %	1.21 %
CHF101	38.64 %	6.94 %	-1.42 %
CHF202	55.86 %	5.43 %	0.64 %
DEM101	43.18 %	3.48 %	3.49 %
DEM202	44.62 %	3.48 %	3.22 %
GBP101	53.41 %	2.24 %	3.67 %
GBP202	51.77 %	2.63 %	1.32 %
JPY101	44.32 %	-1.47 %	-0.52 %
JPY202	44.32 %	-0.78 %	0.02 %

Table 5: The Result of Using ARIMA (AUD101 stands for the ARIMA result of AUD using ARIMA(1,0,1) model and the same rules are applied to other currencies)

6 Conclusion and Further Research

In using neural networks to perform technical forecasting, better results are obtained for Australian Dollar, Swiss Franc, and British Pound and perhaps Swiss Francs and Deutsch Mark. The results for Japanese Yen are the worst in terms of using *strategy 2*. The reason could be that the market for Yen is bigger and more efficient than the market for other currencies. So the traders of the Yen market may depend more on technical analysis and they may act quickly after the signs appear. Hence, technical analysis may not be a good tool for forecasting the trends of Yen. This is similar to our forecasting result using ARIMA models.

The hit rate may be a better standard for determining the quality of the forecast. After all, the return depends a lot on the trading strategies and how the forecasting information are being used for trading advantages. However, the level of hit rates and paper profits also depend on the period of forecast. Hence, constant upgrading of the neural networks is necessary. For the practitioners, the levels of exchange rate and trend can be used depending on their expectation of return and risk. In addition to the above-mentioned two strategies, even more different trading strategies may be used by them. To benefit more practitioners, risk based trading strategies will be taken into consideration in future research. The behavior of each markets will also be studied.

7 Acknowledgments

We thank Mary Aviani, Swee-Yuan Tay and Yili Li for invaluable discussions and helpful comments.

References

[1] Hurst, H. E., "Long Term Storage of Reservoirs", *Transactions of the American Society of Civil Engineers* , 116, 1951

[2] Levin, R. I., *Statistics for management* Prentice Hall, 1994

[3] Refenes, A.N., "Managing exchange rate prediction strategies with neural networks" *Techniques and Applications of Neural Networks*, Liverpool, UK, Sept. 1992

[4] Weigend, A.S. "Generalization by weight-elimination applied to currency exchange rate prediction" *IEEE International Joint Conference on Neural Networks*, Singapore, Nov. 1991

[5] Peters, E. E., *Chaos and Order in the Capital markets: A New View of Cycles, Prices, and Market Volatility*, John Wiley & Sons Inc. 1991

[6] Yao, Jingtao, Poh, H.L., "Equity Forecasting: a Case Study on the KLSE Index" *NNCM'95(3rd International Conference On Neural Networks in the Capital Markets)*, 1995, London.

[7] Zhou, B., "Estimating the Variance Parameter From Noisy High Frequency Financial Data," *MIT Sloan School Working Paper*, No. 3739, 1995.

[8] Zhang, Xiru, "Non-linear predictive models for intra-day foreign exchange trading", *International Journal of Intelligent Systems in Accounting, Finance and Management*, Dec. 1994