Combining Artificial Intelligence with Non-linear Data Processing Techniques for Forecasting Exchange Rate Time Series

[1]Heng-Li Yang , [2]Han-Chou Lin
[1, First Author]*Department of Management Information systems, National Chengchi University, Taipei, Taiwan, yanh@nccu.edu.tw*
[*2,Corresponding Author]*Department of Management Information systems, National Chengchi University, Taipei, Taiwan, actuallife@hotmail.com*

Abstract

Combing back-propagation neural network (BPNN) and empirical mode decomposition (EMD) techniques, this study proposes EMD-BPNN model for forecasting. In the first stage, the original exchange rate series were first decomposed into a finite, and often small, number of intrinsic mode functions (IMFs). In the second stage, kernel predictors such as BPNN were constructed for forecasting. Compared with traditional model (random walk), the proposed model performs best. This study significantly reduced errors not only in the derivation performance, but also in the direction performance.

Keywords: *Back-propagation neural network (BPNN), Hilbert–Huang transform (HHT), Empirical mode decomposition (EMD), Intrinsic mode function (IMF).*

1. Introduction

Policy makers watch foreign exchange rate market carefully because exchange rate is a financial asset and thus is potentially valuable source of timely information about economic and financial conditions. Therefore, by understanding the movement of exchange rate better, the policy makers will be able to extract the relevant information about the economic and financial conditions of the economy. It will enable them to design a better monetary policy for the future which will in turn achieve its desired objective of price stability and greater employment. Therefore, policy makers might wish to forecast exchange rates. Similarly, firms or investors might wish to forecast exchange rates to make asset allocation decisions.

Owing to the high risk associated with the international transactions, exchange rate forecasting is one of the challenging and important fields in modern time series analysis. The determinants of exchange rate have grown manifold making its behavior complex, nonlinear and volatile. Indeed, some researchers also believe that modeling the behavior or the prediction of exchange rate is not feasible. They claim that the best prediction value for tomorrow’s exchange rate is the current value of the exchange rate and the actual exchange rate follows a random walk.

Meese and Rogoff(1983) indicate that no single economic model of exchange rates is better in predicting bilateral exchange rates during floating exchange rates than the simple random walk model. However, their results are not surprising and can easily be explained by the fact that all the models investigated in their work are linear, whereas it is widely agreed that exchange rate movements are nonlinear (Hsieh 1989; De Grauwe, Dewachter et al. 1993; Brooks 1996; Drunat, Dufrenot et al. 1996) during floating exchange rates. Hence, exchange rate data contain nonlinearities that cannot be fully accounted for or approximated well by linear models.

Recently, one amongst nonlinear models, artificial neural networks (ANNs) have been increasingly and successfully employed for modeling time series. Unlike conventional statistical models, ANNs are data-driven, nonparametric, weak models which let ‘‘the data speak for themselves’’. In past decades, ANNs have been explored by many researchers for financial forecasting (Hill, O’Connor et al. 1996; Zhang and Hu 1998; Yao and Tan 2000; Zimmermann, Neuneier et al. 2001; Kamruzzaman and Sarker 2003; Wang 2005; Erkam, Gulgun et al. 2011; K. and K. 2011; Ling-Feng, Su-Chen et al. 2011). Using neural networks to model and predict financial market has been the subject of recent empirical and theoretical investigations by academics and practitioners alike. However, ANNs suffer from several weaknesses, such as the need for a large number of controlling parameters, difficulty in obtaining a stable solution and the danger of over-fitting.

To remedy the above shortcomings, some hybrid methods have been used and obtain the best performances. For example, Wang, Yu et al. (2004) developed a hybrid AI system framework by means of a systematic integration of ANN and rule-based expert system, with web text mining. Amin-Naseri and Gharacheh (2007) proposed a hybrid AI approach integrating feed-forward neural networks, genetic algorithm, and k-means clustering and obtain satisfactory results. The basic idea of the above hybrid methods is to overcome the drawbacks of individual models and to generate a synergetic effect in forecasting.

In terms of the above ideas, an empirical mode decomposition (EMD) based neural network learning is proposed for exchange rate forecasting. EMD based on Hilbert–Huang Transform (HHT), a new technique in dealing with noise and nonlinearity data, will be applied to decompose exchange rate series into a finite and often small number of intrinsic mode functions (IMF). The main advantage of selecting EMD as a decomposition tool is very suitable for decomposing nonlinear and nonstationary time series; it has been reported to have worked better, in describing the local time scale instantaneous frequencies, than the wavelet decomposition and Fourier decomposition (Huang, Shen et al. 1999; Li 2006).

Thus, the major of this study is to integrate EMD and one of artificial neural network models, back-propagation neural network (EMD-BPNN) to attempts to increase the accuracy for the prediction of exchange rates. The effectiveness of the methodology was verified by experiments comparing the random walk model for daily Taiwan/US dollar (NTD/USD) exchange rate. The results show that the proposed EMD-BPNN model provides better prediction of exchange rates and provide the prompt information.

2. Methodology

2.1. Empirical mode decomposition (EMD)

Empirical mode decomposition method is developed from the simple assumption that any signal consists of different simple intrinsic mode oscillations. The essence of the method is to identify the intrinsic oscillatory modes (IMFs) by their characteristic time scales in the signal and then decompose the signal accordingly. The characteristics time scale is defined by the time lapse between the successive extremes.

To extract the IMF from a given data set, the sifting process is implemented as follows. First, identify all the local extrema, and then connect all of the local maxima by a cubic spline line as the upper envelope. Then, repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelope. The upper and lower envelopes should cover all the data between them. Their mean is designated $m_1(t)$, and the difference between the data and $m_1(t)$ is $h_1(t)$, i.e.:

$$x(t) - m_1(t) = h_1(t) \tag{1}$$

Ideally, $h_1(t)$ should be an IMF, for the construction of $h_1(t)$ described above should have forced the result to satisfy all the definitions of and IMF, we demand the following conditions: (i) $h_1(t)$ should be free of riding waves i.e. the first component should nod display under-shots or over-shots riding on the data and producing local extremes without zero crossing. (ii) To display symmetry of the upper and lower envelops with respect to zero. (iii) Obviously the number of zero crossing and extremes should be the same in both functions.

The sifting process has to be repeated as many times as it is required to reduce the extracted signal to an IMF. In the subsequent sifting process steps, $h_1(t)$ is treated as the data; then:

$$h_1(t) - m_{11}(t) = h_{11}(t) \tag{2}$$

Where $m_{11}(t)$ is the mean of the upper and lower envelops of $h_1(t)$. This process can be repeated up to k times; $h_{1k}(t)$ is then given by

$$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t) \tag{3}$$

After each processing step, checking must be done on whether the number of zero crossings equals the number of extrema. The resulting time series is the first IMF, and then it is designated as $c_1(t) = h_{1k}(t)$. The first IMF component from the data contains the highest oscillation frequencies found in the original data $x(t)$. This first IMF is subtracted from the original data, and this difference, is called a residue $r_1(t)$ by:

$$x(t) - c_1(t) = r_1(t) \tag{4}$$

The residue $r_1(t)$ is taken as if it was the original data and we apply to it again the sifting process. The process of finding more intrinsic modes $c_i(t)$ continues until the last mode is found. The final residue will be a constant or a monotonic function; in this last case it will be the general trend of the data.

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t) \tag{5}$$

Thus, one achieves a decomposition of the data into n-empirical IMF modes, plus a residue, $r_n(t)$, which can be either the mean trend or a constant. (Huang, Shen et al. 1998) have defined IMFs as a class of functions that satisfy two conditions:

In the whole data set, the number of extrema and the number of zero-crossings must be either equal or differ at most by one;

At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

2.2. Artificial Neural Network

An ANN is a biologically inspired form of distributed computation. It simulates the functions of the biological nervous system by a composition of interconnected simple elements (artificial neurons) operating in parallel. An element is a simple structure that performs three basic functions: input, processing and output. ANNs can be organized into several different connection topologies and learning algorithms (Lippmann 1987). The number of inputs to the network is constrained by the problem type, whereas the number of neurons in the output layer is constrained by the number of outputs required by the problem type. Moreover, the number of hidden layers and the sizes of the layers are decided by the designer.

ANNs apply many learning rules, of which back-propagation neural network (BPNN) is one of the most commonly used algorithms in financial research. We therefore use the BP in this paper. BPNN trains multilayer feed-forward networks with differentiable transfer functions to perform function approximation, pattern association and pattern classification. It is the process by which the derivatives of network error, with respect to network weights and biases, are computed to perform computations backwards through the network. Computations are derived using the chain rule of calculus. There are several different BPNN training algorithms with a variety of different computation and storage requirements. No single algorithm is best suited to all the problems. All the algorithms use the gradient of the performance function to determine how to adjust the weights to minimize the performance. For example, the performance function of feed forward networks is the Mean Square Error.

The basic BPNN algorithm adjusts the weights in the steepest descent direction (negative of the gradient); that is, the direction in which the performance function decreases most rapidly. The training process requires a set of examples of proper network inputs and target outputs. During the training, the weights and biases of the network are iteratively adjusted to minimize the network performance function

2.3. Overall process of the EMD-based BPNN

Suppose there is a time series $x(t)$, t=1, 2,…, N, in which one would like to make the l-step ahead prediction, i.e. $x(t+l)$. For example, l=1 means one single-step ahead prediction and l=30 represents 30-step ahead prediction. Depending on the previous techniques and methods, an EMD-based back-propagation neural network ensemble paradigm can be formulated, as illustrated in Figure. 1.

As can be seen from Fig. 1, the proposed EMD-based BPNN model is generally composed of the following two steps:

(1) The original time series $x(t)$, $t = 1,2,...,N$ is decomposed into n IMF components, $c_j(t)$ $j = 1,2,...,n$, and one residual component $r_n(t)$ via EMD.

(2) For each extracted IMF component and the residual component, the BPNN model is used as a forecasting tool to model the decomposed components, and generates the final prediction result for the original time series.

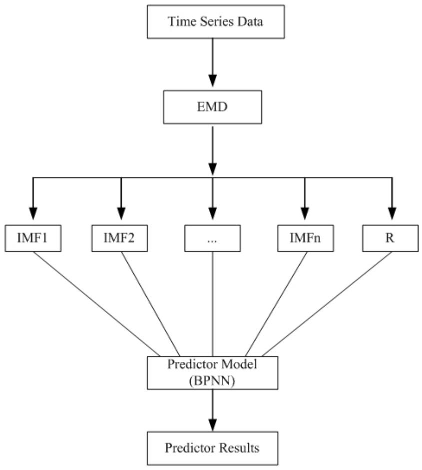

Figure 1. Procedure of the proposed EMD-BPNN predictor model

3. Experimental results and analysis

3.1. Research data and evaluation criteria

Daily values of exchange rates Taiwan/Dollar (NTD/USD) extracted from Datastream provided by Morgan Stanley Capital International (MSCI) were used in this study. The data set is for the period of January, 2001–December, 2010, for a total 2352 observations. In addition, the data are divided into training and testing sets. Many studies have applied a convenient ratio to separate training (in-sample) from testing (out-of-sample) data between the ratios 7 : 3 and 9 : 1(Zhang 2004). This study follows the choice: the data from Jan. 2001 to Jan. 2010 are used for training, while the data from Feb. 2010 to Dec. 2010 are used for testing. The ratio adopted in this study is 9 : 1, which is a ratio that lies in-between.

Each data point was scaled by Eq. (6) within the range of (0, 1). This scaling for original data points helps to improve the forecasting accuracy (Chang and Lin 2001):

$$\frac{X_t - X_{\min}}{X_{\max} - X_{\min}} 0.7 + 0.15 \qquad (6)$$

where X_t is the exchange rate at time t, $X_{\max}$ is the the maximum of exchange rates during the period of data source and $X_{\min}$ is the minimum of exchange rates during the period of data source.

To measure the forecasting performance, five criteria are used for evaluation of level prediction forecasting. We select the following statistics: MSE (mean-squared error)、RMSE (root-mean-square error)、MAPE (mean-absolute percentage error)、MAE (mean-absolute error) and DS (Directional Symmetry)(Tay and Cao 2001; Huang and Wu 2008; T., C. et al. 2008). MSE、RMSE、MAE、MAPE and DS are calculated by using Eqs, showed in Table 1. The former four criteria measure the correctness of a prediction in terms of levels and the deviation between the actual and predicted values. The smaller the values, the closer the predicted time-series values will be to the actual values. Although predicting the levels of price changes (or first differences) is desirable, in many cases the sign of the change is equally important. Most investment analysts are usually far more accurate at predicting directional changes in an asset price than predicting the actual level. DS provides an indication of the correctness of the predicted direction given in the form of percentages (a large value suggests a better predictor).

Table 1. Performance criteria and formulas

MSE = $\frac{1}{N}\sum\left\lvert x_t - \hat{x}_t\right\rvert^2$	RMSE = $\sqrt{\frac{1}{N}\sum\left\lvert x_t - \hat{x}_t\right\rvert^2}$	MAE = $\frac{1}{N}\sum_{t=1}^{N}\left\lvert x_t - \hat{x}_t\right\rvert$
MAPE = $\frac{\sum\left\lvert\frac{x_t - \hat{x}_t}{x_t}\right\rvert}{N}\times 100$	DS = $\frac{100}{N}\sum_{t=1}^{N} d_t$, $d_t = \begin{cases} 1 & (x_t - x_{t-1})(\hat{x}_t - \hat{x}_{t-1}) \geq 0 \\ 0 & otherwise \end{cases}$	

where x_t denotes the real stock index on the "t"th day, $\hat{x}_t$ represents the predicted stock index, and N is the number of days.

In order to compare the forecasting capability of the proposed EMD-based BPNN (EMD-BPNN) methodology with traditional approaches, a random walk model is used as the benchmark model.

3.2. Data preprocessing using EMD

According to previous steps shown in Section 2.3, we start to perform the prediction experiments. First, using the EMD technique, the exchange rate series can be decomposed into ten independent IMFs and one residue. Using EMD, we can get graphical representations of the decomposed results for the exchange rate, as illustrated in Figure 2.

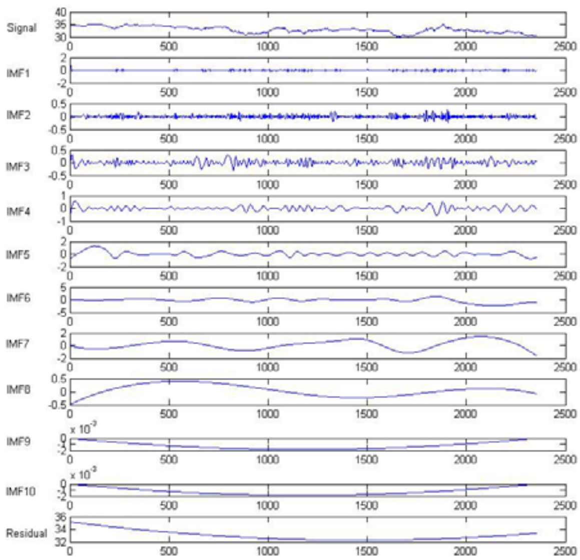

Figure 2. The decomposition of exchange rate (NTD/USD) by EMD

3.3. BPNN implementation

The BPNN technique in this experiment is implemented using the Matlab 7.3's ANN toolbox. Several learning techniques, such as the quasi-Newton method, Levenberg-Marquardt algorithm and conjugate gradient methods could also be used. For efficiency, however, we use the Levenberg–Marquardt algorithm. This study follows (Kim 2003) to varies the number of nodes in the hidden layer and stopping criteria for training. 5, 11, 22 hidden nodes and 50, 100, 200 epochs are employed for each stopping criteria. The learning rate is 0.001(default). The activation function of the hidden layer is sigmoid and the output node uses the linear transfer function. This study allows 11 input nodes because 11 input variables (i.e. numbers of IMF and one residual). Table 2 show the results based on different hidden nodes and learning epochs. 5 hidden nodes and 100 learning epochs are optimal, and the testing MAPE is about 0.15.

Table 2. EMD-BPNN model selection result

Hidden nodes	Epochs	Training MAPE(%)	Testing MAPE(%)
5	50	0.14	0.16
	100	0.14	0.15
	200	0.13	0.26
11	50	0.13	0.32
	100	0.13	0.3
	200	0.13	6.93
22	50	0.12	0.73
	100	0.12	0.49
	200	0.12	2.98

3.4. Results

Using the above settings, the evaluation of prediction results for the exchange rate series is shown in Table 3 via five criteria, representing the relative percentage error of the two forecasting models, including traditional forecasting model (Random Walk) for the testing and training data, and optimal parameters of each models. In general, the proposed EMD-BPNN model outperformed the random walk model in derivation performance criteria. The values of testing MSE, RMSE, MAE, MAPE and DS for EMD-BPNN are 0.0038, 0.061, 0.0465, 0.1470 and 0.647, respectively. The corresponding values for random walk are 0.00624, 0.0079, 0.0591, 0.1874 and 0.5106. Thus, it can be concluded that the proposed model provides a better forecasting result than the Random Walk model in terms of prediction error and prediction accuracy.

Table 3. Comparison of the forecasting results from the two models. (By testing data)

Model	Optimal parameters	MSE	RMSE	MAE	MAPE(%)	DS
EMD-BPNN	Epochs = 50 hidden node = 100	0.0038	0.061	0.0465	0.1470	0.647
Random Walk	None	0.00624	0.079	0.0591	0.1874	0.5106

4. Conclusions

This study applied empirical mode decomposition (EMD) based BPNN to the time series of exchange rate and forecasted exchange rate, namely EMD-BPNN model. We have examined the feasibility of the proposed model compared to a traditional model, random walk model. In this paper, the daily exchange rate (NTD/USD) extracted from Morgan Stanley Capital International (MSCI) (Jan. 2001 – Dec. 2010) were used as experimental data.

We evaluated performance of another model by not only in the derivation performance, but also in the direction performance. It was clear from the empirical results, the proposed model EMD-BPNN gives better accuracy in the both performances not only in trainings data, but also in testing data. Moreover, the experimental results also suggested that within the forecasting fields of exchange rates, the EMD-BPNN was typically a reliable forecasting tool, with the forecasting capacity more precise than that of the random walk model. Therefore, the use of the proposed model may help policy makers in extracting useful information about the economic and financial conditions.

Future research should apply other AI methods and data analytic techniques to the forecasting of exchange rates. In our current models, we only select decomposition IMFs of one-dimensional exchange rate series as input variables. Enhancing the performance of prediction models by including other efficient input variables (such as some macroeconomic variables) and trying to use diverse data for feasibility, the relationships between different markets (or countries) and other information about the markets should also be the subject of future research.

5. References

[1] Amin-Naseri, M. R. and E. A. Gharacheh, "A hybrid artificial intelligence approach to monthly forecasting of crude oil price time series". The Proceedings of the 10th International Conference on Engineering Applications of Neural Networks, CEUR-WS284, pp. 160-167, 2007.
[2] Brooks, C., "Testing for non-linearity in daily sterling exchange rates". Applied Finance Economics vol. 6, pp. 307-317, 1996.
[3] Chang, C. C. and C. J. Lin, "LIBSVM: A library for support vector machines. Department of Computer Science and Information Engineering, National Taiwan University". http://www.csie.ntu.edu.tw/cjlin/papers/libsvm.pdf , 2001.
[4] De Grauwe, P., H. Dewachter, et al., "Exchange rate theory: Chaotic models of foreign exchange markets", Blackwell, 1993.
[5] Drunat, J., G. Dufrenot, et al. (1996). Stochastic or chaotic dynamics in high frequency exchange rates? J. Wiley & Sons.
[6] Erkam, G., K. Gulgun, et al., "Using artificial neural network models in stock market index prediction ". EXPERT SYSTEMS WITH APPLICATIONS vol. 38 no. 8, pp. 10389-10397, 2011.
[7] Hill, T., M. O'Connor, et al., "Neural network models for time series forecasts". Management Science vol. 42, pp. 1082-1092, 1996.
[8] Hsieh, D. A., "Testing for nonlinear dependence in daily foreign exchange rates". Journal of Business vol. 62, pp. 329-368, 1989.
[9] Huang, N. E., Z. Shen, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". Proceeding of Royal Society London, Series A vol. 454, pp. 903-995, 1998.
[10] Huang, S.-C. and T.-K. Wu, "Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting ". Expert Systems with Applications, vol. 35, no. 4, pp. 2080-2088, 2008.
[11] Huang, W., Z. Shen, et al., "Nonlinear Indicial Response of Complex Nonstationary Oscillations as Pulmonary Hypertension Responding to Step Hypoxia". Proc of the National Academy of Sciences, vol. 96, no., pp. 1934-1839, 1999.
[12] K., D. A. and B. V. K., "Exchange rate forecasting: comparison of various architectures of neural networks ". NEURAL COMPUTING & APPLICATIONS vol. 20 no. 3, pp. 355-363, 2011.
[13] Kamruzzaman, J. and R. Sarker, "Forecasting of currency exchange rate using ANN: A case study", Proceedings of the IEEE International Conference on Neural Networks & Signal Processing (ICNNSP'03),, Nanjing, China, 2003.
[14] Kim, K. j., "Financial time series forecasting using support vector machines". Neurocomputing,

vol. 55, pp. 307-319, 2003.
[15] Li, X., "Temporal structure of neuronal population oscillations with empirical mode decomposition", Physics Letters A, vol. 356, no. 237-241, 2006.
[16] Ling-Feng, H., H. Su-Chen, et al., "Enhanced stock price variation prediction via DOE and BPNN-based optimization ". EXPERT SYSTEMS WITH APPLICATIONS vol. 38 no. 11, pp. 14178-14184, 2011.
[17] Lippmann, R. P., "An introduction to computing with neural nets", IEEE ASSP Magazine, pp. 36-54, 1987.
[18] Meese, R. A. and K. Rogoff, "Empirical exchange rate models of the seventies: Do they fit out of sample?" Journal of international Economics, vol. 14, pp. 3-24, 1983.
[19] T., C.-H., S.-T. C., et al., "Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices ", Physica A: Statistical Mechanics and its Applications, vol. 387, no. 13, pp. 3192-3200, 2008.
[20] Tay, F. E. H. and L. Cao, "Application of support vector machines in financial time series forecasting", The International Journal of Management Science, vol. 29, pp. 309-317, 2001.
[21] Wang, L., Support vector machines: Theory and applications, New York: Springer, 2005.
[22] Wang, S. Y., L. Yu, et al., "A novel hybrid AI system framework for crude oil price forecasting". Lecture Notes in Computer Science, vol. 3327, pp. 233-242, 2004.
[23] Yao, J. and C. L. Tan, "A case study on using neural networks to perform technical forecasting of forex", Neurocomputing, vol. 34, pp. 79-98, 2000.
[24] Zhang, G. and M. Y. Hu, "Neural network forecasting of the British Pound/US Dollar exchange rate", International Journal of Management Science, vol. 26, no. 4, pp. 495-506, 1998.
[25] Zhang, G. P., "In G. P. Zhang & Pa. Hershey (Eds.), Business forecasting with artificial neural networks: an overview, neural networks in business forecasting", Idea Group Publishing, pp. 1-22, 2004.
[26] Zimmermann, H., R. Neuneier, et al., "Multi-agent modeling of multiple FX-markets by neural networks", IEEE Transactions on Neural Networks, vol. 12, no. 4, pp. 735-743, 2001.