International Journal of Information Technology & Decision Making
Vol. 3, No. 1 (2004) 145–165

FORECASTING FOREIGN EXCHANGE RATES WITH ARTIFICIAL NEURAL NETWORKS: A REVIEW

WEI HUANG

Institute of Systems Science, Academy of Mathematics and Systems Sciences Chinese Academy of Sciences, Beijing 100080, People's Republic of China

School of Knowledge Science, Japan Advanced Institute of Science and Technology 1-1, Asahidai, Ishikawa 923-1292, Japan

K. K. LAI

Department of Management Sciences, City University of Hong Kong Tat Chee Avenue, Kowloon, Hong Kong

Y. NAKAMORI

School of Knowledge Science, Japan Advanced Institute of Science and Technology 1-1, Asahidai, Ishikawa 923-1292, Japan

SHOUYANG WANG*

Institute of Systems Science, Academy of Mathematics and Systems Sciences Chinese Academy of Sciences, Beijing 100080, People's Republic of China
Tel: 86-10-62651381
Fax: 86-10-62568364
sywang@amss.ac.cn

Forecasting exchange rates is an important financial problem that is receiving increasing attention especially because of its difficulty and practical applications. Artificial neural networks (ANNs) have been widely used as a promising alternative approach for a forecasting task because of several distinguished features. Research efforts on ANNs for forecasting exchange rates are considerable. In this paper, we attempt to provide a survey of research in this area. Several design factors significantly impact the accuracy of neural network forecasts. These factors include the selection of input variables, preparing data, and network architecture. There is no consensus about the factors. In different cases, various decisions have their own effectiveness. We also describe the integration of ANNs with other methods and report the comparison between performances of ANNs and those of other forecasting methods, and finding mixed results. Finally, the future research directions in this area are discussed.

Keywords: Artificial neural networks; exchange rate; forecasting.

*Corresponding author. This author is also with School of Business, Hunan University

1. Introduction

The foreign exchange market is the largest and most liquid of the financial markets, with an estimated $1 trillion traded every day. Exchange rates are amongst the most important economic indices in the international monetary markets. For large multinational firms, which conduct substantial currency transfers in the course of business, being able to accurately forecast exchange rate movements can result in substantial improvement in the firm's overall profitability.

Exchange rates are affected by many highly correlated economic, political and even psychological factors. These factors interact in a very complex fashion. Exchange rate series exhibit high volatility, complexity and noise that result from an elusive market mechanism generating daily observations.[49] Evidence has clearly shown that while there is little linear dependence, the null hypothesis of independence can be strongly rejected, demonstrating the existence of non-linearities in exchange rates.[11]

Much research effort has been devoted to exploring the nonlinearity of exchange rate data and to developing specific nonlinear models to improve exchange rate forecasting. Parametric nonlinear models such as the autoregressive random variance (ARV) model,[44] autoregressive conditional heteroscedasticity (ARCH),[16] general autoregressive conditional heteroskedasticity (GARCH),[1] chaotic dynamic[31] and self-exciting threshold autoregressive[4] models have been proposed and applied to foreign exchange rate forecasting. While these models may be good for a particular situation, they perform poorly for other applications. The pre-specification of the model form restricts the usefulness of these parametric nonlinear models since many other possible nonlinear patterns can be considered. One particular nonlinear specification will not be general enough to capture all the nonlinearities in the data. Some nonparametric methods have also been proposed to forecast exchange rates.[7,28,29] However, nonparametric methods investigated in these studies are still unable to improve upon a simple random walk model in out-of-sample predictions of exchange rates.

There has been growing interest in the adoption of the state-of-the-art artificial intelligence technologies to solve the problem. One stream of these advanced techniques focuses on the use of artificial neural networks (ANNs) to analyze the historical data and provide predictions on future movements in the foreign exchange market. An ANN is a system loosely modeled on the human brain, which detect the underlying functional relationships within a set of data and perform tasks such as pattern recognition, classification, evaluation, modeling, prediction and control. ANNs are particularly well suited to finding accurate solutions in an environment characterized by complex, noisy, irrelevant or partial information. Several distinguishing features of ANNs make them valuable and attractive in forecasting. First, as opposed to the traditional model-based methods, ANNs are data-driven self-adaptive methods in that there are few *a priori* assumptions about the models for problems under study. Second, ANNs can generalize. Third, ANNs are universal functional approximators. Finally, ANNs are nonlinear.[59]

The idea of using ANNs for forecasting exchange rates is not new. Weigend *et al.*[53] find that neural networks are better than random walk models in predicting the DEM/USD exchange rate. Refense *et al.*[37] apply a multi-layer perceptron network to predict the exchange rate between USD/DEM, and to study the convergence issue related to network architecture. Refense[36] develops a constructive learning algorithm to find the best neural network configuration in forecasting DEM/USD. Podding[33] studies the problem of predicting the trend of the USD/DEM, and compares results to those obtained through regression analysis. Pi[32] proposes a test for dependence among exchange rates. Shin[41] applies an ANN model and moving average trading rules to investigate return predictability of exchange rates. Zhang and Hutchinson[62] report the experience of forecasting the tick-by-tick CHF/USD. Kuan and Liu[24] use both feed-forward and recurrent neural networks to forecast GBP, CAD, DEM, JPY, CHF against USD. Wu[55] compares neural networks with ARIMA models in forecasting Taiwan/USD exchange rates. Hann and Steurer[15] mark comparisons between the neural network and linear model in USD/DEM forecasting. Episcopos and Davis[10] investigate the problem of predicting daily returns based on five Canadian exchange rates using ANNs and a heteroskedastic model, EGARCH. Tenti[48] proposes the use of recurrent neural networks in order to forecast exchange rates. Other earlier examples using ANN in exchange rates application include Zhang[61] and Yao *et al.*[56]

Considerable research effort has gone into ANNs for forecasting exchange rates. In this paper, we attempt to provide a survey of research in this area. Forecasting exchange rates using ANNs is a process that can be divided into several steps. Our goal in this paper is to find out consensus and disagreements in each step. Hence, the comparisons of various methods used by different researchers go along the whole forecasting process. For the consensus areas, guidelines are summarized. With the disagreements, we analyze the reasons and point out the advantages and disadvantages of various methods.

The paper is organized as follows. Section 2 covers input selection. Sec. 3 deals with preparing data. In Sec. 4, we give a brief presentation of ANN architecture. Section 5 describes the integration of ANNs with other methods. The comparison between performances of ANNs and those of other forecasting methods is reported in Sec. 6. Finally, conclusions and directions for future research are discussed in Sec. 7.

2. Input Selection

There are two kinds of inputs — fundamental inputs and technical inputs. Fundamental inputs include consumer price index, foreign reserve, GDP, export and import volume, interest rates, etc. Technical inputs include the delayed time series data, moving average, relative strength index, etc. Besides the above two kinds of inputs, individual forecast results could be used as inputs when using ANNs as combined forecasting tools. In order to provide improved volatility forecasts, Hu and Tsoukalas[17] combine GARCH, EGARCH, IGARCH and MAV volatility forecast through ANNs. A preliminary effort to maximize the output performance is conducted by ensuring adequate domain knowledge representation from input variable.[47,51]

While Walczak *et al.*[52] claim that multivariate inputs are necessary, most neural network inputs for exchange rate prediction are univariate. Univariate inputs utilize data directly from time series being forecast, while multivariate inputs utilize information from outside the time series in addition to the time series itself. Univariate inputs rely on the predictive capabilities of the time series itself, corresponding to a technical analysis as opposed to a fundamental analysis. For a univariate time series forecasting problem, the network inputs are the past, lagged observations of the data series and the output is the future value. Each input pattern is composed of a moving window of a fixed length along the series. In this sense, the feed-forward network used for time series forecasting is a general autoregressive model. The question is how many lag periods should be included in predicting the future. Some authors designed experiments to help selecting the number of input nodes while others adopted some intuitive of empirical ideas. Mixed results are often reported in the literatures. The lack of systematic approaches to neural network model building is probably the primary cause of inconsistencies in the reported findings.

Ideally, we desire a small number of lag periods that can unveil the unique features embedded in the data. The inclusion of excessive periods will adversely affect the training time of the network, and the algorithm will likely be trapped in local optimal solutions. On the other hand, if the lag is smaller than required, forecasting accuracy will be jeopardized because the search is restricted to a subspace. Too few or too many lag periods affect either the learning or prediction capability of the network. It is desirable to reduce the number of input nodes to an absolute minimum of essential nodes.

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as well as several extensions have been used as information-based in-sample model selection criteria in selecting neural networks for foreign exchange rate time series forecasting.[34] However, the in-sample model selection criteria are not able to provide a reliable guide to out-of-sample performance and there is no apparent connection between in-sample model fit and out-of-sample forecasting performance.

Huang *et al.*[18] propose a general approach called Autocorrelation Criterion (AC) to determine lag structures in the applications of ANNs to univariate time series forecasting. They apply the approach to the determination of input variables for foreign exchange rate forecasting and conduct comparisons between AC and information-based in sample model selection criterion. Experiment results show that AC outperforms information-based in sample model selection criterion in terms of forecasting performance.

We suggest practitioners to employ Autocorrelation Criterion in the case of univariate input. It does not require any assumptions, completely independent of particular class of model. The selection of input variables is data-driven, making full uses of information among sample observations even if the underlying relationships are unknown or hard to describe. Thus, it is well suited for time series problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations. It nevertheless provides a practical way to solve input selection for neural networks in time series forecasting.

Multivariate inputs are based on economics and finance theory. El Shazly *et al.*[8] use fundamental inputs including the one month Eurorate on US dollar deposit, the one month Eurorate on the foreign currency deposit, the spot exchange rate, and the one month forward premium on the foreign currency. El Shazly *et al.*[9] use inputs including the 90-day Euro deposit rate on the US dollar (INT US); the 90-day Euro deposit rate on the British pound (INTBP), German mark (INTDM), Japanese yen (INTJY), and the Swiss franc (INTSF); the spot exchange rate of the foreign currency: SBP, SDM, SJY, and SSF expressed in direct form; the 90-day forward exchange rate on the foreign currency: FDBP, FDDM, FDJY, FDSF; and the 90-day future exchange rate on the foreign currency: FTBP, FTDM, FTJY, FTSF. The above input variables selection comes from Interest Rate Parity, a principle by which forward exchange rates reflect relative interest rates on default risk-free instruments denominated in alternative currencies. Currencies of countries with high interest rates are expected by the market to depreciate over time, and currencies of countries with low interest rates are expected to appreciate over time.

Leung *et al.*[25] use the MUIP relationship[39] as the theoretical basis for multivariate specification. The MUIP relationship can be modeled and written as follows:

$$\begin{aligned} e_t = \alpha_0 &+ \alpha_1(r_t^* - r_t) + \alpha_2(\pi_t^* - \pi) + \alpha_3(p_t^* - p_t) + \alpha_4(ca_t/ny_t) \\ &+ \alpha_5(ca_t^*/ny_t^*) + \mu_t \end{aligned}$$

where e is the natural logarithm of the exchange rate, defined as the foreign currency price of domestic currency. r, π, p and (ca/ny) represent the logarithm of the nominal short-term interest rate, expected price inflation rate, the logarithm of the price level, and the ratio of current account to nominal GDP for the domestic economy respectively. Asterisks denote the corresponding foreign variables. μ is the error term. The variables (ca/ny) and (ca^*/ny^*) are proxies for the risk premium.

Tenti[48] uses inputs including the compound returns of the last n periods (where $n = 1, 2, 3, 5, 8$), the running standard deviation of the k last periods (where $k = 13, 21, 34$), and technical indicators such as the average directional movement index (ADX), trend movement index TMI), rate of change (ROC), and Ehlers leading indicator (ELI). Lisi and Schiavo[26] use both the past observation of the series itself and those of an "auxiliary" variable chosen among the remaining series. For example, the lagged FRF/USD, GBP/USD are used to predict future FRF/USD.

According to Walczak and Cerpa's[51] suggestions, multivariate inputs can be determined through the following steps. Firstly, perform standard knowledge acquisition. Get as more explanatory variables to foreign exchange rates as possible from economics and finance theory. The primary purpose of the knowledge acquisition phase is to guarantee that the input variable set is not under-specified, providing all relevant domain criteria to the ANNs. Once a base set of input variables defined through knowledge acquisition, the set can be pruned to eliminate variables that contribute noise to the ANNs and consequently reduce the ANNs generalization performance. Smith[43] claims that ANNs input variables need to be predictive, but should not be correlated. Correlated variables degrade ANNs performance by interacting with each other as well as other elements to produce a biased effect. A first pass filter to help identify noise variables is to calculate the correlation of pairs of variables (Pearson correlation matrix). Alternatively, a chi-square test may be used for categorical variables. If two variables have a high correlation, then one of these two variables may be removed from the set of variables without adversely affecting the ANNs performance. Additional statistical techniques may be applied, depending on the distribution properties of the data set. Stepwise regression (multiple or logistic) and factor analysis provide viable tools for evaluating the predictive value of input variables and may serve as a secondary filter to the Pearson correlation matrix.

Multivariate input has advantages in long term forecasting, unveiling the movement trend of foreign exchange rates. But it needs more data and time. Some explanatory variables are not available in time requirement. Univariate input has not such problems. The practitioners can benefit from a net reduction in the development costs, since less data is required. However, it lacks of economic explanations, which weakens forecasting credibility.

3. Preparing Data

Due to the fact that only relatively little preliminary knowledge is required to train artificial neural networks and on account of the black box character, data is often presented to the networks without any further processing steps being taken. However, the degree of care invested in preparing the data is of decisive importance to the networks learning speed and the quality of approximation it can attain. Every hour invested in preparing the data may save days in training the networks.

The first questions to be considered here are of a very general nature:

(1) Is sufficient data available, and does this data contain the correct information?
(2) Does the available data cover the range of the variables concerned as completely as possible?
(3) Are there borderline cases that are not covered by the data?
(4) Does the data contain irrelevant information?
(5) Are there transformations or combinations of variables (e.g. ratios) that describe the problem more effectively than the individual variables themselves?

Once all these points have been clarified, the data needs to be transformed into an appropriate form for the networks. Various normalization methods are generally employed to this end. Tenti[48] normalizes inputs to zero mean and two standard deviations. In Hu *et al.*'s[17] study, all inputs to ANNs are linearly normalized to [0,1]. El Shazly *et al.*[9] suggest that data should be manipulated and converted to the required format for further processing. In Lisi and Schiavo's[26] study, the log-differenced data are scaled linearly in the range of 0.2–0.8 in order to adapt them to the output range of the sigmoid activation function. Qi and Zhang[34] apply natural logarithm transformation to raw data to stabilize the series. An ADF test shows that the transformed time series contains a unit root, thus the first order difference is applied.

There is no consensus on whether data normalization should be used. For example, it is still unclear that whether there is a need to normalize the input because the arc weights can undo the scaling. Shanker *et al.*[40] investigate the effectiveness of linear and statistical normalization methods for classification problems. They find that data normalization methods do not necessarily lead to better performance particularly when the networks and sample size are large. El Shazly *et al.*[8] apply normalization and transformation to initial runs and then discard them. They find that although the difference data set speeds the training time by reducing the noise during training, the networks when tested yield poor forecasts. Based on the objective of improving the testing performance rather than speeding up training time, they decide to use raw data during training. Zhang and Hu[59] find no significant difference between using normalized and original data, based on their experience with the exchange rate data. Hence, raw data are used in that study.

Although normalization of the data is not compulsory, it is sometimes unavoidable. If for example, a function is valid only for a limited range, e.g. the sigmoid function (0.0–1.0) or the tanh function (-1.0–1.0), the network will be unable to generate any output values outside of this range. The target output data for the training and test phases must therefore be normalized.

In principle, it is not absolutely necessary to normalize the input data, as the networks input layer is assigned a linear function. However, it is nevertheless inadvisable not to normalize data when using multivariate inputs. As a result of normalization, all variables acquire the same significance for the learning process. If normalization is not carried out, variables with greater values will be given preference.

Generally, a data set is divided into two, the training set and the test set. The training set is used for ANNs model development and the test set is used to evaluate the forecasting ability. Sometimes a third set, called the validation set, is used to avoid the overfitting problem or to determine the stopping point in the training process.

There is no general solution to splitting the training set and test set. The Brainmaker software randomly selects 10% of the facts from the data set and uses them for testing. Yao *et al.*[57] suggest that historical data are divided into three portions: training, validation and testing sets. The training set contains 70% of the collected data, while the validation and the testing sets contain 20% and 10% respectively.

The division is based on a rule of thumb derived from the authors' experience. Other researches just give the division directly, not touching on the reasons for it.

Sample size is another factor that can affect artificial neural networks forecasting ability. Neural networks researchers have used various sizes of training sets, ranging from one year to sixteen years.[21,36,46,48,52,58] Large samples are often claimed to be optimal in training neural networks due to the large set of parameters involved in the network. To test if there is a significant difference between large and small training samples in modeling and forecasting exchange rates, Zhang and Hu[58] use two training sample sizes. The large sample consists of 887 observations from 1976 to 1992, and the small one includes 261 data points from 1998 to 1992. Their result is that the large sample outperforms the smaller sample. Most of the researchers typically use all of the data in building neural networks forecasting model once they have obtained their training data, with no attempt at comparing data quantity effects on the quality of the produced forecasting models.

However, Kang,[22] in a comprehensive study of neural network time series forecasting, finds that neural networks forecasting models do not necessarily require a large data set to perform well. Walczak[50] examines the effect of different sizes of training sample sets on forecasting exchange rates. His research results indicate that for financial time series, two years of training data is frequently all that is required to produce optimal forecasting accuracy. He claims that given an appropriate amount of historical knowledge, neural networks can forecast future exchange rates with 60% accuracy, while neural networks trained on a larger training set have a worse forecasting performance. In addition to high-quality forecasts, the reduced training set sizes reduce development cost and time.

Huang *et al.*[19] propose to determine the optimal quantity of training data by using change-point detection. The behavior of exchange rates is evolving over time. Therefore, we can conjecture that the movement of exchange rates has a series of change points, which divide data into several homogeneous groups that take heterogeneous characteristics from each other.

4. Architectures of ANNs

Three classes of ANNs architectures have been employed for forecasting foreign exchange rates. In this section, we give a brief presentation and conduct some comparisons.

4.1. *Feedforward*

In feedforward ANNs, the connections between units do not form cycles. Feedforward ANNs usually produce a response to an input quickly.

4.1.1. *Multi-layer perceptrons (MLP)*

MLP[38] is perhaps the most popular network architecture in use, which is relatively easy to implement. An MLP is typically composed of several layers of nodes

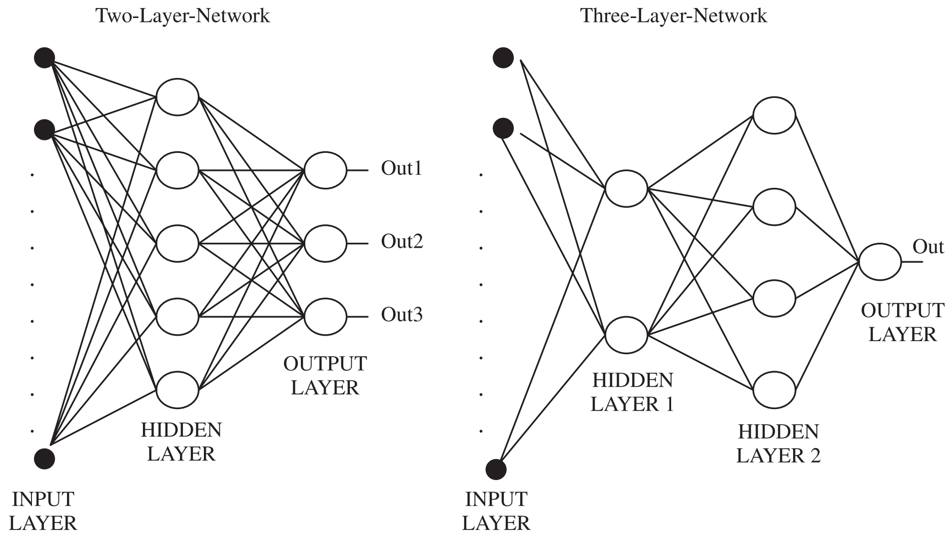

Fig. 1. Examples of multi-layer perceptron neural network architectures.

(see Fig. 1). The network thus has a simple interpretation as a form of input-output model, with the weights and thresholds (biases) the free parameters of the model. Although it has been shown theoretically that the MLP has a universal functional approximating capability and can approximate any nonlinear function with arbitrary accuracy, no universal guideline exists in choosing the appropriate model structure for practical applications. Thus, a trial-and-error approach or cross-validation experiment is often adopted to help find the best model. Typically a large number of neural network architectures are considered. The one with the best performance in the validation set is chosen as the winner, and the others are discarded.

4.1.2. *Radial basis function networks (RBFNs)*

RBFNs[30] have static Gaussian function as the nonlinearity for the hidden layer processing elements. The Gaussian function responds only to a small region of the input space where the Gaussian is centered. The key to successful implementation of these networks is to find suitable centers for the Gaussian functions. This can be done with supervised learning, but an unsupervised approach usually produces better results.

The advantage of the radial basis function network is that it finds the input to output map using local approximators. Usually the supervised segment is simply a linear combination of the approximators. Since linear combiners have few weights, these networks train extremely fast and require fewer training samples.

4.1.3. *Learning vector quantization (LVQ)*

LVQ[12,23] is a precursor of the well-known self-organizing maps (also called Kohonen feature maps) and like them it can be seen as a special kind of artificial neural network. A neural network for learning vector quantization consists of two layers: an input layer and an output layer. It represents a set of reference vectors, the coordinates of which are the weights of the connections leading from the input neurons to an output neuron. Hence, one may also say that each output neuron corresponds to one reference vector. This kind of ANNs architecture can only be used for classification. Hence, we cannot employ it in forecasting foreign exchange rates value.

4.1.4. *General regression neural networks (GRNNs)*

GRNNs[45] are memory-based feed-forward networks based on the estimation of probability density functions. GRNNs featuring fast training times, can model non-linear functions, and have been shown to perform well in noisy environments given enough data. The GRNN topology consists of four layers: the input layer, pattern layer, summation layer and output layer. Each layer of processing units is assigned a specific computational function when nonlinear regression is performed. The only adjustable parameter in a GRNN is the smoothing factor for the kernel function. The optimization of the smoothing factor is critical to the GRNN's performance and is usually found through iterative adjustments and the cross-validation procedure.

The advantages of GRNN include

(1) Fast training times.
(2) Can handle both linear and non-linear data.
(3) Adding new samples to the training set does not require re-calibrating the model.
(4) Only one adjustable parameter thereby making overtraining less likely.

The disadvantages include:

(1) Has trouble with irrelevant inputs (i.e. suffers from the dimensionality curse).
(2) No intuitive method for selecting the optimal smoothing parameter.
(3) Requires many training samples to adequately span the variation in the data.
(4) Requires that all the training samples be stored for future use (i.e. prediction).

4.2. ***Feedback***

In feedback ANNs, there are cycles in the connections. Each time an input is presented, ANNs must iterate for a potentially long time before it produces a response. Feedback ANNs are usually more difficult to train than feedforward ANNs.

4.2.1. *Recurrent neural networks (RNNs)*

Recurrent neural networks (RNNs), in which the input layer's activity patterns pass through the network more than once before generating a new output pattern, can learn extremely complex temporal patterns. Recurrent architecture has been proved to be superior to the windowing technique of overlapping snapshots of data, which is used with standard back-propagation. In fact, by introducing time-lagged model components, RNNs may respond to the same input pattern in a different way at different times, depending on the input sequence. The main disadvantage of RNNs is that they require substantially more connections, and more memory in simulation, than standard back-propagation networks. RNNs can yield good result because of the rough repetition of similar patterns present in exchange rate time series. These regular but subtle sequences can provide a beneficial forecast ability.

4.3. *Competitive*

4.3.1. *Fuzzy ARTMAP network*

A fuzzy ARTMAP network[3] is a fuzzy ART[2] network that adds a single output layer to generate an error signal to the fuzzy ART network that is made up of the input, complement and category layers. The addition of the output layer for the error signal transforms the network from an unsupervised network to a supervised network where the network learns from examples in which the real category is known.

4.3.2. *Modular*

Modular ANNs[20] essentially make use of multiple individual back-propagation networks (BPNs) that compete to learn different aspects of the problem. The networks use an expert gating mechanism to choose which of the BPNs (called a local expert) does best on a particular input observation, essentially assigning different regions of the data space to different local experts. The general idea is that the error at each local expert is weighted by its posterior probability (obtained as training takes place) that it was responsible for in the current output vector. The gating networks learns by trying to match its prior probabilities to the posterior probabilities found in each local expert.

MLP is used most frequently for exchange rate prediction, because it has an inherent capability of arbitrary input-output mapping. However, other types of ANNs are also used.

Tenti[48] perform tests with three variations of RNNs. The first architecture used (RNN1) has one hidden and one recurrent layer. The output layer is fed back into the hidden layer, by means of the recurrent layer, showing the resulting output of the previous pattern. In the second version (RNN2), similar to that of Fransconi *et al.*,[13] the hidden layer is fed back into itself through an extra layer of recurrent nodes. In the third version (RNN3), patterns are processed from the input layer through a recurrent layer of nodes, which holds the input layer's contents as they existed when previous patterns were trained, and then are fed back into input layer.

Leung *et al.*[25] examine GRNNs forecast ability and compare its performance with a variety of forecasting techniques, including the multi-layered feed-forward network.

Davis *et al.*[6] present a variety of neural networks forecasting models applied to Canadian–US exchange rate data. Networks such as back-propagation, modular, radial basis functions, linear vector quantization, fuzzy ARTMAP and genetic reinforcement learning are examined. It is important to note that they predict direction shifts on Canadian–US exchange rate data rather than absolute price. Different types of classification networks have characteristics that may prove effective for specific classification data.

The selection of ANNs architecture is an open problem. ANNs designers must use the constraints of the training data set and development cost for determination. We suggest practitioners to employ MLP which is relatively easy and costs less to implement.

5. The Integration of ANNs with Other Methods

The desire to further enhance the performance of neural network prediction has led to the development of hybrid systems that combine neural networks with other methods. The integration of ANNs with other technologies, such as wavelet analysis, genetic algorithm, or fuzzy logic can improve the applications of ANNs. Although each technology has its own strengths and weaknesses, these technologies are complementary. Weaknesses of one technology can be overcome by strengths of another by achieving a systematic effect. Such an effect can create results that are more efficient, productive, and effective than the sum of their parts.

Genetic algorithm (GA) is a class of probabilistic search techniques based on biological evolution. Each point in the solution space is coded as a binary string called a chromosome. For instance, the co-ordinate (10,5,3) is encoded as

$$\underbrace{1\ 0\ 1\ 0}_{10}\ \underbrace{0\ 1\ 0\ 1}_{5}\ \underbrace{0\ 0\ 1\ 1}_{3}$$

When a new generation exists, each member is ranked according to its fitness. From this, a new population must be created. Essentially this is a "survival of the fittest solution", and the members used for mating are chosen with a probability proportional to their fitness.

A technique called crossover is employed to maximize retention of the good points of the previous generation. This is analogous to biological mating in which a child may be superior to both parents if it inherits good genes from both parents. In the computing process, this is achieved by swapping corresponding bits in pairs of chromosomes according to a given crossover rate; for instance, the last three bits of one chromosome may be swapped with the last three bits of another chromosome.

If the population does not contain all of the traits needed to solve a problem, no amount of crossover will work. As a result, a single bit is flipped very infrequently. This is called mutation, and solves one of the problems of neural networks — that we arrive at local minima. Mutation provides a way out by preventing a bit converging on a single value throughout the entire population. Mutation must be kept to a minimum to prevent loss of good chromosomes.

The inclusion of GA search techniques was undertaken for two reasons. The first relates to the potential GA offer in terms of adaptiveness. The flexibility, robustness and simplicity that GA offers render them very attractive in that respect. The second reason stems from the difficulty in optimizing neural network applications. By operating on entire populations of candid solutions in parallel, GA is much less likely to get stuck at a local optimum.

Wavelet analysis is used to process information effectively at different scales. It is very useful for feature detection from complex and chaotic time series. In particular, the specific local properties of wavelets can be useful in describing the signals with discontinuous or fractal structures in the financial market. It also allows the removal of noise-dependent high frequencies, while conserving the signal bearing high frequency terms. However, one of the most critical issues in the application of the wavelet analysis is to choose the correct wavelet thresholding parameters.

El Shazly *et al.*[9] design a hybrid system combining neural networks with genetic training to forecast the three-month spot exchange rate. Once the network is trained, tested and identified as being "good", a GA is applied to it in order to optimize its performance. The process of genetic evolution works on the neuron connection of a trained network by applying two procedures: mutation and crossover. The application of hybrid systems seems to be well suited for the forecasting of financial data.

Shin *et al.*[42] propose an integrated thresholding design of the optimal or near-optimal wavelet transformation (WT) by GA to represent a significant signal most suitable in ANN models. The model is applied to forecast the Korean won/USD returns one day ahead of time. In this study, the multi-scale signal representation of ANNs is supported by a wavelet transform as the multi-signal decomposition technique to detect the features of significant patterns. A strategy is devised using WT to construct a filter that is significantly matched to the frequency of the time series within the combined model. The experimental results show the enhanced filtering or signal multi-resolution power of wavelet analysis by GA in the performance of the ANNs. This study also finds that the hybrid system of wavelet transformations and ANNs by GA is much better than other ANNs that use other three-wavelet thresholding algorithms (cross-validation, best level, and best basis) to increase forecasting performance.

6. Performance Comparison with Other Forecasting Methods

There are inconsistent reports on the performance of ANNs for forecasting exchange rates when compared with other forecasting methods. Table 1 summarizes the literature on the relative performance of ANNs.

Weigend *et al.*[53] find that neural networks are better than random walk models (RW) in predicting the DEM/USD exchange rate. Wei *et al.*[54] claim that ANNs' forecasting performance is better than those of AR(p), ARMA(p, q), ARIMA(p, d, q). Lisi *et al.*[26] make a comparison between ANNs and chaotic models in forecasting exchange rate prediction. ANNs perform slightly better than chaotic models, in term of NMSE; nevertheless, the two models are statistically equivalent. Yao and Tan[57] show that irrespective of NMSE, gradient or profit, ANNs are much better than traditional ARIMA model when forecasting the exchange rates between USD and five other major currencies, AUD, CHF, DEM, GBP and JPY. Leung *et al.*[25] point out that GRNNs generally outperform parametric multivariate transfer functions and the random walk models.

Episcopos and Davis[10] suggest that neural networks are similar to EGARCH, but superior to random walk models in terms of in-sample fit and out-of-sample prediction performance. Hann and Steurer[15] compared neural network models with linear monetary models in forecasting USD/DEM. Out-of-sample results show that, for weekly data, neural networks are much better than linear models and naïve predictions of a random walk model with regard to Theil's U measure, the hit rate, the annualized returns and the Sharp ratio. However, if monthly data are used, neural networks do not show much improvement over linear models. Monthly data usually contain more irregularities (seasonality, cyclicity, nonlinearity, noise).

Zhang and Hutchinson[62] find mixed results for neural networks in comparison with those from random walk models using different sections of the data set. Kuan and Liu[24] examine the out-of-sample forecasting ability of neural networks on five exchange rates against the USD, including GBP, CAD, DEM, JPY and CHF. For the GBP and JPY, they demonstrate that neural networks have significant market timing ability and/or achieve significantly lower out-of-sample RMSE than the random walk model across three testing periods. For the other three exchange rates, neural networks are not shown to be superior in forecasting performance. Their results also show that different network models perform quite differently in out-of-sample forecasting. Hu *et al.*[17] compare combining the performance of ANNs with those of various forecasting methods. Using different performance measurements and different data stages, they get different results. ANNs are not always better than other forecasting tools. Zhang and Hu[58] find that neural networks predict much better than random walk model when using large training samples. Small training samples will make ANNs fail to outperform the random walks for longer forecast horizons. They suggest possible structural changes in exchange rate data. Therefore, as more observations are available, they should be used to revise the forecasting neural networks models to better reflect change in the underlying pattern.

Table 1. The relative performance of ANNs with traditional forecasting methods.

Researchers	Data	ANNs Type	Traditional Forecasting Method	Performance Measure	Conclusions
Episcopos and Davis[10]	USD, DEM, FRF, JPY, GBP against CAD	MLP	EGARCH, RW	RMSE	Similar to EGARCH; Better than RW
Hann and Steurer[15]	DEM/USD	MLP	Linear model, RW	Theil's U measure, Hit rate, the annualized returns and the Sharp ratio	Better in weekly data; Similar in monthly data
Hu and Tsoukalas[17]	BEF/LUF, GBP, DKK, NLG, FRF, GRD, IEP, ITL, PTE, ESP, USD against DEM	MLP	MAV, GARCH, EGARCH, IGARCH, OLS, AVE	RMSE, MAE	Mixed results
Kuan and Liu[24]	GBP, CAD, DEM, JPY and CHF against USD	MLP, RNNs	RW	RMSE	Mixed results
Leung *et al.*[25]	GBP, JPY, CAD against USD	GRNNs	Multivariate transfer function, RW	MAE, RMSE	Better
Lisi and Schiavo[26]	FRF, DEM, ITL, GBP against USD	MLP	Chaotic model, RW	NMSE	Better
Wei and Jiang[54]	GBP/USD	MLP	AR, ARMA, ARIMA	RMSE	Better
Weigend *et al.*[53]	DEM/USD	MLP	RW	ARV	Better
Yao and Tan[57]	AUD, CHF, DEM, GBP, JPY against USD	MLP	ARIMA	NMSE, Correctness of gradient prediction	Better
Zhang and Hu[58]	GBP/USD	MLP	RW	RMSE, MAE, MAPE	Mixed results
Zhang and Hutchinson[62]	CHF/USD	MLP	RW	RMSE	Mixed results

MAE: mean absolute error.
RMSE: root mean square error.
NMSE: normalized mean square error.
MAPE: mean absolute percentage error.
ARV: average relative variance.

It is important to note that most studies use a single neural network model in modeling and predicting exchange rates. As data-dependent neural networks tend to be more unstable than traditional parametric models, performance of the keep-the-best (KTB) approach can vary dramatically with different models and data. Random variations resulting from data partitioning or subtle shifts in the parameters of the time series generating process can have a large effect on the learning and generalization capability of a single neural network model. These may be the reasons why neural networks perform differently for different exchange rate series and different time frequencies with different data partitions.

7. Conclusions

In this paper, we present a survey of forecasting exchange rates using artificial neural networks. ANNs have been shown to be a promising tool for forecasting financial time series. Several design factors significantly impact the accuracy of neural network forecasts. These factors include selection of input variables, preparing data, ANNs architecture. There is no consensus on these factors. In different cases, various decisions have their own effectiveness. There is no formal systematic model building approach. The integration of neural networks with other technologies is reported. We also discuss the relative performance of ANNs compared with other forecasting methods, finding mixed results.

Model uncertainty comes from three main sources: model structure, parameter estimation and data. The nonlinear nonparametric nature of ANNs may cause more uncertainties in model building. This learning and generalization dilemma or tradeoff has been extensive, and is still an active research topic in the field. To improve generalization performance of neural network models, we may need to go beyond the model selection methods. Efforts can be made along the lines of hint,[14] Bayesian regularization,[27] Vapnik-Chervonenkis dimension analysis,[5] and support vector machine.[35]

Neural network ensembles seem promising for improving predictions over the KTB approach because they do not solely rely on the performance of a single neural network model. Zhang *et al.*[60] examine three ensemble approaches. The first approach is to combine neural networks trained with different initial random weights but with the same data. The second approach is to combine different neural network architectures within an ensemble. The third approach is to combine networks trained with different sample data. Many other ensemble methods can be considered. For example, one potential method is based on bootstrapping samples randomly generated from the original whole training time series. While computationally expensive, ensemble models based on bootstrapping samples may provide further insights and evidence on the effectiveness of the ensemble method for out-of-sample forecasting. Research efforts should also be devoted to the methods that can further reduce the correlation effect in combining neural networks and to quantifying the impact that shifts in the data generation parameters have on the various approaches.

The exchange rates forecasted include Australian Dollar (AUD), Belgian/ Luxembourg Franc (BEF/LUF), British Pound (GBP), Canadian Dollar (CAD), Danish Krone (DKK), Deutsche Mark (DEM), Dutch Guilder (NLG), French Franc (FRF), Greek Drachma (GRD), Irish Punt (IEP), Italian Lira (ITL), Japanese Yen (JPY), Korean won, Portuguese Escudo (PTE), Russian rouble, Spanish Peseta (ESP), Swiss Franc (CHF) and US Dollar (USD). Among them, USD, GBP, JPY, DEM are forecast most frequently. Contrary to Diebold and Nason's[7] opinion that there is little variation in results from one exchange rate to another when nonparametric methods are used, results in one exchange rate cannot be generalized to another. At least, caution is still needed when generalizing.

While some studies have found encouraging results using this artificial intelligence technique to predict the movements of established financial markets, it is interesting to verify the persistence of this performance in the emerging markets. These rapidly growing financial markets are usually characterized by high volatility, relatively smaller capitalization, and less price efficiency, features which may hinder the effectiveness of those forecasting models developed for established markets. So future research should extend to other exchange rates.

There are only two studies[9,42] on the integration of ANNs and other technologies. Their research results show improvement in forecasting performance. Since the neural network is considered to have great potential as a powerful forecasting tool, its integration with other technologies should improve its overall performance.

For practitioners, trading driven by a certain forecast with a small forecast error may not be as profitable as trading guided by an accurate prediction of the direction of movement. Therefore, predicting the direction of change of foreign exchange rates and its return is also significant in the development of effective market trading strategies.

Future research should attempt to formulate a hybrid neural network model for forecasting as follows: the model integrates ANNs with more complementary technologies to enhance its self-adaptation to different situations. More statistical analyses should be provided in determining some parameters like sample size and frequency, etc. We need to find out which kind of data segments best capture the underlying behavior of market changes. Appropriate sample frequency should be investigated to provide enough information on underlying relationship in exchange rates as well as to limit noise incorporation.

Acknowledgement

This project is supported by NSFC, CAS and the City University of Hong Kong.

References

1. T. Bollerslev, Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72** (1990) 498–505.
2. G. A. Carpenter, S. Grossberg and J. H. Reynolds, ART-MAP: Supervised real time learning and classification of nonstationary data by a self-organizing neural network, *Neural Networks* **4** (1991) 565–588.
3. G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen, Fuzzy ARTMAP: An adaptive resonance architecture for incremental learning of analog maps, *IJCNN*, June 1992, Baltimore, Vol. III, pp. 309–314.
4. D. Chappel, J. Padmore, P. Mistry and C. Ellis, A threshold model for French franc/Deutsch mark exchange rate, *Journal of Forecasting* **15** (1996) 155–164.
5. V. Cherkassky, F. M. Mulier and X. Shao, Model complexity control for regression using VC generalization bounds, *IEEE Transactions on Neural Networks* **10** (1999) 1075–1089.
6. J. T. Davis, A. Episcopos and S. Wettimuny, Predicting direction shifts on Canadian-US exchange rates with artificial neural networks, *International Journal of Intelligent Systems in Accounting, Finance and Management* **10** (2001) 83–96.
7. F. X. Diebold and J. A. Nason, Nonparametric exchange rate prediction, *Journal of International Economics* **28** (1990) 315–322.
8. M. R. El Shazly and H. E. El Shazly, Comparing the forecasting performance of neural networks and forward exchange rates, *Journal of Multinational Financial Management* **7** (1997) 345–356.
9. M. R. El Shazly and H. E. El Shazly, Forecasting currency prices using a genetically evolved neural network architecture, *International Review of Financial Analysis* **8**(1) (1999) 67–82.
10. A. Episcopos and J. Davis, Predicting returns on Canadian exchange rates with artificial neural networks and EGARCHM-M model, *Neural Computing and Application* **4** (1996) 168–174.
11. H. Fang, K. S. Lai and M. Lai, Fractal structure in currency futures price dynamics, *Journal of Futures Markets* **14** (1994) 169–181.
12. L. Fausett, *Fundamentals of Neural Networks* (Prentice Hall, Englewood Cliffs, NJ, 1994).
13. P. Fransconi, M. Gori and G. Soda, Local feedback multilayered networks, *Neural Computation* **4** (1992) 120–130.
14. R. Garcia and R. Gencay, Predicting and hedging derivative securities with neural networks and a homogeneity hint, *Journal of Econometrics* **94** (2000) 93–115.
15. T. H. Hann and E. Steurer, Much ado about nothing? Exchange rate forecasting: Neural networks versus linear models using monthly and weekly data, *Neurocomputing* **10** (1996) 323–339.
16. D. A. Hsieh, Modeling heteroscedasticity in daily foreign-exchange rates, *Journal of Business and Economic Statistics* **7** (1989) 307–317.
17. M. Y. Hu and C. Tsoukalas, Combining conditional volatility forecasts using neural networks: An application to the EMS exchange rates, *Journal of International Financial Markets, Institutions and Money* **9** (1999) 407–422.
18. W. Huang, Y. Nakamori and S. Y. Wang, A general approach based on autocorrelation to determine input variables of neural networks for time series forecasting, submitted to *IEEE Transaction on System, Human and Cybernetics*, 2002a.
19. W. Huang, Y. Nakamori and S. Y. Wang, Using change-point detection to seek optimal training set for neural networks in foreign exchange rates forecasting,

submitted to *International Journal of Computational Intelligence and Organization*, 2002b.

20. R. A. Jacobs, M. I. Jordan, S. J. Nowlan and G. E. Hinton, Adaptive mixtures of local experts, *Neural Computation* **3** (1991) 79–87.
21. A. M. Jamal and C. Sundar, Modeling exchange rates with neural networks, *Journal of Applied Business Research* **14**(1) (1998) 1–5.
22. S. Kang, An investigation of the use of feedforward neural network for forecasting, PhD Thesis, Kent State University, Kent, OH 44240, USA, (1991).
23. T. Kohonen, Learning vector quantization, *Neural Networks* **1**(suppl 1) (1988) p. 303.
24. C. M. Kuan and T. Liu, Forecasting exchange rates using feedforward and recurrent neural networks, *Journal of Applied Econometrics* **10** (1995) 347–364.
25. M. T. Leung, A. S. Chen and H. Dauk, Forecasting exchange rate using general regression neural networks, *Computer and Operations Research* **27** (2000) 1093–1110.
26. F. Lisi and R. A. Schiavo, A comparison between neural networks and chaotic models for exchange rate prediction, *Computational Statistical and Data Analysis* **30** (1999) 87–102.
27. D. J. C. Mackay, Bayesian interpolation, *Neural Computation* **4** (1992) 415–447.
28. R. A. Meese and A. Rose, An empirical assessment of non-linearities in models of exchange rate determination, *Review of Economic Studies* **58** (1991) 601–619.
29. B. Mizrach, Multivariate nearest-neighbor forecasts of EMS exchange rates, *Journal of Applied Econometrics* **7** (1992) S151–S164.
30. J. Moody and C. J. Darken, Fast learning in networks of locally tuned processing units, *Neural Computation* **1** (1989) 281–294.
31. D. A. Peel and P. Yadav, The time series behavior of spot exchange rates in the German hyper-inflation period: Was the process chaotic? *Empirical Economics* **20** (1995) 455–463.
32. H. Pi, Dependence analysis and neural network modeling of currency exchange rates, *Proceedings of First International Workshop on Neural Networks in Capital Market*, London, 1993.
33. A. Podding, Short term forecasting of the USD/DEM exchange rate, in *Proceedings of First International Workshop on Neural Networks in Capital Market*, London, 1993.
34. M. Qi and G. Zhang, An investigation of model selection criteria for neural network time series forecasting, *European Journal of Operational Research* **132** (2001) 666–680.
35. S. Raudys, How good are support vector machine (should be "machines?"), *Neural Networks* **13** (2000) 17–19.
36. A. N. Refenes, Constructive learning and its application to currency exchange rate forecasting, in *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, eds. R. R. Trippi and E. Turban (Probus Publishing Company, Chicago, 1993), pp. 777–805.
37. A. N. Refenes, M. Azema-Barac, L. Chen and S. A. Karoussos, Currency exchange rate prediction and neural network design strategies, *Neural Computing and Application* **1** (1993) 46–58.
38. D. E. Rumelhart, G. E. Hinto and R. J. Williams, Learning internal representation by error propagation, in *Parrallel Distributed Processing*, eds. D. E. Rumelhart and J. L. McClelland, Vol. 1 (MIT Press, Cambridge, MA, 1986), pp. 318–362.
39. N. Sarantis and C. Stewart, Monetary and asset market models for sterling exchange rates: A cointegration approach, *Journal of Economic Integration* **10** (1995) 335–371.
40. M. Shanker, M. Y. Hu and M. S. Hung, Effect of data standardization on neural network training, *Omega* **24**(4) (1996) 385–397.

41. S. H. Shin, Essays on asset return predictability, PhD Dissertation, Massachusetts Institute of Technology, (0753), 1993.
42. T. Shin and I. Han, Optimal signal multi-resolution by genetic algorithm to support artificial neural networks for exchange–rate forecasting, *Expert Systems with Application* **18** (2000) 257–269.
43. M. Smith, *Neural Networks for Statistical Modeling* (Van Nostrand Reinhold, New York, 1993).
44. M. K. P. So, K. Lam and W. K. Li, Forecasting exchange rate volatility using autoregressive random variance model, *Applied Financial Economics* **9** (1999) 583–591.
45. D. Specht, A general regression neural network, *IEEE Transactions on Neural Networks* **2** (1991) 568–576.
46. E. Steurer, Nonlinear modelling of the DEM/USD exchange rate, in *Neural Networks in the Capital Markets*, ed. A. Refenes (Wiley, New York), pp. 199–211.
47. A. Tahai, S. Walczak and J. T. Rigsby, Improving artificial neural network performance through input variable selection, in *Applications of Fuzzy Sets and the Theory of Evidence to Accounting II*, eds. P. Siegel, K. Omer, A. deKorvin and A. Zebda (JAI Press, Stamford, CN, 1998), pp. 277–292.
48. P. Tenti, Forecasting foreign exchange rates using recurrent neural networks, *Applied Artificial Intelligence* **10** (1996) 567–581.
49. P. Theodossiou, The stochastic properties of major Canadian exchange rates, *The Financial Review* **29**(2) (1994) 193–221.
50. S. Walczak, An empirical analysis of data requirements for financial forecasting with neural networks, *Journal of Management Information System* **17**(4) (2001) 203–222.
51. S. Walczak and N. Cerpa, Heuristic principles for the design of artificial neural networks, *Information and Software Technology* **41**(2) (1999) 107–117.
52. S. Walczak, A. Tahai and K. Karim, Improved cash flows using neural network models for forecasting foreign exchange rates, in *Applications of Fuzzy Sets and the Theory of Evidence to Accounting II*, eds. P. Siegel, K. Omer, A. deKorvin and A. Zebda (JAI Press, Stamford, CN, 1998), pp. 293–310.
53. A. S. Weigend, B. A. Huberman and D. E. Rumelhart, Predicting sunspots and exchange rates with connectionist networks, in *Nonlinear Modeling and Forecasting*, eds. M. Casdagli and S. Eubank (Addison-Wesley, Redwood City, CA, 1992), pp. 395–432.
54. W. X. Wei and Z. H. Jiang, Artificial neural network forecasting model for exchange rate and empirical analysis, *Forecasting* **2** (1995) 67–69.
55. B. Wu, Model-free forecasting for nonlinear time series (with application to exchange rates), *Computational Statistics and Data Analysis* **19** (1995) 433–459.
56. J. T. Yao, H. L. Poh and T. Jasic, Foreign exchange rates forecasting with neural networks, in *Proceedings of the International Conference on Neural Information Processing*, Hong Kong, 1996, pp. 754–759.
57. J. T. Yao and C. L. Tan, A case study on using neural networks to perform technical forecasting of forex, *Neurocomputing* **34** (2000) 79–98.
58. G. Zhang and M. Y. Hu, Neural network forecasting of the British Pound/US Dollar exchange rate, *Journal of Management Science* **26**(4) (1998) 495–506.
59. G. Zhang, B. E. Patuwo and M. Y. Hu, Forecasting with artificial neural networks: The state of the art, *International Journal of Forecasting* **14** (1998) 35–62.
60. G. P. Zhang and V. L. Berardi, Time series forecasting with neural network ensembles: An application for exchange rate prediction, *Journal of the Operational Research Society* **52** (2001) 652–664.

61. X. R. Zhang, Non-linear predictive models for intra day foreign exchange trading, *International Journal of Intelligent System in Accounting, Finance and Management* **3** (1994) 293–302.
62. X. Zhang and J. Hutchinson, Simple architectures on fast machines: Practical issues in nonlinear time series prediction, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, eds. A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Reading, MA, 1994), pp. 219–241.